Dement Neurocogn Disord. 2021 Oct;20(4):62-69
https://doi.org/10.12779/dnd.2021.20.4.62
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Original Article

A Comparison of the Performance between the 60- and 120-Second Conditions of the Korean-Color Word Stroop Test: Color Reading (K-CWST: CR)

Sun Hwa Lee [1], **Yeonwook Kang** [2,3], **Minji Song** [2,4], **Geon Ha Kim** [5], **Jee Hyang Jeong** [1]

[1]Department of Neurology, Ewha Womans University Seoul Hopital, Seoul, Korea
[2]Department of Psychology, College of Social Sciences, Hallym University, Chuncheon, Korea
[3]Department of Neurology, Hallym University Sacred Heart Hospital, Anyang, Korea
[4]Department of Neurology, Hallym University Sacred Heart Hospital, Chuncheon, Korea
[5]Department of Neurology, Ewha Womans University Mokdong Hopital, Seoul, Korea

Received: Sep 10, 2021
Revised: Oct 15, 2021
Accepted: Oct 18, 2021

Correspondence to
Jee Hyang Jeong
Department of Neurology, Ewha Womans University Seoul Hospital, 260 Gonghang-daero, Gangseo-gu, Seoul 07804, Korea.
E-mail: jjeong@ewha.ac.kr

ORCID iDs
Sun Hwa Lee
https://orcid.org/0000-0001-5998-6984
Yeonwook Kang
https://orcid.org/0000-0003-4426-9127
Minji Song
https://orcid.org/0000-0002-7687-5239
Geon Ha Kim
https://orcid.org/0000-0001-5395-1924
Jee Hyang Jeong
https://orcid.org/0000-0001-7945-6956

Funding
This research was supported by the Original Technology Research Program for Brain Science through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2018M3C7A1057137) and by the grants of the Korea Health Technology R&D Project through the Korea Health Industry Development

ABSTRACT

Background and Purpose: The Korean-Color Word Stroop Test: Color Reading (K-CWST: CR) included in the Seoul Neuropsychological Screening Battery, 2nd Edition (SNSB-II) examines inhibitory control deficit. It provides normative data for both 60- and 120-second conditions, but the validity of the 60-second condition has not yet been proven. This study examined the validity of the 60-second condition by observing concordance between the performances in cognitively normal, MCI, and mild dementia groups.
Methods: There were 1,336 patients performed the SNSB-II, including the K-CWST: CR. Based on the cognitive test results, activities of daily living, and clinical interview, the patients were assigned to normal cognition (n=104), MCI (n=884), or mild dementia (n=348) groups. Abnormal performance on the K-CWST: CR was operationally defined as 1SD below the normative mean. The receiver operating characteristic curve analyses were conducted to compare the discriminability between the 60- and 120-second conditions.
Results: The percentages of abnormal performance in the MCI group were 41.5% and 42.3%, and those in the mild dementia group were 82.7% and 82.4% for the 60- and 120-second conditions, respectively. The areas under the curve for the 60- and 120-seconds were as follows; 0.80 and 0.81 in differentiating normal from MCI; 0.95 and 0.96 in normal from mild dementia; and 0.77 and 0.77 in MCI from mild dementia.
Conclusions: The 60-second condition of the K-CWST showed very similar results, not statistically different from the 120-second condition. Therefore, the 60-second condition could be used interchangeably with the 120-second condition in a clinical setting.

Keywords: Stroop Test; Mild Cognitive Impairment; Dementia

INTRODUCTION

The Stroop test, also known as the color-word naming test, has been used widely since it was originally developed by Stroop.[1] The Stroop test is used in multiple areas of clinical and research fields dealing with cognitive functions and personality traits. In clinical setting,

Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HU20C0271 and HU21C0016).

Conflict of Interest

The authors have no potential conflicts of interest.

Author Contributions

Conceptualization: Lee SH. Data curation: Lee SH. Song M, Kang Y. Formal analysis: Lee SH, Song M; Methodology: Lee SH, Kang Y; Project administration: Jeong JH, Kim GH; Supervision: Jeong JH, Kim GH, Kang Y; Validation: Lee SH, Kang Y; Writing - original draft: Lee SH; Writing - review & editing: Lee SH, Kang Y, Song M, Kim GH, Jeong JH.

the Stroop test has been used to measure several cognitive constructs, including sustained attention and inhibition.[2] Its basic paradigm is based on the time difference between simply reading word regardless of written color (congruent condition) and naming the colors of color words written in non-matched colors (incongruent condition). The delayed reaction time in the incongruent condition compared to the congruent condition is known as Stroop interference. The subjects must inhibit automatized responses of reading words and select appropriate response and it elicits time delay.[3]

Age-related deficits in a variety of cognitive functions could be accounted for inefficient selective attention, which in turn results in decreased inhibitory processing skills.[4] Since a number of studies demonstrated that the Stroop effect is greater for older adults, the Stroop test has been used to examine age-related declines in attention and inhibitory control.[5,6] The contribution of inhibitory deficits to the decline of executive function in the early stages of Alzheimer's disease (AD) has been consistently reported in several studies using different versions of the Stroop test.[7-10] Similar findings were also reported in the Korean clinical population.[11,12]

Currently, there are several standardized versions of commercial Stroop test developed for the elderly in Korea. Each includes some type of control and interference tasks, but it differs in the number of colors and items used, and how items are presented and scored. Since 2017, the Korean government has implemented the Dementia National Responsibility System policy to provide financial support for dementia evaluations. The selected neuropsychological batteries are the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K),[13] Seoul Neuropsychological Screening Battery, 2nd Edition(SNSB-II),[14] and the Literacy Independent Cognitive Assessment (LICA).[15] According to the data provided by Health Insurance Review and Assessment Service, the annual usage of neuropsychological tests in 2020 was the highest for the SNSB-II (66.5%), followed by the CERAD-K (23.9%) then the lowest for the LICA (6.7%).[16] They all include different versions of the Stroop test. Considering the SNSB-II is the most frequently used in dementia evaluation, the K-CWST would be the most frequently used version in Korea. The validity of the K-CWST for the 120-second conditions for the elderly has already been proven.[17] Lee and her colleagues also examined the effectiveness of interference score, interference rate, correct response rate, and a total number of correct responses as indicators of stoop effect on the K-CWST. The total number of correct responses of K-CWST: CR was found to be the most effective indicator.[17]

The Stroop test of CERAD-K consists of two control tasks, word reading and color reading conditions, and one interference task, each with maximum time limit of 45-second for each task. Both LICA and SNSB-II stoop tasks have one control task and one interference task, but each task has a different maximum time limit. It is 180-second for the LICA and 120-second for the SNSB-II, but the SNSB-II provides normative data for 60-second condition. It is certainly difficult for the elderly and particularly harder for those with cognitive impairment to undergo long hours of cognitive testing.[18] Reduced time conditions would increase the efficiency and usefulness of the cognitive test. However, the concurrent validity of the 60-second condition with the 120-second condition was not reported yet. This study aimed to investigate the validity of the 60-second condition by observing the concordance between the performances of the two-time conditions in the cognitively normal, MCI, and mild dementia groups.

METHODS

Participants

This is a retrospective analysis of in-and out-patients who visited the neurology department for cognitive assessments. Data collection was based on clinical chart reviews between 2017 and 2020. The study was approved by the ethics board of the hospital, where the data were collected. Patients' medical history or pathological diseases were not addressed in the study since it was not directly related to the purpose of the study. The target patients were 50 years of age or older. The main inclusion criterion was a complete administration of the SNSB-II, including the K-CWST: CR. A total of 1,757 patients were considered for inclusion. To minimize any possible confounding variables other than inhibitory control ability which may result in poor performance on the K-CWST: CR, we set the following exclusion criteria: 1) Illiterate to less or equal to 3 years of education, 2) presence of visual or perceptual difficulties, 3) hearing or comprehension difficulties and 4) those with severe dementia, whose global Clinical Dementia Rating (CDR) score is greater or equal to 2, since they were more likely to forget the instructions while testing and have a high incidence of color confusion.[19] Finally, 1,336 patients were analyzed in this study (**Fig. 1**). The mean age of them was 70.83 years (standard deviation [SD]=9.70, range: 50–90 years), and the mean year of education was 10.13 years (SD=3.83, range: 4–23 years). There were 723 females and 613 males.

We made an operational definition of abnormal performances as below 1SD from the normative mean on the neuropsychological tests of the SNSB-II. Neurologists rated the CDR and the Korean-Instrumental Activities of Daily Living (K-IADL)[20] through interviewing patients and informants. Based on the O'Bryant et al.'s criteria[21] of Normal Cognition (NC) [CDR sum of boxes (CDR-SOB)=0], MCI (CDR-SOB=0.5-2.5) and mild dementia (DE) [CDR-SOB≥3.0] and the K-IADL score,[20] the patients were classified into normal cognition (NC, n=104), MCI (n=884), and DE (n=348) groups.

Measurements

The K-CWST consisted of word-reading (K-CWST: WR) and color-reading (K-CWST: CR) cards. Each card was A4 size and had 112 four-color names (red, yellow, blue, and black) written in unmatched colored ink on a gray background. Each card was presented in four columns,

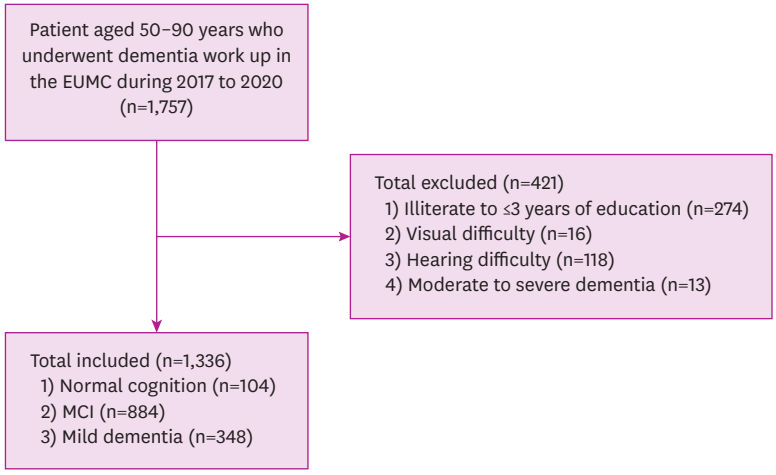

Fig. 1. Flow diagram of patient inclusion/exclusion.
EUMC: Ewha Womans University Medical Center.

https://doi.org/10.12779/dnd.2021.20.4.62

with 28 words written on each column. The main variable for the present study was the total correct score for K-CWST: CR in the 60- and 120-second conditions found as the most effective indicator of the Stroop effect.[17] The scores were converted into z-scores using normative data stratified by age and education.[11] Z scores of -1 or less were defined as abnormal.

Statistical analysis

Descriptive statistics (mean, SD) were reported for the relevant variables. Continuous variables were analyzed using the analysis of variance (ANOVA) and categorical variables with the χ^2 test. The percentages of abnormal performances for each time condition were identified. McNemar tests were used for non-parametric analysis. The receiver operating characteristic (ROC) curve analyses were conducted to compare the discriminability between the 60- and 120-second conditions. Descriptive statistical analysis, ANOVA, χ^2 test, and McNemar tests were conducted using the IBM SPSS statistics 18.0 program (SPSS Inc., Chicago, IL, USA).

Ethics statement

This study was approved by the Institutional Review Board (No. SEUMC 2021-07-047) of the Clinical Research Ethics Review Board of Ewha Womans University Seoul Hospital.

RESULTS

The demographical characteristics of each group were demonstrated in **Table 1**. The mean ages of NC (65.62±10.1 years) and MCI (69.72±9.47 years) groups were not significantly different, but they were significantly younger than the DE group (75.21±8.54 years). The sex ratio was uneven only for the NC group, where female was 64.4%, but it was similar for MCI and DE groups. The mean years of education were not significantly different across groups.

The percentages of abnormal performance, based on 1SD below the normative mean, were analyzed on the K-CWST: CR in each group for the 60- and 120-second conditions. In NC group, abnormal performance percentages for 60-second and 120-second conditions were 4.8% and 3.8%, respectively (χ^2=1.00, p=1.000). In MCI group, they were 41.5% and 42.3% for 60-second and 120-second conditions, respectively (χ^2=0.49, p=0.488). In DE group, they were 82.7% and 82.4% for 60-second and 120-second conditions, respectively (χ^2=0.00, p=1.000).

ROC curves in **Fig. 2A-C** illustrated the sensitivity and specificity of each time condition in differentiating groups. The areas under the curve (AUC) values for the 60- and 120-second conditions in differentiating NC from MCI groups were 0.80 (95% confidence interval [CI], 0.76–0.84) and 0.81 (95% CI, 0.77–0.85) (**Fig. 2A**). The AUC values for the 60- and 120-second conditions in NC from DE groups were 0.95 (95% CI, 0.93–0.96) and 0.96 (95% CI, 0.94–0.97) (**Fig. 2B**). The AUC values for the 60- and 120-second conditions in MCI from DE groups were 0.77 (95% CI, 0.74–0.79) and 0.77 (95% CI, 0.74–0.80) (**Fig. 2C**). It suggests that the two-time conditions have nearly the same sensitivity and specificity in differentiating NC, MCI, and DE groups.

Table 1. Demographical characteristics of the participants

Variables	Normal Cognition (n=104)	MCI (n=884)	Dementia (n=348)	F or χ^2	p-value
Age (yr)	65.62±10.12	69.72±9.47	75.21±8.54	F=61.4	<0.001
Female/male (% of female)	67/37 (64.4)	485/399 (54.7)	171/177 (49.9)	χ^2=8.12	0.017
Education (yr)	10.62±3.62	10.11±3.74	10.03±4.10	F=0.96	0.381

Values are presented as mean ± standard deviation and the number (%) of patients.

https://doi.org/10.12779/dnd.2021.20.4.62

DND Dementia and Neurocognitive Disorder

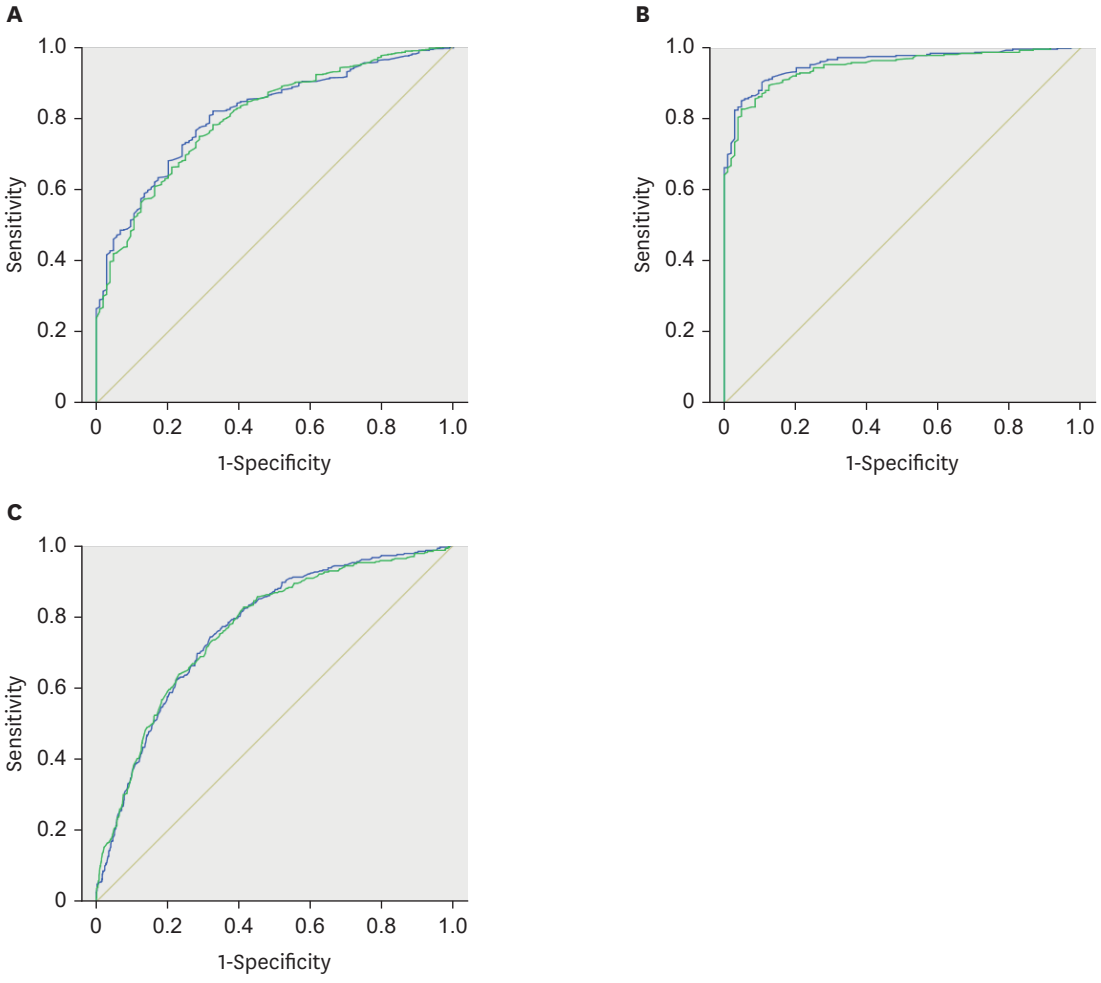

Fig. 2. ROC curves of the 60- and 120-seconds in differentiating each group: (A) normal cognition from MCI, (B) normal cognition from dementia, (C) MCI from dementia.
Represented in z-score, the color of lines (yellow=reference line, green=60-second, blue=120-second).

DISCUSSION

The present study attempted to verify the validity of the 60-second condition for normal elderly, MCI, and mild dementia groups based on retrospective data analysis. Previous studies consistently reported strong effects of age[22-24] and education level[22,25] on the Stroop performance while the effect of sex was controversial.[22,25-27] Therefore, we used the age, education level, and sex-stratified normative data for the K-CWST: CR in the 60- and 120-seconds time conditions provided by the SNSB-II.[14] The 60-second condition of the K-CWST showed very similar results with no statistical difference from the 120-second condition in the percentages of abnormal performance and differentiating among normal, MCI, and mild dementia groups.

In our data, although the numbers are a few, normal cognition group also showed abnormal performances in the K-CWST: CR. The percentage was slightly higher for the 60-second compared to 120-second condition from 4.8% to 3.8%. There is evidence that shorter test duration is better for identifying individuals who have difficulty with the stoop task. Klein and colleagues[28] found that healthy older adults showed greater interference than

https://doi.org/10.12779/dnd.2021.20.4.62

younger adults in the Stroop task with 100 items. They reported this effect was particularly pronounced on the first 40 items than on the last 60 items. The authors interpreted it as reflecting the greater novelty of the task on the initial items that required more effortful processing. Response for the Stroop task could be more automated as the examinee moves towards the latter items. Therefore, it became less sensitive to difficulties with response inhibition as the test proceeded and exposure to the stimuli increased.[23] However, for both MCI and dementia groups in the present study, very similar percentages of abnormal performance were found for the two-time conditions. It may be due to pathological conditions such as a neurodegenerative disease that make it challenging to be automatized for the Stroop task within the 120 second time limit.

Currently used neuropsychological test batteries require at least 1 to 2 hours to complete. Even older adults with normal cognition frequently complain that it is difficult to complete the whole neuropsychological test battery. It would be harder for those with neurological diseases. Patients with increased susceptibility to interference, particularly those with neurodegenerative diseases such as Alzheimer's disease, must invest a considerable amount of effort to complete the Stroop test.[28] This makes testing a problematic and stressful experience and may elicit confounding factors other than interference effect, which is the primary purpose to assess with the Stroop test. Therefore, it is essential to shorten the Stroop test to improve the differentiating power and reduce the total testing time. On the other hand, however, an abridged version with a time limit of 60 seconds may reduce the negative impact of sustained efforts because it takes 50% less time to be administered. We found that both time conditions have almost the same sensitivity and specificity in differentiating normal cognition, MCI, and mild dementia. Therefore, these results suggest that the 60-second condition could replace the 120-second condition regardless of the patient's cognitive ability with the advantage of reducing the examination time and patient's fatigue.

This study had some limitations because it was mainly based on retrospective data. First, all the data was from a single medical center, although the sample size was quite large. Second, the results of this study showed that the 60-second condition could replace the 120-second condition, but it is not known whether 60-second is the most optimal time limit. The optimal time limit should be found through a new prospective study using more diverse time limits, such as 30, 40, and 50-second conditions shorter than 60-second.

In summary, the present study demonstrated the concurrent validity of the 60-second condition and the 120-second condition of the K-CWST: CR in the normal cognition, MCI, and dementia groups. Therefore, we conclude that the 60-second condition could be used in a clinical setting interchangeably with the 120-second condition. However, given that the discriminant power of the two conditions is the same, it is recommended to adopt the 60-second condition since it is beneficial in requiring less administration time and is more efficient in detecting response inhibition difficulties.

REFERENCES

1. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18:643-662.
CROSSREF
2. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997;121:65-94.
PUBMED | CROSSREF

3. Kahneman D, Chajczyk D. Tests of the automaticity of reading: dilution of Stroop effects by color-irrelevant stimuli. J Exp Psychol Hum Percept Perform 1983;9:497-509.
PUBMED | CROSSREF
4. Hasher L, Zacks RT. Working memory, comprehension, and aging: a review and a new view. Psychol Learn Motiv 1988;22:193-225.
CROSSREF
5. Kramer AF, Humphrey DG, Larish JF, Logan GD, Strayer DL. Aging and inhibition: beyond a unitary view of inhibitory processing in attention. Psychol Aging 1994;9:491-512.
PUBMED | CROSSREF
6. Logan GD. Attention and automaticity in Stroop and priming tasks: theory and data. Cognit Psychol 1980;12:523-553.
PUBMED | CROSSREF
7. Bondi MW, Serody AB, Chan AS, Eberson-Shumate SC, Delis DC, Hansen LA, et al. Cognitive and neuropathologic correlates of Stroop Color-Word Test performance in Alzheimer's disease. Neuropsychology 2002;16:335-343.
PUBMED | CROSSREF
8. Binetti G, Magni E, Padovani A, Cappa SF, Bianchetti A, Trabucchi M. Executive dysfunction in early Alzheimer's disease. J Neurol Neurosurg Psychiatry 1996;60:91-93.
PUBMED | CROSSREF
9. Amieva H, Lafont S, Rouch-Leroyer I, Rainville C, Dartigues JF, Orgogozo JM, et al. Evidencing inhibitory deficits in Alzheimer's disease through interference effects and shifting disabilities in the Stroop test. Arch Clin Neuropsychol 2004;19:791-803.
PUBMED | CROSSREF
10. Bayard S, Erkes J, Moroni CCollège des Psychologues Cliniciens spécialisés en Neuropsychologie du Languedoc Roussillon (CPCN Languedoc Roussillon). Victoria Stroop Test: normative data in a sample group of older people and the study of their clinical applications in the assessment of inhibition in Alzheimer's disease. Arch Clin Neuropsychol 2011;26:653-661.
PUBMED | CROSSREF
11. Ahn HJ, Chin J, Park A, Lee BH, Suh MK, Seo SW, et al. Seoul Neuropsychological Screening Battery-dementia version (SNSB-D): a useful tool for assessing and monitoring cognitive impairments in dementia patients. J Korean Med Sci 2010;25:1071-1076.
PUBMED | CROSSREF
12. Park KW, Kim HS, Cheon SM, Cha JK, Kim SH, Kim JW. Dementia with Lewy bodies versus Alzheimer's disease and Parkinson's disease dementia: a comparison of cognitive profiles. J Clin Neurol 2011;7:19-24.
PUBMED | CROSSREF
13. Lee JH, Lee KU, Lee DY, Kim KW, Jhoo JH, Kim JH, et al. Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. J Gerontol B Psychol Sci Soc Sci 2002;57:47-53.
PUBMED | CROSSREF
14. Kang Y, Jahng S, Na D. Seoul Neuropsychological Screening Battery, (SNSB-II). Seoul: Human Brain Research & Consulting Co., 2012.
15. Kang SG, Cho SJ, Ryu SH, Choi SH, Han SH, Shim YS, et al. Normative study of the literacy independent cognitive assessment in illiterate and literate elderly Koreans. Psychiatry Investig 2015;12:305-315.
PUBMED | CROSSREF
16. Health Insurance Review and Assessment Service. Medical treatment (examination/surgery, etc.) statistics [Internet]. Wonju: Health Insurance Review and Assessment Service; 2020 [cited 2021 Aug 16]. Available from: http://opendata.hira.or.kr/op/opc/olapDiagBhvInfo.do.
17. Lee JH, Kang Y, Na D. Efficiencies of Stroop Interference Indexes in Healthy Older Adults and Dementia Patients. Korean J Clin Psychol 2000;19:807-818.
18. Staub B, Doignon-Camus N, Després O, Bonnefond A. Sustained attention in the elderly: what do we know and what does it tell us about cognitive aging? Ageing Res Rev 2013;12:459-468.
PUBMED | CROSSREF
19. Fisher LM, Freed DM, Corkin S. Stroop Color-Word Test performance in patients with Alzheimer's disease. J Clin Exp Neuropsychol 1990;12:745-758.
PUBMED | CROSSREF
20. Chin J, Park J, Yang SJ, Yeom J, Ahn Y, Baek MJ, et al. Re-standardization of the Korean-Instrumental Activities of Daily Living (K-IADL): clinical usefulness for various neurodegenerative diseases. Dement Neurocogn Disord 2018;17:11-22.
PUBMED | CROSSREF

21. O'Bryant SE, Waring SC, Cullum CM, Hall J, Lacritz L, Massman PJ, et al. Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. Arch Neurol 2008;65:1091-1095.
PUBMED | CROSSREF

22. Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. Assessment 2006;13:62-79.
PUBMED | CROSSREF

23. Troyer AK, Leach L, Strauss E. Aging and response inhibition: Normative data for the Victoria Stroop Test. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2006;13:20-35.
PUBMED | CROSSREF

24. Uttl B, Graf P. Color-Word Stroop test performance across the adult life span. J Clin Exp Neuropsychol 1997;19:405-420.
PUBMED | CROSSREF

25. Moering RG, Schinka JA, Mortimer JA, Graves AB. Normative data for elderly African Americans for the Stroop Color and Word Test. Arch Clin Neuropsychol 2004;19:61-71.
PUBMED | CROSSREF

26. Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC. Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. Clin Neuropsychol 1996;10:262-278.
CROSSREF

27. Lucas JA, Ivnik RJ, Smith GE, Ferman TJ, Willis FB, Petersen RC, et al. Mayo's older African Americans normative studies: norms for Boston naming test, controlled oral word association, category fluency, animal naming, token test, wrat-3 reading, trail making test, Stroop test, and judgment of line orientation. Clin Neuropsychol 2005;19:243-269.
PUBMED | CROSSREF

28. Klein M, Ponds RW, Houx PJ, Jolles J. Effect of test duration on age-related differences in Stroop interference. J Clin Exp Neuropsychol 1997;19:77-82.
PUBMED | CROSSREF