

Cohort-Specific Optimization of Models Predicting Preclinical Alzheimer's Disease, to Enhance Screening Performance in the Middle of Preclinical Alzheimer's Disease Clinical Studies

K. Sato^{1,2}, T. Mano², R. Ihara³, K. Suzuki⁴, Y. Niimi⁵, T. Toda², T. Iwatsubo¹, A. Iwata³, for Alzheimer's disease Neuroimaging Initiative, Japanese Alzheimer's disease Neuroimaging Initiative, and The A4 Study Team

1. Department of Neuropathology, Graduate School of Medicine, The University of Tokyo, Japan; 2. Department of Neurology, The University of Tokyo Hospital, Japan; 3. Department of Neurology, Tokyo Metropolitan Geriatric Medical Center Hospital, Japan; 4. Division of Neurology, Internal Medicine, National Defense Medical College, Japan; 5. Unit for Early and Exploratory Clinical Development, The University of Tokyo Hospital, Japan

Corresponding Author: Dr. Atsushi Iwata, Department of Neurology, Tokyo Metropolitan Geriatric Medical Center Hospital, 35-2 Sakaecho Itabashi-ku, Tokyo 173-0015, Japan, Phone: 81-3-3964-1141, FAX: 81-3-3964-2963, E-mails: iwata@m.u-tokyo.ac.jp

Abstract

BACKGROUND: Models that can predict brain amyloid beta ($A\beta$) status more accurately have been desired to identify participants for clinical trials of preclinical Alzheimer's disease (AD). However, potential heterogeneity between different cohorts and the limited cohort size have been the reasons preventing the development of reliable models applicable to the Asian population, including Japan.

OBJECTIVES: We aim to propose a novel approach to predict preclinical AD while overcoming these constraints, by building models specifically optimized for ADNI or for J-ADNI, based on the larger samples from A4 study data.

DESIGN & PARTICIPANTS: This is a retrospective study including cognitive normal participants (CDR-global = 0) from A4 study, Alzheimer Disease Neuroimaging Initiative (ADNI), and Japanese-ADNI (J-ADNI) cohorts.

MEASUREMENTS: The model is made up of age, sex, education years, history of AD, Clinical Dementia Rating-Sum of Boxes, Preclinical Alzheimer Cognitive Composite score, and APOE genotype, to predict the degree of amyloid accumulation in amyloid PET as Standardized Uptake Value ratio (SUVR). The model was at first built based on A4 data, and we can choose at which SUVR threshold configuration the A4-based model may achieve the best performance area under the curve (AUC) when applied to the random-split half ADNI or J-ADNI subset. We then evaluated whether the selected model may also achieve better performance in the remaining ADNI or J-ADNI subsets.

Result: When compared to the results without optimization, this procedure showed efficacy of AUC improvement of up to approximately 0.10 when applied to the models "without APOE;" the degree of AUC improvement was larger in the ADNI cohort than in the J-ADNI cohort.

CONCLUSIONS: The obtained AUC had improved mildly when compared to the AUC in case of literature-based predetermined SUVR threshold configuration. This means our procedure allowed us to predict preclinical AD among ADNI or J-ADNI second-half samples with slightly better predictive performance. Our optimizing method may be practically useful in the middle of the ongoing clinical study of preclinical AD, as a screening to further increase the prior probability of preclinical AD before amyloid testing.

Key words: Amyloid beta, preclinical Alzheimer's disease, machine learning, predictive model.

Introduction

Preclinical Alzheimer's disease (AD), which corresponds to positive brain amyloid beta ($A\beta$) accumulation in healthy individuals without an evidence of cognitive decline (1-3), is getting focused as the target of clinical trials aiming to develop disease-modifying therapies for AD (4). Positive amyloid accumulation on amyloid positron emission tomography (PET) or lowered levels of $A\beta$ 42 in the cerebrospinal fluid (CSF) are used as the gold standard to include participants into clinical trials for preclinical AD (1).

It is estimated that approximately one-third of cognitive normal elderly individuals have positive $A\beta$ (5), which means that if randomly selected, it is necessary to screen 3 times more clinically eligible participants by PET amyloid imaging or CSF lumbar puncture to determine if they are actually amyloid positive or not. Indeed, in the A4 study in which 1,000 participants were included to conduct a double-blinded randomized clinical trial of solanezumab versus a placebo (6), more than 10,000 clinically normal individuals were initially screened, and then the eligible 3,300 participants were further screened by PET amyloid imaging.

If we have some predictive index that can increase the prior probability for the positive $A\beta$ accumulation, the above cost/labor-consuming screening processes could become more efficient with a smaller number of participants requiring PET screening (7, 8). For example, an earlier study reported predicting $A\beta$ of cognitive normal participants from an Alzheimer's disease Neuroimaging Initiative (ADNI) cohort (9) used demographic features of age, sex, education, APOE ϵ 4 status, and cognitive scores, increasing the positive predictive value to 0.65 compared to the reference prevalence of 0.41 (7).

Meanwhile, in case of a Japanese cohort such as the Japanese Alzheimer's disease Neuroimaging Initiative (J-ADNI) (10-12) cohort, there is a concern in deriving similar predictive models from this cohort due to the

limited number of eligible cognitive normal participants. There are fewer than 100 participants included without lack of the necessary data in the J-ADNI (10, 12), so it is considered difficult to construct statistically robust models trained and validated within the Japanese cohort alone to date.

On the other hand, it might be also unsatisfying to apply the models derived from the external population out of the Japanese cohort directly, due to the potential heterogeneity of study participants among different cohorts. In other words, models derived from Anti-Amyloid treatment in Asymptomatic Alzheimer's disease (A4), ADNI, or Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL) cohort data (13) might not always be applicable to the J-ADNI cohort as they are, since the variable importance of each feature in the model can differ depending on the cohort, due to the difference in the distribution of participants' basic demographics. For example, baseline age and education or even the proportion of those with positive $A\beta$ are shown to be significantly different between ADNI and J-ADNI cohorts (10). These problems might have prevented the development of clinical models that effectively predict preclinical AD in a Japanese cohort.

As one of the solutions to overcome these constraints, here we propose to utilize models trained based on the A4 cohort data, which is a large dataset with more than 3,000 participants as of late 2019. Since the data characteristics of A4 participants and Japanese cohort (i.e. J-ADNI here) participants could somehow differ as we mentioned above, we optimized the A4-based models, thereby making the models more suitable to the J-ADNI cohort. Our proposing procedure is composed of two stages: the first is to generate numerous patterns of prediction models based on the A4 data with the varying standard uptake value ratio (SUVr) thresholds, and the second is to find the most appropriate SUVr threshold configuration among them so that the model based on the SUVr configuration would perform best in the random-half of J-ADNI (or ADNI) dataset. The SUVr threshold is the critical cut-off to determine if there is amyloid accumulation in the PET or not (14) but is not always strictly established in the A4 study cohort, so adjusting the SUVr threshold leads to the varied allocation of amyloid positive/negative binary status in each case of the original A4 data. This is the operational procedure made solely for the purpose of identifying the best-performing models for other cohort data, and then we evaluate whether the obtained model based on the determined SUVr threshold can also take the better performance in the remaining J-ADNI (or ADNI) subset. Such 'optimization' procedure might allow us to build more flexible models, thereby enhancing the applicability of the obtained models to any external cohorts such as J-ADNI or ADNI. Practically, our proposed method might be useful as a predictive index available in the actual clinical study settings for preclinical AD, e.g., as a

screening to increase the prior probability of preclinical AD just in the middle of ongoing preclinical AD studies.

Methods

Data acquisition and preprocessing

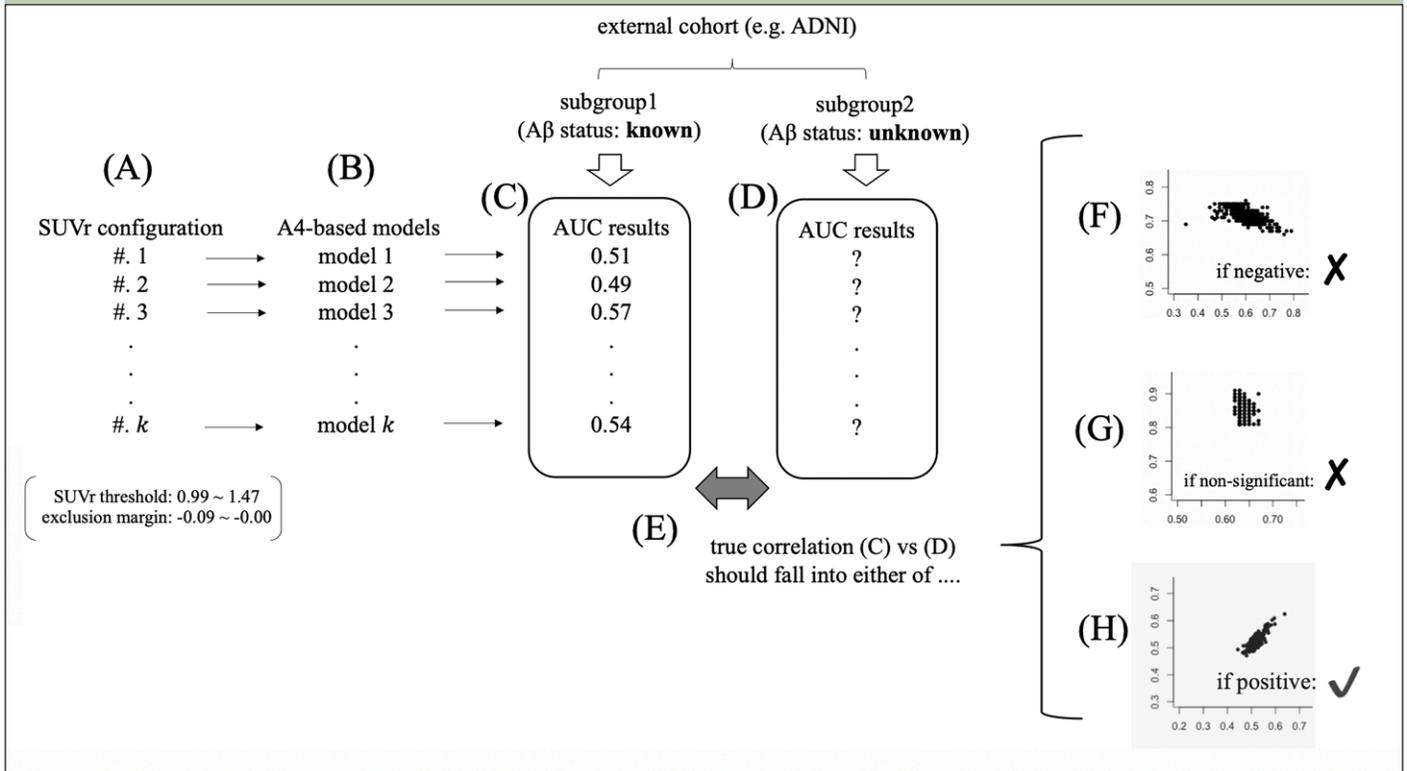
This study was approved by the University of Tokyo Graduate School of Medicine institutional ethics committee (ID: 11628-(3)). Informed consent is not required because this was observational study using publicly available data. We used the datasets of the A4 study and ADNI obtained from the Laboratory of Neuro Imaging (LONI) (<https://ida.loni.usc.edu>) in October 2019 and the J-ADNI dataset obtained from National Bioscience Database Center (NBDC) (<https://humandbs.biosciencedbc.jp/en/hum0043-v1>) in June 2018 with the approval of the data access committee. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org.

In this study, we used the data of cognitive normal participants. General inclusion criteria for the cognitive normal participants were determined in reference to an earlier study on the preclinical Alzheimer cognitive composite (PACC) (3), defined as follows: participants ages 65 to 85 years old (* 60 to 84 years old for cases from the J-ADNI cohort) at the time of screening with a global Clinical Dementia Rating (CDR-global) score of 0, with MMSE score (27-30) and Delayed Recall score on the Logical Memory IIa subtest (8-15) for participants with 13 or more years of education, or with MMSE (25-30) and Delayed Recall score (6-13) for participants with 12 years or less of education.

To determine $A\beta$ accumulation status in A4 study cohort, with/without (binary) positive $A\beta$ -PET (florbetapir) at a varying threshold level of Standardized Uptake Value ratios (SUVr) (value corresponding to the 'Composite_Summary' in the 'A4_PETSUVR.csv' file) was used (Supplemental Table 1). Meanwhile, in the ADNI data, due to the limited number of eligible participants with missing data, we used CSF $A\beta_{42} < 192$ pg/mL (values of median batch in the 'UPENNBIOMK_MASTER.csv' file) as the criterion for positive $A\beta$ accumulation (15). In the J-ADNI cohort, cases with CSF $A\beta_{42} < 333$ pg/mL (values in the 'pub_csf_apoe.tsv' file) (10) or with positive findings on the visual assessment of PiB-PET results (as listed in the 'pub_petqc.tsv' file) (16) were determined as positive $A\beta$.

We used the following clinical and laboratory features, which are available commonly in A4, ADNI, and J-ADNI datasets, as exploratory variables to include into the

Figure 1. Conceptual outline of our proposed method



We at first built a large number of models based on varying SUVr configurations (A, B), then we evaluated the performance of these models in each of the half-split external cohort (= ADNI or J-ADNI here) subgroups of cognitive normal participants (C, D). We supposed that we know the A β status of each case in subgroup1, while we do not know the statuses in subgroup2. When we compare the predictive performance results' distribution across different SUVr configurations (from 1 to k) between the external cohort half-split subgroups (e.g. ADNI or J-ADNI), the actual correlation should fall into the either significantly negative (F), non-significant (G), or significantly positive (H).

models: age at baseline, sex (male or female: binary), education years, with/without parental history of AD (binary), with/without elevated Clinical Dementia Rating sum of boxes (CDR-SB) at baseline (≥ 0.5 or not: binary), with/without APOE $\epsilon 4$ allele(s) (binary), and the baseline PACC score. Other features such as brain MRI or blood test results as used in our previous studies (11, 17, 18) were not included because they are not always available from A4, ADNI, and J-ADNI cohorts in a unified manner. Since the A4 study dataset up to 2019 contains baseline data alone and the participants' sequential changes have not been available, we also used the baseline data alone from the ADNI and J-ADNI datasets. The parental history of AD was regarded as positive if there was a statement that the participant's father or mother had been diagnosed with AD, and it was regarded as negative if there was no such statement or the data were missing.

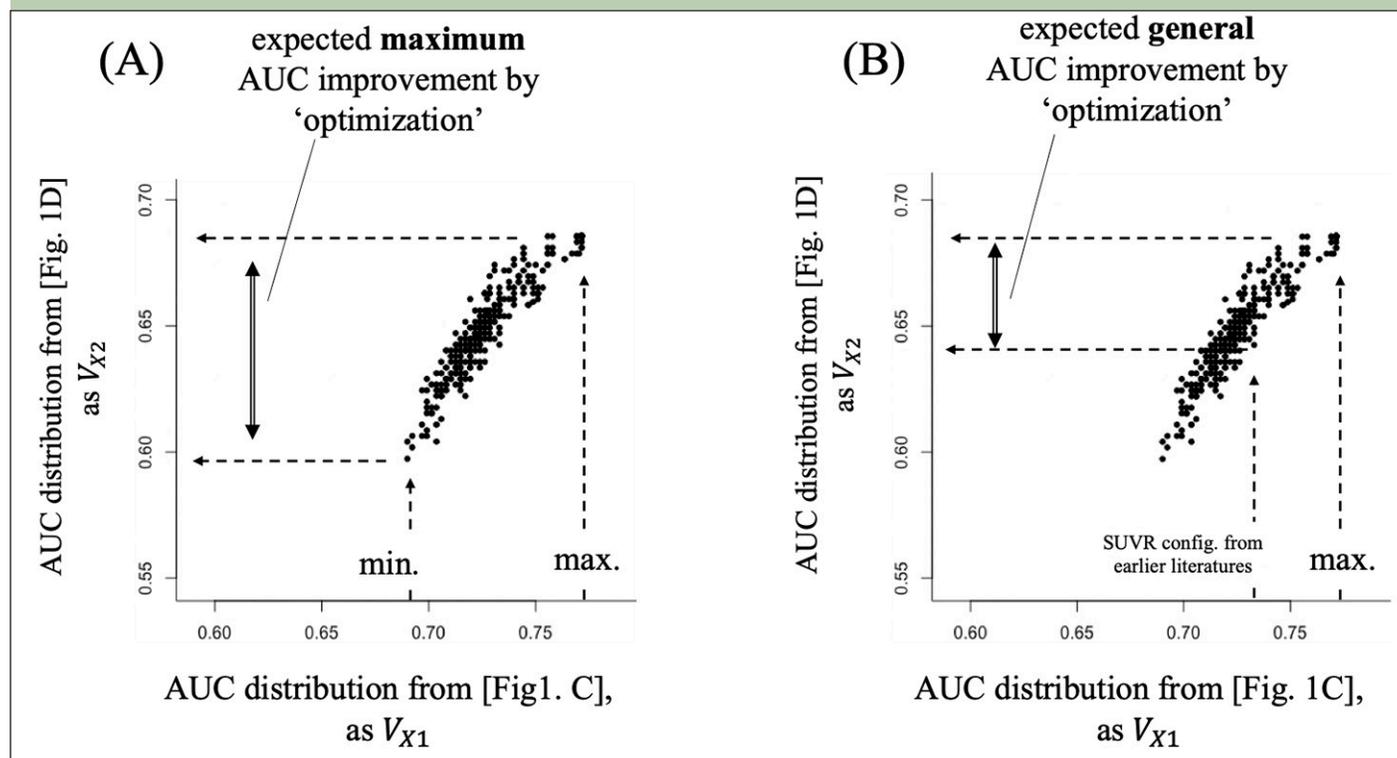
The PACC score (3) is the composite score, which is calculated from the sum of Z scores from 4 items: (1) the Total Recall score from the Free and Cued Selective Reminding Test (FCSRT), (2) Delayed Recall score on the Logical Memory IIa subtest from the Wechsler Memory Scale, (3) Digit Symbol Substitution Test score from the Wechsler Adult Intelligence Scale-Revised, and (4) MMSE total score [3]. Since the PACC score was not calculated in the ADNI and J-ADNI studies, we calculated the virtual PACC score by using the score of "LDELTOTAL: for

Logical Delayed, the score of "DIGITSCOR" for the Digit Symbol Substitution Test, and the total MMSE score. Furthermore, instead of using the FCSRT test score, which was not conducted in ADNI and J-ADNI studies, we used the delayed recall scores of ADAS-cog13 (Q4) in ADNI and J-ADNI datasets as in the earlier study (3). The Z scores of each of the 4 items were calculated within each ADNI or J-ADNI cohort in reference to the data of the cognitive normal cohort as allocated at baseline ("DX_bl" of "CN" (cognitive normal) or "SMC" (subjective memory complaints) in ADNI and the "COHORT" of "NL": (normal) in J-ADNI.

Missing data were handled by using the list-wise method: samples with missing data in the above modeling features were excluded from the analysis. Eventually, we included $n = 3233$ unique eligible cases of the A4 study cohort, $n = 86$ eligible cases of the ADNI cohort, and $n = 50$ eligible cases of the J-ADNI cohort.

Concepts of our proposed method

Here we explain how to demonstrate the practical effectiveness of our proposed ideas. First, we built a large number of models based on the varying SUVr configurations (Figures 1A, 1B) to predict positive A β within the A4 cohort data. Then, we evaluated the performance of these models, as calculated by

Figure 2. Evaluation of performance improvement

If we could observe that the actual correlation is consistently positive using the several datasets for evaluation (e.g. ADNI and J-ADNI here) as in the Figure 1H, the “optimization” to take the SUVr configuration of the model achieving the highest performance in one subgroup would also result in the near-highest performance in the rest of another subgroup with unknown $A\beta$ status (A, B).

area under the curve (AUC) as a performance metric of binary prediction models available regardless of the threshold value, in each of the half-split subgroups of the external cohort from A4 comprised of cognitive normal participants (Figure 1C, 1D). Suppose we know the $A\beta$ status of each case in subgroup1 (Figure 1C), while we do not know the status of each case in subgroup2 (Figure 1D). When we compare the distribution of predictive performance results across all SUVr configurations (from 1 to k here) between the subgroups (Figure 1E), the true correlation should fall into the significantly negative (Figure 1F), non-significant (Figure 1G), or significantly positive (Figure 1H) categories. If we can observe that the actual correlation is consistently positive using the various datasets for evaluation (e.g. ADNI and J-ADNI here), to find which model’s SUVr configuration achieves the highest performance in one subgroup, this would also result in the near-highest performance in the rest another subgroup with unknown $A\beta$ status (Figure 2A & 2B). We call the procedure to find the SUVr configuration with the highest AUC in one subgroup the ‘optimization of models’.

The significant-positive correlation (Figure 1F) is the prerequisite for this optimization. Although the half-split subgroups derived from the same cohort might tend to have a positive correlation due to the similar variance in their participants’ demographical data, such a tendency is not always validated, especially in cohorts which are

far smaller (e.g. ADNI or J-ADNI cognitive normal cases) than the A4 cohort. If the correlation between subgroups occasionally becomes negative (Figure 1F) or non-significant (Figure 1G), the optimization will not work. Therefore, our goal in this study was to confirm that the correlation between the half-split validation subgroups (Figures 1C versus 1D) is reproducibly significantly-positive (Figure 1F-H) and then to assess the degree of AUC improvement by employing this optimization procedure (Figures 2A & 2B), using the ADNI and J-ADNI datasets as validation.

Processing workflow: model training and performance evaluation

A detailed data processing workflow is outlined in Supplemental Figure 1. The target of A4-cohort predictive models is whether they are with/without positive $A\beta$ -PET (florbetapir) (binary) which are determined at varying SUVr threshold levels. In the model training, the SUVr threshold continuously varied by 0.01 from 0.99 to 1.47, corresponding to the [mean - 0.5 SD] and the [mean + 2 SD] of SUVr distribution in all the A4 data. Furthermore, we excluded the $A\beta$ -negative cases with an SUVr barely lower than the threshold, between which the margin range is varied, in order to exclude possible false-negative cases. This exclusion procedure substantially also acts to exclude possible false-positive cases, clarifying

the difference between cases with and without positive $A\beta$. Simultaneously adjusting with the above SUVR threshold, the “exclusion range” (Supplemental Figure 1A) is also adjusted continuously by 0.01 from 0 to 0.09, where 0.09 corresponds to [0.5 SD] of A4-SUVR. Taken together, cases whose SUVR is higher than the [threshold value] are defined as $A\beta$ -positive, and the cases whose SUVR is lower than the [threshold value – exclusion range value] are defined as $A\beta$ -negative (Supplemental Figure 1A). We here define this way of varying $A\beta$ allocation and the eligible case inclusion as “SUVR configuration,” which is used to generate a large number of models (Figure 1B). This SUVR configuration can be changed into 48 SUVR threshold patterns *10 exclusion range patterns = 480 combination patterns in total.

Since the small proportion of cases within the exclusion range is eliminated, the eligible A4 dataset A_k , which is from the A4 cohort cases ($n = 3233$), is slightly different depending on each SUVR configuration k ($k=1,2,\dots,480$) (Supplemental Figure 1B). Then a randomly selected 70% of A_k were further picked up as the A4 training subgroup A'_k ; using this A'_k subgroup, we trained a model M_k predictive for positive $A\beta$ (Supplemental Figure 1C). For the model M_k , we separately constructed 2 types of models, one of them including APOE $\epsilon 4$ status into its features (denoted as “model with APOE”), and another not including APOE $\epsilon 4$ status into the model (“model without APOE”) (Supplemental Figure 1C). This is because APOE $\epsilon 4$ is one of the strongest determinants of the CSF $A\beta_{42}$ level (19), while a model without APOE $\epsilon 4$ status would be more convenient to use as a screening index. The training was conducted with 10-fold cross validation and by a penalized generalized linear regression (GLM) algorithm using R package “caret” (20). Automated optimization of penalized GLM hyperparameters was conducted with grid-search by the caret function.

Then the predictive performance of the model M_k was validated in the ADNI and J-ADNI cohort data, out of the original A4. We split the ADNI and J-ADNI cohorts into half subgroups (“subgroup1” & “subgroup2”) randomly (Supplemental Figure 1G, in Figures 1C & 1D) while retaining equal proportions of $A\beta$ positive between the half subgroups using the “caret” package function (“createDataPartition”), then we aimed to compare the performance between ADNI subgroups or between J-ADNI subgroups. The predictive performance was measured with the metric of area under the curve (AUC), which is calculated by the predicted probability for the positive $A\beta$ of each case in the applied dataset (Supplemental Figure 1D).

Since the randomly sampled A'_k subgroup yields a slightly different model (Supplemental Figure 1C) every sampling time, we repeated the above processing steps (B-D: circled with gray color) 5 times in each k (shown with dagger mark [+]). We named the median from 5 times of AUC results as the $v_{xi,k}$ (Supplemental Figure 1E),

which means it is derived from the k -th configuration-based model M_k applied to the subgroup X_i .

As the configuration can vary for 480 types as described above, the full validation results ($k=1, 2, \dots, 480$) for one subgroup are represented by a vector with a length of 480. For example, when one cohort X (= ADNI or J-ADNI) data are split into subgroup X_1 and subgroup X_2 , vectors representing the results for these subgroups, which correspond to the result list of Figures 1C and 1D, are described as follows (Supplemental Figure 1F):

$$\begin{aligned} V_{X_1} &= [v_{X_1,1}, v_{X_1,2}, \dots, v_{X_1,480}] \\ V_{X_2} &= [v_{X_2,1}, v_{X_2,2}, \dots, v_{X_2,480}] \end{aligned}$$

Then we measured the correlation between V_{ADNI1} and V_{ADNI2} and between V_{JADNI1} and V_{JADNI2} .

The above process (steps A-F) was repeatedly performed for each ADNI and J-ADNI half-split subgroup (Supplemental Figure 1G), which are randomly separated 30 times in total (shown with the asterisk [*]), eventually yielding 30 sets of $[V_{ADNI1}, V_{ADNI2}, V_{JADNI1}, V_{JADNI2}]$.

Next, we again explain how the ‘optimization’ is conducted using Figure 1 & 2, the example scatter plot of V_{X_1} (plotted on X-axis) versus V_{X_2} (plotted on Y-axis) across all 480 patterns of SUVR configurations in one randomization time (*). On this plot, the Pearson correlation between the V_{X_1} and V_{X_2} was $R = 0.967$ ($p < 0.001$). When we choose the k_a of which v_{X_1,k_a} takes maximum among the V_{X_1} , the performance AUC with the same k_a -th SUVR configuration (= v_{X_2,k_a}) would also be approximately the highest among V_{X_2} . In other words, based on the assumption that the correlation between the vectors V_{X_1} versus V_{X_2} is significantly-positive (as in Figure 1H), we can optimize the predictive model in reference to the half subgroup X_1 alone so that the model takes the most of the best performance both in X_1 and the rest from the half subgroup X_2 , of which performance distribution is unknown to us, by choosing the k of which $v_{X_1,k}$ is the highest among V_{X_1} .

And when we choose the k_b of which v_{X_1,k_b} takes minimum among the V_{X_1} , the performance AUC with the same k_b -th SUVR configuration (= v_{X_2,k_b}) would also be approximately the lowest among V_{ADNI2} ; the difference between the v_{X_2,k_a} and v_{X_2,k_b} just corresponds to the theoretically-maximum AUC improvement expected to be achievable by the present “optimization” procedure (Figure 2A).

Furthermore, we compared the optimized result and the non-optimized result based on the conventionally-used SUVR configuration (e.g., threshold of 1.15 (21)). Supposing an i -th SUVR configuration with a threshold of 1.15 and exclusion range of 0, we measured the difference between the above v_{X_2,k_a} and the resulting AUC $v_{X_2,i}$ of the i -th configuration in subgroup2. This difference just corresponds to the AUC improvement expected to be achieved by using this optimization procedure [Figure 2B], compared to the conventional settings when not using “optimization” as in earlier studies.

Statistical analysis

All data handling and statistical analysis were performed using the software R 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). For numerical data, we used median and interquartile ranges (IQR) for summarization and the Wilcoxon rank sum test or analysis of variance (ANOVA) test for comparisons between groups. For categorical data, we used frequency and percentage for summarizing and Fisher's exact test for the group comparison. For calculating correlations between two numerical vectors, we used Pearson's correlation. A P-value less than 0.05 was regarded as statistically significant if not mentioned otherwise.

Results

Overview of the demographical distribution of the included cohorts

Basic demographics are shown in Supplemental Table 1, revealing slight differences among the data of the 3 included cohorts (A4, ADNI, and J-ADNI). The J-ADNI cohort participants had a significantly younger median age, were more predominantly male, and had fewer years of education than the other 2 cohorts. There was no significant difference among the 3 cohorts in the distribution of CDR-SB, parental history of AD, APOE ϵ 4 status, and the baseline PACC.

In addition, we also evaluated the performance of each single feature for predicting positive A β in each of the included cohorts. A heatmap of AUC result values as the predictive performance of the corresponding features (in columns) in each corresponding cohort (in rows) is shown in Supplemental Figure 2A. For the A4 cohort on this heatmap, a SUVr threshold of 1.15 was used (21). Each of the features except for APOE has a different level of association with the positive A β status, depending on the cohort.

Cohort-specific "optimization" of models

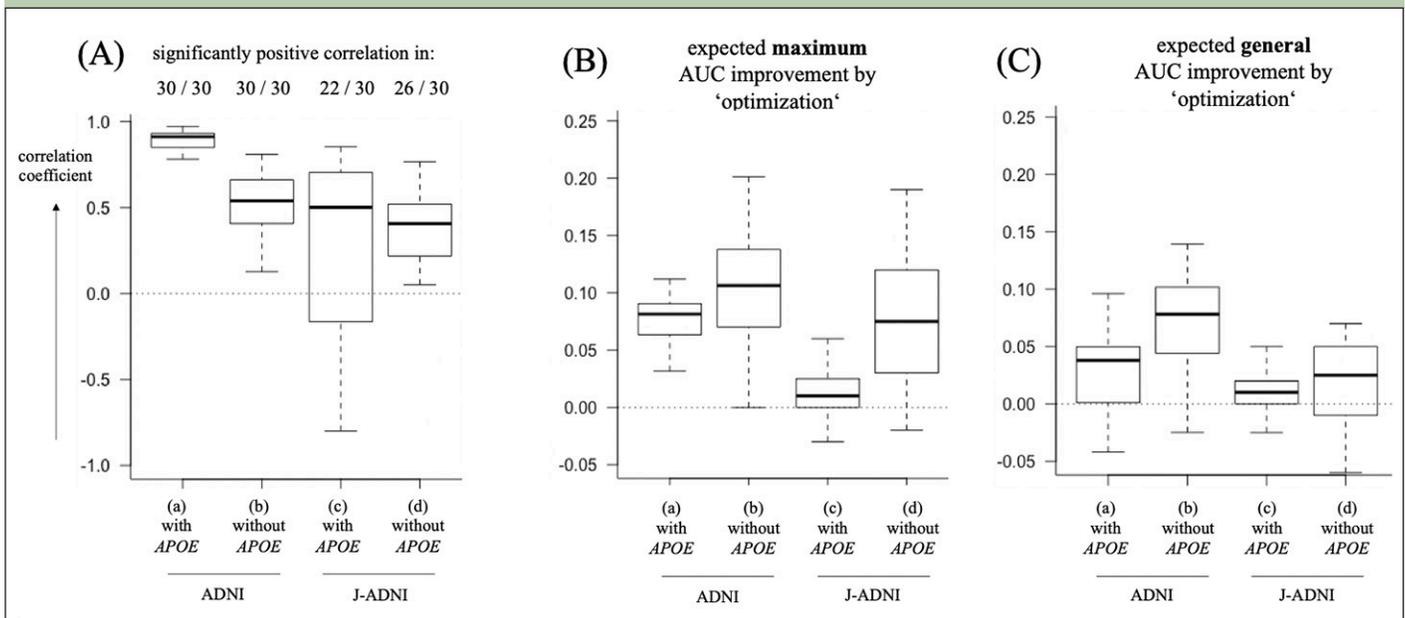
Next, we obtained the predictive performance of models based on the varying SUVr configuration evaluated with AUC in the ADNI and J-ADNI subgroups. We visualized the examples of the result vectors V_{ADNI1} , V_{ADNI2} , V_{JADNI1} , and V_{JADNI2} , summarizing the AUC from 480 different SUVr configurations (48 types of SUVr thresholds \times 10 types of exclusion ranges: Supplemental Figure 1A) by converting them into heatmap matrices for clarify (Supplemental Figure 2B). Each cell in the heatmaps represent the performance

AUC value of the model based on the corresponding SUVr configuration, where the row denotes the SUVr threshold and the column denotes the exclusion range. For the results from ADNI (Supplemental Figure 2B, left) or J-ADNI (Supplemental Figure 2B, right) data, we can see that the AUC performance results distribute differently depending on the SUVr configuration, and that the AUC performance results distribute differently largely depending on the cohort.

By choosing the darkest cell in the heatmap from the ADNI-subgroup1 (Supplemental Figure 2B), we can select the SUVr for each model's performance as that which is the highest in the ADNI subgroup1. As there was a positive correlation of $R = 0.767$ ($p < 0.001$) between the AUC heatmap of the ADNI subgroup1 and ADNI subgroup2 (Supplemental Figure 2B, left), the selected SUVr configuration would also take the near-best performance when applied to the rest of ADNI subgroup2. The same is true to the pair of J-ADNI subgroups (Supplemental Figure 2, right), between which there was a positive correlation of $R = 0.493$ ($p < 0.001$). This "optimization" procedure is generally cohort-specific since each cohort has specific spatial distribution of the resulting AUC heatmaps. Conversely, by choosing the lightest cell in the heatmap from ADNI subgroup1, the selected SUVr configuration would also show the near-lowest performance when applied to the rest of the ADNI subgroup2. The difference between the near-highest and the near-lowest AUC within subgroup2 corresponds to the "expected maximum AUC improvement achievable by optimization," the difference between the worst AUC when the "optimization" was not used, and the best AUC when the "optimization" was used.

Now the set of [V_{ADNI1} , V_{ADNI2} , V_{JADNI1} , V_{JADNI2}] (as in Supplemental Figure 2B) is repeatedly obtained for 30 times of ADNI and J-ADNI randomization (Supplemental Figure 1G, [*]), at first based on the model "with APOE": Figure 3A shows the distribution of the obtained correlation coefficients (as in Figure 1F-H) between V_{ADNI1} and V_{ADNI2} (summarized as Figure 3A, [a] & [b]) or between V_{JADNI1} and V_{JADNI2} (summarized as Figure 3A, [c] & [d]), repeated 30 times in total. In the ADNI cohort of models "with APOE" (Figure 3A [a]), Pearson's correlation coefficient between the ADNI subgroup1 and subgroup2 was a mean of 0.897 (the mean's 95% CI: 0.877 - 0.917), and the correlation coefficient > 0 and p -value < 0.05 were simultaneously observed in 30/30 of randomization (*) trials, fully meeting the prerequisite of our "optimization" method. The expected maximum AUC improvement width was a mean of 0.077 (the mean's 95% CI: 0.069 - 0.085) (Figure 3B [a]), and the expected AUC improvement when compared to the AUC in a model of SUVr threshold 1.15 was a mean of 0.033 (95% CI: 0.022 - 0.043) (Figure 3C [a]), e.g. AUC

Figure 3. Correlation coefficients between the resultant AUC vectors from half-split subgroups and the degree of AUC improvement



Box plots show the distribution of the obtained correlation coefficients (as in Figure 1F-H) between V_ADNI1 versus V_ADNI2 (A, [a] & [b]), or between V_JADNI1 and V_JADNI2 (A, [c] & [d]), repeated 30 times in total. Each box corresponds to the range between the lower and upper quartiles (Q1 and Q3, respectively), and the range between whiskers corresponds to the data distribution within the range of [Q1 - 1.5*IQR, Q3 + 1.5*IQR]. In the ADNI cohort (A, [a] & [b]), 30/30 of results both with models "with APOE" (A, [a]) or "without APOE" (A, [b]) showed a significantly positive correlation between V_ADNI1 versus V_ADNI2. In the J-ADNI cohort, 22/30 results of models "with APOE" (A, [c]) were significantly positive, and 26/30 results of models "without APOE" (A, [d]) were significantly positive. The expected maximum AUC improvement achievable by the "optimization" (B), and the expected AUC improvement achievable by the "optimization" when compared to the model based on the SUVR threshold of 1.15 without optimization (C) are plotted. In all models ([a]-[d]), the mean of "expected AUC improvement" was significantly higher than 0 (i.e. its lower 95% CI > 0), and a model "without APOE" in the ADNI cohort had approximately 0.10 of AUC improvement.

value improved from 0.724 to 0.774 in a representative case. Similarly, in the ADNI cohort by models "without APOE" (Figure 3A [b]), the correlation coefficient was a mean of 0.517 (the mean's 95% CI: 0.444 - 0.582), and the correlation coefficient > 0 and p-value < 0.05 were simultaneously observed in 30/30 of randomization trials (*). The expected maximum AUC improvement width was a mean of 0.107 (the mean's 95% CI: 0.086 - 0.129) (Figure 3B [b]), and the expected AUC improvement when compared to the AUC in a model of SUVR with a threshold of 1.15 was a mean of 0.075 (95% CI: 0.057 - 0.093) (Figure 3C [b]), e.g. AUC value improved from 0.61 to 0.69 in a representative case. In comparison, the expected maximum AUC improvement achievable by the "optimization" was greater in models "without APOE" than in models "with APOE" (Figure 3B) in ADNI (Figure 3B [a] versus [b], and Figure 3C [a] versus [b]).

In the J-ADNI cohort models "with APOE" (Figure 3A [c]), the correlation coefficient between J-ADNI subgroup1 and subgroup2 was a mean of 0.301 (the mean's 95% CI: 0.107 - 0.495), and a significant and positive correlation was observed in 22/30 randomization (*) trials, showing occasionally unsuccessful "optimization." The expected maximum AUC improvement width was a mean of 0.011 (the mean's 95% CI: 0.001 - 0.020) (Figure 3B [c]), and the expected AUC improvement when compared to the AUC in a model of SUVR with a threshold 1.15 was a mean of 0.009 (95%

CI: 0.003 - 0.016) (Figure 3C [c]), e.g. AUC value showed few improvement from 0.65 to 0.65 in a representative case. Furthermore, in the J-ADNI cohort models "without APOE" (Figure 3A [d]), the correlation coefficient was a mean of 0.353 (95% CI: 0.258 - 0.448), and a significant and positive correlation was observed in 26/30 randomization trials (*), mostly meeting the "optimization" prerequisite. The expected maximum AUC improvement width was a mean of 0.086 (95% CI: 0.060 - 0.113) (Figure 3B [d]), and the expected AUC improvement when compared to the AUC in a model of the SUVR threshold of 1.15 was a mean of 0.019 (95% CI: 0.007 - 0.030) (Figure 3C [d]), e.g. AUC value improved from 0.61 to 0.64 in a representative case. The models "without APOE" showed a higher expected maximum AUC improvement achievable with the "optimization" than the models "with APOE" (Figure 3B [c] versus [d], and Figure 3C [c] versus [d]).

Discussion

In this retrospective study, we demonstrated our attempts to optimize the A4 study-derived predictive models to be applicable to external cohort datasets, including ADNI and J-ADNI. The proposed method has novelty in that we operationally manipulated the positive A β allocation in the original training data of A4, thereby enabling the achievement of the best-performing model when applied to the external cohorts, including

ADNI and J-ADNI. The obtained AUC had improved mildly when compared to the AUC in case of literature-based predetermined SUVr threshold configuration. This means our ‘optimization’ procedure allowed us to obtain preclinical AD models for ADNI or J-ADNI with slightly better predictive performance. Our method may be practically useful in the middle of ongoing clinical study of preclinical AD, as a screening to further increase the prior probability of preclinical AD among the remaining samples before their amyloid testing.

The motivation of this study was mainly based on the concern as to the direct application of the A4 study-derived models to J-ADNI cohort, due to the differences in the distribution of participants’ baseline demographics such as age, sex, education years, ethnicity, the proportion of positive A β (Supplemental Table 1), or any unexamined clinical, laboratory, or genetic factors. It is known that such differences in the probability distributions of each feature between the training and validation datasets lead to failures in accurate prediction. “Transfer learning” is used in the field of deep learning as one of its solutions, enabling us to apply the trained model to the dataset origin of other domains. Thus, if utilized in our settings, it would enable us to apply the dataset from a different regional population with the smaller sample size (22, 23). However, our approach is based on conventional machine learning and is different from ‘transfer learning’, which we have not used since even the A β status in the original training data (= A4 study cohort) has not been definitely determined yet. If the biologically-corroborated criteria for the A β status are established within the original A4 cohort, transfer learning would be employable for building models effectively applicable to ADNI or J-ADNI datasets.

As expected, the efficacy of “optimization,” which is measured by the degree of AUC improvement compared to the resulting AUC of not using the “optimization,” was higher than 0 in average. The degree of maximum improvement in AUC (Figure 3B) and the degree of AUC improvement compared to the SUVr threshold of 1.15 (Figure 3C) are both approximately 0.10 in models “without APOE” applied to the ADNI cohort (Figure 3B[b], 3C[b]), which means this optimization procedure is expected when applied to the models “without APOE.” Although showing a smaller improvement, the models “without APOE” applied to the J-ADNI cohort also had a higher AUC than in the case of the models with any SUVr configuration (Figure 3B[d]) or with a conventional SUVr threshold 1.15 (Figure 3C[d]). This difference between ADNI and J-ADNI in their degree of AUC improvement may be due to the difference in their size of samples or in the degree of inter-cohort variation as represented by the different amyloid positivity rate.

Generally, the degree of AUC improvement (Figure 3B, 3C) tended to be higher in models “without APOE” ([b], [d]) than in models “with APOE” ([a], [c]), which means the performance is expected to improve by optimization

much larger models “without APOE” than models “with APOE,” probably reflecting the high importance of APOE ϵ 4 status as a variable for predicting positive A β . In addition, when the model “with APOE” was used, only 22/30 of a randomized half-split of the J-ADNI dataset led to a significantly positive correlation between V_{JADNI} and $V_{JADNI'}$ while it was more frequent (26/30) when the model “without APOE” was used. These results suggest that the current optimization methods are more reliably and effectively used in models not including APOE ϵ 4 status as features than those including it.

The current approach to adjust SUVr configuration consisted of the SUVr threshold and the exclusion of cases whose SUVr is barely lower than the threshold, is no more than an operational procedure here and is not biologically-validated in a strict sense. In this point, we need to be careful in the interpretation on the obtained final model or its variables’ importance that it is the “transferred” model and does not have certain biological basis on its own. For example, when we identified one feature (e.g. higher PACC) with high variable importance in the final model, the potential biological association between that feature and the A β positivity may be smaller than in the case of conventional non-transferred models.

Our study has some limitations. First, while the degree/frequency of positive correlation between the result vectors (Figure 3A) might be influenced by the size of the validating cohort datasets or their intra-cohort data variability, as suggested by our results where the efficacy of “optimization” showed smaller improvement and lower reliability when applied to the J-ADNI cohort than to the ADNI cohort, we have not examined the detailed conditions (e.g. sample size) required for the validation of datasets to be eligible for the “optimization” procedure. Further validation may be needed in other external cohorts with various kinds of sample sizes. Second, in the case of the single multi-center clinical trial to which we attempt to apply our method practically, there may be uncertainty whether the two subgroups collected from different facilities truly have a similar distribution in their demographical features, which is the pre-requisite for the external application of the current methods. Also, the extent to which the difference in inter-subgroup feature distribution can be allowed may be uncertain, and the sample size required to alleviate the potentially underlying variance between subgroups may also remain uncertain. Third, the proposed method manipulates the original training data distribution so as to be specifically best-performing in the validation cohort of interest, so the final model is not reversely applicable to the original A4 cohort data or to other cohorts with different demographical distributions. The fourth limitation is related to the PACC calculation in ADNI and J-ADNI: the validity of using ADAS-cog 13 (Q4) as a substitution of FCSRT, and the validity of setting “NL” cohort data as a reference of PACC calculation. And the fifth is that the proposed method takes a certain

amount of computational times, since model training and validation are repeatedly needed: 30 times of ADNI or J-ADNI splits for each [5 times of A4 training subgroup splits and model validations for each k (480 patterns in total)], eventually requiring us to calculate $30 \times 5 \times 480 = 72,000$ times of model training and validation. This is actually one of the reasons why we used penalized GLM as the prediction algorithm here, which takes shorter computational time than other types of algorithms such as random forest or support vector machine, and it is designed to have a smaller risk of over-fitting to the training data. If possible, other algorithms should also be tried (24). And lastly, used 3 cohorts referred to different modality of amyloid tests (i.e., florbetapir-PET in A4, CSF in ADNI, and CSF and PiB-PET in J-ADNI), possibly lowering the applicability of our method.

To conclude, we proposed a novel method to obtain preclinical A β predictive models specifically optimized to the cohort of interest in order to achieve extrapolative application out of the original training data. This optimization procedure showed efficacy of up to 0.10 of AUC improvement when used in combination with the models “without APOE.” Our method may be practically useful in the mid of the actual clinical study of preclinical AD, as a screening to further increase the prior probability of preclinical AD before amyloid testing.

Funding: This study was supported by Japan Agency for Medical Research and Development grants JP21dk0207057, JP21dk0207048, and JP20dk0207028.

Description about the ADNI: Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Description about the A4 study: The A4 Study is a secondary prevention trial in preclinical Alzheimer’s disease, aiming to slow cognitive decline associated with brain amyloid accumulation in clinically normal older individuals. The A4 Study is funded by a public-private-philanthropic partnership, including funding from the National Institutes of Health-National Institute on Aging, Eli Lilly and Company, Alzheimer’s Association, Accelerating Medicines Partnership, GHR Foundation, an anonymous foundation and additional private donors, with in-kind support from Avid and Cogstate. The companion observational Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study is funded by the Alzheimer’s Association and GHR Foundation. The A4 and LEARN Studies are led by Dr. Reisa Sperling at Brigham and Women’s Hospital, Harvard Medical School and Dr. Paul Aisen at the Alzheimer’s Therapeutic Research Institute (ATRI), University of Southern California. The A4 and LEARN Studies are coordinated by ATRI at the University of Southern California, and the data are made available through the Laboratory for Neuro Imaging at the University of Southern California. The participants screening for the A4 Study provided permission to share their de-identified data in order to advance the quest to find a successful treatment for Alzheimer’s disease. We would like to acknowledge

the dedication of all the participants, the site personnel, and all of the partnership team members who continue to make the A4 and LEARN Studies possible. The complete A4 Study Team list is available on: a4study.org/a4-study-team.

Conflicts of interest: The authors have no conflict of interest to disclose.

Ethical standards: This study was approved by the University of Tokyo Graduate School of Medicine institutional ethics committee (ID: 11628-(3)).

References

- Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH. Toward defining the preclinical stages of Alzheimer’s disease: recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimers Dement*. 2011 May;7(3):280-92.
- Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD, Lesnick TG, Pankratz VS, Donohue MC, Trojanowski JQ. Tracking pathophysiological processes in Alzheimer’s disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol*. 2013 Feb;12(2):207-16.
- Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, Weiner M, Aisen PS; Australian Imaging, Biomarkers, and Lifestyle Flagship Study of Ageing; Alzheimer’s Disease Neuroimaging Initiative; Alzheimer’s Disease Cooperative Study. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol*. 2014 Aug;71(8):961-70.
- Cummings J. The National Institute on Aging-Alzheimer’s Association Framework on Alzheimer’s disease: Application to clinical trials. *Alzheimers Dement*. 2019 Jan;15(1):172-178.
- Jansen WJ, Ossenkuppe R, Knol DL, Tijms BM, Scheltens P, Verhey FR, Visser PJ; Amyloid Biomarker Study Group, Aalten P, Aarsland D, Alcolea D, Alexander M, Almdahl IS, Arnold SE, Baldeiras I, Barthel H, van Berckel BN, Bibeau K, Blennow K, Brooks DJ, van Buchem MA, Camus V, Cavedo E, Chen K, Chetelat G, Cohen AD, Drzezga A, Engelborghs S, Fagan AM, Fladby T, Fleisher AS, van der Flier WM, Ford L, Förster S, Fortea J, Foskett N, Frederiksen KS, Freund-Levi Y, Frisoni GB, Froelich L, Gabryelewicz T, Gill KD, Gkatzima O, Gómez-Tortosa E, Gordon MF, Grimmer T, Hampel H, Hausner L, Hellwig S, Herukka SK, Hildebrandt H, Ishihara L, Ivanoiu A, Jagust WJ, Johannsen P, Kandimala R, Kapaki E, Klimkiewicz-Mrowiec A, Klunk WE, Köhler S, Koglin N, Kornhuber J, Kramerberger MG, Van Laere K, Landau SM, Lee DY, de Leon M, Lisetti V, Lleó A, Madsen K, Maier W, Marcusson J, Mattsson N, de Mendonça A, Meulenbroek O, Meyer PT, Mintun MA, Mok V, Molinuevo JL, Møllergård HM, Morris JC, Mroczko B, Van der Mussele S, Na DL, Newberg A, Nordberg A, Nordlund A, Novak GP, Paraskevas GP, Parnetti L, Perera G, Peters O, Popp J, Prabhakar S, Rabinovici GM, Ramakers IH, Rami L, Resende de Oliveira C, Rinne JO, Rodrigue KM, Rodríguez-Rodríguez E, Roe CM, Rot U, Rowe CC, Rütter E, Sabri O, Sanchez-Juan P, Santana I, Sarazin M, Schröder J, Schütte C, Seo SW, Soetewey F, Soininen H, Spuru L, Struyfs H, Teunissen CE, Tsolaki M, Vandenberghe R, Verbeek MM, Villemagne VL, Vos SJ, van Waalwijk van Doorn JL, Waldemar G, Wallin A, Wallin ÅK, Wiltfang J, Wolk DA, Zboc M, Zetterberg H. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA*. 2015 May 19;313(19):1924-38.
- Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, Aisen P. The A4 study: stopping AD before symptoms begin? *Sci Transl Med*. 2014 Mar 19;6(228):228f13.
- Insel PS, Palmqvist S, Mackin RS, Nosheny RL, Hansson O, Weiner MW, Mattsson N. Assessing risk for preclinical β -amyloid pathology with APOE, cognitive, and demographic information. *Alzheimers Dement (Amst)*. 2016 Aug 3;4:76-84.
- Ansart M, Epelbaum S, Gagliardi G, Colliot O, Dormont D, Dubois B, Hampel H, Durrleman S; Alzheimer’s Disease Neuroimaging Initiative* and the INSIGHT-preAD study. Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis. *Stat Methods Med Res*. 2020 Jan;29(1):151-164.
- Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR Jr, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW. Alzheimer’s Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010 Jan;74(3):201-9.
- Iwatsubo T, Iwata A, Suzuki K, Ihara R, Arai H, Ishii K, Senda M, Ito K, Ikeuchi T, Kuwano R, Matsuda H; Japanese Alzheimer’s Disease Neuroimaging Initiative, Sun CK, Beckett LA, Petersen RC, Weiner MW, Aisen PS, Donohue MC; Alzheimer’s Disease Neuroimaging Initiative. Japanese and North American Alzheimer’s Disease Neuroimaging Initiative studies: Harmonization for international trials. *Alzheimers Dement*. 2018 Aug;14(8):1077-1087.
- Iwata A, Iwatsubo T, Ihara R, Suzuki K, Matsuyama Y, Tomita N, Arai H,

- Ishii K, Senda M, Ito K, Ikeuchi T, Kuwano R, Matsuda H; Alzheimer's Disease Neuroimaging Initiative; Japanese Alzheimer's Disease Neuroimaging Initiative. Effects of sex, educational background, and chronic kidney disease grading on longitudinal cognitive and functional decline in patients in the Japanese Alzheimer's Disease Neuroimaging Initiative study. *Alzheimers Dement* (N Y). 2018 Jul 12;4:765-774.
12. Ihara R, Iwata A, Suzuki K, Ikeuchi T, Kuwano R, Iwatsubo T; Japanese Alzheimer's Disease Neuroimaging Initiative. Clinical and cognitive characteristics of preclinical Alzheimer's disease in the Japanese Alzheimer's Disease Neuroimaging Initiative cohort. *Alzheimers Dement* (N Y). 2018 Nov 26;4:645-651.
 13. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, Masters C, Milner A, Pike K, Rowe C, Savage G, Szoek C, Taddei K, Villemagne V, Woodward M, Ames D; AIBL Research Group. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr*. 2009 Aug;21(4):672-87.
 14. Clark CM, Schneider JA, Bedell BJ, Beach TG, Bilker WB, Mintun MA, Pontecorvo MJ, Hefti F, Carpenter AP, Flitter ML, Krautkramer MJ, Kung HF, Coleman RE, Doraiswamy PM, Fleisher AS, Sabbagh MN, Sadowsky CH, Reiman EP, Zehntner SP, Skovronsky DM; AV45-A07 Study Group. Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA*. 2011 Jan 19;305(3):275-83.
 15. Shaw LM, Vanderstichele H, Knapiak-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ; Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009 Apr;65(4):403-13.
 16. Yamane T, Ishii K, Sakata M, Ikari Y, Nishio T, Ishii K, Kato T, Ito K, Senda M; J-ADNI Study Group. Inter-rater variability of visual interpretation and comparison with quantitative evaluation of 11C-PiB PET amyloid images of the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) multicenter study. *Eur J Nucl Med Mol Imaging*. 2017 May;44(5):850-857.
 17. Sato K, Mano T, Ihara R, Suzuki K, Tomita N, Arai H, Ishii K, Senda M, Ito K, Ikeuchi T, Kuwano R, Matsuda H, Iwatsubo T, Toda T, Iwata A; Alzheimer's Disease Neuroimaging Initiative, and Japanese Alzheimer's Disease Neuroimaging Initiative. Lower Serum Calcium as a Potentially Associated Factor for Conversion of Mild Cognitive Impairment to Early Alzheimer's Disease in the Japanese Alzheimer's Disease Neuroimaging Initiative. *J Alzheimers Dis*. 2019;68(2):777-788.
 18. Sato K, Mano T, Matsuda H, Senda M, Ihara R, Suzuki K, Arai H, Ishii K, Ito K, Ikeuchi T, Kuwano R, Toda T, Iwatsubo T, Iwata A; Japanese Alzheimer's Disease Neuroimaging Initiative. Visualizing modules of coordinated structural brain atrophy during the course of conversion to Alzheimer's disease by applying methodology from gene co-expression analysis. *Neuroimage Clin*. 2019 Jul 25;24:101957.
 19. Lautner R, Palmqvist S, Mattsson N, Andreasson U, Wallin A, Pålsson E, Jakobsson J, Herukka SK, Owenius R, Olsson B, Hampel H, Rujescu D, Ewers M, Landén M, Minthon L, Blennow K, Zetterberg H, Hansson O; Alzheimer's Disease Neuroimaging Initiative. Apolipoprotein E genotype and the diagnostic accuracy of cerebrospinal fluid biomarkers for Alzheimer disease. *JAMA Psychiatry*. 2014 Oct;71(10):1183-91.
 20. Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. (2018). caret: Classification and Regression Training. R package version 6.0-81. (<https://CRAN.R-project.org/package=caret>)
 21. Pascoal TA, Mathotaarachchi S, Shin M, Park AY, Mohades S, Benedet AL, Kang MS, Massarweh G, Soucy JP, Gauthier S, Rosa-Neto P; Alzheimer's Disease Neuroimaging Initiative. Amyloid and tau signatures of brain metabolic decline in preclinical Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2018 Jun;45(6):1021-1030.
 22. Yosinski J., Clune J., Bengio Y., Lipson H. NIPS; 2014. How Transferable are Features in Deep Neural Networks? pp. 3320-3328.
 23. Wee CY, Liu C, Lee A, Poh JS, Ji H, Qiu A; Alzheimers Disease Neuroimage Initiative. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. *Neuroimage Clin*. 2019;23:101929.
 24. Sato K, Ihara R, Suzuki K, Niimi Y, Toda T, Jimenez-Maggiore G, Langford O, Donohue MC, Raman R, Aisen PS, Sperling RA, Iwata A, Iwatsubo T. Predicting amyloid risk by machine learning algorithms based on the A4 screen data: Application to the Japanese Trial-Ready Cohort study. *Alzheimers Dement* (N Y). 2021 Mar 24;7(1):e12135.

How to cite this article: K. Sato, T. Mano, R. Ihara et al. Cohort-Specific Optimization of Models Predicting Preclinical Alzheimer's Disease, to Enhance Screening Performance in the Middle of Preclinical Alzheimer's Disease Clinical Studies. *J Prev Alz Dis* 2021;4(8):503-512; <http://dx.doi.org/10.14283/jpad.2021.39>