Journal of Nursing Measurement, Volume 25, Number 3, 2017

Reliability and Validity of the Center for Epidemiologic Studies Depression Scale in a Population-Based Cohort of Middle-Aged U.S. Adults

Theodore D. Cosco, PhD

University of Oxford, United Kingdom

Matthew Prina, PhD

King's College London, United Kingdom

Brendon Stubbs, PhD

King's College London, United Kingdom
NHS Foundation Trust, London, United Kingdom

Yu-Tzu Wu, PhD

University of Exeter, United Kingdom

Background and Purpose: Globally, depressive symptoms are a leading contributor to years lived with disability. The Center for Epidemiological Studies–Depression (CES-D) scale has been used extensively to quantify depression; yet, its psychometric properties remain contentious. This study examined the reliability and factor structure of the CES-D in the MacArthur Foundation's Midlife in the United States Study (MIDUS), a nationally representative cohort study of noninstitutionalized, English-speaking adults aged 24–74 years. Methods: Internal consistency (Cronbach's alpha) and confirmatory factor analysis (CFA) were used to examine the reliability and factor structure of the CES-D. Results: There were 1,233 participants who were included in the analysis (mean age = 57.3 years [*SD* = 11.5], 56.7% female). Cronbach's alpha of .90 was observed. The 4-factor model had the best model fit. Conclusions: High internal consistency was demonstrated alongside a replication of the original 4-factor structure. Continued use of the CES-D in noninstitutionalized populations is warranted.

Keywords: depression; Center for Epidemiological Studies–Depression (CES-D) scale; psychometric; confirmatory factor analysis; internal consistency

The Center for Epidemiological Studies–Depression (CES-D) scale (Radloff, 1977) is one of the most popular psychometrics used to capture depression (Shafer, 2006). The CES-D is a short self-report scale developed from items used in previously developed depression scales with the intent to capture symptoms associated with

http://dx.doi.org/10.1891/1061-3749.25.3.476

depression in the general population (Radloff, 1977). Due in part to its comparability with other popular scales, for example, the Beck Depression Inventory (Podorefsky, McDonald-Dowdell, & Beardslee, 2001), and the fact it is freely available, the CES-D has grown in popularity since its development in the late 1970s. In addition, the CES-D has been used in a range of populations, including young people (Luyckx et al., 2016), older adults (Ermer & Proulx, 2016), those with physical health comorbidities (Quiñones, Markwardt, & Botoseneanu, 2016), and other mental health conditions (Cohen & Ryu, 2015).

Despite the widespread usage of the CES-D, the psychometric properties of the scale, particularly the latent structure, have come under increasing scrutiny (Callahan & Wolinsky, 1994; Schroevers, Sanderman, van Sonderen, & Ranchor, 2000). During the development of the scale, the authors originally posited a four-factor structure, capturing depressed affect (DA), for example, feeling blue or sad; somatic/vegetative factors (SV), for example, feeling bothered, sleep disturbances, loss of appetite; interpersonal (IP), for example, feeling disliked or lonely; and positive affect (PA), for example, feeling happy or hopeful (Radloff, 1977). These items were not, however, developed a priori to reflect contemporary diagnostic criteria (Carleton et al., 2013). The latent structure of a scale is an integral component of construct validity, that is, the ability of a metric to capture the intended phenomenon. If the latent structure of a scale cannot be replicated in a sample beyond the scope of a population from which the scale was developed, one cannot conclusively infer that the same phenomenon is being captured. Previous examinations of the latent structure of the CES-D have posited one-, two-, and three-factor alternative structures (Carleton et al., 2013); therefore, the four-factor structure of the CES-D must be replicated in other samples to conclude that symptoms of depression are being captured in the same way as in the development of the scale.

Many studies examining the latent structure of the CES-D have employed exploratory factor analytic techniques, such as principal component analysis (Shafer, 2006). Exploratory factor analysis techniques, as the name suggests, are intended to employ data-driven methods to identify the number of latent factors in the absence of a hypothesis. Criticism of the methods employed in exploratory techniques have been suggested with respect to the ability to extract the correct number of factors, for example, Kaiser criterion (Steger, 2006), in an unbiased manner, and with respect to rotation methods, for example, principal components with orthogonal rotation (Osborne, 2008). In instances where a metric has an established factor structure, confirmatory factor analysis (CFA) can be employed to test the fit of the data to an existing structure. Given that the CES-D has been developed under the assumption of a four-factor structure, CFA is a more appropriate method of testing the factor structure than exploratory methods. Furthermore, a recent meta-analysis suggests that CFA should be used to assert the underlying factor structure of the CES-D (Shafer, 2006). Previous psychometric validations of the CES-D have posited various factor structures, ranging from one to four factors (Carleton et al., 2013).

In addition to the construct validity of a scale, internal consistency or reliability, of a scale is important in establishing confidence in a scales' ability to consistently capture the same phenomena (Cronbach, 1951). Cronbach's alpha was developed to provide a means of measuring the internal consistency of a scale, that is, the degree to which all of the items in a scale capture the same construct (Cronbach, 1951).

Given the aforementioned uncertainties, this study examines the reliability of the CES-D and tests the hypothesis that the original CES-D factor structure is maintained in a representative population-based sample of middle-aged American adults.

METHODS

Study Population

The MacArthur Foundation's Midlife in the United States Study (MIDUS), is a nationally representative cohort study of 7,108 noninstitutionalized, English-speaking adults aged 24–74 years through a random digit dialing. Telephone interviews were conducted between 1995 and 1996 to collect extensive information on behavioral, social, psychological, and emotional factors related to age difference in physical and mental health. Full details of the study are available at http://midmac.med.harvard.edu/research.html.

The MIDUS follow-up wave was conducted between 2004 and 2009 through different research projects and focused on subsample of the original cohort. This study was based on 1,255 participants in the biomarker project, which focused on identifying biopsychosocial pathways to diverse health conditions. The sample included 1,054 people from the MIDUS surveys and 201 participants from MIDUS Milwaukee sample, a new sample of African Americans from Milwaukee, Wisconsin. In addition to a wide range of biomedical measures, this project included the CES-D to assess depressive symptoms.

This analysis conducts a secondary analysis of 1,233 participants from the MIDUS with complete CES-D data.

Measures

The CES-D is a short screening tool designed to evaluate the presence and severity of depressive symptomatology. In the initial development of the scale, items were chosen from existing, previously validated, depression scales, for example, Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Components capturing depressed mood, feelings of guilt, feelings of worthlessness, psychomotor retardation, loss of appetite, and sleep disturbances were chosen from literature on depressive symptomology at the time (Radloff, 1977). The frequency of 20 depressive symptoms is rated on a 4-point Likert scale (0–3): *rarely* or *none of the time*, *some* or *a little of time*, *occasionally* or *a moderate amount of time*, *all of the time*. The range of total score is between 0 and 60, with higher scores indicating greater levels of depression. A CES-D score $\geq$16 has been identified as an indicator of depression (Radloff, 1977). The original CES-D model includes four factors: DA, SV, PA, and IP. In its original psychometric evaluation, the CES-D was found to have acceptable levels of internal consistency, test–retest stability, and concurrent validity with other depression scales.

Analysis

Internal consistency reliability was investigated using Cronbach's alpha and a value of .70 or above indicates good reliability (Cronbach, 1951). CFA was used to examine the factor structure of CES-D. Following the methods outlined in Thombs, Hudson, Schieir, Taillefer, and Baron (2008) and Rhee et al. (1999), this study tested five potential factor structure models: a single-factor model (Model 1; Sheehan, Fifield, Reisine, & Tennen, 1995), a two-factor model including general depression and PA (Model 2; Gomez & McLaren, 2015), a three-factor model including PA, IP, and a combined depressive affect and SV factor (Model 3A; Guarnaccia, Angel, & Worobey, 1989), a second three-factor

model including IP, SV, a combined depressive affect and PA factor (Model 3B; Fifield & Reisine, 1992) and Radloff's four-factor model (Model 4; Radloff, 1977). All factor loadings were standardized. Goodness of fit for these five models were assessed using the following indices: the Tucker-Lewis index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). A good-fitting model should have TLI and CFI values greater than .95, RMSEA .06 or lower, and SRMR .08 or lower (de Vet, Terwee, Mokkink, & Knol, 2011). All analyses were conducted in Stata 14.0.

RESULTS

Among the 1,233 participants, the mean age was 54.5 years (SD = 11.7) with a range between 35 and 86 years. The percentage of women was 56.7%. Almost half of the sample smoked (47.4%), whereas just more than two thirds had hypertension (37.2%) and 43.0% had arthritis. Further self-reported participant demographics of the participants are presented in Table 1.

Reliability of Center for Epidemiological Studies–Depression Scale

The estimated Cronbach's alpha was .90 in the overall sample. For each factor, coefficient alpha was .86 in DA, .75 in SV, .79 in PA and .58 in IP.

Confirmatory Factor Analysis

Table 2 shows factor standardized factor loadings of different models and the indices of goodness of fit. The four-factor model had the best model fit when compared to other models, demonstrating the highest values for TLI and CFI and the lowest values of RMSEA and SRMR. All factor loadings in Model 4 were statistically significant, and the range was between .47 and .85. The strongest correlation was between DA and SV factors (.79), and the weakest was between PA and IP (.50; Table 3).

TABLE 1. Study Population (N = 1,233)

	Men (N = 534)	Women (N = 699)	Total (N = 1,233)
Age: M (SD)	55.1 (11.9)	54.0 (11.5)	54.5 (11.7)
Ever smoke: n (%)	281 (52.6)	303 (43.4)	584 (47.4)
Stroke: n (%)	23 (4.3)	29 (4.2)	52 (4.2)
Heart disease: n (%)	82 (15.4)	54 (7.8)	136 (11.1)
Hypertension: n (%)	188 (35.7)	266 (38.4)	454 (37.2)
Cholesterol problems: n (%)	245 (46.6)	270 (39.2)	515 (42.4)
Diabetes: n (%)	67 (12.6)	79 (11.3)	146 (11.9)
Cancer: n (%)	64 (12.1)	103 (14.7)	167 (13.6)
Arthritis: n (%)	206 (39.8)	307 (45.4)	513 (43.0)

TABLE 2. Standardized Factor Loadings of Center for Epidemiological Studies–Depression (CES-D) Scale Items and Goodness of Fit Indicators in Models Tested

CES-D Item	Model 1: One Factor	Model 2: Two Factors	Model 3A: Three Factors	Model 3B: Three Factors	Model 4: Four Factors
Blues	DEP = .78	DA + SV + IP = .80	DA + PA = .79	DA + SV = .80	DA = .81
	(.76, .81)	(.77, .82)	(.77, .82)	(.77, .82)	(.78, .83)
Depressed	DEP = .80	DA + SV + IP = .80	DA + PA = .81	DA + SV = .81	DA = .82
	(.77, .82)	(.78, .83)	(.78, .83)	(.78, .83)	(.80, .84)
Failure	DEP = .64	DA + SV + IP = .64	DA + PA = .65	DA + SV = .64	DA = .65
	(.61, .68)	(.61, .68)	(.61, .68)	(.60, .68)	(.61, .68)
Fearful	DEP = .51	DA + SV + IP = .52	DA + PA = .51	DA + SV = .52	DA = .52
	(.47, .56)	(.48, .56)	(.47, .56)	(.48, .56)	(.48, .57)
Lonely	DEP = .67	DA + SV + IP = .68	DA + PA = .67	DA + SV = .67	DA = .67
	(.64, .70)	(.64, .71)	(.63, .70)	(.64, .71)	(.64, .71)
Cry	DEP = .55	DA + SV + IP = .57	DA + PA = .56	DA + SV = .57	DA = .58
	(.51, .59)	(.52, .61)	(.51, .60)	(.53, .61)	(.54, .62)
Sad	DEP = .74	DA + SV + IP = .75	DA + PA = .76	DA + SV = .75	DA = .78
	(.71, .77)	(.72, .77)	(.73, .79)	(.72, .78)	(.75, .80)
Bothered	DEP = .48	DA + SV + IP = .48	SV = .51	DA + SV = .48	SV = .51
	(.43, .53)	(.44, .53)	(.47, .56)	(.43, .53)	(.47, .56)
Appetite	DEP = .45	DA + SV + IP = .46	SV = .48	DA + SV = .46	SV = .48
	(.41, .50)	(.41, .50)	(.43, .53)	(.41, .50)	(.43, .53)

Mind	DEP = .43	DA + SV + IP = .44	SV = .52	DA + SV = .44	SV = .52
	(.38, .48)	(.39, .49)	(.48, .57)	(.39, .49)	(.48, .57)
Effort	DEP = .54	DA + SV + IP = .54	SV = .65	DA + SV = .54	SV = .65
	(.50, .58)	(.50, .59)	(.61, .68)	(.50, .59)	(.61, .68)
Sleep	DEP = .39	DA + SV + IP = .40	SV = .47	DA + SV = .40	SV = .47
	(.35, .44)	(.35, .45)	(.42, .52)	(.35, .45)	(.42, .56)
Talked less	DEP = .46	DA + SV + IP = .47	SV = .51	DA + SV = .47	SV = .51
	(.41, .51)	(.42, .51)	(.46, .56)	(.42, .51)	(.46, .56)
Get going	DEP = .61	DA + SV + IP = .61	SV = .71	DA + SV = .61	SV = .71
	(.57, .65)	(.57, .65)	(.68, .75)	(.57, .65)	(.68, .75)
Good	DEP = .40	PA = .50	DA + PA = .40	PA = .50	PA = .49
	(.35, .45)	(.45, .54)	(.35, .45)	(.45, .54)	(.45, .54)
Hopeful	DEP = .47	PA = .61	DA + PA = .48	PA = .61	PA = .61
	(.43, .52)	(.57, .65)	(.44, .53)	(.57, .65)	(.57, .65)
Happy	DEP = .66	PA = .85	DA + PA = .67	PA = .85	PA = .85
	(.63, .69)	(.82, .87)	(.63, .70)	(.82, .87)	(.82, .87)
Enjoy	DEP = .63	PA = .81	DA + PA = .64	PA = .81	PA = .81
	(.59, .67)	(.78, .83)	(.60, .67)	(.78, .83)	(.78, .83)

(Continued)

TABLE 2. Standardized Factor Loadings of Center for Epidemiological Studies–Depression (CES-D) Scale Items and Goodness of Fit Indicators in Models Tested *(Continued)*

CES-D Item	Model 1: One Factor	Model 2: Two Factors	Model 3A: Three Factors	Model 3B: Three Factors	Model 4: Four Factors
Unfriendly	DEP = .41	DA + SV + IP = .42	IP = .57	IP = .57	IP = .57
	(.36, .46)	(.37, .47)	(.51, .62)	(.52, .62)	(.52, .63)
Disliked	DEP = .52	DA + SV + IP = .52	IP = .72	IP = .71	IP = .71
	(.47, .56)	(.47, .56)	(.66, .77)	(.66, .77)	(.66, .77)
			Goodness of Fit Indicators		
Chi-square (*df*)	1,671.4 (170)	1,030.1 (169)	1,307.7 (167)	949.6 (167)	669.8 (164)
p Value	<.001	<.001	<.001	<.001	<.001
CFI	.83	.90	.87	.91	.94
TLI	.81	.89	.85	.90	.93
RMSEA	.09	.06	.07	.06	.05
SRMR	.06	.04	.05	.04	.04

Note. DEP = Depression; DA = depressed affect; SV = somatic/vegetative factors; IP = interpersonal; PA = positive affect; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

TABLE 3. Correlation Between the Four Factors in Model 4 (With 95% Confidence Intervals)

	DA	SV	PA	IP
DA	—			
SV	.79*	—		
	[.76, .83]			
PA	.70*	.62*	—	
	[.66, .74]	[.57, .67]		
IP	.69*	.65*	.50*	—
	[.63, .74]	[.58, .71]	[.43, .57]	

Note. DA = depressed affect; SV = somatic/vegetative factors; PA = positive affect; IP = interpersonal.
**p* values < .001.

DISCUSSION

Our study demonstrates that the original four-factor model proposed by Radloff in the original development of the CES-D provided the best fitting model in this population-based cohort of mid- to later life adults in the United States. Furthermore, the CES-D demonstrated high overall reliability as well as acceptable reliability in the DA, SV, and PA factors. The IP factor did not demonstrate adequate internal consistency.

In the interpretation of these results, one must be cognizant of the strengths and limitations of the data and analysis methods. Data in the MIDUS were collected via telephone interview, which may have resulted in social-desirability bias and/or a reluctance to report depressive symptoms. As with any self-reported data, one must acknowledge these limitations in the interpretation of the results. In the analysis of these data, we have compared the factor structure of several previously hypothesized latent structures of the CES-D, to identify the best fitting model and provide evidence for or against the continued validity of the original structure using previously used methods (Rhee et al., 1999; Thombs et al., 2008). There are many more structures of the CES-D that were not tested in this study; however, the theoretical underpinnings, popularity, and prior demonstrations of the tested structures warranted the examination of these specific structures.

The internal consistency of the CES-D demonstrates adequate reliability. The overall scale demonstrates a high level of internal consistency, as do three of the four factors within the CES-D. Cronbach's alpha is directly affected by the length of a scale; if a scale is too short the value of alpha is reduced. This is because of the nature of the test, which measures the interrelatedness of the constituent components; if more components are present capturing the same construct, alpha is generally higher. Therefore, given that the IP factor has only two components, it is not surprising that alpha is low. Previous studies have demonstrated low IP alphas (Thombs et al., 2008) with concerns raised the validity of a 2-item scale (Carleton et al., 2013), concerns echoed in this study. These results suggest that the two IP factors are weakly related; however, within the context of the full CES-D, an alpha value well above the acceptable limit was demonstrated.

Results from the CFA suggest that the CES-D is capturing depressive symptomology as per the original development of the scale. This would indicate that the CES-D is still a relevant and valid measure of depression the U.S. population of adults and that its continued use is warranted. Similarly, the results from the reliability test suggest that the CES-D is consistently capturing depression and that this is a reliable depression scale.

Relevance to Nursing Practice, Education, or Research

The results from this analysis suggest that the CES-D demonstrates adequate psychometric robustness for continued use as a metric for quantifying depression. Through continued examination of the CES-D validation, studies are able to provide invaluable insights into the effectiveness of psychometrics. The literature surrounding the latent structure of the CES-D has been mixed, suggesting that there may be issues with psychometric robustness; however, results from this study suggest that in a sample of middle-aged noninstitutionalized Americans, the CES-D performs as it was intended in its development. The replication of the factor structure identified in the development of the scale suggests that this scale is capturing depression in the same manner as it was when the scale was originally developed. Furthermore, the high level of internal consistency suggests that the scale is reliably capturing depression. Continued use of the CES-D in noninstitutionalized populations is warranted. These results have important implications for the continued use of the CES-D, highlighting the high levels of reliability and validity demonstrated by this popular depression metric.

REFERENCES

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Callahan, C. M., & Wolinsky, F. D. (1994). The effect of gender and race on the measurement properties of the CES-D in older adults. *Medical Care*, *32*(4), 341–356.

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*(3), e58067. http://dx.doi.org/10.1371/journal.pone.0058067

Cohen, C. I., & Ryu, H. H. (2015). A longitudinal study of the outcome and associated factors of subsyndromal and syndromal depression in community-dwelling older adults with schizophrenia spectrum disorder. *The American Journal of Geriatric Psychiatry*, *23*(9), 925–933. http://dx.doi.org/10.1016/j.jagp.2014.06.011

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. http://dx.doi.org/10.1007/bf02310555

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge, United Kingdom: Cambridge University Press.

Ermer, A. E., & Proulx, C. M. (2016). Unforgiveness, depression, and health in later life: The protective factor of forgivingness. *Aging & Mental Health*, *20*(10), 1021–1034. http://dx.doi.org/10.1080/13607863.2015.1060942

Fifield, J., & Reisine, S. (1992). Characterizing the meaning of psychological distress in rheumatoid arthritis. *Arthritis Care and Research*, *5*(3), 184–191.

Gomez, R., & McLaren, S. (2015). The Center for Epidemiologic Studies Depression Scale: Support for a bifactor model with a dominant general factor and a specific factor for positive affect. *Assessment*, *22*(3), 351–360. http://dx.doi.org/10.1177/1073191114545357

Guarnaccia, P. J., Angel, R., & Worobey, J. L. (1989). The factor structure of the CES-D in the Hispanic Health and Nutrition Examination Survey: The influences of ethnicity, gender and language. *Social Science & Medicine*, *29*(1), 85–94. http://dx.doi.org/10.1016/0277-9536(89)90131-7

Luyckx, K., Rassart, J., Goossens, E., Apers, S., Oris, L., & Moons, P. (2016). Development and persistence of depressive symptoms in adolescents with CHD. *Cardiology in the Young*, *26*(6), 1115–1122. http://dx.doi.org/10.1017/S1047951115001882

Osborne, J. W. (2008). *Best practices in quantitative methods*. Thousand Oaks, CA: Sage.

Podorefsky, D. L., McDonald-Dowdell, M., & Beardslee, W. R. (2001). Adaptation of preventive interventions for a low-income, culturally diverse community. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(8), 879–886. http://dx.doi.org/10.1097/00004583-200108000-00008

Quiñones, A. R., Markwardt, S., & Botoseneanu, A. (2016). Multimorbidity combinations and disability in older adults. *The Journals of Gerontology*, *71*(6), 823–830. http://dx.doi.org/10.1093/gerona/glw035

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. http://dx.doi.org/10.1177/014662167700100306

Rhee, S. H., Petroski, G. F., Parker, J. C., Smarr, K. L., Wright, G. E., Multon, K. D., . . . Komatireddy, G. R. (1999). A confirmatory factor analysis of the Center for Epidemiologic Studies Depression Scale in rheumatoid arthritis patients: Additional evidence for a four-factor model. *Arthritis Care and Research*, *12*(6), 392–400.

Schroevers, M. J., Sanderman, R., van Sonderen, E., & Ranchor, A. V. (2000). The evaluation of the Center for Epidemiologic Studies Depression (CES-D) scale: Depressed and positive affect in cancer patients and healthy reference subjects. *Quality of Life Research*, *9*(9), 1015–1029. http://dx.doi.org/10.1023/A:1016673003237

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology*, *62*(1), 123–146. http://dx.doi.org/10.1002/jclp.20213

Sheehan, T. J., Fifield, J., Reisine, S., & Tennen, H. (1995). The measurement structure of the Center for Epidemiologic Studies Depression Scale. *Journal of Personality Assessment*, *64*(3), 507–521. http://dx.doi.org/10.1207/s15327752jpa6403_9

Steger, M. F. (2006). An illustration of issues in factor extraction and identification of dimensionality in psychological assessment data. *Journal of Personality Assessment*, *86*(3), 263–272. http://dx.doi.org/10.1207/s15327752jpa8603_03

Thombs, B. D., Hudson, M., Schieir, O., Taillefer, S. S., & Baron, M. (2008). Reliability and validity of the Center for Epidemiologic Studies Depression Scale in patients with systemic sclerosis. *Arthritis Care and Research*, *59*(3), 438–443. http://dx.doi.org/10.1002/art.23329

Correspondence regarding this article should be directed to Theodore D. Cosco, PhD, University of Oxford, Oxford Institute of Population Ageing, 66 Banbury Road, Oxford, OX2 6PR, United Kingdom. E-mail: theodore.cosco@ageing.ox.ac.uk