NeuroImage 238 (2021) 118233

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Progression along data-driven disease timelines is predictive of Alzheimer's disease in a population-based cohort

Vikram Venkatraghavan[a,1], Elisabeth J. Vinke[a,c,1], Esther E. Bron[a], Wiro J. Niessen[a,b], M. Arfan Ikram[c], Stefan Klein[a,2], Meike W. Vernooij[a,c,2,*], for the Alzheimer's Disease Neuroimaging Initiative[#]

[a] Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam, The Netherlands
[b] Quantitative Imaging Group, Dept. of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands
[c] Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
Disease progression modeling
Event-based model
Alzheimer's disease
APOE
Population study

ABSTRACT

Data-driven disease progression models have provided important insight into the timeline of brain changes in AD phenotypes. However, their utility in predicting the progression of pre-symptomatic AD in a population-based setting has not yet been investigated. In this study, we investigated if the disease timelines constructed in a case-controlled setting, with subjects stratified according to *APOE* status, are generalizable to a population-based cohort, and if progression along these disease timelines is predictive of AD. Seven volumetric biomarkers derived from structural MRI were considered. We estimated *APOE*-specific disease timelines of changes in these biomarkers using a recently proposed method called co-initialized discriminative event-based modeling (co-init DEBM). This method can also estimate a disease stage for new subjects by calculating their position along the disease timelines. The model was trained and cross-validated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, and tested on the population-based Rotterdam Study (RS) cohort. We compared the diagnostic and prognostic value of the disease stage in the two cohorts. Furthermore, we investigated if the rate of change of disease stage in RS participants with longitudinal MRI data was predictive of AD. In ADNI, the estimated disease timeslines for $\epsilon 4$ non-carriers and carriers were found to be significantly different from one another ($p < 0.001$). The estimate disease stage along the respective timelines distinguished AD subjects from controls with an AUC of 0.83 in both *APOE* $\epsilon 4$ non-carriers and carriers. In the RS cohort, we obtained an AUC of 0.83 and 0.85 in $\epsilon 4$ non-carriers and carriers, respectively. Progression along the disease timelines as estimated by the rate of change of disease stage showed a significant difference ($p < 0.005$) for subjects with pre-symptomatic AD as compared to the general aging population in RS. It distinguished pre-symptomatic AD subjects with an AUC of 0.81 in *APOE* $\epsilon 4$ non-carriers and 0.88 in carriers, which was better than any individual volumetric biomarker, or its rate of change, could achieve. Our results suggest that co-init DEBM trained on case-controlled data is generalizable to a population-based cohort setting and that progression along the disease timelines is predictive of the development of AD in the general population. We expect that this approach can help to identify at-risk individuals from the general population for targeted clinical trials as well as to provide biomarker based objective assessment in such trials.

1. Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease that affects roughly 3% of the world's elderly population (above 60 years old) (World Health Organization, 2017). A major genetic risk factor for AD is the presence of $\epsilon 4$ allele of *APOE* (Van Cauwenberghe et al., 2016). Furthermore, *APOE* $\epsilon 4$ has also been shown to affect the clinical (Holmes, 2002; Weintraub et al., 2019) and biological phenotypes

* Corresponding author.
E-mail address: m.vernooij@erasmusmc.nl (M.W. Vernooij).

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

[1] Denote equal contributions.
[2] Denote equal contributions.

https://doi.org/10.1016/j.neuroimage.2021.118233.
Received 31 October 2020; Received in revised form 11 April 2021; Accepted 1 June 2021
Available online 4 June 2021.

of AD (Ferreira et al., 2020), making it a key factor in understanding the pathophysiology of AD.

Neuroimaging biomarkers play an important role in disentangling these phenotypes (Ryan et al., 2018; Young et al., 2018). They could also play an important role in finding disease modifying treatments (Devi and Scheltens, 2018). There has been evidence that selection of the study population at its pre-symptomatic stage is also crucial for the success of potential modifying treatments for AD (Sevigny et al., 2016; Sperling et al., 2013). Hence there is a crucial need for a way to objectively assess the progression of pre-symptomatic AD (or lack thereof).

Biomarkers extracted from neuroimaging data in combination with machine learning approaches have been shown to objectively assess the progression of AD in research cohorts (Marinescu et al., 2020) as well as in clinical cohorts (Kloeppel et al., 2015). However, machine learning approaches are not explainable by default and the lack of transparency in such approaches could hinder clinical decision making (Wachter et al., 2018).

Disease progression models are data-driven approaches that are interpretable by design and can thus aid not only in predicting AD but also in explaining the decision and facilitating transparency and trust (Holzinger et al., 2017). In recent years, many disease progression models have emerged to provide insight into neurodegenerative diseases such as AD (Donohue et al., 2014; Fonteijn et al., 2012). Such insights have also been shown to aid in objective assessment of AD progression (Koval et al., 2018). An example of such a model is the discriminative event-based model (DEBM) (Venkatraghavan et al., 2019), which estimates a timeline of AD related biomarker abnormality events in a data-driven way. This model was recently extended further to identify *APOE* genotype-specific differences in AD biomarker progression, where the biomarkers, including volumetric measures obtained from MRI, were found to progress along different timelines depending on *APOE* status (Venkatraghavan et al., 2021). However, the generalizability of such models to population-based cohorts and their utility in predicting the progression of pre-symptomatic AD in a population-based setting have not yet been investigated.

In this work, we investigate if i) *APOE*-specific disease timelines constructed in a case-controlled setting are generalizable to a population-based cohort, and ii) if progression along these disease timelines is predictive of AD. For constucting the *APOE*-specific disease timelines, we use a recently developed approach called co-initialized (co-init) DEBM (Venkatraghavan et al., 2021) meant for obtaining disease timelines in stratified cross-sectional datasets. We demonstrate the potential of the method's fine-grained disease stage estimation in predicting the subjects with pre-symptomatic AD in the general population.

2. Methods

We first describe the inclusion criteria for participants and the method for obtaining the volumetric biomarkers in the case-controlled Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and the population-based Rotterdam study (RS) dataset. This is followed by the description of co-init DEBM used to construct *APOE*-specific disease timelines of volumetric biomarkes from baseline scans of the participants in the ADNI. We validated the disease timelines constructed on ADNI by assessing their generalizability to the population-based RS cohort, and by predicting the participants at-risk of becoming symptomatic in the RS cohort.

2.1. Participants

2.1.1. ADNI

We considered the baseline measurements of 335 cognitively normal (CN), 565 non-AD, 167 incident-AD and 223 AD participants (prevalent-AD) who had imaging data available in ADNI1, ADNIGO and ADNI2 studies[3] The non-AD cases were defined as ADNI participants who were either mild cognitively impaired (MCI) or had subjective memory complaints at the time of the baseline MRI scan, and did not develop AD within 3 years of follow-up. The incident-AD cases presented with MCI at baseline but developed AD within 3 years. The prevalent-AD and incident-AD subjects were defined by their clinical diagnosis of AD according to NINCDS-ADRDA's criteria for AD (Dubois et al., 2007; Petersen et al., 2010). Characteristics of the subjects and their volumetric measures in the ADNI dataset included in our study are shown in Table 1(a).

2.1.2. Rotterdam study

We considered participants from the population-based RS cohort, a prospective longitudinal study among community-dwelling subjects aged 45 years and over (Ikram et al., 2020). Participants were screened for dementia at baseline and at follow-up examinations with the Mini-Mental State Examination and the Geriatric Mental Schedule organic level. Those with a Mini-Mental State Examination score < 26 or Geriatric Mental Schedule score > 0 underwent further investigation and informant interview, including the Cambridge Examination for Mental Disorders of the Elderly. In addition, the entire cohort was continuously under surveillance for dementia through electronic linkage of the study database with medical records from general practitioners and the regional institute for outpatient mental health care. Available information on cognitive testing and clinical neuroimaging was used when required for diagnosis of dementia subtype. A consensus panel led by a consultant neurologist established the final diagnosis of AD according to NINCDS-ADRDA criteria for AD.

In this work, we included participants from the RS who had at least one MRI scan, who completed cognitive testing, and were interviewed for the presence of subjective cognitive complaints at the time of the MRI. The included participants were categorized into 4 groups: participants that were cognitively normal at the time of the scan (CN), participants that had subjective memory complaints and/or objective cognitive impairment (de Bruijn et al., 2014), but who did not develop AD at follow-up (non-AD), participants with AD at the time of the scan (prevalent-AD) and participants who developed AD after the MRI scan (incident-AD). Unlike in ADNI, we did not set a threshold of conversion within 3 years to be included as an incident-AD participant, since we wanted to assess the utility of our method in monitoring the progression of both pre-clinical and prodromal AD subjects. Participants with clinical stroke were excluded.

In our experiments, we used two subsets of the RS cohort: the *generalizibility set* and the *prediction set*. The generalizibility set consisted of the last MRI scan available for each partipant in the RS cohort. This subset consisted of 998 CN, 2710 non-AD, 97 incident-AD, and 25 prevalent-AD cases and were used for experiments validating the generalizability of the *APOE*-specific disease timelines constructed using co-init DEBM. The characteristics of the subjects in this subset are shown in Table 1(b). The prediction set consisted of the last two MRI scans available for each participant, which were used to assess the progression (or lack thereof) of pre-symptomatic AD in the participants. This subset consisted of 183 CN, 852 non-AD and 31 incident-AD cases. For the incident-AD cases, both the included scans were performed before the AD diagnosis. Participants with prevalent-AD were excluded in this subset. The characteristics of the subjects in this subset are shown in Table 1(c). A scatter plot illustrating the longitudinal sampling in this prediction set is shown in Figure 1.

[3] ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). For up-to-date information, see www.adni-info.org.

Table 1
Characteristics of the ADNI dataset (a), the generalizability set of the RS dataset (b), and the prediction set of the RS dataset (c). * indicates values at last scan.

ADNI dataset	CN	non-AD	incident-AD	prevalent-AD
Number of subjects	335	565	167	223
Number of women, %	174, 51.9	268, 47.4	68, 40.7	104, 46.6
Age (years)	74.3 ± 5.6	71.82 ± 7.2	73.1 ± 7.1	74.0 ± 7.9
Number of *APOEε*4 carriers, %	92, 27.5	238, 42.1	121, 72.5	151, 67.7
Intracranial volume (ml)	1504.0 ± 155.8	1520.9 ± 152.8	1546.2 ± 180.2	1524.2 ± 183.9
Total brain volume (ml)	1030.7 ± 98.7	1043.3 ± 100.0	1017.7 ± 111.7	991.8 ± 114.1
Ventricle volume (ml)	38.4 ± 18.1	41.0 ± 21.3	49.1 ± 23.9	51.4 ± 21.9
Hippocampus volume (ml)	7.3 ± 0.9	7.1 ± 1.0	6.3 ± 1.0	6.0 ± 1.0
Precuneus volume (ml)	16.7 ± 2.2	17.4 ± 2.4	16.2 ± 2.6	15.4 ± 2.5
Middle temporal gyrus volume (ml)	20.4 ± 2.7	20.4 ± 2.7	18.5 ± 2.9	17.6 ± 3.0
Fusiform gyrus volume (ml)	17.5 ± 2.1	17.6 ± 2.2	16.3 ± 2.4	15.5 ± 2.4
Entorhinal cortex volume (ml)	4.0 ± 0.7	3.9 ± 0.8	3.4 ± 0.8	3.2 ± 0.8
Time before AD diagnosis (years)*			1.4 ± 0.7	
		(a)		
RS dataset - generalizability set	CN	non-AD	incident-AD	prevalent-AD
Number of subjects	998	2710	97	25
Number of women, %	500, 50.1	1200, 44.3	39, 40.2	10, 40.0
Age (years)	67.4 ± 8.3	70.9 ± 9.3	79.6 ± 5.7	80.2 ± 6.3
Number of *APOEε*4 carriers, %	255, 25.6	745, 27.5	45, 46.4	11, 44.0
Intracranial volume (ml)	1512.3 ± 157.6	1475.8 ± 155.3	1437.5 ± 156.6	1403.0 ± 163.9
Total brain volume (ml)	1050.3 ± 107.5	1012.6 ± 105.6	936.6 ± 94.9	884.5 ± 105.0
Ventricle volume (ml)	33.7 ± 17.3	36.5 ± 19.3	49.1 ± 21.1	59.9 ± 28.3
Hippocampus volume (ml)	7.9 ± 0.8	7.6 ± 0.8	6.7 ± 0.9	6.0 ± 1.0
Precuneus volume (ml)	18.2 ± 2.1	17.6 ± 2.0	16.8 ± 1.9	15.4 ± 2.2
Middle temporal gyrus volume (ml)	20.6 ± 2.7	19.9 ± 2.7	17.6 ± 2.5	16.2 ± 2.7
Fusiform gyrus volume (ml)	17.7 ± 2.2	17.2 ± 2.1	15.8 ± 2.0	14.5 ± 2.7
Entorhinal cortex volume (ml)	3.7 ± 0.6	3.6 ± 0.7	3.1 ± 0.8	2.6 ± 0.7
Time before AD diagnosis (years)*			2.8 ± 2.3	
		(b)		
RS dataset - prediction set	CN	non-AD	incident-AD	
Number of subjects	183	852	31	
Number of women, %	95, 51.9	412, 48.4	10, 32.3	
Age (years)*	73.3 ± 5.5	75.5 ± 6.4	78.4 ± 6.8	
Follow-up time (years)	3.5 ± 1.3	3.5 ± 1.4	2.9 ± 0.9	
Number of *APOEε*4 carriers, %	39, 21.3	225, 26.4	13, 41.9	
Intracranial volume (ml)*	1522.8 ± 156.6	1478.9 ± 156	1419.4 ± 126.9	
Total brain volume (ml)*	1038.7 ± 100.7	998.4 ± 98.3	926.6 ± 91.4	
Ventricle volume (ml)*	39.7 ± 20.2	41.1 ± 21.6	44.9 ± 17	
Hippocampus volume (ml)*	7.8 ± 0.8	7.4 ± 0.8	6.7 ± 0.9	
Precuneus volume (ml)*	18.0 ± 2.0	17.5 ± 1.9	16.5 ± 1.8	
Middle temporal gyrus volume (ml)*	20.3 ± 2.6	19.5 ± 2.4	17.5 ± 2.4	
Fusiform gyrus volume (ml)*	17.5 ± 2.1	17.0 ± 2.1	15.6 ± 2.1	
Entorhinal cortex volume (ml)*	3.7 ± 0.7	3.6 ± 0.7	3.0 ±0.6	
Time before AD diagnosis (years)*			2.4 ± 1.8	
		(c)		

2.2. MRI Acquisition and imaging biomarker extraction

The imaging biomarkers used in this study were estimated from T1-weighted (T1w) MRI scans. ADNI participants were scanned on a 1.5T ($N = 497$) or a 3T ($N = 793$) MRI system from GE, Philips, or Siemens, using magnetization prepared - rapid gradient echo (MP-RAGE) sequence (voxel size: $1.0 \times 1.0 \times 1.0$ mm^3). RS participants were scanned on a single 1.5T MRI system from GE, using gradient recalled echo (GRE) sequence (voxel size: $0.49 \times 0.49 \times 1.6$ mm^3). Details of the MRI acquisition protocol can be found in Jack et al. (2015, 2008) (ADNI) and Ikram et al., 2015 (RS). The MRI scans were analyzed with FreeSurfer software v6.0 cross-sectional stream (http://surfer.nmr.mgh.harvard.edu). Outputs were visually checked for the ADNI dataset. In the RS dataset, an automated quality metric was used to exclude scans with insufficient quality, which was visually verified in a randomly selected subset of both selected and rejected scans (Lamballais et al., 2020).

The selected imaging markers were the same markers as that of Archetti et al., 2019, namely volumetric measures of: total brain, ventricles, hippocampus, precuneus, middle temporal gyrus, fusiform gyrus and entorhinal cortex. The volumes were defined as the summed volumes of the structure in the left and right hemisphere. To take into account the confounding effects of age, sex, and intracranial volume, linear regressions were performed before constructing the disease timelines. The volumetric measures of CN subjects in ADNI were used to regress against age, sex and intracranial volume to estimate their confounding effects parameterized by their respective slopes and intercepts. These estimates were used for confounding factor correction in the remaining subjects in ADNI as well as in the RS cohort. The resultant volumetric measures will be referred to as biomarkers in the remainder of the manuscript.

2.3. Construction of APOE-specific disease timelines using co-init DEBM

The co-init DEBM model introduced in (Venkatraghavan et al., 2021) constructs genotype-specific AD related disease timelines of biomarker changes, based on *cross-sectional* datasets. Such an estimation from cross-sectional data is feasible because, in a cohort consisting of subjects encompassing a wide spectrum of severity, early biomarkers have a higher prevalence of abnormal biomarker values as compared to biomarkers

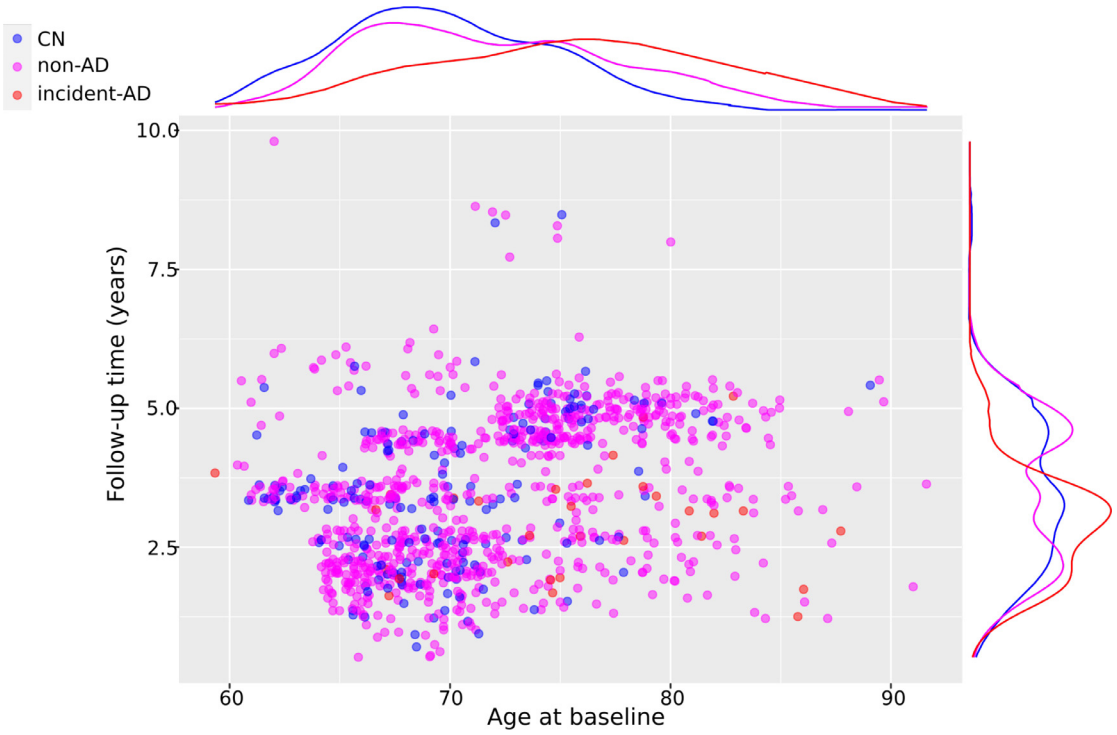


Fig. 1. Longitudinal sampling in the prediction set of the RS dataset. The x-axis represents the age of the participant at baseline and the y-axis represents time difference between the baseline and follow-up scan. The plot on the top of the figure shows the kernel density estimates of the age of the participants for the different diagnostic classes and the one on the right shows the kernel density estimates of the follow-up time.

that become abnormal later in the disease timeline. The co-init DEBM model estimates this timeline without strictly considering the diagnostic labels of the subjects. The model uses a coupled mixture model to jointly fit normal and abnormal distributions in the dataset stratified by (*APOE*) genotypes. The model assumes that the normal and abnormal biomarker distributions in the different genotypes can be approximately represented by Gaussians. It also assumes that the different genotypes' abnormal (and normal) biomarker distributions are close to each other.

After the estimation of the normal and abnormal biomarker distributions, the model computes the probability of abnormality of each biomarker for each subject in the training dataset. Based on the assumption that a biomarker that becomes abnormal earlier in the disease timeline would be more abnormal than the biomarker that becomes abnormal later, it estimates a subject-specific ordering of biomarker changes in each subject of the dataset. A generalized Mallows model is used to average the subject-specific biomarker ordering over the subjects within each genotypic group of the training set, to construct average disease timeline for *APOE* $\epsilon 4$ non-carriers and carriers. Along with the sequence of the biomarker abnormality events, the model also estimates the relative positioning of such events with respect to each other (event-centers). Absolute magnitudes for these event-centers are irrelevant as they only convey relative (temporal) distances and in this study, they were normalized such that the first event and the last event coincided at a value of 0.1 and 0.9 respectively.

To construct the disease timelines, the co-init DEBM was trained on CN, incident-AD, and prevalent-AD subjects from ADNI. The non-AD subjects in ADNI were excluded for training the model, to reduce the chances of disorders unrelated to AD affecting the estimated timelines. The variance in the estimated disease timeline was computed using 100 independent bootstrap samples. In order to evaluate if the estimated orderings in *APOE* $\epsilon 4$ non-carriers and carriers were significantly different from one another, we used permutation testing and estimated the distribution of the Kendall's Tau distance under the null hypothesis. To compute this distribution, we generated 1,000 random permutations of the two groups. We then computed the one-sided p-values for the actual Kendall's Tau distances between the orderings of the two groups, calculated as the proportion of sampled permutations where the distance was greater than or equal to the actual distance.

2.4. *Estimating* APOE-*specific disease stages*

After training the co-init DEBM model, the constructed *APOE*-specific disease timelines were used to estimate the disease stage at multiple timepoints for subjects of the RS cohort. For estimating the disease stages of ADNI subjects, we used a 10-fold cross validation. The training set was used for constructing the disease timelines and the disease stages were estimated in the test set, including the non-AD subjects excluded in the training phase. Disease stage quantifies the severity of the disease in a subject by positioning them along the pre-constructed disease timelines and is normalized between 0 and 1. The estimated disease stages were used in two sets of experiments.

Experiment 1: *Assessing the generalizability of co-init DEBM from ADNI to RS*

In this experiment, we tested the generalizability of the co-init DEBM model trained on ADNI by evaluating the diagnostic and prognostic value of its predicted disease stages in the RS cohort. First we performed a visual assessment by constructing normalized histograms of the estimated *APOE*-specific disease stages for the different diagnostic classes in ADNI and the generalizability set of the RS cohort.

Complementing this visual analysis, for assessing the diagnostic value we used the estimated disease stages to distinguish prevalent-AD from two different reference groups in ADNI and in the generalizability set of the RS cohort. First, only the CN subjects were included in the reference group. To emulate a reference group of participants more representative of the general aging population than the CN group, we used a combined set of CN and non-AD subjects as the second reference group. We computed the area under the receiver operating curve (AUC) for distinguishing the diagnostic classes, and compared the AUCs obtained in

Table 2
Generalizability assessment: The AUCs for distinguishing the different diagnostic classes using the estimated disease stages and their corresponding 95% confidence intervals. The confidence intervals of the AUCs were determined using bootstrap resampling while stratifying the diagnostic classes to maintain their relative proportions. Co-init DEBM AUC represents the AUCs obtained when separate disease timelines were estimated for *APOE* ϵ4 non-carriers and carriers, whereas DEBM AUC represents the AUCs obtained when a combined disease timeline was estimated. N_R and N_C represent the number of subjects in the reference group and number of cases respectively.

Reference group	Cases	No. of Subjects		Co-init DEBM AUC		DEBM AUC	
		ADNI (N_R, N_C)	RS (N_R, N_C)	ADNI	RS	ADNI	RS
APOE ϵ4 non-carriers							
CN	prevalent-AD	243, 72	743, 14	0.86 (0.81-0.91)	0.85 (0.71-0.98)	0.85 (0.80-0.90)	0.79 (0.64-0.94)
CN + non-AD	prevalent-AD	570, 72	2708, 14	0.83 (0.78-0.88)	0.83 (0.70-0.97)	0.81 (0.76-0.86)	0.79 (0.64-0.94)
CN	incident-AD	243, 46	743, 52	0.83 (0.77-0.90)	0.70 (0.62-0.78)	0.83 (0.77-0.89)	0.63 (0.55-0.72)
CN + non-AD	incident-AD	570, 46	2708, 52	0.81 (0.74-0.88)	0.68 (0.60-0.75)	0.80 (0.73-0.86)	0.64 (0.55-0.72)
APOE ϵ4 carriers							
CN	prevalent-AD	92, 151	255, 11	0.89 (0.85-0.94)	0.85 (0.74-0.96)	0.92 (0.87-0.96)	0.84 (0.71-0.97)
CN + non-AD	prevalent-AD	330, 151	1000, 11	0.83 (0.79-0.86)	0.85 (0.74-0.95)	0.83 (0.80-0.87)	0.85 (0.72-0.98)
CN	incident-AD	92, 121	255, 45	0.87 (0.82-0.92)	0.63 (0.54-0.72)	0.88 (0.83-0.93)	0.62 (0.52-0.72)
CN + non-AD	incident-AD	330, 121	1000, 45	0.79 (0.74-0.83)	0.62 (0.54-0.71)	0.79 (0.74-0.83)	0.62 (0.52-0.72)

ADNI and RS. The confidence intervals of these AUCs were measured using bootstrap resampling while stratifying the diagnostic classes to maintain their relative proportions. For assessing the prognostic value, we used the estimated disease stages to distinguish incident-AD from the aforementioned two reference groups in ADNI and in the generalizability set of RS cohort. We computed the AUCs and their confidence intervals for distinguishing these diagnostic classes and compared values obtained in ADNI and RS.

To compare the generalizability of a model that stratifies based on *APOE* carriership, with that of a model that does not, we repeated the experiment described above using disease timeline estimated in ADNI subjects, without stratifying for *APOE*. Furthermore, we computed the correlation of the estimated disease stages with time to dementia diagnosis for incident-AD subjects in ADNI as well as in RS. Lastly, we computed the Spearman correlation of the estimated disease stages with MMSE for subjects in ADNI as well as in RS.

Experiment 2: *Predicting AD based on longitudinal data in the RS cohort*

In this experiment, we assess if the evolution of the disease stages derived from longitudinal neuroimaging data is predictive of AD in the prediction set of the RS cohort. This experiment is further divided into three parts. In the first part, we build longitudinal trajectories of the disease stages and observe the differences in CN, non-AD and incident-AD subjects. In the second part, we assess the prognostic value of the rate of change of disease stages. Lastly, we assess the marginal utility of the follow-up scans in AD prognostication.

Exp. 2.1: We used the disease stages obtained in the prediction set of the RS cohort for building the trajectories of disease stages in the two *APOE* ϵ4 based groups. The trajectories were estimated using linear mixed models with random intercepts and slopes. The time variable in these linear mixed models was follow-up time in years since the first MRI of the subject. To allow different slopes for different diagnostic classes, an interaction between follow-up time and the diagnosis was integrated in the model. Covariates that were accounted for in the model were sex, age at the time of the first MRI, and the interaction of age and follow-up time to allow slope differences for different ages.

Exp. 2.2: We used the rate of change of disease stages (delta disease stage) in the prediction set of the RS cohort to distinguish incident-AD from two different reference groups. As in Experiment 1, the two reference groups selected were CN, and a combined set of CN and non-AD subjects. We computed the AUCs and their confidence intervals for distinguishing these diagnostic classes. For comparison, the AUCs while using the rate of change of the volumetric measures (normalized to their respective intracranial volumes) for distinguishing the same two classes were computed.

Exp. 2.3: Lastly, to evaluate the marginal utility of the follow-up scans for identifying incident-AD subjects, we used the estimated disease stage at the last MRI scan of the subjects in the prediction set of the RS cohort to distinguish incident-AD from the aforementioned two different reference groups. We computed the AUCs and their confidence intervals for distinguishing these diagnostic classes. As a comparison, the AUCs based on participants' age as well as of each individual volumetric imaging biomarker were also computed.

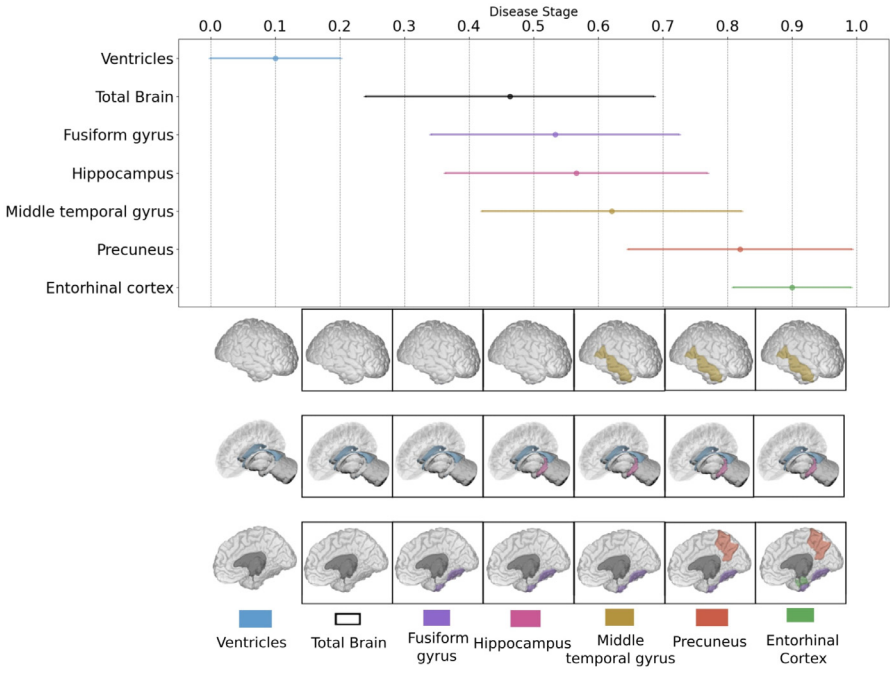
3. Results

Figure 2 shows the *APOE*-specific disease timelines constructed for the ϵ4 non-carriers and carriers in the ADNI dataset. It shows the centers of the biomarker abnormality events along the timeline representing their relative positioning with respect to each other. It can be seen that the disease timelines of *APOE* ϵ4 non-carriers and carriers were quite different. The permutation testing further confirmed that the disease timelines of ϵ4 non-carriers and carriers were indeed significantly different ($p < 0.001$). Most noticeably, ventricular volume and total brain volume were estimated as early biomarkers for *APOE* ϵ4 non-carriers, whereas hippocampal volume and volume of the entorhinal cortex were estimated as early biomarkers for *APOE* ϵ4 carriers. It can also be seen in Fig. 2 that the uncertainty estimates in *APOE* ϵ4 non-carriers were greater than in *APOE* ϵ4 carriers.

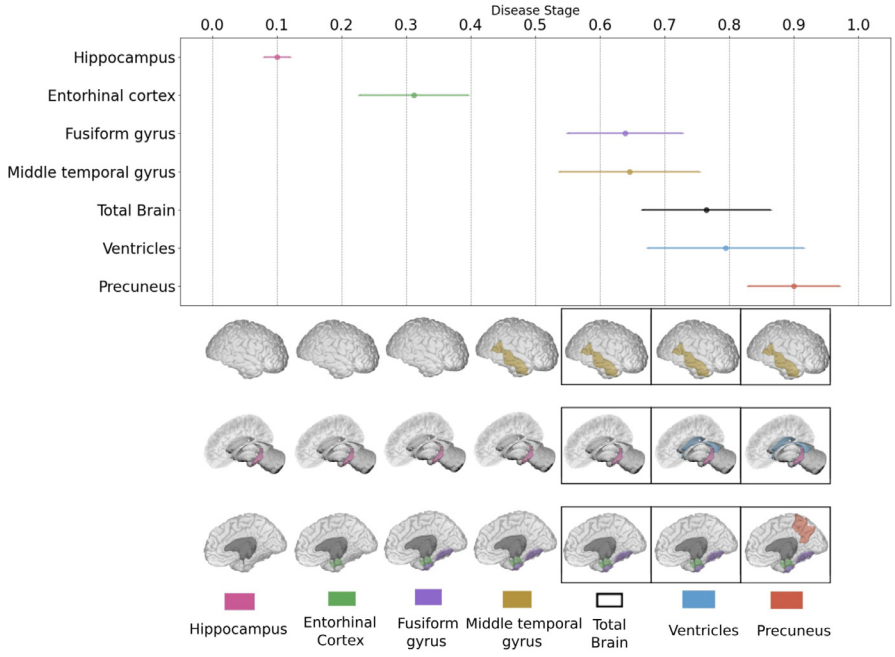
Experiment 1: *Assessing the generalizability of co-init DEBM from ADNI to RS*

The normalized histograms of the estimated *APOE*-specific disease stages for the different diagnostic classes in ADNI and the generalizability set of RS are shown in Fig. 3. It can be seen that the distributions of the disease stages of the four diagnostic classes in ADNI were largely similar to those in the generalizability set of RS. The CN and non-AD subjects were positioned towards the left side of the spectrum, whereas the prevalent-AD were positioned predominantly towards the right. It can also be seen that for a proportion of prevalent-AD subjects in the *APOE* ϵ4 non-carrier group, the model had estimated a low disease stage in both ADNI and RS cohorts. A noticeable difference between ADNI and RS was that a substantial proportion of incident-AD subjects in RS was positioned towards the left side of the histograms in both *APOE* ϵ4 non-carriers and carriers.

The AUCs for distinguishing the different diagnostic classes using the estimated disease stages are shown in Table 2, along with their confidence intervals. It can be observed that the performance of the disease stages obtained using co-init DEBM in distinguishing prevalent-AD from the set of CN and non-AD subjects in ADNI (AUC = 0.83 for both *APOE* ϵ4 non-carriers and carriers) was comparable to that in RS (AUC = 0.83 for *APOE* ϵ4 non-carriers and AUC = 0.85 for ϵ4 carriers). It should however be noted that the confidence intervals were larger in the RS cohort. It can also be observed that incident-AD subjects were harder to distinguish than prevalent-AD in the RS cohort (Co-init DEBM: AUC = 0.68



(a)



(b)

Fig. 2. Disease timelines of *APOE* $\epsilon 4$ non-carriers (a) and carriers (b) estimated using co-init DEBM in ADNI. The plot on top of each subfigure shows the event-centers of the different regions and their respective standard deviation estimated from a batch of 100 independent bootstrap samples. The 3D visualization (Marinescu et al., 2019) at the bottom of each subfigure highlights the region that becomes abnormal at the corresponding disease stage. Total brain volume becoming abnormal is depicted by a black-box surrounding the brain at the corresponding disease stage. The vertical positioning of the biomarkers in the event-center part of each subfigure shows the estimated disease timeline in the *APOE* genotype, which is different for non-carriers and carriers.

for $\epsilon 4$ non-carriers and AUC = 0.62 for $\epsilon 4$ carriers), but not in ADNI (Co-init DEBM: AUC = 0.81 for $\epsilon 4$ non-carriers and AUC = 0.79 for $\epsilon 4$ carriers). It can also be seen in Table 2 that, while AUCs in ADNI are comparable for both DEBM and Co-init DEBM and in *APOE* $\epsilon 4$ carriers in RS, the AUCs for distinguishing the different groups in RS *APOE* $\epsilon 4$ non-carriers is higher for co-init DEBM.

Furthermore, the estimated disease stages showed a significant Pearson correlation with time to diagnosis for *APOE* $\epsilon 4$ carrier incident-AD subjects in both ADNI ($R = 0.31$, $p = 0.0006$) and RS cohorts ($R = 0.29$, $p = 0.04$). However, the correlation was found to be insignificant for *APOE* $\epsilon 4$ non-carrier incident-AD subjects in both ADNI ($R = 0.04$, $p = 0.8$) and RS cohorts ($R = 0.1$, $p = 0.4$). Lastly, the obtained disease stages had a significant Spearman correlation with MMSE in both ADNI non-carriers ($R = -0.41$, $p < 0.001$) and carriers ($R = -0.48$, $p < 0.001$) as well as in RS non-carriers ($R = -0.08$, $p < 0.001$) and carriers ($R = -0.06$, $p = 0.05$).

Experiment 2: *Predicting AD based on longitudinal data in the RS cohort*

Exp. 2.1: In Fig. 4, the trajectories of disease stage over time as estimated by linear mixed models are shown for the CN, non-AD and incident-AD groups of the prediction set of RS. The interaction between the incident-AD diagnosis and follow-up time was statistically significant in both *APOE* $\epsilon 4$ non-carriers and carriers (CN vs. incident-AD $p = 0.0032$ and $p = 0.0041$ respectively; non-AD vs. incident-AD $p = 0.0039$

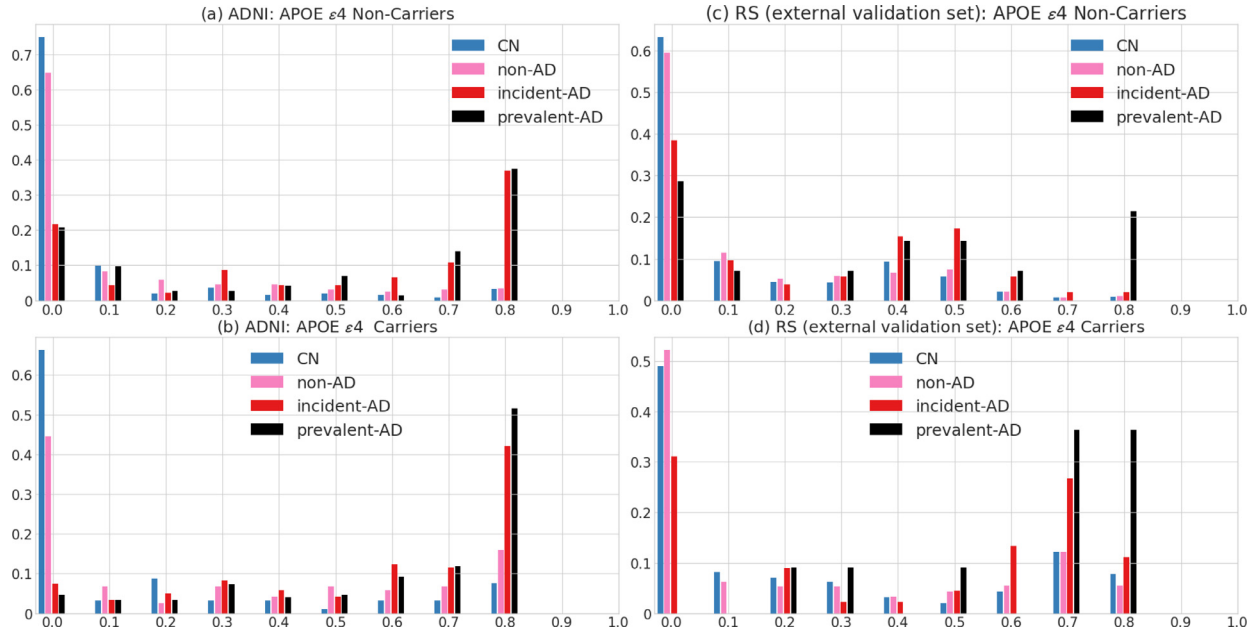


Fig. 3. Normalized histograms of the estimated *APOE*-specific disease stages for the different diagnostic classes in ADNI and the generalizability set of RS. The normalized histograms of disease stages are shown for (a) *APOE* $\epsilon 4$ non-carriers in ADNI, (b) *APOE* $\epsilon 4$ carriers in ADNI (c) *APOE* $\epsilon 4$ non-carriers of the generalizability set in RS, and (d) *APOE* $\epsilon 4$ carriers of the generalizability set in RS. The x-axis represents the disease stage based on the *APOE*-specific disease timeline by the co-init DEBM model, and the y-axis represents the relative percentage of subjects in each diagnostic class, meaning that the relative percentages of all disease stages of one diagnostic category add up to one. Estimated disease stage is a continuous variable and was discretized (binned) for visualization purposes only.

and $p = 0.0032$ respectively), meaning that incident-AD subjects showed a significant increase in disease stage compared to CN and non-AD subjects.

Exp. 2.2: In the left column of Fig. 5, the AUCs and the corresponding 95% confidence intervals for distinguishing incident-AD using two MRI scans based on longitudinal follow-up of participants are shown for *APOE* $\epsilon 4$ non-carriers and carriers. It can be observed that for distinguishing incident-AD from the reference group, delta disease stage consistently performed the best for both the genotypes. It outperformed the rates of changes of volumetric measures, with respect to the obtained AUC. It can also be observed that distinguishing incident-AD from CN and non-AD subjects in the reference group was harder than distinguishing incident-AD from CN alone, as reflected by the lower AUCs for almost all the measures used.

Exp. 2.3: The right column of Fig. 5 shows that age was an important predictor for incident-AD. Age distinguished incident-AD well from CN subjects (AUC of 0.73 for both $\epsilon 4$ non-carriers and carriers), but the performance of age as a predictor dropped substantially when distinguishing incident-AD from CN and non-AD subjects (AUC of 0.64 for $\epsilon 4$ non-carriers and 0.65 for $\epsilon 4$ carriers). When only the last MRI scan was used for incident-AD prediction from a reference group of CN and non-AD subjects, volumes of hippocampus and entorhinal cortex were good indicators in *APOE* $\epsilon 4$ carriers (AUC of 0.79 and 0.81 respectively) but not for *APOE* $\epsilon 4$ non-carriers (AUC of 0.64 and 0.63 respectively). Similarly, total brain volume and ventricle volume were good indicators of incident-AD in *APOE* $\epsilon 4$ non-carriers (AUC of 0.73 and 0.68 respectively), but not for $\epsilon 4$ carriers (AUC of 0.64 and 0.59 respectively). Disease stage estimated using the *APOE*-specific disease timeline performed well consistently in both the *APOE* genotypes (AUC of 0.74 for $\epsilon 4$ non-carriers and 0.76 carriers). The marginal utility of an additional MRI scan can be observed by comparing the left column of Fig. 5 with the right column of Fig. 5. It can be seen that delta disease stage was much better for incident-AD prediction from a reference group of CN and non-AD subjects (AUC of 0.81 for $\epsilon 4$ non-carriers and 0.88 for carriers) than any measure obtained using only the last MRI scan.

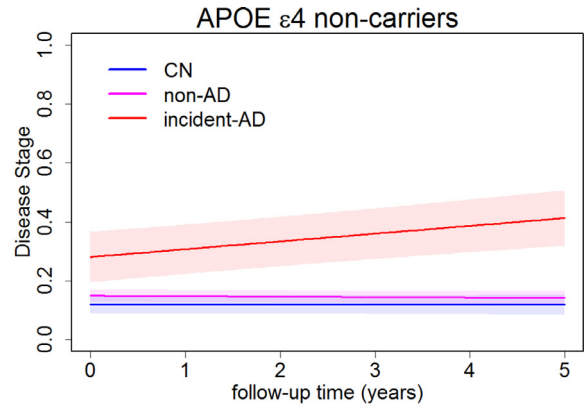
4. Discussion

In this work, we constructed *APOE*-specific disease timelines in a case-controlled setting and validated their generalizability to a population-based setting. We assessed that progression along these timelines is predictive of AD in the general population. In this section, we discuss the insights we obtained from our results.

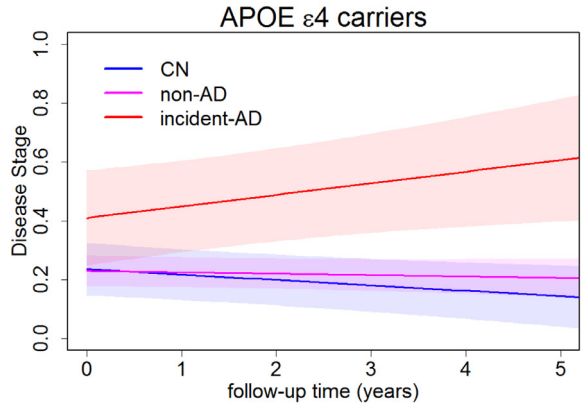
4.1. *Generalizability of the* APOE*-specific disease timelines*

The disease timelines estimated for *APOE* $\epsilon 4$ non-carriers and carriers were significantly different from one another and highlighted the *APOE*-genotype-specific differences in the loss of structural integrity as AD progresses. Ventricular volume and total brain volume were early biomarkers for $\epsilon 4$ non-carriers, and hippocampal volume and volume of the entorhinal cortex were early biomarkers for $\epsilon 4$ carriers. We observed in the normalized histograms that for a proportion of prevalent-AD subjects in the $\epsilon 4$ non-carriers group, the model had estimated a low disease stage. This observation, in combination with the greater uncertainty of the event-centers in that group suggests that there is intra-genotype heterogeneity among the $\epsilon 4$ non-carriers.

The disease timelines were estimated after correcting for the confounding effect of age, assuming a linear relationship of volumetric biomarkers with respect to age. Non-linear biomarker relationship with age such as the one observed in Vinke et al. (2018), could have an adverse effect in the generalizability of the model to the RS cohort, particularly due to the observed differences in the mean age of the participants in the reference group and the groups of incident-AD and prevalent-AD. In spite of these differences, we observed that the normalized histograms of disease stages in the different diagnostic classes were visually largely similar for ADNI and RS. An important difference between the two cohorts was that the model estimated a low disease stage for a substantial proportion of incident-AD subjects in RS, but not in ADNI. Complementing the qualitative analysis, we also observed that the disease stages obtained using co-init DEBM could distinguish prevalent-AD subjects from CN and non-AD subjects almost equally well in both ADNI and



(a)



(b)

Fig. 4. Average disease stage trajectories of participants within the prediction set of RS. The trajectories are shown separately for CN, non-AD and incident-AD subjects within the *APOE* $\epsilon 4$ non-carriers group (a) and the *APOE* $\epsilon 4$ carriers group (b). 95% confidence intervals are shown as shaded regions around the trajectories.

RS cohorts. However, we noticed a lower performance in distinguishing incident-AD from CN and non-AD subjects in RS as compared to ADNI. Three possible explanations for these differences between ADNI and RS are given below.

First, the incident-AD group in ADNI only consisted of prodromal AD subjects with the mean time to AD diagnosis of 1.4 years, whereas the incident-AD group in RS consisted of prodromal and preclinical AD subjects with the mean time to AD diagnosis of 2.8 years. We observed in Experiment 1 that the obtained disease stages of incident-AD subjects correlated with time to AD diagnosis for *APOE* $\epsilon 4$ carriers, making AD harder to detect in the preclinical phase than in the prodromal phase. Hence the difference in the mean time to diagnosis in the two datasets is expected to be a factor contributing to the observed lower performance in the RS cohort.

Secondly, the prodromal AD subjects in ADNI were clinically defined amnestic MCI subjects who have a much higher *a priori* chance of developing AD symptoms than in the general population, making the prediction in the latter cohort a more difficult problem.

Thirdly, a factor contributing to the performance difference could be that ADNI excluded subjects with severe cardiovascular risk factors whereas the RS did not. Hence the probability of co-morbidity of vascular pathology was higher in the RS incident-AD subjects than in the corresponding ADNI set, which could have led to the drop in performance.

In spite of these factors, biologically, one could expect a Normal distribution of AD severity among incident-AD subjects in a population, whereas the observed distribution in Fig. 3 is not. A possible explanation for this apparent anomaly is that, although the biological progression of AD is heterogeneous with differences between subjects even within each genotype, the staging is performed on the basis of a mean disease timeline per genotype. Progression of subjects that is not along the estimated mean disease timeline is not accounted for in our approach, and the corresponding stages are usually an under-estimation of the true biological staging.

We observed that the correlation of MMSE with the obtained disease stages in RS was substantially lower than that in ADNI. One of the reasons for this lower correlation in RS could be that MMSE is a measure of general cognition, not specific to AD and there could be numerous other

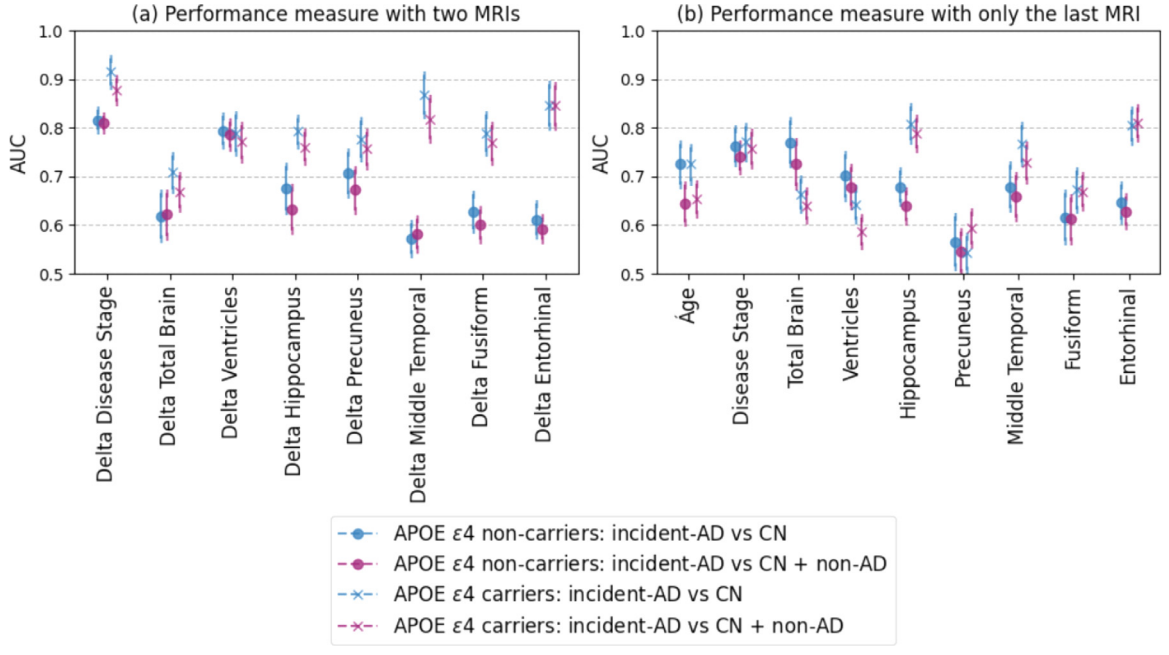


Fig. 5. Predicting incident-AD subjects in the RS cohort. Figure (a) shows the AUCs for distinguishing incident-AD while using data from two MRI scans based on longitudinal follow-up of the participants. Figure (b) shows the AUCs for distinguishing incident-AD using only the last MRI scan available for each participant.

factors affecting its value in a population-based cohort. Furthermore, in AD the MMSE range is expected to be much broader than in the general population. This in combination with less prevalent-AD cases within RS compared to that of ADNI, could explain the lower correlation as well. Moreover, the correlation of MMSE with the obtained disease stage was similar for non-carriers and carriers as measured along their respective disease timelines.

Given the high AUCs for all other classification tasks, the comparable disease stage histograms in ADNI and RS, and the possible explanations given above for the specific differences related to incident-AD prediction, we conclude that the *APOE*-specific disease timelines obtained by co-init DEBM are generalizable from a case-controlled to a population-based setting. Moreover, in our experiment, we observed that co-init DEBM was more generalizable to RS cohort than DEBM for *APOE* $\epsilon 4$ non-carriers, and equally generalizable for $\epsilon 4$ carriers. This could be because ADNI cohort is more enriched for $\epsilon 4$ carriers than the population-based RS cohort and not stratifying based on *APOE* skewed the estimated timeline more towards $\epsilon 4$ carriers.

However, for precise classification of subjects into either diagnostic category, a cut-off point for disease stage needs to be defined. We expect the cut-off point to be different in a case-controlled setting versus a population-based setting. Estimating this cut-off point in a population should ideally be estimated using an independent validation set taking several factors into consideration such as the *a-priori* prevalence of AD in the cohort, and the risks associated with false positives and negatives of this classification.

4.2. Predicting pre-symptomatic AD in the RS cohort

We observed that a participant's age distinguished incident-AD well from CN. This is in line with earlier studies that identified age as an important predictor (Park et al., 2019; Stephan et al., 2015). However, we also observed that the predictive performance of age deteriorated when the reference group was less healthy, *i.e.*, when distinguishing incident-AD from a combined reference group also consisting of subjects with subjective or objective cognitive decline unrelated to AD. This is in line with the expectation that age is poor in distinguishing cognitive decline due to AD and cognitive decline due to other causes.

The predictive performance of the volumetric biomarkers from a single MRI scan depended on the *APOE* $\epsilon 4$ carriership. We observed that hippocampus and entorhinal cortex were good predictors in *APOE* $\epsilon 4$ carriers. Interestingly, those biomarkers were estimated to be early in the corresponding disease timeline. Similarly, total brain volume and ventricle volume were good predictors in *APOE* $\epsilon 4$ non-carriers which were also the early biomarkers in its disease timeline. These results suggest that for predicting pre-symptomatic AD, early biomarkers play an important role and that it is important to understand the genotype-specific differences. However, it must be noted that in this study, the clinical diagnosis of AD was not confirmed further with the participant's amyloid-β status. Hence part of the differences observed in the disease timelines of *APOE* $\epsilon 4$ non-carriers and carriers could be attributed to the presence of greater heterogeneity in the non-carriers with respect to participant's pathologic diagnosis.

Lastly, we assessed the marginal utility of longitudinal MRI scans in identifying individuals at-risk of developing AD symptoms. We observed that participants with incident-AD showed a significant increase ($p < 0.005$) in disease stage over time as compared to CN and non-AD participants, in both *APOE* $\epsilon 4$ non-carriers and carriers. The rate of change of disease stage distinguished incident-AD subjects better than the disease stage at only the last scan, clearly highlighting the added value of longitudinal MRI scans, particularly in pre-symptomatic subjects. The rate of change of disease stage was also a better predictor of incident-AD than any other volumetric biomarker used in this study. This showed that the progression along the *APOE*-specific disease timeline can be used to identify subjects in a population at-risk of developing AD.

In this study, we only used imaging biomarkers because cerebrospinal fluid biomarkers in a pre-clinical setting are usually not available. Recent breakthroughs in blood-based biomarkers (Palmqvist et al., 2020) could help in obtaining fluid biomarkers in the pre-clinical phase of the disease. Previous work on DEBM (Venkatraghavan et al., 2019) and co-init DEBM (Venkatraghavan et al., 2021) had shown that the model is capable of incorporating biomarkers from multiple modalities for constructing the disease timelines. We expect that our current approach of predicting pre-symptomatic AD in the general population would be applicable also in the presence of fluid biomarkers, should they become available in the future.

5. Conclusion and future work

We conclude that data-driven disease timelines estimated by co-init DEBM are generalizable to population-based cohorts and that progression of individuals along such timelines is predictive of incident AD. Although the current study only considered volumetric biomarkers as inputs, it can be extended to fluid-based biomarkers, if these would become available in a population based study. Due to its robustness and explainability, we expect that our model can help identify at-risk individuals from the general population for targeted clinical trials as well as provide biomarker based objective assessment in such trials.

Data and Code Availability

Alzheimer's Disease Neuroimaging Initiative (ADNI) data and Rotterdam Study (RS) data were used in preparation of this article. ADNI data is publicly available and can be downloaded from http://adni.loni.usc.edu/. The source code for the data-driven model used in this paper is available online under the GPL 3.0 license: https://github.com/88vikram/pyebm/ RS data and code the specific experiments in the article can be obtained upon request. Requests should be directed towards the management team of the Rotterdam Study (secretariat.epi@erasmusmc.nl), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

Credit authorship contribution statement

Vikram Venkatraghavan: Conceptualization, Methodology, Formal analysis, Writing - original draft. **Elisabeth J. Vinke:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Esther E. Bron:** Writing - review & editing. **Wiro J. Niessen:** Writing - review & editing. **M. Arfan Ikram:** Writing - review & editing. **Stefan Klein:** Conceptualization, Writing - review & editing, Supervision. **Meike W. Vernooij:** Conceptualization, Writing - review & editing, Supervision.

Acknowledgement

This study is part of the EuroPOND initiative, which is funded by the European Union's Horizon 2020 research and innovation programme

under grant agreement No. 666992. This study is also a part of TKI-LSH Health Holland Alzheimer Nederland project (No. LSHM18049). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (project: ORACLE, grant agreement No. 678543). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. E.E. Bron acknowledges support from Dutch Heart Foundation (PPP Allowance, 2018B011). E.E. Bron and W.J. Niessen are supported by Medical Delta Diagnostics 3.0: Dementia and Stroke.

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

References

Archetti, D., Ingala, S., Venkatraghavan, V., Wottschel, V., Young, A.L., Bellio, M., Bron, E.E., Klein, S., Barkhof, F., Alexander, D.C., Oxtoby, N.P., Frisoni, G.B., Redolfi, A., 2019. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. Neuroimage: Clinical 24, 101954.

de Bruijn, R.F.A.G., Akoudad, S., Cremers, L.G.M., Hofman, A., Niessen, W.J., van der Lugt, A., Koudstaal, P.J., Vernooij, M.W., Ikram, M.A., 2014. Determinants, MRI correlates, and prognosis of mild cognitive impairment: the rotterdam study. J. Alzheimers Dis. 42 (Suppl 3), S239–S249.

Devi, G., Scheltens, P., 2018. Heterogeneity of Alzheimer's disease: consequence for drug trials? Alzheimer's Res. Therapy 10, 122.

Donohue, M.C., Jacqmin-Gadda, H., Goff, M.L., Thomas, R.G., Raman, R., Gamst, A.C., Beckett, L.A., Jack, C.R., Weiner, M.W., Dartigues, J.F., Aisen, P.S., 2014. Estimating long-term multivariate progression from short-term data. Alzheimer's Dementia 10, S400–S410.

Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 6, 734–746.

Ferreira, D., Nordberg, A., Westman, E., 2020. Biological subtypes of Alzheimer disease. Neurology 94, 436–448.

Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C., Alexander, D.C., 2012. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. Neuroimage 60, 1880–1889.

Holmes, C., 2002. Genotype and phenotype in Alzheimer's disease. Br. J. Psychiatry 180, 131–134.

Holzinger, A., Biemann, C., Pattichis, C. S., Kell, D. B., 2017. What do we need to build explainable AI systems for the medical domain?arXiv:1712.09923

Ikram, M.A., Brusselle, G., Ghanbari, M., Goedegebure, A., Ikram, M.K., Kavousi, M., Kieboom, B.C.T., Klaver, C.C.W., de Knegt, R.J., Luik, A.I., Nijsten, T.E.C., Peeters, R.P., van Rooij, F.J.A., Stricker, B.H., Uitterlinden, A.G., Vernooij, M.W., Voortman, T., 2020. Objectives, design and main findings until 2020 from the rotterdam study. Eur. J. Epidemiol. 35, 483517.

Ikram, M.A., van der Lugt, A., Niessen, W.J., Koudstaal, P.J., Krestin, G.P., Hofman, A., Bos, D., Vernooij, M.W., 2015. The Rotterdam Scan Study: design update 2016 and main findings. Eur. J. Epidemiol. 30, 1299–1315.

Jack Jr, C.R., Barnes, J., Bernstein, M.A., Borowski, B.J., Brewer, J., Clegg, S., Dale, A.M., Carmichael, O., Ching, C., DeCarli, C., Desikan, R.S., Fennema-Notestine, C., Fjell, A.M., Fletcher, E., Fox, N.C., Gunter, J., Gutman, B.A., Holland, D., Hua, X., Insel, P., Kantarci, K., Killiany, R.J., Krueger, G., Leung, K.K., Mackin, S., Maillard, P., Malone, I.B., Mattsson, N., McEvoy, L., Modat, M., Mueller, S., Nosheny, R., Ourselin, S., Schuff, N., Senjem, M.L., Simonson, A., Thompson, P.M., Rettmann, D., Vemuri, P., Walhovd, K., Zhao, Y., Zuk, S., Weiner, M., 2015. Magnetic resonance imaging in Alzheimer's disease neuroimaging initiative 2. Alzheimer's Dementia 11, 740–756.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Kloeppel, S., Peter, J., Ludl, A., Pilatus, A., Maier, S., Mader, I., Heimbach, B., Frings, L., Egger, K., Dukart, J., Schroeter, M., Perneczky, R., Haussermann, P., Vach, W., Urbach, H., Teipel, S., Huell, M., Abdulkadir, A., 2015. Applying automated mr-based diagnostic methods to the memory clinic - a prospective study. J. Alzheimers Dis. 47, 939–954.

Koval, I., Schiratti, J.B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., Durrleman, S., 2018. Spatiotemporal propagation of the cortical atrophy: population and individual patterns. Front Neurol 9, 235.

Lamballais, S., Vinke, E.J., Vernooij, M.W., Ikram, M.A., Muetzel, R.L., 2020. Cortical gyrification in relation to age and cognition in older adults. Neuroimage 212, 116637.

Marinescu, R., Eshaghi, A., Alexander, D., Golland, P., 2019. Brainpainter: a software for the visualisation of brain structures, biomarkers and associated pathological processes. arXiv preprint arXiv:1905.08627

Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Eshaghi, A., Toni, T., Salaterski, M., Lunina, V., Ansart, M., Durrleman, S., Lu, P., Iddi, S., Li, D., Thompson, W. K., Donohue, M. C., Nahon, A., Levy, Y., Halbersberg, D., Cohen, M., Liao, H., Li, T., Yu, K., Zhu, H., Tamez-Pena, J. G., Ismail, A., Wood, T., Bravo, H. C., Nguyen, M., Sun, N., Feng, J., Yeo, B. T. T., Chen, G., Qi, K., Chen, S., Qiu, D., Buciuman, I., Kelner, A., Pop, R., Rimocea, D., Ghazi, M. M., Nielsen, M., Ourselin, S., Sorensen, L., Venkatraghavan, V., Liu, K., Rabe, C., Manser, P., Hill, S. M., Howlett, J., Huang, Z., Kiddle, S., Mukherjee, S., Rouanet, A., Taschler, B., Tom, B. D. M., White, S. R., Faux, N., Sedai, S., de Velasco Oriol, J., Clemente, E. E. V., Estrada, K., Aksman, L., Altmann, A., Stonnington, C. M., Wang, Y., Wu, J., Devadas, V., Fourrier, C., Raket, L. L., Sotiras, A., Erus, G., Doshi, J., Davatzikos, C., Vogel, J., Doyle, A., Tam, A., Diaz-Papkovich, A., Jammeh, E., Koval, I., Moore, P., Lyons, T. J., Gallacher, J., Tohka, J., Ciszek, R., Jedynak, B., Pandya, K., Bilgel, M., Engels, W., Cole, J., Golland, P., Klein, S., Alexander, D. C., 2020. The Alzheimer's disease prediction of longitudinal evolution (TADPOLE) challenge: results after 1 year follow-up. arXiv:2002.03419

Palmqvist, S., Janelidze, S., Quiroz, Y.T., Zetterberg, H., Lopera, F., Stomrud, E., Su, Y., Chen, Y., Serrano, G.E., Leuzy, A., Mattsson-Carlgren, N., Strandberg, O., Smith, R., Villegas, A., Sepulveda-Falla, D., Chai, X., Proctor, N.K., Beach, T.G., Blennow, K., Dage, J.L., Reiman, E.M., Hansson, O., 2020. Discriminative accuracy of plasma phospho-tau217 for alzheimer disease vs other neurodegenerative disorders. JAMA 324, 772–781.

Park, K., Sung, J., Kim, W., An, S., Namkoong, K., Lee, E., Chang, H., 2019. Population-based dementia prediction model using korean public health examination data: a cohort study. PLoS ONE 14.

Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Weiner, M.W., 2010. Alzheimer's disease neuroimaging initiative (ADNI). Neurology 74, 201–209.

Ryan, J., Fransquet, P., Wrigglesworth, J., Lacaze, P., 2018. Phenotypic heterogeneity in dementia: a challenge for epidemiology and biomarker studies. Front. Public Health 6, 181.

Sevigny, J., Chiao, P., Bussière, T., Weinreb, P.H., Williams, L., Maier, M., Dunstan, R., Salloway, S., Chen, T., Ling, Y., O'Gorman, J., Qian, F., Arastu, M., Li, M., Chollate, S., Brennan, M.S., Quintero-Monzon, O., Scannevin, R.H., Arnold, H.M., Engber, T., Rhodes, K., Ferrero, J., Hang, Y., Mikulskis, A., Grimm, J., Hock, C., Nitsch, R.M., Sandrock, A., 2016. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature 537, 50–56.

Sperling, R.A., Karlawish, J., Johnson, K.A., 2013. Preclinical Alzheimer disease-the challenges ahead. Nat. Rev. Neurol. 9, 54–58.

Stephan, B.C.M., Tzourio, C., Auriacombe, S., Amieva, H., Dufouil, C., Alpérovitch, A., Kurth, T., 2015. Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. BMJ 350.

Van Cauwenberghe, C., Van Broeckhoven, C., Sleegers, K., 2016. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet. Med. 18, 421–430.

Venkatraghavan, V., Bron, E.E., Niessen, W.J., Klein, S., 2019. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. Neuroimage 186, 518–532.

Venkatraghavan, V., Klein, S., Fani, L., Ham, L.S., Vrooman, H., Ikram, M.K., Niessen, W.J., Bron, E.E., 2021. Analyzing the effect of APOE on Alzheimer's disease progression using an event-based model for stratified populations. Neuroimage 227, 117646.

Vinke, E.J., de Groot, M., Venkatraghavan, V., Klein, S., Niessen, W.J., Ikram, M.A., Vernooij, M.W., 2018. Trajectories of imaging markers in brain aging: the rotterdam study. Neurobiol. Aging 71, 32–40.

Wachter, S., Mittelstadt, B., Russell, C., 2018. Counterfactual explanations without opening the black box: automated decisions and the gdpr. arXiv:1711.00399

Weintraub, S., Teylan, M., Rader, B., Chan, K.C., Bollenbeck, M., Kukull, W.A., Coventry, C., Rogalski, E., Bigio, E., Mesulam, M.M., 2019. APOE is a correlate of phenotypic heterogeneity in Alzheimer disease in a national cohort. Neurology.

World Health Organization, 2017. Global action plan on the public health response to dementia 2017–2025.

Young, A.L., Marinescu, R.V.V., Oxtoby, N.P., Bocchetta, M., Yong, K., Firth, N., Cash, D.M., Thomas, D.L., Dick, K.M., Cardoso, J., van Swieten, J., Borroni, B., Galimberti, D., Masellis, M., Tartaglia, M.C., Rowe, J.B., Graff, C., Tagliavini, F., Frisoni, G., L., J.R., Finger, E., Medonça, A., Sorbi, S., Warren, J.D., Crutch, S., Fox, N.C., Ourselin, S., Schott, J.M., Rohrer, J.D., Alexander, D.C., GENFI, ADNI, 2018. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. Nat. Commun. 9.