Abstract: Measuring mental health accurately is an important endeavor for screening purposes. Depression scales, such as the Center for Epidemiologic Studies Depression (CES-D) scale, have been well-established among different populations. Yet, little work has been done to examine the reliability and validity of the CES-D among older American Indians and Alaska Natives. The purpose of our study was to examine the factor structure, reliability, and concurrent validity of the full 20-item and abbreviated 12-item CES-D scale with a sample of older American Indians. Our findings demonstrate excellent internal reliability and concurrent validity of the full and abbreviated CES-D scales in our study sample.

INTRODUCTION

By 2020, depression is predicted to become the second leading contributor, behind heart disease, to the global burden of disease and physical disability (World Health Organization, 2012). Among community-dwelling older adults, the prevalence of clinically significant depressive symptomatology typically ranges from 8 to 16% (Blazer, 2003). The prevalence of depressive symptomatology has been shown to vary across older racial and ethnic populations (Dunlop, Song, Lyons, Manheim, & Chang, 2003). Although there is not a single definitive source of the prevalence of depression among American Indians and Alaska Natives (AI/ANs), collectively what data are available suggest disproportionately high rates of mental illness, including depression (Beals et al., 2005; Henry J. Kaiser Family Foundation, 2016; National Center for Health Statistics, 2014; Substance Abuse and Mental Health Services Administration, 2007).

Older AI/ANs, compared to older adults of other races and ethnicities, have been found to have higher prevalence rates of depressive symptomatology (Curyto et al., 1998). Specifically,
using the Centers for Epidemiological Studies—Depression (CES-D) scale (Radloff, 1977), researchers found the prevalence of clinically significant depressive symptomatology to be 18% among AIs aged ≥55 years in one tribe (Curyto et al., 1998).

Improving our understanding of depressive symptomatology among older AI/ANs will become more important as this population increases in size. Specifically, the number of AI/ANs aged ≥65 years is projected to more than triple between 2012 and 2050 and the number of AI/ANs aged ≥85 years is projected to have a more than sevenfold increase during the same time (Ortman, Velkoff, & Hogan, 2014). The increase in population combined with the lifelong disparities that AI/AN populations face (Goins et al., 2015) highlight an urgent need to understand depressive symptomatology in this population.

A meta-analysis comparing the factor structure of the CES-D scale across racial and ethnic populations suggests that great variation exists in how depressive symptomatology is conceptualized and expressed (Kim, DeCoster, Huang, & Chiriboga, 2011). Only one study has assessed the psychometric properties of the CES-D scale with older AIs (Chapleski, Lamphere, Kaczynski, Lichtenberg, & Dwyer, 1997). This study examined the factor structure of both the full 20-item scale and an abbreviated 12-item scale with 277 Great Lake region AIs aged ≥55 years. A 3-factor structure was determined to be the best fit model for the full scale, and the abbreviated scale was found to be an equally reliable and valid scale.

Measurement is a key foundational component of research. If a construct is not measured well, then concerns regarding scientific validity emerge. The literature on research and clinical measurement provides a thorough discussion of the potential sources of influences across different racial, ethnic, and cultural groups with respect to psychological assessments (Reynolds & Suzuki, 2013; US Department of Health and Human Services, 2001). Given this, mental health may be conceptualized differently in AI communities than in mainstream Western cultures (Beals et al., 2005; Hodge, Limb, & Cross, 2009). Thus, the purpose of this study was to examine the psychometric properties of a commonly used mental health measure with a sample of older AIs from a single tribe. Specifically, analyses procedures will be used to determine the factor structure, reliability, and concurrent validity of the full CES-D and the abbreviated scale with our sample.
METHODS

Analytic Sample

Data for these analyses were collected as part of the Native Elder Care Study, a cross-sectional study of community-dwelling older members of a federally recognized AI tribe (Goins Garrouste, Leading Fox, Geiger, & Manson, 2011). The tribe’s institutional review board, health board, tribal council, tribal elder council, and West Virginia University institutional review board approved the project. All study participants provided informed consent and received a $20 gift card. The Oregon State University institutional review board approved the secondary data analyses for this study.

From 2006 to 2008, using in-person interviewer-administered surveys, data were collected on demographic characteristics, physical functioning, mental and physical health, personal assistance needs, and psychosocial resources. Inclusion criteria for this study included being an enrolled tribal member, aged ≥55 years, residing in the tribe’s service area, non-institutionalized, and having passed a cognitive screen. Study inclusion criteria were determined by our tribal partners. For instance, the tribal partners requested that the age criteria be dropped from ≥65 years to ≥55 years and that the tribal partners were only interested in obtaining such information from members that were community-dwelling in their service area since the intention was to use the data to reassess the tribe’s array of community-based services. The Time and Change Test was used as a cognitive screener (Inouye, Robison, Froehlich, & Richardson, 1998). Tribal partners preferred this measure given its brevity and less bias due to participants’ socioeconomic characteristics compared to the more commonly used Mini Mental State Examination.

According to the tribal enrollment records, 1,430 persons were potentially eligible for study inclusion based on residential location and age. This list was randomized and the names and contact information were given to interviewers. Equal numbers of respondents were sought for the age groups 55-64, 65-74, and ≥75 years with a targeted sample size of 500. Randomly selected persons were recruited by telephone or home visit by an interviewer. Of the 633 persons assessed for eligibility, 50 were deemed ineligible. Of these 50 individuals, three resided outside of the tribe’s service area; 14 were in a nursing home; 19 were deceased; and 14 did not pass the cognitive screen. Most interviews were conducted in the participant’s home (87%), and the
remaining were conducted in a tribal office building. Seventy-eight persons refused to participate, yielding an 87% response rate and a final sample size of 505 with 491 who had complete responses to the full 20-item CES-D scale.

**Measures**

The CES-D scale measures depressive symptomatology, of which the full CES-D scale consists of 20 items (Radloff, 1977) and an abbreviated version consists of 12 items (Liang, Van Tran, Krause, & Markides, 1989); both were assessed as part of this study. The full scale’s reliability and validity has been demonstrated among older adults and across different racial and ethnic groups (Kim et al., 2011). Similarly, the 12-item abbreviated version has been validated with older AIs (Chapleski et al., 1997). The full version of the CES-D scale is comprised of four domains (i.e., depressed affect, positive affect, somatic symptoms, and perceptions regarding interpersonal relationships; Radloff, 1977). The CES-D scale asks respondents how often they felt each symptom in the past week, with a response scale of 0 to 3 (0 = rarely or none of the time, 1 = some or a little of the time, 2 = occasionally or a moderate amount of time, 3 = most or all of the time). Positive affect items are reverse coded with the full scale total sum score ranging from 0 to 60.

Independent variables included in the analyses were age (55-64, 65-74, ≥75), gender (male, female), educational attainment (<12 years, >12 years), and marital status (married/life partner, not married). Four measures were predicted to correlate with CES-D, thus assessing convergent validity. These measures included chronic pain, physical disability, social support, and self-efficacy. Chronic pain was assessed with the Chronic Pain Grade, which categorizes five grades of chronic pain ranging from pain free to most severe pain (Von Korff, Ormel, Keefe, & Dworkin, 1992). Physical disability was measured as the number of activities of daily living (ADL) and instrumental activities of daily living (IADL) limitations reported. The ADLs included bathing/showering, dressing, eating, transferring, walking, toileting, grooming, and getting outside (Fillenbaum, 1985). The IADLs included using the telephone, light housework, heavy housework, preparing meals, shopping, managing money, managing medications, and transportation (Lawton & Brody, 1969). Given that depression has been consistently associated with both physical disability and chronic pain (Bair, Robinson, Katon, & Kroenke, 2003; Bruce,
we expected that higher scores on the two CES-D scales would be positively associated with higher scores on both the Chronic Pain Grade and a count of ADL and IADL limitations.

Social support was measured with the Medical Outcomes Study Social Support Survey (Sherbourne & Stewart, 1991). This is a 19-item survey with a 5-point response selection (0 = none of the time, 1 = a little of the time, 2 = some of the time, 3 = most of the time, 4 = all of the time) and a sum score range of 0 to 76. The internal consistency of this scale was very high (α = 0.96). Self-efficacy was measured with the General Self-Efficacy Scale (Jerusalem & Schwarzer, 1992). We examined general self-efficacy using a 9-item scale with a 4-point response selection (0 = not at all true, 1 = hardly true, 2 = moderately true, 3 = exactly true) and a sum score range of 0 to 27. The internal consistency of this scale was also high (α = 0.90). Research has demonstrated a strong and consistent inverse relationship between poor mental health and social support (Fiori, Antonucci, & Cortina, 2006; Golden, Conroy, & Lawlor, 2009; Conte, Schure, & Goins, 2014). We expected that higher scores on the two CES-D scales would be negatively associated with higher scores on the Medical Outcomes Study Social Support Survey. Also, in light of the research that has shown a similar inverse relationship between depression and self-efficacy (Bandura, Pastorelli, Barbaranelli, & Caprara, 1999; Blazer, 2010), we expected that higher scores on the CES-D scales would also be inversely associated with the General Self-Efficacy Scale.

Statistical Analyses

Fourteen cases of those completely missing data on the CES-D scale were excluded. First, independent t tests were used to compare the CES-D scale scores of the 14 missing cases with the rest of the sample across sociodemographic characteristics. Results indicated the 14 cases were more likely to be older (p < 0.001) and have ≤12 years of education (p <0.001). Second, chi-square tests were used to analyze differences across sociodemographic characteristics of those with clinically significant depressive symptoms (CES-D score of ≥16) to those without (CES-D score of <16). The sample was randomly split into two analytic samples to use one for the exploratory factor analysis (EFA) and the other for the confirmatory factor analysis (CFA). The EFA was used to examine the factor structure and the CFA to confirm it. The EFA sample consisted of 246 cases of which 185 persons had complete data on the full CES-D scale. The CFA sample consisted of 245 cases of which 179 persons had complete data...
on the full CES-D scale. Overall, 26% of respondents had one or more CES-D scale item(s) missing. Of the 26% \(n = 128\) with some missing CES-D scale items, less than 30% \(n = 38\) had six or more missing items.

For the EFA, factor loadings were generated using an oblique rotation to determine the number of factors of the 20 items, with loadings of \(\geq .40\) indicative of sufficient salience in determining the factor structure. For the CFA, models using maximum likelihood estimation (listwise deletion) were run. Best model was assessed with Goodness of Fit Indices using the 1) chi-square statistic with non-significant values indicating a good fit, 2) comparative fit index with values \(>.95\) indicating a good fit, 3) root mean square error of approximation with values \(<.08\) indicating a good fit, and 4) standardized root mean square residual with values \(<.08\) demonstrating adequate fit (Acock, 2013).

The entire sample data \(N = 491\) was used to assess reliability and validity, using Cronbach’s alpha reliability test and Pearson product moment correlations. Expected direct correlations with chronic pain and physical disability and expected indirect correlations with social support and self-efficacy were assessed. StataCorp statistical analysis software version 12 was used for all analyses (Stata Statistical Software, 2007).

**RESULTS**

The prevalence of clinically significant depression (\(\geq 16\)) in this sample was 13.24% \(n = 65\). Those in the youngest age category, 50-64 years, were more likely to have clinically significant depression (CES-D score \(\geq 16\)) compared to those aged 65-74 and \(\geq 75\) years (19.9% versus 8.2% and 12.1%, respectively; \(p = 0.005\)). Those with 12 years or less of education were more likely to have clinically significant depression compared to those with some college and those with a college degree (15.7% versus 7.9%, \(p = 0.19\)). We found no statistically significant differences in clinical depression by gender or marital status.

Table 1 shows the EFA factor loadings for each of the scale items. The EFA for the full scale indicated a two-factor solution with 16 of the 20 items loading onto the first factor and 3 of the 20 items loading onto the second factor. With the exception of item 4 (“I felt that I was as good as other people”), all items loadings ranged from .403 to .832. The respective Eigenvalues for both factors was 6.61 and 1.05, with factor 1 representing 83.0% of the model variance and
factor 2 representing 17.0% of the model variance. The abbreviated scale fit with a one-factor solution with item loadings ranging from .413 to .772. The Eigenvalue was 4.25.

Table 1
Exploratory Factor Analysis, 2-Factor Structure of the Full 20-Item CES-D Scale and 1-Factor Structure of the 12-Item Abbreviated CES-D Scale (n = 185)

<table>
<thead>
<tr>
<th>Scale Items</th>
<th>Full 20-Item Scale</th>
<th>Abbreviated 12-Item Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor 1</td>
<td>Factor 2</td>
</tr>
<tr>
<td>1</td>
<td>I was bothered by things that usually don’t bother me</td>
<td>.666</td>
</tr>
<tr>
<td>2</td>
<td>I did not feel like eating; my appetite was poor</td>
<td>.536</td>
</tr>
<tr>
<td>3</td>
<td>I felt I could not shake off the blues even with help from my family and friends</td>
<td>.715</td>
</tr>
<tr>
<td>4</td>
<td>I felt that I was as good as other people</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>I had trouble keeping my mind on what I was doing</td>
<td>.562</td>
</tr>
<tr>
<td>6</td>
<td>I felt depressed</td>
<td>.832</td>
</tr>
<tr>
<td>7</td>
<td>I felt everything I did was an effort</td>
<td>.477</td>
</tr>
<tr>
<td>8</td>
<td>I felt hopeful about the future</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>I though my life had been a failure</td>
<td>.604</td>
</tr>
<tr>
<td>10</td>
<td>I felt fearful</td>
<td>.763</td>
</tr>
<tr>
<td>11</td>
<td>My sleep was restless</td>
<td>.460</td>
</tr>
<tr>
<td>12</td>
<td>I was happy</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>I talked less than usual</td>
<td>.461</td>
</tr>
<tr>
<td>14</td>
<td>I felt lonely</td>
<td>.752</td>
</tr>
<tr>
<td>15</td>
<td>People were unfriendly</td>
<td>.514</td>
</tr>
<tr>
<td>16</td>
<td>I enjoyed life</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>I had crying spells</td>
<td>.655</td>
</tr>
<tr>
<td>18</td>
<td>I felt sad</td>
<td>.724</td>
</tr>
<tr>
<td>19</td>
<td>I felt that people dislike me</td>
<td>.403</td>
</tr>
<tr>
<td>20</td>
<td>I could not get “going”</td>
<td>.474</td>
</tr>
</tbody>
</table>

Note. CES-D = Center for Epidemiologic Studies Depression scale. Items 4, 8, 12, and 16 were reverse coded; *Item 4 was dropped based on analysis results; †Item 8 was dropped based on analysis results. For the 3-factor model, Items 12 and 16 comprised the third factor.

Table 2 presents the Goodness of Fit Indices comparisons from the CFA for a 2-factor, 3-factor, and 4-factor model structure for the full 20 item CES-D scale and a 1-factor solution for the abbreviated CES-D scale. The indices confirmed a two-factor structure as the best model fit for the full scale (Chi-squared = 164.02; CFI = 0.969; RMSEA = 0.046; SRMR = 0.044). However, the fit statistics show that the abbreviated scale provides a superior fit (Chi-squared = 66.52; CFI = 0.980; RMSEA = 0.046; SRMR = 0.036) compared to the full scale factor models.
Table 2
Confirmatory Factor Analysis Fit Statistics for the CES-D Scale (n = 179)

<table>
<thead>
<tr>
<th></th>
<th>χ²</th>
<th>df</th>
<th>χ² p value</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full 20-Item Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Factor Model</td>
<td>164.02</td>
<td>119</td>
<td>0.004</td>
<td>0.969</td>
<td>0.046</td>
<td>0.044</td>
</tr>
<tr>
<td>3-Factor Model</td>
<td>151.46</td>
<td>82</td>
<td>&lt;0.001</td>
<td>0.942</td>
<td>0.067</td>
<td>0.061</td>
</tr>
<tr>
<td>4-Factor Model</td>
<td>203.67</td>
<td>122</td>
<td>&lt;0.001</td>
<td>0.940</td>
<td>0.062</td>
<td>0.054</td>
</tr>
<tr>
<td><strong>12-Item Model</strong></td>
<td>66.52</td>
<td>47</td>
<td>0.032</td>
<td>0.980</td>
<td>0.046</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Note: CES-D = Center for Epidemiologic Studies Depression scale; χ² = chi-squared value; CFI = comparative fit index; RMSEA = root mean squared error; SRMR = standardized root mean square residual.

The Cronbach reliability test indicated an overall mean score of 0.8716 for the full CES-D scale. The alpha score for item 4 suggested that dropping this item would improve the overall scale reliability to 0.8824. Similarly, the reliability score for the abbreviated scale was 0.8332. The alpha score for item 8 suggested that dropping this item would improve the overall scale reliability to 0.8466.

Table 3 presents the Pearson product moment correlations with chronic pain, physical disability, social support, and self-efficacy. The full CES-D scale demonstrated direct correlations with chronic pain (r = .326, p < 0.001) and physical disability (r = .366, p < 0.001). It was inversely correlated with social support (r = -.281, p < 0.001) and self-efficacy (r = -.315, p < 0.001). Comparable correlations were demonstrated with the abbreviated scale.

Table 3
Full 20-Item CES-D Factor and 12-Item Scale Correlations with Chronic Pain, Physical Disability, Social Support, and Self-Efficacy (n = 491)

<table>
<thead>
<tr>
<th></th>
<th>Full CES-D Scale</th>
<th>Abbreviated 12-Item Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor 1</td>
<td>Factor 2</td>
</tr>
<tr>
<td>Chronic Pain</td>
<td>r</td>
<td>p value</td>
</tr>
<tr>
<td></td>
<td>.326</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Physical Disability</td>
<td>.366</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Social Support</td>
<td>-.281</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>-.315</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note: CES-D = Center for Epidemiologic Studies Depression scale; Factor 1 = depressed affect, somatic activity, and interpersonal combined; Factor 2 = 3 items of the positive affect.
DISCUSSION

The intent of this study was to examine the factor structure, reliability, and concurrent validity of the full and abbreviated 12-item CES-D scales with a sample of older AIs. This study demonstrated excellent internal reliability of the full (alpha = 0.8824) and 12-item (alpha = 0.8466) CES-D scales among this sample of older AIs. Additionally, the CESD scales’ significant direct moderate correlations with chronic pain and physical disability demonstrated concurrent validity and indirect moderate correlations with social support and self-efficacy demonstrated divergent validity in this population. Previous studies examining the psychometric properties of the CES-D scale have shown varying factor structures with the full version in certain populations (Chapleski et al., 1997; Kim et al., 2011; Perreira, Deeb-Sossa, Harris, & Bollen, 2005). Our findings support the reliability and validity of a two-factor solution of the full CES-D scale and one-factor solution with the abbreviated scale in older AIs.

These results can be contrasted to findings from Chapleski and colleagues (1997) who concluded the 3-factor and 4-factor models to be superior to a 2-factor model in the full scale. Results from our CFA model indicate better model fit for the 2-factor structure. Similarly, results from another sample of adult AIs indicated comparable fit between a 3-factor and 4-factor model of the CES-D (Somervell et al., 1992). Our study corroborates the findings from Chapleski and colleagues (1997) showing the reliability and validity of the 12-item scale.

The implications of our findings are limited in that participants from this study were aged ≥55 years who resided near or on tribal lands of one AI tribe in the Southeast. It is important to note that the AI/AN populations are extremely diverse with 566 federally recognized tribal entities (Bureau of Indian Affairs, 2016), and therefore, this study may only apply to a subset of AI/AN populations. Thus, caution should be exercised in generalizing our findings to other older AI/ANs. Future research with other older AI/AN populations would benefit from examining the psychometric properties of their mental health measures in order for the community stakeholders, investigators, and others who rely on this information to feel confident in the measures used. A new revised version of the CES-D measure (CESD-R) was developed but not yet validated among different populations by the time of this study (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004; Van Dam & Earleywine, 2011). Thus, we did not to test this new measure with our study population. Also, to our knowledge, only abbreviated versions of the original measure have been examined.
Substantial work remains with respect to moving our understanding from prevalence to identifying risk and protective factors of poor mental health in older AIs. Such research would ultimately contribute to the development or modification of intervention efforts. Ensuring that our measures of mental health produce consistent and accurate estimates is fundamental to this research. Future inquiry using this validated tool promises to enhance our understanding of mental health and ways to improve it in AI/AN populations.

REFERENCES

Acock, A. (2013). Discovering Structural Equation Modeling Using Stata. College Station, TX: StataCorp LP.


Stata Statistical Software. (2007). Version 12 [Computer software]. College Station, TX: StataCorp LP.


**ACKNOWLEDGEMENTS**

We would like to thank the tribe and its study participants for their role in making this study possible. The authors have no conflict of interest. This study was funded in part from the National Institute of Aging, NIH (# AG022336).

**AUTHOR INFORMATION**

Dr. Marc Schure is an assistant professor in the department of Health and Human Development at Montana State University in Bozeman, MT.

Dr. R. Turner Goins is an Ambassador Jeanette Hyde Distinguished Professor in the College of Health and Human Sciences at Western Carolina University in Cullowhee, NC.