Geriatrics Gerontology International

DOI: 10.1111/ggi.13552

ORIGINAL ARTICLE

EPIDEMIOLOGY, CLINICAL PRACTICE AND HEALTH

Development of a novel convenient Alzheimer's disease assessment scale, the ABC Dementia Scale, using item response theory

Yumi Umeda-Kameyama,[1] Takahiro Mori,[2] Kenji Wada-Isoe,[3] Takashi Kikuchi,[4] Shinsuke Kojima,[4] Tatsuo Kagimura,[4] Akinori Ueki,[5] Toshiyuki Watabe,[6] Chiaki Kudoh,[7] Masahiro Akishita[1] and Yu Nakamura[2] and ABC dementia scale study group

[1]Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[2]Department of Neuropsychiatry, Faculty of Medicine, Kagawa University, Kagawa, Japan
[3]Division of Neurology, Department of Brain and Neurosciences, Faculty of Medicine, Tottori University, Tottori, Japan
[4]Translational Research Informatics Center, Foundation for Biomedical Research and Innovation, Hyogo, Japan
[5]Ueki Dementia and Geriatric Psychiatry Clinic, Hyogo, Japan
[6]Watabe Clinic, Yamagata, Japan
[7]KUDOH CHIAKI Clinic for Neurosurgery & Neurology, Tokyo, Japan

Correspondence
Yumi Umeda-Kameyama MD PhD, Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8665, Japan.
Email: yumeda-tky@umin.ac.jp

Received: 22 August 2018
Accepted: 25 September 2018

Aim: The present study aimed to assess the interrater reliability and construct the validity of a novel, convenient informant-based Alzheimer's disease assessment scale to prepare its final version.

Methods: For the assessment, site investigators, co-medicals and, if available, medical staff other than doctors or co-medicals interviewed study informants to assess individuals using this scale. We then analyzed the interrater reliability and construct validity using factor analysis and item response characteristics.

Results: In this study, 427 eligible participants were enrolled. We first examined the interrater reliability, and found that the lower limit of the confidence interval of each item was never <0.4 (except for the item "delusion of theft"). After deleting this item, the 14 items of this scale were organized into three domains (activities of daily living, behavioral and psychological symptoms of dementia, and cognitive function) through factor analysis. After discussion of the similarity of two items and their integration into one item, we confirmed that the final version of the 13-item scale showed almost the same degree of interrater reliability and construct validity as the former version of this scale.

Conclusions: The final version of this novel Alzheimer's disease assessment scale had high interrater reliability and construct validity. We named it the ABC (activities of daily living, behavioral and psychological symptoms of dementia, and cognitive function) Dementia Scale. Further studies on its validation are required. **Geriatr Gerontol Int 2019; 19: 18–23.**

Keywords: construct validity, dementia assessment scale, factor analysis, interrater reliability.

Introduction

A recent survey showed that there are >4 million older individuals with dementia in Japan, which accounts for approximately 15% of the total population, and close to 4 million older persons have mild cognitive impairment.[1] To assess the severity or treatment effect of dementia, the Mini-Mental State Examination scale[2] or revised version of Hasegawa's Dementia Scale[3,4] are widely used in Japan. However, these scales are limited to the assessment of cognitive function. Although other scales for assessing behavioral and psychological symptoms of dementia (BPSD) and activities of daily living (ADL) are important, special training for examiners is required to use certain instruments for assessing these symptoms (e.g. Neuropsychiatric Inventory[5] or Disability Assessment for Dementia)[6] and there is no scale that can conveniently and comprehensively assess all symptoms.[7]

In Germany, a comprehensive evaluation scale was developed, named the Relevant Outcome Scale for Alzheimer's Disease (ROSA); it is a 16-item, 21-point scale that can evaluate cognitive function, ADL, BPSD, communication skills and quality of life. It requires approximately 15 min for completion.[8,9] However, the ROSA requires an additional baseline assessment of disease severity and is not suitable for clinical practice for aged care.

Therefore, we attempted to develop a novel and convenient AD assessment scale with illustrations. The present study aimed

to assess the interrater reliability and construct validity of this scale to prepare its final version.

Methods

Goal and requirements of this scale

We first defined the ultimate goals of this scale, including high reliability, validity and sensitivity, as below:

To assess comprehensive AD function (i.e. not only cognitive function, but also BPSD and ADL).
To assess regardless of disease severity.
To assess conveniently in a short time.
To be used not only by specialist physicians, but also by non-specialists or healthcare workers.
To assess without requiring specialized training in its use.

To achieve these goals, we designated the following requirements for the scale.

1. The scale must have <15 items. The score of each item is indicated on a 9-point rating scale (from 1 to 9 [1 = severest to 9 = mildest]). Furthermore, an illustration should be attached to the anchor point of each item (score 1, 3, 5, 7, 9), representing item severity.
2. We must carry out a quantitative assessment of reliability, validity and sensitivity using statistical methods, such as factor analysis or item response characteristics, which are usually employed for the validation of educational or psychological tests.

We planned to develop this scale using the following three steps.

Step 1: Establishment of the first draft of this scale (scale draft 1).
Step 2: Establishment of the final draft.
Step 3: Verification of reliability, validity and sensitivity of the final draft.

Study participants

Participants were outpatients who had been examined at 41 clinics and hospitals across Japan. They were those clinically diagnosed with AD based on the diagnostic criteria of the *Diagnostic and Statistical Manual of Mental Disorders* Fourth Edition Text Revision,[10] National Institute on Aging-Alzheimer's Association workgroup,[11] or National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association.[12] As non-randomized controls, patients with cognitive concerns without dementia (mild cognitive impairment equivalent) were also enrolled. All participants and/or their legal representatives, as well as all the study informants, gave written informed consent.

Individuals were excluded if they had other types of dementia (e.g. vascular dementia, Lewy body dementia, frontotemporal dementia). Specialists in neuropsychiatry, neurology, geriatrics and dementia judged the severity of individuals' AD based on a clinical decision. Unique identifier: NCT02267486 (https://clinicaltrials.gov/ct2/show/NCT02267486).

Establishment of the prototype of the ADL, BPSD and cognitive function Dementia Scale

The ADL, BPSD and cognitive function (ABC) Dementia Scale (ABC-DS; The Japanese version of the ABC-DS can be obtained from https://ctportal.tri-kobe.org/studies/ququ/scale.html; a PDF version of the ABC-DS in English, French, Chinese, or Korean can be downloaded from the Mapi Research Trust at: http://mapi-trust.org/our-resources/questionnaires-distributed-by-the-mapi-research-trust/) is similar to the ROSA in that it comprehensively assesses cognitive function, ADL and BPSD; hence, the items were newly created with reference to the end-points of the ROSA. However, as several items of the ROSA were not suitable for the Japanese population, they were not incorporated in this scale. Finally, after adding several novel items to the ROSA, we determined the draft 1 scale, which comprised 17 items. We used the design of a behavioral observation scale through which the rater interviewed study informants (e.g. family caregiver) and assessed participants.

In the preceding study where a total of 543 participants were assessed using the draft 1 scale (Tables 1, S1), we identified a ceiling effect and inappropriate characteristics of the item characteristic curves for two items ("Delusions about being in the wrong home" and "Circadian rhythm"). After discussion of these results by the committee of specialists, we deleted these two items from the draft 1 scale, and the wording of the anchor points was revised for all items to produce the draft 2 scale, which comprised 15 items (Table 1).

Assessment procedure and raters

In the present study, eligible participants were assessed using the draft 2 scale. For the assessment, site investigators, co-medicals and, if available, medical staff other than doctors or co-medicals interviewed study informants to assess participants.

Statistical analysis

We analyzed the scale's interrater reliability and construct validity (using factor analysis and item response characteristics), as described below. Interrater reliability analyses were carried out using SAS version 9.3 (SAS Institute, Cary, NC, USA). For factor analysis and item response characteristics, R version 3.1.0 (10 April 2014; available as a free download from http://www.r-project.org) was used.

Interrater reliability
The extent of similarity between the "assessment by physicians and co-medicals (nurse, public health nurses, clinical psychologists and care workers)" and "assessment by individuals other than physicians and co-medicals (other than the aforementioned qualified personnel)" for the scores of each item in this scale were assessed using the weighted κ coefficient and its 95% confidence interval. We focused on whether the lower limit of the 95% confidence interval of the weighted κ coefficient exceeded 0.4. The classification of level of similarity was expressed using the Altman classification.[13]

Construct validity: Factor analysis
Using an oblique rotation (promax rotation) to recognize the correlations between domains, three or four domains were assessed to determine the number of domains necessary to explain the data appropriately. The factor loadings and cumulative contribution ratio of each model were listed. Furthermore, we interpreted the medical meaning of each domain and named it.

In this situation, although an eigenvalue of ≥1 and a cumulative contribution ratio of ≥70% was initially used as the reference standard, the overall domain number and constituent items were determined by taking into consideration the extent of the increase in the cumulative contribution ratio, the number of items forming each domain and the medical interpretation of that domain.

Construct validity: Item response characteristics
For the assessment of item response characteristics, we assessed item characteristic curves, item information curves and test characteristic curves. The item characteristic curve of each item was graphed using the graded response model, calculated discrimination parameter, parameters of difficulty and amount of item

Table 1 Domain and item details of scale draft 1 (17 items), draft 2 (15 items) and draft 3 (13 items)

Draft 1	Draft 2	Draft 3	Domain	Contents of item
q1	Q6	Q6	C	Recent memory (forgotten objects) How is the patient when he or she forgets the location of a familiar item? (e.g., patient ID card, hearing aids, insurance card, wallet, shoes, hat, glasses, jacket, house key)
q2	Q7	Q7	C	Recent memory (events) How is the patient when he or she remembers day-to-day events taking place around him or her? (e.g. hospital visits, visitors, shopping, outings, telephone calls)
q3	Q4	Q4	A	Execution (remote control operation) How is the patient when using a remote control to watch television?
q4	Q5[†]	—	A	Execution (phone operation) How is the patient when using a phone?
q5	Q3	Q3	A	Communication How is the patient when he or she wants to communicate something?
q6	Q9	Q9	B	Aggression How is the patient when something goes against his/her wishes?
q7	Q8	Q8	B	Restlessness How is the patient when required to sit quietly? (e.g. on the bus, on the train, during medical examinations)
q8	Q10[‡]	—	B	Delusion of theft How is the patient when he or she has a delusion that his/her money or things have been stolen?
q9	—	—	B	Delusions about being in the wrong home How is the patient when he or she has a delusion that he or she is not in his/her own home even when he or she is?
q10	Q11	Q11	B	Cooperativeness How is the patient when asked to do something? (e.g. to take a bath/shower, take medicine, getting dressed/undressed, go out)
q11	Q1	Q1	A	Daily activities How is the patient when changing clothes?
q12	Q2	Q2	A	Motivation How is the patient when he or she is preparing him/herself? (e.g. dressing and undressing, brushing teeth, shaving, make-up, combing hair)
q13	—	—	A	Circadian rhythm How is the patient when he or she sleeps?
q14	Q13	Q13	A	Meal conditions How is the patient when he or she has a meal?
q15	Q14	Q14	A	Incontinence How is the patient when he or she has a desire to urinate?
q16	Q12	Q12	C	Taking medicines How is the patient when he or she needs to hold medication in an appropriate place and take it?
q17	Q15	Q15	C	Caregiving burden How is the patient when you supervise him/her?

[†]The expert committee discussed the similarity of item Q4 "Complicated activity (remote control operation)" and Q5 "Complicated activity (phone (c) 2014 ABC Dementia Scale Study Group Steering Committee and Working Group operation)" because of the recent development of cellular phones. Therefore, the expert committee integrated Q5 and Q4 in the modified draft 2 to create Draft 3. [‡]A ceiling effect and problems with the interrater reliability were identified in item Q10. Therefore, the expert committee deleted Q10 from draft 2 to create the modified draft 2. A, activity of daily living domain; B, behavioral and psychological symptoms of dementia-related domain; C, cognitive function-related domain.

information. After confirming that the discrimination parameter was ≥0.2 and that the absolute value of difficulty was <5 (within the 99% + confidence interval of capability), the appropriateness of the characteristic of each item was considered.

However, due to the limitation of the statistical software, we first integrated the original 9-point scale into a five-level scale for analysis, as follows: "scores 1 and 2," "scores 3 and 4," "score 5," "scores 6 and 7," and "scores 8 and 9" were summarized into levels 1, 2, 3, 4 and 5, respectively.

In addition, it was confirmed that there was no contradiction in the order of the item characteristic curves (probability distribution curves), and that there was a relatively sufficient amount of item information.

For each participant, a test characteristic curve that plotted the raw data point total for each domain on the vertical axis, and the capability of dementia levels (model-based estimates) θ on the horizontal axis was drawn. Furthermore, a smooth regression curve was shown to indicate the change in the average scores.

Ethics

The study was carried out in accordance with the Declaration of Helsinki, following approval from the ethics committee in each institution. This research was also carried out in line with the Japanese Ministry of Health, Labor and Welfare's "Ethical Guidelines for Medical Research Involving Human Subjects."

Results

Assessments of the draft 2 scale (15 items)

Participant characteristics and average score of each item

In the present study, 427 eligible participants (88 with severe AD, 133 with moderate AD, 129 with mild AD and 77 with mild cognitive impairment equivalent; Table S2) were registered. The number of male and female participants were 164 (38.4%) and 263 (61.6%), respectively. The average age was 80.8 years (males 79.7 years and females 81.4 years).

Table 2 Maximum, minimum and mean score of the 15 items of the draft 2 scale

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
◆ Severe AD	Max.	9	9	9	9	9	7	6	9	9	9	9	9	9	9	9
	Min.	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
	Mean	3.48	4.39	4.56	2.74	3.24	2.06	1.58	7.19	7.05	8.57	6.75	1.7	6.65	4.5	3.11
◆ Moderate AD	Max.	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
	Min.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Mean	6.64	6.15	6.61	5.86	5.72	3.81	2.74	8.35	7.38	8.35	7.38	3.75	8.19	7.41	5.59
◆ Mild AD	Max.	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
	Min.	1	3	3	3	3	1	1	1	3	3	1	1	5	5	3
	Mean	8.09	7.71	7.66	7.14	7.16	5.25	4.27	8.64	7.68	8.55	7.81	5.58	8.58	8.38	7.00
◆ MCI equivalent	Max.	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
	Min.	3	3	7	5	5	1	1	5	3	5	3	1	5	5	1
	Mean	8.65	8.34	8.45	7.95	8.19	6.16	6.48	8.68	8.01	8.84	8.57	7.06	8.74	8.64	8.18

AD, Alzheimer's disease; Max., maximum; MCI, mild cognitive impairment; Min., minimum.

Table 3 Interrater reliability of draft 2 scale (15 items)

	Physician *vs* co-medical (n = 427)			Physician *vs* non-co-medical (n = 84)		
	κ coefficient	Upper limit of 95% CI	Lower limit of 95% CI	κ coefficient	Upper limit of 95% CI	Lower limit of 95% CI
Q1: Daily activities	0.83	0.87	0.79	0.85	0.91	0.79
Q2: Motivation	0.69	0.74	0.64	0.79	0.87	0.71
Q3: Communication	0.71	0.75	0.66	0.67	0.78	0.57
Q4: Execution (remote control operation)	0.78	0.82	0.75	0.79	0.87	0.70
Q5: Execution (phone operation)	0.83	0.86	0.79	0.88	0.94	0.83
Q6: Recent memory (forgotten objects)	0.73	0.77	0.69	0.69	0.78	0.60
Q7: Recent memory (events)	0.77	0.81	0.74	0.72	0.80	0.64
Q8: Restlessness	0.74	0.81	0.67	0.71	0.88	0.53
Q9: Aggression	0.73	0.78	0.68	0.76	0.84	0.67
Q10: Delusion of theft	0.68	0.78	0.57	0.63	0.95	0.31
Q11: Cooperativeness	0.69	0.75	0.64	0.74	0.84	0.64
Q12: Taking medicines	0.78	0.82	0.74	0.72	0.82	0.61
Q13: Meal conditions	0.75	0.81	0.69	0.80	0.90	0.71
Q14: Incontinence	0.80	0.84	0.76	0.84	0.91	0.77
Q15: Caregiving burden	0.79	0.83	0.75	0.76	0.84	0.68
	Average weighed κ coefficient = 0.75			Average weighed κ coefficient = 0.76		

Value of κ <0.20 ■, Poor; 0.21–0.40 ■, Fair; 0.41–0.60 ■, Moderate; 0.61–0.80 ■, Good; 0.81–1.00 ■, Very good;

When examining the average, maximum and minimum score of each item, the average value was >8 points at all severity levels of dementia, and a ceiling effect was identified in the item "Delusion of theft(Q10)" even after revising the wording of the anchor point (Table 2).

Interrater reliability

On the same day, the physician, co-medicals and, if available, staff other than co-medicals interviewed the same study informant (most of whom were family caregivers). We examined the interrater reliability of ratings of "physician *vs* co-medical" and "physician *vs* non-co-medical" by using the weighted κ coefficient and 95% confidence intervals.

The lower limits of the 95% confidence interval for the weighted κ coefficient of "Q10: "Delusion of theft" were 0.57 and 0.31 for "physician *vs* co-medical" and "physician *vs* non-co-medical," respectively, implying that there was not a sufficient interrater reliability (Table 3). In other items, the lower limit of the confidence interval was not <0.4, fulfilling the standard defined earlier.

As mentioned above, a ceiling effect and problems in the interrater reliability were identified in item Q10. Therefore, for the analysis of the construct validity, we used the draft 2 scale from which item Q10 was deleted (so that it contained just 14 items).

Construct validity of the modified draft 2 scale (14 items)

Factor analysis

As a result of a factor analysis, the 14 items fell into three domains: seven items fell into domain 1, four items into domain 2 and three items into domain 3. Based on their content, domains 1, 2 and 3 were named the ADL-related, cognitive function-related and BPSD-related domains, respectively (Table 4).

The average factor loadings (corresponding to the correlation coefficient) for ADL-related, cognitive function-related and BPSD-related domains were 0.754, 0.637 and 0.573, respectively. This showed that the items in each domain appropriately reflected the construct of the specific domain.

In contrast, in the four-domain model, none of the items loaded onto domain 4 substantially (Table 4). Therefore, we decided this model was redundant. Furthermore, the cumulative contribution ratio of the three- and four-domain models were 0.540 and 0.550, respectively, showing that the addition of one more domain did not lead to an improvement in the explained variance.

Item response characteristics

First, we obtained an overview of the difficulty parameters of the 14 items. The difficulty parameters must be included within the

Table 4 Factor analysis of modified draft 2 scale (14 items)

	Three-domain model				Four-domain model				
	Factor loadings			Uniqueness factor	Factor loadings				Uniqueness factor
	Domain 1	Domain 2	Domain 3		Domain 1	Domain 2	Domain 3	Domain 4	
Q1	0.884	*	*	0.214	0.896	*	*	*	0.209
Q2	0.589	0.194	*	0.405	0.707	0.174	*	−0.287	0.298
Q3	0.678	0.105	*	0.365	0.669	0.109	*	*	0.367
Q4	0.756	0.135	*	0.295	0.684	0.133	*	0.256	0.236
Q5	0.665	0.220	*	0.332	0.618	0.226	*	0.134	0.320
Q6	0.255	0.566	*	0.381	0.230	0.566	*	*	0.384
Q7	−0.119	0.878	*	0.381	−0.109	0.885	*	*	0.369
Q8	0.357	−0.109	0.348	0.722	0.291	−0.119	0.365	0.199	0.681
Q9	−0.102	*	0.849	0.411	−0.110	*	0.821	*	0.444
Q11	*	0.200	0.522	0.631	*	0.185	0.546	−0.112	0.604
Q12	0.220	0.659	*	0.363	0.216	0.654	*	*	0.366
Q13	0.812	−0.136	*	0.508	0.818	−0.133	*	*	0.509
Q14	0.897	*	*	0.322	0.897	*	*	*	0.324
Q15	0.395	0.446	*	0.343	0.333	0.459	*	0.144	0.328
Cumulative contribution ratio		0.540				0.550			

*0.000. Uniqueness factor = 1 − commonality. Commonality = $\sum$ (factor loading)^2.

Table 5 Construct validity of the item response characteristics of the 14 items of the modified draft 2 scale: difficulty and discrimination parameters

	Difficulty parameter 1	Difficulty parameter 2	Difficulty parameter 3	Difficulty parameter 4	Difficulty parameter 5	Discrimination parameter
Q1	−1.45	−1.17	−0.68	−0.13	0.23	4.04
Q2	−2.03	−1.57	−0.74	0.01	0.37	2.31
Q3	−2.52	−1.91	−0.92	−0.07	0.42	2.63
Q4	−1.23	−0.89	−0.32	0.32	0.72	3.04
Q5	−1.58	−1.17	−0.46	0.34	0.81	2.86
Q13	−3.76	−3.06	−1.95	−1.10	−0.67	2.00
Q14	−1.88	−1.61	−1.03	−0.41	−0.09	2.70
Q6	−1.01	−0.56	0.38	1.41	1.97	2.50
Q7	−0.30	0.09	0.74	1.41	1.82	2.06
Q12	−0.75	−0.29	0.29	0.77	1.14	2.80
Q15	−1.57	−1.23	−0.47	0.33	0.71	2.88
Q8	−4.25	−3.70	−2.78	−1.94	−1.45	1.14
Q9	−2.86	−2.48	−1.69	−0.66	−0.02	2.75
Q11	−2.96	−2.50	−1.76	−1.02	−0.56	1.40

99% confidence interval of the capability level parameter, in other words, between −4 and 4 of the Z-score. Although the difficulty level 1 parameter (level 1: choice rate distribution parameter for 1 point and 2 points on the 9-point scale) of "Q8: Restlessness" was −4.25, we decided that it is not a problem, because it is not a large deviation from the standard −4 (Table 5).

We then took an overview of the discrimination parameter. The discrimination parameter of Q1 "Daily activity" was the highest at 4.04, suggesting that this item most readily reflects changes in the level of dementia based on the response. In contrast, the discrimination parameters of Q8 "Restlessness" and Q11 "Cooperativeness" were low, at 1.14 and 1.40, respectively, suggesting that these items do not easily reflect changes in the dementia level.

Revision of the draft 2 scale

As mentioned above, a ceiling effect and problems in the interrater reliability were identified in item Q10. In addition, the expert committee discussed the similarity of items Q4 "Complicated activity (remote control operation)" and Q5 "Complicated activity (phone operation)" because of the technical advances in cellular phones in recent years. Therefore, we re-analyzed and confirmed the draft 3 scale (13 items) after deleting Q5 and Q10 from the modified draft 2; it showed almost the same level of performance as the former version of this scale. The draft 3 scale was determined as the final version of this scale (Tables S3, S4).

Discussion

For several items, the item response characteristics were not appropriate, forcing us to revise the anchor points or delete the items. In addition, as the BPSD-related domain becomes "inactive" when patients suffer severe AD, there is a possibility that the score in this domain will increase. Therefore, we carefully revised the anchor points of this domain. After the revision, when the difficulty parameters matched the confidence interval of the capability value and the selection of the items was adequate, we considered it reasonable, even if the item response characteristics were not good. Therefore, in the revision of the draft 2 scale, we recommended the selection of three items from the BPSD-related domain from among those with good item response characteristics.

For domain 1, the "ADL-related domain," the use of the Disability Assessment for Dementia[6] (an ADL evaluation scale) seems to be suitable for assessing concurrent validity, whereas the Mini-Mental State Examination[2] and Neuropsychiatric

Inventory Caregiver Distress[5] appear suitable to assess the cognitive function-related and BPSD-related domains, respectively. The Clinical Dementia Rating is widely used as a global assessment of dementia.[14] However, to effectively use the Clinical Dementia Rating and these other three scales, assessors require training in psychological examination. Furthermore, all these scales require approximately an hour for administering. The time taken for this scale in this survey was approximately 10 min. Therefore, it might serve as a useful new comprehensive AD assessment scale that can be administered by anyone over a short period. In addition, unlike the ROSA, a similar tool, the ABC-DS, does not require the assessor to judge the severity of a patient's AD before assessment.

The final version of this scale (Figure S1) —which had 13 items, 5 simple anchor points, and a 9-point scale—was found to have high inter-rater reliability and construct validity. However, validation on a separate cohort is required since the construct validity of the scale is only optimized for a cohort of about 500 subjects (as in this study). Moreover, with respect to the subject group discriminability, it was only possible to discriminate the severity of a patient diagnosed with reference to the Clinical Dementia Rating. Therefore, we are now conducting a validation study of the final version of ABC-DS to evaluate the intra-rater reliability, subject group discriminability, concurrent validity against external reference scales, reactivity, and construct validity (unique identifier: NCT02667665). Finally, the responsiveness of the ABC-DS to drug treatment should also be examined.

Acknowledgements

The authors thank Mr Tsukada for the preparation of the datasets and carrying out a part of analyses, Mr Yamauchi for data management, and Mr Sakamine for many contributions as a project manager. The authors also appreciate the careful review of this manuscript by Hideaki Kaneda, MD, PhD; Mikio Yoshitomi, PhD; Atsuhiko Kawamoto, MD, PhD; and Masanori Fukushima, MD, PhD, director of the Translational Research Informatics (TRI) Center, Foundation for Biomedical Research and Innovation (FBRI), Kobe, Japan.

Disclosure statement

Daiichi Sankyo provided a research fund to this study. However, they did not have any role in the design of the study, data collection, statistical analyses and writing of the article.

References

1 Ministry of Health, Labour and Welfare. *National Dementia Strategy (New Orange Plan)*. Tokyo, Japan. [Cited 14 Jan 2016]. Available from URL: http://www.mhlw.go.jp/stf/houdou/0000072246.html. (In Japanese).
2 Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12**: 189–198.
3 Kato S, Shimogaki H, Onodera A, Ueda H, Oikawa K. Development of the revised version of Hasegawa's dementia scale (HDS-R). *Jpn J Geriatr Psychiatry* 1991; **2**: 1339–1347. (In Japanese).
4 Kim KW, Lee DY, Jhoo JH *et al*. Diagnostic accuracy of mini-mental status examination and revised Hasegawa Dementia Scale for Alzheimer's Disease. *Dement Geriatr Cogn Disord* 2005; **19**: 324–330.
5 Fuh JL, Lam L, Hirono N, Senanarong V, Cummings JL. Neuropsychiatric inventory workshop: behavioral and psychologic symptoms of dementia in Asia. *Alzheimer Dis Assoc Disord* 2006; **20**: 314–317.
6 Gélinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. *Am J Occup Ther* 1999; **53**: 471–481.
7 Robert P, Ferris S, Gauthier S, Ihl R, Winblad B, Tennigkeit F. Review of Alzheimer's disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice? *Alzheimers Res Ther* 2010; **2**: 24.
8 Holthoff VA, Ferris S, Ihl R *et al*. Validation of the relevant outcome scale for Alzheimer's disease: a novel multidomain assessment for daily medical practice. *Alzheimers Res Ther* 2011; **3**: 27.
9 Holthoff V, Ferris S, Gauthier S *et al*. Memantine effects measured with the Relevant Outcome Scale for Alzheimer's disease in an open-label, single-arm, multicenter clinical study. *Int J Geriatr Psychiatry* 2013; **28**: 164–172.
10 *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, Text Revision. Washington, DC: American Psychiatric Association, American Psychiatric Association, 2000.
11 McKhann GM, Knopman DS, Chertkow H *et al*. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7**: 263–269.
12 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984; **34**: 939–944.
13 Altman DG. *Practical Statistics for Medical Research*. London: Chapman & Hall/CRC, 1999.
14 Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry* 1982; **140**: 566–572.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website:

Table S1 Domain and item details for the 17 items of the ABC Dementia Scale draft 1.
Table S2 Characteristics of participants (draft 2 scale).
Table S3 Factor analysis of the modified draft 3 scale (13 items; factor loadings, uniqueness factor, and cumulative contribution ratio).
Table S4 Construct validity: Item response characteristics of the 13 items of the draft 3 scale (difficulty and discrimination parameters).
Figure S1 A sample of the activities of daily living, behavioral and psychological symptoms of dementia, and cognitive function (ABC) Dementia Scale (final version) to evaluate behavioral and psychological symptoms of dementia.

How to cite this article: Umeda-Kameyama Y, Mori T, Wada-Isoe K, et al. Development of a novel convenient Alzheimer's disease assessment scale, the ABC Dementia Scale, using item response theory. Geriatr. Gerontol. Int. 2019;19:18–23. https://doi.org/10.1111/ggi.13552