BRIEF REPORT

Validation of the 8-item Centre for Epidemiological Studies Depression Scale in a cohort of community-dwelling older people: data from The Irish Longitudinal Study on Ageing (TILDA)

R. Briggs[1,2,3] · D. Carey[1] · A. M. O'Halloran[1] · R. A. Kenny[1,2] · S. P. Kennelly[3]

Abstract

Background The aim of this study was to validate the 8-item Centre for Epidemiological Studies Depression Scale (CES-D-8) against the 20-item version (CES-D-20) in a large sample of community-dwelling older people.

Methods Scales were compared for correlation and internal consistency. The ideal cut-off score for the CES-D-8 was determined by comparing scores ranging from 7 to 12 on the CES-D-8 to CES-D-20.

Results 8033 participants were included. The Spearman co-efficient between the scales was 0.8980 indicating high degree of correlation. At a score of 9/24, the sensitivity and specificity of the CES-D-8 were 98 and 83%, respectively. The Cohen's κ for a score of 9 was 0.7855, indicating strong agreement and the ROC area was 0.88.

Conclusion When compared to the CES-D-20, the CES-D-8 is a valid and reliable measure of depressive symptoms in community-dwelling older people, and a score of 9 can be used to identify those with clinically significant symptoms.

Keywords Depression · CES-D · Validation

Introduction

The Centre for Epidemiological Studies Depression Scale (CES-D) is one of the most commonly employed scales used to identify depression in large population studies involving older people [1]. It is free to use and has been validated in an older cohort against structured diagnostic interviews [2, 3].

However, concerns that the 20-item Centre for Epidemiological Studies Depression Scale (CES-D-20) is too time consuming and onerous, especially when embedded within large surveys involving older people, have led to the development of short forms of the scale. While the 10-item Centre for Epidemiological Studies Depression Scale (CES-D-10) has been studied widely [4–6], the 8-item Centre for Epidemiological Studies Depression Scale (CES-D-8) is now increasingly employed in large international studies, including the European Social Survey and the Health and Retirement Study [7, 8]. Previous work has confirmed the reliability of the CES-D-8 in comparison to the CES-D-20 in terms of internal consistency and factor structure, as well as correlations with disability and perceived stress, but as yet, no appropriate cut-off value to define depression cases via the CES-D-8 has been identified [9].

The aim of this study, therefore, is to validate the shortened CES-D-8 scale against the original CES-D-20 in a large sample of community-dwelling older people, to identify an appropriate cut-off value to define cases with clinically significant depressive symptoms.

✉ R. Briggs
briggsr@tcd.ie

1   The Irish Longitudinal Study on Ageing (TILDA), Trinity College Dublin, Dublin, Ireland

2   Mercer's Institute for Successful Ageing, St James's Hospital, Dublin, Ireland

3   Age-related Health Care, Tallaght Hospital, Dublin, Ireland

Methods

Study design

We analysed data from Waves 1 and 3 of The Irish Longitudinal Study on Ageing (TILDA), a study of a nationally representative sample of community dwelling adults aged 50 years and over [10]. Participants were included if they underwent a complete assessment at Wave 1, including a CES-D, as well as collection of other biological and social data.

CES-D

The CES-D-20 consists of 20 items on a Likert scale, yielding a total score of from 0 to 3 for each item [11]. A score of 16 or more on the CES-D-20 is used to define clinically significant depressive symptoms [1].

The CES-D-8 consists of 8 items taken from this original CES-D-20. These items are shown in "Appendix A". The range of the CES-D-8 is, therefore, 0–24. The CES-D-8 was repeated at Wave 3 (4-year follow-up) to assess retest reliability.

Other measures

Relevant biological and sociological data were also collected by self-report at Wave 1, including marital status, third level educational attainment, cardiovascular disease (myocardial infarction, angina, cardiac arrhythmia, or cardiac failure), chronic pain, functional impairment, diabetes and prior stroke.

Statistical analysis

Correlation between the two scales was assessed using Spearman's rank correlation coefficient. This was used in preference to Pearson's correlation as the CES-D scales are ordinal and results are not normally distributed. Spearman's correlation was also used to assess retest reliability at 4-year follow-up.

Cronbach's α was used to measure internal consistency of the scale. Factor analysis with varimax rotation was used to demonstrate the latent structure of the scale.

Due to the clustered nature of household sampling in TILDA, a cross-validation approach was used whereby analyses were repeated on smaller samples containing one participant from each cluster, and results were compared. For example, when analysing all clusters with a minimum of five households, we initially selected clusters that comprised at a minimum six unique respondent households. Next, we drew a random sample of five participants from that cluster—since each sampled participant was an inhabitant from a unique household, the sample did not include any cohabiting respondents (in order to reduce bias from within-household correlation of data). We then took the first sampled individual from each cluster (cluster $N =$ here) as an observation, and re-ran analysis across the subset of $n =$ (here) individuals. We then repeated this process four further times, allowing each remaining sampled individual to serve as a data point for the cluster they were sampled from. In this way, the analysis provided multiple estimates of Cronbach's α, Spearman's ρ and the factor structure that allowed the reliability of data

Table 1 Baseline characteristics of study population

	Not depressed ($n = 7259$)	Depressed ($n = 774$)	
Mean age in years (95% CI)	63.9 (63.7–64.1)	62.5 (61.8–63.1)	$t = 3.9057$, $p < 0.001$
Female (%)	53%	67%	$\chi^2 = 60.3108$, $p < 0.001$
Married (%)	71%	54%	$\chi^2 = 93.1594$, $p < 0.001$
Third level education (%)	30%	23%	$\chi^2 = 15.0575$, $p < 0.001$
Functional impairment (%)	6%	22%	$\chi^2 = 285.3378$, $p < 0.001$
Chronic pain (%)	18%	47%	$\chi^2 = 385.9011$, $p < 0.001$
Cardiac disease (%)	43%	50%	$\chi^2 = 15.0216$, $p < 0.001$
Prior stroke (%)	1%	3%	$\chi^2 = 20.0271$, $p < 0.001$
Diabetes (%)	8%	10%	$\chi^2 = 5.1679$, $p = 0.023$

Depression diagnosis based on a score ≥ 16 on 20-item Centre for Epidemiological Studies Depression Scale

Categorical variables analysed by Chi-square test, binary variables analysed using Student t test

Functional impairment is defined as one or more impairments in instrumental activities of daily living, including cleaning and maintaining the house, managing money, moving within the community and preparing meal

Chronic pain, prior stroke and diabetes are defined by self-report of these conditions

Cardiovascular disease is defined as self-report of angina, congestive cardiac failure, arrhythmia or myocardial infarction

CI confidence interval

to be verified across multiple samples from separate clusters.

For the full sample analysis, the appropriate cut-off score for the CES-D-8 was determined by comparing scores ranging from 7 to 12 against the CES-D-20 cut-off score of 16. Sensitivity, specificity, agreement and receiver operator characteristics were estimated for each of these potential cut-off values. Agreement was measured by percent agreement calculation, as well as by Cohen's κ statistic.

To confirm that depression defined by the CES-D-8 had similar associations with depression defined by the CES-D-20 with respect to variables of interest, binary logistic regression models, reporting odds ratios, were used and compared. This analysis was completed for the full sample analysed.

Results

8033 participants were included in the study. 10% (774/8033) scored ≥ 16 on the CES-D-20 and were therefore defined as having clinically significant depressive symptoms.

The baseline characteristics of the study sample are shown in Table 1.

Agreement between scales

The Spearman correlation co-efficient between the CES-D-20 and the CES-D-8 was 0.8980 (p value < 0.001), indicating a high degree of correlation between the scales. The CES-D-20 and CES-D-8 both also showed excellent internal consistency with a Cronbach α of 0.8757 and 0.8127, respectively.

Table 2 Performance of different cut-off scores of CES-D-8 against CES-D-20 with cut-off score 16

Cut-off score	Prevalence	Sensitivity (%)	Specificity (%)	% agreement	κ	ROC area (95% CI)
≥ 7	15% (1196/8033)	99	61	93.7	0.7091	0.94 (0.93–0.95)
≥ 8	12% (958/8033)	99	71	95.3	0.7557	0.92 (0.91–0.93)
≥ 9	9% (733/8033)	98	83	96.4	0.7855	0.88 (0.87–0.90)
≥ 10	7% (554/8033)	96	90	95.9	0.7331	0.82 (0.80–0.84)
≥ 11	5% (429/8033)	95	96	95.3	0.6617	0.77 (0.75–0.78)

CES-D-8 8-item Centre for Epidemiological Studies Depression Scale, *CES-D-20* 20-item Centre for Epidemiological Studies Depression Scale, *% agreement* percentage agreement between CES-D8 and CES-D-20, *κ* Cohen's κ statistic, *ROC* receiver operating characteristic, *CI* confidence interval

Table 3 Linear regression, comparing odds ratios for predictor variables for depression by CES-D-20 and CES-D-8

	Depression by CES-D-20 (score ≥ 16)			Depression by CES-D-8 (score ≥ 9)		
	OR (95% CI)	p value	z	OR (95% CI)	p value	z
Age ≥ 75 years	0.41 (0.32–0.53)	< 0.001	− 7.02	0.47 (0.37–0.60)	< 0.001	− 6.06
Female	1.58 (1.34–1.87)	< 0.001	5.46	1.56 (1.32–1.84)	< 0.001	5.14
Married	0.49 (0.42–0.58)	< 0.001	− 8.50	0.52 (0.44–0.61)	< 0.001	− 7.74
3rd level education	0.78 (0.65–0.94)	0.007	− 2.68	0.82 (0.68–0.98)	0.033	− 2.13
Functional imp.	2.91 (2.31–3.65)	< 0.001	9.13	3.29 (2.62–4.14)	< 0.001	10.24
Chronic pain	3.22 (2.72–3.80)	< 0.001	13.76	3.00 (2.53–3.560)	< 0.001	12.62
Cardiac disease	1.14 (0.97–1.34)	0.117	1.57	1.17 (0.99–1.38)	0.061	1.87
Stroke	1.67 (1.02–2.73)	0.040	2.06	1.49 (0.90–2.47)	0.118	1.56
Diabetes	1.13 (0.86–1.49)	0.368	0.90	1.30 (0.99–1.70)	0.060	1.88

$N = 8030$

CES-D-20 20-item Centre for Epidemiological Studies Depression Scale, *CES-D-8* 8-item Centre for Epidemiological Studies Depression Scale, *OR* odds ratio, *CI* confidence interval, *Functional imp.* functional impairment

Functional impairment is defined as one or more impairments in instrumental activities of daily living, including cleaning and maintaining the house, managing money, moving within the community and preparing meal

Chronic pain, prior stroke and diabetes are defined by self-report of these conditions

Cardiovascular disease is defined as self-report of angina, congestive cardiac failure, arrhythmia or myocardial infarction

The random samples in our cross-validation analysis showed a range of values for Cronbach's α all of which were close to or above 0.8, indicating high consistency. Similarly the range of values obtained for Spearman's ρ were all close to 0.9. See "Appendix B".

Factor analysis

Factor analysis revealed the internal structure of the CES-D-8. Two factors with eigenvalues ≥ 1 were identified, and they accounted for 58% of the total variance of the data. The six 'negative affect' items of the CES-D-8 loaded on to Factor 1, while the remaining two 'positive affect' items loaded on to Factor 2. See "Appendix C". Factor analysis remained consistent when analysed iteratively using participants selected as a random sub-sample from each cluster (see "Methods").

Cut-off scores

Table 2 shows the performance of the CES-D-8 using different cut-off scores compared to the CES-D 20 (with a cut-off score for 16) for diagnosis of clinically relevant depressive symptoms. At a cut-off score of 9/24, the sensitivity and specificity of the 8-item CES-D were 98 and 83%, respectively. The Cohen's κ for a cut-off score of 9 was 0.7855, suggestive of strong agreement and the ROC area was adequate at 0.88.

Table 3 shows binary logistic regression models, comparing the relationship of some variables of interest with depression diagnosis by CES-D-20 (at a cut off score of 16) and CES-D-8 (at a cut off score of 9). Results are broadly similar across scales, with the CES-D-8 demonstrating similar relationships with social and biological factors such as sex, educational attainment and chronic medical conditions.

Test retest reliability

Three quarters (6013/8033) of the study sample had the CES-D-8 repeated at 4-year follow-up. The Spearman's correlation coefficient for the CES-D-8 performed at baseline and at 4 years was 0.4239 (p value < 0.001), suggesting moderate correlation.

Discussion

This study demonstrates that the CES-D-8 correlates well with the CES-D-20 when administered to a cohort of community-dwelling older adults.

At a cut-off score of 9 the CES-D-8 accurately identifies clinically significant depressive symptoms in this cohort

Table 5 Factor analysis of CES-D-8

Factor	Eigenvalue		Variation
Factor 1	2.91		0.36
Factor 2	1.76		0.22

CES-D-8 Item	Factor loading	
	Factor 1	Factor 2
I felt depressed	0.7022	− 0.3196
I felt everything I did was an effort	0.7507	− 0.1775
My sleep was restless	0.5726	− 0.1158
I was happy	− 0.1882	0.8435
I felt lonely	0.6454	− 0.2504
I enjoyed life	− 0.1615	0.8592
I felt sad	0.7002	− 0.2745
I could not get 'going'	0.7475	− 0.1530

CES-D-8 8-item Centre for Epidemiological Studies Depression Scale

Table 4 Results from cross-validation using samples derived from unique clusters

	Cronbach α	Spearman ρ
	Analysis of 20 groups each containing 58 participants from unique clusters	
Mean (95% CI)	0.787265 (0.76–0.81)	0.88791 (0.87–0.91)
Median	0.78285	0.8947
IQR	0.0708	0.0652
	Analysis of ten groups, each containing 365 participants from unique clusters	
Mean (95% CI)	0.80431 (0.79–0.82)	0.89676 (0.89–0.91)
Median	0.8035	0.89945
IQR	0.0392	0.0172
	Analysis of five groups, each containing 520 participants from unique clusters	
Mean (95% CI)	0.8201 (0.80–0.84)	0.89710 (0.89–0.90)
Median	0.8235	0.8966
IQR	0.0138	0.0017

CI confidence interval, *IQR* interquartile range

when validated against the CES-D-20 with a sensitivity and specificity of 98 and 83%, respectively. Logistic regression models offer further validation, with similar odds ratios demonstrated for variables of interest, such as older age or chronic medical conditions, when comparing the CES-D-8 and CES-D-20 as dependent variables.

The CES-D-8 has previously been shown to have measurement equivalence across different ages, gender and countries [8, 12] and specifically in older people [13]. However, while cut-off scores of 16 and 10 have been established for the CES-D-20 and CES-D-10 [4], respectively, indicating a threshold for clinically significant depressive symptoms, this is the first study to identify an appropriate cut-off score to define cases of clinically significant depressive symptoms using the CES-D-8.

Validated shortened scales allow more rapid assessment and reduce respondent burden in large surveys. While it is often preferable to analyse continuous data from scales such as these, using a defined cut-off score to dichotomize cases into either a depressed or non-depressed group can also be useful as it mirrors decision-making in clinical practice, and allows us to estimate incidence and prevalence figures.

We also show that the internal structure of the CES-D-8 is similar to that of the CES-D-10, with two factors explaining almost 60% of the variance in the data [4]. This is consistent with previous comparisons of factor structure between the CES-D-8 and CES-D-20 [9].

There are some limitations to this study which must be noted. While the CES-D-20 has previously been validated against a structured psychiatric interview, we did not compare the CES-D-8 to this gold standard and a further study validating the CES-D-8 against clinically diagnosed depression would be welcome.

In conclusion, this study demonstrates that when compared to the 20-item CES-D, the 8-item CES-D is a valid and reliable measure of depression symptoms in community-dwelling older adults, and that a cut-off score of 9 can be used to identify those with clinically significant symptoms.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Ethical approval The study was approved by the Faculty of Health Sciences Research Ethics Committee at Trinity College Dublin. All experimental procedures adhered to the Declaration of Helsinki.

Informed consent All participants gave informed written consent.

Appendices

Appendix A: 20-item CES-D items (with 8-item CES-D items indicated by asterisk)

I was bothered by things that usually don't bother me
I did not feel like eating; my appetite was poor
I felt that I could not shake off the blues with help from my family/friends
I felt I was just as good as other people
I had trouble keeping my mind on what I was doing
I felt depressed*
I felt everything I did was an effort*
I felt hopeful about the future
I felt my life had been a failure
I felt fearful
My sleep was restless*
I was happy*
I talked less than usual
I felt lonely*
People were unfriendly
I enjoyed life*
I had crying spells
I felt sad*
I felt people disliked me
I could not 'get going'*

Appendix B

See Table 4.

Appendix C

See Table 5.

References

1. Vilagut G, Forero CG, Barbaglia G, Alonso J (2016) Screening for depression in the general population with the center for epidemiologic studies depression (CES-d): a systematic review with meta-analysis. PloS One 11(5):e0155431
2. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W (1997) Criterion validity of the center for epidemiologic studies depression scale (CES-d): results from a community-based sample of older subjects in the netherlands. Psychol Med 27(1):231–235

3. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB (1997) Center for epidemiologic studies depression scale (CES-d) as a screening instrument for depression among community-residing older adults. Psychol Aging 12(2):277–287

4. Zhang W, O'Brien N, Forrest JI, Salters KA, Patterson TL, Montaner JSG et al (2012) Validating a Shortened Depression Scale (10 Item CES-D) among HIV-Positive People in British Columbia, Canada. PLoS One 7(7):e40793

5. Lee AE, Chokkanathan S (2008) Factor structure of the 10-item CES-d scale among community dwelling older adults in singapore. Int J Geriatr Psychiatry 23(6):592–597

6. González P, Nuñez A, Merz E, Brintz C, Weitzman O, Navas EL et al (2017) Measurement properties of the center for epidemiologic studies depression scale (CES-d 10): findings from HCHS/SOL. Psychol Assess 29(4):372–381

7. Zivin K, Llewellyn DJ, Lang IA, Vijan S, Kabeto MU, Miller EM et al (2010) Depression among older adults in the United States and England. Am J Geriatr Psychiatry 18(11):1036–1044

8. Van de Velde S, Bracke P, Levecque K, Meuleman B (2010) Gender differences in depression in 25 European countries after eliminating measurement bias in the CES-D 8. Soc Sci Res 39(3):396–404

9. O'Halloran AM, Kenny RA, King-Kallimanis BL (2014) The latent factors of depression from the short forms of the CES-D are consistent, reliable and valid in community-living older adults. Eur Geriatr Med 5:97–102

10. Whelan BJ, Savva GM (2013) Design and methodology of the irish longitudinal study on ageing. J Am Geriatr Soc 61(Suppl):2

11. Carleton RN, Thibodeau MA, Teale MJ, Welch PG, Abrams MP, Robinson T et al (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+

12. Missinne S, Vandeviver C, Van de Velde S, Bracke P (2014) Measurement equivalence of the CES-d 8 depression-scale among the ageing population in eleven european countries. Soc Sci Res 46:38–47

13. Karim J, Weisz R, Bibi Z, ur Rehman S (2015) Validation of the eight-item Center for Epidemiologic Studies Depression Scale (CES-D) among older adults. Curr Psychol 34(4):681–692