GeroPsych, 27 (1), 2014, 41–49

Full-Length Research Report

Screening for Depression with the Depression in Old Age Scale (DIA-S) and the Geriatric Depression Scale (GDS15)

Diagnostic Accuracy in a Geriatric Inpatient Setting

Sonja Heidenblut and Susanne Zank

Department of Rehabilitative Gerontology, University of Cologne, Germany

Abstract. *Purpose of the study.* The Depression in Old Age Scale (DIA-S), a new screening tool for geriatric depression, was designed to be both practical and appropriate for use with medically ill geriatric patients. The diagnostic accuracy of the DIA-S and the short form of the Geriatric Depression Scale (GDS15) were tested and compared. *Methods.* Using the Montgomery and Asberg Depression Rating Scale (MADRS) as gold standard, the scales were validated with a sample of $N = 331$ geriatric inpatients. *Results.* ROC curves, AUC outcomes, sensitivity and specificity, and logistic regression models for impact factors on misclassification rates indicate good psychometrical qualities of the DIA-S, whereas the validity of the GDS15 was lower.

Keywords: depression, screening, subgroups, DIA-S, GDS15

Introduction

Depression is one of the most common psychiatric disorders. In elderly populations, it is strongly associated with physical disability and poor general health status. Thus, geriatric inpatients are especially at risk for depression, with prevalence rates ranging up to more than 40% (McCusker et al., 2005). Depression is not only common among this population, it also has a negative influence on patients' functional status, leading to poorer rehabilitation outcomes (McCusker, Cole, Ciampi, Windholz, & Belzile, 2007), more cognitive impairment, earlier institutionalization, and even higher mortality rates (Lasser, Siegel, Dukoff, & Sunderland, 1998). Still, detection rates for depressive disorders in general healthcare facilities are low, ranging from 28% for minor depression to 56% for major depression (McCusker et al., 2008).

The Geriatric Depression Scale in Inpatient Settings

Screening scales for older medical inpatients have certain special requirements. The scales are intended to provide objective and reliable information for medical staff who have no special background knowledge in the field of psychiatric diagnostics. These scales must therefore be easy to administer and interpret. Self-rating scales with fixed response sets are preferable. In order to serve a broad range of clients who often suffer from cognitive or functional impairment, a scale must also be suitable for verbal administration and be easy to understand and respond to. This means that items must be short and simple, and the response set should not be a rating scale but a simple "yes"/"no" alternative. Furthermore, somatic symptoms of depression like sleep disturbances or loss of appetite should not be addressed in screening instruments for geriatric depression because these symptoms are very common among the elderly and not specifically related to depression. For this reason most diagnostic tools that have been proven to be appropriate in other settings cannot be used with geriatric inpatients.

The Geriatric Depression Scale (GDS; Yesavage et al., 1983), the first depression scale developed especially for the elderly, took these points into consideration. Hence the 15-item short-form of the scale (GDS-15; Burke, Roccaforte, & Wengel, 1991) is the most frequently used screening instrument for geriatric depression in various clinical

DOI 10.1024/1662-9647/a000101

Table 1. The English version of the Depression in Old Age Scale (DIA-S)

Item	Original German phrasing	Answer
I am feeling down.	Ich fühle mich bedrückt.	**Yes** No
I worry that I might say or do the wrong thing.	Ich habe Angst, dass ich etwas falsches sagen oder tun könnte.	**Yes** No
I can relax easily.	Ich kann mich gut entspannen.	Yes **No**
My life seems to make little sense.	Mein Leben kommt mir sinnlos vor.	**Yes** No
It's hard to motivate myself.	Es fällt mir schwer, mich aufzuraffen.	**Yes** No
I'm worried about the future.	Ich habe Angst vor der Zukunft.	**Yes** No
I can enjoy my life, even when things are sometimes more difficult.	Ich kann mein Leben genießen, auch wenn mir manches schwerer fällt.	Yes **No**
Difficulties tend to overwhelm me.	Ich fühle mich durch Schwierigkeiten leicht überfordert.	**Yes** No
I tend to brood a lot.	Ich muss viel grübeln.	**Yes** No
Basically I am content with my life.	Ich bin grundsätzlich mit meinem Leben zufrieden.	**Yes** No

Note. The answer in **bold** scores with one point, respectively.

settings. However, as the scale was actually developed for community-dwelling elderly, there are some difficulties in using the GDS15 with inpatient populations. Questions like "Do you prefer staying at home rather than going out and doing new things?", "Do you feel full of energy?", or "Do you often get bored?" are not appropriate for geriatric patients suffering from acute physical impairment, who have sometimes been in the hospital for weeks. These items decrease patients' willingness to respond to the questions asked. They can also lead to false-positive test results because they confuse depressive symptoms with aspects of multimorbidity. Empirical data on the usefulness of the GDS15 with geriatric inpatients is contradictory. In a large study with 2,032 medical inpatients (Incalzi, Cesari, Pedone, & Carbonin, 2003), the GDS15 did not qualify as unidimensional due to poor internal consistency (Cronbach's $\alpha = .46$). A systematic review on the validity of the scale (Wancata, Alexandrowicz, Marquart, Weiss, & Friedrich, 2006) reported satisfactory outcomes in geriatric inpatients (sensitivity: 88%, specificity 79%). However, specific information about sampling methods or whether the interviewers were in fact blind to the expected results of the screening was often not provided for the studies included in the review. The average results for diagnostic accuracy were therefore likely to overestimate the usefulness of the scale.

There are various very short versions of the GDS consisting of 10 (Shah, Phongasthorn, Bielawska, & Katona, 1996), 5 (Kenny Weeks, McGann, King Michaels, & Penninx, 2003), or even 4 items (Goring, Baldwin, Marriott, Pratt, & Roberts, 2004). Validation studies of these scales are problematic in that they either are based on small and specific samples (Rinaldi et al., 2003), compare the scales to longer versions of the GDS (Goring et al., 2004; Kenny Weeks et al., 2003), or show low rates of specificity for the scales (Pomeroy, Clark, & Philp, 2001; Shah et al., 1996). Thus, based on the available data, these scales do not present an alternative for clinical practice. Two further versions of the GDS, the GDS8 (Jongenelis et al., 2007) and the GDS12-R (Sutcliffe et al., 2000), were constructed for nursing-home populations and excluded some of the context-dependent items. However, because they were designed for a different setting than the one in this study, some of their items are still misleading for people with acute medical problems. For this reason we constructed an alternative screening scale for depression in the elderly which would be more appropriate for use in medical inpatient settings (Table 1).

The Depression in Old Age Scale (DIA-S)

The Depression in Old Age Scale (DIA-S; Heidenblut & Zank, 2010), designed as a screening tool for use in clinical practice, is based on the diagnostic criteria for depressive disorders described in the International Classification of Diseases (ICD-10, World Health Organization (WHO), 2010). Scale results can also be used to make diagnoses based on the DSM-5 (American Psychiatric Association (APA), 2013) as symptoms of depression are comparable in both manuals.

The DIA-S was constructed to be brief and easy to apply and interpret, with items that were meant to be context free, so the instrument could be used in different healthcare settings as well. The scale consists of ten short statements about depression that are to be evaluated as true or false with a simple yes/no answer format. Attention was paid to ensure that the statements were unambiguous, clearly phrased, and each oriented toward one symptom. For signs of depression that were likely to be confused with other health problems, the content of the respective item was slightly modified. Thus, instead of asking about "sleep disturbances," the item was rephrased to capture whether the patient felt able to relax; a focus was on a lack of motivation, rather than a "lack of energy"; and, as a cognitive symptom of depression, worrying is mentioned instead of "problems with concentration." Items that focused on loss of appetite or suicidal thoughts were not included, due to poor discriminatory power. In a first validation study (Heidenblut & Zank, 2010), the scale showed good results in

the discriminatory power of the items, the internal consistency (Cronbach's α = .84), and the correlation with the diagnostic criterion (r = .73).

The German version of the scale was translated into English by a native-speaking psychologist who has very advanced language skills in German. The translator was informed about the purpose of the scale and the context it was intended for. After the first translation, the meaning of each item was discussed and, where indicated, modified.

Factors Influencing the Diagnostic Accuracy of Self-Rating Scales for Depression

The test accuracy of self-rating scales for depression can be influenced in various ways. When a scale is validated on patients whose disease status is more obvious and easier to detect than in the population it is intended for, the true detection rate is likely to be overestimated. Although these influences, also known as spectrum or extrapolation biases (Pepe, 2004), can be avoided with careful sampling, there can still be a considerable range in test performance in different subpopulations (Henkel et al., 2004). Women are found to be more likely to report certain depressive symptoms than men are (Sigmon et al., 2005), and some studies report gender-based differences in test accuracy (Camozzato, Hidalgo, Souza, & Chaves, 2007; Semmler & Klumb, 2004). In the field of geriatrics, patients' age as well as cognitive and functional status are important sources of variance that become obvious in the broad range of the diagnostic accuracy of the GDS between different validation settings (inpatients, day patients, outpatients, nursing-home residents, patients with or without dementia, community-dwelling elderly; Wancata et al., 2008). In addition to patient characteristics, chracteristics of the testing situation can also have an impact on test performance. For example, among healthy populations seasonal variations like the duration of daylight have been found to influence complaints about depressive symptoms (Schlager, Schwartz, & Bromet, 1993). Although the issue of seasonal mood variation in the elderly is still controversial (De Craen, Gussekloo, van der Mast, le Cessie, Lemkes, & Westendorp, 2005), it is likely to be relevant in certain subgroups of clients and can thus serve as a moderating factor.

Purpose of the Current Study

The current study compares the diagnostic accuracy of the DIA-S and the GDS15 in a sample of geriatric inpatients. The study sample is representative of the target group in the relevant characteristics (cognitive and functional impairment, general health status, severity of the depressive disorder); the data were collected under conditions similar to those in clinical practice. The impact of patient characteristics such as sex, age, cognitive and functional status, and of setting characteristics such as the duration of daylight on the diagnostic accuracy of the scales were examined as well as the general psychometric. As detection rates in scales can differ between men and women, sensitivity, specificity, and appropriate cutoff scores were also estimated for men and women separately.

Design and Methods

Study Design and Data Collection

Data for the validation of the scales was collected between 2007 and 2009 in three inpatient geriatric healthcare units in Germany. Only patients whose cognitive status was sufficient (MMSE-Score = 15) and who did not suffer from aphasia, delirium, or psychotic disorders were allowed to participate in the study. All patients meeting the criteria were invited by staff members of the respective clinics to participate in the study. In the first stage of data collection, all subjects who had agreed to participate were included. Because depressed participants were more difficult to recruit than patients without depression, we carried out a second stage of data collection during which staff members purposely recommended subjects with clinical signs of depression so as to increase the number of possibly ill subjects in the sample. In the end, we had a total of n = 151 depressed subjects and n = 181 nondepressed subjects in our sample.

Study participants were interviewed twice, the interview sequence being varied. In one part of the study, the GDS15 and the DIA-S were administered orally ba a trained interviewer. When replying to the questionnaire items, patients were asked to consider how they felt over the last 2 weeks. In the other part, a semistructured psychiatric interview was conducted which referred to ten depressive symptoms that were later rated on the Montgomery and Asberg Depression Rating Scale (MADRS). Each part of the study was conducted by an independent interviewer blind to the results of the other interview.

Measures

DIA-S

The development and translation of the DIA-S is described above. The scale ranges from 0 (*no depressive symptoms*) to 10 (*maximal amount of depressive symptoms*). The internal consistency of the scale based on the data of this study was α = .84 (Heidenblut & Zank, 2010). Further characteristics of the diagnostic accuracy of the scale are presented in the results.

GDS15

The 15-item short form of the Geriatric Depression Scale is the most commonly used self-rating scale for geriatric depression. The maximum score is 15 points, whereas a score above 5 points serves as the cutoff for clinically relevant depression. The internal consistency of the scale based on the data of this study was $\alpha = .75$ (Heidenblut & Zank, 2010).

MADRS

The Montgomery and Asberg Depression Rating Scale (MADRS; Montgomery & Asberg, 1979; Neumann & Schulte, 1989) is an interview-based depression rating scale that is very commonly used with geriatric patients in both research and practice. The validity and reliability of the MADRS are supported by numerous studies comparing the MADRS with other established criteria for depression, including the HAMD, BDI, DSM-Diagnoses, and AGECAT (Mottram, Wilson, & Copeland, 2000; Müller, Szegedi, Wetzel, & Benkert, 2000; Uher et al., 2008). The scale was used as gold standard for this study, because it has already been successfully validated among older patients (Mottram et al., 2000) and among patients with considerable somatic comorbidity such as Parkinson's disease (Leentjens, Verhey, Lousberg, Spitsbergen, & Wilmink, 2000). The entire score ranges from 0 to 60 points. For the analyses cutoff scores recommended by Neumann and Schulte (1989) were used, with a score above 13 indicating a mild, a score above 21 a moderate, and a score above 28 a severe form of depressive disorder. In the current study, the outcome of the scale serves as the gold-standard criterion for depression, with participants being divided into depressed or nondepressed categories via the first cutoff at 13 points. All MADR interviews for the gold standard were conducted by a trained psychologist with clinical experience in dementia and neuropsychology. The internal consistency of the scale based on the results of this study is $\alpha = .86$.

Cognitive Impairment

The participants' cognitive impairment was measured with the Mini-Mental State Examination (AGAST, 1997; Folstein, Folstein, & McHugh, 1975).

Functional Impairment

The functional impairment of the patients in basic activities of daily living was measured via the Barthel Index (AGAST, 1997; Mahoney & Barthel, 1965).

Duration of Daylight

The duration of daylight was measured by counting the number of days between the date of the psychiatric interview and midsummer.

Difficulty of Classification

The difficulty of classification dependent on the true disease status of a participant was operationalized as the absolute value of distance between the MADR score and the first cutoff of the MADR Scale.

Sample

Sample characteristics are presented in Table 2. The age and sex distribution of the sample is comparable to the overall population of geriatric inpatients in Germany, as presented in epidemiological studies, whereas cognitive and physical functioning are slightly better, due to the criteria for inclusion in the study (see Renteln-Kruse & Ebert, 2003). The spectrum of depressive disorders also shows the expected proportions, with 84% of the affected group showing symptoms of mild depressive disorder and only 16% suffering from moderate or severe depression. Among those persons with no current depressive disorder, 26% of the participants suffered from subclinical symptoms (with an MADR score between 10–12 points). Complete assessment data were available for a subsample of $n = 195$ participants which sufficiently matches the characteristics of the original sample. This subsample was used for logistic regression analyses to examine impact factors on the diagnostic accuracy of the scales.

Table 2. Characteristics of the total sample and a subsample of participants with complete assessment data

Characteristic	Total sample $N = 331$		Subsample $n = 195$	
	N (%)	*M* (*SD*)	*N* (%)	*M* (*SD*)
Sex				
Male	88 (27%)		62 (32%)	
Female	243 (73%)		133 (68%)	
Depression diagnosis				
Yes	151 (46%)		101 (52%)	
No	180 (54%)		94 (48%)	
Severity coding				
No depression	133 (40%)		69 (35%)	
Subclinical	47 (14%)		25 (13%)	
Mild	127 (38%)		85 (44%)	
Moderate	19 (6%)		14 (7%)	
Severe	5 (2%)		2 (1%)	
Age ($n = 321$)		80.7 (7.6)		79.8 (7.6)
Alternative assessment				
MMSE ($n = 311$)		25.0 (3.8)		25.3 (3.6)
Barthel Index ($n = 211$)		51.7 (24.6)		52.2 (24.9)

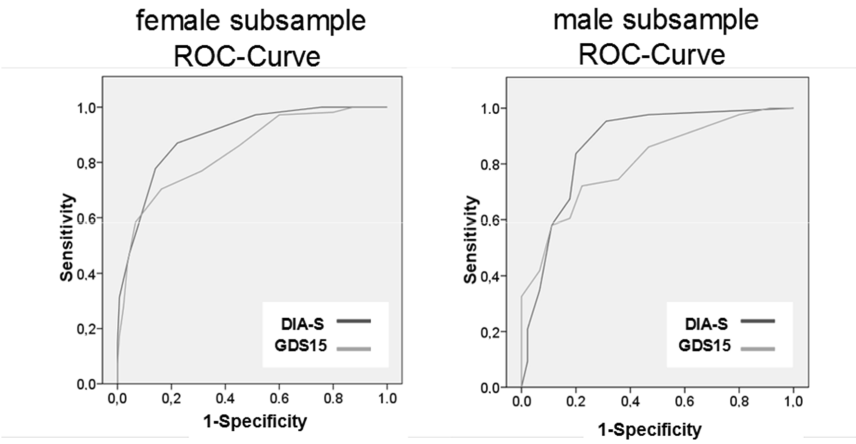

Figure 1. Comparison of receiver operating characteristic (ROC) curves of the tests according to gender.

Notes. Female subsample: n =108 depressed, n = 135 not depressed, male subsample: n = 43 depressed, n = 45 not depressed.

Table 3. Diagnostic accuracy of the DIA-S and the GDS15

		Test accuracy			ROC curve			
								95% CI
Sample	Scale	Cutoff	Sensitivity	Specificity	AUC	Lower	Upper	*d*
All	DIA-S	3.5	.82	.79	.882	.914	.841	
(*N* = 331)	GDS15	5.5	.76	.71	.828	.867	.783	.053*
Women	DIA-S	3.5	.87	.78	.895	.931	.849	
(*n* = 243)	GDS15	5.5	.77	.69	.841	.885	.789	.054*
Men	DIA-S	2.5	.84	.80	.870	.933	.780	
(*n* = 98)	GDS15	3.5	.72	.78	.808	.885	.709	.062

Note. *$p < .05$. ROC curve: receiver operating characteristic curve. AUC: area under the curve. CI: confidence interval.

Results

Diagnostic Accuracy of the Tests

To compare the screening qualities of the two tests, sensitivity, specificity, ROC curves, and AUC outcomes were considered. As screening tools can have gender-related detection rates, ROC curves, and AUC outcomes were also estimated for men and women separately, providing the best DIA-S cutoff scores for each gender (Figure 1).

Test Accuracy

The overall diagnostic accuracy of the new test, considering all possible cutoff scores via a comparison of ROC curves, is significantly higher ($d = .05$, $p < .05$). When the sample was separated according to sex, the effect remained similarly high for both men and women, although a significant outcome could be shown for only the female sample, due to the comparably small subsample of male subjects (see Table 3 for details).

For the total sample, the best cutoff score for the DIA-S was 3.5, with a sensitivity of 82% and a specificity of 79%, whereas the GDS15 showed lower detection rates, with the best cutoff score being 5.5 (sensitivity: 79%; specificity: 71%). There were differentiated gender effects in test efficacy. For the GDS15, the detection rate of depressed subjects was slightly lower in the male subsample (77%), whereas specificity was reduced (69%) in the female sample. For the DIA-S, the diagnostic accuracy in the male sample could be improved by lowering the cutoff score to 2.5. In this way high detection rates could be achieved for both the male (sensitivity: 84%; specificity: 80%) and the female subsamples (sensitivity: 87%; specificity: 79%).

Impact Factors on Detection Rates

Sequential logistic regression analyses were performed to examine possible impact factors on prediction rates of each scale as a two-category outcome (1 = *false prediction*; 0 = *true prediction*) using a subsample of $n = 195$ participants with complete assessment data. The diagnostic accuracy of the scales for this subsample did not differ from the original sample (DIA-S: χ^2 (*ns*;1, $N = 331$) = .277; GDS15: χ^2 (*ns*;1, $N = 331$) = .156).

As the difficulty of prediction varies considerably with the true disease status of a patient, classification difficulty

Table 4. Logistic regression analyses of the diagnostic accuracy of the GDS as a function of patient and setting variables (n = 195)

					95% CI	
Predictor	*B*	*SE B*	Wald	OR	Upper	Lower
Model 1 Difficulty of classification	.153	.051	8.822**	1.165	1.053	1.288
Model 2 Difficulty of classification	.152	.053	8.311**	1.164	1.050	1.290
Age	.051	.024	4.510*	1.053	1.004	1.103
Model 3 Difficulty of classification	.150	.053	8.066**	1.161	1.047	1.288
Age	.049	.024	4.131*	1.051	1.002	1.102
Sex	.211	.374	.318	1.235	.594	2.586
Model 4 Difficulty of classification	.132	.055	5.881*	1.141	1.026	1.270
Age	.056	.025	5.193*	1.058	1.008	1.111
Sex	.171	.382	.200	1.187	.561	2.510
Cognitive impairment	–.102	.053	3.648	.903	.991	1.020
Functional impairment	.006	.007	.597	1.006	.991	1.020
Duration of daylight	.005	.004	1.806	1.005	.998	1.012

Note. **p < .01, *p < .05. CI: confidence interval. A positive odds ratio increases the probability of false classification of the patient by the GDS15.

Table 5. Logistic regression analyses of the diagnostic accuracy of the DIA-S as a function of patient and setting variables (n = 195)

					95% CI	
Predictor	*B*	*SE B*	Wald	OR	Upper	Lower
Model 1 Difficulty of classification	.224	.064	12.165**	1.251	1.1.03	1.419
Model 2 Difficulty of classification	.224	.064	12.286**	1.252	1.1.4	1.419
Age	–.011	.025	.212	.989	.942	1.038
Model 3 Difficulty of classification	.235	.065	13.061**	1.265	1.114	1.438
Age	–.005	.025	.047	.995	.946	1.045
Sex	–.645	.404	2.544	.525	.237	1.159
Model 4 Difficulty of classification	.233	.067	12.107**	1.262	1.107	1.439
Age	–.006	.025	.057	.994	.946	1.045
Sex	–.627	.407	2.375	.534	.241	1.186
Cognitive impairment	.007	.057	.017	1.007	.902	1.125
Functional impairment	.003	.008	.124	1.003	.987	1.019
Duration of daylight	.001	.004	.020	1.001	.992	1.009

Note. **p < .01, *p < .05. CI: confidence interval. A positive odds ratio increases the probability of false classification of the patient by the DIA-S.

based on the MADR score was included in the first model respectively. Patient and setting variables were included stepwise, starting with the patient's age (Model 2) and sex (Model 3), and then entering cognitive and functional impairment and the duration of daylight (Model 4). The degree of multicollinearity in the data was not problematic, with tolerance coefficients between .90 and .96.

Misclassification by the GDS15

A total of n = 53 (27%) participants were misclassified by the GDS15. With the Hosmer-Lemeshow Test, goodness of fit was acceptable for all models $\chi^2_{HL(ns;8;N = 195)}$ = 2.586–12.614, whereas a comparison of log-likelihood ratios showed significant improvement of prediction only for

the first and the second model (Model 1: $\chi^2_{(.01;1;\,N=195)}$ = 9.643, Model 2 $\chi^2_{(.05;1;\,N=195)}$ = 4.898, Nagelkerke R^2 = .10). Regression coefficients, Wald statistics, and odds ratios for the predictors are presented in Table 4. False prediction of the depressive status by the GDS15 (Model 2) was associated with the true disease status (classification difficulty: Wald $\chi^2_{(.01;1;\,N=195)}$ = 8.311) and the age of the participant (Wald $\chi^2_{(.05;1;\,N=195)}$ = 4.131). The risk of a false outcome on the GDS15 was increased by 5% per year of life.

Misclassification by the DIA-S

The disease status predicted by the DIA-S was false for *n* = 37 (19%) participants. Although the goodness of fit was sufficient for all models $\chi^2_{HL(ns;8;N=195)}$ = 5.925–13.419, log-likelihood ratios could only be improved significantly by the addition of the true disease status (Model 1: $\chi^2_{(.01;1;\,N=195)}$ = 14.348, Nagelkerke R^2 = .11, see Table 5 for details), indicating that other patient or setting characteristics had no impact on the performance of the scale.

Discussion

Interpretation of the results

In the current study, the test accuracy of a new screening scale for geriatric depression, the DIA-S, was compared to the GDS15 in a sample of geriatric inpatients. The diagnostic accuracy of the scales was tested in a setting that reflected clinical practice as much as possible. Thus, the sample included a considerable proportion of subjects with minor depression, subclinical syndromes, and cognitive or physical impairment. Whereas all results show a better diagnostic accuracy for the DIA-S, in particular the specificity is superior, indicating that multimorbidity or the inpatient setting have less influence on the new scale. Furthermore, logistic regression analyses confirmed that, aside from the difficulty of classification based on true disease status, there was no systematic impact of patient and setting characteristics on the overall classification rates of the scale. The risk of false classification by the GDS15 was influenced by patients' age, though, unexpectedly, there was no association with functional impairment. This could mean that other aspects of multimorbidity that are more stable than current functional status and thus more strongly associated with old age are more likely to interfere with the detection of depression. Both scales were not influenced by the cognitive status of participants, meaning that the DIA-S proved to be as appropriate as the GDS15 for use with persons with less severe cognitive impairment.

Although the patients' sex had no significant influence on the overall classification rates in the logistic regression, gender-related detection rates could be seen when diagnostic accuracy was differentiated into sensitivity and specificity of the tests. Thus, in male subjects the risk for false-negative results in self-report tends to be higher, whereas the female subsample shows a higher proportion of false-positive responses. This trend leads to higher misclassification rates for the GDS15 in the corresponding subsamples. In the new scale, this effect can be compensated for by lowering the cutoff score in the male subsample, making detection rates similarly high for both sexes. Although the literature (Sigmon et al., 2005; Semmler & Klumb, 2004) recognizes the tendency for women to score higher in self-rating scales of depression regardless of their true disease status, in validation studies it is seldom addressed directly, with results for men and women being reported separately, as we have done here. Nevertheless, our findings indicate that gender-related outcome analyses might improve detection rates in clinical practice.

Limitations of the Current Study

The results of the current study are based on the validation of the DIA-S in a German sample. This means that the English version of the scale, as presented here, has not yet been tested empirically. Thus, further data are necessary to gain an impression of the usefulness of the scale among English-speaking populations.

This study reports differentiated outcomes for the male and female subsamples, respectively. Because the male subsample consisted of less than 100 subjects, these findings can only be seen as a first indication of how scale performance might be moderated by sex; it remains to be seen whether this effect is stable among different samples.

As complete assessment data were not available for the total sample, logistic regression analyses were performed on a subsample of *n* = 195 subjects; however, the characteristics of these subjects were similar to those of the total sample and thus can be seen as a reliable basis for the outcomes.

In this study the DIA-S proved to be a good alternative to the GDS-15, the MADR Scale being used as a gold standard. As this is the first validation study of the DIA-S, the good performance of the new scale could partly be based on some similarity of the MADR-S and the new scale. Further validation using other gold standards such as standardized diagnostic clinical interviews would be useful to provide broader and more reliable empirical evidence of the scale's quality. Another limitation is that this study does not present interrater reliability for the two screening scales. For further analyses of the scales, this would be an important issue that should be addressed. Furthermore, the DIA-S as well as the GDS15 were presented to participants in an oral interview. This way of application was used, because it is the typical way of screening with geriatric inpatients. However, based on the results of our study, conclusions on reliability and validity of the scales are limited to screening by oral interview. Further studies on how people

react to the scales applied in written form or by telephone interview would be necessary to get more information about the quality of the new instrument compared to the GDS15.

Usefulness of the DIA-S in Clinical Practice

During data collection, the DIA-S proved to be a scale that could be easily and quickly administered in an oral interview with the patient. Consisting of only ten items, the scale is considerably shorter than the GDS15 and can also be used with patients in poor general condition. The interviewers reported that the acceptance of the new scale was high, although for the patients talking about their depressive symptoms remained an intimate subject that had to be addressed with sensitivity. For clinical practice we therefore recommend implementing the scale into a broader diagnostic context that is based on existing mutual trust between the patient and the interviewer. Because the DIA-S is a screening instrument that does not replace clinical diagnoses, further diagnostic investigation into the form and severity of the disease, the suicidal tendency, and possible comorbidities would be necessary in the case of a positive screening result.

The results of the current study show the DIA Scale to be a useful screening scale to improve detection of depressive symptoms in clinical practice as well as in research. Further investigation with different samples and in different settings would therefore be worthwhile.

Acknowledgments

We would like to thank Amy Michéle for the translation of the scale.

Declaration of Conflicts of Interest

The authors declare that no conflicts of interest exist.

References

AGAST. (1997). *Geriatrisches Basisassessment. Handlungsanleitungen für die Praxis* [Geriatric Assessment. User manual]. München: Medizin-Verlag.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders, 5th edition (DSM-5)*. Washington, DC: American Psychiatric Publishing.

Burke, W., Roccaforte, W., & Wengel, S. (1991). The short form of the Geriatric Depression Scale: A comparison with the 30-item form. *Journal of Geriatric Psychiatry and Neurology, 4,* 173–177.

Camozzato, A., Hidalgo, M. Souza, S., & Chaves, M. (2007). Association among items from the self-report version of the Hamilton Depression Scale and respondents sex. *Psychological Reports, 101,* 291–301.

De Craen, A. J., Gussekloo, J., van der Mast, R. C., le Cessie, S., Lemkes, J. W., & Westendorp, R. G. (2005). Seasonal mood variation in the elderly: The Leiden 85-plus study. *International Journal of Geriatric Psychiatry, 20,* 269–273.

Folstein, M., Folstein, S., & McHugh, P. R. (1975). Mini-mental state. A practical method for grading the cognitive state for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Goring, H., Baldwin, R., Marriott, A., Pratt, H., & Roberts, C. (2004). Validation of short screening tests for depression and cognitive impairment in older medically ill inpatients. *International Journal of Geriatric Psychiatry, 19,* 465–471.

Heidenblut, S., & Zank, S. (2010). Entwicklung eines neuen Depressionsscreenings für den Einsatz in der Geriatrie [Development of a new screening instrument for geriatric depression]. *Zeitschrift für Gerontologie und Geriatrie, 43*, 170–176.

Henkel, V., Mergl, R., Kohnen, R., Möller, H-J., Hegerl, U., & Allgaier, A.-K. (2004). Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. *General Hospital Psychiatry, 23,* 190–198.

Incalzi, R. A., Cesari, M., Pedone, C., & Carbonin, P. U. (2003). Construct validity of the 15-item Geriatric Depression Scale in older medical inpatients. *Journal of Geriatric Psychiatry and Neurology, 16*, 23–27.

Jongenelis, K., Gerritsen, D. L., Pot, A. M., Beekman, A. T. F., Eisses, A. M. H., Kluiter, H. ... Ribbe, M. W. (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.

Kenny Weeks, S., McGann, P. E., King Michaels, T., & Penninx, B. W. J. H. (2003). Comparing various short-form geriatric depression scales leads to the GDS-5/15. *Journal of Nursing Scholarship, 35,* 133–137.

Lasser, R., Siegel, E., Dukoff, R., & Sunderland, T. (1998). Diagnosis and treatment of geriatric depression. *CNS Drugs, 1*, 17–30.

Leentjens, A., Verhey, F., Lousberg, R., Spitsbergen, H., & Wilmink, F. (2000). The validity of the Hamilton and Montgomery-Asberg Depression Rating Scales as screening and diagnostic tools for depression in Parkinson disease. *International Journal of Geriatric Psychiatry, 15,* 644–649.

Mahoney, F. I., & Barthel, D. (1965). Functional evaluation: The Barthel index. *Maryland State Medical Journal, 14,* 56–61.

McCusker, J., Cole, M., Ciampi, A., Windholz, S., & Belzile, E. (2007). Major depression in older medical inpatients predicts poor physical and mental health status over 12 months. *General Hospital Psychiatry, 29,* 340–348.

McCusker, J., Cole, M., Ciampi, A., Belzile, E., Cepoiu, M., & Sewitch, M. (2008). Recognition of depression in older medical inpatients discharged to ambulatory care settings: A longitudinal study. *General Hospital Psychiatry, 30,* 245–251.

McCusker, J., Cole, M., Dufouil, C., Dendukuri, N., Latimer, E., Windholz, S., & Elie, M. (2005). The prevalence and correlates of major and minor depression in older medical inpatients. *Journal of the American Geriatrics Society, 53,* 1344–1353.

Montgomery, S., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry, 134,* 382–389.

Mottram, P., Wilson, K., & Copeland, J. (2000). Validation of the

Hamilton Depression Rating Scale and Montgomery and Asberg Rating Scales in terms of AGECAT Depression cases. *International Journal of Geriatric Psychiatry, 12,* 1113–1119.

Müller, J. M., Szegedi, A., Wetzel, H., & Benkert, O. (2000). Moderate and severe depression gradations for the Montgomery-Asberg Depression Rating Scale. *Journal of Affective Disorders, 60,* 137–140.

Neumann, N.-U., & Schulte, R.-M. (1989). *Montgomery and Asberg Depression Rating Scale. Deutsche Fassung* [Montgomery and Asberg Depression Rating Scale. German Version]. Erlangen: Perimed Fachbuch Verlagsgesellschaft.

Pepe, M. (2004). *The statistical evaluation of medical tests for classification and prediction.* Oxford: Oxford University Press.

Pomeroy, I., Clark, C., & Philp, I. (2001). The effectiveness of very short scales for depression screening in elderly medical patients. *International Journal of Geriatric Psychiatry, 16*, 321–326.

Renteln-Kruse, W., & Ebert, D. (2003). Merkmale hospitalisierter geriatrischer Patienten: Zwei Kohorten im Vergleich unter Verwendung des Screenings der Arbeitsgemeinschaft Geriatrisches Basisassessment (AGAST). [Characteristics of hospitalized geriatric patients: A comparison of two cohorts using the screening of the Arbeitsgemeinschaft Geriatrisches Basisassessment (AGAST)]. *Zeitschrift für Gerontologie und Geriatrie, 36,* 223–232.

Rinaldi, P., Mecocci, P., Benedetti, C., Ercolani, S., Bregnocchi, M., Menculini, G., . . . Cherubini, A. (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.

Schlager, D., Schwartz, J. E., & Bromet, E. J. (1993). Seasonal variations of current symptoms in a healthy population. *The British Journal of Psychiatry, 163,* 322–326.

Semmler, G., & Klumb, P. L. (2004). Geschlechtsunterschiede in der Prävalenz depressiver Symptomatik: Ein Resultat differentieller Validität der Erhebungsinstrumente? [Gender differences in the prevalence of depressive symptoms: A consequence of differentially valid measurement instruments?]. *Journal of Public Health, 12,* 139–146.

Shah, A., Phongasthorn, V., Bielawska, C., & 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.

Sigmon, S., Pells, J., Boulard, N., Whitcomb-Smith, S., Edenfield, T., Hermann, B., . . . Kubik, E. (2005). Gender differences in self-reports of depression: The response bias hypotheses revisited. *Sex Roles, 53,* 401–411.

Sutcliffe, C., Cordingley, L., Burns, A., Mozley, C. G., Bagley, H., Huxley, P. . . . & Challis, D. (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.

Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A. . . . Aitchison, K. J. (2008). Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychological Medicine, 38,* 289–300.

Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., & Friedrich, F. (2006). The criterion validity of the Geriatric Depression Scale: A systematic review. *Acta Psychiatrica Scandinavia, 114,* 398–410.

World Health Organization (WHO). (2010). *International statistical classification of diseases and related health problems – 10th revision (ICD-10).* Geneva, CH: Author.

Yesavage, J., Brink, T., Rose, T., Lum, O., Huang, V., Adey, M., & Leirer, O. (1983). Development and validation of a geriatric depression scale: A preliminary report. *Journal of Psychiatric Research, 17,* 37–49.

Manuscript submitted: 31 July 2013
Manuscript accepted after revision: 12 December 2013

Dr. Sonja Heidenblut

Department of Rehabilitative Gerontology
University of Cologne
Herbert-Lewin-Str. 2
50931 Cologne
Germany
Sonja.Heidenblut@uni-koeln.de