Measurement properties of the Centers for Epidemiological Studies Depression Scale (CES-D) in a sample of African-American and non-Hispanic White pregnant women

Renée B. Canady[a], **Manfred Stommel**[b], and **Claudia Holzman**[c]

[a]Ingham County Health Department, 5303 S. Cedar, Lansing, MI 48911; fax: (517) 887-4308; office: (517) 887-4466; rcanady@ingham.org

[b]Michigan State University, College of Nursing

[c]Michigan State University, College of Human Medicine, Department of Epidemiology

Abstract

This study investigates the appropriateness of using the CES-D scale for comparing depressive symptoms among pregnant women of different races. Black and white women were matched on education, age, Medicaid status, and marital status/living arrangements. The matching procedure yielded a study sample of 375 in each ethnic group. Using confirmatory factor analysis, the fit of several factor models for the CES-D was evaluated. One CES-D item, "everything was an effort," showed a low item-total correlation (0.04 among blacks; 0.22 among whites) and was excluded from further analysis. After imposing the constraints of equal factor loadings and factor covariance across both groups, a two-factor-model with 19 CES-D items provided a good fit. Only the loading for the "was happy" item displayed a small difference between the two groups. Furthermore, the correlations between the original 20-item and the unbiased 18-item scales were r=0.994 for whites and r=0.992 for blacks. The results suggest that the 20-item CES-D can be used to compare depressive symptoms in white and black pregnant women without introducing significant ethnic/racial bias in the measurement of these symptoms.

Keywords

race comparison; depression & pregnancy; CES-D; ethnic mental health

1. INTRODUCTION

Many women, particularly low-income women and adolescents, experience depressive symptoms during pregnancy (Orr, Sherman, & Prince, 2002; Marcus, Flynn, Blow, & Barry, 2003; Holzman et al., 2005). These depressive symptoms have been linked to risk factors such as drinking, smoking and substance abuse that can lead to unfavorable pregnancy outcomes (Steyn et al., 2006; Zhu & Valbo, 2002). In addition, there may be more direct associations between depressive symptoms in pregnancy and pre-eclampsia (Kurki et al., 2000) and low birth weight (Hoffman & Hatch, 1996). These associations may be especially prevalent among women of lower socioeconomic status (Hoffman & Hatch, 2000) who are disproportionately women of color.

Address correspondences sent to Renée B. Canady, 5303 S. Cedar, Lansing, MI 48911 or rcanady@ingham.org.

Investigators interested in measuring depressive symptoms in pregnancy face the difficult decision of which instrument to select. Historically, studies of depressive symptoms in pregnancy have used various screening tools such as the Beck Depression Inventory (Beck, Steer, & Garbin, 1988), the Edinburgh (Cox, Holden, & Sagovsky, 1987) and the Centers for Epidemiological Studies Depression Scale (CES-D) (Hoffman & Hatch, 2000). While the CES-D scale is one of the more frequently used scales, to date, no study has closely examined the measurement properties of the CES-D scale among pregnant women from diverse racial/ethnic backgrounds.

The effects of depressive symptoms on pregnancy outcome may vary between racial/ethnic groups, and depressive symptoms may partially mediate racial/ethnic differences in adverse pregnancy outcomes (Gaynes et al., 2005). However, before these issues can be adequately studied, it is important to ascertain if there is a cultural bias in the tools used to measure depressive symptoms. Often researchers overlook the possibility that measurement scales may not be "equivalent" or may not have the "same" measurement properties across groups being compared. For example, if African American and White American respondents differ systematically in their responses to some, but not all, of the indicators of depressive symptoms in a standardized instrument, the total scale scores may not provide an unbiased estimate of depressive symptoms across these two groups.

The CES-D scale, a tool that has been in the public domain since 1977 (Radloff, 1977), has often been used to compare prevalence of depressive symptoms in different racial/ethnic groups (Roberts, 1980; Aneshensel, Clark, & Frerich, 1984; Vera et al., 1991; Cole, Kawachi, Maller, & Berkman, 2000; Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004). Previous work on the comparison of the CES-D measurement properties between African Americans and White Americans is limited and has led to mixed results. Nguyen et al. (2004), using confirmatory factor analysis to compare two samples of low-income African Americans to one sample of White Americans, found that the traditional four-factor model provided the best fit in all three groups which included both men and women; however, imposing equality constraints on the factor loadings across the racial groups significantly worsened the fit of the model. In particular, the largest differences were found in the loadings of the "effort" item, which appeared to be a weaker indicator of depressive symptoms among African Americans. By contrast, sleeplessness, loneliness, crying and sadness appeared to contribute more to overall depression scores in African Americans than in White Americans.

Cole et al. (2000) used a proportional odds regression model, conditioned on the total scale score, to estimate item bias between African American and White American responses to the CES-D items. Their sample came from the New Haven EPESE study and included 2340 elderly individuals (age 65+), of whom 20% were African American. Two items, which comprise the interpersonal problems subscale ("people were unfriendly" and "people disliked me"), received more frequent endorsement by African Americans than by White Americans after controlling for overall level of depressive symptoms.

In studies comparing CES-D measurement properties across ethnic/racial groups, the groups often differ with respect to other relevant factors, such as education, age, or income, making it difficult to infer a clear explanation for group differences. Nguyen et al. (2004) attempted to control some of these influences by matching the two African American samples on age and education (within 5 years), but apparently lacked the sample size to match the African Americans to the White American sample. Cole et al. (2000) did not report on any matching procedure. The importance of potential confounding variables in measurement comparisons between ethnic or racial groups should not be underestimated. If, for instance, two ethnic groups differ substantially in their average educational achievement, and educational

achievement is highly correlated with the criterion variable whose measurement properties are compared, then any observed differences in measurement properties between the ethnic groups may be mistakenly attributed to ethnic culture rather than education-related culture.

Analyses of the internal structure of the CES-D scale have often yielded a four-factor model, which includes a 7-item "depressive affect" or "mood" subscale, a 4-item "positive affect" or "well-being" subscale, a 7-item "somatic and retarded activity" subscale, and a 2-item "interpersonal" subscale (Berkman et al., 1986; Hertzog et al., 1990; Nguyen et al., 2004). However, that is not invariably the case. Three-factor models, two-factor and even single-factor models have been shown to be consistent with some data (Beals, Manson, Keane, & Dick, 1991; Guarnaccia, Angel, & Worobey, 1989; Hertzog et al., 1990). A few researchers have found the subscale dimensions to be sufficiently independent to investigate their relations to predictor variables separately (Krause, 1986; Gatz & Hurwicz, 1990; Stommel & Wills, 2004), while others have argued that there is not enough empirical differentiation to warrant partitioning the CES-D scale into multiple subscales (Hertzog et al., 1990). For various reasons, researchers have sometimes excluded a few items from the scale (Radloff, 1977; Ensel, 1986; Liang, Tran, Krause, & Markides, 1989). With a special target population such as pregnant women, some items on the "somatic and retarded activity" subscale may not correlate with more overt indicators of depressed mood, such as "I feel sad." There is a particular concern that responses to such items as "I could not get going" or "everything was an effort" may indicate the physical burden of pregnancy rather than reflect depressive symptoms (Orr et al., 2002).

In this study, we examined the factor structure of the CES-D in a sample of 750 pregnant women (375 African Americans and 375 White Americans) matched on four variables that are known correlates of depressive symptoms: age, education, Medicaid status, and marital status/living arrangement. We were particularly interested in the relevancy of CES-D somatic items for measuring depressive symptoms in pregnant women, and the potential ethnic differences in CES-D measurement properties.

2. METHODS

2.1 Study Sample

This analysis used data from the Pregnancy Outcomes and Community Health (POUCH) Study, which enrolled pregnant women between August 1998 and June 2004 from 52 clinics located in 5 Michigan communities. Women were eligible for the POUCH study after being screened for maternal serum alpha-fetoprotein (MSAFP), a biomarker related to risk of preterm delivery, the major focus of the POUCH study. All POUCH participants were enrolled between the 15th and 27th week of pregnancy. Women were excluded if (1) they lacked proficiency in English; (2) had been diagnosed with diabetes mellitus before the pregnancy; (3) carried multiple fetuses; or (4) carried a fetus with a known chromosomal abnormality or birth defect.

A total of 3,038 women were enrolled in the POUCH Study. Nineteen were lost to follow-up, leaving 3,019, of which 743 were African Americans, and 2,018 were White Americans. Race/ethnicity was determined by maternal self-report in a structured interview. Women were given the option of choosing more than one race/ethnic heritage, and those who did were then asked, "If you could pick only one, which would you pick?" Their response to this question was used to assign a single race/ethnic group for the purpose of these analyses.

For this investigation of the scaling properties of the CES-D, African American and White American POUCH participants were matched using four variables: education (years of formal schooling); age (in years); Medicaid Insurance status (yes or no); and marital status/

living arrangement (living with a spouse, living with a partner, living alone). These matching criteria were chosen because (1) they involve variables which are known predictors of total CES-D scores (Blazer et al., 1998; Jang et al., 2005) and (2) they were correlated with race in the total study sample, in which African Americans, on average, had less formal education, were younger, were more likely to have Medicaid Insurance, and less likely to live with a spouse. If more than one woman was available for a match in either race, the matched pair was randomly selected from the stratum of women with the same combination of the four characteristics described above. This frequency matching procedure (Rothman & Greenland, 1998) yielded a sub-sample of 375 African Americans matched to 375 White Americans. Matching on Medicaid Insurance status did not differentiate between enrollment in Medicaid before pregnancy or during the pregnancy (in Michigan, the latter is based on the more lenient income eligibility criterion of up to 185% of the federal poverty level).

2.2 Measure

At study enrollment, POUCH participants completed in-person interviews and self-administered questionnaires that included measurement of depressive symptoms using the CES-D. Less than 1% of the sample had missing responses on some CES-D items. If fewer than 4 item responses were missing for a particular respondent, scores were substituted via maximum likelihood imputations (Little &Rubin, 2002). Three women with more than 10 missing item responses were excluded from the analysis.

The CES-D instrument contains 20 items addressing depressive symptoms. Respondents indicate how often (within the last week) they experienced those symptoms: "rarely or none of the time" (0); "some or a little of the time" (1); "occasionally or a moderate amount of time" (2); or "most or all of the time" (3). In most studies, researchers employ a total scale score summing the responses of all 20 four-point items (e.g., Ensel & Lin, 1991; Lewinsohn, Rhode, Seeley, & Fischer, 1991). The resulting scores have a potential range of 0 to 60, but tend to be skewed positively in non-psychiatric populations, with most respondents scoring in the lower ranges and mean scale scores not exceeding 10 in the general population (Devins & Orme, 1986; Radloff & Locke, 1986).

2.3 Analysis

Using confirmatory factor analysis routines available in EQS6.1 (Bentler, 2005), the fit of several factor models for the CES-D was evaluated in both the African American and White American samples. Levels of "equivalence" across the two matched groups were tested by addressing the following question: (1) Do the same items load on the same factors (subscales) in each of the two groups? (2) Are the (unstandardized) factor loadings equal across the two groups? And (3) Are the covariances among the latent factors of equal magnitude across the two groups? These three questions represent a hierarchy of constraints that can be tested using the chi-square difference test: if adding constraints does not worsen the fit of the factor model, this would be evidence that the assumption of across-group equivalence in the relevant parameters is justified (Byrne, 1998, Stommel, Wang, Given, & Given, 1992). In addition to the chi-square difference test, other fit indices were used to evaluate the overall fit of the models: the Goodness-of-Fit Index (GFI), the Comparative Fit Index (CFI), the Bentler-Bonnett normed fit index (BBNFI) and the Root Mean Square Error of Approximation (RMSEA). Adequate fit requires the former three indices exceed the value of 0.95, while RMSEAs of less than 0.05 are considered adequate (Hu & Bentler, 1998).

3. RESULTS

Table 1 offers a comparison of the matched sample to the total study sample with respect to the matching criteria. It is apparent that the matched sample is not a representative subgroup of the study sample: its demographic profile resembles that of the black study participants, since matching required the disproportionate elimination of older white women with the result that the remaining white subjects were younger, poorer, less educated and less likely to be married than non-selected whites. The mean CES-D score in the matched sample of N=750 was 16.5, a little bit higher than the 16.0 threshold used to indicate a positive screen for depression. In the entire cohort the mean CES-D was 13.7.

Table 2 shows descriptive information on the CES-D items (means and item-total correlations) and scale reliabilities of the total CES-D scale for African American and White respondents. The internal consistency (reliability) of the item responses was, as expected, quite high within the matched sample (Cronbach's alpha: 0.898). However, one item ("everything was an effort") did not have item-total correlations above the customary cut-off point of 0.3 (Nunnally & Bernstein, 1994) in either the white (r=0.22) or the black (r=0.04) subgroup confirming that, in this sample of pregnant women, responses to this item were essentially unrelated to the other indicators of depressive symptoms. In addition, the mean endorsement levels for this "effort" item were much higher than that for most other CES-D items. The elimination of this item from the scale leads to a marginal improvement in the internal consistency of the scale in both racial groups (Table 2).

As previously noted, many investigators have found that a four-factor model best describes the response pattern to the CES-D items, but this finding is not universal. Initially, the four-factor-model was fitted in the combined matched sample, imposing no constraints on the factor loadings across the racial groups to establish a baseline model against which to test cross-group equality constraints (Byrne, Shavelson & Muthén, 1989). Although the overall fit of this model appears quite acceptable (χ^2=566.79, df=290, p≤0.001; χ^2/df=1.95; GFI=0.913; CFI=0.977; BBNFI=0.966; RMSEA=0.036), inter-factor correlations between the depressive mood and somatic symptom factors were larger than 0.9 and those between the interpersonal factor and the depressive mood/somatic symptoms factors were larger than 0.7, arguing for a simplified factor structure. (Exploratory ML factoring barely suggested 3 factors, with the third factor accounting for an Eigenvalue of just 1.0.) Subsequently, we tested a two-factor model, distinguishing only the 4-item positive-affect factor from a depressive-symptoms factor that is based on the remaining 15 items.

The fit of this model--base model 1 in Table 3 only allows for a single factor loading per item and the covariance between the two factors in both the white and black groups--is marginally worse than the original four-factor model (χ^2/df=2.21 versus χ^2/df=1.95). However, using the Lagrange Multiplier test (Satorra, 1989) to identify unwarranted constraints in four successive modification steps leads to a better fit of the two-factor model. Each of the following models successively provides a better fit to the data, due to the release of the specified constraints: Base model 2 in Table 3 allows for an additional non-zero covariance in both racial groups among the errors associated with the two interpersonal items (CES-D items 15 and 19), which indicates the existence of some residual unique variance associated with these items. Base model 3 adds an unconstrained factor loading in both racial groups of the CES-D 12 ("was happy") item on the depressive symptom factor (Factor 1). Base model 4 allows for unconstrained factor loadings in both racial groups of CES-D item 6 ("felt depressed") on the positive-affect factor (Factor 2), and base model 5 adds unconstrained factor loading in both groups for CES-D item 8 ("felt hopeful about the future") on the depressive-symptoms factor (Factor 1). No further improvements in fit could be made through the release of additional constraints; thus, we accept base model 5 as a

well-fitting, quite parsimonious model that imposes no equality constraints across the two racial groups (χ^2=541.57, df=294, p≤0.001; χ^2/df=1.84; GFI=0.916; CFI=0.980; BBNFI=0.976; RMSEA=0.034).

To test for the invariance of the factor covariance structures across the two racial groups (see Table 3), we imposed the following restrictions: all 20 factor loadings, including the cross-factor loadings for CES-D items 6, 12 and 18, in base model 5 were constrained to be equal across groups, with similar cross-group constraints imposed on the covariance between the two factors and the covariance between the error terms associated with the interpersonal items (CES-D items 15 and 19). Thus, compared to base model 5, this constrained model 6 (see Table 3) has 22 more degrees of freedom and implies a strict interpretation of factor structure equivalence: all free factor loadings and all free covariances (between the two factors and the two error terms associated with the interpersonal items) are hypothesized to be equal across the two racial groups. The resulting model provides a good fit to the data: the increase in the χ^2-statistic by 27.84 is associated with 22 additional degrees of freedom, rendering it non-significant. However, the Lagrange Multiplier test was again used to identify one equality constraint that was inconsistent with the data: the factor loadings of CES-D item 12 ("was happy") on the depressive-symptoms factor (Factor 1) differed significantly between the white and black respondents: it was non- significant (did not differ from zero) in the former and highly significant in the latter (see Table 4). Except for this difference, all other cross-group constraints were consistent with the data resulting in a good overall fit of model 7 as shown in Table 3 and 4 (χ^2=561.91, df=315, p≤0.001; χ^2/df=1.78; GFI=0.914; CFI=0.980; BBNFI=0.978; RMSEA=0.032). Thus, 18 items of the CES-D show no measurement bias in the sense that their internal factor structure is identical for the black and white respondents. While the CES-D item 12 ("was happy") displays a small difference in the factor loadings between the two racial groups, the correlations between the 19-item summated CES-D scale score and the 18-item CES-D scale, from which item 12 is eliminated, are respectively: r=0.998 among whites and r=0.997 among blacks. The correlations of the 18-item scale with the original 20-item scale are r=0.994 for whites and r=0.992 for blacks.

Table 5 shows differences in CES-D scores by race in both the matched analysis sample (N=750) and the total study sample (N=2543), using a linear regression model to compare unadjusted scores and scores adjusted for the matching variables. The results show virtually no difference in the mean CES-D score between the two racial groups in the matched analysis sample (mean difference: 0.05, p>0.94), but a substantial difference in the original study sample (mean difference: 5.1, p<0.01). However, after adjusting for the matching variables, the difference in mean CES-D scores between blacks and whites shrinks to a non-significant 0.8 (p>0.07) in the original study sample. Given the very high correlations between the original 20-item CES-D scale scores and the 'unbiased' 18-item scores, mean score differences in the matched analysis and the original study samples replicate the pattern for the 20-item scale shown in Table 5.

4. DISCUSSION

In this study we focused on the question of whether or not the use of the CES-D scale in survey research involving pregnant women of different races introduces measurement biases. While measurement bias and measurement equivalence are multi-faceted concepts which can be explored using multiple methods (Hui & Triandis, 1985; Knight & Hill, 1998), we chose to focus on the equivalence of the internal structural properties of the CES-D scale in the two comparison groups exploring factorial invariance with the help of confirmatory factor analysis models (Vandenberg & Lance, 2000). However, before applying CFA models to racial comparison groups, we addressed the perennial problem of confounding

factors in racial comparisons (Doucette-Gates, Brooks-Gunn & Chase-Lansdale, 1998) through matching procedures that included age, education, marital status and living arrangements as well as Medicaid status among the matching criteria.

The use of matched samples has some limitations, chief among them being that the application of several matching criteria quickly reduces the available sample size (Stommel & Wills, 2004). Matching also tends to result in an analysis sample that is not representative of the original study sample. In the present study, the constraints of matching across ethnic groups resulted in average maternal characteristics of the retained subjects in the matched sample that differed from those of the overall POUCH cohort. For example, the matching procedure required the elimination of relatively large numbers of older White American women and also some young African Americans, because of a lack of sufficient counterparts in the respective comparison groups. Similarly, we selected few individuals among numerous highly educated white women to match with the few highly educated black women in the cohort. However, it is important to keep in mind that 'representativeness' is not a goal of the matching procedure, but avoiding the error of attributing measurement biases to cultural understandings associated with race, when in fact such differences might be associated with education or class (Doucette-Gates et al., 1998). This is not to deny the fact that the experience of race is lived through social contexts such as education, marital status and socioeconomic status (our matching variables).

In addition to examining race differences in CES-D measurement properties, our study was unique in assessing the properties of CES-D items within a sample of pregnant women. Prior investigations have examined the measurement properties of the CES-D scale in adult and elderly African Americans and White Americans (Cole et al., 2000; Foley, Reed, Mutran & DeVellis, 2002), and teens (Hales et al., 2006). However, these measurement properties have not been evaluated in a multi-ethnic sample of pregnant women.

A major concern among researchers examining depressive symptoms in pregnancy has been the relevance of the somatic indicators within the CES-D scale. In our matched sample, only one item in this subscale, "everything was an effort," showed a corrected item-total correlation in both racial groups, which was low enough (<0.3) to warrant exclusion from the summed rating scale (Nunnally & Bernstein, 1994). In addition, the overall mean response to this "effort" item was higher than for any other CES-D item, a finding, which is not often observed among other populations, which may be assumed at risk for fatigue, such as elderly respondents (Cole et al., 2000). The higher endorsement generated by the "effort" item and the absence of correlation between this item and the remaining CES-D items lead us to conclude that it may not be a useful indicator of depressive symptoms among pregnant women. Careful evaluation of the "effort" item in the CES-D scale is warranted among studies of depressive symptoms in pregnant women.

For the remaining 19 items, we were able to show that an (almost) identical factor model with strict equality constraints across racial groups fit the data remarkably well. The only exception to the imposed cross-group constraints on any of the estimated parameters was the lack of equality of the factor loadings of the CES-D item 12 ("was happy"). This (reverse coded) positive affect item appeared to be viewed by blacks like other indicators of depressive symptoms; white respondents seemed to make a sharper distinction between the positive and negative valences of depression (Schroevers et al., 2000). Despite this difference, we are more impressed by the overall similarity in the factor structure of the CES-D across the racial groups. Thus, it is not a coincidence that the correlations between the 'unbiased' 18- item; 'slightly biased' 19-item; and the original 20-item CES-D scale all exceed the value of 0.99 in both racial groups. Thus, we think it is unlikely that the use of

the original 20-item CES-D scale in racial comparisons among pregnant women introduces racial measurement biases of consequence.

Acknowledgments

REFERENCES

Aneshensel CS, Clark VA, Frerichs RR. Race, ethnicity, and depression: a confirmatory factor analysis. Journal of Personality Social Psychology 1983;2:385.

Beals J, Manson SM, Keane EM, Dick RW. Factorial structure of the Center for Epidemiologic Studies-Depression scale among American Indian college students. Psychological Assessment 1991;3:623–627.

Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck depression inventory: twenty five years of evaluation. Clinical Psychology 1988;8:77–100.

Bentler, PM. Program Manual. Multivariate Software, Inc.; Encino, CA: 2005. EQS6.1 Structural Equation Modeling Software.

Berkman LF, Berkman CS, Kasl S, Freeman DH Jr. Leo L, Ostfeld AM, Cornoni-Huntley J, Brody JA. Depressive symptoms in relation to physical health and functioning in the elderly. American Journal of Epidemiology 1986;124:372–388. [PubMed: 3740038]

Blazer D, Landemann L, Simonsick E, Saunders W. Symptoms of depression among community-dwelling elderly African-American and White older adults. Psychological Medicine 1998;28:1311–20. [PubMed: 9854272]

Byrne BM, Shavelson RJ, Muthen B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin 1989;105(3):456–66.

Byrne, BM. Structural equation modeling with LIS-REL, PRELIS, and SIMPLIS. Lawrence Erlbaum Associates; Mahwah, NJ: 1998.

Cole SR, Kawachi I, Maller SJ, Berkman LF. Test of item-response bias in the CES-D scale: experience from the New Haven EPESE study. Journal of Clinical Epidemiology 2000;53:285–9. [PubMed: 10760639]

Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: Development of the 10 item Edinburgh postnatal depression scale. British Journal of Psychiatry 1987;150:782–786. [PubMed: 3651732]

Devins, GM.; Orme, CM. Center for Epidemiological Studies Depression Scale. In: Kayser; Sweetland, editors. Test Critiques. Vol. 2. Kansas City, MO: 1985. p. 144-160.

Doucette-Gates, A.; Brooks-Gunn, J.; Case-Lansdale, PL. The Role of Bias and Equivalence in the Study of Race, Class and Ethnicity. In: McLoyd, V.; Steinberg, L., editors. Studying Minority Adolescence. Erlbaum; New Jersey: 1998. p. 211-236.

Ensel, WM. Measuring depression: The CES-D scale. In: Lin, N.; Dean, A.; Ensel, WE., editors. Social Support, Life Events, and Depression. Academic Press; New York: 1986. p. 51-70.

Ensel WM, Linn N. The life stress paradigm and psychological distress. Journal of Health and Social Behavior 1991;32:321–341. [PubMed: 1765624]

Foley KL, Reed PS, Mutran EJ, DeVellis RF. Measurement adequacy of the CES-D among a sample of older African-Americans. Psychiatry Research 2002;109:61–69. [PubMed: 11850052]

Gatz M, Hurwicz M. Are old people more depressed? Cross-sectional data on Center for Epidemiologic Studies Depression scale factors. Psychology and Aging 1990;5:284–290. [PubMed: 2378694]

Gaynes, BN.; Gavin, N.; Meltzer-Brody, S.; Lohr, KN.; Swinson, T.; Gartlehner, G.; Brody, S.; Miller, WC. Perinatal Depression: prevalence, screening accuracy, and screening outcomes. Agency for Healthcare Research and Quality; 2005. Evidence Report No 119; AHRQ Publication No. 05-E006-2

Guarnaccia PJ, Angel R, Worobey JL. The factor structure of the CES-D in the Hispanic Health and Nutrition Examination Survey: the influences of ethnicity, gender, and language. Social Science and Medicine 1989;29:85–94. [PubMed: 2740931]

Hales DP, Dishman RK, Motl RW, Addy CL, Pfeiffer KA, Pates RR. Factorial validity and invariance of the Center for Epidemiology Studies Depression (CES-D) Scale in a sample of Black and White adolescent girls. Ethnicity and Disease 2006;16:1–8. [PubMed: 16599341]

Hertzog C, Van Alstine J, Usala PD, Hultsch DF, Dixon R. Measurement properties of the Center for Epidemiologic Studies Depression scale (CES-D) in older populations. Psychological Assessment 1990;2:64–72.

Hoffman S, Hatch MC. Depressive Symptomatology During Pregnancy: Evidence for an Association with Decreased Fetal Growth in Pregnancies of Lower Social Class Women. Health Psychology 2000;19:535–543. [PubMed: 11129356]

Hoffman S, Hatch MC. Stress, social support and pregnancy outcome: A reassessment based on recent research. Paediatric and Perinatal Epidemiology 1996;10:380–405. [PubMed: 8931053]

Holzman C, Eyster J, Tiedje LB, Roman L, Seagull E, Rahbar MH. A life course perspective on depressive symptoms in mid-pregnancy. Maternal Child Health Journal 2006;10:127–138.

Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to unparameterized model specification. Psychological Methods 1998;3:424–53.

Hui CH, Triandis HC. Measurement in cross-cultural psychology: A review and comparison of strategies. Journal of Cross-Cultural Psychology 1985;16(2):131–52.

Jang Y, Borenstein AR, Chiriboga DA, Mortimer JA. Depressive symptoms among African American and White older adults. Journal of Gerontology: Psychological Sciences 2005;60B(6):313–9.

Knight, GP.; Hill, NE. Measurement Equivalence in Research Involving Minority Adolescents. In: McLoyd, V.; Steinberg, L., editors. Studying Minority Adolescence. Erlbaum; New Jersey: 1998. p. 183-210.

Krause N. Stress and sex differences in depressive symptoms among older adults. Journal of Gerontology 1986;41:727–731. [PubMed: 3772048]

Kurki T, Hiilesmaa V, Raitasalo R, Mattila H, Ylikorkala O. Depression and anxiety in early pregnancy and risk for pre-eclampsia. Obstetric Gynecology 2000;95:487.

Lewinsohn PM, Rhode P, Seeley JR, Fischer SA. Age and depression: unique and shared effects. Psychology and Aging 1991;6:247–260. [PubMed: 1863394]

Liang J, Tran TV, Krause N, Markides KS. Generational differences in the structure of the CES-D scale in Mexican Americans. Journal of Gerontology 1989;44:110–20.

Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2nd ed.. Wiley & Sons, Inc.; New York, NY: 2002.

Marcus SA, Flynn HA, Blow FC, Barry KL. Depressive Symptoms among Pregnant Women Screened in Obstetrics Settings. Journal of Women's Health 2003;12:373–380.

Nguyen HT, Kitner-Triolo M, Evans MK, Zonderman AB. Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. Psychiatry Research 2004;126:277–187.

Nunnally, JC.; Bernstein, IH. Psychometric Theory. Third Edition. MacGraw-Hill; New York: 1994.

Orr ST, Sherman AJ, Prince CB. Maternal prenatal depressive symptoms and spontaneous preterm births among African American women in Baltimore, Maryland. American Journal of Epidemiology 2002;165:797–802. [PubMed: 12396996]

Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977;1:385–401.

Radloff, LS.; Locke, BZ. The community mental health assessment survey and the CES-D scale. In: Weissman, MM.; Myers, JK.; Ross, CE., editors. Community Surveys of Psychiatric Disorders. Rutgers University Press; New Brunswick, NJ: 1986. p. 177-189.

Roberts RE. Reliability of the CES-D scale in different ethnic contexts. Psychiatry Research 1980;2:125–134. [PubMed: 6932058]

Stommel M, Wang S, Given CW, Given B. Focus on psychometrics: Confirmatory factor analysis (CFA) as a method to assess measurement equivalence. Research in Nursing and Health 1992;15:399–405. [PubMed: 1529124]

Rothman, KJ.; Greenland, S. Modern epidemiology. 2nd ed. Lippincott Williams & Wilkins; Philadelphia, PA: 1998.

Satorra A. Alternative test criteria in covariance structure analysis: A unified approach. Psychometrica 1989;54:131–51.

Schroevers MJ, Sandermann R, Van Sonderen E, Ranchor AV. 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 2000;9:1015–1029. [PubMed: 11332223]

Steyn K, de-Wet T, Saloojee Y, Nel H, Yach D. The influence of maternal cigarette smoking, snuff use and passive smoking on pregnancy outcomes: the Birth To Ten Study. Paediatric Perinatal Epidemiology 2006;20:90–99.

Stommel M, Wang S, Given CW, Given BA. Confirmatory factor analysis (CFA) as a method to assess measurement equivalence. Research in Nursing and Health 1992;15(5):352–360.

Stommel, M.; Wills, C. Clinical Research: Concepts and Principles for Advanced Practice Nurses. Lippincott, Williams & Wilkins; Philadelphia, PA: 2004.

Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods 2000;3(1):4–70.

Vera M, Alegria M, Freeman D, Robles RR, Rios CF. Depressive symptoms among Puerto Ricans: Island poor compared with residents of the New York City area. American Journal of Epidemiology 1991;134:502–510. [PubMed: 1897506]

Zhu SH, Valbo A. Depression and smoking during pregnancy. Addictive Behavior 2002;27:649.

Table 1

Socio-demographic Characteristics of Matched African American and White American Pregnant Women (N=750) Table 1 Comparison of Matched Analysis Sample to Total Study Sample on Key Socio-Demographic Variables

	Total Study Sample (N=2753)			Matched Analysis Sample (N=750)		
	Mean	**Std.Dev.**	**Range**	**Mean**	**Std.Dev.**	**Range**
Age:	25.98	5.8	15-47	23.01	4.7	15-41
	N		**%**	**N**		**%**
Living Arrangement:						
Living with Spouse	1384		50.3	142		18.9
Living with Partner	609		22.1	238		31.7
Living Alone	760		27.6	370		49.3
Medicaid Status:						
Not on Medicaid	1449		52.7	166		22.1
On Medicaid	1302		47.3	584		77.9
Education:						
Less than High School	485		17.6	188		25.1
High School Graduate	815		29.6	318		42.4
Some College	785		28.5	186		24.8
College Graduate	421		15.3	38		5.1
Post Bachelor Education	245		8.9	10		2.7
Multiple Carriages:						
Primagravida:	815		30	202		27
Multigravida:	1938		70	548		73

Table 2

CES-D Items Means and Item-Corrected-Total Correlations (ITC) by Race

	Means:			ITCs:	
	Whites	Blacks	p-value	Whites	Blacks
CES-D 1.. was bothered by things …	0.70	0.74		0.52	0.48
CES-D 2.. had a poor appetite …	0.47	0.63	0.01	0.41	0.34
CES-D 3.. could not shake off the blues	0.57	0.63		0.66	0.66
CES-D 4.. felt as good as other people	0.92	0.92		0.42	0.25
CES-D 5.. trouble keeping my mind on what I was doing …	1.03	0.86	0.02	0.43	0.55
CES-D 6.. felt depressed	0.88	0.94		0.76	0.76
CES-D 7.. everything was an effort …	1.29	1.56	0.01	0.22	0.04
CES-D 8.. felt hopeful about the future …	0.94	0.90		0.31	0.16
CES-D 9.. thought life a failure	0.45	0.53		0.66	0.63
CES-D 10.. felt fearful	0.61	0.57		0.57	0.53
CES-D 11.. sleep was restless …	1.36	1.02	0.01	0.32	0.34
CES-D 12.. was happy …	0.77	0.87		0.59	0.57
CES-D 13.. talked less than usual …	0.73	0.83		0.44	0.31
CES-D 14.. felt lonely …	0.87	0.88		0.68	0.72
CES-D 15.. people were unfriendly …	0.52	0.50		0.46	0.33
CES-D 16.. enjoyed life …	0.89	0.76	0.02	0.58	0.42
CES-D 17.. had crying spells …	0.96	1.05		0.61	0.61
CES-D 18.. felt sad …	1.03	1.03		0.76	0.79
CES-D 19.. felt people disliked me …	0.49	0.47		0.61	0.55
CES-D 20.. couldnot get going …	0.92	0.79	0.04	0.60	0.55
	Whites:			Blacks:	
Cronbach's Alpha (20 items):	0.898			0.873	
Cronbach's Alpha (19 items minus item 7):	0.903			0.884	

Table 3

Goodness-of-Fit Tests for Alternative Measurement Models

	χ^2	df	χ^2/df	$\Delta\chi^2$	GFI	CFI	BBNFI	RMSEA
Two-Factor Models (19 CES-D items) without cross-group (black vs. white) constraints:								
1) base model 1 (1 factor loading for each item, covariance between Factor 1 & 2)	666.44	302	2.21	---	0.899	0.970	0.966	0.40
2) base model 2 (model 1 plus covariance between CES-D 15 & 19)	598.25	300	1.99	68.19[*]	0.909	0.975	0.972	0.36
3) base model 3 (model 2 plus additional factor loading of CES-D 12 on Factor 1)	575.27	298	1.93	22.55[*]	0.912	0.977	0.974	0.35
4) base model 4 (model 3 plus additional factor loading of CES-D 6 on Factor 2)	559.95	296	1.89	15.32[*]	0.914	0.978	0.975	0.35
5) base model 5 (model 4 plus additional factor loading of CES-D 8 on Factor 1)	541.57	294	1.84	18.38[*]	0.916	0.980	0.976	0.34
Two-Factor Models (19 CES-D items) with cross-group (black vs. white)constraints:								
1) constrained model 6 (model 5 plus cross-group constraints on all factor loadings and covariances)	569.41	316	1.80	27.84[$]	0.913	0.979	0.977	0.33
2) constrained model 7 (model 6 plus release of cross-group constraint on factor loading of item CES-D 12)	561.91	315	1.78	7.50[*]	0.914	0.980	0.978	0.32

[*] p<0.05

[$] p≥0.05

Table 4

19-item CES-D Scale with <u>Identical</u> (=Constrained) Unstandardized Factor Loadings Among Matched Black (N=375) and White (N=375) Pregnant Women (Standard Errors in Parentheses)

FACTOR 1:		
DEPRESSIVE SYMPTOMS	**FACTOR 1**	**FACTOR 2**
	Loadings (Standard Errors)	
CES-D 1.. was bothered by things …	0.619 (0.051)	
CES-D 2.. had a poor appetite …	0.388 (0.045)	
CES-D 3.. could not shake off the blues …	0.837 (0.054)	
CES-D 5.. trouble keeping my mind on what I was doing …	0.687 (0.057)	
CES-D 6.. felt depressed …	1.000	0.215 (0.056)
CES-D 9.. thought life a failure …	0.817 (0.054)	
CES-D 10.. felt fearful …	0.733 (0.055)	
CES-D 11.. sleep was restless …	0.479 (0.060)	
CES-D 13.. talked less than usual …	0.482 (0.053)	
CES-D 14.. felt lonely …	1.107 (0.066)	
CES-D 15.. people were unfriendly …	0.434 (0.046)	
CES-D 17.. had crying spells …	0.968 (0.064)	
CES-D 18.. felt sad …	1.127 (0.063)	
CES-D 19.. felt people disliked me …	0.672 (0.049)	
CES-D 20.. could not get going …	0.752 (0.056)	
FACTOR 2:		
POSITIVE AFFECT, WELL-BEING SUBSCALE		
CES-D 4.. felt as good as other people		0.758 (0.082)
CES-D 8.. felt hopeful about the future …	0.330 (0.082)	1.000
CES-D 12.. was happy …	whites: 0.049 (0.067)[*]	0.816 (0.078)
CES-D 12.. was happy …	blacks: 0.255 (0.067)[*]	0.816 (0.078)
CES-D 16.. enjoyed life …		0.966 (0.089)

Covariance (Factor1, Factor 2): 0.303 (0.041); Error Covariance of (CES-D15, CES-D19): 0.137 (0.019)

[*] factor loadings for CES-D12 have not been constrained across groups.

Table 5

CES-D Scale Scores for Original 20-item Scale by Race alone and By Race, Medicaid Status, Education, Age and Living Arrangements

	Total Study Sample (N=2753)		Matched Analysis Sample (N=750)	
	Dependent Variable: CES-D scores			
Independent Variable:				
	Coefficient	p-value	Coefficient	p-value
Intercept	12.36	<0.01	16.43	<0.01
Race (1=black, 0=white)	5.09	<0.01	0.05	>0.94
	Dependent Variable: CES-D scores			
Independent Variables:				
	Coefficient	p-value	Coefficient	p-value
Race (1=black, 0=white)	0.80	>0.09	0.05	>0.94
Medicaid Status (1=yes, 0=no)	2.28	<0.01	2.46	<0.02
Education (schooling in years)	−0.62	<0.01	−0.49	>0.07
Age (in years)	0.04	>0.30	0.18	>0.07
Living Arrangement (lives with spouse[*])				
Lives with partner	3.44	<0.01	3.77	<0.01
Lives alone	5.31	<0.01	4.64	<0.01

[*] Reference category for living arrangement