Diagnostic Utility of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV) Among Elders with Alzheimer's Dementia

Made Syanesti Adishesa and Magdalena S. Halim
Department of Psychology
Atma Jaya Catholic University of Indonesia

The aim of this study was to examine the diagnostic utility of the Indonesian version of Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV-ID) in classifying between typical aging and Alzheimer's dementia (AD). We administered the WAIS-IV-ID to 47 AD patients (28 females and 19 males; mean age 68 ± 8 years). Severity of dementia was classified into three categories: mild (20 patients), moderate (13 patients), and severe (14 patients). On the basis of receiver operatic characteristic (ROC) analysis, the areas under the curve (AUCs) of each index are as follows: (a) .83, 95% CI [0.738, 0.895] for Full IQ, (b) .88, 95% CI [0.81, 0.94] for Perceptual Reasoning, (c) .79, 95% CI [0.69, 0.86] for Processing Speed, (d) .78, 95% CI [0.69, 0.86] for Verbal Comprehension, and (e) .71, 95% CI [0.61, 0.8] for Working Memory. These AUC values indicate that the WAIS-IV-ID has moderate accuracy in identifying people with AD. This study also raised awareness for the necessity of a standardized process in translating and using cognitive tests, especially in clinical practices.

Keywords: WAIS-IV, dementia, Alzheimer's disease, AD, diagnostic utility, ROC

Tujuan penelitian ini adalah mengukur performa diagnostik dari Wechsler Adult Intelligence Scale Fourth Edition versi Bahasa Indonesia (WAIS-IV-ID) dalam mengklasifikasikan individu dengan gangguan demensia Alzheimer individu yang mengalami penuaan normal. Alat ukur WAIS-IV-ID diadministrasikan pada 47 pasien dengan gangguan demensia Alzheimer (27 wanita dan 19 pria; rata-rata usia 68 ± 8 tahun). Tingkat keparahan gangguan dibagi menjadi tiga kategori: ringan (20 subjek), sedang (13 subjek), dan berat (14 subjek). Berdasarkan teknik analisis *receiver operating characteristic,* nilai *area under curve* untuk setiap indeks adalah sebagai berikut: (a) .82, 95% CI [0.738, 0.895] untuk Full IQ, (b) .88, 95% CI [0.81, 0.94] untuk Perceptual Reasoning, (c) .79, 95% CI [0.69, 0.86] untuk Processing Speed, (d) .78, 95% CI [0.69, 0.86] untuk Verbal Comprehension, and (e) .71, 95% CI [0.61, 0.8] untuk Working Memory. Nilai AUC ini mengindikasikan bahwa WAIS-IV-ID memiliki tingkat akurasi sedang dalam mengidentifikasikan individu dengan demensia Alzheimer. Studi ini juga menyadarkan perlunya proses standarisasi dalam penerjemahan dan pemanfaatan uji kognitif, terutama dalam praktik-praktik klinis.

Kata kunci: WAIS-IV, Alzheimer, diagnostic utility, ROC

Advances in medical technology and therapies have contributed to increasing life expectancy around the world. In 2011, life expectancy in Indonesia has increased to 69.65 years and elderly citizens make up 7.58% of the total population (Pusat Data dan Informasi Kemenkes RI, 2013). The increasing number of elderly citizens is an indicator of a country's development; however, it also raises new challenges. One of those challenges is degenerative diseases due to the human aging process. Brain deterioration is a part of the degeneration process which could lead to neuropsychological disorders, such as dementia, the most common degenerative disease in elderly.

Dementia is marked by progressive cognitive impairment across multiple domains and significant

impairment in social or occupational functioning (Sadock, Sadock, & Ruiz, 2015). Alzheimer's disease (AD) is the most common etiology for dementia, and accounts for about 50 to 75% of dementias. In AD, plaques and tangles build up in the brain structure, which eventually leads to the death of nerve cells and loss of brain tissue. People with AD also have a shortage of some important chemical in their brain. These chemical messengers help transmit signals around the brain, and the lack of these chemicals causes the signals to be transmitted less effectively. AD is a progressive disease, which means that gradually more parts of the brain are damaged. As this happens, more symptoms develop and also become more severe.

Early detection is a critical point in treating AD, as it is said to be the key to treating the disease before it causes irreversible brain damage (Sadock, Sadock, & Ruiz, 2015). Nevertheless, detecting early symptoms has been found to be a difficult task because they tend to be overlooked and considered an inevitable consequence of aging (Urakami, 2007; Wong, Leung, Fung, Chan, & Lam, 2013). The highly variable trajectories of cognitive decline also make it more difficult to recognize initial symptoms (Wong et al., 2013). Hence, it is important to establish accurate cognitive screening tools to detect AD so as to facilitate early intervention and focused clinical management.

Screening tests are used by neurologists to assist in achieving a more accurate diagnosis of AD. They are typically concise and only require a short amount of time to administer, but information provided by them is limited. For example, one of the most widely used screening tests, Mini-Mental State Examination (MMSE), was reported poorly in detecting cognitive impairment due to its inability to detect complex cognitive deficits (Nasreddine et al., 2005; Pendlebury, Cuthbertson, Welch, Mehta, & Rothwell, 2010). MMSE was also less sensitive in detecting cognitive impairment in highly educated patients or in those with high premorbid functioning (Sadock, Sadock, & Ruiz, 2015). Due to said limitations, screening tests are considered as an initial guideline to further and more detailed assessment (Cullen, O'Neill, Evans, Coen, & Lawlor, 2007). However, in Indonesia screening tests are sometimes used as the main method for assessing cognitive functions. In spite of the fact that decisions based on cognitive tests may have a major impact on diagnosis and treatment planning in AD, very few studies have been done in Indonesia to investigate their accuracy in classifying AD from typical aging. This may lead to misdiagnosis or delayed/incorrect treatment. Other issue that should be noted is that the usage of most cognitive tests (including screening tests) in Indonesia is unauthorized, and details of its translation, standardization, or psychometric properties have not been reported (Suwartono, Halim, Hidajat, Hendriks, & Kessels, 2014).

The limitations of existing screening tests lead to the increasing need of a comprehensive assessment of intelligence, which can provide a better understanding of cognitive functions in AD (Izawa, Urakami, Kojima, & Ohama, 2009). The most commonly used test for intelligence in clinical setting is the Wechsler Adult Intelligence Scale. The recent version of Wechsler
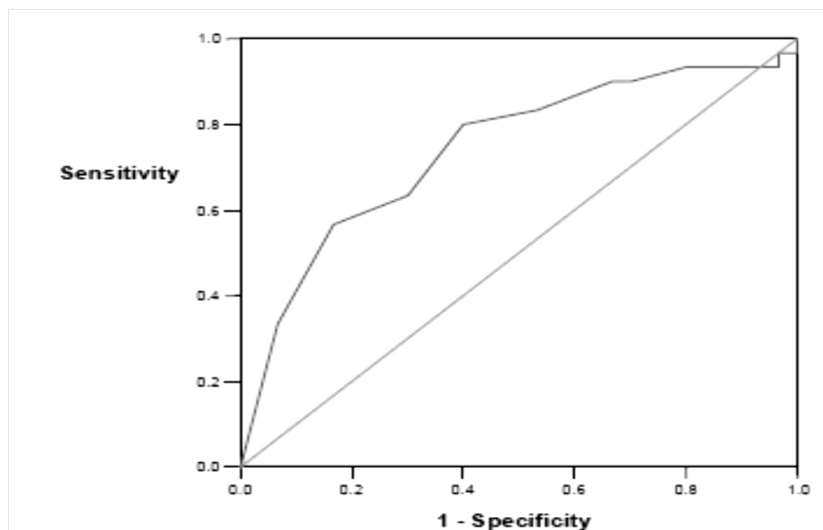

Figure 1. The ROC space.
Source: Pintea, S. & Moldovan. R. (2009).

Adult Intelligence Scale (WAIS-IV) was published in 2008 and translated into Indonesian in 2014 (WAIS-IV-ID; Suwartono et al., 2014). In the Indonesian version, item sequences were reordered due to differences in index difficulties but still stayed close to the original items for content purposes. It showed promising psychometric properties and has been tried out in mild AD sample. The result revealed that people with mild AD had relatively preserved perceptual reasoning (*Median* = 86), followed by verbal comprehension (*Median* = 83), working memory (*Median* = 80), and processing speed (*Median* = 79) as the most impaired cognitive function (Kuswanto & Halim, 2015). Other studies have investigated the usage of previous versions of WAIS in neuropsychological assessment. The results revealed that WAIS-III had good overall diagnostic accuracy (when combined with Wechsler Memory Scale) and proved to be useful in evaluating AD severity (Taylor & Heaton, 2001; Larrabee, Largen, & Levin, 2008).

Given that cognitive assessment result play an important role in diagnosing AD, additional research on its diagnostic utility is necessary. The fundamental measures of diagnostic utility are sensitivity (i.e. true positive rate) and specificity (i.e. true negative rate). However, sensitivity and specificity rely heavily on cutoff score; they change as the cutoff score varies. Therefore, when evaluating a continuous-scale diagnostic test it would be helpful to plot sensitivity and specificity over a range of values of interest, as is done in an ROC (receiver operating characteristic) curve (Zou, O'Malley, & Mauri, 2007). Based on these considerations, this study would use the ROC curve to measure the diagnostic accuracy of WAIS-IV among elders with AD.

Method

Participants

Participants were classified into two categories based on their clinical diagnosis: clinical (with AD) and typical aging (without AD). Those in clinical group were outpatient of hospitals in the city of Bekasi and Tangerang, or residents in a senior living facility in Bogor, and had been diagnosed with AD by neurologists. The neurologists also assigned a severity level of AD to each participant as follows: mild (20 participants), moderate (13 participants), and severe (14 participants). Of 47 participants in the clinical group, 28 were females and 19 were males; mean age was 68

± 8 years.

Participants in the typical aging group were selected from the WAIS-IV-ID standardization sample and matched with the clinical group in terms of age and education level (Suwartono et al., 2014). Of 52 participants in the typical aging group, 43 were females and 9 were males; their mean age was 69 ± 3 years.

Instruments

WAIS-IV-ID is an individually administered standardized and norm-referenced IQ test composed of a standard battery of 15 subtests ($M = 10$; $SD = 3$) that create four index composite scores and a full IQ score (FIQ; $M = 100$; $SD = 15$; Wechsler, 2008). It was adapted into Indonesian by Suwartono et al. (2014) and the final translation was authorized by Pearson Assessment. In this study, the administration of WAIS-IV-ID included the discontinue rule. This means that administration of a subtest is discontinued after a certain amount of consecutive failures. Raw scores are converted using the American norms because the Indonesian version is yet to be completed.

Procedure and Analysis

Two groups of participants were differentiated based on diagnosis obtained from neurologists: clinical (with AD) and typical aging (without AD) group. The diagnostic procedures used to categorize the participants were assumed to be valid. We then administered the WAIS-IV-ID, and data collected from both groups were analyzed on three levels: full IQ score, index scores, and scaled scores for all subtests.

We used these scores as a classifier which relied on a threshold. For example, participants whose full IQ were below the cutoff score would be labeled as 'positive' (with AD), while participants whose full IQ were above would be labeled as 'negative' (without AD). This diagnosis would then be compared to the valid diagnosis obtained from the neurologists. If the valid diagnosis was positive (i.e. participant was in the clinical group) and correctly classified as 'positive', it would be counted as a true positive; if the same outcome was incorrectly classified as 'negative', it would be counted as a false negative. If the valid diagnosis was negative (i.e. the participant was in the typical aging group) and correctly labeled as 'negative', the outcome would be counted as true negative; if the same outcome was incorrectly labeled as 'positive', it would be counted as a false positive (Brown & Davis, 2006). From these outcomes, we calculated

Table 1
Descriptive Statistics for WAIS-IV-ID Subtest Scaled Scores in Clinical and Typical Aging Group

Subtest / Index	Clinical group[a] (*Median*)	Typical aging group[a] (*Median*)	*Z-score*
Block design	8	10	5.01*
Similarity	4	7	4.34*
Digit span	6	7	2.49**
Matrix reasoning	5	8	6.16*
Vocabulary	7	9	3.01*
Arithmetic	7	8	3.78*
Symbol search	5	8	4.06*
Visual puzzle	6	9	5.21*
Information	4	6	4.24*
Coding	3	7.5	4.72*
Letter-number sequencing	7	7	2.49**
Figure weight	6	9	3.67*
Comprehension	3	7	5.82*
Cancellation	1	7	3.53*
Picture completion	3	6	4.73*

Note.   [a] Using Wechsler standard scores ranging from 1 to 20 for subtest ($M = 10$; $SD = 3$)
* $p < .01$, two-tailed test ** $p < .05$, two-tailed test

Table 2
Descriptive Statistics for WAIS-IV-ID Index Scores and Full IQ in Clinical and Typical Aging Group

Index	Clinical group[a] (*Median*)	Typical aging group[a] (*Median*)	*Z-score*
Verbal comprehension index	72	82	4.89*
Perceptual reasoning index	79	94	6.66*
Working memory index	74	89	3.68*
Processing speed index	68	86	4.93*
Full IQ	70	86	5.59*

Note.   [a] Standard Wechsler IQ classification for index scores and full IQ ($M = 100$; $SD = 15$; Wechsler, 2008)
* $p < .05$, two-tailed test

the sensitivity (i.e. the probability that the full IQ score was labeled 'positive' when AD was present) and specificity (i.e. the probability that the full IQ score was labeled 'negative' when AD was not present).

Since sensitivity and specificity vary when the cutoff score is changed, we plot these variations for all possible cutoff scores in the ROC curve (see Figure 1). In other words, the ROC curve (colored blue in the figure) is a representation of the sensitivity (i.e. true positive rate) on the X-axis and 1-specificity (i.e. false positive rate) on the Y-axis.

The green diagonal line where sensitivity equals to 1-specificity represents the performance of a random test. In other words, when the classifier is randomly guessing, it correctly identifies half of the positives and half of the negatives. Therefore, all cutoff points above the random diagonal line are considered to perform better than random guessing (Fawcett, 2006).

To determine the ability of WAIS-IV-ID in discriminating clinical from the typical aging group, we calculated the area under curve (AUC) values with 95% CI. The AUC is the total area under the ROC curve, which is a measure of the overall performance of a diagnostic test, i.e. its diagnostic utility. The larger the area is, the better the performance will be (Westin, 2001). The interpretation of the AUC of a test is the following: the AUC is the probability that a randomly selected individual from the clinical group has a test result indicating greater suspicion than that for a randomly chosen individual from the typical aging group (Zhou, Obuchowski, & McClish, 2002). Regarding the AUC utility in determining the ability of a test to discriminate between groups, Streiner and Cairney (2007) show that the accuracy of tests with AUC between .50 and .70 is low; between .70 and .90 is moderate, and over .90 is high.

We also used the ROC curve to determine the optimal cutoff score. This is the most northwestern point in the ROC space, which has the highest true positive rate and the lowest false positive rate. In other words, the optimal cutoff score is the one which maximizes true positive and true negative.
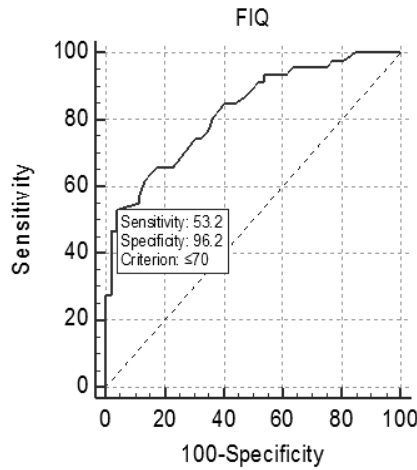
Figure 1. ROC curve of FIQ in elders with AD compared with WAIS-IV-ID standardization sample.

Table 3
AUC Values of WAIS-IV-ID Index Scores and Subtests

Subtest /Index	AUC	Level of Accuracy*
Verbal comprehension index	.78	Moderate
Perceptual reasoning index	.88	Moderate
Working memory index	.71	Moderate
Processing speed index	.78	Moderate
Block design	.79	Moderate
Similarity	.75	Moderate
Digit span	.64	Low
Matrix reasoning	.86	Moderate
Vocabulary	.67	Low
Arithmetic	.72	Moderate
Symbol search	.74	Moderate
Visual puzzle	.88	Moderate
Information	.77	Moderate
Coding	.80	Moderate
Letter-number sequencing	.64	Low
Figure weight	.59	Low
Comprehension	.84	Moderate
Cancellation	.60	Low
Picture completion	.77	Moderate

Note. * Interpreted based on Streiner & Cairney (2007)

Results

Descriptive statistics of full IQ and index scores obtained from all participants are presented in Table 1, whilst descriptive statistics of subtest scaled scores are presented in Table 2.

The obtained scores were analyzed using Mann-Whitney U-test and it was revealed that the clinical group showed significantly lower performance than the typical aging group across all subtests, index scores, and full IQ.

In the clinical group, the lowest subtest scaled scores was obtained for Cancellation, followed by Coding, Comprehension, and Picture Completion with the same median value. In contrast to these areas of weaker performance, the highest subtest scores in the clinical group was Block Design, followed by Vocabulary and Arithmetic.

The result of the ROC analysis comparing 47 elders with AD to 52 participants from the WAIS-IV-ID standardization sample is presented in Figure 2. The AUC of .83, 95% CI [0.73, 0.89] quantifies this visual result. This indicates that the probability that a randomly selected individual from the clinical group has a full IQ indicating greater suspicion than that for a randomly chosen individual from the typical aging group is 83% (Zhou, Obuchowski, & McClish, 2002). The AUC value also indicates that the full IQ of WAIS-IV-ID showed moderate accuracy in identifying elders with AD (Streiner & Cairney, 2007).

The ROC analysis was also used to calculate the optimal threshold. The result showed sensitivity value of .53 and specificity value of .96 when cutoff score was set at ≤ 70. These values indicated that when threshold was set at the optimal point of ≤ 70, full IQ of WAIS-IV-ID could classify 53% participants from the clinical group as positive (with AD) and 96% participants from the typical aging group as negative (without AD).

Diagnostic utilities of all subtests and index scores were calculated and presented in Table 3.

Based on the statistical analysis conducted in this study, the full IQ and all index scores of WAIS-IV-ID showed moderate accuracy in classifying elders with AD. Subtests with the highest AUC values were Visual Puzzle, Matrix Reasoning, Comprehension, and Coding that had AUC above .80.

Discussion

The scores obtained showed that participants from the AD group performed poorly in Cancellation, Coding, Comprehension, and Picture Completion. As a comparison, WAIS-IV was administered to 44 elderly adults with probable AD and the lowest subtest scaled scores were obtained for Symbol Search, Coding, and Information (Kaufman & Lichtenberger, 2009).

ROC analysis showed that full IQ and all index scores of WAIS-IV-ID had moderate accuracy in classifying clinical group from the typical aging group.

From a statistical point of view, these findings further support the promising psychometric properties that WAIS-IV-ID has shown (Suwartono et al., 2014).

As a comparison, a study by Larner (2012) suggests that the English version of Montreal Cognitive Assessment (MoCA) had high accuracy in classifying AD ($n = 150$) with an AUC value of .91. This value indicates that full IQ of WAIS-IV-ID had slightly lower diagnostic utility compared to MoCA (.83 vs .91). Full IQ of WAIS-IV-ID was also less sensitive than MoCA (.53 vs .97) but far more specific (.96 vs .60). Furthermore, the full IQ of WAIS-IV-ID showed similar diagnostic accuracy to another screening test, the Mini Mental State Examination (.83 vs .83; Larner, 2012). In terms of sensitivity, the full IQ of WAIS-IV-ID was slightly lower (.53 vs .65) but more specific (.96 vs .89). Based on this comparison, we could conclude that while MoCA might be more preferable for screening AD with higher diagnostic utility and sensitivity, the WAIS-IV-ID offered a more comprehensive assessment that would prevent over diagnosing due to its high specificity (.96). In other words, high specificity reduced the possibility of misdiagnosing early symptoms as AD, as this could lead to treatments that do no good or perhaps do harm.

Another comparison could be made with a study which explored the sensitivity and specificity of WAIS-III factor scores in neuropsychological assessment (Taylor & Heaton, 2001). This particular study used exploratory and confirmatory factor analyses to identify six constructs measured in WAIS-III: Verbal Comprehension, Perceptual Organization, Processing Speed, Working Memory, Auditory Memory, and Visual Memory. The most sensitive factor scores for AD group were Visual Memory and Auditory Memory (.97), while the least sensitive were Verbal Comprehension (.64). The wide range suggests that some factor scores are more sensitive to AD than others. The AUC values presented in Table 2 support this, as shown by the diagnostic utility of the index scores (ranging from .71 to .88) and the subtests (ranging from .59 to .88). Both of these results showed that variations in the subtest scores or index scores may give us more information regarding the cognitive functions of people with neurological disorder, such as AD. This could also be seen as one of the advantages of using battery test alongside brief cognitive screening test, as the score variations provides more insight to cognitive functions.

It should be noted that while MoCA and MMSE are the two most widely known cognitive tests in Indonesia, there are other cognitive tests being used in screening AD, such as: the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), Boston naming test, clock drawing test, etc. However, very little research is done on the translation method, standardization, psychometric properties, and their accuracy in classifying AD from typical aging (Suwartono et al., 2014). Since WAIS-IV-ID is one of the few widely studied cognitive tests in Indonesia, clinicians would be able to take the information into consideration when using the test and make a more assured decision.

ROC analysis conducted on all subtests also revealed four subtests with AUC values above .80. This indicates that the probability that a randomly selected individual from the clinical group showed results which indicate greater suspicion than that of a randomly chosen individual from the typical aging group is above 80% (Zhou, Obuchowski, & McClish, 2002). While we would not recommend using a sole subtest as a screening tool, poor performance in all these subtests raises greater suspicion of AD and therefore prompts a more thorough assessment.

Limitations and Future Research

This study has several limitations. First, the limited number of participants in this study makes the interpretation of the results should be done with some caution. The second limitation was related to the diagnoses given to participants. Since there is no standardized procedure for diagnosing AD, the examining neurologists used a variety of evaluation method to diagnose AD. Although each participant was given a physical and cognitive evaluation, his or her diagnosis was based on a variety of tests, interviews, behavioral checklists, and clinical judgments. This variation may have an impact on the results of this study.

Future research should continue investigating cognitive tests that contribute to diagnosing AD. Method of diagnosis should be controlled in order to allow unambiguous diagnostic utility results to emerge. Additional research may also be conducted on environmental or other factors that might impact test performances (such as the presence of family member during test administration, living environment, daily habits, etc.), to allow more control when measuring diagnostic utility.

Although the results should be considered preliminary because of its limitations, clinicians should be cautious in interpreting screening tests results as evidence of AD. Since most of cognitive tests used in Indonesia was adapted into Indonesian without

proper translation and standardization; thus, their psychometric properties remained unknown, there may be cultural factors or statistical error that could lead to misdiagnosis or delayed/incorrect treatment. Therefore, more studies should be done comparing psychometric properties of the various cognitive tests used in Indonesia to develop a more accurate approach in diagnosis.

Conclusion

This study aimed to investigate the diagnostic utility of WAIS-IV-ID as a screening tool for individuals with Alzheimer's Dementia (AD). The results revealed a moderate accuracy of WAIS-IV-ID in identifying people with AD. Comparisons with cognitive screening tests showed that while less sensitive, WAIS-IV-ID had higher specificity which could reduce the possibility of overdiagnosing. As a battery test, WAIS-IV-ID also offered more insight to cognitive functions from the variations in the subtest or index scores. Therefore the use of WAIS-IV-ID alongside AD screening tests is highly recommended for a more thorough cognitive assessment. This study also raised awareness for the necessity of a standardized process in translating and using cognitive tests, especially in clinical practices.

References

Brown, C. D., & Davis, H. T. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Systems, 80,* 24-38.

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F., & Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery, & Psychiatry, 78,* 790-799.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 2,* 861-874.

Izawa, Y., Urakami, K., Kojima, T., & Ohama, E. (2009). Wechsler Adult Intelligence Scale, 3rd edition: Usefulness in the Early Detection of Alzheimer's Disease. *Yonago Acta Medica, 52,* 11-20.

Kaufman, E. O. & Lichtenberger, A. S. (2009). *Essentials of WAIS-IV Assessment.* New Jersey: John Wiley & Sons.

Kuswanto, K. G., & Halim, M. S. (2015). *Profil fungsi kognisi berdasarkan WAIS-IV pada individu dengan mild Alzheimer's disease.* Paper submitted for publication.

Larrabee, G. J., Largen, J. W., & Levin, H. S. (2008).

Sensitivity of age-decline resistant ("Hold") WAIS subtests to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 7*(5), 497-504.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 52,* 695-699.

Pendlebury, S. T., Cuthbertson, F. C., Welch, S. J. V., Mehta, Z., & Rothwell, P. M. (2010). Underestimation of cognitive impairment by Mini-Mental State Examination versus the Montreal Cognitive Assessment in patients with transient ischemic attack and stroke: A population-based study. *Stroke, 41,* 1290-1293.

Pintea, S., & Moldovan. R. (2009). The Receiver-Operating Characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies, 9*(1), 49-66.

Pusat Data dan Informasi Kemenkes RI. (2013). *Gambaran Kesehatan Lanjut Usia di Indonesia.* Jakarta: Kementrian Kesehatan Republik Indonesia.

Sadock, B. J., Sadock, V. A., & Ruiz, P. (2015). *Kaplan & Sadock's Synopsis of psychiatry: Behavioral sciences/Clinical psychiatry* (11th ed.). Philadelphia: Wolters Klower.

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to Receiver Operating Characteristic Curves. *The Canadian Journal of Psychiatry, 52,* 121-128.

Suwartono, C., Halim, M. S., Hidajat, L. L., Hendriks, M. P. H., & Kessels, R. P. C. (2014). Development and reliability of the Indonesian Wechsler Adult Intelligence Scale – Fourth edition (WAIS-IV). *Psychology, 5,* 1611-1619.

Taylor, M. J., & Heaton, R. K. (2001). Sensitivity and specificity of WAIS-III/WMS-III demographically corrected factor scores in neuropsychological assessment. *Journal of the International Neuropsychological Society, 7*(7), 867-874.

Urakami, K. (2007). Prevention of dementia. *Psychogeriatrics, 7,* 93-97.

Westin, L. (2001). *Receiver operating characteristic (ROC) analysis: Evaluating discriminance effects among decision support systems.* Retrieved from cs.umu.se/research/reports/2001/018/part1.pdf

Wong, C. H. Y., Leung, G. T. Y., Fung, A. W. T., Chan, W. C., & Lam, L. C. W. (2013). Cognitive predictors for five-year conversion to dementia in community-dwelling Chinese older adults. *International Psycho-*

geriatrics, 25(7), 1125-1134.

Zhou, X. H., Obuchowski, N. A., McClish, D. K. (2002). *Statistical methods in diagnostic medicine.* New Jersey: Wiley-Interscience.

Zou, K. H., O'Malley, J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation, 115,* 654-657.