

## A Novel Dementia Scale for Alzheimer's Disease

Takashi Kikuchi<sup>1\*</sup>, Takahiro Mori<sup>2</sup>, Kenji Wada-Isoe<sup>3</sup>, Yumi Umeda-Kameyama<sup>4</sup>, Tatsuo Kagimura<sup>1</sup>, Shinsuke Kojima<sup>1</sup>, Masahiro Akishita<sup>4</sup> and Yu Nakamura<sup>2</sup>

<sup>1</sup>Translational Research Informatics Center, Foundation for Biomedical Research and Innovation, 1-5-4 Minatojima-Minamimachi, Chuo-ku Kobe, Hyogo 650-0047, Japan

<sup>2</sup>Department of Neuropsychiatry, Faculty of Medicine, Kagawa University, 1750-1, Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0793, Japan

<sup>3</sup>Division of Neurology, Department of Brain and Neurosciences, Faculty of Medicine, Tottori University, 36-1, Nishi-cho, Yonago, Tottori 683-8504, Japan

<sup>4</sup>Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8665, Japan

### Abstract

**Objective:** We established the diagnostic accuracy of the "ABC Dementia Scale" (ABC-DS) for Alzheimer's disease (AD), which concurrently assesses activities of daily living ("A"), behavioral and psychological symptoms of dementia ("B"), and cognitive function ("C"), using a novel scoring approach called the three-dimensional distance (TDD).

**Methods:** The ABC-DS has 13 items with nine ordered categorical levels. Caregivers were interviewed using a semi-structured interview. The construct validity, concurrent validity, test-retest reliability, and responsiveness (score changes over 12 weeks) were assessed.

**Results:** We enrolled 63 participants with probable AD as well as 88, 106, and 55 patients with mild, moderate, and severe AD, respectively. The construct and concurrent validities of each domain score were determined. The TDD accurately discriminated the AD stages and detected score changes indicating disease progression over 12 weeks.

**Conclusion:** The ABC-DS is stable, accurately stages AD severity, and monitors disease progression. The TDD is a useful algorithm for detecting disease progression.

**Keywords:** Activities of daily living; Alzheimer's disease; Cognitive function; Factor analysis; Item response theory; Receiver operating characteristic; Semi-structured interview; Three-dimensional distance; Multi-domain scale

### Introduction

Robert et al. [1] stated that an ideal scale for assessing Alzheimer's disease (AD) must be quickly administered, validated in the context of AD, include multiple AD characteristics, be applicable to all AD severity stages, monitor disease progression, and be sensitive to therapeutic responses. However, the authors concluded that no currently available scale satisfies all of these criteria.

The Relevant Outcome Scale for Alzheimer's disease (ROSA) was designed to assess cognitive function, activities of daily living (ADL), behavioral and psychological symptoms of dementia (BPSD), communication skills, and quality of life with 16 items and 21 levels [2]. Unfortunately, a critical limitation of the scale is that the total scores must be compared directly within a patient or between patient groups with the same disease severity; that is, if a patient's disease severity changes over time, the data must be excluded from statistical analyses. Although a multicenter clinical study statistically validated the ROSA [3], since then, only one report on this scale has been published [4]. Moreover, the clinical relevance of the ROSA and the degree to which it is used in daily practice remain unclear.

In clinical trials, therapeutic efficacy is often evaluated by comparing the total scores of a multi-domain scale, but total scores may not monitor disease progression or severity with sufficient accuracy. Indeed, various changes in ADL, BPSD, and cognitive function occur across the clinical stages of AD (Figure 1); specifically, ADL and cognitive function levels at stage S2 (severe stage) are lower than those at stage S1 (mild stage), while the activity of BPSD is the same at S1 and S2. This is caused by a bell-shape activity curve of BPSD, with levels first increasing and then decreasing as AD progresses [5]. This change in BPSD may cause inaccuracies in total score-based assessments. Additionally, as a

mathematical rule, domain scores should not be summed if they have different qualities. Therefore, we need a new mathematical algorithm that can generate a new type of total score from otherwise incompatible domain scores.

To address this issue, we hypothesized that by plotting a patient's ADL, BPSD, and cognitive function scores in a three-dimensional coordinate plane (Figure 2), we would be able to calculate the mathematical distance of the scores (ADL score, BPSD score, cognitive score) from the origin (0, 0, 0) and use this distance as a type of total score. This distance, referred to as the three-dimensional distance (TDD) score, is calculated as follows:

$$TDD = \sqrt{(ADL \text{ score})^2 + (BPSD \text{ score})^2 + (Cognitive \text{ function score})^2}$$

where ADL, BPSD, and cognitive function scores increase with symptom improvement and decrease towards 0 with symptom deterioration.

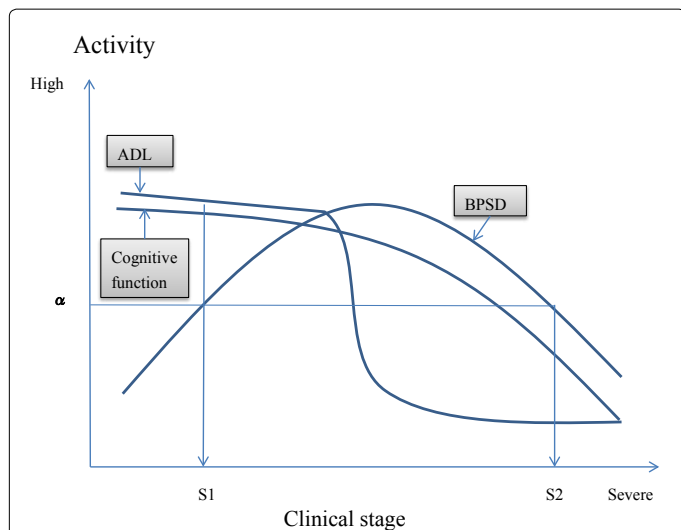
In 2014, we began developing a multi-dimensional scale for AD that initially consisted of 17 items regarding ADL ("A"), BPSD ("B"), and cognitive function ("C"). Since then, we revised the descriptions and deleted items based on a factor analysis and other analyses using Item Response Theory (IRT) [6]. We finally produced a "test version" of the

**\*Corresponding author:** Takashi Kikuchi, Translational Research Informatics Center, Foundation for Biomedical Research and Innovation, 1-5-4 Minatojima-minamimachi, Chuo-ku Kobe, Hyogo 650-0047, Japan, Tel: +81-78-303-9107; Fax: +81-78-303-9094; E-mail: [tkikuchi@tri-kobe.org](mailto:tkikuchi@tri-kobe.org)

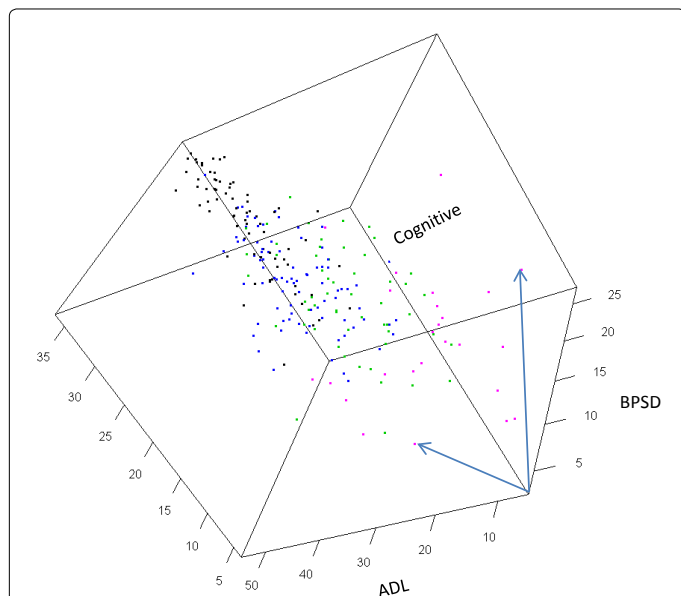
**Received** February 13, 2018; **Accepted** March 02, 2018; **Published** March 09, 2018

**Citation:** Kikuchi T, Mori T, Wada-Isoe K, Umeda-Kameyama Y, Kagimura T, et al. (2018) A Novel Dementia Scale for Alzheimer's Disease. J Alzheimers Dis Parkinsonism 8: 429. doi: [10.4172/2161-0460.1000429](https://doi.org/10.4172/2161-0460.1000429)

**Copyright:** © 2018 Kikuchi T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



**Figure 1:** A conceptual illustration of symptom progression in Alzheimer's disease. Activities of daily living (ADL) and cognitive function scores are different at stages S1 (mild) and S2 (severe), but the behavioral and psychological symptoms (BPSD) of dementia exhibit the same level  $\alpha$  at both stages. The BPSD increase and then decrease as Alzheimer's disease progresses, which may cause inaccuracies when determining the disease stage using a simple total score.



**Figure 2:** Three-dimensional distance (TDD). Domain A is for activities of daily living (ADL), domain B is for the behavioral and psychological symptoms of dementia (BPSD), and domain C is for cognitive function. The scores per patient (Domains A, Domain B, Domain C) are plotted on a three-dimensional coordinate plane. Arrows show examples of the TDD. CDR: Clinical Dementia Rating. AD severity (red dots: CDR 3, green dots: CDR 2, blue dots: CDR 1, black dots: CDR 0/0.5)

scale consisting of 13 items. We will report the details of the study in a future publication. Although we obtained statistical profiles of the test version using factor and IRT analyses during the development phase, the scale profiles require further confirmation in a different sample population. Accordingly, this study aimed to evaluate the diagnostic accuracy of the "ABC-DS." We also discuss the utility of the TDD for longitudinal studies and clinical trials.

## Methods

### Ethical approval

This multicenter observational study was conducted in accordance with the World Medical Association Declaration of Helsinki 1964 and its amendments and subsequent clarifications. The institutional review board approved the study protocol, and all caregivers and participants provided written informed consent. The study was completed as per the ethical guidelines for clinical studies set by the Ministry of Health, Labor, and Welfare of Japan. The trial was registered with the University Hospital Medical Information Network ([www.umin.ac.jp/](http://www.umin.ac.jp/); No.: UMIN000021134).

### Inclusion criteria

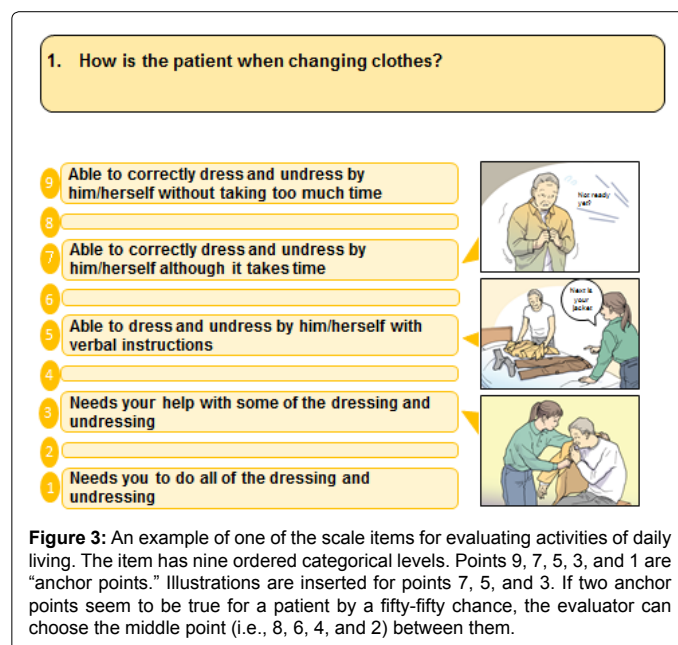
We enrolled patients who were diagnosed with AD using the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision, or with probable AD, as per the National Institute on Aging-Alzheimer's Association workgroup, National Institute of Neurological and Communicative Disorders and Stroke, and/or Alzheimer's Disease and Related Disorders Association guidelines. Patients with dementias other than AD, unstable central nervous system disorders, or psychiatric diseases were not eligible for this study.

### Participants

Participants were enrolled from 22 clinics and hospitals in Japan (see Appendix), based on the above inclusion criteria. Of the 327 participants enrolled, 15 were excluded due to protocol violations of inclusion criteria (three) or measurement procedures (12). Thus, 312 participants were eligible for the baseline assessment. The cohort included 63 participants with probable AD, 88 with mild AD, 106 with moderate AD, and 55 with severe AD.

### The ABC dementia scale

The ABC-DS consists of 13 test items, each with nine levels of ordered categorical scales. The 13 items evaluate ADL, BPSD, and cognitive function (Table 1). A sample item for evaluating ADL is attached to (Figure 3). For scale administration, evaluators interviewed



**Figure 3:** An example of one of the scale items for evaluating activities of daily living. The item has nine ordered categorical levels. Points 9, 7, 5, 3, and 1 are "anchor points." Illustrations are inserted for points 7, 5, and 3. If two anchor points seem to be true for a patient by a fifty-fifty chance, the evaluator can choose the middle point (i.e., 8, 6, 4, and 2) between them.

Item	Domain	Question
Q1	Activities of daily living	<u>Daily activities</u> : When the patient changes his/her clothes, how is he/she?
Q2		<u>Motivation</u> : How willingly does the patient undertake activities of daily life?
Q3		<u>Communication</u> : When the patient wants to communicate with others, how easily can he/she do it?
Q4		<u>Complex acts</u> : When the patient wants to use electric appliances such as a TV or an air conditioner, how well can he/she do it?
Q5	Cognitive function	<u>Recent event memory</u> : How well can the patient recall the place where he/she put his/her belongings?
Q6		<u>Recent event memory</u> : How long can the patient remember daily happenings?
Q7	Behavioral and psychological symptoms of dementia	<u>Restlessness</u> : When the patient is required to sit still, how is he/she?
Q8		<u>Irritability</u> : When the patient feels something is unsatisfactory, how is he/she?
Q9		<u>Cooperativeness</u> : When caregivers ask the patient something, how is he/she?
Q10	Cognitive function	<u>Medication</u> : When the patient takes medication, how much help does he/she need?
Q11	Activities of daily living	<u>Meals</u> : When the patient takes meals, how is he/she?
Q12		<u>Toilets</u> : When the patient uses the toilet, how much help does he/she need?
Q13	Cognitive function	<u>Care-burden</u> : How frequently should caregivers see to the patient?

Table 1: Items of the ABC-DS.

caregivers about their patients' recent episodes using a semi-structured interview [7]. Eligible caregivers were required to spend more than 3 days per week with their patients.

Participants were assessed with the ABC-DS at baseline, week 1, and week 12 by the same evaluators. For statistical analysis, we treated the nine ordered categories of the ABC scale as numeric values [8].

A PDF version of the ABC-DS in English, French, Chinese, and Korean can be downloaded from the homepage of the Mapi Research Trust at: <https://eprovide.mapi-trust.org/instruments/abc-dementia-scale>.

### Other measurements

As standard scales, we also administered the Mini-Mental State Examination (MMSE), Disability Assessment for Dementia (DAD), Neuropsychiatric Inventory-Caregiver Distress Scale (NPI-D), and Clinical Dementia Rating (CDR) (sum of boxes (SOB) and global scores) scales at baseline and week 12.

### Evaluators

The evaluators who administered the standard scales did not administer the ABC-DS. All evaluators recorded the administration duration. Evaluators for the ABC-DS and standard scales were clinicians (13.5%), certified psychologists (3.5%), nurses (41.7%), and medical clerks (41.3%).

### TDD scoring

The TDD for the ABC-DS is the mathematical distance (Euclidean distance) from the origin (0, 0, 0) to the score position of each patient (Domain A score, Domain B score, Domain C score). Domains A, B, and C refer to ADL, BPSD, and cognitive function, respectively. Here, we defined the TDD as follows:

$$TDD = \sqrt{\text{Domain A score}^2 + \text{Domain B score}^2 + \text{Domain C score}^2},$$

Where Domain A score=Q1+Q2+Q3+Q4+Q11+Q12, Domain B score=Q7+Q8+Q9, and Domain C score=Q5+Q6+Q10+Q13. For example, if a patient's score changed from (42,21,28) to (42,19,26), the TDD would change from 54.7 to 52.9. The difference over time was thus -1.8, indicating disease progression. The TDD has been submitted for application for an international patent (PCT/JP2017/035149).

### Factor analyses (construct validity)

When we selected 13 items in the development phase, we expected

		Domain A	Domain B	Domain C	TDD
	Corr	0.674			0.717
DAD	Lower	0.608			0.658
	Upper	0.730			0.767
	Corr		-0.644		-0.600
NPI-D	Lower		-0.705		-0.666
	Upper		-0.574		-0.523
	Corr			0.698	0.736
MMSE	Lower			0.637	0.680
	Upper			0.751	0.783
	Corr	-0.789	-0.498	-0.779	-0.830
CDR-SOB	Lower	-0.828	-0.577	-0.819	-0.862
	Upper	-0.744	-0.410	-0.731	-0.792
	Corr	-0.797	-0.491	-0.771	-0.826
Global CDR	Lower	-0.841	-0.583	-0.816	-0.863
	Upper	-0.753	-0.400	-0.726	-0.789

CDR: Clinical Dementia Rating; CDR-SOB: Clinical Dementia Rating Sum of Box; Corr: Correlation; DAD: Disability Assessment for Dementia; Lower: Lower band of the 95% confidence interval; MMSE: Mini-Mental State Examination; TDD: Three-dimensional Distance; Upper: Upper band of the 95% confidence interval.

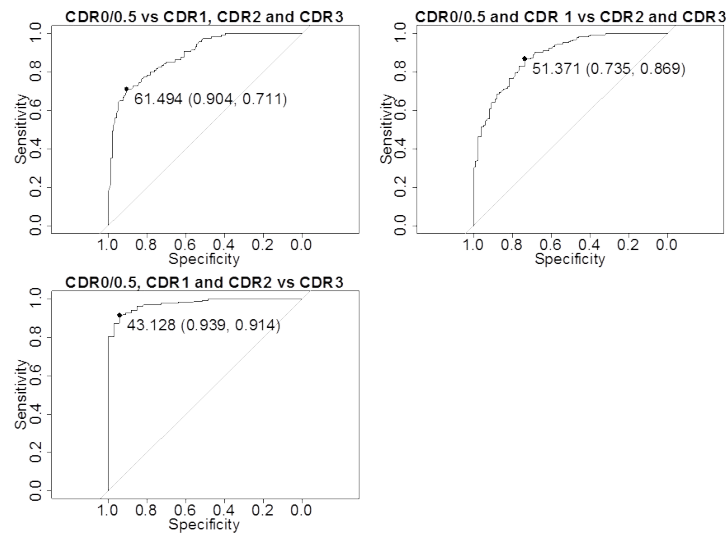
Table 2: Correlation coefficients between the ABC-DS and standard scales.

that the items would assess the ADL, BPSD, and cognitive function domains. We performed factor analyses to confirm this assumption statistically and concluded that the results of a factor analysis would be reasonable if the values for the factor loadings were  $\geq 0.4$  and the cumulative proportion of the contribution was  $\geq 0.5$  [9]. We performed a factor analysis with a *promax* rotation using the statistical software R 3.1.0 with packages *psych*, *psy*, and *polycor* [10,11].

### IRT analysis

IRT is a statistical approach for developing reliable assessment scales that measure various abilities, traits, or behavioral characteristics [12]. Here, we used this approach to inspect the difficulty parameters (locations) and discriminate parameters (steepness) of the item response category characteristic curves (IRCCCs) of graded response models [6,13,14]. Sample curves and IRCCC interpretations are shown in Supplementary (Figure 4). We accepted the models when the locations and steepness of each curve were  $< 4.0$  (absolute value) and  $\geq 0.2$ , respectively [14].

We separately applied a graded response model for each domain, assuming "unidimensionality" within a domain, whereby the items in the domain would measure a single common trait or concept of ADL,



**Figure 4:** Receiver operating characteristic curves assessing the ability of the three-dimensional distance (TDD) to indicate the global Clinical Dementia Rating (CDR). The threshold (specificity, sensitivity) values are shown in each panel.

BPSD, or cognitive function. If an item displays excess covariation or dependence on all other items in the domain, the discrimination parameter may be high relative to other items in the domain [15]. We considered that this assumption of “local independence” was not violated if the discrimination parameter of the item was  $<4.0$ . The statistical analyses were performed using R 3.1.0 with the *ltm* and *irt* packages for the graded response model [14]. Since these R packages are unable to treat nine levels of ordered categorical scales, we converted the original nine levels into five levels as follows: Levels 1&2, Levels 3&4, Level 5, Levels 6&7, and Levels 8&9 were converted into Levels 1,2,3,4, and 5, respectively, accepting a loss of information.

### Test-retest reliability (intra-rater reliability)

We used the ABC-DS scores at baseline and week 1 to evaluate the intra-rater reliability for each item with weighted kappa coefficients [16]. This test assessed the degree of agreement between the two scores given by the same evaluator for each item. If the 95% confidence interval (CI) band of the coefficients included 0.6, then we accepted the result. We also calculated intra-class correlation coefficients for the TDD to evaluate the consistency between two measurements.

### Concurrent validity

Baseline MMSE, NPI-D, DAD, CDR, and ABC-DS scores were used to evaluate concurrent validity. We calculated the correlation coefficients and 95% CIs for Domains A, B, and C corresponding to the DAD, NPI-D, and MMSE, respectively. We also calculated the correlation coefficients of Domains A, B, and C to the global CDR and the CDR-SOB, as well as the correlation coefficients of the TDD to those of other standard scales. We used Spearman coefficients and polyserial coefficients for two continuous variables and continuous and categorical variables, respectively [16].

### Receiver operating characteristic (ROC) curve for the TDD to discriminate the global CDR score

An ROC analysis [17] was used to investigate the sensitivity and specificity of the TDD for discriminating the severity of AD defined by the global CDR (categorical score). We analyzed the ROC curves

to identify the most suitable TDD thresholds for discriminating AD severity using the R 3.1.0 statistical software with the *ROCR* package [18].

We calculated the sensitivities and specificities at the thresholds for “CDR 0/0.5 vs. others,” “CDR 0/0.5 & CDR 1 vs. others,” and “CDR 0/0.5, CDR 1 & CDR 2 vs. CDR 3.” If the values of the TDD were above the threshold, we defined the test as positive, indicating a better stage.

### Changes in scores over the study period (responsiveness)

We used the ABC-DS scores and standard scale scores measured at baseline and 12 weeks for this analysis. Letting  $\Delta$  or responsiveness denote the difference in the scores between the two evaluations, we stratified the  $\Delta$  values by the baseline global CDR and calculated the mean difference, standard deviation, 95% CI, coefficient of variation (CV: standard deviation/mean), and *P* values testing the null hypothesis  $H_0: \Delta=0$ . The accuracy and repeatability of the measurements are high if the absolute value of the CV is small.

## Results

### Patient characteristics

The patients included 126 (40.4%) men and 186 (59.6%) women. The mean (standard deviation) age of the patients was 80.6 (7.1) years, and the mean duration of education was 11.1 (2.6) years, including 9 years of compulsory education in Japan. The mean (SD) of DAD, NPI-D, MMSE, and CDR-SOB were 67.40 (28.34), 11.06 (12.72), 18.26 (7.31), and 7.20 (4.92), respectively.

### Administration duration

The mean administration durations for the ABC-DS and the CDR were 9.96 (4.79) min and 26.4 (9.8) min, respectively.

### Factor analysis

In the development phase, we found that the 13 items of the ABC-DS constructed the ADL, BPSD, and cognitive function domains. We confirmed this result in the present study (Supplementary Table 1). Factors 1, 2, and 3 in (Supplementary Table 2) correspond to Domains



A, C, and B, respectively. Each bold value in the table indicates the factor that gives the item the largest factor loading in the three factors. The bold factor loadings were >0.4, and the cumulative proportion of the contribution by the three factors was 0.585.

We also tested four-factor and five-factor models. The four-factor model was not reasonable, because the fourth domain did not have the item (component) that had the largest factor loading in four domains. The five-factor model was also not practical, because the fourth and fifth domains contained only one component item each. Accordingly, we concluded that more than three factors were redundant. The present results are consistent with those for the development phase (data not shown).

### IRT analysis

The IRCCCs for each domain are shown in Supplementary (Supplementary Figures 1-4). The difficulty (location) parameters of the IRCCCs were within reasonable ranges (between -4 and 4) for the standardized AD severities. The discrimination (steepness) parameters were >0.2 and <4.0 for all items in each domain, except for Q1 (4.1), which had a parameter that was similar to that observed in the development phase. We confirmed that the parameter values were comparable to those determined in the development phase (data not shown).

### Test-retest reliability

For the test-retest reliability evaluation, data were available from 219 of 312 participants; however, one participant was interviewed by different evaluators and therefore excluded from this analysis. We summarized the results in Supplementary (Table 2). The mean kappa coefficient values for Q8 and Q9 were <0.6, but the 95% CI bands included 0.6. The kappa coefficient values for the other items showed moderate or strong similarity between the two measurements. The TDD intra-class correlation coefficient was 0.964.

### Concurrent validity

Correlation coefficients and 95% CIs were calculated to compare Domains A, B, and C, along with the TDD, with the corresponding standard scales (Table 3). The correlation coefficients for "Domain A vs. DAD," "Domain B vs. NPI-D," and "Domain C vs. MMSE" were moderate. The correlations of Domains A and C with the CDR-SOB were strong. The TDD was also strongly correlated with the CDR-SOB and the global CDR.

### ROC of the TDD for discriminating AD severity

Figure 4 shows the ROC curves, along with the threshold, specificity, and sensitivity, for discriminating AD severity for the following: "CDR 0/0.5 vs. others," "CDR 0/0.5 & CDR 1 vs. CDR 2 & CDR 3," and "CDR 3 vs. others." The area under the curve values were 0.887, 0.884, and 0.974, respectively. These results provided the following TDD thresholds of AD severity: CDR 0/0.5 if  $70.29 \geq TDD > 61.49$ ; CDR 1 if  $61.49 \geq TDD > 51.37$ ; CDR 2 if  $51.37 \geq TDD > 43.13$ ; and CDR 3 if  $43.13 \geq TDD > 7.81$ .

### Score changes over the study period (responsiveness)

For the responsiveness evaluation, data were available from 222 of the 312 participants; however, four participants were excluded from this analysis owing to missing paired values for baseline and week 12. We stratified the  $\Delta$  values by the baseline CDR (Table 3). Bold values indicate statistically significant differences ( $P < 0.05$ ) between the baseline and week 12 scores. There were no statistically

#### 1) CDR 2 at baseline.

Variable	n	Mean	LLM	ULM	SD	CV	P value
Domain A	40	-1.55	-2.83	-0.27	3.99	-2.58	<b>.02</b>
Domain B	40	0.70	-0.24	1.64	2.95	4.22	.14
Domain C	40	-1.35	-2.58	-0.12	3.83	-2.84	<b>.03</b>
Arithmetic sum	40	-2.20	-4.55	0.15	7.36	-3.35	.07
TDD	40	-1.37	-2.70	-0.04	4.15	-3.02	<b>.04</b>
DAD	39	-3.96	-6.68	-1.25	8.38	-2.11	<b>.01</b>
NPI-D	40	0.80	-2.68	4.28	10.90	13.62	.65
MMSE	40	-0.65	-1.46	0.16	2.53	-3.89	.11
CDR-SOB	40	0.80	0.33	1.27	1.48	1.84	<b>.00</b>

#### 2) CDR 1 at baseline

Variable	n	Mean	LLM	ULM	SD	CV	P value
Domain A	72	-1.21	-2.11	-0.31	3.82	-3.16	<b>.01</b>
Domain B	72	0.57	-0.06	1.20	2.69	4.72	.08
Domain C	72	-0.90	-1.82	0.01	3.90	-4.32	.05
Arithmetic sum	72	-1.54	-3.09	0.01	6.60	-4.28	.05
TDD	72	-1.11	-2.03	-0.20	3.89	-3.49	<b>.02</b>
DAD	72	-6.60	-9.42	-3.78	11.99	-1.82	<b>&lt;.0001</b>
NPI-D	72	0.26	-1.79	2.32	8.74	33.13	.80
MMSE	72	-0.71	-1.31	-0.10	2.58	-3.64	<b>.02</b>
CDR-SOB	72	0.42	0.13	0.71	1.24	2.97	<b>.01</b>

#### 3) CDR 0/0.5 at baseline

Variable	n	Mean	LLM	ULM	SD	CV	P value
Domain A	93	-0.60	-1.21	0.01	2.95	-4.90	.05
Domain B	93	-0.25	-0.60	0.11	1.72	-6.97	.17
Domain C	93	-0.89	-1.56	-0.23	3.22	-3.61	<b>.01</b>
Arithmetic sum	93	-1.74	-2.81	-0.67	5.19	-2.98	<b>.00</b>
TDD	93	-0.93	-1.55	-0.31	3.00	-3.23	<b>.00</b>
DAD	93	-2.92	-5.58	-0.27	12.90	-4.41	<b>.03</b>
NPI-D	93	0.56	-0.79	1.91	6.56	11.73	.41
MMSE	93	-0.03	-0.58	0.51	2.64	-81.82	.91
CDR-SOB	93	0.06	-0.15	0.27	1.01	16.99	.57

CDR: Clinical Dementia Rating; CDR-SOB: Clinical Dementia Rating Sum of Box; CV: Coefficient of Variation; DAD: Disability Assessment for Dementia; LLM: Lower Limit of the Mean; MMSE: Mini-Mental State Examination; NPI-D: Neuropsychiatric Inventory-Caregiver Distress Scale; SD: Standard Deviation; TDD: Three-dimensional Distance; ULM: Upper Limit of the Mean.

**Table 3:** Score changes over the study period stratified by the baseline Clinical Dementia Rating.

significant differences in  $\Delta$  for any scale if the baseline severity was CDR 3 (Supplementary Table 3). The  $\Delta$  of the total score of 13 items (arithmetic sum) was significant only at CDR 0/0.5 (Table 3). The CV values indicated that the measurement variation in TDD was smaller than that found for the arithmetic sum at CDR 1, 2, and 3 (Table 3).

The  $\Delta$  in the CDR-SOB was statistically significant at CDR 1 and 2; however, the results were inferior to those for the TDD, which successfully detected the differences at CDR 2, 1, and 0/0.5 (Table 3).

### Discussion

The present study confirms the construct validity of all 13 items comprising the three domains, and the concurrent validity of the ABC-DS domain scores with their corresponding standard scale scores as well as with the CDR. The scale's intra-rater reliability, as determined by weighted kappa coefficients, was acceptable, and it detected statistically significant changes in AD severity over the 12-week study period. The correlation coefficients between the TDD and the CDR were strong, and the accurately discriminated AD severity in our ROC analysis.

Collectively, these results indicate that the TDD is similarly informative to the CDR for staging AD and measuring disease progression.

The IRT markedly aided in the ABC-DS development. During the development phase, we repeatedly revised descriptions of the items by examining the parameters of the IRCCCs until the location and steepness of the curves became reasonable. Here, we repeated these IRT analyses and checked that the estimated parameters were similar across the two different sample populations. The results suggested that our item descriptions in the ABC-DS were stable and accurate for assessing AD. It should be noted that the discrimination parameter of Q1 was larger than the conventional threshold of 4.0; hence, this item may have violated local independence. However, this should not seriously affect the scale as a whole because the value was not much larger than the threshold.

We think that the staging and monitoring of AD progression would be most accurate if we concurrently evaluate ADL, BPSD, and cognitive function. To conduct this concurrent evaluation, we introduced the TDD approach. The present study identified three main advantages of the TDD. First, the correlation coefficients between the TDD and the standard scale scores were better than those between the domain scores and the standard scale scores, except for "Domain B vs. NPI-D." Second, the TDD discriminated the severity of AD, as diagnosed by CDR, with satisfactory sensitivity and specificity. Third, the total score (arithmetic sum) of the 13 items in the ABC-DS did not detect significant changes in disease progression over 12 weeks when baseline severity was CDR 1 or 2, whereas the TDD did detect significant changes in these cases.

Most standard dementia assessment scales only evaluate total scores (i.e., a simple sum or an arithmetic sum). However, the sum of item scores can fail to detect true changes in disease progression, as was observed here. Using the ABC-DS as an example, let us consider a patient whose ADL domain score rises by three points between assessments, while the BPSD domain score falls by three points. If we calculate the simple sum of the 13 items, these changes cancel out, and this approach would thus not detect a change in the patient's progression. Total score-based assessments may, therefore, increase the risk of false-negative results. Our TDD scoring system detects otherwise obscured changes by depicting a change in the score positions for ADL, BPSD, and cognitive function. For example, if the score position moves from (27,22,26) to (30,19,26), which represents the same simple total score, then the TDD changes from 43.46 to 44.01. This rationale, in tandem with our present observations, underscores the utility of the TDD for determining or comparing longitudinal treatment effects in clinical research.

This study has some limitations. Several aspects of the ABC-DS require further examination in future research, including scale responsiveness. While we evaluated responsiveness in an observational cohort over a 12-week study period, we did not specify medical treatments and instead observed the natural progression of the participants. Future work should furthermore examine responsiveness in a comparative clinical trial setting, to confirm the utility of the TDD approach and the ABC-DS.

In conclusion, the present study confirms the construct validity, concurrent validity, intra-rater reliability, and responsiveness of the ABC-DS over a 12-week period. Importantly, the TDDs calculated from the ABC-DS represent a useful approach for evaluating patient responses in clinical trials of AD.

#### Acknowledgement

We would like to thank Mr. Hiromi Tsukada for preparing the datasets and

assisting with the analyses, and we appreciate the contribution of Mr. Motohiro Sakamine as a project manager. We would like to thank Dr. Mikio Yoshidomi, Dr. Hideaki Kaneda, and Prof. Masanori Fukushima for reviewing the manuscript. We would like to thank Editage ([www.editage.jp](http://www.editage.jp)) for English language editing.

#### Contributions

The following physicians from the ABC Dementia Scale Study Group collected the study data: Dr. Toshiyuki Watabe (Watabe Clinic), Dr. Akinori Ueki (Ueki Dementia and Geriatric Psychiatry Clinic), Dr. Hideki Yamamoto (Hon-machi clinic of Kita-Osaka Medical Corporative Union), Dr. Kitamura Yuri (Nanohana Clinic), Dr. Naruhiko Maki (Maki Hospital), Dr. Nobuya Kawabata (Yachiyo Hospital), Dr. Hirotake Uchikado (Shonan Inaho clinic), Dr. Naoto Kobayashi (Azuma-Dori Clinic), Dr. Chika Nishimura (Kurumi Clinic), Dr. Chiaki Kudoh (Kudoh Clinic for Neurosurgery and Neurology), Dr. Tomotake Otsuka (Takase Town Nishikagawa Hospital), Dr. Masashi Takita (Takita Memory Mental Clinic), Dr. Kazuyoshi Harada (Harada clinic), Dr. Hideki Ishizu (Jikei Hospital), Dr. Yumi Umeda-Kameyama (Taito Municipal Taito Hospital), Dr. Takashi Ueda (Ueda Neurosurgery), Assoc. Prof. Norifumi Tsuno (The University of Kagawa Hospital), Dr. Mitsuhiro Tsujihata (Nagasaki Kita Hospital), Dr. Kiyoshi Fujita (Seishinkai Okehazama Hospital), Dr. Aoi Yoshiwa (Takada Chuo Hospital), Dr. Masato Yamashita (Yamashita Clinic), and Prof. Masahiro Akishita (The University of Tokyo Hospital).

#### Funding

This work was funded by Daiichi Sankyo Co., Ltd.; however, this company did not have any role in the design of the study, the data collection, statistical analyses, or writing of the paper. The authors did not disclose this manuscript to the company before submission for publication.

#### References

1. Robert P, Ferris S, Gauthier S, Ihl R, Winblad B, et al. (2010) 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* 2: 24.
2. Holthoff VA, Ferris S, Ihl R, Robert P, Winblad B, et al. (2011) Validation of the relevant outcome scale for Alzheimer's disease: a novel multidomain assessment for daily medical practice. *Alzheimers Res Ther* 3: 27.
3. Holthoff V, Ferris S, Gauthier S, Ihl R, Robert P, et al. (2013) 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* 28: 164-172.
4. Pardo C, Alcalde S, García M, Sánchez Magro I3; en nombre del grupo del Estudio ROSA (2017) Validation of the Spanish-language version of the Relevant Outcome Scale for Alzheimer's Disease. *Neurologia* 32: 417-423.
5. Nishikawa T (2010) Alzheimer's disease: Symptoms and clinical progression (in Japanese), in: *Neurological Medicine*, Tokyo: Kagaku-Hyounronsha, p. 277-283.
6. Nguyen TH, Han HR, Kim MT, Chan KS (2014) An introduction to item response theory for patient-reported outcome measurement. *Patient* 7: 23-35.
7. Jamshed S (2014) Qualitative research method-interviewing and observation. *J Basic Clin Pharm* 5: 87-88.
8. Rhemtulla M, Brosseau-Liard PE, Savalei V (2012) When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods* 17: 354-373.
9. Beavers AS, Lounsbury JW, Richards JK, Huck SW, Skolits GJ, et al. (2013) Practical considerations for using exploratory factor analysis in educational research. *Pract Assess Res Eval* 18: 1-13.
10. Everitt BS (2005) *An R and S-Plus® Companion to Multivariate Analysis*. (1st edn), Springer, London.
11. Teetor P (2011) *R Cookbook, Proven Recipes for Data Analysis*. (1st edn), O'Reilly, Sebastopol.
12. Reise SP, Revicki DA (2014) *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. (1st edn), Routledge, New York.
13. Baker FB, Kim SH (2017) *The Basics of Item Response Theory Using R*. (1st edn), Springer, London.
14. Rizopoulos D (2006) *ltm: An R package for latent variable modeling*. *J Stat Soft* 17: 1-25.
15. Hambleton RK, Swaminathan H, Rogers WH (1991) *Fundamentals of Item Response Theory*. Sage Publications, Newbury Park.
16. Agresti A (2002) *Categorical Data Analysis*. (2nd edn), Wiley & Sons, Hoboken.

17. Krzanowski WJ, Hand DJ (2009) ROC Curves for Continuous Data, in: Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, Boca Raton.
18. López-Ratón M, Rodríguez-Álvarez MH, Cadarso-Suárez C, Gude-Sampedro F (2014) OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *J Stat Soft* 61: 1-36.