

## Measurement Article

# Shades of Blue and Gray: A Comparison of the Center for Epidemiologic Studies Depression Scale and the Composite International Diagnostic Interview for Assessment of Depression Syndrome in Later Life

Linh Dang, MPH,<sup>1</sup> Liming Dong, PhD,<sup>1,2</sup> and Briana Mezuk, PhD<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Epidemiology and <sup>2</sup>Center for Social Epidemiology and Population Health, University of Michigan School of Public Health, Ann Arbor. <sup>3</sup>Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor.

\*Address correspondence to: Briana Mezuk, PhD, Center for Social Epidemiology and Population Health, Department of Epidemiology, University of Michigan School of Public Health, Suite 2649B, 1415 Washington Heights, Ann Arbor, MI 48109. E-mail: [bmezuk@umich.edu](mailto:bmezuk@umich.edu)

Received: October 11, 2018; Editorial Decision Date: March 15, 2019

**Decision Editor:** Suzanne Meeks, PhD

## Abstract

**Background and Objectives:** Psychiatric research lacks the equivalent of a thermometer, that is, a tool that accurately measures mental disorder regardless of context. Instead, the psychometric properties of scales that purport to assess psychopathology must be continuously evaluated. To that end, this study evaluated the diagnostic agreement between the eight-item Center for Epidemiologic Studies Depression Scale (CESD-8) and the Composite International Diagnostic Interview—short form (CIDI-SF) in the Health and Retirement Study (HRS).

**Research Design and Methods:** Data come from 17,613 respondents aged >50 from the 2014 wave of the HRS. Kappa coefficients were used to assess the agreement between the 2 instruments on depression classification across a range of thresholds for identifying case status, including variation across subgroups defined by age, race/ethnicity, and gender.

**Results:** The point prevalence of depression syndrome estimated by the CESD was higher than that estimated by the CIDI-SF (CESD: 9.9%–19.5% depending on the cutoff applied to the CESD vs CIDI-SF: 7.7%). Assuming CIDI-SF as the gold standard, the CESD yielded a sensitivity of 56.2%–70.2% and specificity of 84.7%–94.0% across the range of cutoffs. The agreement on depression classification was weak ( $\kappa = 0.32$ – $0.44$ ).

**Discussion and Implications:** Depression cases identified by the CESD have poor agreement with those identified by the CIDI-SF. Conceptually, psychological distress as measured by the CESD is not interchangeable with depression syndrome as measured by the CIDI-SF. Population estimates of depression among older adults based on the CESD should be interpreted with caution.

**Keywords:** CES-D, Depression classification, Validity, Reliability

In psychiatric epidemiology, case identification is important for generating valid and comparable estimates of mental disorders in a population. The meaning of these estimates, whether for service planning or for identifying risk or protective factors, is a function of the validity and reliability

of the case identification tools (Blum, 1962; Eaton, Hall, Macdonald, & Mckibben, 2007; Weissman, Sholomskas, Pottenger, Prusoff, & Locke, 1977; Williams, Tarnopolsky, & Hand, 1980). Self-report questionnaires or structured interviews administered by lay persons are often used to

assess common mental disorders, such as depression syndrome, in population-based research because they are less burdensome and costly than clinical psychiatric interviews (Blum, 1962; Eaton, Mojtabai, Stuart, Leoutsakos, & Kuramoto, 2012; Levine, 2013; Weissman et al., 1977).

The Center for Epidemiologic Studies Depression Scale (CESD) is among the most widely used tools to assess depression syndrome in population-based research (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004; Steffick, 2000). The original CESD is a 20-item instrument assessing depressed affect, positive affect, somatic complaints, and interpersonal relations over a 1-week period (Eaton et al., 2004; Radloff, 1977; Steffick, 2000; Weissman et al., 1977). Despite its popularity in research, the CESD has major limitations. Although the scale was originally designed as a continuous measure of depressive symptomatology (Eaton et al., 2004; Radloff, 1977; Steffick, 2000), it is commonly used for depression case identification based on dichotomized threshold scores (Polsky et al., 2005; Steffick, 2000; Suglia, Demmer, Wahi, Keyes, & Koenen, 2016; Zich, Attkisson, & Greenfield, 1990), and the optimal cutoff score remains unclear (Dozeman et al., 2011; Henry, Grant, & Cropsey, 2018; Poulin, Hand, & Boudreau, 2005; Radloff, 1977; Schein & Koenig, 1997; Steffick, 2000; Vilagut, Forero, Barbaglia, & Alonso, 2016; Zich et al., 1990). Most importantly, the validity of the CESD as a tool for identifying cases of depression syndrome, and differences in the prevalence of depression across population subgroups, is in question.

### Brief History and Overview of Recent Validation Studies of the CESD

The CESD was developed in the Community Mental Health Epidemiology Program (CMH) which consisted of a two sites: Washington County, Maryland (MD) and Kansas City, Missouri (MO; Comstock & Helsing, 1977). The article describing the psychometric properties of the CESD has been cited more than 40,000 times to date (Radloff, 1977), in which the author stated:

The scale was designed for use in studies of the relationships between depression and other variables across population subgroups. To compare results from one subgroup to another, the scale must be shown to measure the same thing in both groups. Therefore, it will be shown that properties of the scale (validity, reliability, factor structure) are similar for the various population subgroups to be studied. (Radloff, 1977, p. 386)

To that end, Radloff's analysis examined the reliability and criterion validity of the CESD in terms of "age, sex, race, and education[al]," to provide a rationale for using the CESD to make meaningful comparisons across these subgroups (Radloff, 1977).

In this validation study, the CESD was administered at the two CMH sites with a total sample size of 3,845

( $n = 295$  of whom were black; all black participants came from the Kansas City site). In terms of reliability, Radloff concluded: "Test-retest correlations were moderate (.40 or above) in all but three groups (Blacks, age under 25, and "need help" [for mental distress])" (Radloff, 1977, p. 400). To assess validity, "true" cases of depression were identified from two psychiatric facilities, one in Washington County MD ( $n = 70$ ) and another in New Haven, Connecticut ( $n = 35$ ); the diagnosis of these "true" cases was determined by psychiatric assessments and then compared with the CESD, which they were also administered. The racial/ethnic composition of these clinical samples are not described, nor is there discussion of whether the validity of the CESD varied by race.

This brief overview indicates that the CMH study was not well-designed for examining the CESD as a case-identification tool in the general population because the sample was insufficient for drawing conclusions about the reliability or validity of this measure across racial groups. Moreover, the CMH results indicated the CESD did not have adequately test-retest reliability across age and race. Indeed, in their comprehensive review of case-identification tools for depression, Eaton and colleagues (2007) identified no interviewer-blinded, English-language validation studies with a sample size greater than 100 that compared the CESD to diagnosis made by a clinical psychiatric interview (Eaton et al., 2007). Several recent investigations have cautioned against the use of the CESD as a stand-alone case-identification tool for depression syndrome (Eaton et al., 2004; Moon et al., 2017; Vilagut et al., 2016), particularly for persons with medical comorbidity, adolescents, and nonwhite minorities (Kim, Chiriboga, & Jang, 2009; Lu, Lindsey, Irsheid, & Nebbitt, 2017; Perreira, Deeb-Sossa, Harris, & Bollen, 2005; Schein & Koenig, 1997; Yang & Jones, 2008).

Finally, little is known about how the psychometric properties of CESD compare to other depression case-identification instruments in the general adult population. Very often the CESD is the *only* measure of mental health included in large population-based health surveys (e.g., National Longitudinal Study of Adolescent to Adult Health, Coronary Artery Development in Young Adults, and Multi-Ethnic Study of Atherosclerosis), and thus such comparisons are not possible.

### Present Study

The goal of this study is to compare the psychometric properties of the CESD as a function of age, gender, and race/ethnicity with a criterion measure of major depression in the Health and Retirement Study (HRS), a nationally representative survey of U.S. adults aged >50. The HRS is unique in that it includes both an 8-item version of CESD and the short-form version of the Composite International Diagnostic Interview (CIDI-SF) module for major depression. This CIDI-SF module is derived from *Diagnostic and*

*Statistical Manual of Mental Disorders (DSM)* criteria and has been extensively validated against clinical psychiatric interviews (i.e., the Structured Clinical Interview for DSM [SCID], Schedules for Clinical Assessment in Neuropsychiatry [SCAN]) in both white and nonwhite populations (Alegria et al., 2009; Andrews, Peters, Guzman, & Bird, 1995; Haro et al., 2006; Jackson, Neighbors, Nesse, Trierweiler, & Torres, 2004; Kessler et al., 2004, 2003; Wittchen, 1994). As a result, the HRS provides an opportunity for comparing the agreement between the CESD-8 and the CIDI-SF, and to assess whether these two instruments perform similarly as depression case-identification tools.

## Methods

### Study Population

The HRS is a longitudinal, nationally representative survey of U.S. adults aged more than 50 years old which oversamples Hispanics, African Americans, and Florida residents. Since 1992, approximately 20,000 respondents are interviewed biennially to assess a range of psychosocial characteristics, health history and health care utilization, and economic circumstances and employment history. Additional details of the study design and sampling method are described elsewhere (HRS, 2008). The HRS is approved by the Institutional Review Board at the University of Michigan (Weir, 2017).

The analytic sample for the present study consisted of 17,613 nonproxy respondents who completed the 2014 HRS Core interview and had complete data on the CESD-8 and the CIDI-SF screening questions. Respondents with missing data on these two assessments were excluded ( $n = 1,135$ ). Those excluded were older, more likely to be male and nonwhite, had lower socioeconomic status, were less likely to be married, had poorer memory, and were more likely to have history of psychiatric problems (all  $p < .05$ ; Supplementary Table 1).

### Measures

#### The Composite International Diagnostic Interview—short form

The Composite International Diagnostic Interview—short form (CIDI-SF) depression module is a fully-structured diagnostic interview based on the *DSM* criteria for major depressive episode (MDE; Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998; Steffick, 2000). It has been validated against semi-structured clinical psychiatric examinations, and has moderate concordance with diagnosis of MDE from these exams (Eaton et al., 2007; Eaton, Neufeld, Chen, & Cai, 2000; Jackson et al., 2004). The CIDI-SF depression module assesses eight symptom domains of MDE (i.e., sadness, anhedonia/lost interest, sleeping disturbances, appetite changes, fatigue, guilt/worthlessness, concentration problems, thinking about death) experienced over the past 12 months, as shown in Supplementary Table 2 (Kessler

et al., 1998; Steffick, 2000). Psychomotor symptoms (e.g., agitation, retardation), while part of the *DSM* criteria for MDE, are not assessed in this module; however, these are also the least common MDE symptoms in general population samples (Mezuk & Kendler, 2012).

To reduce respondent burden, the full CIDI depression module is only administered to respondents endorse either of the two cardinal depressive symptoms—sadness or anhedonia—at sufficiently intensity. This skip-pattern is shown by Supplementary Figure 1. Respondents are first asked whether there was a period of more than 2 weeks when they had depressed mood or anhedonia over the past 12 months. If either item was endorsed, the respondent was further asked about the intensity and duration of the symptom(s) (Steffick, 2000). Respondents whose symptoms lasted most/all of the day and almost every/every day were then “screened in” and completed the full CIDI module.

We constructed a summary score of the CIDI-SF as follows. For respondents who were screened out of the CIDI-SF (i.e., endorsed neither sadness nor dysphoria at sufficient intensity), a score of zero symptoms was assigned. For respondents who were screened in, a summary score was generated by counting the number of endorsed symptoms. This score ranged from one to eight, with higher score indicating more depressive symptoms. Respondents with a summary score of  $\geq 5$  were considered “cases” of past-year MDE, in accordance with *DSM* criteria (Steffick, 2000).

#### The Center for Epidemiologic Studies Depression Scale 8

The Center for Epidemiologic Studies Depression Scale 8 (CESD-8), in its entirety, was administered to all respondents. Respondents were asked about whether they experienced any of the eight symptoms listed in Supplementary Table 2 (e.g., felt depressed, felt everything was an effort) “much of the time during the past week,” each recorded as yes or no (Steffick, 2000). The last two items were reverse coded to generate a summary score (range: 0–8) with higher scores indicating more depressive symptoms.

Consistent with its widespread use as a case-identification tool, these CESD-8 summary scores were then dichotomized to identify “cases” of depression syndrome. However, there is not a consistent threshold score that is universally applied to indicate case status using the CESD-8. A score of  $\geq 3$  has been used in most studies of this instrument (Steffick, 2000). However, recent a study by Polsky and colleagues, using the HRS, applied a more stringent threshold of  $\geq 5$  (Polsky et al., 2005). Other studies also supported a use of more conservative thresholds for older adults (Dozeman et al., 2011; Lyness et al., 1997; Moon et al., 2017; Schein & Koenig, 1997; Schulberg et al., 1985; Steffick, 2000; Zich et al., 1990). Prior work in the HRS has identified that a threshold of  $\geq 4$  corresponds to the commonly used cutoff score of 16 in the original 20-item CESD (Steffick, 2000). Therefore, in this study, we explored three thresholds ( $\geq 3$ ,  $\geq 4$ , and  $\geq 5$ ) to indicate depression case status using the CESD-8.

### Other variables

Other variables included demographic characteristics, memory, and history of psychiatric problems, all factors associated with the likelihood of having depression among older adults (Clark, Aneshensel, Frerichs, & Morgan, 1981; Cole & Dendukuri, 2003; Cole, Kawachi, Maller, & Berkman, 2000; Karim, Weisz, Bibi, & ur Rehman, 2015; La Rue, Swan, & Carmelli, 1995; Perreira et al., 2005; Rabbitt, Donlan, Watson, McInnes, & Bent, 1995; Yang & Jones, 2008). Demographic variables included age (categorized as  $\leq 55$  years, 56–64, and  $\geq 65$  years), sex (male, female), race/ethnicity (white/Caucasian, black/African American, other), marital status (never married, currently married/partnered, widowed, divorced/separated), educational attainment (less than high school, high school, more than high school), employment status (currently working, unemployed, retired, other), and total wealth (reported in U.S. dollars and categorized into quantiles). The categories for age, years of education, and wealth were informed by the distribution of these variables in the HRS sample and selected to ensure adequate sample size for subgroup comparisons.

Memory was assessed using both immediate and 5-min delayed recall of 10 words (Bugliari et al., 2016). Both were treated as continuous measures ranging from 0 to 10. History of psychiatric problems was based on self-report of ever having been told by a doctor that they had “emotional, nervous, or psychiatric problems” (Bugliari et al., 2016).

### Statistical Analysis

Agreement on depression case status between the CIDI-SF and the CESD-8 was assessed by categorizing respondents into four mutually exclusive groups: (a) not a case on both instruments; (b) case according to CESD-8 but not CIDI-SF; (c) case according to CIDI-SF but not CESD-8; and (d) case on both instruments. We compared sample characteristics across these four groups, using chi-squared tests for categorical variables and Student's *t*-tests for continuous variables, and calculated Kappa coefficients ( $\kappa$ ) for the agreement on depression case status classification. We also compared the combined discordant groups (Groups b and c) with respondents classified as a case on both instruments (Group d).

We used cross-tabulations to calculate sensitivity, specificity, false positive rate, and false negative rate of the CESD-8 compared with the CIDI-SF, treating the latter as a temporary “gold standard.” We estimated the prevalence of depression syndrome based on the CIDI-SF and the CESD-8, using the cutoffs described above. We then examined the overall agreement on number of endorsed depressive symptoms between the CESD-8 and the CIDI-SF using cross tabulation, although given that these two instruments have different items (Supplementary Table 2) our main interpretation focuses on agreement in case identification rather than in number of symptoms. Finally, we conducted stratified analyses by age, sex, and race/ethnicity. All steps were repeated using the three cutoff scores of the CESD.

All analyses were conducted using SAS, version 9.4 (SAS Institute), accounting for the HRS survey design. All significance tests were evaluated two-sided at the level of  $p < .05$ .

### Results

Approximately 76% of the 17,613 respondents were considered noncases of depression by both the CESD-8 and the CIDI-SF; 6% were categorized as a case using both instruments. Approximately 17% was classified as a case by CESD-8 only, and 2% was classified as a case by CIDI-SF only (Table 1). Compared with respondents classified as cases by both instruments, those identified by CESD-8 only (using a threshold of  $\geq 3$  symptoms to indicate case status) were older, more likely to be male, more likely to be currently employed, had higher total wealth, poorer memory, and were less likely to ever have history of psychiatric problems (all  $p < .01$ ). Compared with respondents classified as cases by both instruments, those identified by CIDI-SF only were more likely to be non-Hispanic white, have higher education, be employed, had higher total wealth, had better memory, and were less likely to report a history of psychiatric problems (all  $p < .01$ ). We repeated these analyses using CESD cutoff scores of 4 and 5 (Supplementary Tables 3 and 4). Overall, results were consistent with those using threshold of  $\geq 3$ . The only exceptions are that with these higher thresholds respondents classified as having depression by CIDI-SF only were more likely to be married/partnered or widowed and that gender and wealth did not vary by depression case status.

### Agreement on Number of Depressive Symptoms

Table 2 presents the agreement between CESD-8 and CIDI-SF on number of depressive symptoms. The numbers on the diagonal represent the number of respondents who had the same number of symptoms on the two scales. Out of 17,613 respondents, 7,646 (43.4%) had zero symptom on both scales, and only 286 (1.6%) had exact agreement. Over 15% of respondents ( $n = 2,774$ ) with three or more CESD-8 symptoms did not meet the level of symptom intensity necessary to screen into the CIDI-SF module. That is, these respondents were classified as cases of depression based on CESD-8 criteria, whereas they did not even screen into the full CIDI-SF module. In contrast, among respondents who screened into the full CIDI-SF module, only 381 (2.2%) reported five to eight symptoms in CIDI-SF (i.e., met criteria as a case of MDE), but reported  $< 3$  symptoms in CESD-8.

### Agreement on Depression Case Classification

Prevalence of depression syndrome estimated by the CESD-8 was higher than that estimated by the CIDI-SF: CIDI-SF: 7.7% versus CESD-8: 9.9% (threshold  $\geq 5$ ), 13.5% (threshold  $\geq 4$ ), and 19.5% (threshold  $\geq 3$ ). This is



**Table 1.** Sample Characteristics by Depression Classification Based on the CIDI-SF and the CESD-8: Health and Retirement Study, 2014

Characteristics	Total	CIDI-SF			
		Depression <sup>a</sup> (N = 1,399)		No depression (N = 16,214)	
		CESD-8 depression <sup>a</sup> (N = 1,018)	CESD-8 no depression (N = 381)	CESD-8 depression <sup>a</sup> (N = 2,917)	CESD-8 no depression (N = 13,297)
Age, N (weighted %)					
≤55	2,215	192 (12.7)	72 (12.2)	359 (6.9)	1,592 (7.3)
56–64	5,885	465 (52.7)	142 (46.5)	1,005 (37.6)	4,273 (40.6)
≥65	9,513	361 (34.6)	167 (41.4)	1,553 (55.6)	7,432 (52.1)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> = .2 <sup>c</sup>	<i>p</i> <sub>b</sub> < .0001 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Sex, N (weighted %)					
Female	10,458	729 (67.3)	281 (67.5)	1,891 (60.8)	7,557 (52.5)
Male	7,151	289 (32.7)	100 (32.5)	1,025 (39.2)	5,737 (47.5)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> = 1.0 <sup>c</sup>	<i>p</i> <sub>b</sub> = .009 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Race/ethnicity, N (weighted %)					
White/Caucasian	12,533	673 (78.7)	297 (88.9)	1,844 (76.4)	9,719 (84.6)
Black/African American	3,483	212 (12.1)	56 (6.3)	751 (14.5)	2,464 (9.4)
Other	1,545	128 (9.2)	26 (4.8)	311 (9.1)	1,080 (6.0)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> = .003 <sup>c</sup>	<i>p</i> <sub>b</sub> = .4 <sup>d</sup>	<i>p</i> <sub>c</sub> = .002 <sup>e</sup>
Education, N (weighted %)					
Less than high school	3,462	290 (22.6)	61 (12.9)	930 (26.4)	2,181 (11.7)
High school	5,472	312 (32.0)	126 (31.3)	934 (31.3)	4,100 (29.4)
Other	8,592	410 (45.4)	193 (55.8)	1,039 (42.3)	6,950 (58.8)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> = .006 <sup>c</sup>	<i>p</i> <sub>b</sub> = .2 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Marital status, N (weighted %)					
Never married	883	66 (9.0)	22 (9.8)	195 (9.6)	600 (6.1)
Married/partnered	10,996	491 (48.6)	212 (57.3)	1,481 (49.6)	8,812 (67.9)
Widowed	3,181	213 (18.0)	82 (18.1)	705 (21.1)	2,181 (12.5)
Divorced/separated	2,550	247 (24.3)	64 (14.8)	536 (19.7)	1,703 (13.5)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> = .05 <sup>c</sup>	<i>p</i> <sub>b</sub> = .1 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Employment status, N (weighted %)					
Currently working	5,312	155 (14.8)	118 (35.0)	620 (23.7)	4,419 (37.2)
Unemployed	407	41 (3.8)	15 (2.1)	84 (2.2)	267 (1.7)
Retired	10,827	682 (67.1)	215 (55.1)	1,942 (67.4)	7,988 (57.3)
Other	1,067	140 (14.2)	33 (7.8)	271 (6.7)	623 (3.9)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> < .0001 <sup>c</sup>	<i>p</i> <sub>b</sub> < .0001 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Total wealth, N (weighted %)					
First quantile	4,341	482 (42.6)	95 (18.8)	1,130 (32.6)	2,634 (15.4)
Second quantile	4,410	264 (25.4)	106 (28.8)	750 (24.8)	3,290 (22.4)
Third quantile	4,431	160 (17.5)	94 (24.5)	602 (23.0)	3,575 (27.3)
Fourth quantile	4,431	112 (14.5)	86 (27.9)	435 (19.6)	3,798 (34.9)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> < .0001 <sup>c</sup>	<i>p</i> <sub>b</sub> < .0001 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Cognitive measures, mean (SE)					
No. words recall—immediate	5.32 (0.07)	5.85 (0.09)	5.00 (0.06)	5.69 (0.03)	
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> < .0001 <sup>c</sup>	<i>p</i> <sub>b</sub> = .0005 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
No. words recall—delayed		4.15 (0.08)	4.93 (0.10)	3.84 (0.06)	4.71 (0.03)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> < .0001 <sup>c</sup>	<i>p</i> <sub>b</sub> = .002 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Psychiatric problems, N (weighted %)					
Ever	3,531	724 (73.7)	183 (51.7)	983 (34.9)	1,641 (13.1)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> < .0001 <sup>c</sup>	<i>p</i> <sub>b</sub> < .0001 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>
Current	306	58 (5.5)	21 (6.3)	105 (3.5)	122 (0.9)
<i>p</i> Value	<i>p</i> < .0001 <sup>b</sup>	—	<i>p</i> <sub>a</sub> = .7 <sup>c</sup>	<i>p</i> <sub>b</sub> = .09 <sup>d</sup>	<i>p</i> <sub>c</sub> < .0001 <sup>e</sup>

Note: CESD-8 = eight-item Center for Epidemiologic Studies Depression Scale; CIDI-SF = Composite International Diagnostic Interview—short form; SE = standard error.

<sup>a</sup>Depression status defined as: CIDI-SF ≥ 5 or CESD-8 ≥ 3.

<sup>b</sup>*p* Value for difference across four depression status groups.

<sup>c</sup>*p*<sub>a</sub> value for difference between respondents who met the depression criteria on both the CIDI-SF and CESD-8 versus only CIDI-SF criteria.

<sup>d</sup>*p*<sub>b</sub> value for difference between respondents who met the depression criteria on both the CIDI-SF and CESD-8 versus only CESD-8 criteria.

<sup>e</sup>*p*<sub>c</sub> value for difference between respondents who met the depression criteria on both the CIDI-SF and CESD-8 versus neither CIDI-SF nor CESD-8 criteria.

**Table 2.** Agreement Between the CESD-8 and CIDI-SF in the Number of Symptoms Reported for Depressive Syndrome in the 2014 Health and Retirement Study

Number of depressive symptoms using CESD-8	Number of depressive symptoms using CIDI-SF, number of subjects									
	0	1	2	3	4	5	6	7	8	Total
0	7,646	2	8	22	48	41	49	37	15	7,868
1	3,796	2	5	13	24	30	47	37	15	3,969
2	1,686	2	5	16	22	26	36	34	14	1,841
3	1,047	1	2	9	25	21	41	32	16	1,194
4	627	1	6	9	17	18	35	29	14	756
5	440	1	5	16	9	24	39	41	34	609
6	371	0	2	6	13	36	60	67	43	598
7	199	0	1	1	14	29	58	89	74	465
8	90	0	1	0	4	16	40	82	80	313
Total	15,902	9	35	92	176	241	405	448	305	17,613

Note: CESD-8 = 8-item Center for Epidemiologic Studies Depression Scale; CIDI-SF = Composite International Diagnostic Interview—short form.

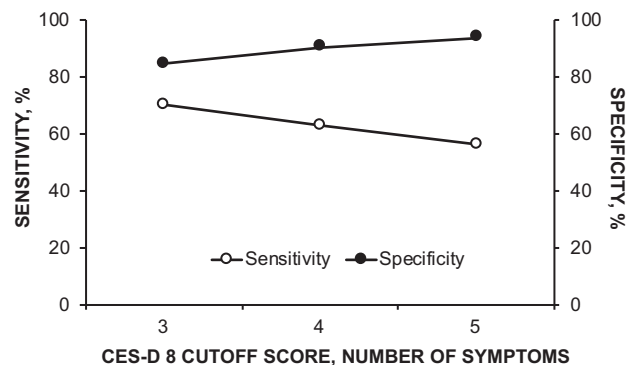
despite the fact that the CIDI-SF assesses symptoms over the past 12 months, whereas the CESD only asks about the past week. Temporary assuming MDE identified by the CIDI-SF as the “gold standard,” the CESD-8 had moderate sensitivity (ranging from 56.2% to 70.2%, depending on threshold) and high specificity (ranging from 84.7% to 94.0%, depending on threshold) at identifying cases. As expected, higher threshold scores resulted in lower sensitivity but higher specificity (Figure 1).

As illustrated in Table 3, the chance-corrected agreement on depression case classification between the CIDI-SF and the CESD-8 was in the poor to fair range ( $\kappa$  coefficients: 0.32–0.44). In addition, agreement between CIDI-SF screening status (a lower threshold than meeting full MDE criteria) and the CESD-8 indicated only fair agreement across the range of threshold scores ( $\kappa_{\text{CES-D} \geq 3} = 0.32$  [95% confidence level (CI): 0.30, 0.34];  $\kappa_{\text{CES-D} \geq 4} = 0.38$  [95% CI: 0.36, 0.40];  $\kappa_{\text{CES-D} \geq 5} = 0.42$  [95% CI: 0.40, 0.44]).

### Variation in Agreement by Age, Sex, and Race/Ethnicity

As shown in Figure 2, agreement on number of depressive symptoms between the two scales was poor/fair across age, sex, and race/ethnicity subgroups ( $\kappa$ : -0.0002 to 0.30). In general, among respondents with at least one symptom, Kappa increased as number of symptoms increased. Highest agreement was observed in respondents with eight symptoms ( $\kappa$  ranged from 0.23 to 0.30, depending on the subgroup). The agreement for one to five symptoms was poor for all subgroups (all  $\kappa < 0.1$ ). In general, agreement tended to be higher among respondents who were male, younger (aged  $\leq 55$ ) and non-Hispanic white (Figure 2).

Table 3 shows the agreement on depression case classification between the CESD-8 and the CIDI-SF was poor to fair for all subgroups (adjusted  $\kappa$  range: 0.23–0.56). The Kappa coefficients increased as the cutoff threshold for the CESD-8



**Figure 1.** Sensitivity and specificity for different cutoff scores of the CESD-8 to identify depression syndrome when compared with the CIDI-SF Major Depressive Episode, Health and Retirement Study, 2014. Sensitivity and specificity of the CESD-8 across three thresholds of symptom counts ( $\geq 3$ ,  $\geq 4$ , and  $\geq 5$ ) to define depression case status relative to MDE defined by the CIDI-SF.

increased. Agreement was similar across the subgroups defined by age, gender, and race/ethnicity, and was marginally higher for younger respondents, women, and non-Hispanic whites, regardless of the CESD-8 threshold applied.

### Discussion

The primary finding of this study is that CESD-8 and CIDI-SF have a poor agreement in the case identification of depression syndrome in the general older adult population. Using the CIDI-SF as the criterion standard, the CESD-8 had only moderate sensitivity (ranging from 56.2% to 70.2%), regardless of the threshold score. Kappa coefficients for agreement on the number of symptoms and depression case status were consistently in the poor to fair range. There was greater agreement among respondents at the two extremes of the measures (i.e., agreement was highest for those with zero or  $\geq 7$  symptoms). These findings were similar across

**Table 3.** Agreement Between CESD-8 and CIDI-SF Depression Status: Overall and Across Subgroups

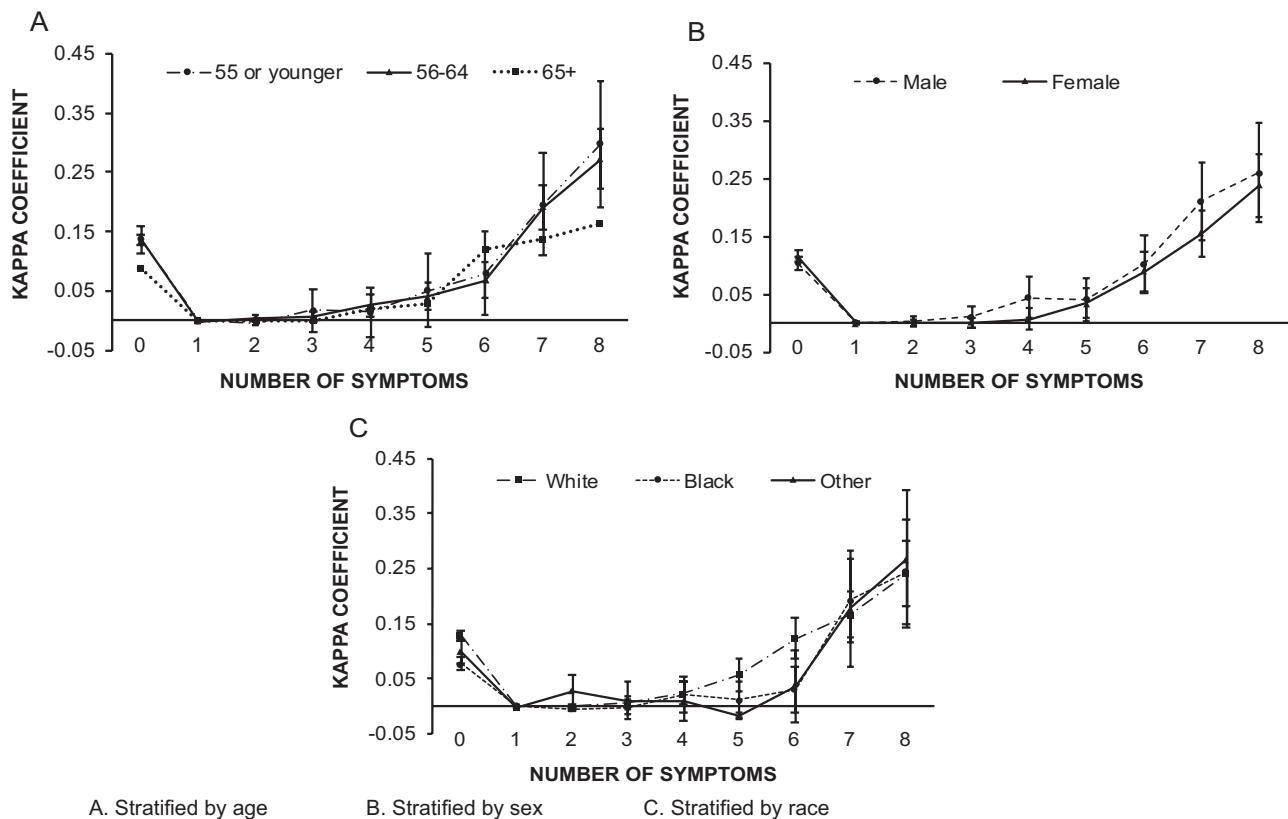
Characteristics	CIDI-SF screening status		Depression syndrome weighted %	Agreement between CIDI-SF and CESD-8			
	Screened in N (weighted %)	Screened out N (weighted %)		Kappa unadjusted	95% CI	Kappa adjusted <sup>a</sup>	95% CI
<b>CIDI-SF Scale<sup>b</sup></b>							
Mean score (SE)	5.94 (0.06)						
Participants with score ≥ 5	1,399 (81.5)		7.7				
<b>CESD-8 Scale<sup>c</sup></b>							
Mean score (SE)	4.12 (0.11)	1.07 (0.02)					
Participants with score ≥ 3	1,161 (65.1)	2,774 (14.8)	19.5	0.30	0.28, 0.32	0.32	0.29, 0.35
Participants with score ≥ 4	1,014 (57.4)	1,727 (9.0)	13.5	0.37	0.35, 0.39	0.40	0.36, 0.43
Participants with score ≥ 5	885 (49.9)	1,100 (5.7)	9.9	0.43	0.40, 0.45	0.44	0.40, 0.48
<b>CESD-8 Scale—by age<sup>c</sup></b>							
<b>≤55 years old</b>							
Mean score (SE)	4.36 (0.28)	1.06 (0.07)					
Participants with score ≥ 3	208 (66.5)	343 (14.6)	21.9	0.37	0.32, 0.41	0.43	0.33, 0.53
Participants with score ≥ 4	187 (63.8)	231 (8.5)	16.3	0.43	0.38, 0.48	0.54	0.45, 0.64
Participants with score ≥ 5	166 (58.2)	144 (6.0)	13.4	0.48	0.43, 0.53	0.56	0.46, 0.65
<b>56–64 years old</b>							
Mean score (SE)	4.31 (0.19)	1.02 (0.03)					
Participants with score ≥ 3	510 (68.2)	960 (13.9)	20.0	0.35	0.33, 0.38	0.39	0.35, 0.44
Participants with score ≥ 4	450 (60.5)	610 (8.5)	14.3	0.42	0.39, 0.45	0.46	0.40, 0.51
Participants with score ≥ 5	398 (53.1)	412 (5.7)	11.0	0.46	0.43, 0.50	0.49	0.43, 0.55
<b>≥65 years old</b>							
Mean score (SE)	3.82 (0.12)	1.10 (0.03)					
Participants with score ≥ 3	443 (61.0)	1,471 (15.6)	18.9	0.23	0.21, 0.25	0.23	0.20, 0.26
Participants with score ≥ 4	377 (52.0)	886 (9.4)	12.6	0.30	0.27, 0.33	0.30	0.26, 0.34
Participants with score ≥ 5	321 (43.8)	544 (5.8)	8.6	0.36	0.33, 0.39	0.35	0.31, 0.40
<b>CESD-8 Scale—by sex<sup>c</sup></b>							
<b>Female</b>							
Mean score (SE)	4.08 (0.11)	1.18 (0.02)					
Participants with score ≥ 3	808 (64.8)	1,812 (16.9)	22.4	0.30	0.28, 0.33	0.33	0.30, 0.36
Participants with score ≥ 4	701 (56.7)	1,149 (10.4)	15.7	0.37	0.34, 0.39	0.40	0.36, 0.43
Participants with score ≥ 5	615 (49.6)	742 (6.8)	11.7	0.42	0.39, 0.44	0.43	0.40, 0.47
<b>Male</b>							
Mean score (SE)	4.19 (0.23)	0.94 (0.03)					
Participants with score ≥ 3	353 (65.8)	961 (12.3)	16.2	0.28	0.25, 0.31	0.30	0.25, 0.36
Participants with score ≥ 4	313 (58.9)	577 (7.4)	11.1	0.37	0.34, 0.41	0.40	0.33, 0.46
Participants with score ≥ 5	270 (50.5)	357 (4.5)	7.8	0.44	0.40, 0.48	0.45	0.38, 0.53
<b>CESD-8 Scale—by race<sup>c</sup></b>							
<b>White/Caucasian</b>							
Mean score (SE)	3.97 (0.12)	0.98 (0.02)					
Participants with score ≥ 3	783 (62.8)	1,734 (13.4)	18.1	0.31	0.29, 0.33	0.33	0.30, 0.36
Participants with score ≥ 4	671 (54.7)	1,025 (7.9)	12.3	0.38	0.36, 0.41	0.40	0.37, 0.44
Participants with score ≥ 5	588 (47.4)	630 (4.9)	8.9	0.44	0.41, 0.47	0.45	0.41, 0.49
<b>Black/African Americans</b>							
Mean score (SE)	4.80 (0.26)	1.48 (0.05)					
Participants with score ≥ 3	236 (77.6)	727 (21.4)	26.4	0.25	0.22, 0.29	0.29	0.23, 0.35
Participants with score ≥ 4	215 (73.1)	477 (14.2)	19.4	0.33	0.29, 0.37	0.38	0.30, 0.46
Participants with score ≥ 5	183 (61.8)	328 (9.5)	14.1	0.38	0.33, 0.42	0.42	0.33, 0.51
<b>Other</b>							
Mean score (SE)	4.89 (0.32)	1.49 (0.08)					
Participants with score ≥ 3	136 (76.2)	303 (21.5)	27.3	0.33	0.28, 0.38	0.32	0.23, 0.41
Participants with score ≥ 4	122 (69.1)	218 (14.7)	20.4	0.39	0.33, 0.44	0.39	0.29, 0.49
Participants with score ≥ 5	108 (62.8)	136 (10.7)	16.1	0.45	0.39, 0.51	0.43	0.33, 0.54

Note: CESD-8 = 8-item Center for Epidemiologic Studies Depression Scale; CIDI-SF = Composite International Diagnostic Interview—short form; CI = confidence interval; SE = standard error.

<sup>a</sup>Adjusted for the HRS survey design and sample weights.

<sup>b</sup>Depression syndrome based on CIDI-SF criteria.

<sup>c</sup>Depression syndrome based on CESD-8 criteria.



**Figure 2.** Cohen's kappa coefficient comparing agreement in number of depressive symptoms across the CESD-8 and CIDI-DF by age (A), gender (B), and race/ethnicity (C) in the 2014 Health and Retirement Survey.

age, gender, and race/ethnicity, with slightly higher (but still relatively poor) agreement among respondents who were younger and male. Overall, our results indicate that depression case-identification measured by the CESD-8 is not interchangeable with case status measured by the CIDI-SF, and that this lack of concordance is not a function of demographic characteristics. This conclusion has implications for both descriptive and analytic epidemiology.

In terms of descriptive epidemiology, the discrepancies we identify here impact the credibility of estimates of the prevalence and distribution of depression in the population. Despite its widespread use as a case-identification tool, our findings provide little evidence that the CESD-8 is an appropriate metric of assessing depression burden in the general population. The CIDI-SF estimates of past-year depression are substantially lower than those estimated by the CESD-8, despite the fact that the CESD-8 only measures current symptomatology. Assuming the CIDI-SF as the "gold standard," the CESD had a high rate of false-positives, consistent with previously studies (Dozeman et al., 2011; Schein & Koenig, 1997; Schulberg et al., 1985; Zich et al., 1990). This makes it a poor stand-alone tool for case identification. If, however, the CESD were instead treated as a first-order screening instrument, one designed to identify individuals who would benefit from more intensive second-order diagnostic testing (say, with a clinician or a diagnostic instrument such as the CIDI-SF), this

high false-positive rate would be seen as a strength (Kessler et al., 2002). Indeed, this latter interpretation is more in line with the original writings regarding the utility of the CESD:

Because depression can be characterized as a feeling, a reaction, a syndrome, and an illness, a study of the distribution of the symptoms common to all these variants will obviously suffer from lack of specificity ... it seems reasonable to infer that persons with clinical depression are concentrated among persons with high levels of related symptoms, and that the epidemiologic pattern of clinical depression could be revealed, at least in broad outline, by the characteristics of persons with many or persistent symptoms of depression. Medical analogies are provided by hypertensive disease or diabetes mellitus. For both of these diseases, the epidemiologic pattern has been delineated not only by examining the distribution of persons with diagnosed disease but also by studying the distribution of one facet of disease, namely elevated blood pressure or blood sugar, respectively. (Comstock & Helsing, 1977, p. 559)

In sum, the intent and design of the measure (e.g., whether as a case-identification tool like the CIDI-SF or a screening assessment to capture population patterns of psychological distress) needs to be taken into consideration when interpreting the results of the descriptive epidemiology of depression assessments.



The implications for analytic epidemiology are complex. If the CESD-8 was simply a less precise measure of depression than the CIDI, we would expect that the effect estimates generated by studies that use it as a predictor to be biased toward the null. However, longitudinal studies examining depression as a predictor of subsequent health outcomes (e.g., diabetes, cardiovascular disease, mortality) generally report similar findings regardless of whether they use the CESD-8 or instruments like the CIDI (Demakakos, Pierce, & Hardy, 2010; Suglia et al., 2016; Vogelzangs et al., 2010; Womack et al., 2016). It is possible that the CESD is tapping into related constructs beyond depression, such as perceived stress or disengagement, which are also hypothesized to contribute to health outcomes in later life. Apathy, hopelessness, and social withdrawal are central to the concept of depression, particularly in later life, and are prominent in measures such as the Geriatric Depression Scale which includes several items to assess these thoughts and feelings (e.g., Do you feel that your life is empty? Do you often get bored? Have you dropped many of your activities and interests?; Adams, Matto, & Sanders, 2004). The CESD differs from the CIDI-SF in that it measures two positive elements of affect (i.e., felt happy, enjoyed life), which are then reverse-coded; however, positive and negative affect are not mutually exclusive (Diener & Emmons, 1984). While there are sound psychometric reasons for including reverse-coded items in scales, others have questioned this approach to measuring depression syndrome (Carleton et al., 2013). In sum, the CESD may be sufficient as a measure of “non-specific psychological distress” as a *risk factor for subsequent outcomes*, including depression syndrome. However, our findings indicate the CESD-8 is not appropriate for studies that seek to identify *correlates of depression* or how those correlates vary across subgroups. Finally, from an intervention and mental health services planning perspective, the clinical significance of depression case status as indexed by the CESD is simply unknown.

Findings should be interpreted considering study limitations. We acknowledge that even clinical assessments of depression status by psychiatrists do not agree completely, reflecting the challenge of case identification in psychiatry more generally (Blum, 1962; Eaton et al., 2007, 2012; Kessler, 2000; Williams et al., 1980). The CIDI-SF depression module has moderate agreement on MDE diagnosis compared with clinical psychiatric interview (Cooper, Peters, & Andrews, 1998; Eaton et al., 2000; Kessler et al., 1998; Wittchen, 1994), with Kappa coefficients in the range of 0.4–0.5 (Alegria et al., 2009; Brugha, Jenkins, Taub, Meltzer, & Bebbington, 2001; Kessler et al., 2004; Williams et al., 2007). Much of the discrepancy between instruments like the CIDI and psychiatric interviews comes from the latter identifying more symptomology (Eaton et al., 2000), that is, the CIDI is likely a conservative measure of MDE.

The study also has several strengths. First, this study examined the agreement of CESD-8 with the CIDI-SF specifically in the context of case identification in a general

population sample. The CESD is the most frequently used, and often the only index of mental health included, in large population cohorts. Our findings therefore have implications for a large body of literature. Second, to the best of our knowledge, this is the largest study to compare the CESD and CIDI-SF in the same cohort. A recent meta-analysis study found only 28 validation studies with both the CESD and other diagnostic instruments; only three of these studies had a sample size >1,000, and none were larger than 2,500 (Vilagut et al., 2016). Third, we examined how the performance of the CESD varied across demographic subgroups. The U.S. older adult is becoming increasingly diverse, and it is important that our measures to assessment mental health are appropriate for this population.

Mental health is an important and understudied component of overall population health, particularly as populations age. In the time since the CESD was first proposed, multiple advances have been made in survey methods for assessing psychiatric disorders in the general population, and the widespread reliance on the CESD in population-based cohorts does not reflect those advances. Large population-based cohorts and surveillance surveys should include multiple measures of mental health in the same way they include multiple measures of physical health. These measures should reflect the most reliable and validated metrics available to have confidence in the inferences drawn from these efforts.

## Supplementary Material

Supplementary data are available at *The Gerontologist* online.

## Funding

This study was supported by a grant from the National Institutes of Health (P30 AG015281), and the Michigan Center for Urban African American Aging Research.

## Conflict of Interest

None reported.

## References

- Adams, K. B., Matto, H. C., & Sanders, S. (2004). Confirmatory factor analysis of the Geriatric Depression Scale. *The Gerontologist*, *44*, 818–826. doi:10.1093/geront/44.6.818
- Alegria, M., Shrout, P. E., Torres, M., Lewis-Fernández, R., Abelson, J. M., Powell, M.,...Canino, G. (2009). Lessons learned from the clinical reappraisal study of the Composite International Diagnostic Interview with Latinos. *International Journal of Methods in Psychiatric Research*, *18*, 84–95. doi:10.1002/mp.280

- Andrews, G., Peters, L., Guzman, A. M., & Bird, K. (1995). A comparison of two structured diagnostic interviews: CIDI and SCAN. *The Australian and New Zealand Journal of Psychiatry*, *29*, 124–132. doi:10.3109/00048679509075901
- Blum, R. H. (1962). Case identification in psychiatric epidemiology: Methods and problems. *The Milbank Memorial Fund Quarterly*, *40*, 253–258. doi:10.2307/3348572
- Brugha, T. S., Jenkins, R., Taub, N., Meltzer, H., & Bebbington, P. E. (2001). A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine*, *31*, 1001–1013. doi:10.1017/S003329170100418
- Bugliari, D., Campbell, N., Chan, C., Hayden, O., Hurd, M., Main, R.,...St.Clair, P. (2016). *RAND HRS data documentation, Version P*. Santa Monica, CA: RAND Center for the Study of Aging. Retrieved from [http://hrsonline.isr.umich.edu/modules/meta/rand/randhrsp/randhrs\\_P.pdf](http://hrsonline.isr.umich.edu/modules/meta/rand/randhrsp/randhrs_P.pdf)
- Carleton, R. N., Thibodeau, M. A., Teale, M. J., Welch, P. G., Abrams, M. P., Robinson, T., & Asmundson, G. J. (2013). The Center for Epidemiologic Studies Depression Scale: A review with a theoretical and empirical examination of item content and factor structure. *PLoS ONE*, *8*, e58067. doi:10.1371/journal.pone.0058067
- Clark, V. A., Aneshensel, C. S., Frerichs, R. R., & Morgan, T. M. (1981). Analysis of effects of sex and age in response to items on the CES-D scale. *Psychiatry Research*, *5*, 171–181. doi:10.1016/0165-1781(81)90047-0
- Cole, M. G., & Dendukuri, N. (2003). Risk factors for depression among elderly community subjects: A systematic review and meta-analysis. *The American Journal of Psychiatry*, *160*, 1147–1156. doi:10.1176/appi.ajp.160.6.1147
- Cole, S. R., Kawachi, I., Maller, S. J., & Berkman, L. F. (2000). Test of item-response bias in the CES-D scale. experience from the New Haven EPESE study. *Journal of Clinical Epidemiology*, *53*, 285–289. doi:doi.org/10.1016/S0895-4356(99)00151-1
- Comstock, G. W., & Helsing, K. J. (1977). Symptoms of depression in two communities. *Psychological Medicine*, *6*, 551. doi:10.1017/S0033291700018171
- Cooper, L., Peters, L., & Andrews, G. (1998). Validity of the Composite International Diagnostic Interview (CIDI) psychosis module in a psychiatric setting. *Journal of Psychiatric Research*, *32*, 361–368. doi:doi.org/10.1016/S0022-3956(98)00021-1
- Demakakos, P., Pierce, M. B., & Hardy, R. (2010). Depressive symptoms and risk of type 2 diabetes in a national sample of middle-aged and older adults: The English longitudinal study of aging. *Diabetes Care*, *33*, 792–797. doi:10.2337/dc09-1663
- Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105–1117. doi:dx.doi.org/10.1037/0022-3514.47.5.1105
- Dozeman, E., van Schaik, D. J., van Marwijk, H. W., Stek, M. L., van der Horst, H. E., & Beekman, A. T. (2011). The Center for Epidemiological Studies Depression Scale (CES-D) is an adequate screening instrument for depressive and anxiety disorders in a very old population living in residential homes. *International Journal of Geriatric Psychiatry*, *26*, 239–246. doi:10.1002/gps.2519
- Eaton, W. W., Hall, A. L., Macdonald, R., & McKibben, J. (2007). Case identification in psychiatric epidemiology: A review. *International Review of Psychiatry (Abingdon, England)*, *19*, 497–507. doi:10.1080/09540260701564906
- Eaton, W. W., Mojtabai, R., Stuart, E. A., Leoutsakos, J.-M. S., & Kuramoto, S. J. (2012). *Assessment of distress, disorder, impairment, and need in the population*. New York, NY: Oxford University Press. Retrieved from <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195390445.001.0001/acprof-9780195390445-chapter-4>
- Eaton, W. W., Neufeld, K., Chen, L. S., & Cai, G. (2000). A comparison of self-report and clinical diagnostic interviews for depression: Diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore Epidemiologic Catchment Area follow-up. *Archives of General Psychiatry*, *57*, 217–222. doi:10.1001/archpsyc.57.3.217
- Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In Mark E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults* (Vol. 3, 3rd ed., pp. 363–377). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., de Girolamo, G., Guyer, M. E., Jin, R.,...Kessler, R. C. (2006). Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *International Journal of Methods in Psychiatric Research*, *15*, 167–180. doi:10.1002/mpr.196
- Henry, S. K., Grant, M. M., & Cropsey, K. L. (2018). Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample. *Journal of Affective Disorders*, *234*, 270–275. doi:10.1016/j.jad.2018.02.071
- HRS Staff. (2008). *HRS design history*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from <https://hrs.isr.umich.edu/sites/default/files/biblio/DesignHistory.pdf>
- Jackson, J. S., Neighbors, H. W., Nesse, R. M., Trierweiler, S. J., & Torres, M. (2004). Methodological innovations in the National Survey of American Life. *International Journal of Methods in Psychiatric Research*, *13*, 289–298. doi:doi.org/10.1002/mpr.182
- Karim, J., Weisz, R., Bibi, Z., & ur Rehman, S. (2015). Validation of the eight-item Center for Epidemiologic Studies Depression Scale (CES-D) among older adults. *Current Psychology*, *34*, 681–692. doi:10.1007/s12144-014-9281-y
- Kessler, R. C. (2000). Psychiatric epidemiology: Selected recent advances and future directions. *Bulletin of the World Health Organization*, *78*, 464–474.
- Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E.,...Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *International Journal of Methods in Psychiatric Research*, *13*, 122–139. doi:<https://doi.org/10.1002/mpr.169>
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T.,...Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in

- non-specific psychological distress. *Psychological Medicine*, 32, 959–976. doi:10.1017/S0033291702006074
- Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H.-U. (1998). The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *International Journal of Methods in Psychiatric Research*, 7, 171–185. doi:10.1002/mpr.47
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., ... Wang, P. S.; National Comorbidity Survey Replication. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, 289, 3095–3105. doi:10.1001/jama.289.23.3095
- Kim, G., Chiriboga, D. A., & Jang, Y. (2009). Cultural equivalence in depressive symptoms in older white, black, and Mexican-American adults. *Journal of the American Geriatrics Society*, 57, 790–796. doi:10.1111/j.1532-5415.2009.02188.x
- La Rue, A., Swan, G. E., & Carmelli, D. (1995). Cognition and depression in a cohort of aging men: Results from the Western Collaborative Group Study. *Psychology and Aging*, 10, 30–33. doi:10.1037/0882-7974.10.1.30
- Levine, S. Z. (2013). Evaluating the seven-item Center for Epidemiologic Studies Depression Scale short-form: A longitudinal U.S. community study. *Social Psychiatry and Psychiatric Epidemiology*, 48, 1519–1526. doi:10.1007/s00127-012-0650-2
- Lu, W., Lindsey, M. A., Irsheid, S., & Nebbitt, V. E. (2017). Psychometric properties of the CES-D among black adolescents in public housing. *Journal of the Society for Social Work and Research*, 8, 595–619. doi:10.1086/694791
- Lyness, J. M., Noel, T. K., Cox, C., King, D. A., Conwell, Y., & Caine, E. D. (1997). Screening for depression in elderly primary care patients. A comparison of the Center for Epidemiologic Studies-Depression Scale and the Geriatric Depression Scale. *Archives of Internal Medicine*, 157, 449–454. doi:10.1001/archinte.1997.00440250107012
- Mezuk, B., & Kendler, K. S. (2012). Examining variation in depressive symptoms over the life course: A latent class analysis. *Psychological Medicine*, 42, 2037–2046. doi:10.1017/S003329171200027X
- Moon, J. R., Huh, J., Song, J., Kang, I. S., Park, S. W., Chang, S. A., ... Jun, T. G. (2017). The Center for Epidemiologic Studies Depression Scale is an adequate screening instrument for depression and anxiety disorder in adults with congenital heart disease. *Health and Quality of Life Outcomes*, 15, 176. doi:10.1186/s12955-017-0747-0
- Perreira, K. M., Deeb-Sossa, N., Harris, K. M., & Bollen, K. A. (2005). What are we measuring? An evaluation of the CES-D across race/ethnicity and immigrant generation. *Social Forces*, 83, 1567–1601. doi:10.1353/sof.2005.0077
- Polsky, D., Doshi, J. A., Marcus, S., Oslin, D., Rothbard, A., Thomas, N., & Thompson, C. L. (2005). Long-term risk for depressive symptoms after a medical diagnosis. *Archives of Internal Medicine*, 165(11), 1260–1266. doi:10.1001/archinte.165.11.1260
- Poulin, C., Hand, D., & Boudreau, B. (2005). Validity of a 12-item version of the CES-D used in the National Longitudinal Study of Children and Youth. *Chronic Diseases in Canada*, 26, 65–72.
- Rabbitt, P., Donlan, C., Watson, P., McInnes, L., & Bent, N. (1995). Unique and interactive effects of depression, age, socioeconomic advantage, and gender on cognitive performance of normal healthy older people. *Psychology and Aging*, 10, 307–313. doi:10.1037/0882-7974.10.3.307
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401. doi:10.1177/014662167700100306
- Schein, R. L., & Koenig, H. G. (1997). The Center for Epidemiological Studies-Depression (CES-D) Scale: Assessment of depression in the medically ill elderly. *International Journal of Geriatric Psychiatry*, 12, 436–446. doi:10.1002/(SICI)1099-1166(199704)12:4<436::AID-GPS499>3.0.CO;2-M
- Schulberg, H. C., Saul, M., McClelland, M., Ganguli, M., Christy, W., & Frank, R. (1985). Assessing depression in primary medical and psychiatric practices. *Archives of General Psychiatry*, 42, 1164–1170. doi:10.1001/archpsyc.1985.01790350038008
- Steffick, D. (2000). *Documentation of affective functioning measures in the Health and Retirement Study*. Institute for Social Research, Ann Arbor, MI: University of Michigan. doi:10.7826/ISR-UM.06.585031.001.05.0005.2000
- Suglia, S. F., Demmer, R. T., Wahi, R., Keyes, K. M., & Koenen, K. C. (2016). Depressive symptoms during adolescence and young adulthood and the development of type 2 diabetes mellitus. *American Journal of Epidemiology*, 183, 269–276. doi:10.1093/aje/kwv149
- Vilagut, G., Forero, C. G., Barbaglia, G., & Alonso, J. (2016). Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): A systematic review with meta-analysis. *PLoS One*, 11, e0155431. doi:10.1371/journal.pone.0155431
- Vogelzangs, N., Seldenrijk, A., Beekman, A. T., van Hout, H. P., de Jonge, P., & Penninx, B. W. (2010). Cardiovascular disease in persons with depressive and anxiety disorders. *Journal of Affective Disorders*, 125, 241–248. doi:10.1016/j.jad.2010.02.112
- Weir, D. R. (2017). *HRS Institutional Review Board information*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from [https://hrs.isr.umich.edu/sites/default/files/biblio/HRS\\_IRB\\_Information%28web%29\\_08\\_2018.pdf](https://hrs.isr.umich.edu/sites/default/files/biblio/HRS_IRB_Information%28web%29_08_2018.pdf)
- Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology*, 106, 203–214. doi:10.1093/oxfordjournals.aje.a112455
- Williams, D. R., González, H. M., Neighbors, H., Nesse, R., Abelson, J. M., Sweetman, J., & Jackson, J. S. (2007). Prevalence and distribution of major depressive disorder in African Americans, Caribbean blacks, and non-Hispanic whites: Results from the National Survey of American Life. *Archives of General Psychiatry*, 64, 305–315. doi:10.1001/archpsyc.64.3.305
- Williams, P., Tarnopolsky, A., & Hand, D. (1980). Case definition and case identification in psychiatric epidemiology: Review and assessment. *Psychological Medicine*, 10, 101–114. doi:10.1017/S0033291700039635
- Wittchen, H. U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI):

- A critical review. *Journal of Psychiatric Research*, 28, 57–84. doi:10.1016/0022-3956(94)90036-1
- Womack, V. Y., De Chavez, P. J., Albrecht, S. S., Durant, N., Loucks, E. B., Puterman, E.,...Carnethon, M. R. (2016). A longitudinal relationship between depressive symptoms and development of metabolic syndrome: The coronary artery risk development in young adults study. *Psychosomatic Medicine*, 78, 867–873. doi:10.1097/PSY.0000000000000347
- Yang, F. M., & Jones, R. N. (2008). Measurement differences in depression: Chronic health-related and sociodemographic effects in older Americans. *Psychosomatic Medicine*, 70, 993–1004. doi:10.1097/PSY.0b013e31818ce4fa
- Zich, J. M., Attkisson, C. C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: The CES-D and the BDI. *International Journal of Psychiatry in Medicine*, 20, 259–277. doi:10.2190/LYKR-7VHP-YJEM-MKM2