Instantiated mixed effects modeling of Alzheimer's disease markers.

R. Guerrero[a,1,*], A. Schmidt-Richberg[a], C. Ledig[a], T. Tong[a], R. Wolz[b,a], D. Rueckert[a], The Alzheimer's Disease Neuroimaging Initiative (ADNI)[2]

[a]*Department of Computing, Imperial College London, UK*
[b]*IXICO plc., UK*

Abstract

The assessment and prediction of a subject's current and future risk of developing neurodegenerative diseases like Alzheimer's disease is of great interest in both the design of clinical trials as well as in clinical decision making. Exploring the longitudinal trajectory of markers related to neurodegeneration is an important task when selecting subjects for treatment in trials and the clinic, in the evaluation of early disease indicators and the monitoring of disease progression. Given that there is substantial intersubject variability, models that attempt to describe marker trajectories for a whole population will likely lack specificity for the representation of individual patients. Therefore, we argue here that individualized models provide a more accurate alternative that can be used for tasks such as population stratification and a subject-specific prognosis. In the work presented here, mixed effects modeling is used to derive a global and individual marker trajectories for a training population. Test subject (new patient) specific models are then instantiated using a stratified "marker signature" that defines a subpopulation of similar

*Corresponding author: Ricardo Guerrero
Email address: reg09@imperial.ac.uk (R. Guerrero)

[1]This project was partially funded by the Innovate UK (formerly Technology Strategy Board - TSB).

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

cases within the training database. From this subpopulation, personalized models of the expected trajectory of several markers are subsequently estimated for unseen patients. These patient specific models of markers are shown to provide better predictions of time-to-conversion to Alzheimer's disease than population based models.

Keywords: Longitudinal modeling, Alzheimer's disease, AD markers, subject stratification.

---

1. Introduction

Alzheimer's disease had a world wide prevalence of around 26.6 million reported cases by 2006, with predictions suggesting an increase to about 100 million by 2050 [6]. Preliminary evidence suggests that intervention to slow the progression of the disease is likely to be most effective in early stages of the disease [44]. Recent clinical trials successfully employ enrichment strategies with multiple markers (e.g. [43]) and a recent post-hoc analysis of the Roche Gantenerumab study showed a treatment effect in a subpopulation that was identified through a longitudinal mixed effects model [27]. Advances in early diagnosis (and treatment) have the potential to make a huge impact on the overall well-being of the population, and reduce the burden to caregivers, family members and health systems.

Current research into modeling Alzheimer's disease markers has led to an increased understanding of the underlying pathophysiology of the disease. Recently, the view on AD diagnosis has shifted from seeing subjects as being healthy or diseased, to a more dynamic process in which clinical and pathological markers gradually begin to change before current diagnosis criteria are met and continue to do so over time. Exploring the longitudinal trajectory of Alzheimer's disease related markers is an important task in the evaluation of early indicators of the disease, patient selection for trials, monitoring disease progression, and validating proposed progression models. Detailed studies into early state longitudinal Alzheimer's disease marker trajectory dynamics, using data-driven methods, have the potential to aid the effort in the development of measures that can accurately and robustly quantify indications of the disease, even before its presymtomatic and preclinal stages. Previously, hypothetical [23, 24] and experimental models [13, 14, 50, 10, 20, 7, 25, 41, 21, 4, 5, 1, 12, 15, 53] of disease progression based on Alzheimer's disease markers, such as cerebrospinal fluid (CSF), imaging

and cognitive markers have been proposed.

Estimating the current progress of patients along the disease trajectory has been the focus of many studies in this field. Recently, several approaches that regard the disease progression trajectory as a continuous process have been developed [13, 14, 50, 10, 20, 7, 25, 41, 21, 4, 5, 1, 12]. Many of the proposed frameworks rely on modeling cognitive scores such as the mini mental state examination (MMSE), Alzheimer's disease assessment scale (ADAS), clinical dementia rating sum of boxes (CDRSB), or Rey's audio visual learning test (RAVLT) among others, as surrogate measures of disease progression. Doody et al. [13] proposed the use of MMSE scores to compute an initial preprogression rate (rate of initial decline prior to first physician visit) used to classify subjects as slow, intermediate or rapid progressors. Using this estimated preprogression rate, Doody et al. [14] tested the predicted survival rate of subjects and their performance on cognition, function and behavior over time. Yang et al. [50] proposed an exponential model of ADAS scores and used this to estimate the disease duration and current pathological stage of a patient. Similarly, Delor et al. [10] computed a disease onset time by adjusting subjects according to CDRSB score. Ito et al. [20] developed a model to describe the longitudinal response in ADAS score, where the rate of progression was found to increase with baseline severity, age and/or ApoE-ϵ4 genotype status.

As noted in the theoretical model proposed by Jack et al. [23, 24], cognitive markers are especially relevant for the advanced stages of the disease, but less sensitive at earlier stages. In this regard, modeling frameworks that employ different types of Alzheimer's disease markers have also been put forward. Caroli et al. [7] tested the hypothetical model of marker dynamics proposed by Jack et al. [24] on real data. Using CSF Aβ 142, tau, hippocampal volume, and fludeoxyglucose-PET markers, it was shown that when individual values were Z-transformed and plotted against ADAS scores, a sigmoid model could be used to describe the population time course of markers. Jedynak et al. [25] proposed a statistical methodology using multiple Alzheimer's disease markers, which produces an Alzheimer's disease progression score for each patient and each time point in the database. Schmidt-Richberg et al. [41] proposed an approach to estimate typical trajectories of several markers over the full course of the disease to estimate current disease progress and disease progression rate.

Mixed effect modeling offers an appealing alternative to model generation, as it permits the incorporation of unseen subject specific random effects. To

this end, Ito and Hutmacher [21] proposed the use of a longitudinal non-linear mixed effects model to predict the CDRSB time course, which is used to estimate the median time for clinically worsening (defined by several thresholds in change of CDRSB) and to determine clinical endpoints to use in therapeutic trials. Bernal-Rusiel et al. [4, 5] proposed spatio-temporal linear mixed effects modeling for mass-univariate analysis of longitudinal neuroimaging data. Adak et al. [1] calculated rates of cognitive decline using hierarchical mixed effects models in which each subject is assumed to have a linear trajectory of cognitive decline. Donohue et al. [12] used self-modeling regression to simultaneously estimate pathological timing and long-term growth curves. Recently, Schiratti et al. [22] proposed a generative statistical (nonlinear mixed-effects) model for longitudinal data, described in a univariate Riemannian manifold setting, to estimate an average disease progression model, subject-specific time shifts and acceleration factors.

Modeling disease progression not as a continuous process but as discrete stages has also been suggested by some authors. Fonteijn et al. [15] proposed the use of an event-based model to determine the order in which markers become abnormal in familiar Alzheimer's and Huntington's disease, to subsequently assign a subject to one of several discrete disease stages. The event-based model was later reformulated and extended by Young et al. [53] for the use in sporadic Alzheimer's disease. However, discretizing the disease progression trajectory is a strong assumption that most likely leads to an over-simplification of the problem and does not allow a time to event prediction. To this end, Jack and Holtzman [23] argue that accurate time-dependent models of Alzheimer's disease markers must be incorporated into diagnosis criteria.

Some of the drawbacks of the mentioned approaches are their reliance on modeling whole populations without considering data grouping and/or that they offer no trivial way to make predictions about the future disease trajectory of unseen, individual subjects. Given that there is substantial intersubject variability in populations, a model that sets to describe a whole population could provide a sub-optimal representation. Modeling the behavior of a subject's marker using an instantiated subpopulation of similar cases might produce a more realistic patient-specific trajectory model. The approach proposed here draws inspiration from and builds upon the methods previously discussed. However, it differs from previous work in that patient specific models are instantiated using a relevant "marker signature" that defines a subpopulation of similar cases in a training database. This

allows to estimate patient specific models (based on an instantiated subpopulation) for each of the markers calculated. The proposed concept of the "marker signature" defines a feature space in which similarities between subjects' marker trajectories can be calculated. The "marker signature" differs for each marker and for each parameter in the parametric modeling function used. The motivation behind the proposed approach stems from noting that in many areas (such as clinical trials) patient stratification offers increased observational power to studies. A description of the validation data used and details into the framework proposed here are given in Section 2, experimental results are provided in Section 3, and finally Section 4 offers a discussion on the implications, strengths and weaknesses of the proposed framework.

2. Materials and methods

2.1. Data

Data used in this work was obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database. ADNI is a large-scale multi-site study that aims at analysing markers from cognitive tests, blood tests, tests of CSF, and MRI/PET imaging with regard to their ability to characterize the progression of Alzheimer's disease. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. Subjects have been recruited from over 50 sites across the U.S. and Canada. To date, ADNI in its three studies (ADNI-1, -GO and -2) has recruited over 1500 adults, aged between 55 and 90 years, to participate in the study. Here, the subset of subjects enrolled in either ADNI-1, ADNI-GO or ADNI-2, that have at least four different time point samples and convert to Alzheimer's disease or MCI (mild cognitive impairment) during the study are considered (as of 19/02/2015). A different number of samples from each subject and markers are available, e.g. not all available time points contain all cognitive or imaging information (features). 78 subjects that reverted to a less severe disease label during the study were excluded. In total, there were 1153 subjects with at least four time points (6407 individual entries). From these subjects, 216 (1309 samples) converted to Alzheimer's disease, including 10 subjects that started as CN for which time-to-conversion to AD was used for alignment (points prior to MCI conversion were also aligned according to AD conversion), 39 (252 samples) converted from CN (cognitive normal) to MCI (excluding subjects that start as CN and convert to Alzheimer's disease) and had available all markers at all time points. Henceforth, these

groups of subjects will be called MCI-AD and CN-MCI converters, respectively. Evaluations were carried out for the 216 MCI-AD converters, using models trained either with only the MCI-AD converter group or with a larger set that combined MCI-AD and CN-MCI groups into a temporally aligned CN-MCI-AD converter set. This set had an average follow up time of $\sim$45 months, with a maximum and minimum of 96 and 18 months, respectively. Markers considered in this work can be divided into two categories: cognitive (or functional), and structural imaging, with the latter further subdivided into volumetric, manifold learning and grading. A summary of the available samples for each markers is given in Table 1 while the available range for each marker/population is given in Table 2. Further details about the markers considered here are specified as follows.

Cognitive markers. ADNI participating subjects were asked to perform a battery of cognitive tests at each visit during the study. The direct total score of each of these tests was used here as a marker. Cognitive tests included in this work comprise MMSE, ADAS11, ADAS13, functional activities questionnaire (FAQ), CDRSB and RAVLT. As noted before, not all results were available at every visit due to unspecified reasons. Thus, the number of available subjects for mixed effects modeling of a specific marker varied from 48-49 (with 332-339 samples) for the CN-MCI conversion group and from 264-269 (with 1697-1740 samples) for the MCI-AD group.

Volumetric. Magnetic resonance (MR) images were segmented into anatomical regions using multi-atlas label propagation with expectation-maximization (MALPEM) [29]. Here, 30 atlases[3] segmented into 134 anatomical regions were transformed to an unsegmented image space using non-rigid registration [38, 19]. Individual atlas label maps were then transformed to the unsegmented image space using the calculated transformations and a nearest neighbor interpolation scheme. The propagated atlas label maps were then fused into a consensus probabilistic segmentation using a local weighting approach. Further refinement using a method that exploits image intensities in an expectation-maximization framework was carried out. From the 134 anatomical structures, all cortical volumes were combined into left and right cortex, while ventricular volumes were combined into a single value. Follow-

[3]Brain atlases corresponding to manual expert segmentations provided by Neuromorphometrics, Inc. (http://Neuromorphometrics.com)

ing suggestions made in the literature [25, 53, 12, 1], a subset of structures consisting of the ventricles, left and right cortex, amygdala, and hippocampus (seven in total), were selected as markers for further analysis. The total number of training subjects available was 41 (with 264 samples) for the CN-MCI model and 224 (with 1379 samples) for the MCI-AD model.

Manifold learning. Manifold learning features from MR images have been shown to contain valuable information about disease severity and progression [18, 16, 17, 49]. Here, the aim is to learn the underlying low-dimensional manifold that best represents the population. This is achieved in three steps: First, a region of interest (ROI) relevant to disease progression is learned using sparse regression [16]. Second, local binary patterns (LBP) are extracted from the ROI to transform MR intensities to a higher-dimensional binary space. Finally, Laplacian Eigenmaps [3] is used to learn a low-dimensional manifold, where the k-neighborhood graph is constrained to allow only one instance per subject into each neighborhood [17]. The manifold is chosen to have two dimensions (D1 and D2). Data points in this two-dimensional manifold follow an approximately quadratic distribution. With this in mind, a third feature P_D1D2 was obtained via a perpendicular projection of the points in $\mathbb{R}^2$ to a fitted quadratic curve. Manifold features were computed for 1063 subjects (5679 images) with at least four visits, out which 41 (264 images) and 224 (1379 images) were CN-MCI and MCI-AD converters, respectively.

Grading. The goal of grading features [9, 47, 46] is the scoring of a test subject by estimating its similarity to different training subjects. There are two main steps in their calculation: ROI learning using sparse regression as described in [18], and disease label propagation from training to test subjects. The relation between training population and test subject is modeled using elastic-net regression. Given the ROI intensities of a test subject $X_t \in R^{k\times 1}$ and n training subjects $\mathbf{X}_\tau \in R^{k\times n}$, the grading score g_t of the test subject can be calculated by minimizing the following cost function [46]:

$$\begin{cases} \hat{\alpha} & = \min\limits_{\alpha} \frac{1}{2} \left\| X_t - \mathbf{X}_\tau \alpha \right\|_2^2 + \lambda_a \|\alpha\|_1 + \frac{1}{2}\lambda_b \|\alpha\|_2^2 \\ g_t & = \sum_{j=1}^{n} \hat{\alpha}_t(j) l_j \Big/ \sum_{j=1}^{n} \hat{\alpha}_t(j) \end{cases} \tag{1}$$

Here, $\hat{\alpha}$ are coding coefficients for the test subject and l_j is the disease label for the jth training subject. CN and Alzheimer's disease subjects were used

Population	Marker	N	S	bl. age	Follow-up
CN-MCI	MMSE	49	339	76.35 (5.17)	61.95 (28.07; 24/96)
	ADAS11	”	”	”	”
	ADAS13	”	338	”	”
	CDRSB	48	332	76.19 (5.09)	62.50 (28.11; 24/96)
	FAQ	”	334	”	”
	RAVLT	49	337	76.35 (5.17)	61.95 (28.07; 24/96)
	Volumetric	41	264	76.03 (5.05)	58.24 (27.44; 24/96)
	Manifold	”	”	”	”
	Grading	”	”	”	”
	ALL	39	252	76.03 (4.83)	59.69 (27.32; 24/96)
MCI-AD	MMSE	268	1737	73.93 (7.12)	47.66 (24.45; 18/108)
	ADAS11	266	1723	73.94 (7.14)	47.75 (24.59; 18/108)
	ADAS13	264	1697	73.91 (7.15)	47.48 (24.57; 18/108)
	CDRSB	268	1736	74.02 (7.11)	47.71 (24.48; 18/108)
	FAQ	269	1740	74.06 (7.06)	47.87 (24.40; 18/108)
	RAVLT	261	1673	73.89 (7.06)	46.83 (24.14; 18/108)
	Volumetric	224	1379	73.73 (7.27)	44.87 (23.16; 18/96)
	Manifold	”	”	”	”
	Grading	”	”	”	”
	ALL	216	1309	73.88 (7.13)	44.25 (23.03; 18/96)

Table 1: Summary of all available markers, where **N** represents the number of available subjects, **S** the number of samples, **bl. age** is the mean (± standard deviation) age of subjects at baseline in years and **Follow-up** is the mean (± standard deviation; maximum / minimum) follow-up time in months. The row named ALL refers to the subset of subject and samples for which all markers were available. Volumetric, manifold and grading markers are grouped as they all have the same amount of subjects and samples available.

for training, where their corresponding l_j was set to 1 or -1, respectively. Disease label propagation was constrained to allow only one time point per subject. Stability selection re-sampling [18] was employed to reduce sampling bias and increase robustness. Images were registered to MNI152 space using free form deformations [38] with different control point spacings (affine, 20mm, 10mm, 5mm and 2.5mm) and grading scores for each of these were used as features for the subsequent analysis. The number of subjects available was the same as for the manifold features.

2.2. Methods

Given a set of markers, extracted from multiple subjects at several time points, the aim is to derive subject specific marker models that in turn can

Marker	CN-MCI					MCI-AD				
	Max	95%	50%	5%	Min	Max	95%	50%	5%	Min
MMSE	30	30	29	26	24	30	29.65	25	16	0
ADAS11	24.67	14	7	2.67	1	70	32.12	14.67	6.33	0
ADAS13	33	21	11	4.33	2	85	45	24	11.33	0
CDRSB	4.5	2	0	0	0	18	10	3	0.5	0
FAQ	14	6	0	0	0	30	26	9	0	0
RAVLT	65	56	38	24	12	71	40	25	11	0
P_D1D2	0.65	0.21	-0.09	-0.29	-0.38	0.93	0.61	0.04	-0.16	-0.36
D1	0.24	0.16	-0.05	-0.16	-0.22	0.44	0.29	0.08	-0.09	-0.21
D2	0.48	0.15	-0.03	-0.16	-0.18	0.74	0.42	-0.06	-0.18	-0.27
Hip. Left	4034	3711	3019	2350	2037	4133	3469	2621	1881	1365
Hip. Right	4213	3992	3246	2533	2095	4478	3688	2778	2008	644
Amy. Left	1540	1403	1042	728	303	1517	1242	976	560	302
Amy. Right	1635	1423	1093	649	536	1748	1300	947	589	244
Cor. Left	311140	296430	252290	215410	199730	332650	298130	250570	202800	145770
Cor. Right	306820	294080	255710	215130	184400	328830	298060	249520	203820	157770
Ventricles	122030	89020	51320	22970	12680	188120	116260	57000	28480	13380
G 2.5mm	0.95	0.84	0.19	-0.36	-0.87	0.99	0.61	-0.36	-0.83	-0.99
G 5mm	0.97	0.88	0.35	-0.66	-0.85	0.96	0.65	-0.39	-0.87	-0.98
G 10mm	0.95	0.81	0.29	-0.65	-0.94	0.92	0.67	-0.35	-0.84	-0.98
G 20mm	0.86	0.73	0.25	-0.56	-0.74	0.97	0.67	-0.19	-0.78	-0.98
G lin	0.91	0.76	0.29	-0.34	-0.61	0.89	0.60	-0.19	-0.77	-0.96

Table 2: Marker value range and percentiles of data distribution according to population. Cognitive, manifold and grading markers are unitless values, while volumetric markers are given in mm^3.

be used to estimate a subject's current and future disease "state". Subjects considered in the proposed framework exhibit worsening in terms of their clinical label (called "converters") at some point during the data collection study (see Section 2.1 for data description). Furthermore, all subjects' markers are aligned with respect to the time at which conversion (clinical worsening) occurred as described in Section 2.2.1. In the work presented here, there are two groups of subjects: MCI-AD and CN-MCI converters. However, a general model that includes subjects from both of these groups, called CN-MCI-AD, is also explored (see Section 2.2.1). Mixed effect modeling is used to derive a training population's global and individual marker trajectories (see Section 2.2.2). The main objective here is to derive individualized models for patients not present during training. From the proposed framework, three types of models are derived: 1) the resulting population model (popMod) from the mixed effects modeling (fixed effects), 2) an instantiated model based on similar cases (nnMod), and 3) the "real" model (realMod), which is the model fitted to each subject in the mixed effects modeling framework (fixed + random effects). Models realMod and popMod are a result of the mixed effects modeling, while nnMod is an estimated model according to Section 2.2.3. Generally, given a new patient, not present during training, only the general population model popMod is available for prediction as the individualized realMod is unavailable. The aim is therefore to examine if the

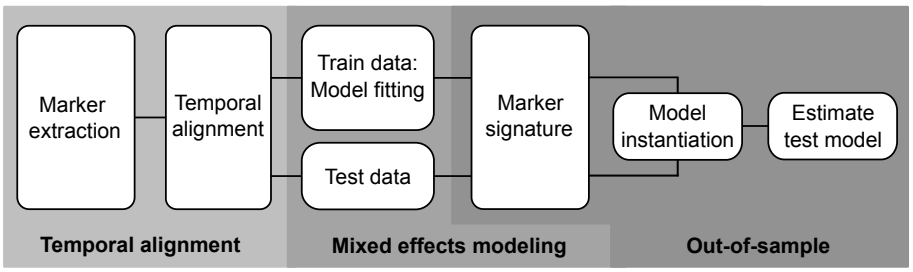

Figure 1: Overview of the proposed framework.

estimated nnMod is a better approximation to the realMod than the popMod in a validation setting, thus providing an individualized model based on a relevant subpopulation. Figure 1 shows an illustrated diagram of the three main steps of the proposed framework.

2.2.1. Temporal alignment of markers

The proposed framework relies on the temporal alignment of markers with respect to the conversion to a more severe disease label (clinical worsening). However, due to the temporal limitations of most population studies, it is often difficult to observe subjects progressing through the complete Alzheimer's disease trajectory. It is more often the case that in population studies subjects progress from CN to MCI or from MCI to Alzheimer's disease. Generally, a different model is needed to explain each of these two phases due to the lack of a consistent temporal alignment. With this in mind, an extended disease trajectory model can be learned by combining the CN-MCI and MCI-AD groups via finding a temporal group alignment variable. Cross-sectional variance of markers generally increases as markers measurements move further along the disease trajectory. Hence, group alignment can be achieved using the cross-sectional variance distribution similarity of both groups in an overlapping disease stage region. This region of overlap is assumed to occur before MCI subjects converts to Alzheimer's disease and after CN subjects have converted to MCI.

Quantile regression models the relation between a set of predictor variables and specific quantiles of the measured variable. It specifies changes in the quantiles of the measurements and can be used to infer the cross-sectional variance distribution of the measured variable. The need for quantile curves arises when measurements are strongly dependent on some covariate, e.g. position along the disease trajectory, so that the reference range changes with

the covariate [8]. Whereas standard linear regression techniques summarize the average relationship between a set of regressors and the measurements based on the conditional mean function, quantile regression considers the relationship between the regressors and measurements using the conditional median function, where the median is the 50th percentile, or quantile q, of the empirical distribution.

Assuming that the distribution of the measurements is not normal but can be normalized via the Box-Cox power transform, then the LMS [8] method (named after the three parameters of the Box-Cox power transform: λ, the mean μ and the coefficient of variation σ) can be used to obtain the quantile curves, and hence the conditional cross-sectional distributions of the measurements. Furthermore, a modified version of the LMS method, that uses the Yeo-Johnson [52] power transform as in [51] to allow the handling of non-positive marker values, can be used to induce Gaussianity in the generally non-Gaussian marker distribution data and obtain the conditional cross-sectional distributions. In the same way as in [41], the conditional cross-sectional distributions, obtained from the modified LMS method [51], for each of the groups (CN-MCI and MCI-AD) can be used for the temporal alignment of markers. The area under the curve of the conditional distributions of the markers at each time (after within-group temporal alignment) are used as goodness of fit, and its maximization determines the temporal shift between the CN-MCI and MCI-AD subject groups. Figure 2 illustrates the conditional distribution matching process used for temporal alignment, where the variance of conditional cross-sectional distributions are illustrated as the three colored lines, with the middle line representing the mean and the outer lines the distributions variance. Figure 2 (a) illustrates how the conditional distribution variance of both groups varies in relation to the time of conversion. These two groups, each aligned to its respective point of conversion (CN-MCI or MCI-AD), can be matched by finding the best matching between the cross-sectional distributions as in Figure 2 (b), to ultimately define the optimal time shift between the points of conversion as in Figure 2 (c).

2.2.2. Mixed effects modeling

Modeling Alzheimer's disease marker relations in longitudinal data remains a challenge. On the one hand, it could be assumed that all measurements are drawn independently from the same general population, however, some measurements are grouped (e.g. longitudinally by subject) and there-

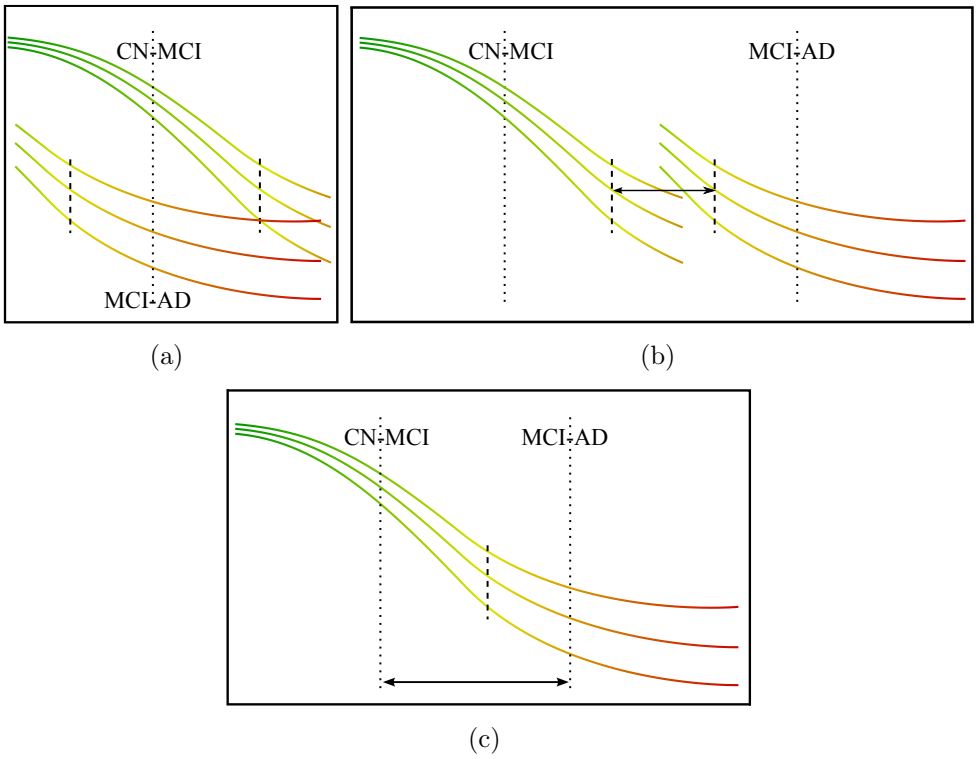

Figure 2: Variance matching and temporal group alignment process. The X- and Y-axis represent time and marker measumerment, respectively. Colored curves represent the variance distribution of the temporally aligned marker values, dotted vertical lines represent group transition to a more severe disease state and the dashed vertical lines indicate the optimal variance matching point. (a) Groups aligned with respect to transition to a more severe disease state. (b) Search for optimal variance matching point. (c) Groups temporarily aligned at optimal time shift.

fore not independent. On the other hand, individual models could be fitted to each subject, treating same subject measurements as dependent, while assuming independence from the rest of the population. This, however, ignores the fact that subjects themselves belong to a population. Mixed effects modeling attempts to reconcile these model design schemes by combining fixed and random effects, where effects are considered variables that form part of the model parameters. Fixed effects are assumed to represent those parameters that are the same for the whole population and amount to traditional regression modeling, where a single model describes the population's behavior. Random effects are group (subject) dependent random variables, which

are modeled as additional error terms and determine the deviation of individual subjects from the population's model. Generally, a mixed effects model that belongs to the jth observation of the ith individual can be written as [31]

$$y_{ij} = f(\boldsymbol{\phi}_i, \mathbf{x}_{ij}) + \epsilon_{ij}, \tag{2}$$

where y_{ij} is the jth response of the ith subject, $\mathbf{x}_{ij}$ is the predictor vector for the jth response on the ith subject, $f(\cdot)$ is a model function, $\boldsymbol{\phi}_i$ a parameter vector of length r, and ϵ_{ij} is simplified as a normally distributed noise term. The parameter vector $\boldsymbol{\phi}_i$ can be split in fixed and random effects and be written as

$$\boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \qquad \mathbf{b}_i \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{D}), \tag{3}$$

where $\boldsymbol{\beta}$ is a p-vector of parameters corresponding to the fixed effects, $\mathbf{b}_i$ is the ith subject's q-vector of random effects, $\mathbf{A}_i$ and $\mathbf{B}_i$ are $r \times p$ and $r \times q$ design matrices, and $\sigma^2\mathbf{D}$ is the covariance matrix. If design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ are identity matrices $\mathbf{I}$, all model parameters have associated random effects and the same fixed effects apply to the whole population. Here, the predictor (dependent) and the target (independent) variables x_{ij} and y_{ij}, respectively, contain the marker measurements and the time-to-conversion of the jth observation (time point) of the ith subject. Additionally, covariance matrix $\mathbf{D}$ and design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ from Equation 3 are considered identity matrices for simplicity. A more sophisticated choice of design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ can be used to add covariates to the model, such as ApoE-ϵ4 genotype. Furthermore, design matrices $\mathbf{A}_i$ and $\mathbf{B}_j$ can also be made to be dependent on both i and j to include subject- and visit-specific such as age or education. Equation 2 can represent either linear or non-linear models by changing the parametric modeling function $f(\cdot)$, which should be determined by the underling disease progression process being modeled.

Types of parametric models. As mentioned, markers considered in this work can be divided into cognitive, volumetric, manifold learning and grading. For a description of each marker type (see Section 2.1).

Theoretically, long-term physiological and psychological marker models have been hypothesized to follow a sigmoid shape [23]. Experimental evidence also suggests that linear models do not sufficiently portrait cognitive or functional decline in disease progression [40, 35, 45]. In this work, sigmoidal

curves were used to describe cognitive and functional decline, as this is a well accepted practice [45, 2, 32]. Sigmoidal models also offer the advantage of fixing the lower and upper model's asymptotes to the scale's limits of individual markers, such that model predictions do not fall outside this bounded scale. Some evidence also suggests a sigmoidal pattern, with an acceleration phase during the early stages of the disease and deceleration at later stages, for the dynamic behavior of cortical thinning and hippocampal volume loss [39, 7]. However, in [30, 12, 41] little evidence of acceleration, that would suggest a non-linear effect, in structural brain atrophy rates was observed. In this work we use a linear function to model the dynamic changes of structural MR volumetric markers. In principle, the dynamic behavior of machine learning derived markers is less well understood, and generally depends on the algorithm and parameter choices. Here, two types of markers fall into this category: manifold learning and grading features. For these, parametric model choices were made empirically, where sigmoid and quadratic models were used for manifold learning features, while sigmoid models were used for grading features.

Model selection for each marker was driven by observations made in the literature previously discussed. Other parametric model choices were considered for each , e.g. Gompertz [48] or Richards [37] in addition to those already mentioned, but were ultimately empirically rejected.

2.2.3. Model personalization

Using mixed effects modeling, individual (realMod) and population (popMod) models can be obtained for a training dataset of subjects and their respective markers. Given an unseen subject with one (or more) time points and at least one measured marker, the objective is to estimate its current "state" and predict its future behavior along the disease trajectory. This is achieved by estimating the subject's current time-to-conversion and modeling the corresponding progression path (nnMod).

Some markers might be more closely associated with each other than others. Therefore, it stands to reason that choosing a relevant set of markers, the "marker signature", to determine subject similarity, and hence define an instantiated subpopulation of similar cases for parameter estimation, can play an important role. Here, the correlation coefficient between all training subjects' realMod models parameters and all markers is used to find those markers that are best correlated with the change in specific marker model parameters. At training time, these coefficients are used to choose the top

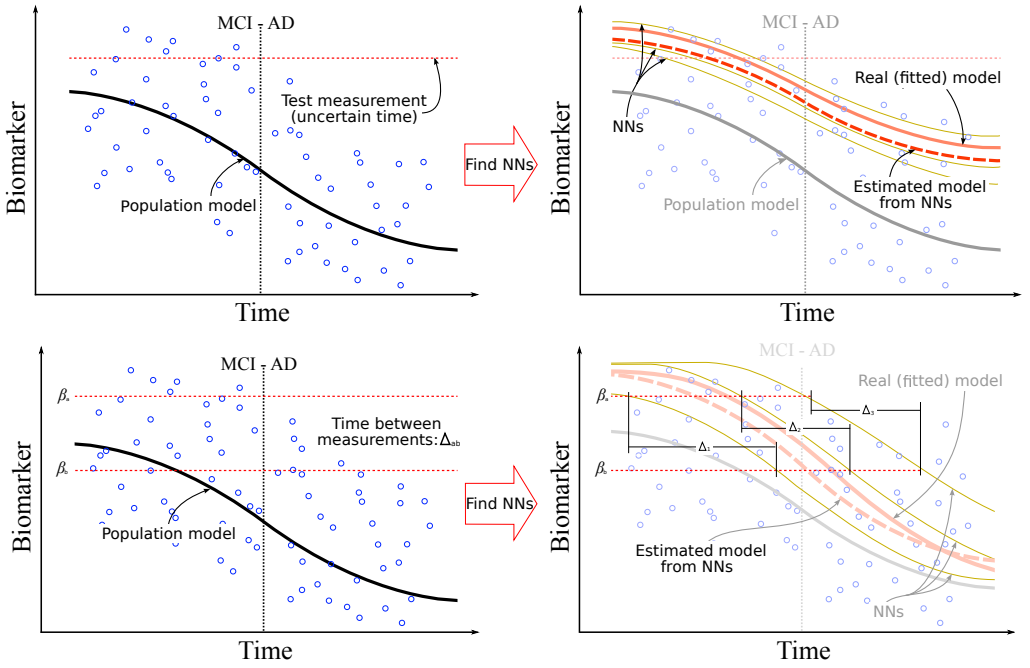

Figure 3: Model instantiation. First row: Example of model estimation from nearest neighbors based on marker signature. Second row: Example of model estimation from nearest neighbors based on marker change over time.

(available) markers to define a "marker signature" (specific to each parameter of each marker), which in turn is used to measure similarity between a test subject and the training set to find the most similar cases. Figure 3 (top row) illustrates the process of estimating a progression model for an unseen subject for one marker using its marker signature. That is, given a marker (or several) measured at a specific time in a test subject, its position along the disease trajectory and the trajectory that this subject will follow are both unknown (Figure 3 top left). The trajectories of training subjects (fitted to observed data) that are found to have a similar "marker signature" (specific to the marker in question) to the previously unseen test subject (and its set of measured markers) are combined in order to estimate a trajectory (Figure 3 top right).

In principle, the obtained "marker signature" might be a good way to approximate the current disease "state", under the assumption that if all Alzheimer's disease markers are similar, the disease state might be similar. However, this is not necessarily the case for disease progression speed esti-

mation (e.g. the slope in a linear model). If more than one time point is available, test subject instantiation can also take into account marker change between visits to find subjects that experienced similar change over a similar time period. Here, a marker β is measured at time points t_a and t_b, where the time difference between measurements is Δ_{ab}. Nearest neighbors are selected as those training models that given marker measurements β_a and β_b produce a measurement time difference Δ_i similar to Δ_{ab}. On the other hand, nearest neighbors can also be selected as those training models that, given time difference Δ_i, produce marker values similar to β_a and β_b. Figure 3 (bottom row) illustrates the process of subpopulation instantiation based on marker change. This means that given an AD marker measured at two (or more) time points in a test subject, with its current position along the disease trajectory unknown (Figure 3 bottom left), the information given by marker change and time difference between time points can be leveraged to find trajectories that behave in a similar fashion and combined in order to estimate a trajectory (Figure 3 bottom right). Both of these measures (marker signature and speed) are normalized and combined to find a test subject's nearest neighbors and estimate its markers' progression models.

Model averaging. Once a new patient's nearest neighbors (similar cases in the database) have been found, individual realMod models from that subpopulation have to be combined to form the personalized nnMod model of the patient. Averaging the subpopulation models to estimate a nnMod for the new patient seems like a natural choice. However, several individual curves of the same type and their arithmetic mean do not necessarily maintain the same form. That is, the arithmetic mean of several curves functions does not necessarily follow a sigmoid shape. Here, the assumption that individual models should behave in the same way as the mean population model is made, and hence averaging several individual curves should result in a curve of the same type. Mathematically, the validity of this assumption can be ensured as follows [36]: Suppose that a marker is modeled by an arbitrary function,

$$y = f(\boldsymbol{\phi}, x) + \epsilon, \tag{4}$$

where y represents the markers evolution, x represents time, and $\boldsymbol{\phi}$ are the model parameters. Then, n individuals belonging to this population can be modeled as

$$y_i = f(\boldsymbol{\phi}_i, x) \ , \ \ i = [1, 2, ..., n] \ . \tag{5}$$

If $\bar{y}$ is considered to represent the mean marker value at time x, then

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{\phi}_i, x) \ . \tag{6}$$

Parameters of individual curves can be expressed as deviations from the mean model in an analogous way as the random effects in a mixed effects model, and thus can be written as $\boldsymbol{\phi}_i = \bar{\boldsymbol{\phi}} + \boldsymbol{\psi}_i$. Here $\bar{\boldsymbol{\phi}}$ are the parameters' values of the mean curve and $\boldsymbol{\psi}_i$ are constants' associated to the deviation from such curve, therefore an individual's model can be written as

$$f(\boldsymbol{\phi}_i, x) = f(\bar{\boldsymbol{\phi}} + \boldsymbol{\psi}_i, x) \ . \tag{7}$$

Expanding Equation 7 by a Taylor's series expansion,

$$f(\bar{\boldsymbol{\phi}} + \boldsymbol{\psi}_i, x) = f(\bar{\boldsymbol{\phi}}, x) + \boldsymbol{\psi}_i f'_{\boldsymbol{\phi}} + \frac{1}{2!} \left(\boldsymbol{\psi}_i^2 f''_{\boldsymbol{\phi}} \right) + ... \tag{8}$$

where $f^n_{\boldsymbol{\phi}} = \frac{\partial^n f}{\partial \boldsymbol{\phi}^n}$. Any new individual's model can be estimated and expressed as a series expansion, where the first term is a function of the same form as all the other represented curves (if $n = 1$, then $\psi_1 = \vec{0}$ and $\bar{\phi} = \phi_1$). The first term of Equation 8 represents the mean curve, while the remaining terms express the individual divergence from the mean curve. Averaging n such curves can be expressed as

$$\bar{y} = f(\bar{\boldsymbol{\phi}}, x) + f'_{\boldsymbol{\phi}} \frac{\sum_i \psi_i}{n} + \frac{1}{2!} \left(f''_{\boldsymbol{\phi}} \frac{\sum_i \boldsymbol{\psi}_i^2}{n} \right) + ... \tag{9}$$

As mentioned before, the underlying assumption in this work is that all individual trajectories (of a specific marker) are represented by the same model as the population (albeit with different parameters). Therefore, if an estimated nnMod model is to be of the same form as realMod and popMod (training models), then only the first term of Equation 9 should be considered. Therefore, averaging several curves is approximated as finding the mean of their parameters.

3. Experiments and results

3.1. Model training

Linear and non-linear mixed effects models were trained using MATLAB R2014a *nlmefit* and *nlmefitsa*, respectively. In these functions, mixed and random effects parameter estimates that maximize a likelihood function are sought. Here, the predictor (dependent) and the target (independent) variables x_{ij} and y_{ij}, respectively, contain the marker measurements and the time-to-conversion of the jth observation (time point) of the ith subject. Additionally, design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ from Equation 3 are considered identity matrices for simplicity.

Non-linear mixed effects model estimation is based on stochastic approximation version of expectation-maximization (SAEM) algorithm [11]. The SAEM algorithm with simulated annealing [28] theoretically ensures convergence close to a global maximum. Therefore, parameter estimates can slightly vary from run to run depending on the exact initial parameter estimates. Following the suggested analysis done in [26], the mean parameter estimates from 10 different (successful) runs (with different initial parameters) are used.

As mentioned, linear models were used for volumetric markers, where random effects were considered in both the slope and offset. Sigmoid models, used for cognitive and grading markers, had their lower and upper asymptotes fixed at the limits of the corresponding cognitive test and to -1 and 1 for the grading markers. Therefore, only random effects associated with the curvature and inflection point were considered. D1 from the manifold markers was also modeled as a sigmoid function, where additionally random effects were considered for the upper and lower asymptotes as there is no theoretical lower or upper bound of the manifold coordinates. Finally, D2 and P_D1D2 from the manifold were modeled as quadratic functions with random effects considered in the linear and quadratic coefficients.

Using the LMS method the conditional distributions of each marker at each time (after temporal alignment) were used to determine the temporal shift between CN-MCI and MCI-AD subject groups. It must be noted that this procedure failed in the case of CDRSB and FAQ markers, as a large portion of marker samples clustered at zero (in the CN-MCI converter subject group). Therefore, these models were omitted from the group shift calculation. CN-MCI conversion was estimated to occur approximately 59 months before MCI-AD conversion. Comparable results (of $\sim$61 months)

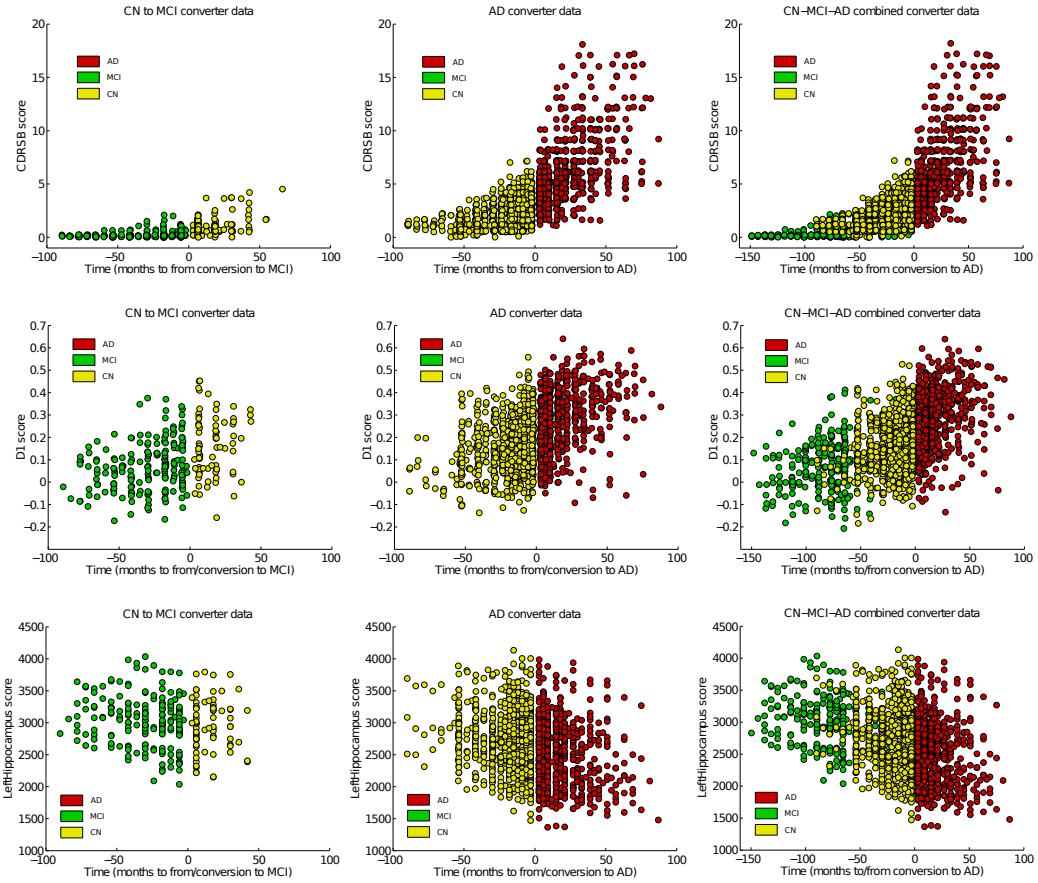

Figure 4: Time-adjusted CDRSB score values of CN-MCI converters, MCI-AD converters and CN-MCI-AD (after temporal alignment). A small amount random noise was artificially added to aid visualization of the cognitive scores.

were found by Schmidt-Richberg [41], where the same technique but a different set of subjects and markers was used. Figure 4 shows some examples of the resulting CN-MCI, MCI-AD and CN-MCI-AD subject groups. Figure 5 displays the resulting popMod, nnMod and realMod models for an example subject from the mixed effects modeling and model personalization for the CDRSB, grading feature, ventricles and D1 markers. Additionally, the measured marker values for the example subject are overlaid as connected arrows.

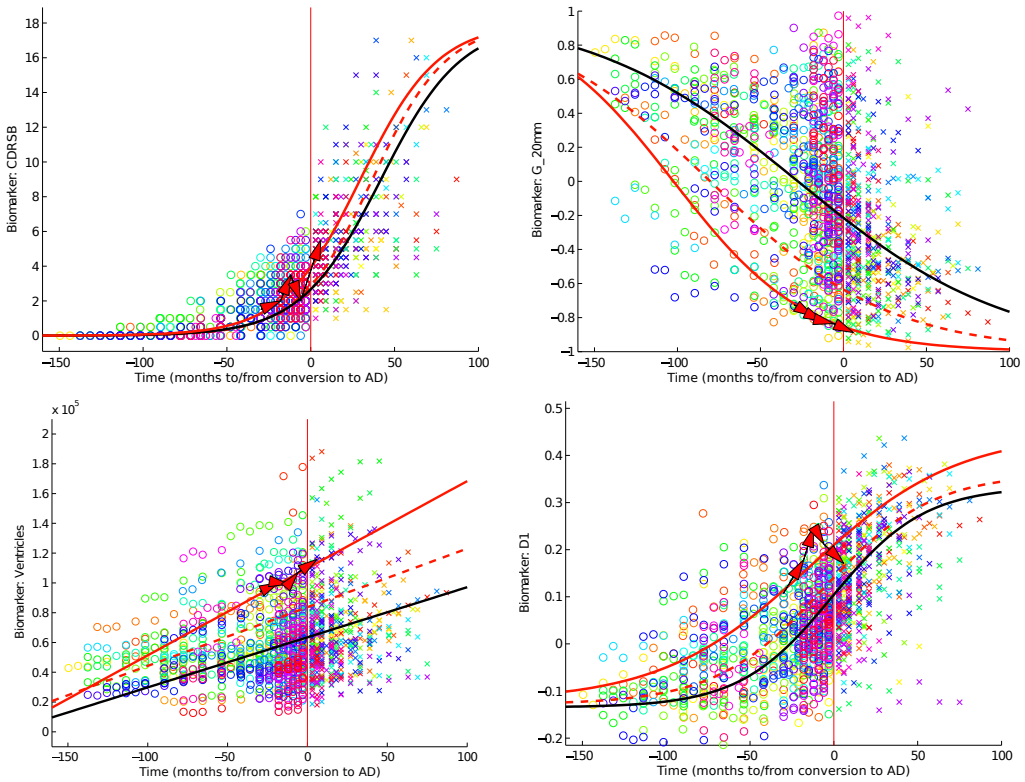

Figure 5: Marker value distribution after time alignment (with respect to time-to-conversion to AD) for CDRSB, patch-based grading, ventricles and D1 (best seen in color). Same color markers indicate same subject. The black curve refers to the fitted mixed effect population model. Vertical red line indicates time of conversion (zero). The red solid curve indicates the "real" fitted mixed effects model for a particular subject. The dashed red curve indicates the estimated nnMod. Arrows connect the measured marker values.

3.2. Parameter exploration

There are two main tuning parameters in the proposed framework, the size of the "marker signature" (the amount of markers that will form part of it) and the number of nearest neighbors sought during subpopulation instantiation to estimate a test subject's nnMod model. Two metrics were explored while varying these parameters: time-to-conversion estimate mean absolute error (Figure 6 top row) and accuracy of the predicted time-to-conversion to detect pre-AD instances less than 24 months from conversion (Figure 6 bottom row). Figure 6 shows the exploration of these tuning parameters using both described metrics with both previously discussed groups

of subjects, MCI-AD and CN-MCI-AD. Given the information provided by Figure 6, tuning parameters were chosen as to provide a good compromise between these two metrics, with the "marker signature" size set to 15 and the number of nearest neighbors set to 10. Both MCI-AD and CN-MCI-AD subject groups' models produce relatively similar results. In Figure 6 top row it can be observed that the optimal "marker signature" sizes for the MCI-AD and CN-MCI-AD populations are between 30-40 and 5-15 markers, respectively. A close inspection of the prefered "marker signature" composition reveled that this varied depending on the choice of population (MCI-AD or CN-MCI-AD). "Marker signature" composition is determined by the correlation coefficient between model parameters of the analysed marker and the measured values of the rest of the markers. There are different amount of samples in either population (1309 and 1561) and hence correlation coefficients vary between them. It was observed that marker preference in the CN-MCI-AD model tended to favor markers usually associated with AD progression, therefore generating more compact "marker signatures".

3.3. Time-to-conversion estimation

Disease progress or state estimation is of great interest to Alzheimer's disease researchers, clinicians and caregivers. Identifying subjects at greatest risk of progressing on a clinical endpoint is relevant in trial design. However, information about the current disease state of a subject and its future trajectory is generally unknown and needs to be estimated. Here, time-to-conversion is used as a surrogate measure of the current disease state. Estimating the time-to-conversion with respect to Alzheimer's disease diagnosis is the main goal here, and two different sets of subjects groups can be used to train the mixed effects models: The MCI-AD conversion group and the CN-MCI-AD group. Time-to-conversion is estimated by finding similar cases of a test subject in the training set, averaging the models of those similar cases to estimate the test subject's individual model, and finally using this model to estimated time-to-conversion (see Section 2.2.3). Three types of model are considered and compared here to make predictions: popMod, derived from the mixed effects modeling (fixed effects), nnMod, based on nearest neighbors (or similar cases), and realMod, again from the mixed effects modeling (fixed + random effects).

Considering that each subject has four or more time points, two scenarios are explored: Cross-sectional estimation, where each instance of the test subject is treated independently to estimate its "current" disease state; and

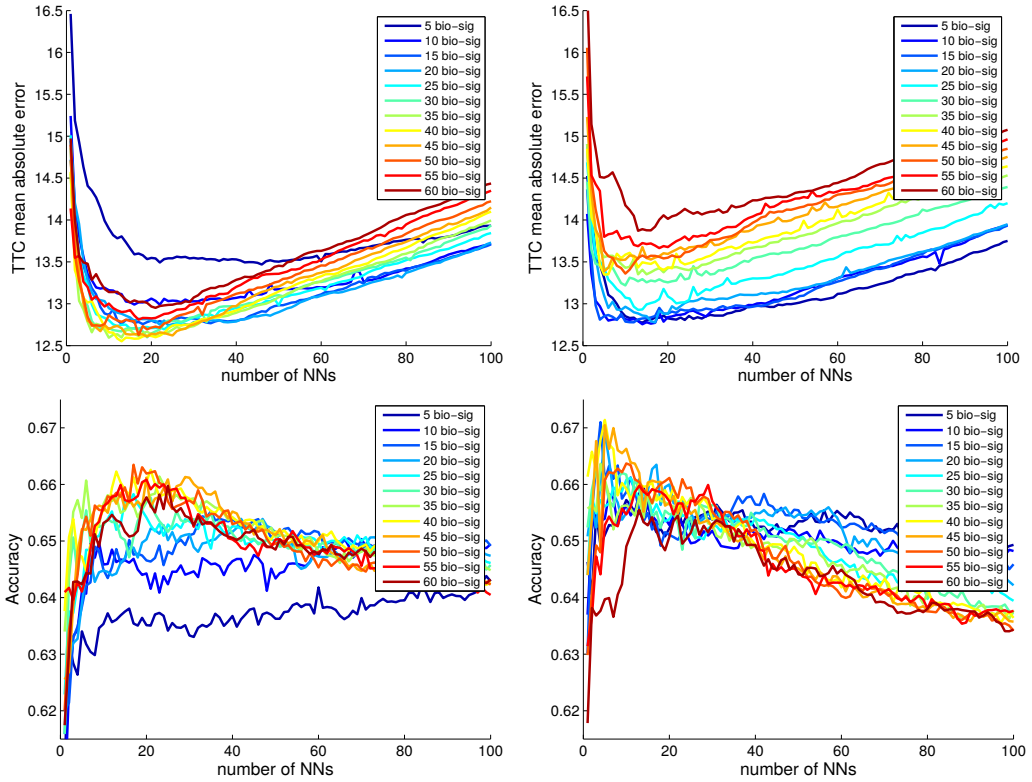

Figure 6: Marker signature size and number of nearest neighbors tuning parameter exploration. Top row: time-to-conversion evaluation. Bottom row: accuracy to detect conversion within 24 months. Left column: using MCI-AD models. Right column: using CN-MCI-AD models.

longitudinal estimation, where one (or more) of the subject's instances is used to estimate its current and future disease states. A leave-one-out experiment was carried out where each subject was removed from the training set and treated as an unseen subject. Table 3 details the results of cross-sectional and longitudinal time-to-conversion estimation using models from all available "core" markers. What are referred here as core markers (see in Table 2) are: Cognitive (MMSE, CDRSB, ADAS11, ADAS13, FAQ and RAVL), manifold (P_D1D2, D1 and D2), volumetric (left and right hippocampus, amygdala and cortex, and ventricles) and grading (using linearly and FFD aligned images at 20, 10, 5 and 2.5 mm). The rest of the markers that make the complete pool from which the "marker signature" can be built consists of the remaining 134 volumetric features and up to 20 manifold dimensions

Model	Experiment	popMod	nnMod	realMod
MCI-AD Core markers	Cross-sectional	19.6 (15.0)*	12.6 (12.0)*†	5.5 (5.8)
	Longitudinal (1 tp)		12.9 (11.9)*†	
	Longitudinal (2 tp)		13.0 (12.3)*†	
	Longitudinal (3 tp)		12.3 (12.1)*†	
CN-MCI-AD Core markers	Cross-sectional	20.5 (15.5)*	13.4 (12.1)*†	5.5 (5.8)
	Longitudinal (1 tp)		12.8 (11.3)*	
	Longitudinal (2 tp)		12.7 (11.9)*†	
	Longitudinal (3 tp)		12.2 (11.8)*†	
MCI-AD No CDRSB/MMSE	Cross-sectional	21.1 (16.2)*	13.3 (12.5)*†	5.7 (6.2)
	Longitudinal (1 tp)		13.6 (12.5)*†	
	Longitudinal (2 tp)		13.6 (12.9)*†	
	Longitudinal (3 tp)		13.0 (12.8)*†	
CN-MCI-AD No CDRSB/MMSE	Cross-sectional	22.1 (16.7)*	14.1 (12.1)*†	5.8 (6.2)
	Longitudinal (1 tp)		13.4 (11.9)*	
	Longitudinal (2 tp)		13.6 (12.4)*†	
	Longitudinal (3 tp)		13.2 (12.3)*†	

Table 3: Cross-sectional and longitudinal (tp: time point) time-to-conversion estimation mean absolute error (standard deviation in parenthesis) of popMod, nnMod and realMod using MCI-AD and CN-MCI-AD models. Statistical significance ($p<0.001$) using a Wilcoxon signed rank test between MCI-AD and CN-MCI-AD indicated by *, and between using all core markers and without CDRSB/MMSE by †.

(as in [41]). In the cross-sectional experiment, estimations for all instances using each type of model (popMod, nnMod and realMod) were made. Here, each marker (and each model type) produces an estimate on the current time-to-conversion, then the predictions from all core markers are averared to give a final estimate. For longitudinal estimation, the baseline visit of each subject was used as the current state in order to find its corresponding nnMod model. Subsequent visits were estimated based on the nnMod model found for the baseline visit. Here, model parameters related to the offset of the current disease state are estimated based on the nearest neighbors according to its "marker signature". Model parameters associated with the slope or curvature of the model are estimated based on the nearest neighbors according to two criteria: marker and temporal goodness of fit (see Section 2.2.3). In the same way as in the cross-sectional experiment, each type of model (popMod, nnMod and realMod) and core marker produces an estimate for every instance on the current time-to-conversion, and predictions from all core markers are averared to give a final estimate.

Since diagnosis in the ADNI protocol is mainly based on MMSE and CDR, and these scores are implicitly used for temporal alignment of the training subjects, there is a potential bias in the results (marked as "Core markers") presented in Table 3. To this end, results excluding these two scores (marked as "No CDRSB/MMSE") are also presented in Table 3. On average, there was an observed decrease in time-to-conversion prediction accuracy between using all core markers and excluding MMSE and CDRSB in the MCI-AD and CN-MCI-AD converter groups of 0.76 and 0.88 months, respectively. The interpretation of this difference needs to disentangle the bias effect from information lost due to the use of less data. To this end, experiments removing any other two cognitive scores (not related to the temporal alignment) produce on average a decline in accuracy of about 0.41 and 0.48 months for MCI-AD and CN-MCI-AD, respectively. Additionally, a statistical comparison, using a Wilcoxon signed rank test, of the results found that 70% of these were not statistically significant ($p<0.001$). These results seem to suggest that values reported in Table 3 could be overestimated by 0.4 months on average while at the same time few of these differences were found to be statistically significant.

Figure 7 shows the error distribution histogram for the longitudinal experiment using one time point. Here, it can be seen that the estimated nnMod models perform better than the average population popMod model in the -30 to 30 month range (closer to the best case scenario of the realMod model), however, they perform worse outside this range. On average, using MCI-AD models, instead of the extended CN-MCI-AD models, produces better results, however, the use of the extended models improves substantially the performance of nnMod for times before -30 months, although at the price of slightly worse results for times after -30. As the data is not uniformly distributed it must be noted that there are far less data points available outside the -30 to 30 month range (see numbers on top of bars in Figure 7) and therefore could be less reliable.

4. Discussion

In the used ADNI data, subject visits on average occur every $\sim$8.7 months and conversion is assumed to occur at the midpoint between the last visit before conversion and the first visit afterwards. Since it is assumed that conversion could have occurred at any point between these two visits with equal probability, it should be noted that there is on average an a-priori mean

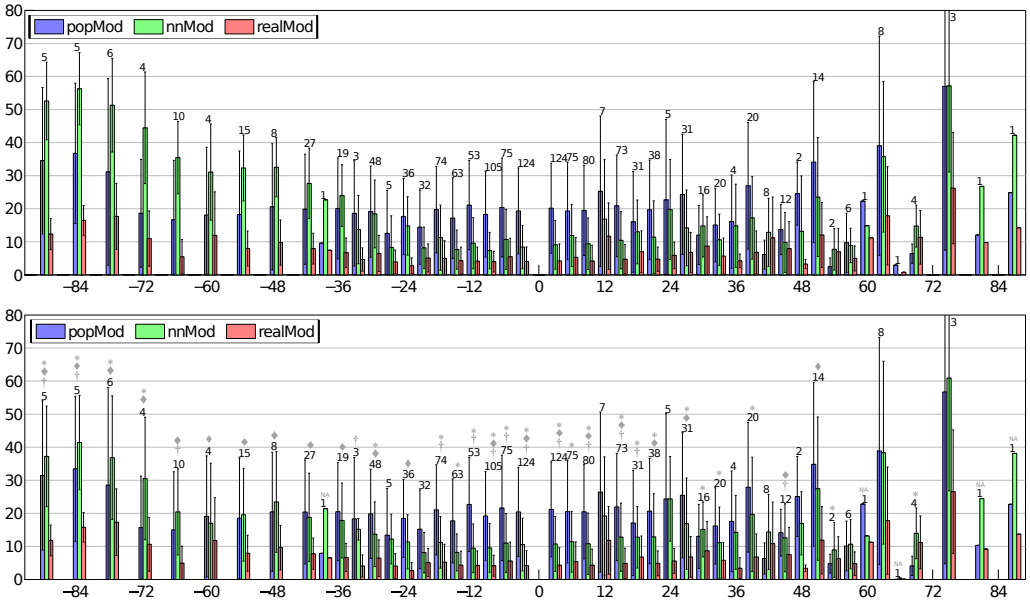

Figure 7: Error distribution histogram for the longitudinal experiment using one time point. X-axis time-to-conversion. Y-axis time-to-conversion estimation mean absolute error. Top figure MCI-AD group, bottom figure CN-MCI-AD group. Whiskers depict the standard deviation of the mean absolute error (drawn on top of bars). On top of wiskers the number of samples available at that time-to-conversion. Symbols on top of wiskers denote statistical significance ($p < 0.05$) between MCI-AD and CN-MCI-AD models predicted error, with * for popMod, ♦ for nnMod, † for realMod and NA where there was only one sample. Note that for some time points there were very few samples, and hence a careful consideration of the statistic results should be given.

absolute error of ∼2.18 (8.7/4) due to the exact conversion time uncertainty. Results presented here should be considered within this context, as this is the absolute minimum expected error due to the average marker sampling rate.

Results from Table 3 suggest that adding more time points to the model estimation process has a very small effect in the overall time-to-conversion estimation accuracy. Additionally, it was observed that the inclusion of several time points generally had a positive effect in the -30 to 30 range, while outside this range results generally degraded. On close inspection of the estimation accuracy of each individual marker on its own, the estimated nnMod performed always better that popMod when only one time point was considered in the model estimation process. These results justify the use of the proposed

"marker signature". Nonetheless, results for the use of several time points were mixed. In this case, all imaging markers generally benefited from the extra information provided by the use of several time points in the estimation process. However, for cognitive features there was no benefit from the use of several time points in the estimation process. In fact, this proved detrimental as in this case the estimated nnMods performed worse than the popMod. A close inspection of the absolute mean individual weighted residuals of models fitted to cognitive scores reveals that on average they are approximately 30% higher than those of other markers. Such behavior could be associated with higher measurement uncertainty (acquisition noise) and therefore estimating model speed from a couple of (noisy) time points measurements might be less beneficial.

some potential areas within the current framework that could be further improved are the model training and how nearest neighbors are selected and handeled. In the current experiments, no additional covariates were added to the mixed effects modeling (design matrices A_i and B_i identity matrices) in order to keep the modeling as simple as possible. However, the methodology permits adding additional subject-specific covariates to the models through design matrices A_i and B_i, or subject- and visit-specific covariates through A_{ij} and B_{ij}. Such covariates could include age, ApoE-ϵ4 genotype or education, all of which were shown to be covariates of the rate of Alzheimer's disease progression [20]. Additionally, some of the markers considered here have a fixed range of possible values, data at the extremes of these markers could be censored during the model training phase [21] in order to improve model fitting. In the work proposed here, the nearest neighbor estimation (Section 2.2.3) does not take into account subject distance for other than to determine neighborhood membership. A simple way of including subject distance information would be to find the weighted average, according to distance, rather than the mean of the parameters. Furthermore, there are two types of neighbors considered here, e.g. from "marker signature" for cross-sectional estimation and from marker "speed" to incorporate longitudinal information. In the current form both of these were treated with equal importance. These could be treated differently, e.g. by weighing parameters, however at the cost of including additional tuning parameters. More complex methods for finding and combining nearest neighbors could be used.

A shortcoming of parametric mixed effects modeling is the implicit trajectory shape assumption being made, where not only the population is assumed to follow a very specific parametric model, but that also individuals

follow the same type of parametric model. Generally, this assumption is not straightforward (or maybe even impossible) to verify in individuals' curves. However, such assumption also ensures compactness and interpretability of the estimated models and hence is usually tolerated (albeit some work focuses on using non-parametric models [12, 42, 41] for population AD marker modeling). In the work proposed here and in a similar way as in mixed effects modeling, when estimating the trajectory of an unseen test subject, the assumption (with its associated caveats) is also made that it should follow the same type of parametric model. The (seemingly) over simplification of the mean curve Taylor expansion from Equation 9 ensures model compactness, interpretability and that test subjects can be easily embedded in the mixed effects model parameter space.

In general, longitudinal modeling of populations markers requires some sort of temporal alignment of subjects. Such temporal alignment can be either approximated from clinical data (e.g. from clinical diagnosis [41], from cognitive score [21] or in special cases from familial history [15]) or estimated from the data (e.g. modeling a time-shift as random effect [22], using predefined sequences of marker abnormality [53] or iteratively from long-term progression curves and subject-specific linear random effects [12]). Moreover, methods that rely on clinical variables for grouping in their analysis can be said to also implicitly use such grouping as a rough form of temporal alignment [4, 5].

Diagnosis of AD is notoriously challenging and conclusive confirmation can only be obtained with histopathologic evidence, therefore the current diagnosis of possible/probable AD [33, 34] is prone to misdiagnosis. The presented approach relies on a clinical diagnosis for temporal pre-alignment, and therefore is also susceptible to errors introduced due to potential misdiagnosis in the data.

One of the main limitations in the presented validation is the bias introduced in the subject inclusion criteria. Here, subjects are only included if conversion to a more severe state of the disease (CN-MCI or MCI-AD) occurs within the (ADNI) study. This reduces the number of available subjects for modeling and more importantly introduces a bias to subjects that progress faster along the disease trajectory. That is, it ignores subjects that might convert but are progressing on the disease trajectory at a slower pace, and hence conversion might fall beyond the study's time frame. As mentioned before and as noted by Donohue et al. [12], this limitation is introduced due to a lack of an obvious biological "time zero" in Alzheimer's disease. Here,

time-to-conversion is used to align subjects, however a more sophisticated alignment in time framework could be used to include those subjects that have thus far not converted.

Acknowledgments

The research discussed here was developed in a joint research project with IXICO plc and funded by an Innovate UK (formerly UK Technology Strategy Board) grant.

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, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimers 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

[1] Adak, S., Illouz, K., Gorman, W., Tandon, R., Zimmerman, E. A., Guariglia, R., Moore, M. M., Kaye, J. A., 2004. Predicting the rate of cognitive decline in aging and early Alzheimer disease. Neurology 63 (1), 108–114.

[2] Ashford, J. W., Schmitt, F. A., 2001. Modeling the time-course of Alzheimer dementia. Current Psychiatry Reports 3 (1), 20–28.

[3] Belkin, M., Niyogi, P., 2002. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Advances in Neural Information Processing Systems 14. Vol. 14. pp. 585–591.

[4] Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., Sabuncu, M. R., 2012. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. NeuroImage 66C, 249–260.

[5] Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., Fischl, B., Sabuncu, M. R., 2013. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. NeuroImage 81, 358–370.

[6] Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H. M., 2007. Forecasting the global burden of Alzheimers disease. Alzheimer's & Dementia 3 (3), 186–191.

[7] Caroli, A., Frisoni, G. B., 2010. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. Neurobiology of aging 31 (8), 1263–74.

[8] Cole, T. J., Green, P. J., 1992. Smoothing reference centile curves: The lms method and penalized likelihood. Statistics in Medicine 11 (10), 1305–1319.

[9] Coupé, P., Eskildsen, S. F., Manjn, J. V., Fonov, V. S., Collins, D. L., 2012. Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. NeuroImage 59 (4), 3736 – 3747.

[10] Delor, I., Charoin, J.-E., Gieschke, R., Retout, S., Jacqmin, P., 2013. Modeling Alzheimer's Disease Progression Using Disease Onset Time and Disease Trajectory Concepts Applied to CDR-SOB Scores From ADNI. CPT: pharmacometrics & systems pharmacology 2, e78.

[11] Delyon, B., Lavielle, M., Moulines, E., 1999. Convergence of a Stochastic Approximation Version of the EM Algorithm. The Annals of Statistics 27 (1), 94–128.

[12] Donohue, M. C., Jacqmin-Gadda, H., Le Goff, M., 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: the journal of the Alzheimer's Association 10 (5 Suppl), S400–10.

[13] Doody, R. S., Massman, P., Dunn, J. K., 2001. A Method for Estimating Progression Rates in Alzheimer Disease. Archives of Neurology 58 (3), 449–454.

[14] Doody, R. S., Pavlik, V., Massman, P., Rountree, S., Darby, E., Chan, W., 2010. Predicting progression of Alzheimer's disease. Alzheimer's research & therapy 2 (1), 2.

[15] 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 (3), 1880–9.

[16] Guerrero, R., Ledig, C., Rueckert, D., 2014. Manifold Alignment and Transfer Learning for Classification of Alzheimer's Disease. In: Machine Learning in Medical Imaging. Vol. 8679. pp. 77–84.

[17] Guerrero, R., Ledig, C., Schmidt-Richberg, A., Rueckert, D., 2015. Group-constrained Laplacian Eigenmaps: Longitudinal AD biomarker learning. In: Machine Learning in Medical Imaging. Vol. 9352. pp. 178–185.

[18] Guerrero, R., Wolz, R., Rao, A. W., Rueckert, D., 2014. Manifold population modeling as a neuro-imaging biomarker: Application to ADNI and ADNI-GO. NeuroImage 94C, 275–286.

[19] Heckemann, R. A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J. V., Hammers, A., 2010. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. NeuroImage 51 (1), 221–227.

[20] Ito, K., Corrigan, B., Zhao, Q., French, J., Miller, R., Soares, H., Katz, E., Nicholas, T., Billing, B., Anziano, R., Fullerton, T., 2011. Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. Alzheimer's & dementia: the journal of the Alzheimer's Association 7 (2), 151–60.

[21] Ito, K., Hutmacher, M. M., 2014. Predicting the time to clinically worsening in mild cognitive impairment patients and its utility in clinical trial design by modeling a longitudinal clinical dementia rating sum of boxes from the ADNI database. Journal of Alzheimer's disease: JAD 40 (4), 967–79.

[22] J.-B. Schiratti, Stéphanie Allassonnière, Alexandre Routier, Olivier Colliot, S. D., 2015. A Mixed-Effects Model with Time Reparametrization for Longitudinal Univariate Manifold-Valued Data. In: Information Processing in Medical Imaging. pp. 564–575.

[23] Jack, C. R., Holtzman, D. M., 2013. Biomarker modeling of Alzheimer's disease. Neuron 80, 1347–58.

[24] Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., Trojanowski, J. Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. The Lancet Neurology 9 (1), 119–128.

[25] Jedynak, B. M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B. T., Raunig, D., Jedynak, C. P., Caffo, B., Prince, J. L., 2012. A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging Initiative cohort. NeuroImage 63 (3), 1478–86.

[26] Kuhn, E., Lavielle, M., 2005. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis 49 (4), 1020–1038.

[27] Lasser, R., Ostrowitzki, S., Scheltens, P., Boada, M., Dubois, B., Dorflinger, E., Balas, B., Nikolcheva, T., Volz, D., Ashford, E., Edgar, C., Garibaldi, G., Fontoura, P., Santarelli, L., 2015. Efficacy and safety of gantenerumab in prodromal Alzheimers disease: Results from scarlet roada global, multicenter trial. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 11 (7), P331–P332.

[28] Lavielle, M., Moulines, E., 1997. A simulated annealing version of the EM algorithm for non-Gaussian deconvolution. Statistics and Computing 7 (4), 229–236.

[29] Ledig, C., Heckemann, R. A., Hammers, A., Lopez, J. C., Newcombe, V. F. J., Makropoulos, A., Lötjönen, J., Menon, D. K., Rueckert, D., 2015. Robust whole-brain segmentation: application to traumatic brain injury. Medical image analysis 21 (1), 40–58.

[30] Leung, K. K., Bartlett, J. W., Barnes, J., Manning, E. N., Ourselin, S., Fox, N. C., 2013. Cerebral atrophy in mild cognitive impairment and Alzheimer disease: rates and acceleration. Neurology 80 (7), 648–54.

[31] Lindstrom, M., Bates, D., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46 (3), 673–687.

[32] Liu, X., Tsai, W. Y., Stern, Y., 1996. A functional decline model for prevalent cohort data. Statistics in medicine 15 (10), 1023–32.

[33] McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E. M., 1984. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34 (7), 939–939.

[34] McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., Phelps, C. H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association 7 (3), 263–9.

[35] Mendiondo, M. S., Ashford, J. W., Kryscio, R. J., Schmitt, F. A., 2000. Modelling mini mental state examination changes in Alzheimer's disease. Statistics in medicine 19 (11-12), 1607–16.

[36] Merrell, M., 1931. The relationship of individual growth to average growth. Human Biology 3 (1), 37–70.

[37] Richards, F. J., 1959. A Flexible Growth Function for Empirical Use. Journal of Experimental Botany 10 (2), 290–301.

[38] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., Hawkes, D. J., 1999. Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging 18 (8), 712–721.

[39] Sabuncu, M. R., Desikan, R. S., Sepulcre, J., Yeo, B. T. T., Liu, H., Schmansky, N. J., Reuter, M., Weiner, M. W., Buckner, R. L., Sperling, R. A., Fischl, B., 2011. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. Archives of neurology 68 (8), 1040–8.

[40] Samtani, M. N., Raghavan, N., Shi, Y., Novak, G., Farnum, M., Lobanov, V., Schultz, T., Yang, E., DiBernardo, A., Narayan, V. A., 2013. Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes. British journal of clinical pharmacology 75 (1), 146–61.

[41] Schmidt-Richberg, A., Guerrero, R., Ledig, C., Molina-Abril, H., Frangi, A. F., Rueckert, D., ADNI, 2015. Multi-stage Biomarker Models for Progression Estimation in Alzheimers Disease. In: International Conference on Information Processing in Medical Imaging (IPMI). Vol. 9123 of Lecture Notes in Computer Science. Springer International Publishing, pp. 387–398.

[42] Schmidt-Richberg, A., Ledig, C., Guerrero, R., Molina-Abril, H., Frangi, A., Rueckert, D., jan 2016. Learning Biomarker Models for Progression Estimation of Alzheimer's Disease. PloS one 11 (4), e0153040.

[43] Sevigny, J., Chiao, P., Williams, L., Miao, X., O'Gorman, J., 2015. Randomized, double-blind, phase 1B study of BIIB037, an anti-amyloid beta

monoclonal antibody, in patients with prodromal or mild Alzheimer's disease. Neurodegenerative Diseases 15 (Suppl. 1), 311.

[44] Siemers, E. R., Sundell, K. L., Carlson, C., Case, M., Sethuraman, G., Liu-Seifert, H., Dowsett, S. A., Pontecorvo, M. J., Dean, R. A., Demattos, R., 2016. Phase 3 solanezumab trials: Secondary outcomes in mild Alzheimer's disease patients. Alzheimer's & dementia : the journal of the Alzheimer's Association 12 (2), 110–120.

[45] Stern, Y., Liu, X., Albert, M., Brandt, J., Jacobs, D. M., Castillo-Castaneda, C. D., Marder, K., Bell, K., Sano, M., Bylsma, F., Lafleche, G., Tsai, W.-Y., 1996. Application of a growth curve approach to modeling the progression of alzheimer's disease. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 51A (4), M179–M184.

[46] Tong, T., Guerrero, R., Ledig, C., Gao, Q., Chen, L., Rueckert, D., 2016. A Novel Grading Biomarker for the Prediction of Conversion from MCI to AD. IEEE Transaction on Biomedical Engineering (in press).

[47] Tong, T., Wolz, R., Coupé, P., Hajnal, J. V., Rueckert, D., 2013. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage 76, 11–23.

[48] Winsor, C. P., 1932. The Gompertz curve as a growth curve. National Academy of Sciences 18 (1).

[49] Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D. P., Rueckert, D., Soininen, H., Ltjnen, J., the Alzheimer's Disease Neuroimaging Initiative, 10 2011. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS ONE 6 (10), e25446.

[50] Yang, E., Farnum, M., Lobanov, V., Schultz, T., Verbeeck, R., Raghavan, N., Samtani, M. N., Novak, G., Narayan, V., DiBernardo, A., 2011. Quantifying the pathophysiological timeline of Alzheimer's disease. Journal of Alzheimer's disease: JAD 26 (4), 745–753.

[51] Yee, T. W., Jul. 2004. Quantile regression via vector generalized additive models. Statistics in medicine 23 (14), 2295–315.

[52] Yeo, I.-K., Johnson, R. A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87 (4), 954–959.

[53] Young, A. L., Oxtoby, N. P., Daga, P., Cash, D. M., Fox, N. C., Ourselin, S., Schott, J. M., Alexander, D. C., 2014. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain: a journal of neurology 137 (Pt 9), 2564–77.