Journal of Alzheimer's Disease 99 (2024) 953–963
DOI 10.3233/JAD-231426
IOS Press

Simplifying Alzheimer's Disease Monitoring: A Novel Machine-Learning Approach to Estimate the Clinical Dementia Rating Sum of Box Changes Using the Mini-Mental State Examination and Functional Activities Questionnaire

Kenichiro Sato[a,b], Yoshiki Niimi[b,c], Ryoko Ihara[d], Kazushi Suzuki[e], Atsushi Iwata[d], Takeshi Iwatsubo[a,b,*], the Alzheimer's Disease Neuroimaging Initiative[1] and the Japanese Alzheimer's Disease Neuroimaging Initiative

[a]*Department of Neuropathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*
[b]*Unit for Early and Exploratory Clinical Development, The University of Tokyo Hospital, Tokyo, Japan*
[c]*Department of Healthcare Economics and Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*
[d]*Department of Neurology, Tokyo Metropolitan Institute for Geriatrics and Gerontology, Tokyo, Japan*
[e]*Division of Neurology, Internal Medicine, National Defense Medical College, Saitama, Japan*

Accepted 27 March 2024
Pre-press 11 May 2024

Abstract.
Background: Primary outcome measure in the clinical trials of disease modifying therapy (DMT) drugs for Alzheimer's disease (AD) has often been evaluated by Clinical Dementia Rating sum of boxes (CDRSB). However, CDR testing requires specialized training and 30–50 minutes to complete, not being suitable for daily clinical practice.
Objective: Herein, we proposed a machine-learning method to estimate CDRSB changes using simpler cognitive/functional batteries (Mini-Mental State Examination [MMSE] and Functional Activities Questionnaire [FAQ]), to replace CDR testing.
Methods: Baseline data from 944 ADNI and 171 J-ADNI amyloid-positive participants were used to build machine-learning models predicting annualized CDRSB changes between visits, based on MMSE and FAQ scores. Prediction performance was evaluated with mean absolute error (MAE) and R^2 comparing predicted to actual ΔCDRSB/Δyear. We further assessed whether decline in cognitive function surpassing particular thresholds could be identified using the predicted ΔCDRSB/Δyear.

*Correspondence to: Takeshi Iwatsubo, Department of Neuropathology, Graduate School of Medicine, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8655, Japan. Tel.: +81 03 3815 5411; E-mail: iwatsubo@m.u-tokyo.ac.jp.

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Results: The models achieved the minimum required prediction errors (MAE < 1.0) and satisfactory prediction accuracy ($R^2 > 0.5$) for mild cognitive impairment (MCI) patients for changes in CDRSB over periods of 18 months or longer. Predictions of annualized CDRSB progression > 0.5, >1.0, or >1.5 demonstrated a consistent performance (i.e., Matthews correlation coefficient > 0.5). These results were largely replicated in the J-ADNI case predictions.

Conclusions: Our method effectively predicted MCI patient deterioration in the CDRSB based solely on MMSE and FAQ scores. It may aid routine practice for disease-modifying therapy drug efficacy evaluation, without necessitating CDR testing at every visit.

Keywords: Alzheimer's disease, clinical dementia rating, disease-modifying therapy, efficacy assessment, machine-learning, prediction

INTRODUCTION

Alzheimer's disease (AD) is a major cause of dementia in the elderly [1]. Several anti-amyloid drugs, such as aducanumab, lecanemab, and donanemab, have been developed as promising disease-modifying therapy (DMT) drugs for AD [2]. Clinical trials have demonstrated that these drugs can suppress cognitive decline in patients with mild cognitive impairment (MCI) or early AD to a certain extent [3, 4]. Notably, lecanemab received FDA approval in July 2023 [5], and it was also approved in Japan in September 2023 [6] and in mainland China in January 2024 [7]. DMT drugs for AD are thus moving beyond the clinical trial stage and are just beginning to be applied in real-world practice, raising expectations but also concerns about the preparedness for the treatment [8].

In real-world settings, the efficacy of DMT drugs may not consistently match the 100% efficacy observed in clinical trials for all patients. This is true when the patient's background is different from that of the RCT population [5]. The efficacy of DMT drugs for patients who started treatment for their mild stage of dementia but progressed to moderate dementia during the course of treatment remains uncertain. Additionally, patients undergoing DMT treatment are at risk of developing adverse effects, including Amyloid-Related Imaging Abnormalities (ARIA) [5, 9, 10], which may occasionally necessitate the intermittent cessation of treatment. Therefore, when administering DMT drugs in real-world settings, it is crucial for clinicians to periodically assess their safety as well as clinical progression over time. While adverse effects such as ARIA can be evaluated using scheduled brain MRI scans [5, 11], practical challenges exist when assessing clinical progression. In many clinical trials of these DMT drugs, the primary outcome measure evaluating clinical progression is the Clinical Dementia Rating sum of boxes (CDRSB), which is derived from the Clinical Dementia Rating (CDR) test that evaluates the clinical severity of dementia [12]. One problem is that CDR testing requires special training for raters, and it takes 30–50 min to complete the assessment for a single patient. CDR testing is not as straightforward as Mini-Mental State Examination (MMSE) testing in regular clinical practice [13]. Therefore, the need for adequate efficacy assessment requiring CDR testing for every routine efficacy evaluation, may act as a barrier to the widespread adoption of DMT drugs.

To address this problem, we propose a machine-learning approach that estimates the change in CDRSB using simpler cognitive/functional batteries, namely: the MMSE and Functional Activities Questionnaire (FAQ). In this study, addressing a clinical scenario in which clinicians aim to assess the efficacy of DMT drugs for their individual patients, we specifically focus on predicting the rate of change in the CDRSB, specifically the annualized changes, rather than solely assessing the CDRSB scores themselves. This approach acknowledges that monitoring the disease's progression over time is critical for assessing clinical status and roughly inferring treatment efficacy on an individual basis. We consider that understanding the annualized changes in the CDRSB may provide a more clinically intuitive measure compared to the raw CDRSB scores, thus enhancing the practical application of our method in real-world clinical settings. The proposed method is intended to replace CDR testing, thereby facilitating the assessment of DMT drug efficacy in clinical practice.

METHODS

Study purpose

This retrospective study used pre-existing data and was approved by a local ethics committee [ID:11754-(1)]. We aimed to introduce a machine-

learning approach suited for future clinical settings, to aid physicians in assessing the efficacy of DMT in patients who are receiving or have previously received DMT drugs. The anticipated clinical scenario is as follows: a patient meets the eligibility criteria for DMT drugs (e.g., evidence of amyloid accumulation in the brain while maintaining a certain cognitive function level). The patient was scheduled for periodic administration of a DMT drug at an outpatient clinic. While safety assessments especially concerning ARIA, were conducted at each clinic visit; efficacy was evaluated on a set schedule, such as once every six months or annually. In this context, our goal was to estimate the degree of change in CDRSB scores after each efficacy evaluation period.

Used data

We used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) study. ADNI study is a comprehensive longitudinal observational research initiative that began in 2004, gathering data from over 2,000 participants, including those classified with cognitively normal (CN), MCI, and dementia [14]. J-ADNI study initiated in Japan in 2008, follows a protocol nearly identical to that of the ADNI and has gathered data from approximately 500 participants [15]. We downloaded the data in June 2023, with permission from the study teams. Participants were included based on the following criteria: evidence of positive amyloid accumulation in the brain, as confirmed by either cerebrospinal fluid (CSF) analysis or amyloid PET (with tracers: florbetapir or florbetaben for ADNI and PiB for J-ADNI), a minimum of two visits to the study site, and a CDR-global score (CDR-GS) of 1 or less at baseline. A CDR-GS of 0 indicates CN or subjective memory complaints (SMC), a CDR-GS of 0.5 generally corresponds to early/late MCI but includes a fraction of mild AD (e.g., approximately 15% among ADNI cases with CDR-GS 0.5 at baseline), and a CDR-GS of 1 or higher corresponds to AD dementia [12]. Amyloid accumulation was considered positive if the participant met any one of the following criteria: for ADNI data, a baseline CSF $A\beta_{42}$ level < 192 [16]; an SUVR > 1.11 in florbetapir PET, or an SUVR > 1.08 in florbetaben PET [17]; for J-ADNI data, a CSF $A\beta$ level < 333 [15] or a positive result in visual assessment from PiB PET. Our analysis only considered participants displaying positive $A\beta$ accumulation at their baseline evaluation.

The following covariables were incorporated into our analysis: age at the time of study participation; sex (binary: male or female); years of education; CDR-GS and CDRSB at baseline and subsequent visits; diagnosis at baseline (denoted as "DX_bl" in the original dataset); marital status (binary: married or not); cohort details (binary: ADNI or J-ADNI); *APOE* genotype (numerically represented by the count of $\varepsilon 4$ alleles); MMSE scores at baseline and subsequent visits; and FAQ scores at baseline and subsequent visits. The CSF p-tau status at baseline (e.g., with/without elevation) was not considered because of the absence of a well-defined threshold in the J-ADNI dataset.

Data preprocessing

Now we revisit the aforementioned assumed clinical scenario. The regimen administered differed depending on the type of DMT drug administered or the patient characteristics. Some patients could have undergone treatment continuously for three years or longer, whereas others could have been treated for only one or two years or less. This variability necessitates the evaluation of efficacy during and after the administration period. Thus, it would be advantageous to estimate the degree of change in CDRSB scores over an arbitrary time span rather than being restricted to every 6 or 12 months.

Consequently, we obtained paired data consisting of any two observational visits from the longitudinal data of the visits of each participant. For illustration (see Fig. 1A), consider a participant with MCI at baseline who attended the ADNI study site five times in total (e.g., at baseline (0 months), 6 months, 12 months, 18 months, and 24 months). From this participant's longitudinal data, we could formulate 10 "previous – current" visit pairs (i.e., 0–6, 0–12, 0–18, 0–24, 6–12, 6–18, 6–24, 12–18, 12–24, and 18–24 months). The intervals between these visit pairs were 6, 12, 18, 24, 6, 12, 18, 6, 12, and 6 months, respectively. We set the maximum time span between visits at 24 months (or 2 years) and capped the last possible visit at 36 months from the baseline. The paired data were then used as inputs for machine learning.

The target variable we sought to predict using machine-learning was the annualized difference in CDRSB between "previous" and "current" visits for each pair (denoted as ΔCDRSB/Δyear in Fig. 1A): the ΔCDRSB/Δyear within each paired data is calculated using the formula:

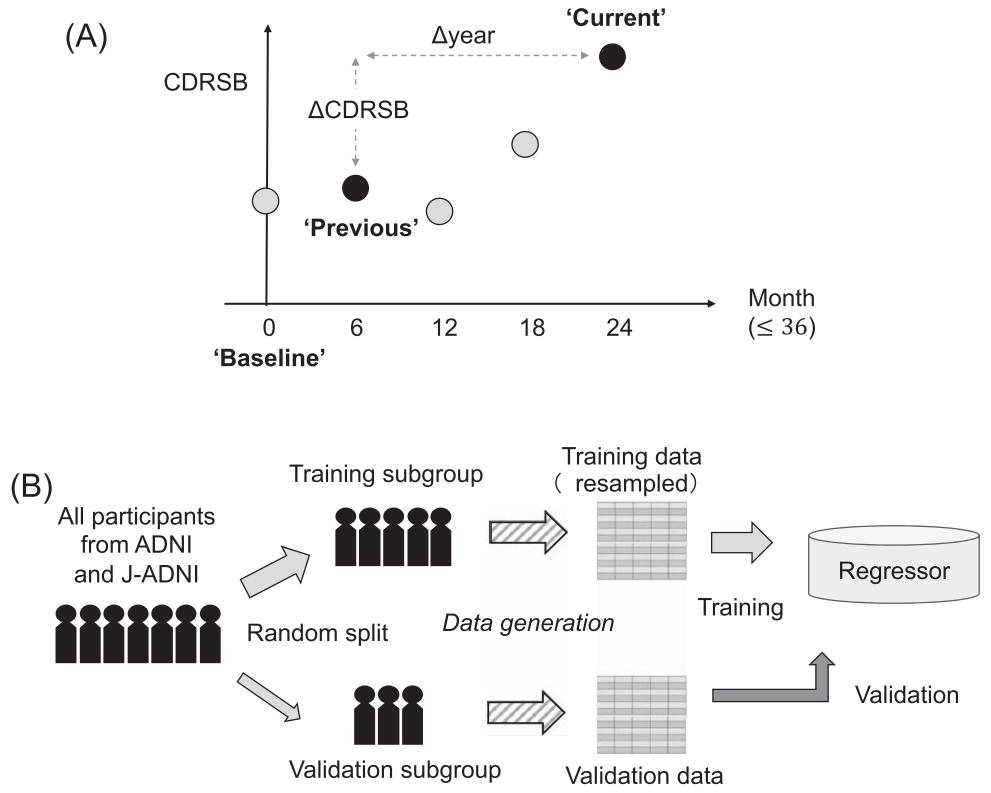

Fig. 1. Outline of target variable and data workflow. We created pairs using two arbitrary observational visits from the longitudinal visit data for each participant. The paired data were then used as observations for machine learning. For example, in this figure (A), the visit at 24 month is designated as the 'current' visit, and the visit at 6 months as the 'previous' visit, making the period between these visits 18 months. The ΔCDRSB/Δyear for this paired data is calculated using the formula: $\Delta CDRSB/\Delta year = (CDRSB_{current} - CDRSB_{previous})/\{(Month_{current} - Month_{previous})/12\}$. The generated paired data were utilized for both training and validation of the machine-learning model. Given that multiple pairs of data arose from a single participant's longitudinal data, we initially randomly partitioned all participants into training and validation groups at a 2 : 1 ratio (B). Paired observation data derived from the training group were denoted as the training dataset, whereas paired observation data from the validation group formed the validation dataset. The training dataset was resampled via replacement. Training and hyperparameter tuning were performed using the R package *caret*. Subsequently, the model performance was assessed using the validation dataset. This process of data splitting, resampling, model training, and validation was iteratively executed 1,000 times with varying random seeds. CDRSB, Clinical Dementia Rating Sum of Boxes; ADNI, Alzheimer's Disease Neuroimaging Initiative; J-ADNI, Japanese Alzheimer's Disease Neuroimaging Initiative.

$$\Delta CDRSB\Delta year = (CDRSB_{current} - CDRSB_{previous})/\{(Month_{current} - Month_{previous})/12\}$$

We used 122 features as explanatory variables (Supplementary Table 1). These consisted of: age at the current visit; sex; number of years of education; marital status; *APOE* genotype; MMSE scores at baseline, previous visit, and current visit; and FAQ scores at baseline, previous visit, and current visit. Both the MMSE and FAQ scores encapsulated the total and individual sub-item scores. Neither the CDRSB nor the CDR-GS from any visit were included, as their inclusion would detract from the practical applicability of the proposed approach.

Machine-learning architecture

All data processing and analyses were conducted using R software. The generated paired data were utilized for both training and validation of the machine-learning model. Given that multiple pairs of data arose from a single participant's longitudinal data, we initially randomly partitioned all participants into training and validation groups in a 2 : 1 ratio (see Fig. 1B). Paired observation data derived from the training group were denoted as the training dataset, whereas paired observation data from the validation group formed the validation dataset. The training dataset was resampled via replacement. Model training employed the "regularized generalized linear regression" regressor (specifically, we

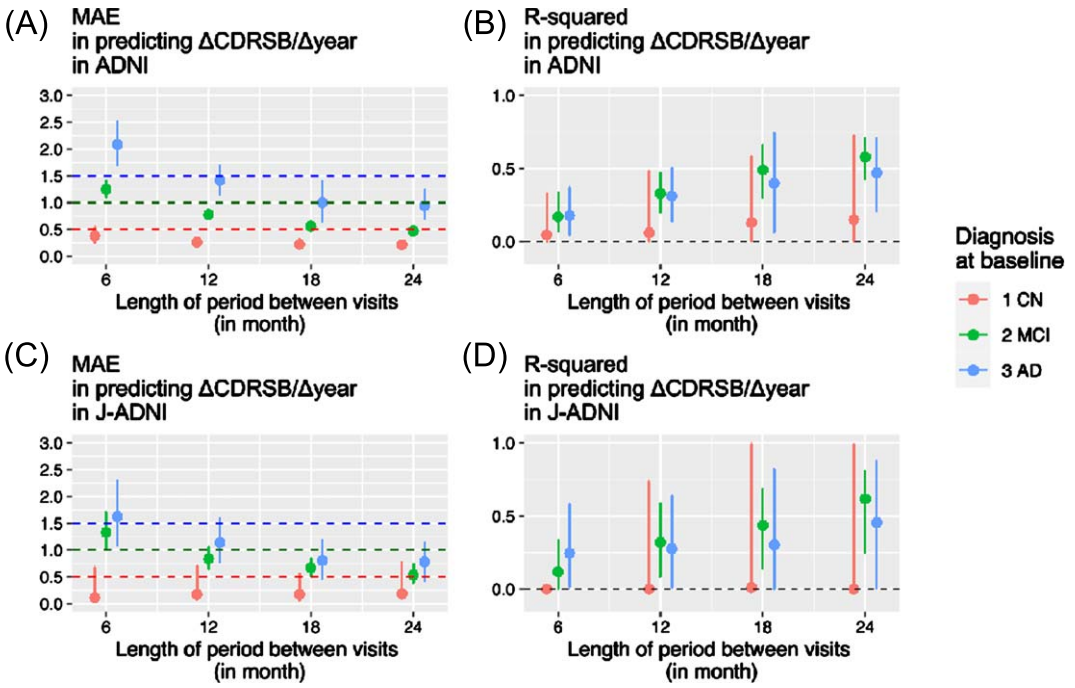

Fig. 2. Prediction performances in MAE and R^2. We assessed the performance of predicting the annualized change in CDRSB, defined as ΔCDRSB/Δyear. The prediction error (MAE; see (A) for ADNI participants and (C) for J-ADNI participants) met the minimum required level (i.e., below +0.5 CN, +1.0, and +1.5 for AD participants) for predicting CDRSB score changes in patients diagnosed with MCI or AD at baseline, especially for periods spanning 18 months or longer between visits. However, the predictions for CN participants, especially those from the J-ADNI cohort, were not as promising. Regarding prediction accuracy, the R^2 values (illustrated in (B) for ADNI participants and (D) for J-ADNI participants) were substantially unsatisfactory in the CN participants, displaying a low median R^2 and broad 95% CI encompassing 0. Moreover, participants also exhibited relatively poor performance, with a 95% CI for R^2 verging at 0. In contrast, participants with MCI at baseline consistently showed relatively better performance in terms of R^2. MAE, mean absolute error; CDRSB, Clinical Dementia Rating Sum of Boxes; MMSE, Mini-Mental State Examination; FAQ, Functional Activities Questionnaire; AD, Alzheimer's disease; MCI, mild cognitive impairment; CN, cognitively normal; CI, confidence interval; ADNI, Alzheimer's Disease Neuroimaging Initiative; J-ADNI, Japanese Alzheimer's Disease Neuroimaging Initiative.

used Elastic Net), abbreviated as "glmnet." Training and hyperparameter tuning with repeated cross-validation were performed using the R package *caret*. Subsequently, the model performance was assessed using the validation dataset. This process of data splitting, resampling, model training, and validation was iteratively executed 1,000 times with varying random seeds. The aggregate prediction performance metrics were collated and presented as medians and 95% confidence intervals (CI).

Evaluation of prediction performance

Prediction accuracy was gauged using the mean absolute error (MAE) and R-squared (R^2) values, comparing the predicted ΔCDRSB/Δyear against the actual ΔCDRSB/Δyear across all validation data. The MAE is defined by the following formula: $MAE = \frac{1}{n}\sum_{i=1}^{n}|predicted - actual|$. Since AD participants at baseline may experience a faster cognitive decline than that experienced by CN participants [18], the expected ΔCDRSB/Δyear may differ depending on the baseline status. Furthermore, the trajectory of CDRSB may not be linear over time. To address this, we visualized predictive performance according to baseline cognitive status (as indicated by baseline diagnosis [see Figs. 2 and 3]), interval length between visits, and originating cohort (i.e., ADNI or J-ADNI).

Regarding the minimum acceptable level of prediction error, we evaluated whether the upper bound of the 95% CI for the obtained MAE was below +0.5 for CN cases, +1.0 for MCI cases, and +1.5 for AD cases. These threshold values are derived from earlier research which indicated that clinically meaningful deterioration in CDRSB was +0.54 from the previous visit (spanning an average of 1.15 years) for CN participants, +0.98 for MCI participants, and +1.63 for AD participants [19]. If the MAE of a model exceeded these thresholds, it could not be used to detect clin-

Table 1
Characteristics of the included participants

Cohort	ADNI	J-ADNI
Total Participants	$N = 944$	$N = 171$
Age at baseline	Median 74.1 (IQR: 69.2∼78.6)	Median 74 (IQR: 68.7∼77.7)
Sex	Male 515: Female 429	Male 84: Female 87
APOE ε4 allele[s]	2 = 144 : 1 = 453 : 0 = 347	2 = 23 : 1 = 83 : 0 = 65
Diagnosis at baseline	AD 237 : MCI 477 : CN 230	AD 76 : MCI 76 : CN 18
CDR-GS at baseline	1 = 122 : 0.5 = 590 : 0 = 232	1 = 22 : 0.5 = 130 : 0 = 19
CDRSB at baseline	Median 1.5 (IQR: 0.5∼3)	Median 2.0 (IQR: 1.0∼3.5)
MMSE at baseline	Median 27 (IQR: 25∼29)	Median 25 (IQR: 23∼27)
FAQ at baseline	Median 2 (IQR: 0∼8)	Median 5 (IQR: 1∼9.5)
Marital status	Married 743: Not 201	Married 149: Not 22

CDR-GS, Clinical Dementia Rating Global score; CDRSB, Clinical Dementia Rating Sum of Boxes; MMSE, Mini-Mental State Examination; FAQ, Functional Activities Questionnaire; AD, Alzheimer's disease; MCI, mild cognitive impairment; CN, cognitively normal; IQR, interquartile range; ADNI, Alzheimer's Disease Neuroimaging Initiative; J-ADNI, Japanese Alzheimer's Disease Neuroimaging Initiative.

ically significant changes. $R^2 = 0.5$ was used as an approximate measure of good prediction accuracy.

Evaluating clinical meaningful change

One crucial practical consideration when estimating ΔCDRSB/Δyear over a specific period is determining a decline in cognitive function surpassing a particular threshold. As previously described, setting the ΔCDRSB/Δyear threshold at +0.5 for CN, +1.0 for MCI, and +1.5 for AD can enable machine-learning models to discern whether a patient has undergone clinically significant deterioration since their last visit.

Prediction accuracy was appraised using the Matthews correlation coefficient (MCC)- ranging between –1 and +1, alongside the positive predictive value (PPV) and negative predictive value (NPV). The MCC is similar to the Pearson correlation in terms of its interpretation and is considered superior to the F1 score or accuracy in evaluating binary classification tasks [20]. An MCC of 0 corresponds to random prediction, –1 indicates perfect inverse prediction, and 1 points to an entirely accurate prediction. We used the MCC = 0.5 as a rough measure of good prediction accuracy. PPV and NPV denote the proportions (ranging from zero to one) of genuinely positive or negative outcomes relative to all positively or negatively predicted instances, respectively.

We adjusted the threshold levels to +0.5, +1.0, +1.5, and +2.0 for sensitivity analyses. Similar to the MAE and R^2 metrics, we visualized the MCC, PPV, and NPV results based on baseline cognitive status, interval between visits, and the originating cohort.

RESULTS

Overview

We included 1,115 unique participants: 944 from the ADNI study and 171 from the J-ADNI study. The basic characteristics of the participants are summarized in Table 1. From their longitudinal data, we obtained 7,282 original paired observations, 5,928 from the ADNI study and 1,354 from the J-ADNI study.

The top-20 predictive variables ranked by average variable importance are shown in Supplementary Figure 1. The total FAQ score during the current visit was the most important variable, followed by the total MMSE score at the current visit. *APOE* genotype was at the 109th percentile of all 122 variables.

Prediction error and accuracy

We then assessed the performance of predicting the annualized change in CDRSB, defined as ΔCDRSB/Δyear. The prediction error (MAE; see Fig. 2A for ADNI participants and Fig. 2C for J-ADNI participants) met the minimum required level (i.e., below +0.5 for CN, +1.0 for MCI, and +1.5 for AD participants) for predicting CDRSB score changes in patients diagnosed with MCI or AD at baseline, especially for periods spanning 18 months or longer between visits. However, the predictions for CN participants, especially those from the J-ADNI cohort, were not as promising.

Regarding prediction accuracy, the R^2 values (illustrated in Fig. 2B for ADNI participants and Fig. 2D for J-ADNI participants) were substantially

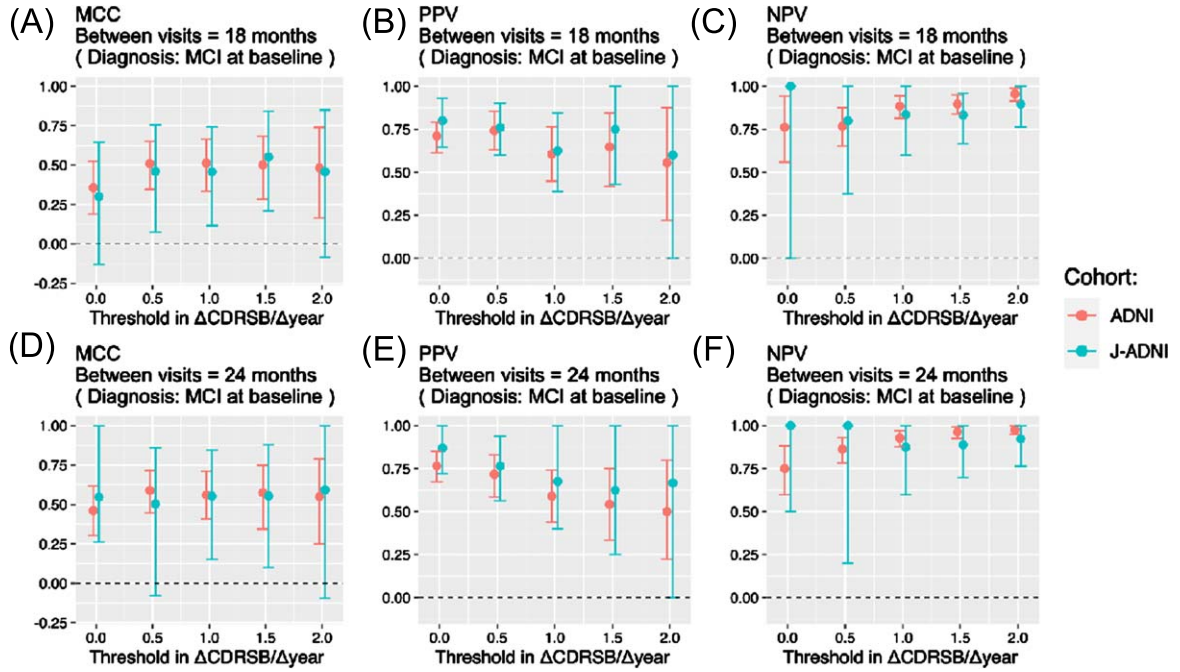

Fig. 3. Prediction performances in participants with MCI at baseline. Based on results in Fig. 2, we focused on predicting CDRSB change in participants with MCI at baseline, to periods spanning 18 months or longer between visits. We obtained MCC (A, D), PPV (B, E), and NPV (C, F) metrics. In A-C, the period between visits is 18 months, and in D-F, it extends to 24 months. Notably, irrespective of the threshold levels on the x-axis, the overall performance, as indicated by MCC, appears relatively stable. Conversely, the PPV and NPV show contrasting variations in their 95%CI ranges based on the x-axis threshold level. For example, when the threshold is set at 1.5 or 2.0, the 95% CI for PPV broadens, while that for NPV tightens. Specifically for MCI cases, thresholds for ΔCDRSB/Δyear exceeding 0.5, 1.0, or 1.5 seem to be the most appropriate threshold, offering reliable performance across MCC, PPV, and NPV. MCC, Matthews correlation coefficient; PPV, positive predictive value; NPV, negative predictive value; CDRSB, Clinical Dementia Rating Sum of Boxes; MCI, mild cognitive impairment; ADNI, Alzheimer's Disease Neuroimaging Initiative; J-ADNI, Japanese Alzheimer's Disease Neuroimaging Initiative.

unsatisfactory in the CN participants, displaying a low median R^2 and broad 95% CI encompassing 0. Moreover, participants also exhibited relatively poor performance, with a 95% CI for R^2 verging at 0. In contrast, participants with MCI at baseline consistently showed relatively better performance in terms of R^2, nearly above 0.5: median 0.491 (95%CI:0.306–0.662) for ADNI cases and 0.436 (95%CI:0.141–0.683) for J-ADNI cases for periods spanning 18 months between visits. For periods spanning 24 months between visits, the median was 0.580 (95%CI:0.428–0.710) for the ADNI cases and 0.618 (95%CI:0.247–0.808) for the J-ADNI cases.

Predicting clinical significant change

Based on these results, we focused on predicting CDRSB changes in participants with MCI at baseline for periods spanning 18 months or longer between visits. The MCC, PPV, and NPV metrics were obtained, as shown in Fig. 3. In Fig. 3A–C, the period between visits was 18 months, and in Fig. 3D–F, it extended to 24 months. Notably, irrespective of the threshold levels on the x-axis, the overall performance, as indicated by MCC, appeared to be relatively stable. The median MCC was larger than 0.5, in ADNI and largely the same in J-ADNI; for ADNI cases, the median MCC was 0.510 for threshold at >0.5, 0.513 for threshold at >1.0, and 0.501 for threshold at >1.5. For J-ADNI cases, the median was 0.461 for the threshold at >0.5, 0.457 for the threshold at >1.0, and 0.552 for the threshold at >1.5.

Meanwhile, PPV and NPV showed contrasting variations in their 95% CI ranges based on the x-axis threshold level. For example, when the threshold was set at >2.0, the 95% CI for the PPV broadened, whereas it increased for the NPV. Conversely, when the threshold was set to >0.0, the 95% CI for the PPV tightened, whereas that for the NPV broadened. For thresholds at >0.5, >1.0, and >1.5, PPV had 0.6–0.7 as median and NPV 0.8 has median 0.8 both for ADNI and J-ADNI cases.

Taken together, specifically for MCI cases, thresholds for ΔCDRSB/Δyear exceeding 0.5, 1.0, or 1.5 are the appropriate threshold, offering a fair performance reliably across MCC, PPV, and NPV.

DISCUSSION

In this study, we proposed a machine-learning approach to predict changes in the CDRSB using primarily MMSE and FAQ scores. Our results demonstrated that our proposed models met the minimum required level of prediction errors (i.e., MAE) and displayed a satisfactory level of prediction accuracy (i.e., R^2) for MCI patients at baseline, provided that the prediction pertained to changes in the CDRSB over a period of 18 months or longer between visits. Moreover, predictions regarding annualized progression in CDRSB > 0.5, > 1.0, or > 1.5 consistently showed good predictive performance. These findings suggest that our proposed approach could be practically valuable in predicting whether MCI patients at baseline have experienced clinically significant deterioration in the CDRSB in subsequent visits, based solely on MMSE and FAQ scores. This has the potential to be widely adopted in daily practice to evaluate the efficacy of DMT drugs without requiring CDR testing at every visit.

In this study, we used data from the ADNI and J-ADNI studies, both of which are observational studies. This means that our machine-learning models were based on non-interventional study data, which were then applied to clinical scenarios in which patients were treated with DMT drugs. We believe this approach is justifiable because, in this study, we primarily focused on cognitive and functional scales (specifically, symptomatic changes along the disease course) rather than changes in AD pathological biomarkers during the longitudinal period, which would be altered by the administration of DMT drugs.

We now discuss in detail how our approach may benefit the treatment of individual patients using DMT drugs in a real-world setting. We made assumptions based on scenarios in which patients treated with DMT drugs periodically underwent efficacy assessments at outpatient clinics. In the Clarity-AD study, a phase-3 trial that demonstrated the efficacy of lecanemab in reducing cognitive decline in MCI and mild AD patients [3], the change in CDRSB was+1.21 over 18 months for the group treated with lecanemab, and $+1.66 > 18$ months for the group treated with placebo. This equates to changes of +0.8/year and +1.1/year in CDRSB, respectively. However, in real-world settings, not every patient receiving lecanemab will benefit from such a decrease in cognitive deterioration, and some may even progress faster than expected despite treatment. At this juncture, if we determine a certain threshold level of change in the CDRSB as an efficacy benchmark prior to starting treatment using our models, we can assess in individual patients whether the drug operates above or below the average efficacy levels reported in clinical trials. For example, drawing from the Clarity-AD study [3], if we set the threshold at >1/year, it can be inferred that MCI patients receiving lecanemab and predicted by our model to have a CDRSB change of more than 1/year might be experiencing efficacy below the trial's average level of +0.8/year. Conversely, since the NPV is as good as the PPV, we can also say that MCI patients treated with lecanemab and not predicted by the model to undergo a CDRSB change of more than 1/year might be experiencing efficacy comparable to or better than the average level (i.e., +0.8/year) observed in the clinical trial.

Referring to TRAILBLAZER-ALZ2 study [4], a phase-3 clinical trial that investigated the efficacy and safety of donanemab in early AD patients including MCI and mild dementia, observed change in CDRSB in donanemab and placebo arms were +1.17/year and +1.65/year, respectively. If we set the threshold at >1.5/year, it can be inferred that MCI patients receiving donanemab, predicted by the model to have a CDRSB change of more than 1.5/year, might experience efficacy below the average level of efficacy (i.e., +1.17/year) observed in the clinical trial.

Additionally, the prediction model could potentially serve as a clinical tool for determining premature discontinuation of the regimen in certain patients, although there are no definitive criteria for deciding when to discontinue DMT treatment. For example, because of the high medical costs associated with DMT medications, the inconvenience of frequent outpatient clinic visits, or excessive concerns about adverse effects, some patients may want premature discontinuation of DMT drugs if they do not perceive adequate efficacy.

The potential usefulness of our proposed approach is expected to be reflected in clinical practice in Japan. Specifically, following the approval of lecanemab in Japan in September 2023, the Clinical Optimal Use Guideline (OUG) for lecanemab in Japan [21] was published in December 2023. The OUG requires measuring the CDR or other corresponding measures for all patients seeking lecanemab treatment as part

of the eligibility criteria (i.e., CDR-GS 0.5 or 1). It also requires that every patient receiving lecanemab be monitored every six months to track their clinical progression. Additionally, the OUG calls for considering the discontinuation of treatment for patients who are unlikely to benefit from it. We believe that our study will also be useful in providing an alternative to CDR testing for the periodic monitoring of Japanese patients receiving lecanemab treatment.

Our study had several limitations as listed below.

1. The ΔCDRSB/Δyear is obtained by dividing ΔCDRSB by the period between visits within a paired observation. This calculation assumes that the longitudinal change in CDRSB is linear. Therefore, we restricted the period between visits to a maximum of 2 years.
2. The regressor used in our model training was the regularized GLM, which exhibited a relatively high-performance accuracy. However, not all regressor types have been examined to date. Therefore, there may be opportunities to improve the prediction accuracy using alternative algorithms, such as gradient boosting.
3. While our models can assess whether an individual receiving a DMT drug experiences cognitive changes above or below the average progression reported in the active arm of clinical trials of the drug, we cannot determine the precise efficacy of the drug.
4. We could not include baseline p-tau status, corresponding to the "T" in the A/T/N classification [22], as we used both ADNI and J-ADNI datasets. Incorporating baseline T status in the model might increase prediction accuracy, as it is known to be one of the prognostic factors of cognitive decline, as measured in the CDRSB [23].
5. In applying machine-learning models derived from observational study data to clinical scenarios in which patients were treated with DMT drugs, we assumed that because of the low frequency of severe symptomatic ARIA, the influence of ARIA on cognitive and functional scales in the medium to long term was virtually negligible at the population level. In short, our models should not be applied to patients who develop symptomatic ARIA and do not recover completely.
6. The model equation we used for regressor - regularized generalized linear regression is more complicated than that of conventional linear regression to calculate manually, so that we have not presented final specific formula to be used for real-world calculation by directly applying actual patient scores in clinical settings. Instead, in the near future, we would like to release some kind of interactive online application that can be used easily.

In the future, we are considering validating the performance and feasibility of our proposed method using real-world data. Specifically, regarding lecanemab in Japan, post-marketing surveillance is mandated for all patients treated with lecanemab [21]. The development of ARIA, *APOE* genotype, comorbidities, concurrent medications, and clinical measures including MMSE, CDRSB, and other scales have been discussed as some of the variables to be collected within the surveillance [24]. This suggests that there may be opportunities to validate our approach in the future. Although it is currently uncertain whether the *APOE* genotype will be available from all patients in the surveillance, the absence of the *APOE* genotype may not critically impact the performance of our model, given that the variable importance of the *APOE* genotype was at the 109th percentile of all 122 variables (Supplementary Figure 1). Furthermore, incorporating the influence of comorbidities, concurrent medications, or adverse events including ARIA into the prediction of clinical progression might improve our model and optimize it for real-world application.

In conclusion, our study introduced a machine-learning approach to predict changes in the CDRSB using primarily the MMSE and FAQ scores. Our approach consistently performed well in predicting whether patients with MCI at baseline would experience clinically significant deterioration in the CDRSB over a period of 18 months or longer. A notable advantage of our model is its simplicity; it only requires the MMSE and FAQ scores along with baseline features. Thus, it has the potential for widespread use in daily practice to evaluate the efficacy of DMT drugs without requiring CDR testing at every visit.

AUTHOR CONTRIBUTIONS

Kenichiro Sato (Conceptualization; Data curation; Formal analysis; Writing – original draft); Yoshiki Niimi (Conceptualization; Writing – review & editing); Ryoko Ihara (Writing – review & editing); Kazushi Suzuki (Writing – review & editing); Atsushi

Iwata (Writing – review & editing); Takeshi Iwatsubo (Supervision).

ACKNOWLEDGMENTS

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 (http://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.

FUNDING

This study was supported by AMED Grant Numbers JP21dk0207048 and JP21dk0207054 and JSPS KAKENHI Grant Number 21K20891.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

The data supporting the findings of this study are openly available at (https://ida.loni.usc.edu/) for ADNI and (https://humandbs.biosciencedbc.jp/en/hum0043-v1) for J-ADNI data.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-231426.

REFERENCES

[1] Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chételat G, Teunissen CE, Cummings J, van der Flier WM (2021) Alzheimer's disease. *Lancet* **397**, 1577-1590.

[2] Cummings J, Zhou Y, Lee G, Zhong K, Fonseca J, Cheng F (2023) Alzheimer's disease drug development pipeline: 2023. *Alzheimers Dement (N Y)* **9**, e12385.

[3] van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, Kanekiyo M, Li D, Reyderman L, Cohen S, Froelich L, Katayama S, Sabbagh M, Vellas B, Watson D, Dhadda S, Irizarry M, Kramer LD, Iwatsubo T (2023) Lecanemab in early Alzheimer's disease. *N Engl J Med* **388**, 9-21.

[4] Sims JR, Zimmer JA, Evans CD, Lu M, Ardayfio P, Sparks J, Wessels AM, Shcherbinin S, Wang H, Monkul Nery ES, Collins EC, Solomon P, Salloway S, Apostolova LG, Hansson O, Ritchie C, Brooks DA, Mintun M, Skovronsky DM; TRAILBLAZER-ALZ 2 Investigators (2023) Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA* **330**, 512-527.

[5] Cummings J, Apostolova L, Rabinovici GD, Atri A, Aisen P, Greenberg S, Hendrix S, Selkoe D, Weiner M, Petersen RC, Salloway S (2023) Lecanemab: Appropriate use recommendations. *J Prev Alzheimers Dis* **10**, 362-377.

[6] Eisai. "LEQEMBI® Intravenous Infusion" (Lecanemab) Approved for the Treatment of Alzheimer's Disease in Japan, https://www.eisai.com/news/2023/news202359.html, Accessed on September 25, 2023.

[7] Biogen. "LEQEMBI®" (Lecanemab) Approved for the Treatment of Alzheimer's Disease in China, https://investors.biogen.com/news-releases/news-release-details/leqembir-lecanemab-approved-treatment-alzheimers-disease-china, Accessed on January 9, 2024.

[8] Brodtmann A, Darby D, Oboudiyat C, Mahoney CJ, Le Heron C, Panegyres PK, Brew B (2023) Assessing preparedness for Alzheimer disease-modifying therapies in Australasian health care systems. *Med J Aust* **218**, 247-249.

[9] Honig LS, Barakos J, Dhadda S, Kanekiyo M, Reyderman L, Irizarry M, Kramer LD, Swanson CJ, Sabbagh M (2023) ARIA in patients treated with lecanemab (BAN2401) in a phase 2 study in early Alzheimer's disease. *Alzheimers Dement (N Y)* **9**, e12377.

[10] Barakos J, Purcell D, Suhy J, Chalkias S, Burkett P, Marsica Grassi C, Castrillo-Viguera C, Rubino I, Vijverberg E (2022) Detection and management of amyloid-related imaging abnormalities in patients with Alzheimer's disease

treated with anti-amyloid beta therapy. *J Prev Alzheimers Dis* **9**, 211-220.

[11] Zhang Y, Wang S, Phillips P, Dong Z, Ji G, Yang J (2015) Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC. *Biomed Signal Process Control* **21**, 58-73.

[12] Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **43**, 2412-2414.

[13] Nosheny RL, Yen D, Howell T, Camacho M, Moulder K, Gummadi S, Bui C, Kannan S, Ashford MT, Knight K, Mayo C, McMillan M, Petersen RC, Stricker NH, Roberson ED, Chambless C, Gersteneker A, Martin R, Kennedy R, Zhang Y, Kukull W, Flenniken D, Fockler J, Truran D, Mackin RS, Weiner MW, Morris JC, Li Y (2023) Evaluation of the electronic clinical dementia rating for dementia screening. *JAMA Netw Open* **6**, e2333786.

[14] 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 (2010) Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* **74**, 201-209.

[15] 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 (2018) Japanese and North American Alzheimer's Disease Neuroimaging Initiative studies: Harmonization for international trials. *Alzheimers Dement* **14**, 1077-1087.

[16] Shaw LM, Vanderstichele H, Knapik-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 (2009) Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* **65**, 403-413.

[17] Sturchio A, Dwivedi AK, Young CB, Malm T, Marsili L, Sharma JS, Mahajan A, Hill EJ, Andaloussi SE, Poston KL, Manfredsson FP, Schneider LS, Ezzat K, Espay AJ (2021) High cerebrospinal amyloid-β 42 is associated with normal cognition in individuals with brain amyloidosis. *EClinicalMedicine* **38**, 100988.

[18] Doraiswamy PM, Sperling RA, Johnson K, Reiman EM, Wong TZ, Sabbagh MN, Sadowsky CH, Fleisher AS, Carpenter A, Joshi AD, Lu M, Grundman M, Mintun MA, Skovronsky DM, Pontecorvo MJ; AV45-A11 Study Group; AV45-A11 Study Group (2014) Florbetapir F 18 amyloid PET and 36-month cognitive decline: A prospective multicenter study. *Mol Psychiatry* **19**, 1044-1051.

[19] Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR (2019) Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. *Alzheimers Dement (N Y)* **5**, 354-363.

[20] Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6.

[21] Ministry of Health, Labour, and Welfare. Optimal Use Guideline for lecanemab, https://www.mhlw.go.jp/content/12404000/001178607.pdf, Accessed on December 13, 2023.

[22] Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R; Contributors (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562.

[23] Yagi T, Kanekiyo M, Ito J, Ihara R, Suzuki K, Iwata A, Iwatsubo T, Aoshima K; Alzheimer's Disease Neuroimaging Initiative; Japanese Alzheimer's Disease Neuroimaging Initiative (2019) Identification of prognostic factors to predict cognitive decline of patients with early Alzheimer's disease in the Japanese Alzheimer's Disease Neuroimaging Initiative study. *Alzheimers Dement (N Y)* **5**, 364-373.

[24] Central Social Insurance Medical Council. Responses to high-cost drugs (for dementia): resources submitted on October 18, 2023, https://www.mhlw.go.jp/content/12404000/001157647.pdf, Accessed on December 13, 2023.