[Environmental Health and Preventive Medicine 11, 177–183, July 2006]

Original Article

Agreement in Depression Determination among Four Self-Rating Depression Scales Applied to Japanese Community-Dwelling Elderly

Shinichi DEMURA[1], Susumu SATO[2], Nobuhiko TADA[3], Jinzaburo MATSUZAWA[4] and Hiroshi HAMASAKI[5]

[1]*Kanazawa University, Graduate School of Natural Science and Technology, Japan*
[2]*Kanazawa Institute of Technology, Life-long Sports Core, Japan*
[3]*Fukui Prefectural University, Japan*
[4]*Fukui University of Technology, Japan*
[5]*Kyoto Pharmaceutical University, Japan*

Abstract

Objective: In this study, we examined the characteristics of depression determination using four representative self-rating depression scales (Geriatric Depression Scale, GDS; Self-rating Depression Scale, SDS; Center for Epidemiologic Studies Depression Scale, CES-D; and Carroll Rating Scale, CRS) applied to Japanese community-dwelling elderly.

Methods: Subjects were 563 community-dwelling independent elderly living in twelve prefectures (330 males, 68.9±6.3 yr; 233 females, 68.1±5.8 yr).

Results: Depression rates determined using SDS (45.8%) and CES-D (68.6%) were higher than those determined using GDS (5.7%) and CRS (14.7%). Although correlations of depression scale scores among the four scales were significant and comparable (r: 0.61 (GDS vs. SDS, $p<0.01$) to 0.78 (SDS vs. CES-D, $p<0.01$)), the agreement in depression determination varied among scales (kappa coefficients: 0.05 (GDS vs. CES-D, $p>0.05$) to 0.46 (SDS vs. CES-D, $p<0.01$)).

Conclusions: Similarities in depression determination were found between GDS and CRS, and between CES-D and SDS. Depression rates determined on the basis of cut-off point for each scale were higher for CES-D and SDS than for GDS and CRS. Depression determination using a four-point rating scale may overestimate a slightly depressive symptom, compared with that using a two-point scale.

Key words: Geriatric Depression Scale (GDS), Self-Rating Depression Scale (SDS), Center for Epidemiologic Studies Depression Scale (CES-D), Carroll Rating Scale (CRS)

Introduction

For the elderly, physical functional decline with aging causes inactivity in daily life, and directly influences their health status (1, 2). In addition, the elderly have most likely experienced life events that negatively affect their mental health, such as retirement, loss of economic and social infrastructure, and bereavement for a deceased spouse, family members and close friends (3–5). Because the deterioration of mental health is considered to negatively affect activities in daily life and physical condition, the assessment and maintenance of mental health are considered to be important, particularly in the case of the elderly (6–10).

The psychological features of the elderly have been assessed from various aspects of life-satisfaction, happiness, self-efficacy, depression and "*ikigai* (purpose of life)". Among them, depression is used as an index of psychological well-being. As mentioned above, considering the personal and social environments of the elderly, the assessment of depression is considered to be important for the elderly.

Depression is categorized as a psychological disease, and it is diagnosed on the basis of criteria, such as the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (11). In addition, self-rating depression scales, such as the Geriatric Depression Scale (GDS) (12), the Self-rating Depression Scale (SDS) (13), the Center for Epidemiologic Studies Depression Scale (CES-D) (14) and the Carroll Rating

Received Dec. 16, 2005/Accepted Apr. 3, 2006
Reprint requests to: Susumu SATO
Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan
TEL: +81(76)248-1100 (ex. 2386), FAX: +81(76)294-6704
E-mail: sssato@neptune.kanazawa-it.ac.jp

Scale (CRS) (15), have been developed. Japanese versions of these scales have also been developed, and their reliability and validity have been examined (16–20). These scales are used not only for the evaluation of the severity of depression in patients, but also for the screening of depressive symptoms in a general population. They determine a depressive symptom on the basis of a cut-off point that is set as the total score for each scale. From the view points of simplicity and versatility, these self-rating scales are useful for assessing depressive symptoms or an indicium in a general elderly population.

In contrast, with advancing age in society, studies of depressive symptoms or an indicium in a general elderly population have been attempted to assess them as a component of quality of life (QOL). Because the depression self-rating scales were developed for the diagnosis of depression, their validities have been examined by case-control studies on the basis of clinical diagnosis or by comparison studies using an existing depression self-rating scale as the validation criterion. However, previous studies of depression in general Japanese elderly population, which examined the prevalence of depression and its relevant factors for general Japanese elderly population, were not sufficient (4). There are few studies that examined either the consistency in depression determination when applying several depression scales to a general elderly population, or studies that compared the sensitivity of several self-rating scales in depression determination (i.e., the difference in depression determination in the same person using different scales) when screening depression symptoms in a general elderly population. These examinations are considered valuable as a descriptive study of a general elderly population. Furthermore, they may provide useful information when using these depression scales for assessing subjective or health-related QOL of general Japanese elderly population.

In this study, we aimed to examine the characteristics of depression determination using four self-rating depression scales, namely, GDS, SDS, CES-D and CRS, in a general community-dwelling Japanese elderly population.

Methods

1. Subjects and data collection

We conducted surveys at Kushiro City in Hokkaido Prefecture, Akita City in Akita Prefecture, Sendai City in Miyagi Prefecture, Morioka City in Iwate Prefecture, Mito City in Ibaraki Prefecture, Iida City in Nagano Prefecture, Gifu City in Gifu Prefecture, Nagoya City in Aichi Prefecture, Kanazawa City in Ishikawa Prefecture, Fukui City in Fukui Prefecture, Kyoto City in Kyoto Prefecture and Yonago City in Tottori Prefecture on the basis of intentional survey sampling. We carried out questionnaire surveys on 1,400 subjects (100 to 300 questionnaires were distributed to each area), and collected data from 1,325 subjects (94.6%). After confirming data loss for gender, age and disease history, valid data from 1,275 subjects (91.1%) were obtained. Furthermore, 563 subjects, who completed the four depression scales, were used for the analysis in this study (40.2%).

The investigators were researchers working at universities in each prefecture. The survey method was selected by each investigator, such as placement survey, educational class setting and mail survey methods, taking into consideration of the state of each subjects. In the case of subjects with poor eyesight, an investigator read questions to them. Prior to the survey, we explained the aim and design of the study to each subject before obtaining their written informed consent. The investigators explained to the subjects that they could refuse to participate in the survey, and that participation would not adversely affect their privacy. This study was approved by the Human Subject Ethical Committee of Kanazawa University.

Subjects were 563 community-dwelling independent elderly people aged 60 to 87 years old (330 males: 68.9±6.3 yr, 233 females: 68.1±5.8 yr). The characteristics of the subjects are shown in Table 1.

Among the subjects, 446 (79.2%) lived with a spouse. Among those living without a spouse, 24 (4.3%) lived alone, and 49 (8.9%) lived with children or other family relations. Among all the subjects, 246 (43.7%) did not work and 180 (32.0%) were working with an income.

Among the subjects, 344 (61.1%) were attending a hospital, and 438 (77.8%) had diseases involving the following: cardiovascular system, 154 (27.4%); endocrine system, 72 (12.8%); digestive system, 27 (4.8%); and musculoskeletal system, 27 (4.8%). Only one subject (0.2%) responded that he/she has a psychiatric disorder. Comparing their physical fitness level with others in the same age group, 357 (63.4%) subjects self-evaluated their fitness level as “average”, 123 (21.8%) as “good” or “above average”, and 80 (14.2%) as “poor” or “very poor”. Similarly, for health condition, 365 (64.8%) subjects self-evaluated their condition as “average”, 141 (25.0%) as “good” or “very good”, and 56 (9.9%) as “poor” or “very poor”.

Five hundred forty-five (96.8%) subjects have breakfast every day, 469 (83.3%) did not have a smoking habit, 225 (40.0%) did not have a drinking habit, and 144 males (25.6%) drank every day. Regarding exercise habit, 275 (48.8%) subjects exercised mote than two to three time a week and 177 (31.5%) did not. Regarding sleeping time, 418 (74.2%) subjects slept more than six hours. Four hundred ninety-two (87.4%) responded of having several close friends and 450 (79.9%) participated in hobbies or community volunteer activities.

2. Self-rating depression scales

This study used GDS (12), SDS (13), CES-D (14) and CRS (15). All scales have a cut-off point for evaluating depression symptoms. The GDS and CRS are two-point rating scales (yes or no), and CES-D and SDS are four-point rating scales (1, rarely or never; 2, sometime; 3, occasionally; and 4, most or all of the time). The values of these response categories are reversed for the positive affect items. The total score was the sum of item scores for each scale.

3. Statistical analyses

1) Examination of characteristics of depression determination using each depression scale

Mean and standard deviation values in total scale score were calculated for each depression scale, and gender and age differences in the total scale score were confirmed by two-way

Table 1 Chracteristics of study sample

Content		Category					Non-response
Sample size		60–69 years old	70–79 years old	80–89 years old	Total		
	Male	184	125	21	330		
	Female	152	70	11	233		
	Total	336	195	32	563		
Family structure		With spouse	Without spouse				
	Total	446 (79.2)	113 (20.1)	(Alone)	(With child(ren) or grandchildren)	(Others)	4 (0.7)
	Male	277 (49.2)	51 (18.3)	8 (1.4)	18 (3.2)	25 (4.4)	2 (0.4)
	Female	169 (30.0)	62 (11.4)	16 (2.8)	31 (5.5)	47 (8.3)	2 (0.4)
Occupation		Working			Do not work		
	Total	306 (54.4)	(With an income)	(Without an income)	246 (43.7)		11 (2.0)
	Male	172 (30.6)	118 (38.6)	54 (17.6)	154 (27.4)		4 (0.7)
	Female	134 (23.8)	62 (20.3)	72 (23.5)	92 (16.3)		7 (1.2)
Attending a hospital		Yes	No				
	Total	344 (61.1)	208 (36.9)				11 (2.0)
	Male	201 (35.7)	125 (22.2)				4 (0.7)
	Female	143 (25.4)	83 (14.7)				7 (1.2)
Disease		With disease	Without disease				
	Total	438 (77.8)	108 (19.2)				17 (3.0)
	Male	266 (47.2)	53 (9.4)				11 (2.0)
	Female	172 (30.6)	55 (9.8)				6 (1.1)
*Multiple responses		Immune sys.	Endocrine sys.	Psychiatric disoder	Nerve sys.	Ocular sys.	
	Total	2 (0.4)	72 (12.8)	1 (0.2)	4 (0.7)	18 (3.2)	
	Male	2 (0.4)	44 (7.8)	1 (0.2)	4 (0.7)	9 (1.6)	
	Female	0 (0.0)	28 (5.0)	0 (0.0)	0 (0.0)	9 (1.6)	
		Auditory sys.	Cradiovascular sys.	Respiratory sys.	Digestive sys.	Musculoskeletal sys.	
	Total	6 (1.1)	154 (27.4)	9 (1.6)	27 (4.8)	27 (4.8)	
	Male	5 (0.9)	90 (16.0)	5 (0.9)	18 (3.2)	16 (2.8)	
	Female	1 (0.2)	64 (11.4)	4 (0.7)	9 (1.6)	11 (2.0)	
Self-evaluated physical fitness		Very poor	Poor	Average	Above average	Good	
	Total	16 (2.9)	64 (11.4)	357 (63.4)	88 (15.7)	35 (6.2)	3 (0.5)
	Male	11 (2.0)	36 (6.4)	197 (35.0)	59 (10.5)	26 (4.6)	1 (0.2)
	Female	5 (0.9)	28 (5.0)	160 (28.4)	29 (5.2)	9 (1.6)	2 (0.4)
Self-evaluated health status		Very poor	Poor	Average	Good	Very good	
	Total	3 (0.6)	53 (9.4)	365 (64.8)	119 (21.1)	22 (3.9)	1 (0.2)
	Male	2 (0.4)	34 (6.0)	197 (35.0)	81 (14.4)	15 (2.7)	1 (0.2)
	Female	1 (0.2)	19 (3.4)	168 (29.8)	38 (6.7)	7 (1.2)	0 (0.0)
Habit of breakfast		Every day	Sometime	Do not eat			
	Total	545 (96.8)	10 (1.7)	7 (1.3)			1 (0.2)
	Male	316 (56.1)	7 (1.2)	6 (1.1)			1 (0.2)
	Female	229 (40.7)	3 (0.5)	1 (0.2)			0 (0.0)
Habit of eating between meals		Do not eat	Sometime	Almost every day			
	Total	85 (15.1)	378 (67.1)	96 (17.0)			4 (0.7)
	Male	69 (12.3)	210 (37.3)	49 (8.7)			2 (0.4)
	Female	16 (2.8)	168 (29.8)	47 (8.3)			2 (0.4)
Smoking habit		Do not smoke	5 cigarettes a day	10 cigarettes a day	20 cigarettes or more a day		
	Total	469 (83.3)	14 (2.5)	30 (5.3)	41 (7.3)		9 (1.6)
	Male	251 (44.6)	12 (2.1)	25 (4.4)	41 (7.3)		1 (0.2)
	Female	218 (38.7)	2 (0.4)	5 (0.9)	0 (0.0)		8 (1.4)
Drinking habit		Do not drink	Very little	Sometimes	Every day		
	Total	225 (40.0)	65 (11.5)	96 (17.0)	158 (31.1)		19 (3.4)
	Male	73 (13.0)	38 (6.7)	65 (11.5)	144 (25.6)		10 (1.8)
	Female	152 (27.0)	27 (4.8)	31 (5.5)	14 (2.5)		9 (1.6)
Frequency of exercise		Almost every day	2 or 3 times a week	1 or 2 times a month	A few times a year	None	
	Total	97 (17.3)	178 (31.6)	75 (16.3)	21 (3.7)	177 (31.5)	15 (2.7)
	Male	73 (13.0)	97 (17.2)	44 (7.8)	16 (2.8)	95 (16.9)	5 (0.9)
	Female	24 (4.3)	81 (14.4)	31 (5.5)	5 (0.9)	82 (14.6)	10 (1.8)
Hours of sleep		Less than 5 hours	5 to 6 hours	6 to 7 hours	7 to 8 hours	More than 8 hours	
	Total	24 (4.3)	118 (21.0)	213 (37.9)	144 (25.6)	61 (10.8)	3 (0.5)
	Male	14 (2.5)	60 (10.7)	118 (21.0)	93 (16.5)	43 (7.6)	2 (0.4)
	Female	10 (1.8)	58 (10.3)	95 (16.9)	51 (9.1)	18 (3.2)	1 (0.2)
Number of friends		Plenty	Several	1 person	None		
	Total	159 (28.2)	333 (59.2)	16 (2.8)	53 (9.4)		2 (0.4)
	Male	88 (15.6)	193 (34.3)	8 (1.4)	40 (7.1)		1 (0.2)
	Female	71 (12.6)	140 (24.9)	8 (1.4)	13 (2.3)		1 (0.2)
Hobbies or volunteer activity		Regular	Sometimes	None			
	Total	296 (53.6)	154 (27.4)	107 (19.0)			6 (1.1)
	Male	183 (32.5)	81 (14.4)	63 (11.2)			3 (0.5)
	Female	113 (20.1)	73 (13.0)	44 (7.8)			3 (0.5)

Values are expressed as n (%).

ANOVA. Then, the rate of people with depression symptoms, which was determined by cut-off point of each scale, was calculated. Significant differences in depression rate among four scales were tested by Cochran's Q-test. When a significant difference was found ($p<0.05$), McNemar's test for proportion was used for multiple comparisons of all pairs. Significant level was controlled by Bonferroni correction for multiple comparisons (significant level was set at 0.05/6=0.00083).

For CES-D and SDS, which are four-point scales, the relative frequency of each category was calculated for each item.

2) Agreement in depression determinations among four scales

The agreement in depression determinations based on the cut-off points of the four scales was examined. Firstly, the consistency in depression determination for each combination of the four scales was examined by cross tabulation analysis (2×2). The frequency and relative frequency of each category and kappa coefficient were calculated. Furthermore, Correlations among the four scale scores were examined using Pearson's correlation coefficient for total scale scores.

Then, the agreement in depression determination among the four scales was examined. The number of subjects showing agreement in depression determination among the four scales, those showing agreement among three scales, and those showing agreement among two scales were calculated.

Results

1. Depressive symptom rate determined using four depression scales

Table 2 shows descriptive statistics and gender and age differences in the total scale score, and depressive symptom

Table 2 Descriptive statistics and gender and age differences in total scale score, and depressive symptom rate in each depression scale

Age group	Total	60–69		70–79		80–89		Two-way ANOVA			Multiple comparisons		Depressive rate (%)
Sex		Male (n=184)	Female (n=152)	Male (n=125)	Female (n=70)	Male (n=21)	Female (n=11)	Gender	Age	Int.	Gender	Age	
GDS	4.4 (2.8)	3.9 (2.8)	4.2 (2.6)	4.8 (3.1)	5.1 (2.9)	5.2 (2.6)	5.4 (3.1)	ns	**	ns		60–69<70–79	5.7%
SDS	38.7 (6.8)	37.7 (6.8)	38.7 (6.6)	39.2 (7.4)	40.0 (7.9)	40.2 (7.1)	39.4 (8.0)	ns	ns	ns			45.8%
CES-D	19.1 (7.4)	18.7 (7.4)	18.9 (7.1)	19.4 (8.1)	19.5 (8.7)	19.7 (6.2)	19.2 (7.8)	ns	ns	ns			68.6%
CRS	10.3 (6.3)	9.3 (6.3)	9.3 (5.5)	11.2 (6.5)	12.4 (6.9)	12.0 (8.2)	10.2 (5.5)	ns	**	ns		60–69<70–79	14.7%

Values are expressed as mean (SD) or %. Int.: Interaction.
GDS: Geriatric depression scale, SDS: Self-rating depression scale, CES-D: Center for epidemiologic studies depression scale, CRS: Carroll rating scale.
Significant differences in depressive rates were found among four scales, and depressive rates were significantly higher in the order of CES-D, SDS, CRS and GDS.

Table 3 Categorical relative frequency of each item of CES-D and SDS

	CES-D					SDS			
Item No.	Category 1	Category 2	Category 3	Category 4	Item No.	Category 1	Category 2	Category 3	Category 4
1	12.6	59.3	25.6	2.5	1	62.9	32.5	4.4	0.2
2	47.6	45.3	6.4	0.7	2	16.5	43.2	29.1	11.2
3	39.1	51.0	9.6	0.4	3	30.4	54.9	13.0	1.8
4	20.8	63.9	12.6	2.7	4	42.3	38.4	16.0	3.4
5	22.7	52.9	22.2	2.1	5	31.3	57.2	7.5	4.1
6	34.1	52.4	11.5	2.0	6	8.9	43.5	34.1	13.5
7	17.9	51.2	29.0	2.0	7	44.8	40.9	12.4	2.0
8	6.7	59.7	29.1	4.4	8	52.2	29.0	15.3	3.6
9	29.0	58.3	11.5	1.2	9	52.6	36.6	9.4	1.4
10	26.6	54.5	17.4	1.4	10	28.2	36.1	32.7	3.0
11	27.7	41.4	27.5	3.4	11	17.9	55.4	23.8	2.8
12	15.5	63.2	19.5	1.8	12	19.7	59.1	17.8	3.4
13	18.5	60.2	20.1	1.2	13	39.6	50.1	8.5	1.8
14	52.2	41.9	5.2	0.7	14	9.4	47.6	37.7	5.3
15	27.5	61.5	10.1	0.9	15	45.5	49.0	5.2	0.4
16	20.2	59.5	17.6	2.7	16	10.8	50.8	33.9	4.4
17	30.4	54.9	13.0	1.8	17	12.1	55.8	29.5	2.7
18	24.0	59.3	14.6	2.1	18	15.1	67.0	16.0	2.0
19	36.8	58.3	4.8	0.2	19	60.2	34.1	5.0	0.7
20	21.8	52.6	24.0	1.6	20	25.2	63.2	11.0	0.5
Mean (%)	26.6	55.1	16.6	1.8	Mean (%)	31.3	47.2	18.1	3.4

Categories 1 to 4 are rating scale of CES-D and SDS scales: Category 1: rarely or none of the time, Category 2: some or a little of the time, Category 3: occasionally or a moderate amount of the time, and Category 4: most or all of the time.
The values of these response categories are reversed for positive affect items.

Table 4 Results of cross tabulations for depression symptom determination among four scales

		SDS		CRS		CES-D	
		Positive	Negative	Positive	Negative	Positive	Negative
GDS	Positive	28 (5.0%)	4 (0.7%)	20 (3.6%)	12 (2.1%)	32 (5.7%)	0 (0.0%)
	Negative	230 (40.9%)	301 (53.5%)	63 (11.2%)	468 (83.1%)	354 (62.9%)	177 (31.4%)
	Agreeement	329 (58.4%)		488 (86.7%)		209 (37.1%)	
	Disagreement	234 (41.6%)		75 (13.3%)		354 (62.9%)	
	r	0.61 **		0.68 **		0.68 **	
	kappa	0.10 **		0.29 **		0.05 ns	
SDS	Positive			76 (13.5%)	182 (32.3%)	243 (43.2%)	143 (25.4%)
	Negative			7 (1.2%)	298 (52.9%)	15 (2.7%)	162 (28.8%)
	Agreeement			374 (66.4%)		405 (71.9%)	
	Disagreement			189 (33.6%)		158 (28.1%)	
	r			0.68 **		0.78 **	
	kappa			0.29 **		0.46 **	
CRS	Positive					80 (14.2%)	3 (0.5%)
	Negative					306 (54.5%)	174 (30.9%)
	Agreeement					254 (45.1%)	
	Disagreement					309 (54.9%)	
	r					0.63 **	
	kappa					0.13 **	

"Positive" and "Negative" mean the result of depression determination assessed by a cut-off point for each scale, and "Positive" and "Negative" are "with depression symptom" and "without depressive symptom", respectively.
r: Pearson's correlation coefficient in total scale score. *kappa*: kappa coefficient calculated from frequensies of cross tabulation. **: $p<0.01$.

Table 5 Agreement in depression symptom determinations among four scales

			n	%
Agreement among all scales			180	32.0%
Assessed as "positive" in all scales			19	10.6%
Assessed as "negative" in all scales			161	89.4%
Agreement among three scales			209	37.1%
	Positive	Negative		
	SDS, CES-D, CRS	GDS	54	25.8%
	GDS, CES-D, CRS	SDS	0	
	GDS, SDS, CRS	CES-D	0	
	GDS, SDS, CES-D	CRS	8	3.8%
	GDS	SDS, CES-D, CRS	0	
	SDS	GDS, CES-D, CRS	10	4.8%
	CES-D	GDS, SDS, CRS	136	65.1%
	CRS	GDS, SDS, CES-D	1	0.5%
Agreement between two scales			172	30.6%
	Positive	Negative		
	SDS, CES-D	GDS, CRS	160	93.0%
	CES-D, CRS	GDS, SDS	6	3.5%
	SDS, CRS	GDS, CES-D	2	1.2%
	CES-D, GDS	SDS, CRS	4	2.3%
		Total	172	30.6%

"Positive" and "Negative" mean "with depressive symptom" and "without depressive symptom," respectively, when assessed by a cut-off point for each scale.

rate for each depression scale. The depression scale scores (mean±SD) were 4.4±2.8 for GDS, 38.7±6.8 for SDS, 19.1±7.4 for CES-D, and 10.3±6.3 for CRS. Significant age differences were found in GDS and CRS, although no significant gender differences were found in all scales. Depression symptom rates assessed on the basis of the cut-off point of each scale were 5.7% for GDS, 45.8% for SDS, 68.6% for CES-D, and 14.7% for CRS. Significant differences in depression symptom rate were found by Cochran's Q-test. As the results of multiple comparisons (McNemar's test), depression symptom rates were in the following order with statistically significant differences: CES-D>SDS>CRS>GDS.

CES-D and SDS which showed a high depression symptom rate are four-point scales (1, rarely or never; 2, sometime; 3, occasionally; and 4, most or all of the time). To confirm response trends, relative frequency of each category was calculated for each item (Table 3). The values of these response categories were reversed for the positive affect items. CES-D and CRS showed high percentages in category 2, and the mean values of items were 55.1% for CES-D and 47.2% for CRS.

Table 4 shows the results of cross tabulations among the four scales. The highest agreement was found between GDS and CRS (the agreement rate was 86.7%), whereas the lowest agreement was found between GDS and CES-D (the agreement rate was 37.1%).

Regarding correlations among the four scales calculated from the total score, the lowest correlation coefficient was obtained between GDS and SDS ($r=0.61$, $p<0.01$), and the highest correlation coefficient was obtained between SDS and CES-D ($r=0.78$, $p<0.01$). These correlation coefficients were relatively comparable.

The kappa coefficients were calculated on the basis of depression determination that was assessed from the cut-off point of each scale. The highest kappa coefficient was obtained between SDS and CES-D (kappa=0.46, $p<0.01$), the lowest kappa coefficient was obtained between GDS and CES-D

(kappa=0.05, p>0.05). Compared with the correlation coefficients calculated from the total score, kappa coefficients varied according to the combinations of these four scales.

Table 5 shows the consistency of depression determinations among the four scales. For depression determination, 32% of the subjects showed depression determination agreement among all four scales, 37.1% showed agreement among three scales, and 30.6% between two scales. Among the 209 subjects showing agreement among three scales, 65.1% were assessed to be "positive for depression using CES-D, and negative using GDS, SDS and CRS", and 25.8% were assessed to be "negative for depression using GDS, and positive using SDS, CES-D and CRS". Among 172 subjects showing agreement between two scales, 93.0% were assessed as "positive for depression using SDS and CES-D, and negative by GDS and CRS".

Discussion

In this study, depression symptom rates determined using SDS (45.8%) and CES-D (68.6%) were higher than those determined using GDS (5.7%) and CRS (14.7%). In previous studies in Japan and other countries, various depression symptom rates were reported (22–26).

Using SDS, Aoki (23) evaluated 903 Japanese elderly subjects living in an agricultural community, and found that the depressive symptom rate was 4.3% in males and 11.4% in females. Yamashita et al. (24) evaluated depression in 113 healthy Japanese elderly subjects using SDS, and reported that the depressive symptom rate was 9.7% in total; 6% (5 of 80 people) of the elderly living with others, and 18.2% (6 of 33 people) of the elderly living alone. Also using CES-D, depression symptom rate of about 5% was reported in previous studies by Shima et al. (16) who examined depression of Japanese adults, and Ihara (25) who evaluated 695 Japanese elderly living in an agricultural community. On the other hand, Iwata et al. (26) who examined race difference in depression assessment using CES-D, reported the following depression symptom rates: Anglo-Americans, 33.0% (95%CI: 27.7–38.9); Native Americans, 54.2% (48.7–59.6); Argentineans, 31.1% (22.9–40.6); and Japanese, 52.2% (46.5–57.8%). Because there are studies showing depressive symptom rates comparable to ours, our results are not necessarily unusual.

Regarding the correlations among these depression scales calculated from the total score, the correlation coefficients were relatively high and comparable among the four scales. However, the agreement in depression determinations varied among the four scales; it was the lowest between GDS and CES-D, and the highest between SDS and CES-D. Moreover, among the subjects showing disagreement in depression determinations, there were few subjects who were determined to be "having depression (positive)" using GDS or CRS and to be "without depression (negative)" using CES-D or SDS. As shown in Table 4, no subjects were determined to be "positive for depression using GDS and negative using CES-D", 0.7% were determined to be "positive for depression using GDS and negative using SDS", 0.5% were determined to be "positive for depression using CRS and negative using CES-D", and 1.2% were determined to be "positive for depression using CRS and negative using SDS". These results indicate that the sensitivities of these four scales in determining depression differ, and similarities are found between GDS and CRS, and between SDS and CES-D.

From the depression rates of the four scales, it is considered that CES-D (68.6%) and SDS (45.8%) tended to more easily determine a person as "positive for depression" than GDS (5.7%) and CRS (14.7%). As for the causes of these findings, there are two possibilities: One is that CES-D and SDS tended to determine a person as "having depression" despite the person not being depressed. Another is that GDS and CRS tend to determine a person as "without depression" despite the person being depressed. Because this study was not designed as a case (depression group)-control study based on a clinical diagnosis, it is not exactly clear which possibility is true. However, it is important to understand that depressive determination rates obtained using these self-rating scales vary when assessing the same Japanese elderly person in general population.

From the response tendencies, the differences in component (contents of items) and rating scale (two-point scale or four-point scale) among the four scales may affect the results obtained in this study. The depression scales, other than the GDS, include questions for physical symptoms, such as asitia disordered sleep and loss of vigor, which are easily affected by physical disorders often recognized with aging (27, 28). On the other hand, among the depression scales used in this study, only CES-D has subscales. It includes more items concerning "depressive affects" and "physical symptoms" than items concerning "human relations" and "positive affect".

GDS and CRS, which had lower depression determination rates, are two-point rating scales, whereas CES-D and SDS are four-point rating scales. In general, a four-point scale can more easily reflect a slightly depressive symptom on the score than a two-point scale. Indeed, the percentages of the response "some or a little of the time (category 2)" in CES-D and SDS were very high (Table 2: the mean relative frequencies of category 2 were 55.1% for CSE-D and 47.2% for SDS). This finding may indicate that many Japanese elderly persons self-rated a slightly depressive symptom not as "rarely or none of the time" but as "some or a little of the time". Combined with the finding that depression rates were low in GDS and CRS, this may indicate several possibilities. One is that the Japanese elderly tend to overestimate a slightly depressive symptom when evaluating depressive symptoms with a multiple-rating scale. Furthermore, the number of elderly persons with a lightly depressive symptom may be increasing in the Japanese society.

There is a limitation in directly generalizing our results to the Japanese elderly. Although this study used 563 community-dwelling elderly persons from twelve prefectures, there are problems in the sampling method and uniformity of the survey method, and further research is needed. On the other hand, the characteristics of the subjects such as family structures, health conditions, lifestyles and social relationships were not considered to be specific. In general, the mental health of the elderly tends to be negatively influenced by physical deterioration with age and the environments of the elderly. From our results, many Japanese elderly persons may potentially have depressive symptoms that as yet do not reach psychotic proportions. The

detection of these depressive symptoms may differ depending on the selected depression scale.

Conclusions

In this study, we examined the consistency of depression determination of four self-rating depression scales applied to 563 Japanese community-dwelling elderly living in twelve prefectures. Although depression scale scores closely correlated among the four scales, the agreement in depressive determinations was differed among the scales. A similarity of the depressive determination tendency was found between GDS and CRS, and between CES-D and SDS. Depression rates determined on the basis of the cut-off point for each scale were higher for CES-D and SDS than for GDS and CRS. Depression determination by a four-point rating scale may overestimate a slightly depressive symptom, compared with a two-point scale. From our results, the Japanese community-dwelling elderly potentially have slightly depressive symptoms.

References

(1) Demura S, Sato S. Functional assessment for the elderly by using activities of daily living (ADL). Japan J Phys Educ, Hlth Sports Sci. 2004;49:519–533. (Article in Japanese)

(2) Demura S, Minami M, Nagasawa Y, Tada N, Matsuzawa J, Sato S. Physical-Fitness declines in older Japanese adults. J Aging Phys Act. 2003;11:112–122.

(3) Demura S, Sato S. Quality of Life (QOL) Assessment for Japanese Elderly: the course of QOL studies and assessments of health related and subjective QOL. Japan J Phys Educ, Hlth Sports Sci. 2006;51. (in press) (Article in Japanese)

(4) Masuchi A, Kishi R. A review of epidemiological studies on the relationship of social networks and support to depressive symptoms in the elderly. Nippon Koshu Eisei Zasshi. 2001;48:435–448. (Article in Japanese)

(5) Hashimoto K, Kurita H, Haratani T, Fujii K, Ishibashi T. Direct and buffering effects of social support on depressive symptoms of the elderly with home help. Psychiatry Clin Neurosci. 1999;53:95–100.

(6) Osada H, Shibata H, Haga H, Yasumura S. Relationship of physical condition and functional capacity to depressive status in person aged 75 years. Nippon Koshu Eisei Zasshi. 1995;42:897–909. (Article in Japanese)

(7) Sato S, Demura S, Matsuzawa J, Kobayashi H, Toyoshima Y. Examination of ADL achievement ability in partially dependent older people: from the viewpoints of aging, perceptual level of health and physical fitness, and susceptibility to disease. Japan J Phys Educ, Hlth Sports Sci. 1999;44:13–24.

(8) Sato S, Demura S, Kobayashi H, Nagasawa Y. The relationship and its change with aging between ADL and daily life satisfaction characteristics in independent Japanese elderly living at home. J Physiol Anthropol Appl Human Sci. 2002; 21:195–204.

(9) Demura S, Sato S. Relationships between depression, lifestyle and quality of life in community dwelling elderly: comparison between gender and age group. J Physiol Anthropol Appl Human Sci. 2003;22:159–166.

(10) Demura S, Minami M, Noda M, Ishikawa Y, Noda Y. Gender and relation of life-style to morale in older people living in regional cities. Nippon Eiseigaku Zasshi. 2002;56:655–663. (Article in Japanese)

(11) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV (4th ed). Washington DC: American Psychiatric Press Inc.; 1994.

(12) Yesavage JA, Blink TL. Geriatric depression screening scale; a preliminary report. J Psychiatry Res. 1983;17:37–49.

(13) Zung WW. A self-rating depression scale. Arch Gen Psychiatry. 1965;12:63–70.

(14) Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Measurement. 1977;1:385–401.

(15) Carroll BJ, Feinberg M, Smouse PE, Rawson SG, Greden JF. The Carroll rating scale for depression: I. Development, reliability and validation. Br J Psychiatry. 1981;138:194–200.

(16) Shima S, Shikano T, Kitamura T, Asai M. A new self-report depression scale. Psychiatry. 1985;27:717–723. (Article in Japanese)

(17) Shima S. CES-D scale. Chiba Test Center: Tokyo; 1977. p. 1–12. (Article in Japanese)

(18) Niino N, Imaizumi T, Kawakami N. A Japanese translation of the Geriatric Depression Scale. Clin Gerontol. 1991;10: 85–87.

(19) Yatomi N. The factor structure and item characteristics of the GDS (Geriatric Depression Scale) short version in a Japanese elderly sample. Ronen Shakai Kagaku. 1994;16:37–45. (Article in Japanese)

(20) Yatomi N, Lian J, Krause N, Akiyama H. Measurement of depression symptoms of the Japanese elderly by the CES-D scale: examination of factor structure and cultural differences. Jpn J Soci Gerontol. 1996;37:37–47. (Article in Japanese)

(21) Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56–62.

(22) O'Hara MW, Kohout FJ, Wallance RB. Depression among the rural elderly: a study of prevalence and correlates. J Nerv Ment Dis. 1985;173:582–589.

(23) Aoki K. Depressive states and their correlates in elderly people. Ronen Seishin Igaku Zasshi. 1997;8:401–410. (Article in Japanese)

(24) Yamashita K, Kobayashi S, Tsunematsu T. Depressed mood and subjective sensation well-being in the elderly living alone on Oki island in Shimane prefecture. Nippon Ronen Igakkai Zasshi. 1992;29:179–184. (Article in Japanese)

(25) Ihara K. Depressive states and their correlates in elderly people living in a rural community. Nippon Koshu Eisei Zasshi. 1993;40:85–94. (Article in Japanese)

(26) Iwata N, Buka S. Race/ethnicity and depressive symptoms: a cross-cultural/ethnic comparison among university students in East Asia, North and South America. Soc Sci Med. 2002;55:2243–2252.

(27) Blink TL, Yesavage J, Lum O, Heersema P, Adey MB, Rose TL. Screening test for geriatric depression. Clinical Gerontol. 1982;1:37–44.

(28) Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. In: Blink TL editor. Clinical Gerontology: A Guide to Assessment and Intervention. NY: The Haworth Press; 1986. p. 165–173.