Development and validation of the Florey Dementia Risk Score web-based tool to screen for Alzheimer's disease in primary care

Yijun Pan,[a,b,d,e,*] Chenyin Chu,[a,b,d,**] Yifei Wang,[a,b] Yihan Wang,[a,b] Guangyan Ji,[a] Colin L. Masters,[a,b] Benjamin Goudey,[a,b,c] and Liang Jin,[a,b,e] AIBL Research Group

[a]The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, 3052, Australia
[b]Florey Department of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, 3052, Australia
[c]The ARC Training Centre in Cognitive Computing for Medical Technologies, The University of Melbourne, Parkville, Victoria, 3052, Australia

eClinicalMedicine 2024;76: 102834
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102834

Summary

Background It is estimated that ~60% of people with Alzheimer's disease (AD) are undetected or undiagnosed, with higher rates of underdiagnosis in low-to middle-income areas with limited medical resources. To promote health equity, we have developed a web-based tool that utilizes easy-to-collect clinical data to enhance AD detection rate in primary care settings.

Methods This study was leveraged on the data collected from participants of the Australian Imaging, Biomarker & Lifestyle (AIBL) study and the Religious Orders Study and Memory and Aging Project (ROSMAP). The study included three phases: (1) constructing and evaluating a model on retrospective cohort data (1407 AIBL participants), (2) performing simulated trials to assess model accuracy (30 AIBL participants) and missing data tolerability (30 AIBL participants), and (3) external evaluation using a non-Australian dataset (500 ROSMAP participants). The auto-score machine learning algorithm was employed to develop the Florey Dementia Risk Score (FDRS). All the simulated trials and evaluation were performed using a web-based FDRS tool.

Findings FDRS achieved an area under the curve (AUC) of approximately 0.82 [95% CI, 0.75–0.88], with a sensitivity of 0.74 [0.60–0.86] and a specificity of 0.73 [0.70–0.79]. The accuracy of the simulated pilot trial for 30 AIBL participants with complete record was 87% (26/30 correct), while it only slightly decreased (80.0–83.3%, depending on imputation methods) for another 30 AIBL participants with one or two missing data. FDRS achieved an AUC of 0.82 [0.77–0.86] of 500 ROSMAP participants.

Interpretation The FDRS tool offers a potential low-cost solution to AD screening in primary care. The present study warrants future trials of FDRS for optimization and to confirm its generalizability across a more diverse population, especially people in low-income countries.

Funding National Health and Medical Research Council, Australia (GNT2007912) and Alzheimer's Association, USA (23AARF-1020292).

Keywords: Alzheimer's disease; Auto-score algorithm; Binary classification; Disease screening; Health equity; Web-based tool

Introduction

Alzheimer's disease (AD) is the major cause of dementia. It is estimated that 59% of older adults with dementia in the United States are undiagnosed or unaware of their diagnosis, suggesting shortcomings in detection of dementia.[1] From a global perspective, a systematic review of 23 studies published prior to 2016 revealed that the pooled rate of undetected dementia

*Corresponding author. Florey Department of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, 3052, Australia.
**Corresponding author. Florey Department of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, 3052, Australia.
E-mail addresses: yijun.pan@unimelb.edu.au (Y. Pan), chenyin.chu1@unimelb.edu.au (C. Chu).
[d]Co-first author.
[e]Co-senior author.

Research in context

Evidence before this study

We searched PubMed for studies published from the database inception to April 20, 2024, using combinations of the terms "binary classification", "machine learning", "Alzheimer's disease", "auto-score", "self-report tool", and "cognitive decline" without language restrictions. Additionally, we searched Google Scholar and reviewed reference lists to identify relevant studies. We excluded studies that developed binary classification models for Alzheimer's disease (AD) and cognitively unimpaired (CU) individuals using cognitive assessments, cerebrospinal fluid biomarkers, and neuroimaging. Most of the included studies focused on the binary classification of CU/AD using self-report tools such as the Australian National University Alzheimer's Disease Risk Index. These tools typically use an evidence-based medicine approach for feature selection. Of note, many of these tools use an excessive number of features or include neuropsychological tests that require the assistance from neuropsychologists, making them unsuitable for AD screening, especially for countries and areas with limited medical resources.

Added value of this study

To our knowledge, this is the first study exploring the use of an auto-score framework for binary classification for AD. This novel method employs random forest feature selection and parsimony analysis to minimize the number of features used in the model, while also incorporates evidence-based medicine information. We have developed a web-based tool, the Florey Dementia Risk Score (FDRS), which has achieved a classification accuracy of 80–87% for older Australians (n = 60) and Americans (n = 500).

Implications of all the available evidence

The developed FDRS is an easy-to-use, machine learning-based tool for the binary classification of AD. By utilizing demographic information, medical history, self-report data, vital signs, and apolipoprotein E genotype data, the FDRS offers a new digital health technology with the potential to improve AD detection rate in primary care settings, especially where diagnostic resources are limited. Further trial is required to validate FDRS in low- and medium-income countries and evaluate its potential to promote health equity by facilitating AD screening in primary care.

was 61.7% [95% CI 55.0–68.0%], and the rate of under-detection was higher in China and India (versus North America and Europe), and in the community setting (versus residential/nursing care).[2] In clinical practice, the diagnosis of AD is based primarily on clinical symptoms, supplemented by neuropsychological tests, advanced imaging and, where possible, biomarkers.[3] PET imaging is the gold standard for quantifying amyloid-beta, the hallmark of AD; however, it is rarely used in the clinic. The technical complexity and high costs of these approaches pose significant challenges to screening and early detection of AD in primary care, especially in low-income countries or remote areas where diagnostic resources are even more limited.[4] In addition, stigma associated with neuropsychological tests may deter some individuals from undergoing the required tests for AD diagnosis.[5,6]

Several digital tools and risk indices have been developed to assess the probability/risk of individuals currently having AD.[7–9] Using data from the Northern Manhattan study, Reitz et al. integrated factors such as sex, education, ethnicity, apolipoprotein E (APOE) genotype, diabetes, hypertension, smoking, high-density lipoprotein, and waist-to-hip ratio to develop a vascular risk score.[7] Although this score helps identify older adults who might be at risk for AD, their output (relative risk) is difficult to interpret clinically, and more importantly, the accuracy of the model was not assessed. Another tool was developed using the German primary care patient registry, tracking 3055 patients across three follow-ups with an 18-month interval. This tool requires age, subjective memory complaints, Mini-Mental State Examination score, depressive symptoms, and instrumental activities of daily living as features of AD and achieved a prediction accuracy of 0.79.[8] However, neuropsychologists need to be involved to use this tool, making it less practical for AD screening in the primary care settings. In addition, the Australian National University AD Risk Index (ANU-ADRI) is a self-reported risk index that uses an evidence-based medicine approach to predict AD occurrence.[9] Demographic data (e.g., age), medical history (e.g., depression, diabetes), and lifestyle factors (e.g., physical activity) were collected. The ANU-ADRI has been tested in three independent AD cohort datasets (the Rush Memory and Aging Project, Kungsholmen Project, and Cardiovascular Health Cognition Study), and achieved an accuracy of 0.7.[10] Although these tools have the potential to facilitate early detection of AD, their practicability and performance require further improvement.

Here, we developed and validated a new web-based diagnostic tool, the Florey Dementia Risk Score (FDRS), which was designed for AD screening in older adults (≥65 years) during their primary care, with clinicians as the intended users. FDRS was developed by leveraging the Australian Imaging, Biomarker & Lifestyle (AIBL) study[11] and was powered by an auto-score machine learning algorithm.[12] Our tool only requires relatively easy-to-collect data, such as age, living arrangement, occupation, heart rate, blood pressure, self-reported difficulty in memory, history of neurological disorders, Geriatric Depression Scale (GDS) score,

and APOE genotype. The performance of FDRS is promising in older Australians, demonstrating a high consistency between FDRS results and clinical diagnoses. We also evaluated FDRS using participants of the Religious Orders Study and Memory and Aging Project (ROSMAP),[13–16] and a consistently good performance was achieved.

Methods

Study design and participants

This study follows the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guideline, which is used to aid the reporting of studies developing prediction models for diagnosis or prognosis using machine learning or regression methods and/or evaluating (validating) their performance.[17] This study is not a clinical trial and therefore was not registered. The AIBL dataset (n = 1407) was used to construct the FDRS model (70% of the participants) and internally evaluated the model (30% of the participants). Additional AIBL participants data (unseen by the model) were used for a simulated trial (n = 30) and a missing value trial (n = 30). The FDRS was also externally evaluated using 500 participants of ROSMAP study. AIBL and ROSMAP study were conducted in Australia (since 2006) and the United States (since 1994), which are two of the most well characterized observational dementia cohorts in the world.

The current study included five data categories as predictors: demographics (D), medical history (M), self-reports (S), vital signs (V), and APOE genotype (G). All predictor data were collected/measured by AIBL and ROSMAP research groups as per their published method.[11,13] Two sets of features, DMSVG and DMSV, were used for model construction. The demographic data included age, sex, living arrangements, primary occupation, retirement status, and marital status. For medical history, the AIBL study employed a questionnaire covering 22 diseases, and each was represented as a binary variable. Self-reported data included the GDS[18] and difficulty in memory. The vital signs included blood pressure and heart rate. The APOE genotypes included ε2/ε2, ε2/ε3, ε3/ε3, ε4/ε2, ε4/ε3, and ε4/ε4.[19] These data can be collected using non- or less invasive approaches/procedures,[20] and therefore they are selected as initial predictors for FDRS.

Out of a total of 2449 AIBL participants, 1407 had all the DMSVG data recorded and were therefore included in the current study. This size of participant data has been successfully used for the development and validation of existing machine learning models for AD.[20] The counts for each feature for these 1407 AIBL participants are listed in Supplementary Materials (A), Table S1. An additional 60 participants were recruited from AIBL for the simulated trials: 30 participants with full DMSVG information and 30 participants with one or two pieces of missing data. To test the generalizability of the FDRS, 500 ROSMAP participants (375 CU and 125 AD) with complete DMSVG information were used for external evaluation. For participants with multiple records at different ages after enrolment, only the last records were included. This approach was used to mitigate correlations between time series of multiple assessments for an individual participant[21] and to avoid imbalanced occurrences of AD and non-AD cases in subsequent analyses.[22] The use of de-identified human data from these participants was consented.

FDRS model development

Details of the model development including exact data handling steps are available in Supplementary Materials (B). Briefly, the development of the FDRS is based on an algorithm and software package called auto-score, which introduces a framework for automating the development of a clinical scoring model for predefined outcomes and systematically presents its structure.[12] The auto-score algorithm comprises six modules: 1) variable ranking by random forest, 2) variable transformation, 3) score derivation, 4) parameter determination by parsimony plot, 5) fine-tuning, and 6) predictive performance evaluation. The first module ranks features by their importance, helping to select relevant features for the development of FDRS. The second module converts continuous variables into categorical ones, allowing for the modeling of nonlinear effects. This approach has been widely used in medical research and can reduce the impact of outliers on the performance. The third module uses multi-logistic regression to create a risk score for outcome prediction, assigning a weight and an integer point value for the categories within each feature. The fourth module employs parsimony analysis to balance model performance and complexity, which is used to determine the final number of features to be used by FDRS. Since the variable transformation in module two is data-driven and lacks domain knowledge, the fifth module fine-tunes the automatically generated interval boundary values for continuous variables by combining, rounding, and aligning them with standard clinical norms (e.g., clinical guidelines). The threshold for the binary classification was thereafter determined by the auto-score algorithm. Finally, the sixth module assesses the developed FDRS on an unseen test set to evaluate its performance after the previous modules. For the current study, the AIBL dataset was randomly divided into a non-overlapping training set (70%) and a test set (30%). The training set was utilized for training, development, and fine-tuning of the FDRS (module 1–5), while the test set served as unseen data to evaluate the performance of the developed FDRS (module 6).

FDRS construction and evaluation

As mentioned above, the data were categorized into five categories, DMSVG. As APOE genotype data (G) may not be readily available for the target users, we have also developed FDRS model using DMSV data (i.e. a model without the use of APOE genotype). The FDRS-DMSV was compared to the standard FDRS model, to evaluate the impact of missing APOE genotype on model performance. We have also developed two baseline models using multi-logistic regression with 1) all features in the AIBL dataset and 2) the selected features (using module 1 and 4), to demonstrate the advantages of using feature selection and auto-score algorithm, respectively. The receiver operating characteristic (ROC) analysis was chosen as the evaluative metric, incorporating several key performance indicators, including area under the curve (AUC) of ROC, sensitivity, and specificity.

Web-based tool development

The FDRS model developed in this study has been embedded into a user-friendly, self-report web-based tool through a co-design process[23] involving health consumers and clinicians. The co-design was held virtually and physically via the Victorian Co-design Research Hub. This tool prompts clinicians to input data into several specifically chosen features. Once the required information is entered, the tool calculates and displays the FDRS with binary classification of AD.

Simulated pilot trial, missing data trial, and external evaluation

To appreciate the real-world applicability of our web-based tool, we randomly selected 30 older Australians from the AIBL study. In addition, as missing data is common in clinical settings,[20] we evaluated its impact on the performance of FDRS using another simulated trial with 30 older Australians with one or two pieces of missing data. For imputation, we attempted three imputation methods for the respective feature values, including 1) mean substitution, 2) the k-nearest neighbor, and 3) multiple imputation by chain equation. Finally, FDRS was externally evaluated on 500 (125 AD, 375 CU) participants of a United State-based cohort study, the ROSMAP, to ensure model generalizability. All participants were randomly selected, and their data had not been previously used in the construction and validation of the FDRS model. The FDRS results were calculated by auto-score algorithm, which is handled by a researcher (GJ) blinded to the participants' clinical diagnoses. The FDRS results of trials/external evaluation were compared against clinical diagnoses as a measure of accuracy.

Software and packages

All data preprocessing and analyses were performed using Python version 3.9 and RStudio version 12.0 + 369. The developed model relies on the auto-score package in the R 3.5.3 programming environment (R Foundation).[24] The *auto-score* package enables the convenient creation of point-based clinical scoring models to predict outcomes, minimizing manual intervention for data processing, parameter tuning, and model optimization.

Ethics statement

The AIBL study was approved by the St Vincent's Health Melbourne Human Research Ethics Committee (HREC Reference number: 028/06), and all participants provided written informed consent prior to study enrolment.[11,25] ROSMAP was approved by the Institutional Review Board of Rush University Medical Centre, and all participants provided informed consent.[13]

The current study analyzes the de-identified secondary data collected by AIBL and ROSMAP, and informed consent or local ethical committee approval was not required.

Role of the funding source

The funder of the current study had no role in the study design, data collection, data analysis, data interpretation, or writing of the study.

Results

FDRS model construction and internal validation using AIBL dataset

We first performed a random forest feature selection to identify the most to least important features for classification. The feature importance ranking table is shown in Fig. 1, with age standing out as the most important feature for AD binary classification. These results are not unexpected, as age is the greatest known risk factor for AD.[26] Blood pressure (systolic and diastolic) and heart rate are also strongly associated with AD. This is also in line with existing clinical and epidemiological evidence.[27,28] In addition, the GDS score, living arrangements, APOE genotype, and history of neurological disorders (other than AD) are also highlighted as important features, which are well supported by clinical and epidemiological observation.[29] Overall, the features selected via the machine learning approach are consistent with the evidence-based medicine.

An ideal FDRS should balance the number of features with its performance. A parsimony plot (Fig. 2) was therefore employed to determine the number of features required by FDRS to achieve the most optimal performance. We noted a dramatic increase in performance when comparing models using two-five features (AUC 0.69–0.71) to those using age as a single feature (AUC 0.59). Notably, incorporating the APOE genotype (the sixth feature) increased the AUC by 0.06. The performance was improved further by using more than six features (AUC 0.75–0.80) and peaked at eleven features

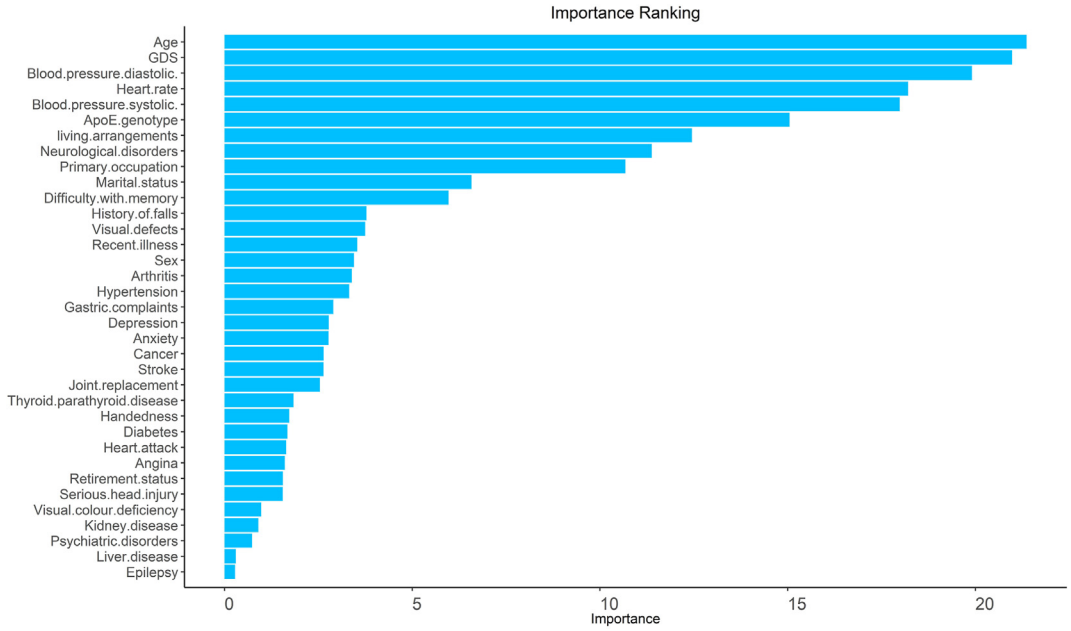

Fig. 1: **Importance ranking of features**. The x-axis displays the importance determined by the random forest feature selection, with a higher numerical value indicating higher importance, while the y-axis lists the name of 35 features collected in the Australian Imaging, Biomarker & Lifestyle study. The plot illustrates the relative importance of various features. Age is the most important feature, while epilepsy is the least important one. *Abbreviations: Geriatric Depression Scale (GDS), apolipoprotein E (APOE).*

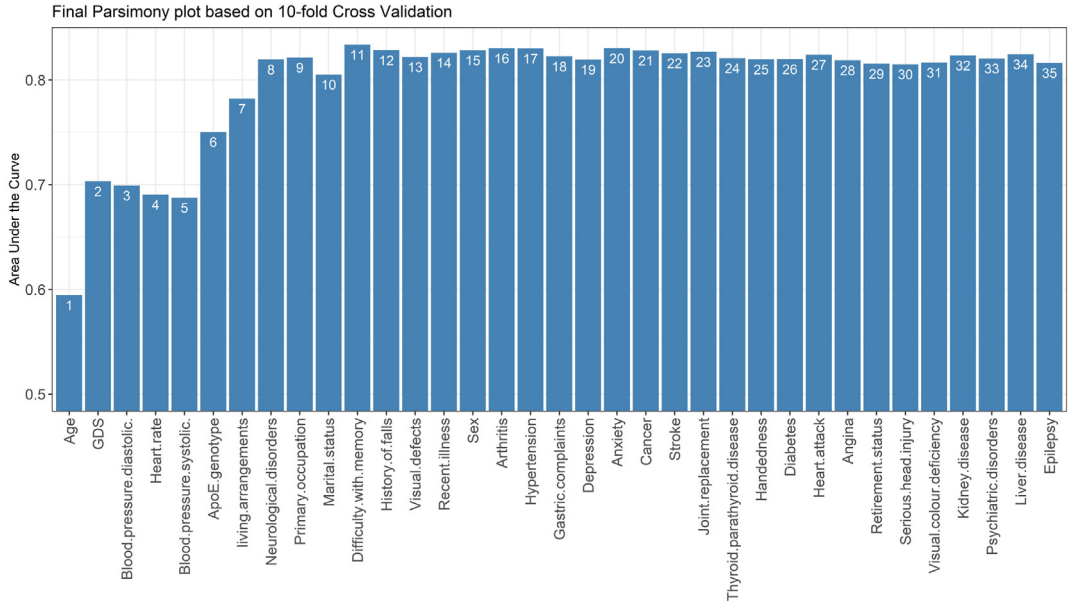

Fig. 2: **Parsimony plot for the Florey Dementia Risk Score (FDRS) model using a cumulative number of features**. This plot was obtained using the training set, which shows the area under curve (AUC) values when increasing number of features are used in the FDRS model. The number within the bar represents the total number of features used, and the height of the bar indicates the mean AUC value from a 10-fold cross validation. For example, the third bar on left means when the first three features are used for FDRS, the AUC is ~0.7 as indicated on the y-axis. *Abbreviations: Geriatric Depression Scale (GDS), apolipoprotein E (APOE).*

(AUC 0.83). Thereafter, the performance remained steady despite adding additional features (AUC 0.80–0.83). Balancing performance and model complexity, we decided to use eleven features for our FDRS model.

After feature selection, multiple logistic regression was used to derive the score for each category/interval of the selected features (Table 1). The maximum FDRS is 100. The cut-off score for AD binary classification (i.e. FDRS = 76) was automatically determined by the auto-score algorithm. The performance metrics on the validation set were as follows: AUC 0.88 [95% CI, 0.85–0.91], sensitivity 0.79 [0.71–0.85], and specificity 0.82 [0.80–0.85]. After that, we manually fine-tuned the interval to ensure clinical relevancy. For example, the GDS score was fine-tuned with clinically relevant ranges [0,5), [5,10), [10,15].[18] For diastolic blood pressure, the intervals are <60, [60,80), [80,90), and ≥90, while for systolic blood pressure, the intervals are <120, [90,150), and ≥150.[30] The updated scores for the selected features after fine tuning and the clinical guidelines used to determine these ranges/intervals are listed in Table 2. The performance of the fine-tuned FDRS model on the validation set was as follows: AUC 0.88 [95% CI, 0.84–0.91], sensitivity 0.74 [0.70–0.81], and specificity 0.84 [0.82–0.87], which was comparable to that of the FDRS without fine tuning. However, incorporating fine-tuning is essential as it integrates clinical information for score derivation, thereby avoiding non-clinically meaningful intervals for some features (e.g., GDS < 0). After fine tuning, the FDRS was evaluated on the test set, and the FDRS cutoff score decreased to 66. This reduction in cutoff score is likely due to fine tuning. The performance of the fine-tuned FDRS model on the test set was as follows: AUC 0.82 [95% CI, 0.75–0.88], sensitivity 0.74 [0.60–0.86], and specificity 0.73 [0.70–0.79]. Overall, the results indicated that FDRS can distinguish CU and AD subjects.

Features	Category/interval	Scores
Age	<77	0
	[77,82)	2
	≥82	3
Geriatric Depression Scale	<0	0
	[0,2)	28
	≥2	31
Blood pressure (diastolic), mm Hg	<79	0
	≥79	1
Heart rate, beats/min	<62	1
	[62,68)	0
	[68,76)	1
	≥76	0
Blood pressure (systolic), mm Hg	<151	2
	≥151	0
Apolipoprotein E genotype	ε2/ε2	0
	ε3/ε2	32
	ε3/ε3	33
	ε4/ε2	34
	ε4/ε3	36
	ε4/ε4	39
Neurological disorders (other than AD)	No	0
	Yes	5
Present living arrangements	Home of Relative	6
	Other	2
	Own (or rented) home alone	0
	Own (or rented) home with spouse/others	3
	Residential Hostel	7
Primary occupation	Clerical/Teaching/Nursing	0
	Domestic Duties/Factory/ Agriculture	2
	Other	1
Difficulty with memory	No	0
	Yes	4
Marital status	Cohabiting	0
	Divorced/Single/Windowed	1
	Married	2
	Separated	4

Table 1: **Score table of features for the Florey Dementia Risk Score model.**

Comparison between the standard FDRS and baseline models

The developed standard FDRS was compared with two baseline models: (1) multi-logistic regression with all 35 features, and (2) multi-logistic regression with the same 11 features used by the standard FDRS. The comparison results have been summarized in Supplementary Materials (C), Table S2. The ROC plots of these two baseline models are shown in Supplementary Materials (C), Figure S1(A) and 1(B). From these results, we can see that compared to baseline model-1 (AUC 0.738 [95% CI, 0.657–0.820]), although the FDRS used fewer features, it achieved a better performance. Compared to baseline model-2 (AUC 0.728 [0.646–0.810]), the FDRS employing auto-score algorithm performed better, although the same eleven features were used. Overall, these comparisons demonstrated that advantages of using the auto-score algorithm in the binary classification of AD.

Comparison between FDRS-DMSV model with the standard FDRS

Similar to the development of the standard FDRS, the development process for FDRS-DMSV includes six modules, which are detailed in Supplementary Materials (D). In Module 1, random forest feature selection was used to rank features by their importance (Figure S2). In Module 4, a parsimony plot was used to determine the number of features required by the FDRS-DMSV (Figure S3). When 15 features were used, the model

Features	Interval	Scores
Age	<73	0
	[73,77)	1
	[77,82)	3
	≥82	5
Geriatric depression scale[18]	[0,5)	0
	[5,10)	5
	[10,15)	4
Blood pressure (diastolic),[30] mm Hg	<60	4
	[60,90)	0
	≥90	3
Heart rate, beats/min	<62	2
	[62,68)	1
	[68,76)	2
	≥76	0
Blood pressure (systolic),[30] mm Hg	<120	2
	[120,150)	3
	≥150	0
Apolipoprotein E genotype	ε2/ε2	0
	ε3/ε2	42
	ε3/ε3	43
	ε4/ε2	45
	ε4/ε3	48
	ε4/ε4	52
Neurological disorders (other than AD)	No	0
	Yes	7
Present living arrangements	Home of Relative	7
	Other	3
	Own (or rented) home alone	0
	Own (or rented) home with spouse/others	4
	Residential Hostel	10
Primary occupation	Clerical	0
	Domestic Duties Factory/ Agriculture	3
	Other	1
	Teaching/Nursing	0
Difficulty with memory	No	0
	Yes	6
Marital status	Cohabiting	0
	Divorced/Single/Windowed	1
	Married	2
	Separated	5

Table 2: **Score table of fine-tuned features for the Florey Dementia Risk Score model.**

achieved an AUC of 0.80 for the ROC, and there was no evident improvement in performance when additional features were added. Therefore, 15 features were used for FDRS-DMSV. The score table of the selected features is shown in Table S3. The FDRS-DMSV achieved an AUC of ROC = 0.84 [95% CI 0.80–0.87], sensitivity = 0.73 [0.66–0.80], and specificity = 0.81 [0.78–0.83] on the validation set. In Module 5, fine tuning was performed for clinical relevancy, and the revised score table for the fine-tuned features is provided in Table S4. The threshold for the FDRS-DMSV for the AD binary classification is 37. FDRS-DMSV achieved an AUC of ROC = 0.82 [95% CI: 0.78–0.86], sensitivity = 0.78 [0.71–0.84], and specificity = 0.74 [0.71–0.76] on the validation set. Finally, the FDRS-DMSV was evaluated on an unseen test set and achieved an AUC of ROC = 0.78 [95% CI: 0.73–0.84], sensitivity = 0.80 [0.70–0.88], and specificity = 0.65 [0.60–0.70].

Simulated pilot trial and missing data trial using the web-based FDRS tool

We have developed a web-based tool based on the FDRS model (a demo video is available in Supplementary Materials (E)). The tool automatically computes the FDRS after the required information is entered. The FDRS framework architecture is graphically presented in Figure S4. The simulated pilot trial was conducted using the web-based tool. Thirty AIBL participants with complete records of all eleven required features were selected. The data of these participants had not been previously used for FDRS model construction and validation. Their FDRS scores were calculated via the web-based tool and are presented in Supplement Materials (F). This trial achieved an accuracy of ~86.7%, with 26/30 correct classification overall, 20/22 correct classification for CU, and 6/8 correct classification for AD (eTable 1). In addition, a simulated pilot trial was conducted for another thirty participants, with one or two missing data for the selected features. Only a slight decrease in accuracy (80.0–83.3%, 24–25/30 correct overall depending on imputation methods) was noted (eTable 2). Overall, the simulated trial results support the potential clinical application of FDRS in the AD binary classification and demonstrate its tolerability for missing data.

External evaluation of FDRS on ROSMAP study participants

The ROC plot for the external evaluation among the 500 selected ROSMAP participants (eTable 3) is shown in Fig. 3. The AUC of the ROC curve was 0.82 [95% CI, 0.77–0.86], similar to the performance on the test set in AIBL. Moreover, the distribution of the FDRS of the 500 participants in ROSMAP is presented in Fig. 4. It is evident that the FDRS for most CU participants (84.5%, 317/375) is below the cutoff score of 66, while the FDRS for most AD participants (72.8%, 91/125) is above the cutoff. Overall, these results demonstrate a good performance of FDRS in an independent non-Australian cohort, supporting the generalizability of FDRS.

Discussion

More than half of people living with dementia in the community are not detected or diagnosed, as many

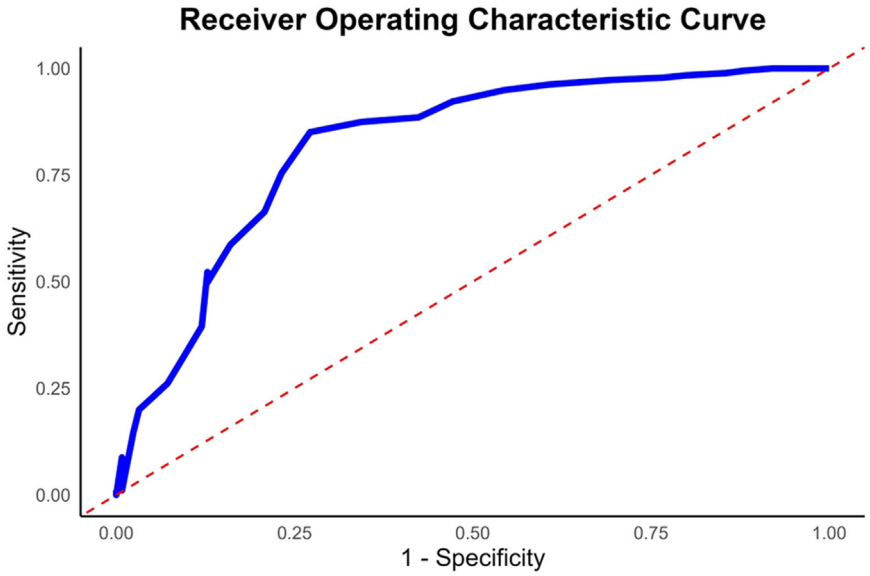

***Fig. 3:* Receiver operating characteristic (ROC) plot on the Religious Orders Study and Memory and Aging Project dataset**. The ROC curve illustrates the trade-off between sensitivity (true positive rate) and specificity (1–false positive rate) for different threshold settings. The blue line is the ROC curve when the cutoff score is 66 for Alzheimer's disease binary classification. The red dashed line represents random guessing for Alzheimer's disease binary classification.

older adults experiencing memory decline may consider it part of the normal aging process. This issue is even more concerning in low- and middle-income countries,[2] where access to diagnostic resources are limited. To improve AD detection rate and address health inequity, we have developed a web-based tool- FDRS, powered by an auto-score algorithm to make an AD binary classification (CU versus AD). This algorithm is composed of six modules. We employed a random forest method for feature ranking (module 1),[31] which allow us to remove unimportant features (Fig. 1). Parsimony analysis (module 4) was used to determine the number of

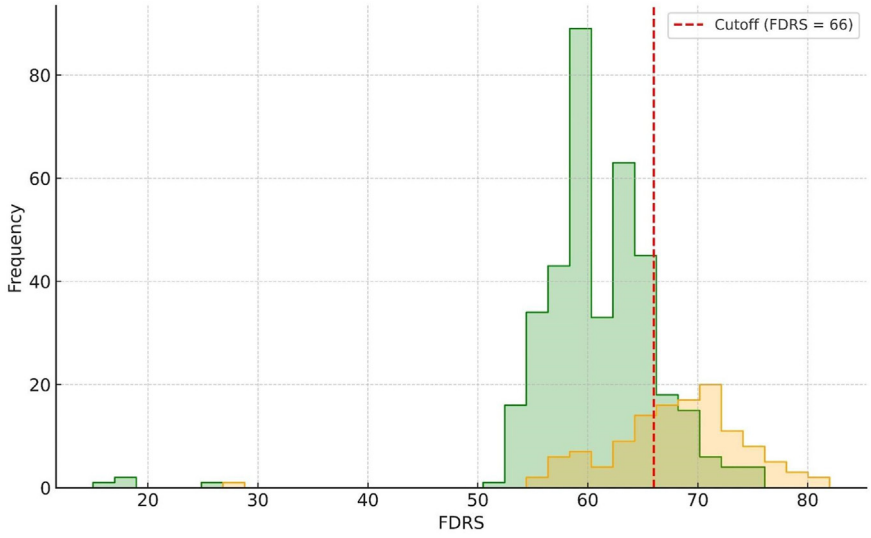

***Fig. 4:* Histogram of frequency distribution of Florey Dementia Risk Score (FDRS) for participants in Religious Orders Study and Memory and Aging Project (ROSMAP)**. The plot shows the distribution of FDRS scores of 500 ROSMAP study participants, with a cutoff value of 66 indicated by the red dashed line separating cognitive unimpaired (CU, left) and Alzheimer's disease (AD, right). The clinical diagnoses of the participants are color-coded, with green for CU and yellow for AD. It appears that most of the CU and AD subjects can be correctly classified.

features required for the optimal performance of FDRS (Fig. 2), balancing model complexity and accuracy.[32] Although using only easy-to-collect data such as age, blood pressure, heart rate, medical history, and APOE genotype, FDRS achieved an AUC of 0.82, a sensitivity of 0.74, and a specificity of 0.73.

As per feature ranking and parsimony analysis, APOE genotype was selected as a feature for FDRS. Diverted scores for various APOE genotypes (ε2/ε2, ε2/ε3, ε3/ε3, ε4/ε2, ε4/ε3, and ε4/ε4) were calculated and shown in Tables 1 and 2. Notably, the APOE genotype is the major contributor to the FDRS score. The ε2 allele is associated with a lower score, while the ε4 allele is associated with a higher score. In addition, the homozygous ε4 genotype had the highest score compared to other APOE genotypes. This observed score pattern is well aligned with clinical findings that the ε3 allele is considered the norm genotype, while the ε2 allele and the ε4 allele decreases and increases the risk of AD, respectively.[33] This is also consistent with a recent study reporting that homozygous APOE ε4 individuals are destined to develop AD.[34]

It must be acknowledged that APOE genotype is not always known by the target users of FDRS, and therefore, we have also developed an FDRS model without the use of APOE genotype (FDRS-DMSV) (Supplementary Materials (D)). The performance of the FDRS-DMSV was slightly lower than the standard FDRS, with an AUC of 0.77, a sensitivity of 0.67, and a specificity of 0.75. The FDRS-DMSV requires five additional features (history of anxiety, arthritis, hypertension, sex, and recent illness) as informed by the parsimony analysis (Figure S3), to compensate for the lack of APOE genotype data. Interestingly, the association of these additional features with AD has been previously demonstrated by our laboratory and others. For example, anxiety is associated with a higher risk of AD,[35–37] which is reflected by a higher FDRS score for people with anxiety. Arthritis is associated with a lower risk of AD (possibly due to the use of non-steroid anti-inflammatory drugs),[38] which is reflected by a lower FDRS score for people with arthritis.[29] The link between hypertension and AD has also been well studied.[39]

The performance of the FDRS model can be influenced by several potential errors and biases. For instance, in the AIBL study, some data for selected features were self-reported, which can introduce recall bias. Additionally, the quality of the collected data may be affected by the design of the questionnaire used in the AIBL study. However, we anticipate that these errors and biases have minimal impact on the results and conclusions, as consistently good performance was observed in the external evaluation on ROSMAP participants. Regarding the missing data trial, the model's performance may be influenced by the different imputation methods used. We compared three methods: mean imputation, k-nearest neighbor, and multiple imputation by chained equations. Interestingly, the performance of the FDRS did not differ significantly between these imputation methods. Whether the impact will become more apparent in a larger trial cohort requires further investigation in future studies.

The FDRS outperformed the baseline models and three existing models for binary AD classification. By comparing the FDRS with baseline model-1, we can clearly see that feature selection was effective, as the FDRS achieved a higher AUC while using fewer features. Comparing the FDRS with baseline model-2, we observe that, while using the same features, the auto-score algorithm achieved a higher AUC than multi-logistic regression. Considering the existing models for the same classification task, the ANU-ADRI[10] used comparable types of features as the FDRS, but it achieved a lower AUC of 0.7. Prediction Score[8] achieved an AUC of 0.79, which is also lower than the FDRS. In addition, it requires neuropsychologists for a Mini-Mental State Examination, which would be a barrier to its widespread application in communities with limited diagnostic resources. We cannot compare FDRS with vascular risk score,[7] as the AUC was not reported. Overall, these comparisons highlight the advantages of FDRS over other tools. The advancement of the FDRS is due to the incorporation of innovative computational strategy (i.e. auto-score machine learning algorithm) into the model development, as other tools were developed solely using principle of evidence-based medicine or biostatistics/epidemiological approaches.

The FDRS is not without limitation. It was trained, validated, and internally tested using data from the AIBL Study, and it is known that the AIBL cohort consists predominantly of highly educated Australians, and strict inclusion/exclusion criteria were applied to participant recruitment.[11,40] This can introduce bias into the model, which leads to a lower classification accuracy when the model is applied to a different population. The generalizability of our FDRS to other countries with different ethnicities, especially low- and middle-income countries, needs to be further investigated. In addition, although the external evaluation of the FDRS on the ROSMAP participants achieved a high accuracy of 81.6% (417/500), comparable to the pilot trial using AIBL participants (86.7%, 26/30), additional external evaluations in other developed countries are desirable. Regardless, the results achieved in this study warrant its further development and validation using other AD cohort datasets and larger-scale trials, and the performance of FDRS should also be directly compared with other models using the same dataset interact in the handling of the input data.

The developed FDRS is an easy-to-use, machine learning-based tool for the binary classification of AD. The intended users do not need to have machine learning expertise to use our web-based tool. By utilizing demographic, medical history, self-report data, vital sign, and

genetic data, FDRS offers a new digital health technology to potentially improve AD detection rate in primary care, that can be used when the diagnostic resources are limited. FDRS is likely to promote health equity and contribute significantly to public health efforts against AD via disease screening in the primary care setting.

Contributors

YP and CC have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. LJ and YW (Yihan Wang) accessed and verified the data. CC, BG, LJ, and YP contributed to the study conceptualization. CLM contributed resources (AIBL data). CC, YFW (Yifei Wang), and YW performed formal analysis. YP was responsible for funding acquisition. BG provided advice on study methodology. YP, LJ, and BG were responsible for project administration and supervision. GJ was responsible for engaging health consumers and clinicians for research co-design, and performed the diagnosis-blinded web-based simulated trial. YP and CC drafted the manuscript, and all authors contributed to writing, reviewing, and editing.

Data sharing statement

AIBL data and ROSMAP data included in this work is available for research purposes and can be requested, respectively, through the AIBL scientific committee http://www.aibl.org.au/, and the Rush Alzheimer's Disease Center Research Resource Sharing Hub https://www.radc.rush.edu/. The code of FDRS is available via GitHub (https://github.com/auccy/Jack.git).

Declaration of interests

All the authors declare no conflict of interests.

Acknowledgements

This work would not have been possible without the contributions of the participants from AIBL and ROSMAP studies, and their investigators and staff. Data used in preparation of this article were obtained from the Australian Imaging, Biomarker and Lifestyle (AIBL) Study database (https://aibl.org.au/). As such, the investigators within AIBL contributed to the design and implementation of AIBL and/or provided data but did not participate in analysis or writing of this report. A complete listing of AIBL investigators can be found at: https://aibl.org.au/about/our-researchers/. The ROSMAP study received support from the National Institute on Aging (Grant Numbers: P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, U01AG61356). RADC resources can be requested at https://www.radc.rush.edu and www.synpase.org. The FDRS model was digitalized to a web-based tool by Min Shen. Dr Yijun Pan's salary is supported by the National Health and Medical Research Council, Australia (GNT2007912) and Alzheimer's Association, USA (23AARF-1020292).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102834.

References

1 Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of dementia: an observational study of patterns in diagnosis and awareness in US older adults. *J Gen Intern Med*. 2018;33:1131–1138.
2 Lang L, Clifford A, Wei L, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. *BMJ Open*. 2017;7(2):e011146.
3 Porsteinsson AP, Isaacson R, Knox S, Sabbagh MN, Rubino I. Diagnosis of early Alzheimer's disease: clinical practice in 2021. *J Prev Alzheimers Dis*. 2021;8:371–386.
4 Barth J, Nickel F, Kolominsky-Rabas PL. Diagnosis of cognitive decline and dementia in rural areas—a scoping review. *Int J Geriatr Psychiatr*. 2018;33(3):459–474.
5 Garand L, Lingler JH, Conner KO, Dew MA. Diagnostic labels, stigma, and participation in research related to dementia and mild cognitive impairment. *Res Gerontol Nurs*. 2009;2(2):112–121.
6 Kerwin D, Abdelnour C, Caramelli P, et al. Alzheimer's disease diagnosis and management: perspectives from around the world. *Alzheimer's Dementia*. 2022;14(1):e12334.
7 Reitz C, Tang M-X, Schupf N, Manly JJ, Mayeux R, Luchsinger JA. A summary risk score for the prediction of Alzheimer disease in elderly persons. *Arch Neurol*. 2010;67(7):835–841.
8 Jessen F, Wiese B, Bickel H, et al. Prediction of dementia in primary care patients. *PLoS One*. 2011;6(2):e16852.
9 Anstey KJ, Cherbuin N, Herath PM. Development of a new method for assessing global risk of Alzheimer's disease for use in population health approaches to prevention. *Prev Sci*. 2013;14:411–421.
10 Anstey KJ, Cherbuin N, Herath PM, et al. A self-report risk index to predict occurrence of dementia in three independent cohorts of older adults: the ANU-ADRI. *PLoS One*. 2014;9(1):e86141.
11 Fowler C, Rainey-Smith SR, Bird S, et al. Fifteen years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer's disease. *J Alzheimers Dis Rep*. 2021;5(1):443–468.
12 Xie F, Chakraborty B, Ong MEH, Goldstein BA, Liu N. AutoScore: a machine learning–based automatic clinical score generator and its application to mortality prediction using electronic health records. *JMIR Med Inform*. 2020;8(10):e21798.
13 Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious orders study and rush memory and aging project. *J Alzheim Dis*. 2018;64(s1):S161–S189.
14 L Barnes L, C Shah R, T Aggarwal N, A Bennett D, A Schneider J. The Minority Aging Research Study: ongoing efforts to obtain brain donation in African Americans without dementia. *Curr Alzheimer Res*. 2012;9(6):734–745.
15 Schneider JA, Aggarwal NT, Barnes L, Boyle P, Bennett DA. The neuropathology of older persons with and without dementia from community versus clinic cohorts. *J Alzheim Dis*. 2009;18(3):691–701.
16 Marquez DX, Glover CM, Lamar M, et al. Representation of older latinxs in cohort studies at the rush Alzheimer's disease center. *Neuroepidemiology*. 2020;54(5):404–418.
17 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
18 Greenberg SA. The geriatric depression scale (GDS). *Best Pract Nurs Care Older Adul*. 2012;4(1):1–2.
19 Knopman DS, Amieva H, Petersen RC, et al. Alzheimer disease. *Nat Rev Dis Prim*. 2021;7(1):33.
20 Wang Y, Liu S, Spiteri AG, et al. Understanding machine learning applications in dementia research and clinical practice: a review for biomedical scientists and clinicians. *Alzheimers Res Ther*. 2024;16(1):175.
21 Sayal M. *Detecting time correlations in time-series data streams*. vol. 122004;vol. 12. Hewlett-Packard Company; 2004.
22 Liu L, Wu X, Li S, Li Y, Tan S, Bai Y. Solving the class imbalance problem using ensemble algorithm: application of screening for aortic dissection. *BMC Med Inf Decis Making*. 2022;22(1):82.
23 Bird M, McGillion M, Chambers E, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. *Res Involv Engagem*. 2021;7:1–12.
24 Xie F, Ning Y, Liu M, et al. A universal AutoScore framework to develop interpretable scoring systems for predicting common types of clinical outcomes. *STAR Protoc*. 2023;4(2):102302.
25 Ellis KA, Szoeke C, Bush AI, et al. Rates of diagnostic transition and cognitive change at 18-month follow-up among 1,112 participants in the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL). *Int Psychogeriatr*. 2014;26(4):543–554.
26 Masters CL. Major risk factors for Alzheimer's disease: age and genetics. *Lancet Neurol*. 2020;19(6):475–476.
27 Saiz-Vazquez O, Puente-Martinez A, Pacheco-Bonrostro J, Ubillos-Landa S. Blood pressure and Alzheimer's disease: a review of meta-analysis. *Front Neurol*. 2023;13:1065335.
28 Kim H-B, Jung YH, Han HJ. Resting heart rate and cognitive decline: a meta-analysis of prospective cohort studies. *J Clin Neurol*. 2022;18(6):619.
29 Nguyen CQN, Ma L, Low YLC, et al. Exploring the link between comorbidities and Alzheimer's dementia in the Australian

Imaging, Biomarker & Lifestyle (AIBL) study. *Alzheimer's Dement*. 2024;16(2):e12593.
30 Qaseem A, Wilt TJ, Rich R, et al. Pharmacologic treatment of hypertension in adults aged 60 years or older to higher versus lower blood pressure targets: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Ann Intern Med*. 2017;166(6):430–437.
31 Kuhn M, Johnson K, Kuhn M, Johnson K. *An introduction to feature selection*. Applied Predictive Modeling; 2013:487–519.
32 Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for clinical predictive modeling. *Int J Med Inf*. 2018;116:10–17.
33 van Duijn CM, de Knijff P, Cruts M, et al. Apolipoprotein E4 allele in a population–based study of early–onset Alzheimer's disease. *Nat Genet*. 1994;7(1):74–78.
34 Fortea J, Pegueroles J, Alcolea D, et al. APOE4 homozygozity represents a distinct genetic form of Alzheimer's disease. *Nat Med*. 2024:1–8.
35 Mendez MF. The relationship between anxiety and Alzheimer's disease. *J Alzheimers Dis Rep*. 2021;5(1):171–177.
36 Pentkowski NS, Rogge-Obando KK, Donaldson TN, Bouquin SJ, Clark BJ. Anxiety and Alzheimer's disease: behavioral analysis and neural basis in rodent models of Alzheimer's-related neuropathology. *Neurosci Biobehav Rev*. 2021;127:647–658.
37 Ma L, Tan EC, Bush AI, et al. Elucidating the link between anxiety/depression and Alzheimer's dementia in the Australian imaging biomarkers and lifestyle (AIBL) study. *J Epidemiol Glob Health*. 2024:1–12.
38 Kukar T, Golde TE. Possible mechanisms of action of NSAIDs and related compounds that modulate γ-secretase cleavage. *Curr Top Med Chem*. 2008;8(1):47–53.
39 Lennon MJ, Makkar SR, Crawford JD, Sachdev PS. Midlife hypertension and Alzheimer's disease: a systematic review and meta-analysis. *J Alzheim Dis*. 2019;71(1):307–316.
40 Huynh ALHWY, Ma L, Low YLC, et al. A comparison of an Australian observational longitudinal Alzheimer's disease cohort to community-based Australian data. *J Alzheim Dis*. 2024;101(3).