



A clinically feasible short version of the 15-item geriatric depression scale extracted using item response theory in a sample of adults aged 85 years and older

Sanna Johansson, Hugo Lövheim, Birgitta Olofsson, Yngve Gustafson & Johan Niklasson

To cite this article: Sanna Johansson, Hugo Lövheim, Birgitta Olofsson, Yngve Gustafson & Johan Niklasson (2021): A clinically feasible short version of the 15-item geriatric depression scale extracted using item response theory in a sample of adults aged 85 years and older, *Aging & Mental Health*, DOI: [10.1080/13607863.2021.1881759](https://doi.org/10.1080/13607863.2021.1881759)

To link to this article: <https://doi.org/10.1080/13607863.2021.1881759>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Feb 2021.



Submit your article to this journal [↗](#)



Article views: 444



View related articles [↗](#)



View Crossmark data [↗](#)

A clinically feasible short version of the 15-item geriatric depression scale extracted using item response theory in a sample of adults aged 85 years and older

Sanna Johansson^a, Hugo Lövheim^b, Birgitta Olofsson^c, Yngve Gustafson^b and Johan Niklasson^a

^aDepartment of Community Medicine and Rehabilitation, Geriatric Medicine, Sunderby Research Unit, Umeå University, Umeå, Sweden;

^bDepartment of Community Medicine and Rehabilitation, Geriatric Medicine, Umeå University, Umeå, Sweden; ^cDepartment of Nursing, Umeå University, Umeå, Sweden

ABSTRACT

Objectives: To extract the items most suitable for a short version of the 15-item Geriatric Depression Scale (GDS-15) in a sample of adults aged ≥ 85 years using item response theory (IRT).

Method: This population-based cross-sectional study included 651 individuals aged ≥ 85 years from the Umeå 85+/Gerontological Regional Database (GERDA) study. Participants were either community dwelling (approximately 70%) or resided in institutional care (approximately 30%) in northern Sweden and western Finland in 2000–2002 and 2005–2007. The psychometric properties of GDS-15 items were investigated using an IRT-based approach to find items most closely corresponding to the GDS-15 cut off value of ≥ 5 points. Receiver operating characteristic curves were used to compare the performance of the proposed short version with that of previously proposed short GDS versions.

Results: GDS-15 items 3, 8, 12, and 13 best differentiated respondents' levels of depressive symptoms corresponding to the GDS-15 cut off value of ≥ 5 , regardless of age or sex, and thus comprise the proposed short version of the scale (GDS-4 GERDA). For the identification of individuals with depression (total GDS-15 score ≥ 5), the GDS-4 GERDA with a cut-off score of ≥ 2 had 92.9% sensitivity and 85.0% specificity.

Conclusion: The GDS-4 GERDA could be used as an optimized short version of the GDS-15 to screen for depression among adults aged ≥ 85 years.

ARTICLE HISTORY

Received 29 August 2020
Accepted 17 January 2021

KEYWORDS

Aged; 80 and over; depression; item response theory; psychometrics; ROC curve; psychiatric status rating scales

Introduction

Depression is a leading cause of disability worldwide, and it is ranked first in the burden of disease in middle- and high-income countries (World Health Organization, 2008). Depressive disorders negatively affect well-being (Bergdahl et al., 2005) and daily functioning (Beekman, Copeland, & Prince, 1999), and are among the most common mental disorders in older adults (Fiske, Wetherell, & Gatz, 2009), although they often are underdiagnosed (Stek, Gussekloo, Beekman, Van Tilburg, & Westendorp, 2004) and not properly treated (Bergdahl et al., 2005). The prevalence of depression has been estimated to be 20–25% among people aged ≥ 85 years and 30–50% among those aged ≥ 90 years (Luppa et al., 2012).

The Geriatric Depression Scale (GDS), a standardized scale developed in 1983 by Yesavage et al. (Yesavage et al., 1982), is reliable for depression screening in older adults. It originally consisted of 30 items (Yesavage et al., 1982), but the 15-item version (GDS-15) (Sheikh & Yesavage, 1986) is now used commonly worldwide. Various short forms of the GDS have been proposed (Weeks, McGann, Michaels, & Penninx, 2003), such as the D'Ath GDS-4 (D'Ath, Katona, Mullan, Evans, & Katona, 1994), van Marwijk GDS-4 (van Marwijk et al., 1995), Hoyl GDS-5 (Hoyl et al., 1999), and

Wongpakaran GDS-6 (Wongpakaran, Wongpakaran, & Kuntawong, 2019).

Item response theory (IRT) is a psychometric statistical modelling (Chiesi et al., 2017) that can be used to compare different scale items relative to each other and depending on the level of a latent trait, which is a variable that cannot be measured directly (e.g. depression or pain) (Thomas, 2011). By analyzing individuals' responses to scale items, the level of a variable of interest (e.g. depressive symptoms) at which a specific item usually starts to render points can be extracted (Chiesi et al., 2017). IRT analyses are used to calculate item parameters such as difficulty and discrimination. For scales measuring depression, item difficulty represents the level of depression at which half of the population scores points on a particular item, and item discrimination describes the strength of the relationship between the item and the depressive symptoms (Thomas, 2011).

Previous short forms of the GDS-15 have been extracted using Classical Test Theory except Wongpakaran et al. (Wongpakaran et al., 2019) who used a complex combination of IRT analysis and Confirmatory Factor Analysis. To our knowledge no one has used item difficulty from the IRT analysis to find items closest to the cut-off of the original GDS-15 version. Using the items closest to the cut-off would probably suffice to identify individuals with or

without depression and render the other items redundant. Further, to our knowledge no short version of GDS-15 have been tested among adults ≥ 85 years of age. Thus, the aim of this study was to use IRT analysis to determine which GDS-15 items best differentiate at the level of depressive symptoms corresponding to the original cut-off value of ≥ 5 points in a sample of adults aged ≥ 85 years, and to investigate whether these items can serve as a feasible, sensitive, and specific shorter version of the GDS for this population.

Methods

Setting and data source

The data used in this study were collected as part of the Umeå 85+/GERontological Regional DAtabase (GERDA) study, a population-based cohort study performed by researchers from Umeå University, Sweden, in collaboration with Åbo Akademi, Vaasa University, and Novia University of Applied Sciences, Finland. In 2000–2002, participants from Västerbotten county, northern Sweden, were recruited. Those who participated in 2000–2002 were asked to participate again in 2005–2007. In addition, new participants were recruited from the same area and the study was also expanded to the Ostrobothnia region of western Finland. Every second person aged 85 years, and all persons aged 90 and ≥ 95 years registered in the Swedish National Tax Board and the Finnish Population Register were invited to participate in the GERDA study. Ethical approval was obtained from the regional Ethics Review Board in Sweden (nos. 99-326, 05-063 M, 09-178 M, and 14-221-31 M), and in Finland (nos. 05-87 and 10-54).

Participants

Of 1489 individuals who were eligible for participation during these periods, 524 were excluded due to death before contact or refusal of study participation. For individuals who participated at both timepoints, only data from 2005–2007 (when participants were oldest) were used. The remaining 965 individuals accepted interview and answered at least 14 of the 15 GDS-15 questions. Of these, 651 were cognitively assessed using the Mini-Mental State Examination, no exclusion criteria were used, and they constituted the final sample.

Procedure

Eligible participants were first sent letters with information about the study and later telephoned by a research assistant to obtain informed consent. In cases of concern about individuals' ability to provide informed consent (e.g. due to cognitive capacity), consent was discussed with the individuals' next of kin. Home visits (ca. 2 h) were then conducted, with health check-ups and structured interviews. The interviews were performed by specially trained physicians, physiotherapists, nurses or medical students. Participation was voluntary and could be terminated at any time. Participants' medical records and information from relatives and caregivers were used to validate information collected during interviews.

Assessments

The GDS-15 has been shown to have high sensitivity and specificity for the detection of depression in older adults when used with the recommended cut-off of ≥ 5 points to indicate depression (Pocklington, Gilbody, Manea, & McMillan, 2016). Item response options are "yes" and "no"; answers indicating depression was given a score of 1 and overall scores range from 0 to 15. The MMSE was used to assess cognitive status. Total MMSE scores range from 0 to 30, with scores ≤ 17 indicating severe cognitive impairment and scores of 18–23 indicating mild cognitive impairment (Tombaugh & McIntyre, 1992). The GDS-15 has previously been shown to be feasible and valid among old adults with cognitive impairment (Conradsson et al., 2013). The Barthel index was used to assess independence in activities of daily living; scores range from 0 to 20, with higher scores indicating greater independence (Mahoney & Barthel, 1965). Impaired vision was defined as the inability to read 5-mm-high letters from a normal reading distance, with or without glasses. Participants who could not hear normal conversation at a distance of 1 m, with or without a hearing aid, were considered having impaired hearing. Participants who did not live with partners or other close relatives were considered to be living alone. Living in institutional care was defined by residence in a residential care facility, nursing home, or group home for people with dementia.

Statistical analyses

Differences between participants with total GDS-15 scores < 5 and ≥ 5 were evaluated using the chi-squared and independent-samples *t* tests. *P* values < 0.05 were considered to indicate significance. To identify items most suitable for a shorter version of the GDS, a two-parameter IRT analysis based on binary logistic regression was performed for each GDS-15 item using equations provided by Lövhheim et al. (Lövhheim, Gustafsson, Isaksson, Karlsson, & Sandman, 2019). The dependent variable was the item score and the independent variable was the total GDS-15 score. Item discrimination and difficulty were calculated as item β and minus constant β /item β , respectively. Item characteristic curves (ICCs) were drawn using the equation $y = 1 / \{1 + e^{\lambda[-(\text{constant } \beta + \text{item } \beta \times x)]}\}$ (Lövhheim et al., 2019). IRT analyses were also performed for subgroups defined by age, sex, and cognitive function to identify any related difference in item suitability for a shorter GDS version. The rationale for the selection of four items for the short version of the GDS was that this approach would allow for optimal sensitivity and specificity while minimizing administration time for clinical practice. Using the full scale cut-off score of ≥ 5 points as the gold standard, receiver operating characteristic (ROC) curves were drawn, and sensitivity and specificity of GDS-4-GERDA compared to GDS-15 were calculated. Cronbach's α was calculated and was used to compare the internal consistency of the full scale and the scale with individual items omitted. Cronbach's alpha values vary between 0 and 1. Alpha below 0.7 usually indicates poor internal consistency and values above 0.9 suggests items are very similar and perhaps fewer items could lead to similar results (Peacock & Peacock, 2011). Corrected item-to total-

Table 1. Characteristics of the participants and comparison with previous studies and of subgroups depending on depression being indicated or not according to total GDS-15 score with cut off of ≥ 5 points as indicating depression.

	Present study				Other versions suggested in previous research			
	Total	GDS < 5	GDS ≥ 5	<i>p</i> -value	D'Ath-4	Van Marwijk-4	Hoyl-5	Wongpakaran-6
Number of people, <i>n</i>	651	453	198		198	586	74	803
Age, mean \pm SD (range)	89.9 \pm 4.4 (84-104)	89.5 \pm 4.3	90.8 \pm 4.7	<i>p</i> < 0.001	73.5 (65-92)	73.5 (65-94)	74.6 (65-87)	69.2 (60-89)
Women, <i>n</i> (%)	438 (67.3)	293 (64.7)	145 (73.2)	<i>p</i> = 0.040	(63.6 %)	(59.6%)	(2.0%)	(70.0%)
MMSE, mean \pm SD	22.3 \pm 5.4	23.1 \pm 5.3	20.7 \pm 5.4	<i>p</i> < 0.001				
Bartell ADL index, mean \pm SD*	17.2 \pm 4.4	17.8 \pm 4.0	15.8 \pm 5.0	<i>p</i> < 0.001				
GDS-15 score, mean \pm SD (range)	3.7 \pm 2.6 (0-14)	2.3 \pm 1.2	7.0 \pm 2.1	<i>p</i> < 0.001				
Education †								
≤ 7 years, <i>n</i> (%)	459 (74.0)	317 (74.1)	142 (74.0)	<i>p</i> = 1.000				
≥ 8 years, <i>n</i> (%)	161 (26.0)	111 (25.9)	50 (26.0)					
Loss of child, <i>n</i> (%) ‡	130 (21.1)	74 (17.4)	56 (29.5)	<i>p</i> < 0.001				
Living alone, <i>n</i> (%) §	526 (80.9)	350 (77.4)	176 (88.9)	<i>p</i> < 0.001				
Living in institutional care, <i>n</i> (%)	217 (33.3)	131 (28.9)	86 (43.4)	<i>p</i> < 0.001				
Experienced loneliness, <i>n</i> (%)	302 (47.9)	162 (36.6)	140 (74.9)	<i>p</i> < 0.001				
Number of medications, mean \pm SD **	6.6 \pm 4.3	5.9 \pm 4.1	8.5 \pm 4.3	<i>p</i> < 0.001				
Diabetes, <i>n</i> (%)**	77 (11.9)	48 (10.6)	29 (14.8)	<i>p</i> = 0.169				
Previous stroke, <i>n</i> (%)**	92 (14.2)	58 (12.8)	34 (17.3)	<i>p</i> = 0.165				
Heart disease of any kind, <i>n</i> (%)**	418 (64.5)	276 (61.1)	142 (72.4)	<i>p</i> = 0.007				
Impaired hearing, <i>n</i> (%) ††	114 (17.7)	71 (15.8)	43 (22.1)	<i>p</i> = 0.075				
Impaired reading vision, <i>n</i> (%)**	117 (18.1)	61 (13.6)	56 (28.6)	<i>p</i> < 0.001				

SD, standard deviation; GDS, Geriatric depression scale (score range 0-15, ≥ 5 indicating depression); AD, activities of daily living (score range 0-20, 20 points representing total independence in ADL); MMSE, mini-mental state examination (score range 0-30, high scores indicating better cognitive function). Almost all participants were of Caucasian race, with a few native Sapmi (formerly "laplander") descent.

*Five cases missing.

†Thirty-one cases missing.

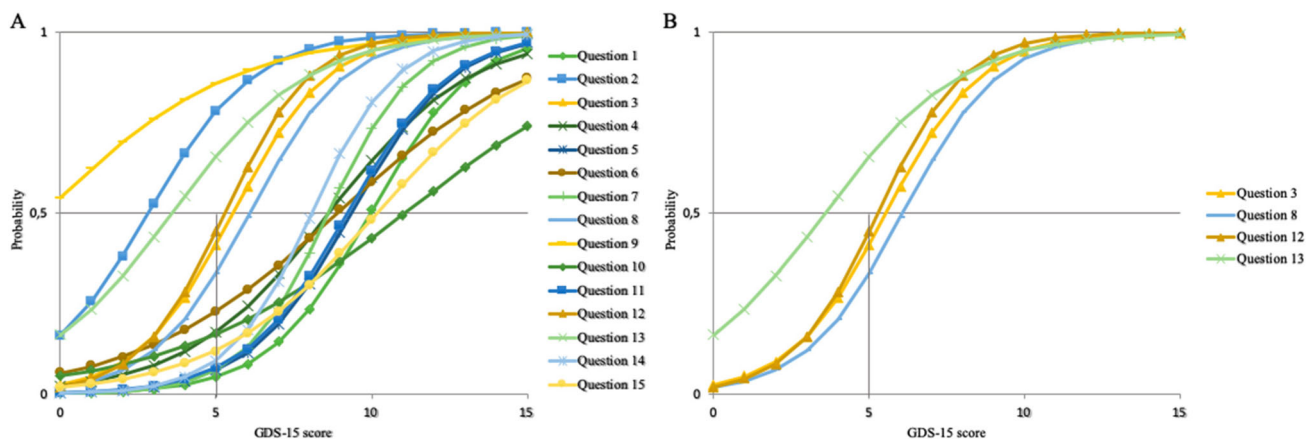
‡Thirty-five cases missing.

§One case missing.

||Twenty-one cases missing.

**3 missing.

††8 missing.

**Figure 1.** Item characteristic curves illustrating the results of the two-parameter item response theory analysis of A) every item of GDS-15 and B) of the items 3, 8, 12 and 13 which were included in the suggested short version, GDS-4 GERDA. The vertical line representing the 15-item Geriatric depression scale (GDS-15) cut off of ≥ 5 points as indicating depression.

correlations were also determined where values below 0.2 would indicate that the item might measure something different from the scale as a whole (Streiner & Norman, 2008). All statistical analyses were performed using IBM SPSS Statistics (IBM Corp., Armonk, NY, version 26 for Macintosh), and figures were drawn using Microsoft® Excel® (version 16.9 for Macintosh).

Results

The study sample comprised 651 participants with a mean age of 89.9 ± 4.4 years. Most participants were women (67.3%), lived alone (80.9%), and did not live in institutional care facilities (66.7%). The mean GDS-15 score was 3.7 ± 2.6 (range, 0–14) and the mean MMSE score was 22.3 ± 5.4 . Compared with participants with GDS-15 scores < 5, those considered to be depressed were older, used more

medications, and were more likely to live in institutional care facilities or alone and to report experiencing loneliness. Depressed participants also had lesser cognitive function and functional levels and were more likely to have impaired reading vision, experience of the loss of a child, and heart disease (Table 1). Excluded individuals did not differ from participants in age, but the excluded sample contained more women than did the included sample (74.2% vs. 67.3%; $p < 0.01$).

IRT analysis results

ICCs are provided in Figure 1a, and item difficulty and discrimination are summarized in Table 2. The difficulty of the GDS-15 items ranged from very low to very high (–0.519 to 11.061), where item 9 (“Do you prefer to stay at home, rather than go out and do new things?”) was identified

Table 2. Item parameters and reliability analysis of Geriatric Depression Scale in the present study and comparison of the suggested 4-item short version (GDS-4 GERDA) with previous versions. The table is sorted according to item difficulty from lowest to highest. Cronbach's alpha of the whole scale was 0.736.

Corrected item to total correlation	Cronbach's alpha if item removed	Difficulty	Discrimination	Item	Items used in other versions suggested in previous research					
					GDS-4 GERDA-study items	GDS-4 D'Ath	GDS-4 Van Marwijk	GDS-5 Hoyl	GDS-6 Wong-pakaran	
All items										
0.108	0.748	-0.519	0.328	9. Do you prefer to stay at home, rather than go out and do new things?			●	●		
0.317	0.727	2.822	0.586	2. Have you dropped many of your activities and interests?			●			
0.276	0.733	3.595	0.456	13. Do you feel full of energy?	●					
									
0.490	0.704	5.288	0.731	12. Do you feel worthless the way you are now?	●			●		●
0.459	0.708	5.548	0.655	3. Do you feel that your life is empty?	●					
0.464	0.708	6.080	0.646	8. Do you often feel helpless?	●					
0.498	0.709	8.076	0.743	14. Do you feel that your situation is hopeless?						
0.479	0.712	8.614	0.729	7. Do you feel happy most of the time?			●			
0.326	0.724	8.625	0.436	4. Do you often get bored?						
0.212	0.736	8.896	0.314	6. Are you afraid that something bad is going to happen to you?						
0.417	0.718	9.920	0.606	11. Do you think it is wonderful to be alive?			●			
0.415	0.718	9.349	0.609	5. Are you in good spirits most of the time?						
0.394	0.721	9.954	0.606	1. Are you basically satisfied with your life?						
0.270	0.729	10.174	0.386	15. Do you think that most people are better off than you are?			●			●
0.156	0.739	11.061	0.266	10. Do you feel you have more problems with memory than most?					●	●

GDS = Geriatric Depression Scale.

The dotted line (".....") between 3.595 and 5.288 in the Difficulty column represents the difficult level of the cut-off of the original GDS-15, i.e. 5 points.

Difficulty = represents the level of depressive symptoms at which half of the population scores points on a particular item. The higher the value, the more severe symptoms.

Discrimination = describes the strength of the relationship between the item and the depressive symptoms. A higher value discriminate better and is therefore preferred.

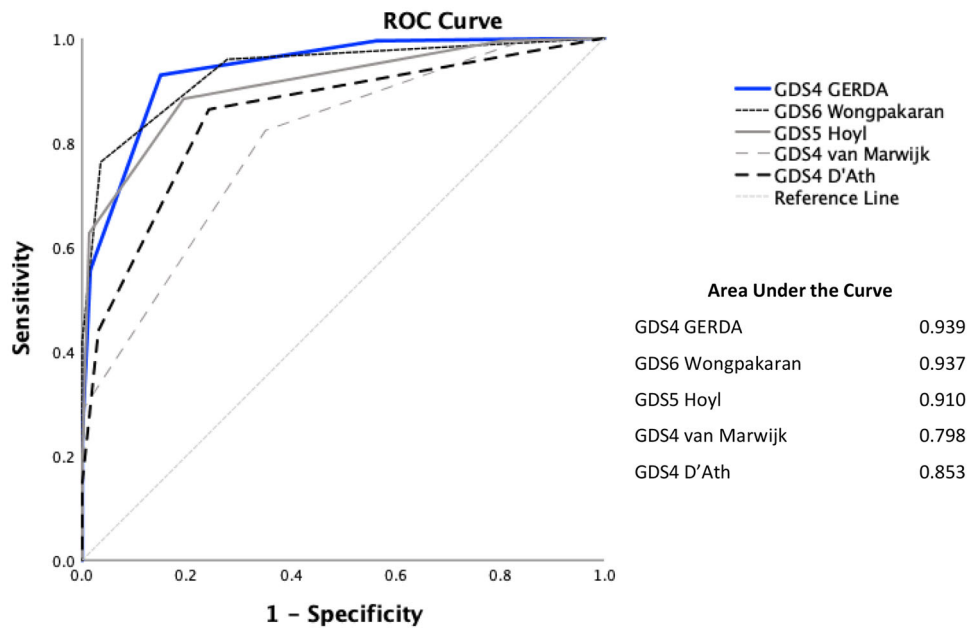


Figure 2. Receiver operating characteristic (ROC) curves for the suggested short version GDS-4 GERDA in comparison to other versions suggested in previous research: GDS-6 Wongpakaran, GDS-4 D'Ath, GDS-4 van Marwijk, and GDS-5 Hoyl. The ROC curves were drawn using the 15-item Geriatric Depression Scale with cut-off of ≥ 5 as indicating depression as gold standard.

Table 3. Sensitivity, specificity, positive predictive value and negative predictive value for the suggested 4-item short version, GDS-4 GERDA, for indicating depression at possible cut-off values.

Cut-off	Participants (n)	Sensitivity	Specificity	Positive predictive value	Negative predictive value
≥ 1	452	0.995	0.437	0.436	0.995
≥ 2	252	0.929	0.850	0.730	0.965
≥ 3	117	0.556	0.985	0.940	0.835
4	46	0.232	1	1	0.749

All values were calculated using the 15-item Geriatric Depression Scale with cut-off of ≥ 5 as indicating depression as golden standard. Participants (n) means the number of participants of the total sample ($n = 651$) that would have scored at or above the suggested cut off value, including true positives and false positives.

with the lowest difficulty and item 10 ("Do you feel you have more problems with memory than most?") identified with the highest difficulty. These two items also had low discrimination values. Items 3 ("Do you feel that your life is empty?"), 8 ("Do you often feel helpless?"), 12 ("Do you feel worthless the way you are now?"), and 13 ("Do you feel full of energy?") corresponded most closely to the GDS-15 cut-off of ≥ 5 points (Figure 1b), with difficulty values ranging from 3.595 to 6.080. These four items also had satisfactory discrimination values (0.456–0.731).

Item difficulty and discrimination values varied only marginally according to age group, sex, and cognitive function (data not shown), except for items 6 ("Are you afraid that something bad is going to happen to you?") and 10 ("Do you feel you have more problems with memory than most?"), which presented large sex-related differences in difficulty (item 6: 15.70 for men, 7.61 for women; item 10: 7.63 for men, 14.24 for women). No differences regarding sex or age group was found in which four items most closely corresponded to the GDS-15 cut-off. For participants with severe cognitive impairment (MMSE score < 18), items 2 ("Have you dropped many of your activities and interests?"), 3 ("Do you feel that your life is empty?"), 8 ("Do you often feel helpless?"), and 12 ("Do you feel worthless the way you are now?") most closely corresponded to the GDS-15 cut-off value; the item 2 ("Have you dropped many of your activities and interests?") difficulty value of 3.74 was marginally closer to the cut-off than the item 13 ("Do you feel full of energy?") value of 3.69.

Reliability

In this sample, the internal consistency of the full GDS-15 scale was acceptable (Cronbach's $\alpha = 0.736$). Corrected item-to-total correlations ranged from 0.108 to 0.498 (Table 2), with the lowest values obtained for items 9 and 10 (the items with lowest and highest difficulty; both < 0.2). Cronbach's α value for the total scale was 0.736 and removing item 9 increased α to 0.748 and item 10 to 0.739 (Table 2).

ROC analysis results

ROC curves were drawn for the proposed four-item version of the GDS (GDS-4 GERDA) and compared with previously proposed short versions (Figure 2). The GDS-4 GERDA had the greatest area under the ROC curve (0.939), indicating better performance than the other GDS short versions.

Sensitivity and specificity of the GDS-4 GERDA

The sensitivity and specificity of the GDS-4 GERDA for the identification of individuals with depression (total GDS-15 score ≥ 5) are shown in Table 3. With a cut-off value of ≥ 1 , the GDS-4 GERDA had 99.5% sensitivity and 43.7% specificity; with a cut-off value of ≥ 2 , it had 92.9% sensitivity and 85.0% specificity.

Discussion

Based on IRT analysis of the GDS-15 items, we found that items 3, 8, 12, and 13 were most suitable for the GDS-4 GERDA in our sample. The inclusion of item 2 could also be considered for application to populations with severe cognitive impairment.

For easy and rapid screening, an instrument should preferentially have a high negative predictive value and high sensitivity; a somewhat larger number of false positive findings is acceptable, given the ability to rule out such cases later in the diagnostic process using more specific methods (Streiner & Norman, 2008). Following this guideline, the GDS-4 GERDA could be used for screening in clinical practice with a cut-off score of ≥ 1 to indicate depression; this approach, however, resulted in the finding of depression in approximately 70% of our sample. GDS-4 GERDA use with a cut-off score of ≥ 2 yielded fewer false-positive results, with the detection of depression in approximately 40% of our sample compared with approximately 30% detected with the GDS-15. Clinical practitioners can memorize the four screening items for easier administration and less time consumption, which could lead to more frequent usage among very old adults.

Item 9 (“Do you prefer to stay at home, rather than go out and do new things?”) had the lowest difficulty value in our sample, suggesting that the oldest adults may prefer to not leave their homes to do new things, regardless of depressive status. This result is in accordance with Wongpakaran et al.’s (Wongpakaran et al., 2019) finding that item 9 is likely culturally biased and unable to assess depression. In contrast, this item is included in the GDS-4 van Marwijk (van Marwijk et al., 1995) and GDS-5 Hojl (Hojl et al., 1999). This discrepancy could be explained by differences in methodology or respondent age and/or cultural background. Item 13 (“Do you feel full of energy?”) is included in the GDS-4 GERDA, but not in any previously proposed short version used for comparison in this study. Wongpakaran et al. (Wongpakaran et al., 2019) found item 13 to have differential functioning among men and women. However, this item is similar to and could be taken to represent the major depression criterion “fatigue or loss of energy nearly every day” appearing in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders: DSM-5 (Arlington, 2013). Our result suggests that item 13 is suitable for the detection of depression in the oldest adults, despite the common association of reduced energy levels with old age and health issues other than depression (Birrer & Vemuri, 2004). The greater sex-based variance of difficulty values for items 6 (“Are you afraid that something bad is going to happen to you?”) and 10 (“Do you feel you have more problems with memory than most?”) than for other items may imply a sex-related difference in the symptomology of depression in the oldest adults. Item 10 also had the highest difficulty value among items and is not included in any short version used for comparison in this study, whereas item 6 is included in the GDS-4 version proposed by D’Ath (D’Ath et al., 1994). The GDS-4 GERDA had a slightly larger area under the ROC curve than did other proposed short versions of the GDS examined in this study but comparison is difficult due to different age, language and culture etc in the compared samples.

Study strengths and limitations

Few studies have evaluated GDS use with adults aged ≥ 85 years, perhaps because this population is difficult to study due to age-related health issues. Adults aged ≥ 85 years have a high prevalence of depression⁶ but also of loneliness, social isolation, negative life events, loss of function, living in institutional care as well as comorbidities (Table 1) which increases the importance of testing psychometric properties of screening tools for depression in this group. Thus, a strength of this study is that the participants were older than those included in previous studies (D’Ath et al., 1994; Hojl et al., 1999; van Marwijk et al., 1995; Wongpakaran et al., 2019), which along with the large sample size supports the feasibility of the GDS-4 GERDA for these oldest adults. Because the GERDA study did not involve exclusion according to health status, cognitive functioning, or living conditions, our sample likely reflects the heterogeneity of the oldest adults in northern Sweden and western Finland during the study period. Thus the sample could be taken to be representative, a strength for the development of an easy-access screening method. The sample also included participants with a wide range of total GDS-15 scores. Further, we also used a novel and systematic approach using IRT analysis to find items in GDS-15 with difficulty close to the scales cut-off, which seems like a more logical method than previous attempts.

A limitation of the study is that the evaluation of the GDS-4 GERDA was performed with the GDS-15 serving as the gold standard; the validity of the scale needs further evaluation against clinical depression diagnoses and other scales assessing depressive symptoms. Such analyses were not possible in this study because GERDA participants’ clinical depression diagnoses were based partially on GDS-15 scores. The use of the Swedish GDS-15 scale may limit the applicability of the results to the scale in other languages, as wording may differ.

Conclusions and implications

We propose the use of the GDS-4 GERDA, an optimized short version of the GDS-15, for easy and rapid screening for depressive symptoms among adults aged ≥ 85 years. Routine use of this tool could reduce the underdiagnosis of depression in this population.

Disclosure of interest

The authors reports no conflict of interest

Author contribution

All authors made substantial contributions to conception and design, and/or acquisition of data, and/or analysis and interpretation of data; participated in drafting the article or revising it critically for important intellectual content; and gave final approval of the version to be submitted.

Sponsor’s role

Funders had no role in trial design, data collection, analysis, or preparation of the manuscript.

Funding

This work was supported by funds from the Interreg IIIA Kvarken-MittSkandia Program (2005–2007) and the Bothnia-Atlantica Program, both funded by the European Union and the European Regional Development Fund. This work was also supported by the Umeå University Foundation for Medical Research, King Gustav V's and Queen Viktoria's Foundation of Freemasons, Västerbotten County Council, the Strategic Research Program in Care Sciences, Norrbotten County Council, Sweden, Joint Committee of County Councils in Northern Sweden "Visare Norr" and the Swedish Research Council 2014 (Grant K2014-99X-22610-01-6).

References

- Arlington, V. (ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Arlington, VA: American Psychiatric Association.
- Beekman, A. T. F., Copeland, J. R. M., & Prince, M. J. (1999). Review of community prevalence of depression in later life. *British Journal of Psychiatry*, *174*(4), 307–311. doi:10.1192/bjp.174.4.307
- Bergdahl, E., Gustavsson, J. M., Kallin, K., von Heideken Wagert, P., Lundman, B., Bucht, G., & Gustafson, Y. (2005). Depression among the oldest old: The Umeå 85+ study. *International Psychogeriatrics*, *17*(4), 557–575. doi:10.1017/S1041610205002267
- Birrer, R. B., & Vemuri, S. P. (2004). Depression in later life: A diagnostic and therapeutic challenge. *American Family Physician*, *69*(10), 2375–2382.
- Chiesi, F., Primi, C., Pigliautile, M., Ercolani, S., Della Staffa, M. C., Longo, A., ... Mecocci, P. (2017). The local reliability of the 15-item version of the Geriatric Depression Scale: An item response theory (IRT) study. *Journal of Psychosomatic Research*, *96*, 84–88. doi:10.1016/j.jpsychores.2017.03.013
- Conradsson, M., Rosendahl, E., Littbrand, H., Gustafson, Y., Olofsson, B., & Lövheim, H. (2013). Usefulness of the Geriatric Depression Scale 15-item version among very old people with and without cognitive impairment. *Aging & Mental Health*, *17*(5), 638–645. doi:10.1080/13607863.2012.758231
- D'Ath, P., Katona, P., Mullan, E., Evans, S., & Katona, C. (1994). Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Family Practice*, *11*(3), 260–266. doi:10.1093/fampra/11.3.260
- Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in older adults. *Annual Review of Clinical Psychology*, *5*(1), 363–389. doi:10.1146/annurev.clinpsy.032408.153621
- Hoyl, M. T., Alessi, C. A., Harker, J. O., Josephson, K. R., Pietruszka, F. M., Koelfgen, M., ... Rubenstein, L. Z. (1999). Development and testing of a five-item version of the Geriatric Depression Scale. *Journal of the American Geriatrics Society*, *47*(7), 873–878. doi:10.1111/j.1532-5415.1999.tb03848.x
- Lövheim, H., Gustafsson, M., Isaksson, U., Karlsson, S., & Sandman, P. O. (2019). Gottfries' cognitive scale for staff proxy rating of cognitive function among nursing home residents. *Journal of Alzheimer's Disease: JAD*, *72*(4), 1251–1260. doi:10.3233/JAD-190599
- Luppa, M., Sikorski, C., Luck, T., Ehreke, L., Konnopka, A., Wiese, B., ... Riedel-Heller, S. G. (2012). Age- and gender-specific prevalence of depression in latest-life-systematic review and meta-analysis. *Journal of Affective Disorders*, *136*(3), 212–221. doi:10.1016/j.jad.2010.11.033
- Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal*, *14*, 61–65.
- Peacock, J., & Peacock, P. (2011). *Oxford handbook of medical statistics*. Oxford, UK: Oxford University Press. doi:10.1093/med/9780199551286.001.0001
- Pocklington, C., Gilbody, S., Manea, L., & McMillan, D. (2016). The diagnostic accuracy of brief versions of the Geriatric Depression Scale: A systematic review and meta-analysis. *International Journal of Geriatric Psychiatry*, *31*(8), 837–857. doi:10.1002/gps.4407
- Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*, *5*(1-2), 165–173.
- Stek, M., Gussekloo, J., Beekman, A., Van Tilburg, W., & Westendorp, R. (2004). Prevalence, correlates and recognition of depression in the oldest old: The Leiden 85-plus study. *Journal of Affective Disorders*, *78*(3), 193–200. doi:10.1016/S0165-0327(02)00310-5
- Streiner, D. L., & Norman, G. R. (2008). *Helath measurement scales: A practical guide to their development and use*. New York, NY: Oxford University Press.
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, *18*(3), 291–307. doi:10.1177/1073191110374797
- Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society*, *40*(9), 922–935. doi:10.1111/j.1532-5415.1992.tb01992.x
- van Marwijk, H. W., Wallace, P., de Bock, G. H., Hermans, J., Kaptein, A. A., & Mulder, J. D. (1995). Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *Br J Gen Pract*, *45*(393), 195–199.
- Weeks, S. K., McGann, P. E., Michaels, T. K., & Penninx, B. W. (2003). Comparing various short-form Geriatric Depression Scales leads to the GDS-5/15. *Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing*, *35*(2), 133–137. doi:10.1111/j.1547-5069.2003.00133.x
- Wongpakaran, N., Wongpakaran, T., & Kuntawong, P. (2019). Evaluating hierarchical items of the geriatric depression scale through factor analysis and item response theory. *Heliyon*, *5*(8), e02300 doi:10.1016/j.heliyon.2019.e02300
- World Health Organization (2008). *The global burden of disease: 2004 update*. Geneva, Switzerland: WHO Press.
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, *17*(1), 37–49. doi:10.1016/0022-3956(82)90033-4