The World Alzheimer Report estimated that 46.8 million people worldwide are living with dementia and projected that the number would increase to 131.5 million by 2050. Although Alzheimer’s disease (AD) is a disorder with significant unmet needs, improvements in the prevention and treatment of dementia have been limited. The lack of success in the development of effective treatments for dementia is an ongoing public health challenge. Because of screening failures, the pharmaceutical industry is disinvesting. To find new approaches that would enhance pre-screening to reduce clinical trial failure rates, global efforts to gather big data are ongoing. In Korea, clinical registries for dementia research have also been developed. However, major deficiencies in regards to existing dementia registries have limited possibilities of data sharing between existing data collection systems (e.g., interoperability) and lack of available data on the costs of operating dementia registries and their cost-eff-
fectiveness.8
Since aggregating data into larger pools is essential to obtain effective data, there have been global attempts to consolidate data from different cohorts.9-12 Currently, however, only a few platforms support the sharing of measurements and derived data, and only a few services provide a combined preprocessed dataset at each variable level after performing data cleansing.13 One key challenge to combining individual data is that the protocols and methods used in each study are different. For this reason, integrating different data is a difficult process.14 Therefore, the aim of this study was a pilot project to evaluate the feasibility of building an integrate dementia platform for converging pre-exist dementia cohorts from individual variable levels.
Eligible cohorts satisfied the following conditions: 1) dementia cohorts built with national funding; 2) prospective cohorts; and 3) multicenter cohorts. After experts reviewed the potential for integrating a cohort, we contacted data owners to request access to their data and the sharing of the data to build a platform. The following four cohorts were identified as potentially useful cohorts to conduct this pilot study: 1) Clinical Research Center for Dementia of South Korea (CREDOS) (identifier: NCT01198093), 2) Korean Brain Aging Study for Early Diagnosis and Prediction of Alzheimer's disease (K-BASE),15 3) Environmental Pollution-induced Neurological Effects (EPINEF) study,16 and 4) a prospective registry in Dementia Platform Korea project (DPKR) (identifier: KCT0005516) (Table 1). After obtaining approval for data sharing, we received the baseline data and variable catalogues from each cohort. The Institutional Review Board (IRB) of Samsung Medical Center approved this study (approval number: IRB 2018-07-016) and waived the requirement for informed consent as only de-identified data were used in this study.

In our study, we selected important domains in dementia based on the Korea National Health and Nutrition Survey (KNHANES).17 The domains included health surveys (KNHANES).

Table 1. Characteristics of the Included Cohorts

| Characteristics | CREDOS (n=18240) | K-BASE-VI (n=385) | EPINEF (n=200) | DPKR (n=355) |
|-----------------|-----------------|----------------|---------------|-------------|
| Recruitment period | 2005–2015 | 2015–2019 | 2014–2019 | 2018–2020 |
| Number of hospitals* | 59 | 9 | 7 | 13 |
| Cognitive status | | | | |
| Normal | 2069 (11.3) | 173 (44.9) | 200 (100) | 71 (20) |
| MCI | 6127 (33.6) | 88 (22.9) | 0 (0) | 134 (37.7) |
| Dementia | 7512 (41.2) | 75 (19.5) | 0 (0) | 89 (25.1) |
| Unknown | 2532 (13.9) | 49 (12.7) | 0 (0) | 61 (17.2) |
| Age at baseline (yr) | 71.7±8.9 | 71.1±8.7 | 67.9±6.7 | 71.2±8.8 |
| Sex | | | | |
| Male | 6047 (33.2) | 132 (34.3) | 103 (51.0) | 119 (33.5) |
| Female | 12192 (66.8) | 207 (53.8) | 97 (48.0) | 190 (53.5) |
| Unknown | 1 (0) | 46 (11.9) | 0 (0) | 46 (13.0) |

CREDOS, Clinical Research Center for Dementia of South Korea; MCI, Mild Cognitive Impairment; DPKR, a prospective registry in Dementia Platform Korea project; EPINEF, Environmental Pollution-induced Neurological Effects; K-BASE, the Korean Brain Aging Study for the Early Diagnosis and Prediction of Alzheimer’s disease. Values are presented as a n (%) or mean±SD. *Participated in the cohort.
tionnaires), neuropsychological tests, and physical examinations (laboratory, imaging, and other tests). Among several variables, neuroimaging could not be integrated due to differences in file storage and transfer format. Since imaging and blood tests involve different methods of standardization, these variables were not included in this study. Sociodemographic characteristics, health behavior, comorbidities, family history, clinical assessment of QoL, and mental health, and neuropsychological tests were available (Table 2).

The overall process of data integration involved eight steps (Fig. 1): 1) setting up the rules for data cleaning, 2) preparing a data integration plan, 3) cleaning each dataset, 4) generating derived variables for the integrated dataset, 5) appending cleaned variables, 6) checking errors using the integration database, 7) generating a codebook and dashboard, and 8) uploading the integrated cohort for storage.

1) Set up and rules for data cleaning: all processes and rules for decision making during data cleaning are documented in Stata 14.0 (Stata Corp, College Station, TX, USA).

2) Preparing a plan for data integration: to integrate the cohorts, we identified variables that contained similar information from each study. Each variable was then extracted from raw data tables of each study and assigned to the correct column, placing the data for each subsequent study sequentially in the same column.

3) Cleaning each dataset: we reviewed the descriptive analysis results (distribution, frequency of each category) for all variables in the dataset to find errors. To identify logical errors, we confirmed variables based on instances of a potential hierarchical relationship (e.g., smoking status – amount of smoke).

4) Generating derived variables for the integrated cohort: to ensure that each cohort contained the same information that could be analyzed together, several processes were carried out to integrate and harmonize the data, including the following: (i) transforming each dataset to the same database programs (e.g., csv, dta); (ii) formatting heterogeneity variables to the same format (e.g., date: from dd-mm-yy to yyyy-mm-dd and gender from M/F to 1/2); (iii) evaluating syntactical heterogeneity (the meaning of the data captured is the same across sources, but words used to capture the information are different between different datasets); (iv) determining content heterogeneity (capture), wherein a whole variable is captured in one study, but not in another; (v) determining response heterogeneity (level of granularity), wherein some datasets had more response options than others in the same questionnaires. We generated derived variables for the integrated cohort. These data were then harmonized and cleaned further. For example, literacy was asked in five categories in CREDOS and three categories in K-BASE-VI, resulting in three categories of derived variables. In addition, when one cohort included categorical variables while another had continuous variables, we created a categorical variable from a continuous variable to combine the variable. Variables included in two or more cohorts were created as derived variables.

| Table 2. Collected Variables by Cohort |
|---------------------------------------|
| Variables                             | Cohort |
|                                       | CREDOS | K-BASE-VI | EPINEF | DPKR |
| Sociodemographic                      |        |           |       |      |
| Age (or birth year)                   | Yes    | Yes       | Yes   | Yes  |
| Sex                                   | Yes    | Yes       | Yes   | Yes  |
| Housing type                          | Yes    | Yes       | Yes   | Yes  |
| Education                             | Yes    | Yes       | Yes   | Yes  |
| Literacy                              | Yes    | Yes       | Yes   | Yes  |
| Job                                   | Yes    | Yes       | Yes   | Yes  |
| Married                               | No     | Yes       | Yes   | Yes  |
| Health behavior                       |        |           |       |      |
| Smoking                               | Yes    | Yes       | Yes   | Yes  |
| Alcohol                               | Yes    | Yes       | Yes   | Yes  |
| Physical activity                     | Yes    | Yes       | Yes   | Yes  |
| Comorbidity                           |        |           |       |      |
| Hypertension                          | Yes    | Yes       | Yes   | Yes  |
| Diabetes                              | Yes    | Yes       | Yes   | Yes  |
| Hyperlipidemia                        | Yes    | Yes       | Yes   | Yes  |
| Stroke                                | Yes    | Yes       | Yes   | Yes  |
| Heart disease                         | Yes    | Yes       | Yes   | Yes  |
| Cancer                                | Yes    | Yes       | Yes   | Yes  |
| Depression                            | Yes    | Yes       | Yes   | Yes  |
| Lung                                  | Yes    | Yes       | No    | Yes  |
| Family history                        |        |           |       |      |
| Dementia                              | Yes    | Yes       | Yes   | Yes  |
| Stroke                                | Yes    | Yes       | Yes   | Yes  |
| Parkinson                             | Yes    | Yes       | Yes   | No   |
| Clinical assessment                   |        |           |       |      |
| Depression                            | Yes    | Yes       | No    | Yes  |
| GDS-15                                | No     | No        | Yes   | No   |
| GDS-30                                | No     | Yes       | No    | No   |
| Anxiety                               |        |           |       |      |
| BAI                                   | No     | No        | No    | Yes  |
| Stress                                |        |           |       |      |
| KNHANES: short form                   | No     | No        | No    | Yes  |
| Nutrition examination                 |        |           |       |      |
| MDAI                                  | No     | No        | No    | No   |
| MNA                                   | Yes    | No        | No    | Yes  |
| SNAQ                                  | No     | Yes       | No    | No   |
| EBS                                   | No     | Yes       | No    | No   |
| Sleep                                 |        |           |       |      |
| PSQI                                  | No     | Yes       | No    | Yes  |
| SSS                                   | No     | Yes       | No    | No   |
| ESS                                   | No     | Yes       | No    | No   |
| Quality of Life                       |        |           |       |      |
| SF-36                                 | No     | No        | No    | Yes  |
| Neuropsychological tests              |        |           |       |      |
| Cognitive screening questionnaires    |        |           |       |      |
| KDSQ                                  | Yes    | No        | No    | No   |
| SMCQ                                  | No     | Yes       | No    | No   |
| KAD8                                  | No     | No        | No    | Yes  |
5) Appending cleaned variables: if the same variables were integrated, then that variable was appended.

6) Checking data integration errors: after appending the cleaned data, we checked for integration errors. If the same variable was not integrated in each cohort, the coding value was to remain missing.

7) Generating a codebook: we generated a codebook to describe the contents, structure, and layout of the collected data. The codebook provided information on variable name, variable label, question text, value, value label, summary statistics, and missing data. For summary statistics, depending on the type of variable, unweighted summary statistics were provided for quick reference. For categorical variables, for instance, frequency counts showing the number of times a value occurred and the percentage of cases that the value represented for the variable were appropriate. For continuous variables, minimum, maximum, and median values were relevant.

8) Uploading the integrated cohort for storage: the integrated data were upload to the dementia databank web portal. Researchers can access the data after going through a certification process. Application for access can be made through the Data Portal: http://dementiasplatform.kr/.

A total of 29916 patients were included in the platform with 348 integrated variables. On average, each variable had missing information on 16.8% of the data. Among these 348 variables,
marital status (94.3%), cohabitation (85.7%), and family history of Parkinson (96.7%) had missing rates higher than 80%. Missing rates were 50% to 80% for smoking status (59.2%), drinking status (58.4%), vigorous physical activity (75.5%), moderate physical activity (71.3%), walking (67.9%), family history of dementia (61.3%), family history of stroke (61.3%), Pittsburgh Sleep Quality Index (61.2%), mini nutritional assessment (55.9%), and mini-dietary assessment index (62.3%). On the other hand, age (4.5%), sex (1.1%), education (3.7%), and neuropsychological tests (0.9%) had missing rates less than 5% (Table 3).

Among participants, 13.9% (n=4156), 31.5% (n=9412), and 44.2% (n=13227) of patients had normal cognition, mild cognitive impairment, and dementia, respectively (Table 3). The mean age was 72.4 years. Females accounted for 65.7%. Those with college or higher education and those without problems in reading or writing accounted for 12.3% and 46.8%, respectively.

| Status of dementia | No. of available | Values |
|--------------------|------------------|--------|
| Normal             | 4156 (13.9)      |        |
| Mild cognitive impairment | 9412 (31.5) |        |
| Dementia           | 13227 (44.2)     |        |
| Unknown            | 3121 (10.4)      |        |

| Age (yr) | 28578 | 72.4 ± 8.7 |
| Sex      | 29593 |
| Male     | 10138 (34.3) |
| Female   | 19455 (65.7) |

| Education level | No. of available | Values |
|-----------------|------------------|--------|
| None            | 7789 (29.3)      |        |
| Elementary school | 7280 (27.4) |        |
| Middle school   | 2919 (11.0)      |        |
| High school     | 4927 (18.5)      |        |
| College or higher | 3668 (13.8) |        |

| Literacy | No. of available | Values |
|----------|------------------|--------|
| None     | 1035 (4.1)       |        |
| Problem reading or writing | 10522 (41.2) |        |
| No problem | 13984 (54.7) |        |

| Married* | No. of available | Values |
|----------|------------------|--------|
| Single   | 2 (0.1)          |        |
| Married  | 1288 (75.5)      |        |
| Divorce or separated | 66 (3.9) |        |
| Bereaved | 328 (19.2)       |        |
| Other    | 23 (1.3)         |        |

| Cohabitation | No. of available | Values |
|--------------|------------------|--------|
| Living alone | 659 (15.4)       |        |
| Only spouse  | 1889 (44.2)      |        |
| Spouse and other family | 666 (15.3) |        |
| Family without spouse | 787 (18.4) |        |
| Other        | 287 (6.7)        |        |

| Smoking worker (yes) | No. of available | Values |
|----------------------|------------------|--------|
| Never smoker         | 8821 (72.2)      |        |
| Ex-smoker            | 2554 (20.9)      |        |
| Current              | 836 (6.9)        |        |

| Alcohol status | No. of available | Values |
|---------------|------------------|--------|
| Never drinker | 7228 (58.1)      |        |
| Ex-drinker    | 2231 (17.9)      |        |
| Current drinker | 2978 (24.0) |        |

| Physical activity | No. of available | Values |
|-------------------|------------------|--------|
| Vigorous          | 6141             | 1141 (18.6) |
| Moderate          | 8871             | 4138 (46.6) |
| Walking           | 9903             | 6150 (62.1) |

| Comorbidity*      | No. of available | Values |
|-------------------|------------------|--------|
| Hypertension      | 27627            | 13978 (50.6) |
| Diabetes          | 27616            | 5891 (21.3) |
| Hyperlipidemia    | 27594            | 5220 (18.9) |
| Stroke            | 27254            | 2368 (8.0) |
| Heart disease     | 27606            | 4010 (14.5) |
| Cancer            | 27562            | 1726 (6.3) |
| Depression        | 27590            | 4187 (15.2) |

| Variables         | No. of available | Values |
|-------------------|------------------|--------|
| Family history*   |                  |        |
| Dementia          | 11884            | 2629 (22.1) |
| Stroke            | 11793            | 2410 (20.44) |
| Parkinson         | 984              | 23 (2.3) |
| CGA-NPI           | 3258             | 9.1 ± 9.8 |
| Pittsburgh Sleep Quality Index | 1060 | 4.6 ± 3.0 |
| Nutritional assessment | 1201 | 6.3 ± 1.0 |
| Mini Dietary Assessment Index | 10517 | 36.2 ± 5.8 |
| Geriatric Depression Scale | 25184 |        |
| Mild              | 12726 (50.5)     |        |
| Moderate          | 7173 (28.5)      |        |
| Severe            | 5285 (21.0)      |        |

| Neuropsychological tests | No. of available | Values |
|--------------------------|------------------|--------|
| Cognitive screening questionnaires | 25820 |        |
| Cognitive unimpaired      | 19711 (76.3)     |        |
| Cognitive impaired        | 6109 (23.7)      |        |
| Mini-mental State Examination, <20 | 28144 | 15771 (56.0) |
| Boston Naming Test, <1SD  | 16972            | 7362 (43.4) |
| Figure copy, <1SD         | 21205            | 8781 (41.4) |
| Verbal delayed recall, <1SD | 22256 | 14073 (63.2) |
| Visual delayed recall, <1SD | 20266 | 12081 (63.2) |
| Animal fluency, <1SD      | 21942            | 13245 (60.4) |
| Stroop Test, <1SD         | 19023            | 10080 (53.0) |

| Clinical Dementia Ratings | No. of available | Values |
|---------------------------|------------------|--------|
| None                      | 2142 (7.8)       |        |
| Questionable              | 15484 (56.4)     |        |
| Mild                      | 6694 (24.4)      |        |
| Moderate                  | 2573 (9.4)       |        |
| Severe                    | 557 (2.0)        |        |
| Profound                  | 11 (0.0)         |        |
| Terminal                  | 5 (0.0)          |        |

CGA-NPI, Caregiver-Administered Neuropsychiatric Inventory; SD, standard deviation. Values are presented as a n (%) or mean ± SD. *Mutually not.
We established a dementia platform databank by integrating pre-existing dementia cohorts in Korea. In the dementia area, other data platforms are also available. The most popular platforms are the Dementias Platform UK (DPUK) Data Portal,\(^6\) the EU Joint Programme for Neurodegenerative Disease Research (JPND) Global Cohort Directory,\(^3\) the Integrative Analysis of Longitudinal Studies of Aging and Dementia (IALSA) Network,\(^11\) and the Global Alzheimer’s Association Interactive Network (GAIN).\(^12\) DPUK included 35 cohorts. Of these cohorts, 22 (n=1399082) have uploaded full or partial datasets, and 13 (n=2062162) will upload on a per project basis.\(^9\) The JPND Global Cohort Directory (http://www.neurodegenerationresearch.eu/jpnd-global-cohort-portal/) provides contact details for 175 cohorts (n=3586109), whilst the IALSA Network (http://www.ialsa.org/) provides details for 110 cohorts (n=1485410). More sophisticated and convenient data discovery tools are provided by GAIN with 47 cohorts (n=480020). GAIN also offers centralized processing for selected datasets.\(^3\) EMIF-AD offers a comprehensive data harmonization program for a selection of their 60 catalogued cohorts (n=139595) and 18 electronic health records datasets (n=6500000).\(^31\) Our pilot platform sought to integrate data from each variable level in pre-existing dementia cohorts, and we found that integration was difficult if each cohort had difference measurements. In the UK, the ROADMAP project supported by the Medical Research Council has attempted an approach similar to ours to optimize evidence of Alzheimer’s disease based on data integration.\(^32\) Data Cube, an integrated data platform, includes information on clinical diagnosis; disease severity and progression; cognitive and functional ability; independence; behavioral and neuropsychiatric symptoms; medical investigations; healthcare and social services utilization; therapeutic treatment; disease-related life events; QoL for the patient, caregiver, and family members; mortality; and comorbidities.

Data Cube suggests combining domains from different data sources for use in research studies. However, even though individual cohorts have the same domain and all of them are of high quality, we found there were several barriers to integrating individual cohorts. First is missing values due to the variability of study variables across cohorts. Among all variables, anxiety, stress, nutrition, sleep, and QoL were not collected from some cohorts. Thus, these variables could be used only for a limited dataset. Second, even a domain may be available, sometimes it collected from different measurements, and this could lead to pre-analytical variability. For example, physical activity and neuropsychological tests were measured using different questionnaires across cohorts. Thus, they could not be appended. Recently, an item response theory has been used to generate the same scores from a completely different test built to evaluate the same construct, assuming that a respondent has similar latent traits.\(^33\) Once we use these types of methods, it will be easier to integrate data across cohorts.\(^34\) Third, we were uncertain of the accuracy of the information obtained from pre-existing datasets.\(^35-38\)

Although many researchers are trying to combine pre-existing cohorts, the process of integrating past data has not proven easy. Therefore, researchers should consider their choice of data elements and strive for quality assurance guided by reliability and validity, in addition to achieving the study purpose. Also, to aid in data integration, researchers should establish a protocol with considerations at the cohort establishment stage for the ability of the data to be integrated in future applications.

**ACKNOWLEDGEMENTS**

This research was supported by funding (2018-ER6203-02) from the Research of Korea Centers for Disease Control and Prevention.

**AUTHOR CONTRIBUTIONS**

Conceptualization: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, Sang Won Seo, Juhee Cho, and Danbee Kang. Data curation: Bo Kyoung Cheon, Min Jung Hahn, Sang Won Seo, Duk I. Na, Jaelim Cho, and Seong Hye Choi. Formal analysis: Minwoong Kang, Bo Kyoung Cheon, and Min Jung Hahn. Funding acquisition: Sang Won Seo. Investigation: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, and Danbee Kang. Methodology: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, and Danbee Kang. Project administration: Sang Won Seo. Resources: Sang Won Seo, Duk I. Na, Jaelim Cho, and Seong Hye Choi. Software: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, and Danbee Kang. Supervision: Danbee Kang. Validation: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, and Danbee Kang. Visualization: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, and Danbee Kang. Writing—original draft: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, and Danbee Kang. Writing—review & editing: Minwoong Kang, Bo Kyoung Cheon, Min Jung Hahn, Sang Won Seo, Juhee Cho, Soo Yong Shin, and Danbee Kang. Approval of final manuscript: all authors.

**ORCID iDs**

Minwoong Kang https://orcid.org/0000-0002-5162-3641
Bo Kyoung Cheon https://orcid.org/0000-0002-5613-517X
Min Jung Hahn https://orcid.org/0000-0001-7688-8198
Sang Won Seo https://orcid.org/0000-0003-2568-0797
Juhee Cho https://orcid.org/0000-0001-9081-0266
Soo Yong Shin https://orcid.org/0000-0002-4100-6120
Duk I. Na https://orcid.org/0000-0002-0998-7592
Jaelim Cho https://orcid.org/0000-0002-4524-0310
Seong Hye Choi https://orcid.org/0000-0002-4180-8626
Danbee Kang https://orcid.org/0000-0003-0244-7714

**REFERENCES**

1. Prince M, Comas-Herrera A, Knapp M, Guerchet M, Karagianni-dou M. World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future. London: Alzheimer’s Disease International; 2016. p.110-115.
2. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer’s..."
disease drug development pipeline: 2019. Alzheimer’s Dement (N Y) 2019;5:272-93.

3. Marsden G, Mestre-Ferrandiz J. Dementia: the R&D landscape [Internet]. London: The Office of Health Economics; 2015 [accessed on 2020 July 28]. p.12-6. Available at: https://www.ohe.org/publications/dementia-rd-landscape.

4. Hsuain M. Big data: could it ever cure Alzheimer’s disease? Brain 2014;137:2623-4.

5. Saykin AJ, Shen L, Yao X, Kim S, Nho K, Risacher SL, et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. Alzheimers Dement 2015;11:792-814.

6. Ferguson AR, Nielson JL, Cragnin MH, Bandrowski AE, Martone ME. Big data from small data: data-sharing in the ‘long tail’ of neuroscience. Nat Neurosci 2014;17:1442-7.

7. Poldrack RA, Gorgolewski KJ. Making big data open: data sharing in neuroimaging. Nat Neurosci 2014;17:1510-7.

8. Krysinska K, Sachdev PS, Breitner J, Kivipelto M, Kukul W, Brudat H. Dementia registries around the globe and their applications: a systematic review. Alzheimers Dement 2017;13:1031-47.

9. Bauermeister S, Orton C, Thompson S, Barker RA, Bauermeister JR, Ben-Shlomo Y, et al. The dementias platform UK (DPUK) data portal. Eur J Epidemiol 2020;35:601-11.

10. Lerche S, Liepelt-Scarfone I, Alves G, Barone P, Behnke S, Ben-Shlomo Y, et al. Methods in neuroepidemiology characterization of European longitudinal cohort studies in Parkinson’s disease-report of the JPND working group BioLoC-PD. Neuroepidemiology 2015;45:282-97.

11. Kaye J, Hofer SM. Integrative analysis of longitudinal studies on aging and dementia (IALSA). Innov Aging 2017;1:1275.

12. Toga AW, Neu SC, Bhatt P, Crawford KL, Ashish N. The global Alzheimer’s association interactive network. Alzheimers Dement 2016;12:49-54.

13. Neu SC, Crawford KL, Toga AW. Sharing data in the global Alzheimer’s association interactive network. Neuroimage 2016;124:1168-74.

14. Doan A, Halevy A, Ives Z. Principles of data integration. Burlington, MA: Morgan Kaufmann; 2012.

15. Hwang J, Jeong JH, Yoon SJ, Park KW, Kim EJ, Yoon B, et al. Clinical and biomarker characteristics according to clinical spectrum of Alzheimer’s disease (AD) in the validation cohort of Korean brain aging study for the early diagnosis and prediction of AD. J Clin Med 2019;8:341.

16. Cho J, Sohn J, Noh J, Jang H, Kim W, Cho SK, et al. Association between exposure to polyyclic aromatic hydrocarbons and brain cortical thinning: the environmental pollution-induced neurological effects (EPINEF) study. Sci Total Environ 2020;737:140097.

17. Korea Disease Control and Prevention Agency. KNHANES VI [Internet]. Cheongju: Korea Disease Control and Prevention Agency; 2015 [accessed on 2015 February 24]. p.167-218. Available at: https://knhanes.kdcda.go.kr/knhanes/sub03/sub03_02_05.do.

18. Kang SJ, Choi SH, Lee BH, Jeong Y, Hahn DS, Han JW, et al. Caregiver-administered neuropsychiatric inventory (CGA-NPI). J Geriatr Psychiatry Neurol 2004;17:32-5.

19. Yang DW, Cho BL, Chey JY, Kim SY, Kim BS. The development and validation of Korean dementia screening questionnaire (KDSQ). J Korean Neurol Assoc 2002;20:135-41.

20. Kang Y, Na D, Hahn S. Seoul neuropsychological screening battery. Incheon: Human Brain Research & Consulting Co; 2003.

21. Oh JY, Yang YJ, Kim BS, Kang JH. Validity and reliability of Korean version of international physical activity questionnaire (IPAQ) short form. J Korean Acad Fam Med 2007;26:532-41.