

Accuracy of 12 short versions of the Geriatric Depression Scale to detect depression in a prospective study of a high-risk population with different levels of cognition

Simona Sacuiu,^{1,2}  Nazib M. Seidu,¹ Robert Sigström,^{1,2} Therese Rydberg Sterner,¹ Lena Johansson,¹ Stefan Wiktorsson,¹ and Margda Waern^{1,3}

¹Neuropsychiatric Epidemiology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden

²Department of Psychiatry Cognition and Old Age Psychiatry, Sahlgrenska University Hospital, Region Västra Götaland, Mölndal, Sweden

³Psychosis Department, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden

ABSTRACT

Objectives: To determine the accuracy of 12 previously validated short versions of the Geriatric Depression Scale (GDS) to detect major depressive disorder (MDD) in a high-risk population with and without global cognitive impairment.

Design: Cross-sectional study.

Setting: Five hospitals, Western Sweden.

Participants: Older adults (age ≥ 70 years, $n = 60$) assessed at a home visit 1 year after hospital care in connection with suicide attempt.

Measurements: Depression symptoms were rated using the established 15-item GDS. Eleven short GDS versions identified by a recent systematic review were derived from this administered version. Receiver operating characteristic curves and area under the curve (AUC) for the identification of MDD diagnosed according to Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, were obtained for each version. The Youden Index optimal criterion was used to determine the appropriate cutoffs. Analyses were repeated after stratification by cognitive status (Mini Mental State Examination score ≤ 24 and >24) for the best performing GDS short versions and the established 15-item GDS.

Results: The 7-item GDS according to Broekman *et al.* (2011), with a cutoff 3, was the most accurate among the 12 short versions (AUC 0.90, 95% confidence interval 0.80–1.00), identifying MDD with sensitivity 88% and specificity 81%. The cutoff score remained consistent in the presence of global cognitive impairment, which was not the case for the standardized 15-item GDS.

Conclusion: The Broekman 7-item GDS had high accuracy to detect MDD in this prospective clinical cohort at high risk for MDD. Further testing of GDS short versions in diverse settings is required.

Key Words: depression, scales, prospective clinical study, cognitive impairment

Introduction

The prevalence of major depression in late life ranges between 4.6% and 9.3% (Luppa *et al.*, 2012). Major depression is a contributor to decreased functional level and quality of life (Daly *et al.*, 2010). Screening instruments that are easy to administer and have high

sensitivity and specificity for the detection of major depression can play an important role in secondary preventive interventions.

The Geriatric Depression Scale (GDS) was originally developed as a 30-item screening tool for depression in older adults (Yesavage *et al.*, 1982). Efforts have been undertaken to cement the validity of GDS, to better its accuracy in diverse populations and settings, and to improve its efficiency by decreasing the number of redundant items. The GDS 15-item version (Sheikh and Yesavage, 1986) is currently one of the most widely used instruments for detection of depression in older adults. A recent

Correspondence should be addressed to: Simona Sacuiu, Neuropsychiatric Epidemiology, Institute of Neuroscience and Physiology, University of Gothenburg; and Department of Psychiatry Cognition and Old Age Psychiatry, Sahlgrenska University Hospital, SE-431 41 Mölndal, Sweden. Phone + 46 31 342 7002. Email: simona.sacuiu@neuro.gu.se. Received 13 Jun 2019; revised version received 28 Sep 2019; accepted 06 Oct 2019. First published online 21 November 2019.

systematic review of brief GDS versions further identified 1-, 4-, 5-, 7-, 8-, and 10-item versions (Pocklington *et al.*, 2016). In that review, a meta-analysis was applied to the standardized version of 15-item GDS and demonstrated pooled sensitivity 89% and specificity 77% to detect depression when using the established 15-item GDS cutoff score 5. The values of these pooled accuracy parameters are not very high, most likely reflecting differences in clinical and community samples included in the meta-analysis.

A 12-item version was developed for administration in institutionalized patients and showed only slightly lower sensitivity, but higher specificity to detect depression compared to the 15-item GDS using the same cutoff (score 5) (Sutcliffe *et al.*, 2000). The accuracy of other brief versions was not determined in that study. Moreover, it remains to be elucidated whether a given scale retains accuracy to detect depression in prospectively followed clinical cohorts when affective pathology may be in remission.

Depression often coexists with cognitive impairment in old adults (Van der Mussele *et al.*, 2013), and it has been shown that the accuracy of longer GDS versions is affected by severe cognitive impairment (Sheikh and Yesavage, 1986). Therefore, it remains to be clarified whether abbreviated GDS versions could be sensitive enough to detect major depression, irrespective of the presence of cognitive impairment.

We aimed to assess the existent GDS short versions for their accuracy to detect major depressive disorder (MDD) according to the Diagnostic and Statistical Manual of Mental Disorder, fourth edition (DSM-IV) (APA, 1994). For short scales with highest accuracy, we also aimed to determine whether accuracy to discriminate MDD was affected by global cognitive impairment.

Materials and methods

Participants

Data were obtained in a prospective study on attempted suicide in individuals 70 years and older (range 70–91 years) (Wiktorsson *et al.*, 2011). Briefly, consecutive patients admitted to emergency departments at five hospitals in western Sweden in connection with a suicide attempt were recruited during 2003–2006. The ability to comprehend study aims and interview content and to give informed consent was determined by the attending physician. There was no formal testing of patients' capacity to participate in the research study. Patients were excluded due to terminal illness ($n = 2$), severe dementia ($n = 2$), and insufficient knowledge of the Swedish language

($n = 1$), leaving 140 patients eligible for the study. Of these, 7 patients were discharged without receiving study information, 16 died, and 28 declined participation. Thus, 103 patients were included (73.6% participation rate), but 6 of them did not complete the interview, thus leaving 97 participants assessed with GDS (Wiktorsson *et al.*, 2010). At 1-year follow-up, there were 14 deceased, 1 nontraceable, and 22 refusals. Sixty individuals were alive and accepted a psychiatric assessment at follow-up, including GDS (Wiktorsson *et al.*, 2011).

In accordance with the Declaration of Helsinki, all participants gave their informed and written consent for the study. The study was approved by the Research Ethics Committee at the University of Gothenburg.

Procedures

The interviews were performed by a psychologist (SW) who read aloud all study questions for the participants, including self-report questionnaires. Interviews were carried out in participants' homes ($n = 48$), nursing homes ($n = 9$), psychiatric wards ($n = 2$), and at a psychiatric outpatient department ($n = 1$). The median time from hospitalization to the follow-up interview was 391 days. The following scales and questionnaires were used during interview:

The established 15-item GDS (Sheikh and Yesavage, 1986) was used as a standard self-report screening instrument for clinical depression. The scale has 15 "yes/no" questions, and a score 1 was assigned to all "yes" answers and to "no" answers in items 1, 5, 7, 11, and 13 to indicate depressive symptoms (score range 0–15). A score of 5 is considered the standard cutoff score that indicates depression (Shah *et al.*, 1996). For the purpose of this study, we tested the accuracy to detect MDD of the 15-item GDS along with 11 shorter versions of the GDS all derived from the 15-item version: a 12-item version (Sutcliffe *et al.*, 2000), two 10-item versions (D'Ath *et al.*, 1994; van Marwijk *et al.*, 1995), an 8-item version (Allgaier *et al.*, 2011; Jongenelis *et al.*, 2007), a 7-item version (Broekman *et al.*, 2011), two 5-item versions (Cheng *et al.*, 2010; Hoyle *et al.*, 1999), two 4-item versions (D'Ath *et al.*, 1994; van Marwijk *et al.*, 1995), and two 1-item versions (D'Ath *et al.*, 1994; van Marwijk *et al.*, 1995) (Table 1).

The Comprehensive Psychopathological Rating Scale (CPRS) (Åsberg *et al.*, 1978) was used to assess past month psychopathology. The Montgomery Åsberg Depression Rating Scale (MADRS) (Montgomery and Åsberg, 1979) was derived from the CPRS and was employed at initial assessment and at follow-up to capture change in burden of depressive symptoms over time. MADRS includes 10 items rated 0 (no symptom) to 6 (severe symptom) with a maximum score of 60.

Table 1. Items comprised by GDS versions

NO.	GDS ITEM	GDS VERSIONS BY NUMBER OF ITEMS												ANSWER THAT WAS ATTRIBUTED SCORE 1
		15	12	10 ¹	10 ²	8	7	5 ¹	5 ²	4 ¹	4 ²	1 ¹	1 ²	
1.	Are you basically satisfied with your life?	X	X	X	X	X	X	X		X	X		X	No
2.	Have you dropped many of your activities and interests?	X	X	X	X						X			Yes
3.	Do you feel that your life is empty?	X	X	X		X	X		X	X		X		Yes
4.	Do you often get bored?	X	X		X	X	X	X						Yes
5.	Are you in good spirits most of the time?	X	X		X	X	X							No
6.	Are you afraid that something bad is going to happen to you?	X	X	X						X				Yes
7.	Do you feel happy most of the time?	X	X	X	X	X	X		X	X	X			No
8.	Do you often feel helpless?	X	X	X	X	X	X	X						Yes
9.	Do you prefer to stay at home, rather than going out and doing new things?	X			X			X			X			Yes
10.	Do you feel you have more problems with memory than most?	X		X										Yes
11.	Do you think it is wonderful to be alive now?	X	X			X			X					No
12.	Do you feel pretty worthless the way you are now?	X	X		X			X	X					Yes
13.	Do you feel full of energy?	X	X	X	X									No
14.	Do you feel that your situation is hopeless?	X	X	X		X			X					Yes
15.	Do you think that most people are better off than you are?	X		X	X		X							Yes

GDS, Geriatric Depression Scale.

Relevant items for GDS versions were selected according to the following references: 15 items, Sheikh and Yesavage (1986) (established standardized scale); 12 items, Sutcliffe (2000); 10¹ items, D'Ath *et al.* (1994); 10² items, van Marwijk *et al.* (1995); 8 items, Allgaier *et al.* (2011); 7 items, Broekman *et al.* (2011); 5¹ items, Hoyl (1999); 5² items, Cheng (2010); 4¹ items, D'Ath *et al.* (1994); 4² items, van Marwijk *et al.* (1995); single items 1¹, D'Ath *et al.* (1994) and 1², van Marwijk *et al.* (1995).

The Cumulative Illness Rating Scale for Geriatrics (CIRS-G) (Miller *et al.*, 1992) was used to rate serious physical illness in 13 organ systems using scores 0–4. Participants were considered to have serious physical illness if assigned scores 3 (“severe/constant disability and/or uncontrollable chronic problems”) or 4 (“extremely severe illness and/or functional impairment”) on any of the 13 organ categories. A senior psychiatrist (MW) reviewed all ratings.

The Mini Mental State Examination (MMSE) (Folstein *et al.*, 1975) was used to assess global cognitive function. For the purpose of this study, participants with an MMSE score ≤ 24 were categorized as cognitively impaired (Creavin *et al.*, 2016). No imputation was used for missing points due to physical or sensory handicap, e.g. visual impairment, and MMSE scores ranged from 14 to 30.

Psychiatric diagnoses

The research diagnosis of major depression according to the DSM-IV (APA, 1994) was established using an algorithm based on symptoms according to the CPRS (Sjoberg *et al.*, 2013). History of alcohol use disorder (past and current) was registered at

baseline if either alcohol misuse or dependence was acknowledged by any of three sources: interview with the patient, case records, or the regional hospital discharge register (Morin *et al.*, 2013).

Statistical methods

Participants and nonparticipants (deceased and refusals) at follow-up were compared on demographic and clinical variables registered during hospitalization. As Shapiro–Wilk normality testing showed non-normal distributions for age, MMSE, MADRS, and GDS scores, Mann–Whitney test was employed to compare participants and nonparticipants on these continuous numeric variables. Fishers's exact test was applied to test for differences in proportions regarding sex, marital status, education level, antidepressant prescription, MDD, alcohol use disorder, and serious physical illness. In ancillary analyses at follow-up, we also compared age and MADRS score across the subgroups with different levels of global cognitive function (MMSE score ≤ 24 vs. MMSE score >24) using Mann–Whitney test. Distribution of sexes, education level, and serious physical illness were compared in the two subgroups using Fishers's exact test. Two-tailed statistical testing was considered

significant at the α level 0.05 in these analyses (significant p -value <0.05).

We used receiver operating characteristic (ROC) with area under the curve (AUC) to estimate the diagnostic accuracy for different GDS versions to detect depression in this prospective sample using “gold standard” MDD diagnosed according to DSM-IV criteria (Murphy *et al.*, 1987). Youden Index was used to identify optimal cutoffs based on the ROC according to a parametric method (Fluss *et al.*, 2005). The Youden Index was computed as $J = \max_{\text{cut-off}} (\text{sensitivity}_{\text{cut-off}} + (\text{specificity}_{\text{cut-off}} - 1))$. Values of the Youden Index range between 0 (poor accuracy) and 1 (best accuracy). Accuracy of each GDS version was further evaluated by computing sensitivity, specificity, and positive and negative likelihood ratios (LR) using the optimal cutoffs identified using the Youden Index. Sensitivity was computed as the probability of a positive test result in individuals with MDD and specificity as the probability of a negative test result in individuals without MDD. Positive LR was computed as the ratio of probabilities of having a positive test result in individuals with MDD vs. probabilities of a positive test in those without MDD. Negative LR was computed as the ratio of probabilities of having a negative test result in individuals with MDD vs. probabilities of a negative test in the nondepressed. Large positive LR values (closer to infinity) and small negative LR values (closer to 0) indicate accurate diagnostic tests. We also computed Cronbach’s α as a measure of internal consistency of the GDS versions. Finally, we applied equivalence testing among the different GDS versions using two one-sided tests (TOST, i.e. null-hypothesis and equivalence testing of differences between mean values) (Lakens *et al.*, 2018).

All statistical analyses were run using the R program (R studio version 3.5.1).

Results

No baseline differences were observed in participants and nonparticipants (deceased and refusals) at 12-month follow-up regarding sociodemographic factors (age, sex, married or cohabiting status, mandatory education) (Table 2). Most clinical characteristics (MMSE score, major depression, antidepressant prescription at discharge, serious physical illness) also did not differ, but participants had a higher frequency of alcohol use disorder and lower MADRS mean score than nonparticipants.

At 1-year follow-up, 28.3% of participants ($n = 17$) fulfilled DSM-IV criteria for MDD. The median MADRS score for participants at follow-up was 9 (interquartile range (IQR) 13.2, 1st–3rd quantile

range 3–16.2) indicating no to moderate depressive symptoms. As expected, many were on antidepressant treatment (86.7%, $n = 52$). Thirty-seven participants (63.4%) had a serious physical illness as evaluated by the CIRS-G.

AUC and Youden Indices were similar for many of the GDS versions tested (Table 3). The AUC for the 7-item GDS according to Broekman (Broekman *et al.*, 2011) was slightly numerically greater in comparison to AUCs for all the other versions, but the 10-item GDS according to Van Marwijk (van Marwijk *et al.*, 1995) had the highest Youden Index (Table 3).

According to the Youden Index, the 4-item GDS according to D’Ath (D’Ath *et al.*, 1994) was as accurate as the Broekman 7-item version in detecting MDD in this sample.

The 15-item GDS did not outperform these three versions in detecting MDD (Figure 1). However, internal consistency according to Cronbach’s α was poor for both 4-item GDS versions, and for the 5-item version according to Hoyl (Hoyl *et al.*, 1999). All other GDS versions showed good internal consistency (0.8–0.9) (Table 3). The value of Cronbach’s α was similar for the Broekman 7-item version (0.84) and the established 15-item GDS (0.86), which may indicate that the Broekman 7-item version retains the unidimensional property of the scale.

The optimal cutoff GDS-15 score was 9 in this sample, as determined by the Youden Index. Overall accuracy parameters for the 15-item GDS did not improve when we applied the standard cutoff score 5 (sensitivity 88%, specificity 65%). Although the best sensitivity at follow-up (94%) was achieved by the 4-item version by Van Marwijk (van Marwijk *et al.*, 1995), the specificity was poor (Table 3). Low AUC and Youden Index estimates were observed for the two 1-item versions.

Although the GDS versions were equivalent (nonsignificant p -values in equivalence tests), GDS scores were statistically different for the majority of the GDS versions (significant p -values in null-hypothesis tests of differences between mean values), and TOST results were inconclusive (see Supplementary Table S1).

The accuracy of selected GDS versions to detect MDD in subgroups with or without cognitive impairment

The most accurate short versions of the GDS at follow-up, the 10 items by Van Marwijk (van Marwijk *et al.*, 1995), the 7 items by Broekman (Broekman *et al.*, 2011), and the 4 items by D’Ath (D’Ath *et al.*, 1994), were tested further for their accuracy to detect MDD in cognitively intact

Table 2. Baseline sociodemographic and clinical characteristics by participation status at 1-year follow-up

BASELINE CHARACTERISTICS	PARTICIPANTS, N = 60		NONPARTICIPANTS ^a , N = 37		P-VALUE
	n (%)		n (%)		
Women	32 (53.3)		21 (56.8)		0.835
Married or cohabiting	19 (31.7)		16 (43.2)		0.281
Mandatory education	33 (55.0)		20 (54.1)		1.000
Antidepressant prescription ^b	55 (91.7)		34 (91.9)		1.000
MDD	40 (66.7)		26 (70.3)		0.824
Alcohol use disorder	20 (33.3)		5 (13.5)		0.034*
Serious physical illness ^c	37 (61.7)		21 (56.8)		0.674
	MEDIAN	IQR (1ST—3RD QU)	MEDIAN	IQR (1ST—3RD QU)	
Age (years)	80.0	9.0 (75.0–84.0)	80.0	9.0 (75.0–83.5)	0.823
MMSE score	26.0	5.0 (24.0–28.7)	26.0	5.0 (23.0–28.0)	0.477
MADRS score	26.5	15.7 (16.2–32.0)	30.5	21.2 (18.7–40.0)	0.040*
GDS version score					
15-item Sheikh	8.5	7.0 (4.0–11.0)	8.0	8.0 (3.5–11.0)	0.822
12-item Sutcliffe	7.0	6.0 (3.0–9.0)	7.0	7.0 (2.5–9.5)	0.974
10-item D’Ath	5.0	5.0 (2.2–7.0)	5.0	5.0 (2.5–7.0)	0.734
10-item Van Marwijk	5.5	5.0 (3.0–8.0)	6.0	5.0 (3.0–7.5)	0.890
8-item Allgaier	4.0	5.0 (1.0–6.0)	4.0	5.0 (1.0–6.0)	0.702
7-item Broekman	3.0	4.0 (1.0–5.0)	3.0	4.0 (1.0–4.0)	0.629
5-item Hoyl	3.0	2.0 (2.0–4.0)	3.0	2.0 (2.0–4.0)	0.820
5-item Cheng	3.0	3.0 (1.0–4.0)	3.0	4.0(1.0–4.5)	0.768
4-item D’Ath	2.0	3.0 (0.2–3.0)	2.0	2.0 (1.0–3.0)	0.997
4-item Van Marwijk	2.0	2.0 (1.0–3.0)	2.0	2.0 (1.0–3.0)	0.766
1-item D’Ath	1.0	1.0 (0.0–1.0)	1.0	1.0 (0.0–1.0)	0.290
1-item Van Marwijk	0.0	1.0 (0.0–1.0)	0.0	1.0 (0.0–1.0)	0.426

MDD, major depressive disorder; IQR, Interquartile Range; 1st–3rd Qu, First to Third Quantile; MMSE, Mini Mental State Examination; MADRS, Montgomery Åsberg Depression Rating Scale; GDS, Geriatric Depression Scale.

Differences between participants and nonparticipants were tested by Fishers’s exact test for distribution of sexes, marital status, education, antidepressant prescription, MDD, alcohol use disorder, and serious physical illness; and Mann–Whitney test for not normally distributed variables age, MMSE, MADRS, and GDS scores. GDS score range was 0 to maximum in all versions, except GDS 15-item participants score range 0–14 and nonparticipants 1–13; 12-item nonparticipants score range 0–11; 10-item D’Ath participants score range 0–9; and 10-item Van Marwijk nonparticipants score range 1–10.

Missing scores among nonparticipants at follow-up: MMSE *n* = 2 and MADRS *n* = 1.

^a Nonparticipants: 14 deceased and 23 refusals (including *n* = 1 nontraceable).

^b At hospital discharge.

^c Defined as score 3 or 4 on any somatic category on the Cumulative Illness Rating Scale for Geriatrics.

* *p*-values ≤ 0.05 are considered statistically significant (two-tailed significance at α level 0.05).

Table 3. Diagnostic accuracy of different GDS versions to detect major depression in a prospective clinical sample

GDS VERSION	MEDIAN	IQR (1ST–3RD QU)	CRONBACH'S α	AUC (95% CI)	YOUDEN INDEX y	CUTOFF	SENS	SPEC	POSITIVE LR (95% CI)	NEGATIVE LR (95% CI)
15-item Sheikh	4.5	7.0 (2.0–8.7)	0.86	0.87 (0.77, 0.98)	0.64	9	71	93	10.1 (3.3, 31.4)	0.3 (0.1, 0.7)
12-item Sutcliffe	3.0	6.0 (1.0–7.0)	0.88	0.87 (0.77, 0.98)	0.66	6	82	84	5.1 (2.5, 10.3)	0.2 (0.1, 0.6)
10-item D'Ath	2.5	4.0 (1.0–5.0)	0.81	0.88 (0.78, 0.98)	0.66	6	71	95	15.2 (3.8, 60.8)	0.3 (0.2, 0.6)
10-item Van Marwijk	3.5	5.0 (2.0–6.7)	0.79	0.89 (0.78, 0.99)	0.72	7	76	95	16.4 (4.1, 65.3)	0.2 (0.1, 0.6)
8-item Allgaier	1.5	5.0 (0.0–4.7)	0.87	0.88 (0.78, 0.99)	0.67	3	88	79	4.2 (2.3, 7.7)	0.1 (0.0, 0.5)
7-item Broekman	1.0	4.0 (0.0–4.0)	0.84	0.90 (0.80, 1.00)	0.70	3	88	81	4.7 (2.5, 9.1)	0.1 (0.0, 0.5)
5-item Hoyl	2.0	2.0 (1.0–3.0)	0.54	0.86 (0.75, 0.96)	0.63	3	76	86	5.5 (2.5, 12.0)	0.3 (0.1, 0.6)
5-item Cheng	1.0	3.0 (0.0–3.0)	0.81	0.87 (0.77, 0.97)	0.60	3	76	84	4.7 (2.3, 9.7)	0.3 (0.1, 0.7)
4-item D'Ath	1.0	2.0 (0.0–2.0)	0.67	0.88 (0.79, 0.98)	0.70	2	88	81	4.7 (2.5, 9.1)	0.1 (0.0, 0.5)
4-item Van Marwijk	2.0	2.0 (1.0–3.0)	0.31	0.86 (0.77, 0.96)	0.55	2	94	60	2.4 (1.6, 3.5)	0.1 (0.0, 0.7)
1-item D'Ath	0.0	1.0 (0.0–1.0)	Not applicable	0.80 (0.70, 0.91)	0.60	1	88	72	3.2 (1.9, 5.3)	0.2 (0.0, 0.6)
1-item Van Marwijk	0.0	0.0 (0.0–0.0)	Not applicable	0.79 (0.67, 0.91)	0.58	1	65	93	9.3 (2.9, 29.2)	0.4 (0.2, 0.7)

GDS, Geriatric Depression Scale; IQR, Interquartile Range; 1st–3rd Qu, First to Third Quartile; AUC, Area Under the Curve; CI, Confidence Interval; Sens, Sensitivity; Spec, Specificity; LR, Likelihood Ratio.

GDS score range was 0 to maximum in all versions.

(MMSE score 25–30) and cognitively impaired individuals (MMSE score 14–24). The established GDS-15 was also tested for comparison.

There were no differences between the two subgroups regarding age (cognitively intact median age 80 years, IQR 8, 1st–3rd quantile range 75.5–83.5; cognitively impaired median age 83 years, IQR 9, 1st–3rd quantile range 77.0–86.0; Mann–Whitney test p -value 0.174); MADRS score (cognitively intact median MADRS score 8, IQR 12.5, 1st–3rd quantile range 2.2–14.7; cognitively impaired median MADRS score 12.5, IQR 15.7, 1st–3rd quantile range 5.5–21.2; Mann–Whitney test p -value 0.184), and sex distribution (53.3% women cognitively intact vs. 53.3% women cognitively impaired); education (48.9% more than mandatory education among those cognitively intact vs. 33.3% in cognitively impaired; Fisher's Exact test p -value 0.375) or severe physical illness (56.8% in cognitively intact vs. 86.7% in cognitively impaired; Fisher's Exact test p -value 0.219).

Table 4 shows results for subgroups with and without cognitive impairment regarding diagnostic accuracy of GDS versions. Only the 7-item GDS according to Broekman and 4-item GDS according to D'Ath retained their cutoff scores in both subsamples, but sensitivity decreased from 90% to 86% among those with cognitive impairment. Among cognitively impaired patients, the established 15-item version showed accuracy similar to that of the Broekman 7-item version, but the cutoff for the 15-item version suggested by the Youden Index (7) was higher than the standard cutoff (5) and different from optimal cutoffs suggested by the Youden Index in the total sample (9). Internal consistency according to Cronbach's α was not affected by the level of global cognitive impairment.

Discussion

We found satisfactory accuracy for the brief 7-item GDS according to Broekman (Broekman *et al.*, 2011) to detect MDD in a diagnostically heterogeneous sample of previously hospitalized suicide attempters at 1-year follow-up when many were in remission. The presence of cognitive impairment seemed not to affect the cutoff score for the Broekman 7-item scale, which was the case for the established 15-item version. However, the small size of the sample makes any definitive conclusion difficult.

In this prospective clinical sample at high risk for depression, the accuracy to detect MDD was higher for three short versions, i.e. the 10 items by Van Marwijk (van Marwijk *et al.*, 1995), the 7 items by Broekman (Broekman *et al.*, 2011), and the 4 items by D'Ath (D'Ath *et al.*, 1994), compared to the

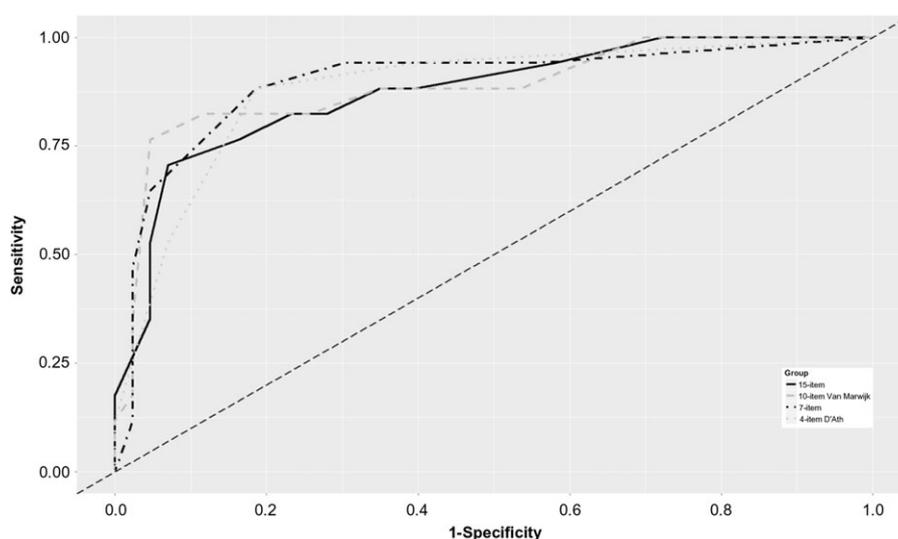


Figure 1. ROC for the identification of major depression with short versions of the GDS at 1-year follow-up ($n = 60$). Most accurate GDS versions 7 items by Broekman (dot-dash dark line), 10 items by Van Marwijk (broken light line), and 4 items by D'Ath (dot light line) as determined by the AUC and the Youden Index at follow-up are depicted. ROC for the established 15-item version by Sheikh (solid dark line) is shown for comparison.

established version of 15-item GDS. The values obtained for the accuracy parameters were moderate across scales, with the Broekman 7-item scale showing highest numerical values in most parameters. The cutoff score (9) for the 15-item GDS identified by the optimization statistic Youden Index in this clinical sample was higher than the standard cutoff (5), and the 15-item GDS did not outperform the Broekman 7-item scale. Our results suggest that the Broekman 7-item GDS with an optimal cutoff score 3 might be applicable in similar clinical settings. This version may be preferred due to its shorter format. Moreover, the internal validity of the Broekman 7-item scale was as good as for the GDS versions with more items.

Studies of brief GDS versions are scarce and differ in methodology, making comparisons difficult. Our finding contrasts with that of a community-based study using the Broekman 7-item GDS scale that identified a cutoff at 1 to detect MDD according to the DSM-IV with high sensitivity (93%) and specificity (91%) (Broekman *et al.*, 2011). While application of the cutoff score 1 in our sample improved sensitivity (94%), specificity at this cutoff was unacceptable (44%). Our results seem to indicate that the short scales have lower accuracy in clinical samples at high risk of affective psychopathology compared to community samples. Furthermore, 5-item scales tested in our sample showed moderate sensitivity (76%) and high specificity (84–86%) at cutoff 3, in contrast with previous clinical studies. A study of the 5-item GDS by Cheng suggested cutoff 2 determined using Youden Index to detect depression in a clinical population

(sensitivities 72–81% and specificities 55–58% for different old age groups) (Cheng *et al.*, 2010). Another study of the accuracy of the 5-item GDS according to Hoyl (Hoyl *et al.*, 1999) tested using an a priori chosen cutoff score 2 in three settings, i.e. hospital, outpatient, and nursing home, demonstrated highest sensitivity 97% among hospitalized patients in the acute geriatric ward (specificity 74%) (Rinaldi *et al.*, 2003). The accuracy of the short GDS scales varies when used in different populations. Taken together, these findings suggest that short GDS scales may be useful in clinical populations but standardized cutoffs may be difficult to establish.

At the cutoff score 3, the Broekman 7-item GDS had slightly better accuracy than the established 15-item GDS in our prospective clinical sample, but only among those who were cognitively intact. The versions tested after stratification by cognitive level had similar accuracies in those with global cognitive impairment, but the optimal cutoff varied in the 15- and 10-item GDS versions for those with and without impairment.

The short GDS scales tested after stratification by cognitive level performed better in those without cognitive impairment than in those impaired, in line with previous studies (Friedman *et al.*, 2005). Others have reported on the accuracy of the 15-item GDS being negatively affected by global cognitive impairment if the standard cutoff was retained (Chiesi *et al.*, 2018; de Craen *et al.*, 2003). However, the small size of the group with global cognitive impairment ($n = 15$) makes clear-cut inferences difficult.

Table 4. Diagnostic accuracy of different GDS versions by global cognitive status in a prospective clinical sample

GDS VERSION	IQR		CRONBACH'S α	AUC (95% CI)	YOUDEN INDEX γ	CUTOFF	SENS	SPEC	POSITIVE LR (95% CI)	NEGATIVE LR (95% CI)
	MEDIAN	(1ST-3RD QU)								
<i>MMSE > 24 n = 45</i>										
15-item Sheikh	4.0	6.0 (2.0-8.0)	0.85	0.88 (0.77, 1.00)	0.66	8	80	86	5.6 (2.3, 13.3)	0.2 (0.1, 0.8)
10-item Van Marwijk	3.0	4.0 (2.0-6.0)	0.78	0.88 (0.74, 1.02)	0.74	7	80	94	14.0 (3.5, 55.7)	0.2 (0.1, 0.7)
7-item Broekman	1.0	4.0 (0.0-4.0)	0.84	0.93 (0.86, 1.01)	0.73	3	90	82	5.2 (2.5, 11.2)	0.1 (0.0, 0.8)
4-item D'Ath	1.0	2.0 (0.0-2.0)	0.62	0.89 (0.79, 0.98)	0.67	2	90	77	3.9 (2.1, 7.5)	0.1 (0.0, 0.8)
<i>MMSE \leq 24 n = 15</i>										
15-item Sheikh	7.0	8.0 (2.0-10.0)	0.89	0.87 (0.67, 1.06)	0.61	7	86	75	3.4 (1.0, 11.8)	0.2 (0.0, 1.2)
10-item Van Marwijk	5.0	5.0 (2.0-7.0)	0.83	0.92 (0.77, 1.07)	0.73	6	86	87	6.8 (1.1, 44.0)	0.2 (0.0, 1.0)
7-item Broekman	3.0	5.0 (0.0-5.0)	0.83	0.85 (0.62, 1.08)	0.61	3	86	75	3.4 (1.0, 11.8)	0.2 (0.0, 1.2)
4-item D'Ath	1.0	3.0 (0.0-3.0)	0.79	0.91 (0.73, 1.00)	0.86	2	86	100	Infinite	0.1 (0.0, 0.9)

MMSE, Mini Mental State Examination; GDS, Geriatric Depression Scale; IQR, Interquartile Range; 1st-3rd, Qu First to Third Quantile; AUC, Area Under the Curve; CI, Confidence Interval; Sens, Sensitivity; Spec, Specificity; LR, Likelihood Ratio. GDS score range was 0 to maximum in all versions, except for the 15-item GDS in those with MMSE > 24 (range 0-14).

To summarize, our findings shed further light on the diagnostic accuracy of the brief versions of the GDS, as called for by the authors of a recent review of the evidence (Pocklington et al., 2016). Furthermore, although we can draw no firm conclusions regarding the accuracy of the brief scales in persons with cognitive dysfunction, our results provide leads for much-needed future research that takes cognitive impairment into consideration (Chiesi et al., 2018).

Methodological considerations

The strengths of the study are the prospective clinical sample and the fact that the same licensed psychologist made assessments during hospitalization and at follow-up. Study limitations include the special nature of this clinical sample (high risk for MDD) and the small sample size. GDS scores in our sample were not normally distributed, which makes it difficult to interpret the results of the available TOST that rely on mean value distributions. Although we acknowledge a bias in obtaining the items of the shorter GDS versions from administering the 15-item GDS, this method allows a comparison of different versions using the same sample. We thus eliminate confounding factors that are difficult to control when comparing different samples. A further consideration is the fact that the cohort was assessed over a decade ago. Nevertheless, the results may be considered representative for individuals older than 70 years currently diagnosed with MDD using DSM-IV criteria, since no radical changes with impact on mental healthcare have occurred in the catchment area during the last decade. There may be a selection bias and a healthy survivor effect as those who took part in the follow-up interviews had lower MADRS scores at baseline. However, a history of alcohol use disorder was more prevalent among those who took part in the follow-up examination.

Another aspect that has to be discussed is that despite almost all participants receiving antidepressants at the 1-year follow-up, one in four individuals exhibited depressive symptomatology fulfilling criteria for major depression. Treatment-refractory depression is common in old age (Knochel et al., 2015). Although the short GDS scales detected MDD with low to moderate accuracy, short scales with high sensitivity (88% for the Broekman 7-item version in our sample) could be useful in the follow-up care of high-risk populations. The scale had good internal consistency.

In conclusion, the brief 7-item version of GDS according to Broekman (Broekman et al., 2011) could be useful in the follow-up of MDD in high-risk populations due to consistency in cutoff score in relation to global cognitive function. Further studies are warranted to compare the accuracy of these

instruments and to standardize the cutoffs in older adult populations with diverse affective pathology.

Conflicts of interest

The authors have no conflicts of interest to declare.

Source of funding

The study was financed by grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement (ALFGBG-715841 to M. Waern and ALFGBG-637271 to S. Sacuiu), the Swedish Research Council (521-2004-6080, 521-2008-3139, 521-2013-2699, 2016-01590), the Swedish Council for Working Life and Social Research (2003-0153, 2016-07097), AGECAP 2013-2300, Söderström König Foundation, Thuring Foundation, Hjalmar Svensson Research Fund, Organon Foundation, Axel Linder Foundation, Wilhelm and Martina Lundgren Foundation, and Konung Gustav V: och Drottning Victorias Frimurarestiftelse. The funding agencies and sponsors had no role in formulation of research question(s), choice of study design, data collection, data analysis, and decision to publish.

Description of authors' roles

Simona Sacuiu was responsible for designing the study, analyzing the data, and writing the paper.

Nazib M. Seidu was responsible for carrying out the statistical analysis and assisted with writing the paper.

Robert Sigström assisted with interpretation of the results and writing the paper.

Therese Rydberg Sterner assisted with interpretation of the results and writing the paper.

Lena Johansson assisted with interpretation of the results and writing the paper.

Stefan Wiktorsson collected the data and assisted with data analysis and writing the paper.

Margda Waern was responsible for the financing of the project, collection of data, designing the study, and writing the paper.

Acknowledgments

The authors wish to thank all study participants and hospital staff on participating wards. We also want to thank Tom Marlow and Vasileios Nikolopoulos for the initial statistical analysis and literature search.

Supplementary material

To view supplementary material for this paper, please visit <https://doi.org/10.1017/S1041610219001650>

References

- Allgaier, A. K., Kramer, D., Mergl, R., Fejtkova, S. and Hegerl, U. (2011). Validity of the Geriatric Depression Scale in nursing home residents: comparison of GDS-15, GDS-8, and GDS-4. [German] Validität der Geriatrischen Depressionsskala bei Altenheimbewohnern: Vergleich von GDS-15, GDS-8 und GDS-4. *Psychiatrische Praxis*, 38, 280–286.
- APA (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association.
- Åsberg, M., Montgomery, S. A., Perris, C., Schalling, D. and Sedvall, G. (1978). A comprehensive psychopathological rating scale. *Acta Psychiatrica Scandinavica*, 57, 5–27. doi:10.1111/j.1600-0447.1978.tb02357.x
- Broekman, B. F., Niti, M., Nyunt, M. S., Ko, S. M., Kumar, R. and Ng, T. P. (2011). Validation of a brief seven-item response bias-free geriatric depression scale. *The American Journal of Geriatric Psychiatry*, 19, 589–596. doi:10.1097/JGP.0b013e3181f61ec9.
- Cheng, S.-T. et al. (2010). The geriatric depression scale as a screening tool for depression and suicide ideation: a replication and extension. *The American Journal of Geriatric Psychiatry*, 18, 256–265. doi:10.1097/JGP.0b013e3181bf9edd.
- Chiesi, F. et al. (2018). Does the 15-item Geriatric Depression Scale function differently in old people with different levels of cognitive functioning? *Journal of Affective Disorders*, 227, 471–476. doi:10.1016/j.jad.2017.11.045.
- Creavin, S. T. et al. (2016). Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews*. doi:10.1002/14651858.CD011145.pub2.
- Daly, E. J. et al. (2010). Health-related quality of life in depression: a STAR*D report. *Annals of Clinical Psychiatry*, 22, 43–55.
- D'Ath, P., Katona, P., Mullan, E., Evans, S. and 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, 260–266.
- de Craen, A. J. M., Heeren, T. J. and Gussekloo, J. (2003). Accuracy of the 15-item geriatric depression scale (GDS-15) in a community sample of the oldest old. *International Journal of Geriatric Psychiatry*, 18, 63–66. doi:10.1002/gps.773.
- Fluss, R., Faraggi, D. and Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal*, 47, 458–472.
- Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

- Friedman, B., Heisel, M. J. and Delavan, R. L.** (2005). Psychometric properties of the 15-Item Geriatric Depression scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients: psychometric properties of the GDS-15. *Journal of the American Geriatrics Society*, 53, 1570–1576. doi:10.1111/j.1532-5415.2005.53461.x
- Hoyl, M. T. *et al.*** (1999). Development and testing of a five-item version of the Geriatric Depression Scale. *Journal of the American Geriatrics Society*, 47, 873–878.
- Jongenelis, K. *et al.*** (2007). Construction and validation of a patient- and user-friendly nursing home version of the Geriatric Depression Scale. *International Journal of Geriatric Psychiatry*, 22, 837–842. doi:10.1002/gps.1748.
- Knochel, C. *et al.*** (2015). Treatment-resistant late-life depression: challenges and perspectives. *Current Neuropharmacology*, 13, 577–591. doi:10.2174/1570159X1305151013200032.
- Lakens, D., Scheel, A. M. and Isager, P. M.** (2018). Equivalence testing for psychological research: a tutorial. *AMPPS* 2018, 259–269. doi:10.1177/2515245918770963.
- Luppa, M. *et al.*** (2012). Age- and gender-specific prevalence of depression in latest-life—systematic review and meta-analysis. *Journal of Affective Disorders*, 136, 212–221. doi:10.1016/j.jad.2010.11.033.
- Miller, M. D. *et al.*** (1992). Rating chronic medical illness burden in geropsychiatric practice and research: application of the Cumulative Illness Rating Scale. *Psychiatry Research*, 41, 237–248.
- Montgomery, S. A. and Åsberg, M.** (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, 134, 382–389.
- Morin, J., Wiktorsson, S., Marlow, T., Olesen, P. J., Skoog, I. and Waern, M.** (2013). Alcohol use disorder in elderly suicide attempters: a comparison study. *The American Journal of Geriatric Psychiatry*, 21, 196–203. doi:10.1016/j.jagp.2012.10.020.
- Murphy, J. M., Berwick, D. M., Weinstein, M. C., Borus, J. F., Budman, S. H. and Klerman, G. L.** (1987). Performance of screening and diagnostic tests. Application of receiver operating characteristic analysis. *Archives of General Psychiatry*, 44, 550–555.
- Pocklington, C., Gilbody, S., Manea, L. and 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, 837–857. doi:10.1002/gps.4407.
- Rinaldi, P. *et al.*** (2003). Validation of the five-item geriatric depression scale in elderly subjects in three different settings. *Journal of the American Geriatrics Society*, 51, 694–698.
- Shah, A., Phongsathorn, V., Bielawska, C. and Katona, C.** (1996). Screening for depression among geriatric inpatients with short versions of the geriatric depression scale. *International Journal of Geriatric Psychiatry*, 11, 915–918. doi: 10.1002/(Sici)1099-1166(199610)11:10<915:Aid-Gps411>3.0.Co;2-H.
- Sheikh, J. I. and Yesavage, J.** (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, 5, 165–173. doi:10.1300/J018v05n01_09.
- Sjoberg, L., Ostling, S., Falk, H., Sundh, V., Waern, M. and Skoog, I.** (2013). Secular changes in the relation between social factors and depression: a study of two birth cohorts of Swedish septuagenarians followed for 5 years. *Journal of Affective Disorders*, 150, 245–252. doi:10.1016/j.jad.2013.04.002.
- Sutcliffe, C. *et al.*** (2000). A new version of the geriatric depression scale for nursing and residential home populations: the geriatric depression scale (residential) (GDS-12R). *International Psychogeriatrics*, 12, 173–181.
- Van der Mussele, S. *et al.*** (2013). Prevalence and associated behavioral symptoms of depression in mild cognitive impairment and dementia due to Alzheimer's disease. *International Journal of Geriatric Psychiatry*, 28, 947–958. doi:10.1002/gps.3909.
- van Marwijk, H. W., Wallace, P., de Bock, G. H., Hermans, J., Kaptein, A. A. and Mulder, J. D.** (1995). Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *British Journal of General Practice*, 45, 195–199.
- Wiktorsson, S., Marlow, T., Runeson, B., Skoog, I. and Waern, M.** (2011). Prospective cohort study of suicide attempters aged 70 and above: one-year outcomes. *Journal of Affective Disorders*, 134, 333–340. doi:10.1016/j.jad.2011.06.010
- Wiktorsson, S., Runeson, B., Skoog, I., Ostling, S. and Waern, M.** (2010). Attempted suicide in the elderly: characteristics of suicide attempters 70 years and older and a general population comparison group. *The American Journal of Geriatric Psychiatry*, 18, 57–67. doi:10.1097/JGP.0b013e3181bd1c13.
- Yesavage, J. A. *et al.*** (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, 17, 37–49.