Development of the Korean version of Alzheimer’s Disease Assessment Scale (ADAS-K)

J. C. Youn¹²³, D. Y. Lee⁴, K. W. Kim⁴, J. H. Lee⁵, J. H. Jhoo³⁶, K. U. Lee⁵, J. Ha²⁷ and J. I. Woo¹²³*

¹Department of Neuropsychiatry, Seoul National University, College of Medicine and Seoul National University Hospital, Seoul, Korea
²Neuroscience Research Institute of Medical Research Center, Seoul National University, and Clinical Research Institute of Seoul National University Hospital, Seoul, Korea
³BK 21 Life Science, Seoul National University, College of Medicine, Seoul, Korea
⁴Department of Neuropsychiatry, Kyunggi Provincial Hospital for the Elderly, Yongin, Kyunggido, Korea
⁵Department of Neuropsychiatry, Kangwon National University College of Medicine, Chunchon, Kangwondo, Korea
⁶Department of Neuropsychiatry, Metro Hospital, Anyang, Kyunggido, Korea
⁷The Korean Association for Dementia, Seoul, Korea

SUMMARY

Objective The purpose of this study was the development of the Korean Version of Alzheimer’s Disease Assessment Scale (ADAS-K).

Method ADAS-K was administered to 84 AD patients as well as 105 non-demented control subjects. Three aspects of reliability were tested. To evaluate the validity of ADAS-K, discriminant validity and concurrent validity were tested. To evaluate the sensitivity of ADAS-K to disease severity, all subjects, AD patients and control subjects, were grouped by CDR scale and their mean scores on ADAS-K were compared.

Result ADAS-K demonstrated high levels of reliability. Mean ADAS-K scores for AD patients were significantly different from the control group (p < 0.01). Furthermore, ADAS-K exhibited significant correlations with other tests and scales (range 0.45–0.85, p < 0.01). In ROC curve analysis, ADAS-K displayed high diagnostic efficacy and the optimal cut-off point was selected between 18/19. ADAS-K was able to discriminate the degree of AD severity according to CDR classification. Our results suggested that ADAS-K-cog was sensitive to very mild AD.

Conclusion We demonstrated that ADAS-K is a reliable and valid instrument not only for AD diagnosis but also for evaluation of its severity. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS — ADAS-K; ADAS-K-cog; ADAS-K-noncog; reliability; validity

INTRODUCTION

Alzheimer’s disease (AD), which is caused by progressive neurodegeneration, affects various domains of cognitive and noncognitive function. Thus, any instrument for evaluation of Alzheimer’s disease needs to measure multiple domains of cognitive and noncognitive function (Rosen et al., 1984).

Alzheimer’s Disease Assessment Scale (ADAS) was initially developed for severity evaluation of multiple major dysfunctions arising in Alzheimer’s disease patients, but it also has diagnostic applications (Rosen et al., 1984; Zec et al., 1992). The scale consists of 11 cognitive (ADAS-cog) and 10 noncognitive items (ADAS-noncog), which can be used separately. ADAS-cog includes performance based short neuropsychological tests, while ADAS-noncog includes observation based behavioral rating items. ADAS has not only advantages of single instrument but also covers major characteristics of AD.

In addition to characterizing the core abnormalities of AD, ADAS has proven to be sensitive to AD
progression, especially at the moderate stage (Rosen et al., 1984; Stern et al., 1994). Based on this sensitivity, ADAS has been adopted for the evaluation of treatment efficacy in many anti-dementia drug trials (Knapp et al., 1994; Rogers and Friedhoff, 1996; Rosler et al., 1999) and become a standard instrument for assessment of the drug inducing cognitive effect (Doraiswamy et al., 1997). The purpose of this study was the development of the Korean version of ADAS through translation of the original English version into Korean while maintaining its basic structure. We also tried to elucidate the psychometrical properties of ADSA-K by assessing its reliability and validity.

**METHOD**

1. **Translation of ADAS into Korean**

Translation and back-translation performed by bilingual psychiatrists. In translating ADAS into Korean, the overall structure of the 21 item original English version (Rosen et al., 1984) was essentially maintained. However, the English words in the verbal memory tests (Word Recall Task and Word Recognition Task) and the Object Naming test were not translated into semantically equivalent Korean ones, but rather words were selected considering the relative frequency and imagery in Korean language (Lee, 1976; Seo, 1997). For Word Recall Task, 10 high frequency and high imagery nouns were chosen as in English version. In the case of Object Naming test, three groups were chosen according to word frequency (high, medium, and low frequency) as in the English version. In the four 12-word lists of Word Recognition Task, low or medium frequency words were not excluded, rather an attempt was made to balance the semantic categories of the words between one target list and three confounding lists.

2. **Subjects**

The participants in this study consisted of 84 AD patients and 105 non-demented elderly controls. AD cases were recruited from three different institutes: the Dementia Special Clinic of Seoul National University Hospital, Kyunggi Provincial Hospital for the Elderly, and Kwanak-gu Public Health Center in Seoul city. Control subjects were recruited from the Dementia Early Detection program in Kwanak-gu Public Health Center. On entry into the study, informed consent was obtained from each subject or caregiver. All subjects were evaluated according to the clinical assessment protocol of the Korean Version of the Consortium to Establish a Registry for Alzheimer’s Disease Assessment battery (CERAD-K) (Lee et al., 2002).

Patients with severe neurological, medical, or psychiatric disorders other than AD were excluded. All patients were 50 years old or older, could speak Korean and had reliable informants who could provide adequate clinical histories. Subjects who had no formal education but could read Korean language were included, although illiterate subjects were excluded. The diagnosis of dementia was made by DSM-IV (APA, 1994) and then the diagnosis of AD was made according to the criteria of the National Institute of Neurological and Communicative Disorders and Stroke, Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) (McKhann et al., 1984). Both probable and possible AD was included. The final diagnosis was assigned by a panel discussion among at least four psychiatrists with expertise in dementia research. Most of the AD patients were receiving medication such as acetylcholinesterase inhibitors. All subjects in the control group were defined as being in a non-demented state.

The demographic characteristics of both groups are presented in Table 1. No significant differences were found in age or male/female ratio between the two groups. The mean educational level of AD patients was significantly higher than non-demented control subjects ($p < 0.01$). In the AD group, 15 (17.9%) cases were staged as very mild (CDR 0.5), 37 (44.0%) as mild (CDR 1), 28 (33.3%) as moderate

| Table 1. Demographic characteristics of AD patients and non-demented subjects |
|-----------------|-----------------|
| **non-demented subjects** $(n = 105)$ | **AD patients** $(n = 84)$ |
| **Age (years)** | **Sex** | **Education (years)** | **Clinical Dementia Rating scale** |
| Mean ± SD | Range | | 0 | 0.5 | 1 | 2 | 3 |
| 71.9 ± 5.3 | 56–86 | 4.2 ± 4.1 | 0–16 | 66 (63.5) | 39 (36.5) | 37 (44.0) | 28 (33.3) | 4 (4.8) |
| 73.3 ± 8.6 | 50–90 | 6.8 ± 5.1 | 0–18 | |

*Note: number (percent) of subjects unless otherwise indicated.*
(CDR 2), and four (4.8%) as severe (CDR 3). In the non-demented control group, 66 (63.5%) cases exhibited no cognitive impairment (CDR 0) while the other 39 (66.5%) were cognitively impaired to a mild degree without having dementia (CDR 0.5).

3. Reliability Testing

Three aspects of reliability were examined; internal consistency of scale, inter-rater reliability and test–retest reliability. Internal consistency was examined by Crohnbach’s alpha. In order to evaluate Inter-rater reliability, eight AD patients and seven non-dementia controls were rated with ADAS-K simultaneously by two independent researchers (tester and rater). ADAS-K was administered again to ten AD patients and non-dementia controls by the same rater at 4–6 weeks after the initial assessment to determine the test–retest reliability. For evaluating inter-rater and test–retest reliabilities, Pearson correlation coefficients were calculated.

4. Validity testing

To evaluate the discriminant validity, the mean ADAS-K (total, cognitive, and non-cognitive) scores were compared between the two groups by ANCOVA after adjusting for age, sex and educational level.

As reference tests or severity indices, also administered alongside ADAS-K were the Korean MMSE version of the Consortium to Establish a Registry for Alzheimer’s Disease (MMSE-KC) (Lee et al., 2002), the Korean Version of Short Blessed Test (SBT-K) (Lee et al., 1999), Clinical Dementia Rating (CDR) (Hughes et al., 1982) and Global Deterioration Scale (GDS) (Reisberg et al., 1982). Pearson correlation coefficients between ADAS-K, MMSE-KC and SBT-K and Spearman correlation coefficients between ADAS-K, CDR and GDS were calculated to examine the concurrent validity of ADAS-K. To evaluate the concurrent validity further and to identify the diagnostic utility, Receiver Operating Characteristics (ROC) curve analysis was undertaken for ADAS-K.

To evaluate the sensitivity of ADAS-K to variations in severity level, all subjects were divided into six groups (two non-demented, four AD) based on CDR scale grades. Among those subjects with CDR 0.5 staging, some could be diagnosed as AD with a very mild AD classification. The other CDR 0.5 subjects were classified with mild cognitive impairment. Because education level varied among the six different stages, the mean ADAS-K scores for the 6 different severity levels were compared by ANCOVA after adjusting for education level.

All statistical analyses were done using SPSS for window, version 10.

RESULTS

1. Reliability

The results of the reliability analysis are presented in Table 2. ADAS-K and its two subscales were found to be highly reliable as a whole, although ADAS-K-cog exhibited higher reliabilities than ADAS-K-noncog. Chronbach’s coefficients alpha for ADAS-K-total and ADAS-K-cog in AD patients were both above 0.8, indicating excellent degrees of internal consistency. Chronbach’s coefficients alpha in the non-dementia subject group (0.41–0.64) were lower than those in AD patients (0.61–0.88). All inter-rater reliabilities and test–retest reliabilities, as measured by Pearson correlation coefficients, were statistically significant (p < 0.01) except that of ADAS-K-noncog in the AD group (0.58, p = 0.078). Pearson correlation coefficients for ADAS-K-total and ADAS-K-cog in the non-dementia subject group were generally lower than those in the AD group.

2. Validity

ADAS-K clearly discriminated the AD group from the control group. The mean test scores of ADAS-K were compared between the two groups by ANCOVA after adjusting for age, sex and educational level.

Table 2. Results of reliability test for non-demented control subjects and Alzheimer’s disease patients

<table>
<thead>
<tr>
<th></th>
<th>Non-demented control subjects</th>
<th>Alzheimer’s disease patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td>IRa</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>TRa</td>
<td>0.80</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Note: IC, internal consistency; IR, inter-rater reliability; TR, test–retest reliability.

*All correlations are significant at 0.01 level by Pearson correlation analysis except* (nonsignificant).
patients were significantly higher than those of the control group for ADAS-K-total and also for its two subscales (cognitive and non-cognitive) \( (p < 0.01, \text{Table 3}) \).

Pearson correlation coefficients between ADAS-K and the cognitive tests (MMSE-KC and SBT-K) and Spearman correlation coefficients between ADAS-K and the severity scales (CDR and GDS) were statistically significant in the AD group (Table 4, \( p < 0.01 \)). Correlation coefficients for ADAS-K-total and ADAS-cog ranged from 0.642 (ADSA-K-total and SBT-K) to 0.758 (ADS-K-total and MMSE-KC). Similar to the results of reliability analysis, correlation coefficients for ADAS-K-noncog were lower than ADAS-K-cog and ranged from 0.319 (ADAS-K-noncog and SBT-K) to 0.363 (ADS-K-noncog and CDR). The estimated correlation coefficients between ADAS-K and MMSE-KC were negative due to the opposite direction of the two scoring methods.

ADAS-K was demonstrated to have a high diagnostic ability. From ROC curve analysis, the area under curve (AUC), indicating the overall performance of the test, could be calculated (Hanley and McNeil, 1982). AUC ranges from 0 to 1, with an AUC value above 0.85 generally considered to indicating good diagnostic ability (Metz, 1978). AUC of ADAS-K was 0.945, confirming the excellent diagnostic ability of this scale (Table 5, Figure 1). The optimal cut-off point could be determined at the level 18/19, at which sensitivity and specificity were 0.863 and 0.874, respectively. Because ADAS-K-cog could be used independently, ROC curve analysis was also performed for ADS-K-cog. AUC of ADS-K-cog was 0.942. The optimal cut-off point was selected between levels 15 and 16, where sensitivity and specificity were 0.904 and 0.825, respectively.

Mean scores on ADAS-K for the 6 different stages according to CDR indices were significantly different (\( df = 6, F = 95.9, p < 0.001, \text{Figure 2} \)). In post-hoc multiple comparison analyses (Scheffe’s test) for ADAS-K-total, all of the different groups differed significantly from one another, except for the no cognitive impairment group (CDR 0) vs the mild cognitive impairment group (CDR 0.5) comparison and except for the mild cognitive impairment vs very mild AD (CDR 0.5 AD) comparison. For ADAS-K-cog, all of the different groups differed significantly from one another, except for the cognitive unimpaired group (CDR 0) vs the mild cognitive impairment group (CDR 0.5) comparison.

### Table 3. Mean scores of ADAS-K, MMSE-KC and SBT-K in non-demented control subjects and Alzheimer’s disease patients*

<table>
<thead>
<tr>
<th></th>
<th>Non-demented control subjects</th>
<th>Alzheimer’s disease patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAS-K-total</td>
<td>12.3 ± 5.4</td>
<td>35.0 ± 15.4</td>
</tr>
<tr>
<td>ADAS-K-cog</td>
<td>11.5 ± 5.2</td>
<td>31.7 ± 13.4</td>
</tr>
<tr>
<td>ADAS-K-noncog</td>
<td>0.8 ± 1.2</td>
<td>3.4 ± 3.5</td>
</tr>
<tr>
<td>MMSE-KC</td>
<td>24.5 ± 2.8</td>
<td>14.7 ± 6.9</td>
</tr>
<tr>
<td>SBT-K</td>
<td>5.0 ± 5.3</td>
<td>20.4 ± 7.6</td>
</tr>
</tbody>
</table>

*Mean scores of AD patients are significantly different from non-dementia subjects \( (p < 0.001, \text{by Student t-test, two-tailed}).

ADAS-K, Alzheimer’s Disease Assessment Scale; MMSE-KC, MMSE in CERAD-K; SBT-K, Korean Version of Short Blessed Test; ADAS-total, whole scale of ADAS; ADAS-K-cog, cognitive subscale of ADAS-K; ADAS-K-noncog, noncognitive subscale of ADAS-K.

### Table 4. Correlations between ADAS-K and cognitive tests (MMSE-KC, SBT-K), global scales (CDR, GDS) in Alzheimer’s disease patient

<table>
<thead>
<tr>
<th></th>
<th>ADAS-K-total</th>
<th>ADAS-K-cog</th>
<th>ADAS-K-noncog</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE-KC</td>
<td>-0.758</td>
<td>-0.769</td>
<td>-0.359</td>
</tr>
<tr>
<td>SBT-K</td>
<td>0.642</td>
<td>0.664</td>
<td>0.319</td>
</tr>
<tr>
<td>CDR</td>
<td>0.707</td>
<td>0.692</td>
<td>0.363</td>
</tr>
<tr>
<td>GDS</td>
<td>0.667</td>
<td>0.656</td>
<td>0.329</td>
</tr>
</tbody>
</table>

*All correlations are significant at \( p < 0.01, \text{by Pearson correlation (MMSE-KC, SBT-K) and Spearman correlation analysis (CDR, GDS)}.\)

CDR, Clinical Dementia Rating scale; GDS, Global Deterioration Scale; (for other abbreviations: see Table 3).

### Table 5. Sensitivity and specificity of ADAS-K-cog for the diagnosis of AD patients

<table>
<thead>
<tr>
<th>Test</th>
<th>Cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAS-K-total</td>
<td>15/16</td>
<td>0.932</td>
<td>0.806</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>16/17</td>
<td>0.904</td>
<td>0.825</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17/18</td>
<td>0.863</td>
<td>0.845</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18/19</td>
<td><strong>0.863</strong></td>
<td><strong>0.874</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19/20</td>
<td>0.849</td>
<td>0.903</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20/21</td>
<td>0.822</td>
<td>0.903</td>
<td></td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>11/12</td>
<td>0.986</td>
<td>0.602</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>12/13</td>
<td>0.959</td>
<td>0.650</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13/14</td>
<td>0.945</td>
<td>0.728</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14/15</td>
<td>0.918</td>
<td>0.816</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15/16</td>
<td><strong>0.904</strong></td>
<td><strong>0.825</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16/17</td>
<td>0.877</td>
<td>0.854</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17/18</td>
<td>0.836</td>
<td>0.893</td>
<td></td>
</tr>
</tbody>
</table>

*Bold characters indicate optimal cut-off point as determined by ROC analysis.

AUC, Area Under Curve.
Figure 1. Receiver Operator characteristics (ROC) curves of ADAS-K, SBT-K and MMSE-KC

![ROC curves](image1)

Figure 2. Means scores on ADAS-K according to CDR severity scaling. *: mean score on ADAS-K-total was not significantly different from CDR 0 and very mild AD. #: mean score on ADAS-K-cog was not significantly different from CDR 0. Except * and #, all mean scores on ADAS-K-total and ADAS-K-cog differed significantly from one another by post hoc analysis (Scheffe comparison, $p < 0.05$)

![Means scores](image2)
DISCUSSION

Our study demonstrated that ADAS-K is a reliable and valid instrument for the diagnosis and evaluation of AD. It was also established the value of ADAS-K-cog as a test for early detection of dementia.

When translating a cognitive test from the original language version into a different language version, cross-cultural issues are important. Thus, we tried not only to retain the original structure of ADAS-K but also to evaluate the reliability and validity of our translated version. Each test was translated under the concept of equivalence (Flaherty et al., 1988). Because the word lists incorporated in memory test and naming test are most liable to be influenced by cultural differences, careful consideration was given to the construction of the word lists. The word lists for recall and recognition tasks and the list for object naming were not merely literally translated but were rather selected from word frequency data on the written Korean language (Seo, 1997; Lee, 1976). Thus the structure, method of application and goal of each test could be maintained.

Previous studies have reported that ADAS-noncog has lower reliabilities than ADAS-cog (Weyer et al., 1997; Inzitari et al., 1999; Tsolaki et al., 1997; Penacaspanoia, 1997)—the finding confirmed by our own results. The relatively lower reliabilities observed for ADAS-noncog could be explained by several factors. Firstly, the weak item construction affects internal consistency. Correlations between noncognitive items were weak in general and sometimes were nonsignificant (Mohs, 1996). Secondly, the subjectivity mainly affects inter-rater reliability. For ADAS-noncog, the scoring of observation-based items could increase the degree of inconsistency. Thirdly, the temporal change of behavioral symptoms in AD patients affects the results of test–retest reliability. In our ten subject investigation into test–retest reliability, most patients showed little change of ADAS-K-noncog over 1 month, but three patients manifested a marked decrease in ADAS-K-noncog score (above 5 points) over the same period. In addition to this spontaneous change, various medications including psychiatric drugs demonstrated an effect on patient behavior in our results.

Although ADAS was not originally designed as a diagnostic instrument, it has potential diagnostic applications (Zec et al., 1992). ADAS-K was able to clearly discriminate AD patients from non-dementia control subjects. In addition, ADAS-K revealed high correlations with other popular screening tests. In ROC analysis, ADAS-K displayed high diagnostic efficacy (AUC 0.945 for ADAS-K-total, AUC 942 for ADAS-K-cog). The optimal cut-off point 15/16 (sensitivity 90%, specificity 83%) of ADAS-K-cog was similar to that reported from a previous study featuring a Greek version (Tsolaki et al., 1997).

Because ADAS is known as a sensitive instrument for AD progression (Rosen et al., 1984; Stern et al., 1994), it is important to confirm whether ADAS-K is able to reflect the severity of Alzheimer’s disease. ADAS-K not only exhibited high correlations with global severity scales but also differentiated each stage from among the four stages in the AD group. Our results indicated that ADAS-K reflects the severity of AD. But a more definitive answer awaits the results of a comprehensive longitudinal study.

In the development of new anti-dementia drugs, the early detection of dementia has become a new focus for research. It remains unclear whether ADAS could be used for early detection. Some authors have suggested that ADAS might be relatively insensitive to very early phase of AD (Mohs et al., 1997). But other researchers have documented that ADAS-cog is sensitive to very mild AD (Zec et al., 1992). From our results, it is supported that ADAS-K-cog could discriminate very mild AD from non-demented state. However, in our study ADAS-K-cog could not discriminate mild cognitive impairment from no cognitive impairment. Mild cognitive impairment is known as transitional period between normal aging and AD (Petersen et al., 1999). Thus there is a need for the delayed recall of memory, attention and executive function, which are known as sensitive tests to the development of AD, to be tested in addition to ADAS-cog in at-risk populations like mild cognitive impairment (Mohs et al., 1997).

It should be pointed out that the majority of our subjects had a lower educational level which could affect the neuropsychological test results and lead to the over diagnosis of dementia (Geerings et al., 1999). However, our diagnosis did not merely rely on a single neuropsychological test but was based on the complete clinical data received from the patients and informants. Thus the possibility of over diagnosis should have been minimized by our diagnostic methods. Nevertheless, the high proportion of the population with low educational levels reflects the real clinical situation in Korea. In the 1995 Korean census (Korean National Statistical Office, 2001), of the 4 135 287 subjects over 60 years, 3 243 424 (78.4%) had less than six years of education. In a similar finding, 87 subjects (83.7%) in our study sample had less than six years of education.

In our study, a reference test to examine the concurrent validity of ADAS-K-noncog was not undertaken.
because a reliable behavioral scale for AD is rare in Korea. However, we consider that global severity scales like CDR and GDS can provide an effective supplement.

REFERENCES


