Psychometric properties of a short form of the Center for Epidemiologic Studies Depression (CES-D-10) scale for screening depressive symptoms in healthy community dwelling older adults

Mohammadreza Mohebbi[a,b,*,1], Van Nguyen[a,j,1], John J. McNeil[c], Robyn L. Woods[c], Mark R. Nelson[d], Raj C. Shah[e], Elsdon Storey[c], Anne M. Murray[f,g], Christopher M. Reid[c,h], Brenda Kirpach[f], Rory Wolfe[c], Jessica E. Lockery[c], Michael Berk[b,c,i], on behalf of the ASPREE Investigator Group[2]

[a] Deakin University, Biostatistics Unit, Geelong, Australia
[b] Deakin University, School of Medicine, IMPACT Strategic Research Centre, Barwon Health, Geelong, Australia
[c] Department of Epidemiology & Preventive Medicine, Monash University, Melbourne, VIC, Australia
[d] Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia
[e] Department of Family Medicine and Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, United States
[f] Berman Center for Outcomes and Clinical Research, Minneapolis Medical Research Foundation, Hennepin County Medical Center, Minneapolis, MN, United States
[g] Division of Geriatrics, Department of Medicine, Hennepin County Medical Center and University of Minnesota, Minneapolis, MN, United States
[h] School of Public Health, Curtin University, Perth, WA, Australia
[i] Orygen, the National Centre of Excellence in Youth Health, the Department of Psychiatry and the Florey Institute for Neuroscience and Mental Health, University of Melbourne, Australia
[j] Deakin University, School of Nursing and Midwifery, Geelong, Australia

ARTICLE INFO

ABSTRACT

Background: The 10-item Center for the Epidemiological Studies of Depression Short Form (CES-D-10) is a widely used self-report measure of depression symptomatology. The aim of this study is to investigate the psychometric properties of the CES-D-10 in healthy community dwelling older adults.

Methods: The sample consists of 19,114 community-based individuals residing in Australia and the United States who participated in the ASPREE trial baseline assessment. All individuals were free of any major illness at the time. We evaluated construct validity by performing confirmatory factor analysis, examined measurement invariance across country and gender followed by evaluating item discrimination bias in age, gender, race, ethnicity and education level, and assessing internal consistency.

Results: High item–total correlations and Cronbach's alpha indicated high internal consistency. The factor analyses suggested a unidimensional factor structure. Construct validity was supported in the overall sample, and by country and gender sub-groups. The CES-D-10 was invariant across countries, and although evidence of marginal gender non-invariance was observed there was no evidence of notable gender specific item discrimination bias. No notable differences in discrimination parameters or group membership measurement non-invariance were detected by gender, age, race, ethnicity, and education level.

Conclusion: These findings suggest the CES-D-10 is a reliable and valid measure of depression in a volunteer sample. No noteworthy evidence of invariance and/or item discrimination bias is observed across gender, age, race, language and ethnic groups.

1. Introduction

A systematic review of depression prevalence in elderly populations showed that the prevalence of major depression ranges from 0.9% to 9.4% in private households and from 14% to 42% in institutional living; and the prevalence of clinically relevant depressive symptoms in similar settings varies between 7.2% and 49% [1]. Another systematic review on depression prevalence in later life (≥ 75 years) illustrated that the prevalence of major depression ranged from 4.6% to 9.3%, and that of depressive disorders from 4.5% to 37.4% [2]. Depression is a major

contributor to healthcare costs in older populations, and is projected to be the leading cause of disease burden in older populations by the year 2020 [3,4]. The prevalence of depression in patients aged ≥ 65 years may be as high as 40% in hospitalised and nursing home patients, and 8–15% in community settings [5].Depression in the elderly is associated with an increased risk of mortality, dementia and substantial psychosocial disability [6], resulting in an economic burden of $15 billion in Australia [7] and $83 billion in the United States [8].

The Center for Epidemiologic Studies Depression Scale (CES-D) has been widely used to assess depressive symptoms in community and population-based epidemiological studies [9]. The scale's validity and internal consistency in the detection of both clinical and non-clinical depressive symptoms have been established. It has however been suggested that the length of the 20-item CES-D could be halved without appreciable loss to reliability and validity. Various short and/or simplified forms of the 20-item CES-D have been evaluated [10–14]. The Boston form (10 dichotomously scored items), the Iowa form (11 items with three response options) developed by Kohout et al. [15] and the four-category response 10-item form (CES-D-10) developed by Andresen et al. [10] are most commonly used. The Andresen version, CES-D-10, has strong reliability and excellent sensitivity and specificity in screening for major depression in older adults [14]. Construct validity of the short form of the CES-D has been examined in Singaporean older adults in community settings [16], Chinese elderly in community dwelling [17] and older Chinese in social centres [18].While the published validity studies of the CES-D-10 illustrated acceptable factorial validity there were indications that the factorial structure has not been consistently determined. For example while studies among adults in Zulu, Xhosa and Afrikaans in South Africa [19] and the USA Hispanics population [20] concluded a one factor solution had the best model fit, studies in Canadian adolescents [21] and Singaporian elderly [16] resulted in a two-factor model and validation studies in older Chinese populations [17,18] reported two-factor and three-factor models of the CES-D-10 respectively. These contradictory findings may be due in part to the use of: i) individuals with different cultural background; ii) differences in study sample age ranges; iii) participant characteristics (e.g. a psychiatric sample as compared with community-based participants) or; iv) small sample size (sample size in the studies with factorial validation in elderly populations was 231, 742 and 1013 respectively (16–18)). In such situations, performing confirmatory factor analyses (CFA), a commonly approach for the evaluation of the construct validity of psychometric inventories, on a large sample of community-based elderly individuals with diverse ethnic and cultural backgrounds [22], is a unique opportunity to clarify this issue.

Reise, Widaman and Pugh [23] further recommend the use of measurement invariance tests within the CFA framework to examine the invariance of the instrument's psychometric properties across different groups. The goal of the present study was to investigate the internal consistency and construct validity of the CES-D-10, relying on a CFA approach in healthy community-dwelling older Australian and American adults who participated in the ASPirin in Reducing Events in the Elderly (ASPREE) trial [24]. ASPREE is a placebo-controlled trial of low-dose aspirin to determine whether 5 years of daily 100-mg enteric-coated aspirin extends disability-free and dementia-free life in a healthy elderly population and whether these potential benefits outweigh the risks. We also aimed to evaluate measurement invariance across the two countries and sexes and examine item-response bias analyses of the exogenous variables: age, gender, ethnicity, race and education.

2. Methods

2.1. Participants

This study included all 19,114 community-based individuals who participated in the baseline measurements of the ASPREE trial and were subsequently randomised. The participants were recruited from general practice services in Australia and community-based centres in the United States (U.S.). Recruitment ended in December 2014 with 16,703 Australian and 2411 American participants. Readers are referred to the work of the ASPREE Investigator Group [24] and Berk et al. [25] for details regarding the research settings, recruitment strategies, inclusion and exclusion criteria and ethical aspects of the study. In short, participants aged from 70 years old (Australians and U.S. non-(racial) minorities) or 65 years (U.S. - (racial) minorities) and were free of cardiovascular disease, dementia and physical disability. There were no exclusion criteria based on depressive symptoms. CES-D-10 overall score ranged from 0 (4277 cases) to 30 (2 cases) and 1906 (9.9%) of participants had CES-D-10 of 8 or above. Recruitment by age group was 65–74 years 11,163 (58%), 75–84 years 7219 (38%) and 85 + years 732 (4%), with 10,782 (56%) female. There were 1664 (9% of total cohort) US minority participants, of whom 54% (901) were African American and 29% (488) from the U.S. Latino/Hispanic population. A total of 10,477 (55%) had 12 years or more of formal education and 856 (4%) spoke a first language other than English. Further details on demographics and other baseline characteristics can be found in McNeil et al. [26].

2.2. Measures

The 10-item version of the Center for Epidemiologic Studies Short Depression Scale (CES-D-10) was used [10]. All items included four response categories indicating the frequency of depressive symptoms. Of the ten, eight items focussed on positive symptoms while the other two (items 5 and 8) assessed negative symptoms of depression. In brief, subjects responded to each item of the scale by rating the frequency of each mood or symptom 'during the past week' on a four-point scale. A score is assigned by totalling all items (after reversing the positive mood items).

2.3. Data analysis and results

We hypothesised a priori that representing depression by CES-D-10 score (depression score) can be explained by a single first-order factor. This model was compared with various alternative models. The single factor CFA was first estimated on all participants. Hu and Bentler's [27] and Hair et al.'s (2010) guidelines for model fit indices' cut-offs were used. In particular, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Goodness of Fit Index (GFI) above 0.95 were taken to manifest a good level of model fit. A Root Mean Squared Error of Approximation (RMSEA) value of 0.06 or lower and Standardised Root Mean Squared Residual (SRMR) < 0.09 were considered to indicate a satisfactory fit.

The CES-D-10's internal consistency was assessed using both Cronbach's alpha and composite reliability. Cronbach's alphas were obtained from factor analysis. Composite reliability was calculated from the squared sum of standardised factor loadings divided by the total of the squared sum of standardised factor loadings and the sum of error variance for a factor [28]. A threshold of 0.7 for both reliability coefficients was used to indicate the consistency of all items to measure a factor [28,29].

Measurement invariance tests were utilised to examine the invariance of the CES-D-10 between male and female participants, and between Australia and America. A series of nested hierarchies of hypotheses within the CFA framework was tested to address the cross-group invariance of the CES-D-10. As suggested by Meade et al. [30] a cut off of 0.002 or lower for absolute differences in CFI (|ΔCFI|, i.e. differences in CFI obtained when an unconstrained model was compared with a model with measurement invariance constraints) was used