Alzheimer's Research & Therapy

RESEARCH

Open Access

Independent validation and outlier analysis of EuroPOND alzheimer's disease staging model using ADNI and real-world clinical data

Mandy M. J. Wittens[1,2,3], Diana M. Sima[4], Arne Brys[4], Hanne Struyfs[1], Ellis Niemantsverdriet[1], Ellen De Roeck[1,5], Christine Bastin[6,7], Florence Benoit[8], Bruno Bergmans[9], Jean-Christophe Bier[10], Peter Paul de Deyn[11,12], Olivier Deryck[28], Bernard Hanseeuw[13,14,15], Adrian Ivanoiu[14,15], Gaëtane Picard[16], Eric Salmon[17], Kurt Segers[19], Anne Sieben[20,21,22], Evert Thiery[18], Jos Tournoy[23,24], Anne-Marie van Binst[25], Jan Versijpt[2,3], Dirk Smeets[4], Maria Bjerke[1,3,26], Maura Bellio[27], Neil P. Oxtoby[27], Daniel C. Alexander[27], Annemie Ribbens[4] and Sebastiaan Engelborghs[1,2,3*]

Abstract

Background Event-based modeling (EBM) traces sequential progression of events in complex processes like neurodegenerative diseases, adept at handling uncertainties. This study validated an EBM for Alzheimer's disease (AD) staging designed by EuroPOND, an EU-funded Horizon 2020 project, using research and real-world datasets, a crucial step towards application in multi-center trials.

Methods The training dataset comprised 1737 subjects from ADNI-1/GO/2, using the EuroPOND EBM toolbox. Testing datasets included a research cohort from University of Antwerp (controls, CN ($n=46$), subjective cognitive decline, SCD ($n=10$), mild cognitive impairment, MCI ($n=47$), AD dementia, ADD ($n=16$)) and a real-world cohort from 9 Belgian Dementia Council memory clinics (CN ($n=91$), SCD ($n=66$), (non-amnestic) naMCI ($n=54$), aMCI ($n=255$), and ADD ($n=220$). Biomarkers included: 2 clinical scores (Mini Mental State Examination (MMSE), Rey Auditory Verbal Learning Test (RAVLT)); 3 CSF-biomarkers ($A\beta_{1-42}$, P-tau_{181}, total-Tau); and 4 magnetic resonance imaging (MRI) biomarkers (volumes of the hippocampi, temporal, parietal, and frontal cortices) computed with icobrain dm. The naMCI and aMCI groups were compared by EBM stage proportions, and the model's effectiveness at patient level was evaluated.

Results The research cohort's maximum likelihood event sequence comprised CSF $A\beta_{1-42}$, P-tau_{181}, T-tau, RAVLT, MMSE, and cortical volumes. The clinical cohort's order was frontal cortex volume, MMSE, and remaining cortical regions. aMCI subjects showed higher staging than naMCI, with 54% in the two most advanced stages compared to 38% in naMCI. In the research cohort, 10 outliers were identified with potential mismatches between assigned stages and clinical or biomarker profiles, with CN ($n=4$) and SCD ($n=2$) subjects assigned in stage 4, one control in stage 9 with abnormal imaging, and three aMCI cases in stage 0 despite clinical or volumetric signs of impairment.

*Correspondence:
Sebastiaan Engelborghs
Sebastiaan.Engelborghs@vub.be

Full list of author information is available at the end of the article

Conclusions This study highlights the generalizability of EuroPOND's AD EBM model across research and real-world clinical datasets, supporting its use in multi-center trials. aMCI subjects generally reside in more advanced stages than naMCI, who may not necessarily have AD, demonstrating utility for precision recruitment/screening.

Highlights

- EBM for AD staging is generalizable and reliable across research and real-world clinical datasets.
- Amnestic MCI subjects exhibited higher EBM scores than non-amnestic subjects, indicating utility for precision recruitment and screening in clinical trials.
- Tailored EBM models for distinct AD subtypes and diverse neurodegenerative diseases are imperative for accurate staging and understanding varied disease trajectories.

Keywords Alzheimer's disease, Biomarkers, Magnetic resonance imaging, Automated volumetry, Event-based modelling

Introduction

Alzheimer's disease (AD) is a slowly progressive irreversible neurodegenerative brain disorder, which is considered a major global health problem in today's society [1]. AD pathology is characterized by a long-lasting asymptomatic phase, where biomarker abnormalities can occur years, if not decades, prior to symptom onset [2]. The characterization of this preclinical stage is challenging, partly due to non-specific preclinical findings and debates about their relevance in terms of progression and prognosis [3]. The preclinical stage is followed by a prodromal stage exhibiting variable risk and rate of progression, ranging from subtle changes in memory and thinking ability to obvious symptoms of brain dysfunction. The heterogeneity of this pre-dementia stage is a well-known phenomenon that has also been extensively studied, with differences in conversion rates to the AD dementia (ADD) stage between the distinctive subtypes (read. single or multi-domain and amnestic or non-amnestic mild cognitive impairment (MCI)), being of particular interest [4].

To increase understanding on AD onset and progression, there has been a rapid evolution in the development and validation of several biomarkers reflecting underlying AD pathophysiology, including biochemical [5], genetic [6], metabolic [7], neuroanatomical [8] and neuropsychological biomarkers, now pertaining important roles in the detection, diagnosis, prognosis and monitoring of AD. After the hypothetical AD cascade model of Jack et al. [9, 10], a categorical classification scheme based on biomarker positivity was introduced and recently updated [11–13]. It considers early changing biomarkers such as amyloid-PET, $A\beta_{1-42,}$ the $A\beta_{1-42}$/ $A\beta_{1-40}$ ratio and phosphorylated (p)-Tau_{181} in cerebrospinal fluid (CSF), as well as accurate plasma biomarkers such as p-Tau_{217} [14–16]. It also includes later-changing biomarkers such as tau-PET, and biofluid biomarkers, which can offer prognostic insights. This classification provides a biomarker-based standardization for defining and differentiating true AD from non-AD pathology, while simultaneously considering common co-pathologies, cognitive reserve and resistance [11].

Other neurodegenerative brain disorders, such as frontotemporal dementia and Lewy body disease, often share AD symptomatic features and pathology [17], but the severity of change, together with the chronological order in which events occur, can be disease-specific, and thus enable discrimination [18–20]. The overlap with other types of neurodegenerative diseases, heterogeneity of clinical presentation and pathophysiology of AD, together with the lack of consensus between AD specialists regarding clinical terminology across the AD continuum, has put forward a growing need for an unbiased system for characterizing AD stages. Construction of a model that objectively evaluates the current disease state of a patient, using information that is directly extracted from real-world and subsequently determining the correct sequence of biomarker changes, could facilitate earlier and more accurate diagnosis and give more insight in the dynamics and variability of disease progression, a vital model component considering the heterogeneity of AD [21].

Several probabilistic data-driven models have already been proposed to visualize AD staging patterns [21–23]. The scope of data driven modelling is to create a holistic image of the entire disease timeline, staging asymptomatic as well as symptomatic patients and enabling identification of potential therapeutic targets. Event-based modeling (EBM) is a promising data-driven approach for establishing the most likely sequence of events in progressive processes such as neurodegenerative brain diseases, including uncertainty in the sequence, with each biomarker event comprising a significant change in patient state [22, 24]. EBM modeling is based on the principle that more individuals from a cohort containing a disease stage spectrum display data-driven abnormality in biomarkers that change early in disease progression [24, 25]. It is designed as such that a-priori characterization in different clinical groups is not required. A-priori clinical staging often results in a three-stage division,

separating pre-symptomatic, mild cognitive impairment and a dementia stage. It has however been recognized that the concept of cognitive decline in AD is characterized as a long and continuous process, which also applies to the continuous nature of biomarkers measures that already start before clinical manifestations. Regarding the complex, idiopathic and multifactorial nature of AD, meaning there is always a chance that predefined cognitively healthy controls might be preclinical patients, a-priori clinical staging would limit the temporal resolution of the model and would result in a less fine-grained ordering of events [26, 27].

The EuroPOND project (EU Horizon 2020, grant No. 666992) developed several models to track disease progression in neurological diseases. These tools build disease progression signatures, analyze variability within and between populations, and stage individual patients using diverse medical data (e.g., clinical scores, biofluid biomarkers, and imaging). Without requiring (but acknowledging its importance) an estimate of time to onset, disease duration nor disease stage, a discrete EBM model can construct a clear-cut quantitative disease signature that can significantly improve disease understanding, aid in patient selection in clinical trials and be useful with respect to prognostic methods in routine clinical practice. Despite the existence of continuous disease progression models as a possible alternative [28–30], discrete EBM models remain popular in practice due to their simplicity, ability to handle uncertainty and missing data, fewer required parameters, effectiveness with smaller datasets, and alignment with established clinical staging systems [31].

EBMs built on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset obtained characteristic biomarker orderings and demonstrated a good ability to classify cognitively normal and AD subjects [25]. This study aims to assess the validity of the EuroPOND EBM for AD staging, trained on a cross-sectional ADNI dataset. The EBM tool's effectiveness will be assessed in both a research cohort and an independent clinical dataset, with specific objectives to: (1) identify and analyze differences in EBM profiles between amnestic and non-amnestic MCI patients and (2) identify outliers and determine how these findings can be used at patient level, as a step towards using these models in multi-center trials.

Materials and methods

Datasets

Training dataset – ADNI

An EBM of AD progression was trained on a dataset of 1737 baseline records from ADNI-1/GO/2 subjects (mean age in years ± SD; 73.7 ± 7.2, mean Mini Mental State Examination (MMSE) score ± SD; 27.2 ± 2.6), using the EuroPOND EBM toolbox. The ADNI database was launched in the year 2003 by the National Institute on Aging (NIA), Food and Drug Administration (FDA), National Institute of Biomedical imaging and Bioengineering (NIBIB), and other private pharmaceutical companies and non-profit organizations. The main goal of ADNI is to improve and standardize protocols through longitudinal multi-site data collection, intending to define biomarkers for usage in clinical trials and track disease progression through clinical measures, neuroimaging, and chemical biomarkers.

The following diagnostic groups were included: cognitively normal (CN, $n = 417$), subjective cognitive decline (SCD, $n = 106$), early mild cognitive impairment (eMCI, $n = 310$), late MCI (lMCI, $n = 562$) and 342 ADD patients. Categorization in early or late MCI was determined by different levels of impairment according to the Wechsler Memory Scale Logical Memory II test. For early MCI with > 16 years of education (YOE), the assigned score needed to be between 9 and 11 for eMCI's and <-8 for lMCI. Individuals with 8–15 YOE were assigned to a score of 5–9 for eMCI, and <- 4 for lMCI. Finally, for 0–7 YOE, a score of 3–6 was classified as 'early MCI' while a score equal to or below 2 corresponded to lMCI. Clinical scores and MRIs were available in 99.4% of the cases; CSF biomarkers were available in 23% of the cases, but were missing at random across diagnostic groups, therefore the EBM software was able to deal with the missing data.

The following 9 biomarkers were considered based on availability and previously published literature: 2 clinical scores (MMSE and Rey Auditory Verbal Learning Test (RAVLT)); 3 CSF-biomarkers ($A\beta_{1-42}$, P-tau_{181} and total-Tau); and 4 regional magnetic resonance imaging (MRI) biomarkers (volumes of the hippocampi, temporal, parietal and frontal cortices) computed with icobrain dm and normalized for head size. For more information on the ADNI initiative and diagnosis protocols we refer to: http://www.adni-info.org.

Testing dataset – research cohort

A multi-modal AD dataset was acquired from 119 subjects of a memory clinic-based research cohort who participated in a study at the University of Antwerp, Belgium (mean age in years ± SD; 66.9 ± 9.8, mean MMSE score ± SD; 26.6 ± 3.7). According to clinical evaluation, the study population consisted of CN ($n = 46$), SCD ($n = 10$), MCI due to AD ($n = 47$) and ADD ($n = 16$), subjects. The MCI patients consisted of 8 non-amnestic (naMCI) and 39 amnestic (aMCI) subjects. Patient classification was effectuated in compliance with the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria for 'MCI due to AD' and 'Dementia due to AD'. Controls were not cognitively impaired. The battery of tests included the RAVLT, MMSE, and a combination of additional assessments guided by clinical evaluation of

each participant, alongside CSF biomarkers, FDG and amyloid PET, MR imaging, and clinical follow-up, as was previously described elsewhere [32], Only baseline data was used, and the 9 biomarkers previously mentioned for the ADNI trainings dataset were extracted.

To correct for batch-to-batch variability in absolute CSF values, a global pre-processing step of rescaling the 3 CSF biomarkers using the min-max range in each cohort was applied.

Testing dataset – real-world clinical dataset

The real-world clinical dataset consisted of 686 subjects, a subset of the 'retrospective Belgian multi-center MRI biomarker study in dementia' (REMEMBER, $n = 887$), recruited from 9 different memory clinics that are members of the Belgian Dementia Council (BeDeCo) (mean age in years ± SD; 74.2 ± 9.0, mean MMSE score ± SD; 24 ± 4). The study population consisted of CN ($n = 91$), SCD ($n = 66$), MCI (MCI; $n = 309$, aMCI; $n = 255$ naMCI; $n = 54$) and ADD ($n = 220$) subjects. Patient classification was effectuated as described in Sect. 2.1.2. Controls were not cognitively impaired based on the neuropsychological examination, had no cognitive complaints and were recruited amongst spouses of patients. From this dataset, one clinical score (MMSE score) and the 4 neuroimaging biomarkers (volumes of the hippocampi, temporal, parietal and frontal cortices) used to train the ADNI dataset were available and thus extracted. Specific details regarding recruitment and clinical diagnostic criteria have been previously published [33–35].

Subjects from the research cohort and real-world clinical dataset were staged within the EBM trained on ADNI. The results were used to compare between different diagnostic groups. Additionally, the non-amnestic and amnestic MCI groups were compared in terms of proportions assigned to each EBM stage.

Ethical committee

The REMEMBER study was approved by the ethics committee of University of Antwerp / Universitair Ziekenhuis Antwerpen, Antwerp (N°16/2/18) and by the ethics committees of Algemeen Ziekenhuis Sint-Jan Brugge-Oostende, Brugge (N°1992); Centre Hospitalier Universitaire Brugmann (CHU Brugmann), Brussels (N°2016/84); Centre Hospitalier Universitaire Liège (CHU Liège), Liège (N°2012/274); Cliniques Universitaires de Bruxelles (ULB), Hôpital Erasme, Brussels (N°P2016/187); Cliniques Universitaires Saint-Luc (UCL), Brussels (N°2016/07jui/261); Cliniques St-Pierre Ottignies, Ottignies (N°OM045); Universitair Ziekenhuis Brussel, Brussels (N°2016/183); and Ziekenhuis Netwerk Antwerp (ZNA), Antwerp (N°4730).

Informed written consent was obtained from all participants from the University of Antwerp prospective research cohort, that was approved by the ethics committee of University of Antwerp / Universitair Ziekenhuis Antwerpen, Antwerp (N°15/38/394).

MRI acquisition and processing

Image acquisition

Research cohort MR imaging was performed on each subject on a 3T whole body scanner with a 32-channel head coil (Siemens Trio/PrismaFit, Erlangen, Germany). To obtain 176 axial slices (without slice gap) and 1.0 mm nominal isotropic resolution (FOV = 192 × 256 mm), the 3D MP-RAGE (TR/TE = 2200/2.45 ms) was used. A 3D T1w MR sequence (slice thickness; mean (SD) 1.69 (±) 1.76) was obtained from all participants.

REMEMBER dataset T1-weighted (T1w) MRI sequences from respective radiology departments of the participating memory clinics were available for each subject (MRI systems: GE medical systems (1.5T and 3.0T), Philips (1.5T and 3.0T), and Siemens (1.5T and 3.0T)).

Image analysis

An automated brain imaging morphometry analysis was performed by icobrain dm (v.5.0.0). For this study, volumes of the hippocampi (HIP), "lateral" temporal (TL), parietal (PL), and frontal (FL) cortices were extracted. Please note that for the temporal cortex, cortical gray matter segmentation occurred based on an atlas without inclusion of substructures (e.g. amygdala, hippocampus). Brain volumes scaled for head size were adjusted to account for age and sex by using icobrain's healthy reference population, which are obtained from MR images of 1903 healthy subjects (female ($n = 1069$) and male ($n = 834$) subjects) available from several public collections on which the icobrain software is applied. For each brain structure, the age- and sex-matched median volume computed using icobrain's healthy reference population is subtracted from the patient volume to obtain age- and sex-adjusted volumes. Specific details regarding icobrain dm's cortical lobe and hippocampal segmentation processing steps have been previously published [33–36].

The event-based model

To estimate the maximum likelihood of events, including uncertainty in the sequence, an enhanced version of the EBM model designed by Fonteijn et al., 2012 was applied[18], enabling sequence determination without relying on a-priori clinical knowledge nor biomarker cut-offs. Each EBM stage corresponds to the accumulation of a new biomarker event, therefore, for the selected 9 biomarkers, there are 10 EBM stages (from 0 to 9 respectively). The EBM stage 0 corresponds to no biomarker

event having occurred, while stage 9 corresponds to the occurrence of all events. Each biomarker event indicates the change from a 'normal' (read. non-pathological) to an 'abnormal' (read. pathological) event as seen in AD. In addition, since the ADNI training dataset contains CSF biomarkers, while the real-world clinical testing dataset does not, controls who are amyloid positive were identified and relabeled to nonzero (and also not 1). This indicates the model was trained using amyloid-negative controls, to avoid the controls being contaminated by amyloid-positive individuals that might already have AD pathology but do not show any symptoms yet. In this way, the model can be applied to the real-world clinical testing dataset as well.

Once the maximum likelihood sequence has been determined using the EBM, the stage for a particular subject is assigned using the highest probability by the model, i.e. the stage, that maximizes the probability of the data given the maximum likelihood event sequence. For mathematical details regarding the EBM procedure, we refer to Fonteijn 2012 and Young et al., 2014 [26].

Statistical analysis

R environment (R-Studio, v.1.0.136) for statistical computing and graphics was used for all data processing. Demographic information was reported as mean, standard deviation (SD), median and interquartile range (IQR) where applicable [R package: "arsenal" (tableby and write2word)], with a significance level of 0.05.

Outliers

Clinical diagnosis A subject that was at least two event stages away from the average stage per diagnostic group, was classified as an outlier. The average stage was calculated through taking the median and interquartile range (IQR). To identify the cause of a subject being an outlier, each subject was analyzed as a case study and individual measurements were reported, which were verified by a clinician (S.E.) (Table 3).

CSF biomarkers Preset clinical cut-offs used in the clinic at the time of analysis, but not applied in the EBM, were employed to give an indication of biomarker measure abnormality and will be further referred to as 'abnormal'. The following cut-offs for a healthy individual were used: $A\beta_{1-42}$: >639 pg/mL, T-tau: <297 pg/mL and P-tau_{181}: <57 pg/mL).

Brain volumes Abnormal brain volumes were determined as brain segmentation volumes corresponding to the lowest percentile (< 1%) of healthy volumes of icobrain dm's age-matched healthy reference population.

Additional variables Additional baseline information (if available) consisted of the $A\beta_{1-42/1-40}$ ratio (cut-off: normal value: > 0.120), follow up neuropsychological examination, FDG and amyloid-PET.

Results

Testing datasets

Testing dataset 1 - research cohort

The study population demographics for the research dataset are presented in Table 1. To determine the uncertainty in the maximum likelihood event sequence, bootstrapping resampling of the data (100x) was performed on the ADNI data. Subsequently, an EBM was fitted on the bootstrapped data and a positional variance diagram was plotted for each combination of biomarkers from the testing datasets, showing the maximum likelihood order of events (Fig. 1, x-axis), where each subsequent stage corresponded to the accumulation of an additional biomarker event (Fig. 1, y-axis). In addition, the percentages, and numbers of subjects from each group per EBM stage were reported (Fig. 2; Table 2).

In the research cohort, the positional variance diagram showed a larger uncertainty in biomarker addition/accumulation/sequence associated to EBM stages 2 and 3, indicated by a spread away from the diagonal. Most of the $A\beta_{1-42}$ entries resided in stage 2, whereas the majority of the P-tau_{181} resided in stage 3. However, a proportion of entries for $A\beta_{1-42}$ resided in a later stage than most $A\beta_{1-42}$ entries; stage 3. In contrast to $A\beta_{1-42}$, a proportion of the entries for P-tau_{181} resided in an earlier stage than the majority of P-tau_{181} entries; stage 2.

Maximum likelihood event sequence

The EBM trained on cross-sectional ADNI data confirmed previous findings [25, 26]. The maximum likelihood event sequence for the 9 considered biomarkers was: CSF $A\beta_{1-42}$, CSF P-tau_{181}, CSF T-tau, RAVLT, MMSE, hippocampal volume, volumes of temporal cortex, parietal cortex and frontal cortex. In Young et al., 2014, cognitive scores were preceded by hippocampal *atrophy rates* computed in individuals using longitudinal MRI, but cross-sectional hippocampal and other brain volumes were staged after the cognitive scores, consistent with our results.

Applied to the research cohort, the model provided a plausible distribution of subjects across EBM stages (Fig. 1): 72% of the CN subjects had no abnormal biomarkers (EBM stage 0) and 26% were assigned to stages between 1 and 4 (CSF and MMSE abnormality). When looking at the SCD subjects, all were assigned to stages between 0 and 4. The MCI subjects were scattered across all EBM stages in an increasing fashion, with 6% in EBM stage 0 (no abnormal biomarker), 10% in stages 1–3 (CSF), 18% in stages 4–5 (+ cognition), 19% in stages 6–8

Table 1 Study population – Research cohort clinical and demographic characteristics

	CN (*N* = 46)	SCD (*N* = 10)	aMCI (*N* = 39)	naMCI (*N* = 8)	ADD (*N* = 16)	Total (*N* = 119)	*p* value
Gender (%F)							0.013
F	32 (69.6%)	4 (40.0%)	21 (53.8%)	5 (62.5%)	4 (25,0%)	66 (55.5%)	
Age at BL* (years)							< 0.001
Mean (SD)	60.0 (6.8)	62.6 (10.2)	72.1 (8.4)	74.8 (6.9)	72,7 (7.9)	66.9 (9.8)	
MMSE from 0 to 30							< 0.001
N	46	10	38	8	15	117	
Mean (SD)	29 (1)	28 (2)	25 (3)	27 (2)	21 (4)	26 (4)	
$A\beta_{1-42}$							< 0.001
N	44	10	39	8	16	117	
Mean (SD)	1109.1 (198.7)	1023.1 (366.0)	719.5 (300.8)	868.6 (419.8)	631.1 (145,0)	890.1 (326.8)	
T-tau							< 0.001
N	44	10	39	8	16	117	
Mean (SD)	211.6 (84.8)	236.7 (85.7)	483.1 (229.7)	338.1 (194.0)	614.0 (305.5)	367.9 (241.8)	
$P\text{-}tau_{181}$							< 0.001
N	44	10	39	8	16	117	
Mean (SD)	45.7 (15.0)	47.8 (15.9)	77.1 (30.4)	58.8 (24.5)	88.7 (35.6)	63.1 (30.0)	
RAVLT							0.129
N	1	4	17	4	4	30	
Mean (SD)	30.0 (0.0)	37.8 (11.4)	24.8 (9.2)	24.3 (8.4)	23.0 (3.6)	26.4 (9.5)	
Hippocampus (mL)							< 0.001
Mean (SD)	9.3 (0.8)	8.5 (0.7)	8.1 (1.2)	8.3 (0.8)	7.4 (0.8)	8.5 (1.2)	
Temporal cortex (mL)							< 0.001
Mean (SD)	153.3 (9.8)	146.2 (9.4)	136.0 (13.2)	139.8 (13.1)	128.7 (14.9)	142.8 (15.0)	
Frontal cortex (mL)							< 0.001
Mean (SD)	220.0 (17.0)	215.9 (22.1)	187.7 (20.2)	182.4 (24.5)	189.4 (17.8)	142.8 (15.0)	
Parietal cortex (mL)							< 0.001
Mean (SD)	143.8 (11.1)	131.4 (12.3)	123.9 (12.6)	118.0 (20.0)	111.2 (13.2)	1407.2 (79.4)	

Description of the data as mean and standard deviation (SD). Analysis: Chi-square test (categorical variables; gender), ANOVA analysis (continuous variables; **Age at BL** (in years), **MMSE** (from 0 to 30), Aβ values (**$A\beta_{1-42}$**), total tau (**T-tau**), phosphorylated Tau (**$P\text{-}tau_{181}$**), Rey's auditory verbal learning test (**RAVLT**) and segmented brain volumes normalized for head size (**hippocampus**, **temporal**, **frontal** and **parietal** cortices). *BL = baseline

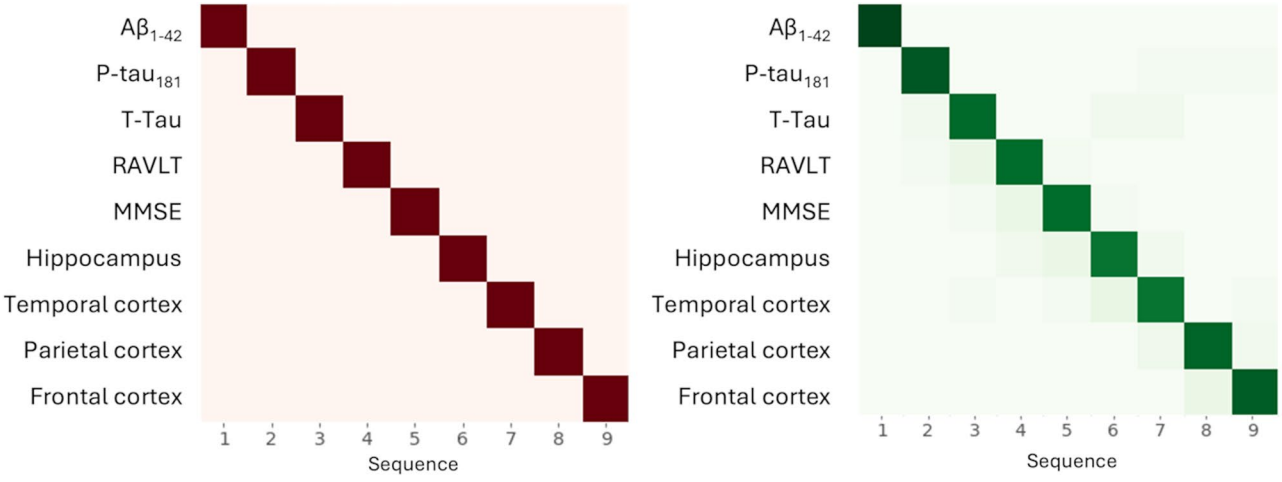

Fig. 1 Positional variance diagram - Research cohort. Positional variance plots of the research dataset were visualized, showing the distribution of the maximum likelihood event sequence. Each entry represents the proportion of bootstrapped biomarker samples, color-coded through cell-shading. White indicates 0, which becomes increasingly more shaded with increased proportion of biomarker samples, until, if 1, becoming red (left) or green (right). **MMSE** = Mini-Mental State Examination. **RAVLT** = Rey Auditory Verbal Learning Test

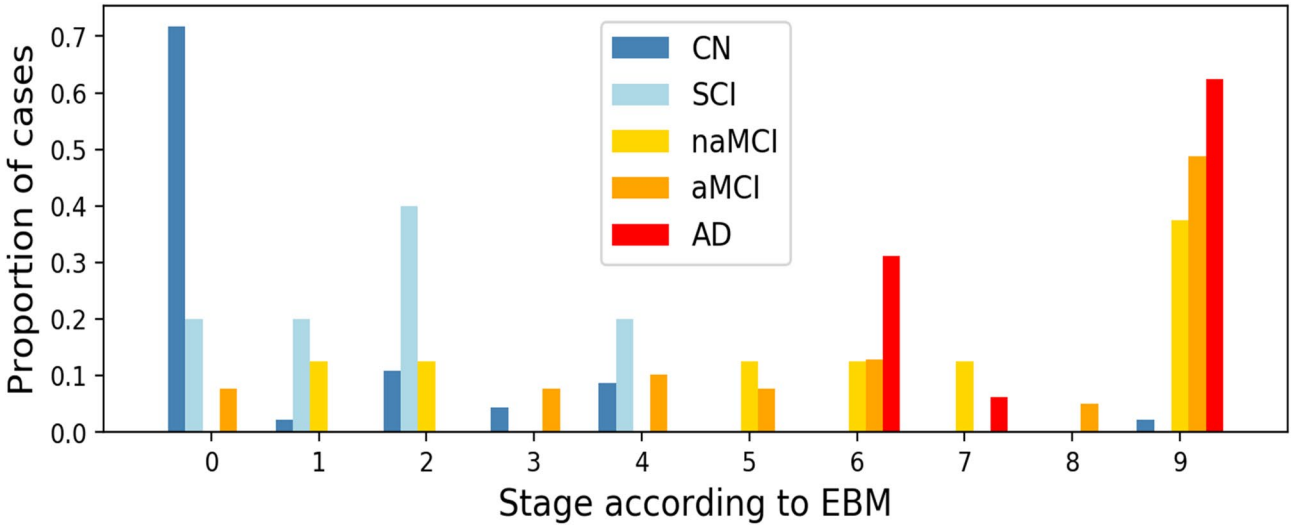

Fig. 2 Distribution for each group per EBM stage – Research cohort. Proportion of subjects from each group per EBM stage, with MCI split as non-amnestic (naMCI) and amnestic (aMCI). Each EBM stage on the x-axis corresponds to the occurrence of a new biomarker transition event. Stage 0 being no events having occurred and stage 9 is when all events have occurred. Events are ordered by the maximum likelihood event sequence for the ADNI population. **CN** = controls, **SCD** = subjective cognitive decline, **MCI** = mild cognitive impairment, **aMCI** = amnestic mild cognitive impairment, **naMCI** = non-amnestic mild cognitive impairment, **ADD** = Alzheimer's disease dementia

(+ hippocampi/temporal/parietal cortex volumes), and 47% in stage 9 (+ frontal). Comparing the staging results of non-amnestic and amnestic MCI subjects, a clear trend towards higher staging was observed in amnestic subjects, with 54% being assigned stages 8–9 as opposed to 38% of the non-amnestic subjects. All ADD subjects had at least stage 6 (CSF + cognition + hippocampi), with 62% being assigned to stage 9.

Real-world clinical dataset

Testing dataset 2 – real-world clinical dataset

The study population demographics for the real-world clinical dataset are presented in Table 3. The positional variance plots of the independent real-world clinical dataset (Fig. 3) were visualized as described in Sect. 3.1.1. The independent clinical dataset showed an uncertainty for the biomarkers MMSE (stage 2) and the hippocampus (stage 3), where a small proportion of the entries were situated in a later stage (stage 3) for MMSE, and an earlier stage (stage 2) for hippocampal volume, with respect to the location in the maximum likelihood sequence of most of their entries. The percentages and numbers of subjects from each group per EBM stage were also reported (Fig. 4; Table 4).

Maximum likelihood event sequence

An EBM was fitted to data from ADNI with the following 5 measurements: MMSE, hippocampus, frontal, parietal, temporal cortex. The most likely sequence of events was found to be: MMSE, hippocampus, temporal cortex, parietal cortex, and frontal cortex volumes. When applied to the independent REMEMBER cohort, staging results showed an association of EBM stage with clinical diagnosis, with CN and SCD predominantly in stage 0 (no abnormal event), ADD largely in stages 2–5, and MCI quite evenly distributed over the EBM stages. Comparing the staging results of non-amnestic and amnestic MCI subjects, again a trend towards higher staging was observed in amnestic subjects, with 24% being assigned stage 5 as opposed to 8% of the non-amnestic subjects. In addition, a higher proportion of cases was observed for non-amnestic subjects in stages 0–1 (22%), as opposed to amnestic subjects (17%).

Event based modeling analysis at patient level - research cohort

A total of 10 outliers were identified from the research cohort, according to the procedure mentioned in Sect. 2.6. To identify the cause of a subject being an outlier, each subject was analyzed as a case study, and individual measurements were reported and verified by a clinician (S.E.). Outliers consisted of four cognitively healthy controls (CN1, CN2, CN3 and CN4) and two SCD (SCD1 and SCD2) subjects assigned to stage 4, one cognitively healthy control (CN0) assigned to stage 9, and three aMCI (aMCI1, aMCI2 and aMCI3) patients that were assigned to stage 0 (Table 2, highlighted in red). The probability of the outliers being put in a certain stage can be found in Supp 1.

Cognitively healthy controls

As shown in Table 5, CN1, assigned to stage 4, showed a clinically abnormal T-tau value (341 pg/mL). Follow-up neuropsychological examination reported normal

Table 2 Percentage and numbers of subjects from each group per EBM stage–Research cohort

EBM STAGE	STAGE 0		STAGE 1		STAGE 2		STAGE 3		STAGE 4		STAGE 5		STAGE 6		STAGE 7		STAGE 8		STAGE 9	
GROUP	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#
CN	72	33	2	1	11	5	4	2	9	**4**	0	0	0	0	0	0	0	0	2	**1**
SCD	20	2	20	2	40	4	0	0	20	**2**	0	0	0	0	0	0	0	0	0	0
MCI	6	3	2	1	2	1	6	3	9	4	9	4	13	6	2	1	4	2	47	22
naMCI	0	0	12	1	12	1	0	0	0	0	12	1	12	1	12	1	0	0	38	3
aMCI	8	**3**	0	0	0	0	8	3	10	4	8	3	13	5	0	0	5	2	49	19
ADD	0	0	0	0	0	0	0	0	0	0	0	0	31	5	6	1	0	0	62	10

CN = controls, **SCD** = subjective cognitive decline, **MCI** = mild cognitive impairment, **aMCI** = amnestic mild cognitive impairment, **naMCI** = non-amnestic mild cognitive impairment, **ADD** = Alzheimer's disease dementia. Each stage (0–9) corresponds to the occurrence of a new biomarker transition event. Stage 0 corresponds to no events having occurred and stage 9 is when all events have occurred. Pink cells are considered outliers

cognitive and neurological functioning. CN2, assigned to stage 4, showed clinically abnormal P-tau$_{181}$ (65.4 pg/mL) and T-tau (390 pg/mL) values and reported normal cognitive and neurological functioning during follow-up. CN3, assigned to stage 4, showed a clinically abnormal P-tau$_{181}$ value (60.1 pg/mL). CN4, assigned to stage 4, without clinically abnormal biomarker changes. CN0, assigned to stage 9, showed hippocampal (6.561 mL), frontal cortex (162.713 mL) and parietal cortex volumes (103.930 mL) abnormal for age and head size when compared to icobrain dm's healthy reference population. There were no CSF biomarkers for CN0, but there was a positive amyloid-PET available. Follow up neuropsychological examination was not available for subjects CN3, CN4 and CN0.

Subjective cognitive decline

SCD1, assigned to stage 4, showed clinically abnormal CSF P-tau$_{181}$ value (58.3 pg/mL) and parietal cortex volume for age and head size (120.284 mL) when compared to icobrain dm's healthy reference population. SCD2, assigned to stage 4 displayed a pathological CSF biomarker profile compatible with AD according to the preset clinical cut-offs used in this study.

Mild cognitive impairment

aMCI1, assigned to stage 0, displayed an absence of clinically abnormal biomarkers, except for a hippocampal volume (7.747 mL) abnormal for the subject's age and head size when compared to icobrain dm's healthy reference population. aMCI2 and aMCI3, both assigned to stage 0, displayed an absence of clinically abnormal biomarkers. Neuropsychological testing at baseline for aMCI3 stated the presence of a verbal amnestic imprinting syndrome e causa ignota (unknown cause), resulting in the indication of a single domain amnestic MCI, however noting that according to scientific criteria the patient was too young for this diagnosis.

Discussion

This study examined the utility of an EBM staging model for AD, trained on a cross-sectional ADNI dataset, and tested/deployed on two independent datasets, while simultaneously investigating potential differences between EBM profiles of amnestic and non-amnestic MCI patients.

Research cohort

The maximum likelihood of the sequence of events generated by the EBM model was in line with the well-known hypothetical model suggested by Jack et al. 2010[9]. In addition, in Young et al. 2014 [26], cognitive scores were preceded by hippocampal *atrophy rates* computed in individuals using longitudinal MRI, but cross-sectional

Table 3 Study population – real-world clinical dataset clinical and demographic characteristics

	CN (*N* = 91)	SCD (*N* = 66)	naMCI (*N* = 54)	aMCI (*N* = 255)	ADD (*N* = 220)	Total (*N* = 685)	*p* value
Gender (%F)							
F	53.8	48.5	63.0	46.7	63.2	54.4	0.004
Age at BL* (years)							
Mean (SD)	67.3 (8.6)	68.1 (10.2)	74.1 (8.9)	74.9 (7.5)	78.2 (7.9)	74.2 (9.0)	<0.001
MMSE from 0 to 30							
N	71	59	49	243	210	632	
Mean (SD)	21 (4)	28 (2)	27 (3)	25 (3)	21 (4)	24 (4)	<0.001
Stage							
mean (SD)	0 (1)	1 (1)	2 (2)	3 (2)	4 (1)	3 (2)	<0.001
Hippocampus (mL)							
Mean (SD)	9.177 (0.753)	8.746 (0.901)	8.953 (3.246)	7.956 (1.122)	7.506 (2.074)	8.128 (1.783)	<0.001
Temporal cortex (mL)							
Mean (SD)	137.899 (12.554)	131.806 (15.850)	128.733 (20.608)	122.404 (18.440)	131.037 (19.892)	135.836 (19.782)	<0.001
Frontal cortex (mL)							
Mean (SD)	214.761 (16.569)	213.333 (22.839)	199.700 (25.914)	197.250 (23.014)	196.670 (32.781)	201.127 (26.994)	<0.001
Parietal cortex (mL)							
Mean (SD)	137.899 (12.554)	131.8096 (15.850)	128.733 (20.608)	122.404 (18.440)	120.849 (25.675)	125.363 (21.209)	<0.001

Description of the data as mean and standard deviation (SD). Analysis: Chi-square test (categorical variables; gender), ANOVA analysis (continuous variables; **Age at BL** (in years), **MMSE (**from 0 to 30), and segmented brain volumes normalized for head size (**hippocampus**, **temporal**, **frontal** and **parietal** cortices). *BL = baseline

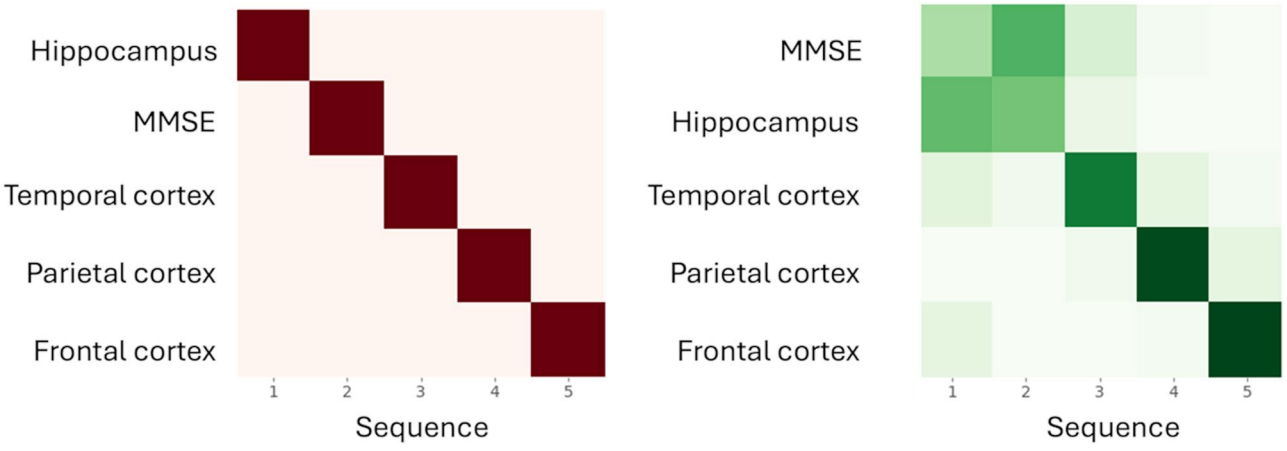

Fig. 3 Positional variance diagram - Real-world clinical dataset. Positional variance plots of the independent clinical dataset were visualized, showing the distribution of the maximum likelihood event sequence. Each entry represents the proportion of bootstrapped biomarker samples, color-coded through cell-shading. White indicates 0, which becomes increasingly more shaded with increased proportion of biomarker samples, until, if 1, becoming red (left) or green (right). **MMSE** = Mini-Mental State Examination. **RAVLT** = Rey Auditory Verbal Learning Test

hippocampal and other brain volumes were staged after the cognitive scores, consistent with our results. In the hypothetical model, Aβ accumulation was introduced first, followed by synaptic dysfunction, tau mediated neuronal injury, brain structure abnormalities, changes in cognition and finally changes in clinical function. The EBM model showed T-tau as one of the earliest abnormalities, following Aβ accumulation and P-tau changes. It is recognized that amyloid beta and tau changes have already changed before onset of clinical symptoms[22,23]. The cognition test scores, however, were staged prior to brain structure changes. MMSE, even though being a proprietary instrument for dementia screening, only shows an intermediate predictive accuracy in risk models. In addition, MMSE cut-offs are often debated and dependent on age and education [37–39]. Episodic memory, assessed by RAVLT, has however been suggested to be more sensitive than MMSE in predicting disease progression in case of AD[24]. The results observed for CN and AD were as expected. Many subjects of the CN group (72%) were categorized in the stage corresponding to no abnormal biomarkers (stage 0), while for the AD subjects, 93% were categorized in stage 6 to 9, displaying at least one abnormal MRI biomarker (read. hippocampal atrophy) together with abnormal CSF biomarkers and cognitive scores. For the MCI patients, it was observed

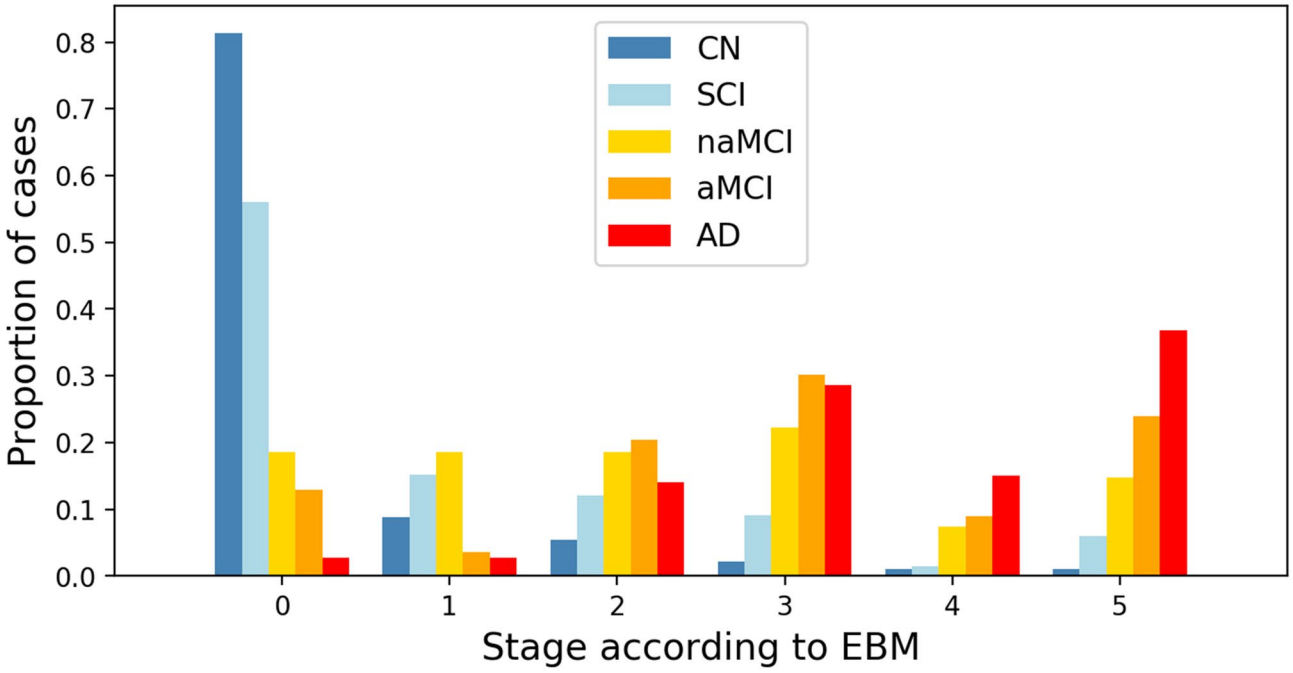

Fig. 4 Distribution for each group per EBM stage – Real world clinical dataset. Proportion of subjects from each group per EBM stage, with MCI split as non-amnestic (naMCI) and amnestic (aMCI). Each EBM stage on the x-axis corresponds to the occurrence of a new biomarker transition event. Stage 0 corresponds to no events having occurred and stage 5 is when all events have occurred. Events are ordered by the maximum likelihood event sequence for the ADNI population. **CN** = controls, **SCD** = subjective cognitive decline, **MCI** = mild cognitive impairment, **aMCI** = amnestic mild cognitive impairment, **naMCI** = non-amnestic mild cognitive impairment, **ADD** = Alzheimer's disease dementia

that most of the MCI patients that were diagnosed with amnestic MCI are in the most advanced stages, while a big proportion of the non-amnestic MCI patients were categorized in earlier stages. However, it needs to be taken into consideration that the biomarker RAVLT is essentially a memory related biomarker, which is more impaired in aMCI.

Real-world clinical dataset

The study demonstrated that the EBM, initially trained on a large cross-sectional dataset could effectively stage disease progression in a real-world clinical setting encompassing patients from multiple memory clinics. This suggests that the underlying principles and biomarker associations learned from the training dataset generalize well to new and diverse patient populations, indicating robustness and reliability in different clinical contexts. However, the independent clinical dataset revealed uncertainties in the staging of certain biomarkers, notably the MMSE and hippocampus. This uncertainty was reflected in the positional variance plots, where a small proportion of entries deviated from the expected staging sequence. For instance, some MMSE entries were situated in a later stage (stage 3) compared to the majority, while hippocampus entries were positioned in an earlier stage (stage 2). This discrepancy suggests variability in the progression patterns of these biomarkers within this cohort, highlighting the complexity of disease dynamics in real-world clinical settings. Despite the uncertainties in biomarker staging, the EBM demonstrated an association between stage progression and clinical diagnosis within the REMEMBER cohort. Specifically, individuals diagnosed with CN and SCD were predominantly categorized in stage 0, indicating no abnormal biomarker events. In contrast, ADD patients were largely distributed across stages 2–5, reflecting the presence of abnormal biomarker changes consistent with disease pathology. MCI subjects showed a more even distribution across EBM stages, with a trend towards higher staging in amnestic MCI cases compared to non-amnestic MCI cases. This association underscores the utility of the EBM in stratifying patients based on disease severity and clinical phenotype, facilitating more targeted interventions and management strategies.

Research cohort case studies

The individual cases showed how the EBM model puts attention to subjects that were not classified in accordance with their clinical diagnosis. Several reasons might exist for this observation, including the possibility of subjects being preclinical AD patients, the phenomenon of MCI normalization, the presence of depression or another major psychiatric disorder as a possible cause of

Table 4 Percentage and numbers of subjects from each group per EBM stage – Real-world clinical dataset

EBM STAGE	STAGE 0		STAGE 1		STAGE 2		STAGE 3		STAGE 4		STAGE 5	
GROUP	%	#	%	#	%	#	%	#	%	#	%	#
CN	81	74	9	8	6	5	2	2	1	**1**	1	1
SCD	56	37	15	10	12	8	9	6	2	**1**	6	4
MCI	14	43	6	19	20	62	29	89	9	26	22	69
naMCI	11	10	11	10	11	10	13	12	46	43	8	8
aMCI	13	33	4	9	20	52	30	77	9	23	24	61
ADD	3	6	3	6	14	31	29	63	15	33	37	81

CN = controls, **SCD** = subjective cognitive decline, **MCI** = mild cognitive impairment, **aMCI** = amnestic mild cognitive impairment, **naMCI** = non-amnestic mild cognitive impairment, **ADD** = Alzheimer's disease dementia

cognitive dysfunction and possible progression into other neurodegenerative diseases.

Subject CN0 was assigned to stage 9 due to abnormal hippocampal, frontal cortex, and parietal cortex volumes. CSF biomarkers were not available for this subject. However, a positive amyloid-PET at baseline, which was not part of the selected biomarkers that were used to create the EBM model, strengthened the hypothesis that the model, even in the case of missing data, classifies accordingly. According to the NIAAA-criteria, in case of absence of CSF biomarkers, the combination of a positive amyloid-PET accompanied by an MRI showing hippocampal atrophy or FDG-PET corresponding with a pathological AD biomarker profile one could suggest that subject CN0 might be a preclinical AD.

SCD1, who was also classified in stage 4 due to an abnormal P-tau$_{181}$ value and parietal cortex volume, showed a conversion to non-amnestic multi domain MCI during a follow up neuropsychological examination after one year. Furthermore, even though SCD1 did not reveal any abnormal CSF biomarker values, additional calculation of the A$\beta_{1-42/1-40}$ ratio (0.08 pg/mL) was below the 0.120 pg/mL cut-off, indicating the possibility of a preclinical AD stage. SCD2, who was classified in stage 4 as well due to a pathological CSF biomarker profile, also progressed to a single domain amnestic MCI during follow-up. These results were in line with the reported biomarker measurements used in the EBM model.

Considering that the biomarker model was built on AD biomarkers, there might be non-AD subjects present amongst the non-amnestic MCI subjects that can progress into other neurodegenerative diseases. In addition, even with the presence of AD-characteristic memory complaints, the possibility to develop into other neurodegenerative diseases is also very much present for the amnestic MCI subjects. For example, aMCI1 revealed an abnormal hippocampal volume, however, was assigned to stage 0. Since follow-up neuropsychological testing indicated a normal cognitive and psychological functioning, it is possible that the underlying pathology might not be AD-related, but due to another disease displaying common pathology (such as early/preclinical hippocampal sclerosis that could not be detected at a neuropsychological evaluation yet). In addition, even though subject aMCI1 had a MMSE score of 30, the neuropsychological examination showed deficits in one cognitive domain: logical memory (> 1.5 SD), which could be an indication that the MMSE score might not be sensitive enough.

The presence of anxiety, apathy and depression is not uncommon in MCI patients but could also be the cause of their cognitive deficit. For aMCI2, categorized as stage 0, additional FDG-PET testing showed a compatibility with atrophy supported by enlargement of the lateral ventricles, but no indication of an association to AD. In

Table 5 Individual measurements of EBM model outliers

OUTLIERS	CN1 STAGE 4	CN2 STAGE 4	CN3 STAGE 4	CN4 STAGE 4	CN0 STAGE 9	SCD1 STAGE 4	SCD2 STAGE 4	aMCI1 STAGE 0	aMCI2 STAGE 0	aMCI3 STAGE 0
Gender (M/F)	F	F	F	F	F	F	M	F	M	F
Age at BL* (years)	60	69	47	52	83	63	68	77	73	51
MMSE from 0 to 30	27	27	25	26	27	25	27	30	30	30
Education (years)	18	21	21	24	19	19	18	21	30	18
$A\beta_{1\text{-}42}$	904	880	1285	931	NA	1452	443	1403	1230	1117
T-tau	341	390	267	163	NA	279	340	257	159	155
P-tau_{181}	49.1	65.4	60.1	32.6	NA	58.3	62.1	45.5	30.5	36.5
RAVLT	NA	NA	NA	NA	30	38	42	43	28	NA
Hippocampus (mL)	11.604	10.177	9.982	10.416	6.561	8.331	8.001	7.747	7.936	8.692
Temporal cortex (mL)	161.088	151.692	149.880	161.329	134.071	143.580	134.248	130.529	124.676	159.881
Frontal cortex (mL)	216.920	218.029	234.974	235.613	162.713	201.641	227.367	196.090	167.873	232.269
Parietal cortex (mL)	144.466	145.919	147.575	146.247	103.930	120.284	131.248	123.433	116.014	139.365
FU NPO diagnosis	CN	CN	NA	NA	NA	md naMCI	sd aMCI	CN	sd aMCI	CN
$A\beta_{1-42/1-40}$ ratio	NA	NA	NA	NA	NA	0.086	0.041	NA	0.119	0.110
Amyloid-PET	NA	NA	NA	NA	AD-positive*	NA	NA	NA	NA	NA

*According to the NIAAA-criteria in case of absence of CSF biomarkers (positive amyloid-PET accompanied by an MRI showing hippocampal atrophy or FDG-PET corresponding with a pathological AD biomarker profile). **CN** = controls, **SCD** = subjective cognitive decline, **MCI** = mild cognitive impairment, **aMCI** = amnestic mild cognitive impairment, **naMCI** = non-amnestic mild cognitive impairment

addition, neuropsychological testing at baseline indicated the presence of a severe depression (GDS 30/30), which could explain the subject's symptoms.

Even though MCI is often seen as a station on a one-way journey towards dementia, represented as cognitive dysfunction due to mental illness, reversion to normal after an MCI diagnosis occurs regularly[25]. In fact, several longitudinal studies revealed that up to half of the patients diagnosed with MCI during an initial visit to the clinic, returned to normal at follow-up examinations [26]. This might be due to a lack of employing biomarkers at time of the diagnosis, combined with the biological heterogeneity of the MCI syndrome, as might have been the case with aMCI3, assigned to stage 0 due to an absence of biomarker changes. The follow up neuropsychological examination concluded a borderline normal functioning according to age and education, whereas the neuropsychological evaluation the year after concluded a normal cognitive and psychological functioning.

Another important consideration is the presence of alternative or co-occurring conditions that may transiently affect cognitive performance. For instance, depressive symptoms can lead to cognitive impairments that closely resemble MCI. When effectively treated, cognitive function may improve, leading to reversion. Furthermore, in some cases, fluctuating cognition may be an early symptom of Dementia with Lewy Bodies (DLB), which is characterized by attentional and executive variability [40–42]. In these individuals, cognitive performance may differ significantly across visits, giving the appearance of reversion when, in fact, the underlying disease course is progressive but inconsistent in early stages.

Strengths and limitations

One of the strengths of this study was that an EBM staging model for AD was tested on both a research cohort and a real-world clinical dataset containing specific subclasses, enabling the evaluation of EBM profiles of amnestic and non-amnestic MCI patients. However, in the research cohort, the non-amnestic MCI group only contained 9 subjects. In addition, one of the limitations of this EBM staging is that it applies a 'one size fits all' approach, which complicates the interpretation of borderline cases. Nevertheless, EBM-staging models can aid in including high-risk patients in clinical trials, and increase our understanding of the several factors that drive the different trajectories of AD.

Another important consideration is the increasingly recognized heterogeneity of the ADNI cohort, partly due to (undetected) limbic-predominant age-related TDP-43 (LATE) or concomitant α-synuclein pathology [43, 44]. This variability can pose significant challenges for disease progression modeling. EBM models trained on such heterogeneous samples may inadvertently capture patterns specific to certain subtypes, which can lead to overfitting, where the model performs well within the training cohort but lacks robustness and fails to generalize to other populations or patients with different disease trajectories.

Furthermore, the presence of multiple underlying pathologies can introduce bias in how the model associates biomarkers, clinical features, or neuroimaging data with disease stages. For instance, LATE can clinically mimic AD but is driven by distinct neuropathology (TDP-43 vs. Aβ/tau) and tends to follow a slower, predominantly amnestic progression [43, 45]. Without accounting for such variation, EBM models risk

misclassification or misinterpretation of disease dynamics, limiting their clinical utility.

In addition, the EBM faces challenges in accurately staging disease progression in all cases, as evidenced by the uncertainties observed in the REMEMBER cohort.

Here, the EBM models were used as proof-of-concept applications for the validation of the EBM technique. However, it needs to be noted that the EBM's reliance on specific biomarkers may limit its applicability in settings where certain biomarkers are not routinely assessed. Thus, the inclusion of specific biomarkers in EBM models also depends on their availability. CSF biomarkers can provide valuable diagnostic information, and the RAVLT may offer greater sensitivity than the MMSE for detecting cognitive decline (e.g., verbal memory impairment) [46]. Nevertheless, CSF biomarkers are not always accessible due to practical constraints associated with lumbar puncture, such as medical contraindications or patient refusal. In this context, it is also important to note that recent revisions to the ATN framework challenge the assumption of equivalence between biomarkers within the same category. Whereas the original 2018 NIA-AA framework grouped CSF and imaging biomarkers as interchangeable within each AT(N) category, newer evidence indicates that this equivalence does not always hold. Therefore, the revised criteria acknowledge potential discrepancies between imaging and biofluid biomarkers, further underlining the need to tailor model choice to biomarker availability and clinical applicability [11]. It also needs to be noted that it does not inherently make use of longitudinal data, especially when, even though this was not the case in this study, individuals have varying numbers of timepoints, therefore not considering within-individual correlations which can lead to a considerable bias.

The EBM's intuitive design and simplicity lie in its ability to present a straightforward sequence of biomarker events. However, this simplicity comes at the cost of losing the temporal dimension, a critical drawback when it comes to identifying patients likely to progress within the timeframe of a clinical trial.

Finally, variability in biomarker trajectories and patient heterogeneity necessitates careful interpretation and validation in different clinical contexts. However, well-managed heterogeneity—such as through stratification, subtype modeling, and multimodal inputs—can actually enhance generalizability by teaching the model about real-world variation. These approaches can refine the identification of distinct pathological subgroups, leading to more personalized predictions. When combined with methods to correct sample imbalance (e.g., stratified sampling, oversampling, undersampling, and synthetic data generation), EBM can provide a promising framework for building models that are both technically robust and sensitive to the biological complexity of real-world clinical populations.

Conclusion

This study showed that the event-based model for AD staging is generalizable, meaning that it can be trained on large cross-sectional historical datasets such as ADNI, and still have reliable staging results in new independent data, provided that the same biomarkers are used. This provides confidence towards using these models in multi-center trials, for instance, as a screening tool. Furthermore, we show that amnestic MCI subjects score in general higher than non-amnestic subjects, demonstrating utility for precision recruitment/screening.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13195-025-01788-6.

Supplementary Material 1

Acknowledgements

The authors would like to acknowledge Dr. J.C. Lemper, Dr. E. Mormont, and Dr. E. Triau, and extend their appreciation to the patients and caregivers whose essential contributions have been invaluable to this manuscript.

Author contributions

MMJW: conceptualization, investigation, resources, data curation, formal analysis, validation, visualization, and writing – original draft. DMS: methodology, software, validation, formal analysis, and writing – review and editing. CB, FB, BB, JB, PPdD, OD, AI, GP, ES, KS, AS, ET, JT, AvB, JV: data curation, and writing – review and editing. DS: methodology, resources, software, review and editing. MBj: formal analysis, and writing, review and editing. MBe, NPO, DCA: methodology, software, supervision, validation, formal analysis SE: conceptualization, resources, supervision, project administration, funding acquisition, and writing – review and editing. All authors critically revised and approved the content of the final manuscript before submission.

Funding

This research was in part supported by the agency of Flanders Innovation & Intrepreneurship (VLAIO), the Flemish Agency for Innovation by Science and Technology (IWT 140262), the Interreg V programme Flanders-The Netherlands of the European Regional Development Fund (ERDF) (Herinneringen/Memories project), the European Union's Horizon 2020 research and innovation program under grant agreement numbers 666992 (EuroPOND) and 765148 (TRABIT). For the University of Liège center, this work was supported by a French Speaking Community Concerted Research Action (ARC-06/11–340) and a Belgian InterUniversity Attraction Pole (P6/29). SD is funded by an industrial grant (Baekeland, HBC.2019.2579) from VLAIO.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The REMEMBER study was approved by the ethics committee of University of Antwerp / Universitair Ziekenhuis Antwerpen, Antwerp (N°16/2/18) and by the ethics committees of Algemeen Ziekenhuis Sint-Jan Brugge-Oostende, Brugge (N°1992); Centre Hospitalier Universitaire Brugmann (CHU Brugmann), Brussels (N°2016/84); Centre Hospitalier Universitaire Liège (CHU Liège), Liège (N°2012/274); Cliniques Universitaires de Bruxelles (ULB), Hôpital Erasme, Brussels (N°P2016/187); Cliniques Universitaires Saint-Luc (UCL), Brussels

(N°2016/07jui/261); Cliniques St-Pierre Ottignies, Ottignies (N°OM045); Universitair Ziekenhuis Brussel, Brussels (N°2016/183); and Ziekenhuis Netwerk Antwerp (ZNA), Antwerp (N°4730). The research was conducted in accordance with the Declaration of Helsinki and informed written consent was obtained from all participants from the University of Antwerp prospective research cohort, that was approved by the ethics committee of University of Antwerp / Universitair Ziekenhuis Antwerpen, Antwerp (N°15/38/394). icobrain dm is a proprietary software, developed by icometrix for the automated quantification of brain volumes and white matter hyperintensities.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Dep. of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[2]Dep. of Neurology, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium
[3]Neuroprotection & Neuromodulation (NEUR) Research Group, Center for Neurosciences (C4N), Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, Brussel 1090, Belgium
[4]icometrix, Leuven, Belgium
[5]Department of Neurology and Memory Clinic, ZAS-Hoge Beuken, Antwerp, Belgium
[6]CRC Human Imaging , GIGA Research, University of Liège, Liège, Belgium
[7]Fonds de la Recherche Scientifique-FNRS, Liège, Belgium
[8]Geriatrics Department, Brugmann University Hospital, Université Libre de Bruxelles, Brussels, Belgium
[9]Neurology Department, AZ St-Jan Brugge, Ghent University and Ghent University Hospital, Bruges, Gent, Belgium
[10]Neurology Department, H. U. B. - Erasme Hospital, Université Libre de Bruxelles (ULB), Brussels, Belgium
[11]Laboratory of Neurochemistry and Behaviour, Experimental Neurobiology unit, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[12]Department of Neurology and Alzheimer Research Center, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands
[13]WELBIO department, WEL Research Institute, Wavre 1300, Belgium
[14]Institute of Neuroscience, Université Catholique de Louvain, Brussels 1200, Belgium
[15]Department of Neurology, Clinique Universitaires Saint-Luc, Brussels 1200, Belgium
[16]Department of Neurology, Clinique Saint-Pierre, Ottignies, Belgium
[17]GIGA Cyclotron Research Centre, University of Liege, Liège, Belgium
[18]Department of Neurology, University Hospital Ghent, Ghent University, Ghent, Belgium
[19]Neurology & Geriatrics Dpt, Brugmann University Hospital, Van Gehuchtenplein 4, Brussels 1020, Belgium
[20]Neuropathology lab, IBB-NeuroBiobank BB190113, Born Bunge Institute, Antwerp, Belgium
[21]Department of Pathology, Antwerp University Hospital - UZA, Antwerp, Belgium
[22]Laboratory of Neurology, Translational Neurosciences, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium
[23]Gerontology & Geriatrics, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium
[24]Department of Geriatric Medicine, UZ Leuven, Leuven, Belgium
[25]Radiology department, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium
[26]Laboratory of Neurochemistry, Dept of Clinical Chemistry, Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium
[27]Department of Computer Science, UCL Hawkes Institute (Centre for Medical Image Computing), University College London, London, UK
[28]Department of Neurology, AZ Sint-Lucas, Brugge, Belgium

Received: 18 November 2024 / Accepted: 8 June 2025
Published online: 16 June 2025

References

1. Collaborators GBDD. Global, regional, and National burden of alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol. 2019;18(1):88–106.
2. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimers Dement. 2011;7(3):280–92.
3. Parnetti L, Chipi E, Salvadori N, D'Andrea K, Eusebi P. Prevalence and risk of progression of preclinical alzheimer's disease stages: a systematic review and meta-analysis. Alzheimers Res Ther. 2019;11(1):7.
4. Bertens D, Vos S, Kehoe P, Wolf H, Nobili F, Mendonça A, et al. Use of mild cognitive impairment and prodromal AD/MCI due to AD in clinical care: a European survey. Alzheimers Res Ther. 2019;11(1):74.
5. Blennow K, Zetterberg H. Biomarkers for alzheimer's disease: current status and prospects for the future. J Intern Med. 2018;284(6):643–63.
6. Huynh RA, Mohan C. Alzheimer's disease: biomarkers in the genome, blood, and cerebrospinal fluid. Front Neurol. 2017;8:102.
7. Chetelat G, Arbizu J, Barthel H, Garibotto V, Law I, Morbelli S, et al. Amyloid-PET and (18)F-FDG-PET in the diagnostic investigation of alzheimer's disease and other dementias. Lancet Neurol. 2020;19(11):951–62.
8. Sheikh-Bahaei N, Sajjadi SA, Manavaki R, Gillard JH. Imaging biomarkers in alzheimer's disease: A practical guide for clinicians. J Alzheimers Dis Rep. 2017;1(1):71–88.
9. Jack CR Jr., Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. Lancet Neurol. 2010;9(1):119–28.
10. Jack CR Jr., Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013;12(2):207–16.
11. Jack CR Jr., Andrews JS, Beach TG, Buracchio T, Dunn B, Graf A, et al. Revised criteria for diagnosis and staging of alzheimer's disease: alzheimer's association workgroup. Alzheimers Dement. 2024;20(8):5143–69.
12. Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: an unbiased descriptive classification scheme for alzheimer disease biomarkers. Neurology. 2016;87(5):539–47.
13. Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of alzheimer's disease. Alzheimers Dement. 2018;14(4):535–62.
14. Aguillon D, Langella S, Chen Y, Sanchez JS, Su Y, Vila-Castelar C, et al. Plasma p-tau217 predicts in vivo brain pathology and cognition in autosomal dominant alzheimer's disease. Alzheimers Dement. 2023;19(6):2585–94.
15. Della Monica C, Revell V, Atzori G, Laban R, Skene SS, Heslegrave A, et al. P-tau217 and other blood biomarkers of dementia: variation with time of day. Transl Psychiatry. 2024;14(1):373.
16. Ashton NJ, Brum WS, Di Molfetta G, Benedet AL, Arslan B, Jonaitis E, et al. Diagnostic accuracy of a plasma phosphorylated Tau 217 immunoassay for alzheimer disease pathology. JAMA Neurol. 2024;81(3):255–63.
17. Karantzoulis S, Galvin JE. Distinguishing alzheimer's disease from other major forms of dementia. Expert Rev Neurother. 2011;11(11):1579–91.
18. Wei G, Irish M, Hodges JR, Piguet O, Kumfor F. Disease-specific profiles of apathy in alzheimer's disease and behavioural-variant frontotemporal dementia differ across the disease course. J Neurol. 2020;267(4):1086–96.
19. Landin-Romero R, Kumfor F, Leyton CE, Irish M, Hodges JR, Piguet O. Disease-specific patterns of cortical and subcortical degeneration in a longitudinal study of alzheimer's disease and behavioural-variant frontotemporal dementia. NeuroImage. 2017;151:72–80.
20. Ye BS, Lee S, Yoo H, Chung SJ, Lee YH, Choi Y, et al. Distinguishing between dementia with lewy bodies and alzheimer's disease using metabolic patterns. Neurobiol Aging. 2020;87:11–7.
21. Venkatraghavan V, Bron EE, Niessen WJ, Klein S. Alzheimer's disease neuroimaging I. disease progression timeline Estimation for alzheimer's disease using discriminative event based modeling. NeuroImage. 2019;186:518–32.
22. Fonteijn HM, Modat M, Clarkson MJ, Barnes J, Lehmann M, Hobbs NZ, et al. An event-based model for disease progression and its application in Familial alzheimer's disease and huntington's disease. NeuroImage. 2012;60(3):1880–9.
23. Iturria-Medina Y, Sotero RC, Toussaint PJ, Mateos-Pérez JM, Evans AC. Early role of vascular dysregulation on late-onset alzheimer's disease based on multifactorial data-driven analysis. Nat Commun. 2016;7:11934.

24. Young AL, Oxtoby NP, Garbarino S, Fox NC, Barkhof F, Schott JM, et al. Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. Nat Rev Neurosci. 2024;25(2):111–30.
25. Archetti D, Ingala S, Venkatraghavan V, Wottschel V, Young AL, Bellio M, et al. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in alzheimer's disease. Neuroimage Clin. 2019;24:101954.
26. Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, et al. A data-driven model of biomarker changes in sporadic alzheimer's disease. Brain. 2014;137(Pt 9):2564–77.
27. Archetti D, Young AL, Oxtoby NP, Ferreira D, Martensson G, Westman E, et al. Inter-Cohort validation of sustain model for alzheimer's disease. Front Big Data. 2021;4:661110.
28. Bilgel M, Jedynak BM. Predicting time to dementia using a quantitative template of disease progression. Alzheimers Dement (Amst). 2019;11:205–15.
29. O'Connor A, Weston PSJ, Pavisic IM, Ryan NS, Collins JD, Lu K, et al. Quantitative detection and staging of presymptomatic cognitive decline in Familial alzheimer's disease: a retrospective cohort analysis. Alzheimers Res Ther. 2020;12(1):126.
30. Staffaroni AM, Quintana M, Wendelberger B, Heuer HW, Russell LL, Cobigo Y, et al. Temporal order of clinical and biomarker changes in Familial frontotemporal dementia. Nat Med. 2022;28(10):2194–206.
31. Wijeratne PA, Eshaghi A, Scotton WJ, Kohli M, Aksman L, Oxtoby NP, et al. The Temporal event-based model: learning event timelines in progressive diseases. Imaging Neurosci (Camb). 2023;1:1–19.
32. Niemantsverdriet E, Ottoy J, Somers C, De Roeck E, Struyfs H, Soetewey F, et al. The cerebrospinal fluid Abeta1-42/Abeta1-40 ratio improves concordance with Amyloid-PET for diagnosing alzheimer's disease in a clinical setting. J Alzheimers Dis. 2017;60(2):561–76.
33. Wittens MMJ, Sima DM, Houbrechts R, Ribbens A, Niemantsverdriet E, Fransen E, et al. Diagnostic performance of automated MRI volumetry by icobrain Dm for alzheimer's disease in a clinical setting: A REMEMBER study. J Alzheimers Dis. 2021;83(2):623–39.
34. Niemantsverdriet E, Ribbens A, Bastin C, Benoit F, Bergmans B, Bier JC, et al. A retrospective Belgian Multi-Center MRI biomarker study in alzheimer's disease (REMEMBER). J Alzheimers Dis. 2018;63(4):1509–22.
35. Wittens MMJ, Denissen S, Sima DM, Fransen E, Niemantsverdriet E, Bastin C, et al. Brain age as a biomarker for pathological versus healthy ageing - a REMEMBER study. Alzheimers Res Ther. 2024;16(1):128.
36. Rakic M, Vercruyssen S, Van Eyndhoven S, de la Rosa E, Jain S, Van Huffel S, et al. Icobrain Ms 5.1: combining unsupervised and supervised approaches for improving the detection of multiple sclerosis lesions. Neuroimage Clin. 2021;31:102707.
37. Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res. 2009;43(4):411–31.
38. Chapman KR, Bing-Canar H, Alosco ML, Steinberg EG, Martin B, Chaisson C, et al. Mini mental state examination and logical memory scores for entry into alzheimer's disease trials. Alzheimers Res Ther. 2016;8(1):9.
39. Salis F, Costaggiu D, Mandas A. Mini-Mental state examination: optimal Cut-Off levels for mild and severe cognitive impairment. Geriatrics. 2023;8(1):12.
40. O'Dowd S, Schumacher J, Burn DJ, Bonanni L, Onofrj M, Thomas A, et al. Fluctuating cognition in the lewy body dementias. Brain. 2019;142(11):3338–50.
41. Matar E, Shine JM, Halliday GM, Lewis SJG. Cognitive fluctuations in lewy body dementia: towards a pathophysiological framework. Brain. 2019;143(1):31–46.
42. Matar E, Ehgoetz Martens KA, Phillips JR, Wainstein G, Halliday GM, Lewis SJG, et al. Dynamic network impairments underlie cognitive fluctuations in lewy body dementia. Npj Parkinson's Disease. 2022;8(1):16.
43. Grothe MJ, Moscoso A, Silva-Rodriguez J, Lange C, Nho K, Saykin AJ, et al. Differential diagnosis of amnestic dementia patients based on an FDG-PET signature of autopsy-confirmed LATE-NC. Alzheimers Dement. 2023;19(4):1234–44.
44. Barba L, Abu-Rumeileh S, Barthel H, Massa F, Foschi M, Bellomo G, et al. Clinical and diagnostic implications of alzheimer's disease copathology in lewy body disease. Brain. 2024;147(10):3325–43.
45. Nelson PT, Schneider JA, Jicha GA, Duong MT, Wolk DA. When alzheimer's is LATE: why does it matter?? Ann Neurol. 2023;94(2):211–22.
46. Jannati A, Toro-Serey C, Gomes-Osman J, Banks R, Ciesla M, Showalter J, et al. Digital clock and recall is superior to the Mini-Mental state examination for the detection of mild cognitive impairment and mild dementia. Alzheimers Res Ther. 2024;16(1):2.

Publisher's note