# Estimating long-term multivariate progression from short-term data

Michael C. Donohue[a,*], Hélène Jacqmin-Gadda[b], Mélanie Le Goff[b], Ronald G. Thomas[c],
Rema Raman[a,c], Anthony C. Gamst[a,c], Laurel A. Beckett[d], Clifford R. Jack, Jr.[e],
Michael W. Weiner[f], Jean-François Dartigues[g], Paul S. Aisen[c], for the Alzheimer's Disease
Neuroimaging Initiative[†]

[a]*Department of Family and Preventive Medicine, Division of Biostatistics and Bioinformatics, University of California San Diego, La Jolla, CA, USA*
[b]*INSERM, U897, Biostatistics Department, Bordeaux, France*
[c]*Department of Neuroscience, University of California San Diego, La Jolla, CA, USA*
[d]*Department of Public Health Sciences, Biostatistics Unit, University of California Davis, Davis, CA, USA*
[e]*Department of Radiology, Mayo Clinic, Rochester, MN, USA*
[f]*Center for Imaging of Neurodegenerative Diseases, University of California, San Francisco, CA, USA*
[g]*INSERM, U897, Aging Department, Bordeaux, France*

**Abstract**

**Motivation:** Diseases that progress slowly are often studied by observing cohorts at different stages of disease for short periods of time. The Alzheimer's Disease Neuroimaging Initiative (ADNI) follows elders with various degrees of cognitive impairment, from normal to impaired. The study includes a rich panel of novel cognitive tests, biomarkers, and brain images collected every 6 months for as long as 6 years. The relative timing of the observations with respect to disease pathology is unknown. We propose a general semiparametric model and iterative estimation procedure to estimate simultaneously the pathological timing and long-term growth curves. The resulting estimates of long-term progression are fine-tuned using cognitive trajectories derived from the long-term "Personnes Agées Quid" study.

**Results:** We demonstrate with simulations that the method can recover long-term disease trends from short-term observations. The method also estimates temporal ordering of individuals with respect to disease pathology, providing subject-specific prognostic estimates of the time until onset of symptoms. When the method is applied to ADNI data, the estimated growth curves are in general agreement with prevailing theories of the Alzheimer's disease cascade. Other data sets with common outcome measures can be combined using the proposed algorithm.

**Availability:** Software to fit the model and reproduce results with the statistical software R is available as the *grace* package. ADNI data can be downloaded from the Laboratory of NeuroImaging.
© 2014 The Alzheimer's Association. All rights reserved.

*Keywords:* Multiple outcomes; Semiparametric regression; Self-modeling regression; Progression curves; Growth curves

## 1. Introduction

Several methods exist for estimating smooth progression or growth curves from serial observations of individuals over some biologically common time span. For example, generalized linear or nonlinear mixed effects models [1] can be used to model height, weight, or pharmacokinetics over time from some event of interest. The event might be birth or an intervention. However, we often study diseases that occur over long periods of time by sampling populations at different stages of

disease and taking short-term longitudinal "snapshots." Epidemiologic studies with biologically heterogeneous subpopulations may not have an obvious biological event that can serve as a reference "time zero." Such a time zero is required to fit the standard mixed-effects model. Also, the standard nonlinear mixed-effects models and software assume similar features on both the subject and population levels [1–3]. Short-term follow-up with relatively few observations may require much more simple subject-level features.

A motivating example is Alzheimer's disease (AD), which is believed to develop decades before the onset of symptoms. The Alzheimer's Disease Neuroimaging Initiative (ADNI) [4], has followed volunteers diagnosed as cognitively normal (CN), early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI), and probable mild AD. Maximum follow-up is currently as long as about 6 years, and data collection is ongoing. The ADNI battery includes serial magnetic resonance imaging (MRI) measures of regional brain volumes, positron emission tomography (PET) measures of brain function and amyloid accumulation, other biological markers, and clinical and neuropsychological assessments. Time of onset of dementia, a potential time zero, is recorded for subjects with or transitioning to dementia, but these times can be unreliable and subjective. Furthermore, CN individuals and individuals with mild cognitive impairment may not be followed long enough to observe clinical transitions.

Jack and colleagues [5,6] proposed a long-term model of the AD pathological cascade, and hypothesized specifically the trajectory of several key biomarkers during the decades preceding the onset of dementia symptoms. The model (Fig. 1) proposes that the AD cascade begins many years before the onset of symptoms, with amyloid plaque deposition in the brain followed by neurofibrillary tau tangles; cognitive, clinical, and functional decline are relatively late features of the disease. This hypothesized model (Fig. 1) is shaping the field of AD research. Drug development and observational

studies have shifted focus to earlier stages of the disease, selecting subjects based on biomarkers instead of symptomatic impairment. Ideally, we would test the hypothesized model of Fig. 1 by enrolling a large cohort of CN subjects and by collecting biomarkers and cognitive and functional assessment results for decades. The subset who progress to AD could be used to model the long-term biomarker progression of the disease. Until such a study is conducted, we are limited to analyzing shorter term studies, such as ADNI.

Self-modeling regression (SEMOR) is an approach for fitting sets of curves under the assumption of a common shape [7]. A subclass of SEMOR, shape-invariant models [8–11], accommodate unknown location and scale parameters for both the outcome and the time covariate, and model the common shape with regression splines. Kneip and Gasser [7] relaxed some of the parametric assumptions by using kernel smoothers to estimate the common shape. Others have modeled the common shape with free-knot regression splines [12], smoothing splines [13], and penalized splines [14–16].

To our knowledge, SEMOR has been applied only to data sets in which each subject has similar follow-up. SEMOR approaches assume a common shape throughout the population, and estimate subject-level curves with similar features as the population curve. Our goal is to estimate population curves for decades of AD progression on an array of outcome measures; however, subject-level data comprise, at most, nine observations over 6 years. We propose a SEMOR model with simple, linear subject-level effects, while modeling long-term features with nonparametric monotone smoothing. Subjects are shifted backward or forward in time according to performance across the panel of outcomes. Long-term progression curves for the multiple outcomes, and subject-specific random effects and time shifts are estimated iteratively until convergence of the residual sum of squares (RSS).

There have been studies of AD progression with long-term follow-up, but these tend to lack the novel biomarkers
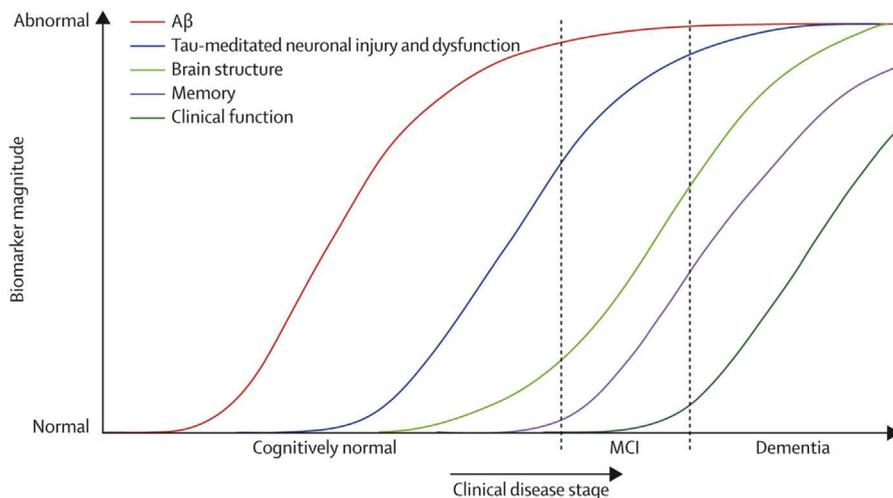


Fig. 1. Dynamic biomarkers of the Alzheimer's disease cascade hypothesized by Jack and colleagues [5]. Aβ, amyloid β; MCI, mild cognitive impairment.

of prime interest in early stages of the disease. For instance, the "Personnes Agées Quid" (PAQUID) study has followed 3777 French individuals age 65 years or older studied from 1988 until the present [17]. The PAQUID data set lacks the imaging and cerebrospinal fluid (CSF) biomarkers that ADNI collected, but provides invaluable long-term Mini-Mental State Examination (MMSE) trajectories [18]. We can use these trajectories to fine-tune the results of the algorithm applied to ADNI data, and transform time to represent time to dementia onset.

## 2. Model assumptions

We assume several outcomes $Y_{ij}$ arise during time $t$ for individual $i = 1, \ldots, n$ and outcome $j = 1, \ldots, m$ according to

$$Y_{ij}(t) = g_j(t + \gamma_i) + \alpha_{0ij} + \alpha_{1ij}t + \varepsilon_{ij}(t). \quad (1)$$

Furthermore, we assume each $g_j$ is a continuously differentiable monotone function, $\gamma_i$'s have mean zero and variance $\sigma_\gamma^2$, $(\alpha_{0ij}, \alpha_{1ij})$ are bivariate Gaussian with mean zero and covariance matrix $\sum_j$, and $\varepsilon_{ij}(t)$'s are independent Gaussian residual errors with mean zero and outcome-specific variance $\sigma j$. To simplify notation, we think of $t$ as both a covariate and a continuous valued index. "Short-term" observation time is represented by observed covariate $t$. In a panel study such as ADNI, $t$ would correspond to the study time clock. "Long-term" progression time is represented by $t + \gamma_i$, where $\gamma_i$ is the unknown subject-specific time shift. If subjects age uniformly, with identical ages at different stages of progression of the underlying disease features, "long-term" progression time would be the subjects' age; in fact, however, disease manifests at different ages, so this corresponds to an unknown "health age," which may be shifted left or right relative to actual age.

Fig. 2A depicts simulated data generated according to equation (1). The logistic function $g_1(t) = \frac{1}{1 + \exp(-t)}$, the linear function $g_2(t) = \frac{t}{12 + 0.5}$, and the quadratic $g_3(t) = \frac{(t + 6)^2}{72}$ generated the three outcomes. For each of the 100 subjects, we sampled subject-specific time shifts, $\gamma_0$, uniformly from the interval –5 to 5. The unshifted observation times were $t = -1, -0.5, 0, 0.5, 1$. The random intercepts and slopes for each subject and outcome are distributed according to a bivariate Gaussian with mean zero, variance 0.01, and covariance 0.005. The residual variance is also Gaussian with variance 0.01. We chose the different long-term shapes to test whether our semiparametric method could recover them without supervision. The observation times and long-term scatter were chosen to mimic ADNI roughly. The variance parameters were chosen so that the long-term trends were reasonably apparent by visual inspection of Fig. 2A

The long-term trends are obvious in Fig. 2A because the data are plotted with the simulated time shifts. However, the time shifts are not observed in data such as ADNI's. Rather, the data are observed as in Fig. 2B. The goal of the algorithm

proposed in the next section is to estimate both the time shift parameters and the long-term curves. The algorithm will leverage the assumption that the long-term trends are monotone, and will pool information across outcomes to estimate the subject-specific time shifts.

The restriction that $\gamma_i$, $\alpha_{0ij}$, and $\alpha_{1ij}$ each have mean zero, helps ensure identifiability (i.e., that the parameters of the model are uniquely determined). Without the random slope term $\alpha_{1ij}$, our model is a simplification of the classical shape invariant model (SIM) for each outcome. The SIM includes two rescaling parameters and two shift parameters. Our model excludes the SIM rescaling parameters but includes an additional random slope term. Without the random slope term, identifiability of our model is established in Kneip and Gasser [7] under the normalizing condition that shift parameters $\gamma_i$ and $\alpha_{0ij}$ have mean zero, which we maintain. To ensure identifiability in our model with a random slope, $\alpha_{1ij}$, we simply require the additional restriction that the mean of $\alpha_{1ij}$ is zero. Following Kneip and Gasser [7], the restrictions on the mean of $\gamma_i$, $\alpha_{0ij}$, and $\alpha_{1ij}$, and the assumption that $g_j$ is a continuously differentiable monotone function for each outcome, ensure identifiability.

## 3. The algorithm

The algorithm reduces the highly dimensional and complex problem into simpler problems. Each of the unknown parameters ($g_j$, $\gamma_i$, and $\alpha$) is estimated in turn using the current estimates of the other parameters. This process is iterated until convergence of the RSS. The algorithm uses three different types of *partial residuals*, using the language of generalized additive model estimation [19], which we denote $R_{ij}^g(t)$, $R_{ij}^\alpha(t)$, $R_{ij}^\gamma(t)$ (Table 1). If we assume that model (1) is correct, then each of the partial residuals provides an unbiased estimate of one of the unknown parameters. Specifically, conditional expectations of the partial residuals are equivalent, or at least approximately equivalent, to the target parameters (Table 1).

We begin the algorithm by initializing $\gamma_i = 0$ and iterating the following.

1. Given $\gamma_i$, estimate the monotone functions $g_j$ by setting $\alpha_{0ij} = \alpha_{1ij} = 0$ and iterating the following subroutine.
   (a) Estimate $g_j$ by a monotone smooth of $R_{ij}^g(t)$.
   (b) Estimate $\alpha_{0ij}$, $\alpha_{1ij}$ by the linear mixed model of $R_{ij}^\alpha(t)$. Repeat steps a and b until convergence of the RSS for the $j$th outcome: $\text{RSS}_j = \sum_{it} [Y_{ij}(t) - g_j(t + \gamma_i) - \alpha_{0ij} - \alpha_{1ij}t]^2$.
2. Given current set of $g_j$, set $\alpha_{0ij} = \alpha_{1ij} = \varepsilon_{ij}(t) = 0$, and estimate each $\gamma_i$ with the average of $R_{ij}^\gamma(t)$ over all $j$ and $t$. Repeat steps 1 and 2 until convergence of the total RRS equals $\Sigma_{ijt}[Y_{ij}(t) - g_j(t + \gamma_i) - \alpha_{0ij} - \alpha_{1ij}t]^2$.

Step 1 involves $m$ parallel subroutines for fitting $g_j$ and the subject-specific $\alpha_{0ij}$ and $\alpha_{1ij}$ for each outcome $j = 1, \ldots, m$. We begin each of the $m$ parallel subroutines by setting
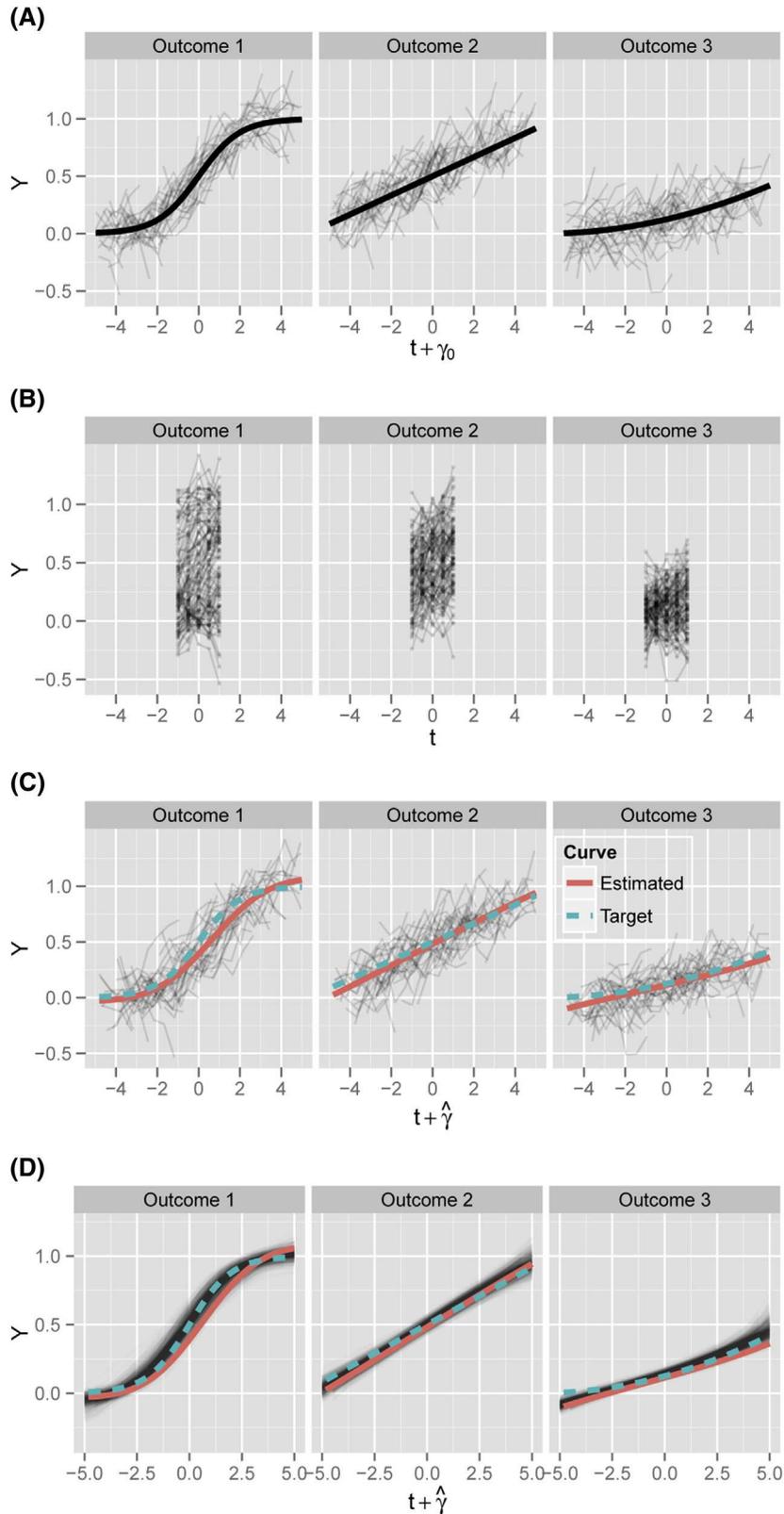
Fig. 2. (A) Simulated long-term data. The three monotone functions depicted in bold are logistic, linear, and quadratic. Long-term trends are easily apparent because data are plotted with the unknown time shifts. The simulated data are not derived from real data and are intended for demonstration only. (B) Simulated short-term observations. Long-term trends are obscured because we observe the data in the short term, without the unknown time shifts. (C) Long-term curves estimated from short-term observations. The algorithm described in section 3 estimates long-term curves (red line), with good fidelity to the true target curves (dashed green line). (D) Fitted curves from 1000 simulations. We repeated the experiment, generating new data and fitting the curves 1000 times, then plotted the fitted curves in black.

Table 1
Partial residuals for each target parameter and their conditional expectations

| Partial residual | Conditional expectation |
|---|---|
| Long-term smooth curve: $g_j$ | |
| $R_{ij}^g(t) = Y_{ij}(t) - \alpha_{0ij} - \alpha_{1ij}t$ | $E(R_{ij}^g(t)\|g_j, t, \gamma_i) = g_j(t + \gamma_i)$ |
| Subject- and outcome-specific intercept and slope: $\alpha_{0ij}, \alpha_{1ij}$ | |
| $R_{ij}^\alpha(t) = Y_{ij}(t) - g_j(t + \gamma_i)$ | $E(R_{ij}^\alpha(t)\|\alpha_{0ij}, \alpha_{1ij}, t) = \alpha_{0ij} + \alpha_{1ij}t$ |
| Subject-specific time shift: $\gamma_i$ | |
| $R_{ij}^\gamma(t) = t - g_j^{-1}(Y_{ij}(t))$ | $E(R_{ij}^\gamma(t)\|\gamma_i) \approx g_j^{-1}(g_j(t + \gamma_i)) - t = \gamma_i$ |

NOTE. The algorithm estimates each target parameter of our model via the partial residuals defined here. Under the assumptions of the model, we see that the conditional expectations of the partial residuals are equivalent to the parameters of interest. The equivalence is approximate for $E(R_{ij}^\gamma(t)\|\gamma_i)$, because here we integrate over the function $g_j^{-1}$. The approximation is reasonable provided $\alpha_{0ij}$, $\alpha_{1ij}$, and $\varepsilon_{ij}$ are small and $g_j^{-1}$ is not too steep.

$\alpha_{0ij} = 0$ and $\alpha_{1ij} = 0$. To estimate $g_j$, we use a monotone B-spline smoother [20] through the scatterplot of $[t + \gamma_i, R_{ij}^g(t)]$. This is accomplished using the R package *fda* [21]. Using the model, $R_{ij}^g(t)$ is distributed independently and identically about $g_j(t)$, so we need not model within-subject correlations at this step. To estimate $\alpha_{0ij}$ and $\alpha_{1ij}$, we minimize $R_{ij}^\alpha(t)$ by fitting a linear mixed-effect model of $R_{ij}^\alpha(t) = \alpha_{0ij} + \alpha_{1ij}t + \varepsilon_{ij}(t)$ using *lme4* [22]. Steps a and b are repeated with the same $\gamma_i$ until convergence of RSS_j. The result is *m* smooth curves, $g_1, \ldots, g_m$, and $m \times n$ sets of random effects estimates for the *m* outcomes and *n* individuals. Plots of the fits and residuals at each iteration are produced with *ggplot2* [23].

In step 2, we invert the outcome variables and estimate the time shift parameters $\gamma_i$ by taking the average of $R_{ij}^\gamma(t)$ over all outcomes and times for each individual. This is the only step that pools data derived from all outcomes at once. To down-weight the influence of more variable outcomes, one could use a weighted average with weights inversely proportional to each outcome's residual variance.

## 4. Simulations

Data simulated as described in Section 2 are depicted in Fig. 2. We submitted these data to the algorithm. Each curve was estimated with the same monotone B-spline smoother with five equally spaced knots and fifth-degree polynomial splines. The resulting fitted curves are shown to have good fidelity with the true logistic, linear, and quadratic curves (Fig. 2C).

We also plotted the true simulated time shifts against the estimated time shifts (not shown). The agreement was not perfect, but, as hoped, the regression line through this scatterplot lies close to the identity line. The RSS for each of the outcomes converged in 10 iterations to a tolerance of 0.1% of the RSS. Code to reproduce the results and an animation demonstrating convergence is available [24].

## 5. ADNI and PAQUID results

Fig. 3 shows longitudinal plots of some of the key variables that have been collected during the course of ADNI. Amyloid plaque accumulation in the brain is associated with decreased CSF and increased Pittsburgh compound B (PiB) and florbetapir uptake on PET. Fig. 3 also includes CSF tau and phosphorylated tau (p-tau); FreeSurfer volumetric MRI data for hippocampal, whole-brain, and ventricular volume; fluorodeoxyglucose uptake on PET; the 13 item Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS13) including delayed word recall and number cancellation tasks; the MMSE; the Alzheimer's Disease Cooperative Study Functional Activities Questionnaire; the Rey Auditory Visual Learning Test, and the Clinical Dementia Rating Scale Sum of Boxes (CDRSB). CSF measures were collected only on a subset of ADNI volunteers, as evidenced by the relative sparsity of CSF data. Florbetapir PET imaging and the EMCI cohort were added relatively late in the study. Additional details on these measures in ADNI are available [4].

One of the primary motivations for this work was to derive a data-driven version of the progression curves hypothesized by Jack and colleagues [5]. The hypothesized figure (Fig. 1) shows the key markers of disease progressing on a common vertical scale from normal to abnormal, with clinical disease stage on the horizontal scale. The percentile scale is a natural choice to attain a common scale. Therefore, before submitting ADNI measures to our algorithm, we first transformed them to a percentile scale. Because the diagnostic groups are not represented equally, we use a weighted percentile transformation. The resulting scale is percentile normalized to range from 0 (least severe observed value) to 100 (most severe observed value). Percentiles were calculated using the empirical cumulative distribution function, derived by weighting according to the inverse of the proportion of observations from each diagnostic category (CN, EMCI, LMCI, and AD). Table 2 provides counts of subjects and observations by baseline diagnostic category. For example, we observed N=7216 MMSE scores across all diagnostic groups and time points. Among those N=7216 MMSE observations, N=2094 (29.0%) were from CN, N=979 (13.6%) were from EMCI, N=3074 (42.6%) were from LMCI, N=1069 (14.8%) were from AD. Because diagnostic groups are not represented equally, we used the inverse proportions as weights when computing the empirical cumulative distribution function. Table 3 provides the raw values that correspond to the resulting percentiles.
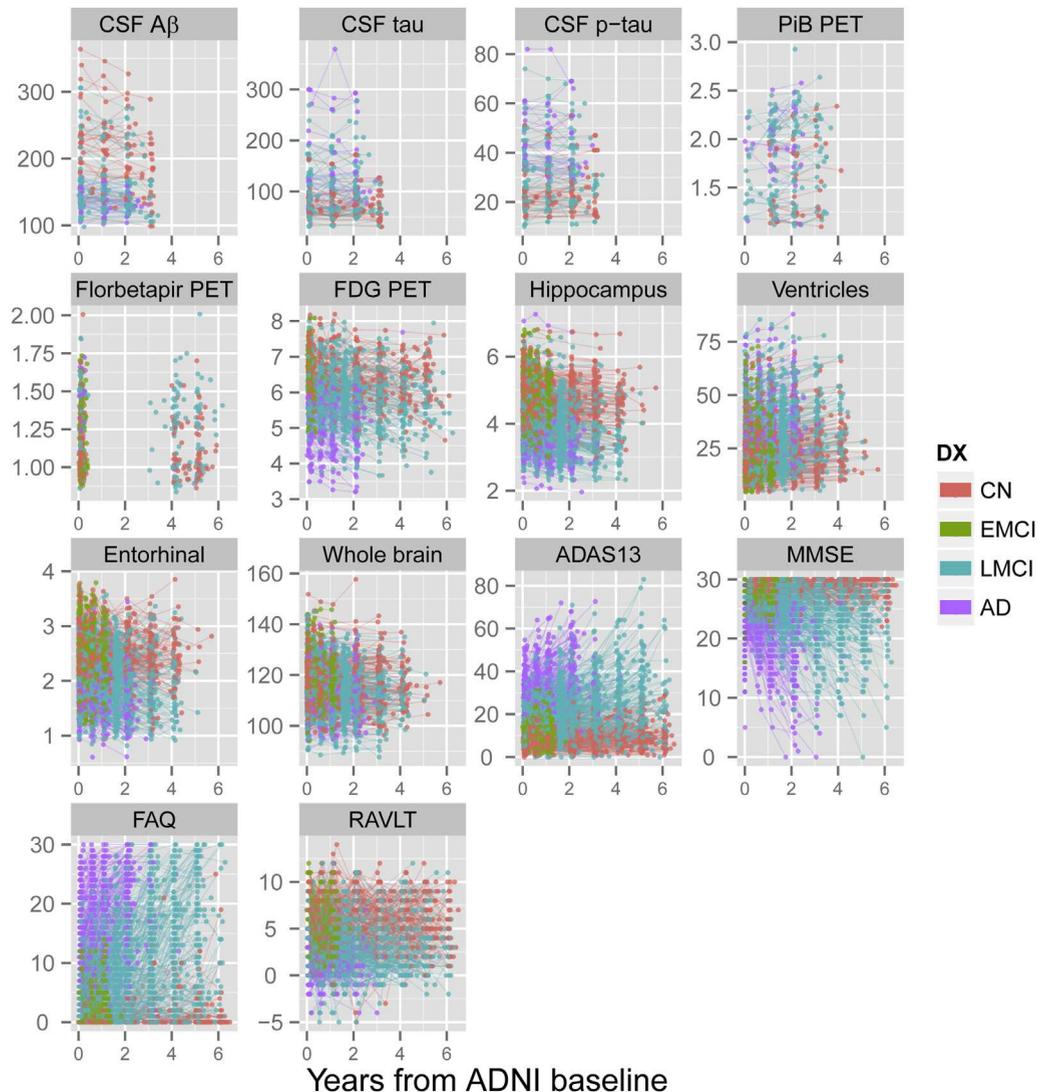
Fig. 3. The Alzheimer's Disease Neuroimaging Initiative (ADNI) battery consists of a rich panel of biomarkers and assessments collected at 6-month intervals for as long as 6 years. Subjects began the study as one of four diagnostic categories. The early mild cognitive impairment (EMCI) cohort was enrolled relatively recently, so there is relatively less data from that cohort. Cerebrospinal fluid (CSF) measures are not collected from every ADNI volunteer. Some measures, such as florbetapir positron emission tomography (PET), have not been collected for as long. There is no obvious biological or clinical reference time point. The *x*-axis is time since the first ADNI visit. CSF measures are in pg/mL, PET measures are standardized uptake value ratios, and hippocampus and ventricles are percent intracranial volumes. Aβ, amyloid-β; p-tau, phosphorylated tau; PiB, Pittsburgh compound B; FDG, fluorodeoxyglucose; ADAS13, the 13 item Alzheimer's Disease Assessment Scale–Cognitive Subscale; MMSE, Mini-Mental State Examination; FAQ, Alzheimer's Disease Cooperative Study Functional Activities Questionnaire; RAVLT, Rey Auditory Visual Learning Test; CN, cognitively normal; LMCI, late mild cognitive impairment; AD, Alzheimer's disease.

ADNI includes many CN subjects, individuals with mild cognitive impairment, and even misdiagnosed probable AD subjects who may not have AD pathology. This is because physicians do not have access to biomarker results at the time of diagnosis. These subjects do not provide any information about the long-term trends of AD. Therefore, we applied our algorithm to the subset 388 ADNI participants with some evidence of abnormal accumulation of amyloid in the brain (amyloid+) using published thresholds for CSF amyloid-β (Aβ; 192 pg/mL), PiB PET (1.5 standardized uptake value ratio in the region relative to the cerebellum), and florbetapir PET (1.1 standardized uptake value ratio in the region relative to the cerebellum) [25–27]. The group consisted of 100

CN subjects, 137 individuals with EMCI, 225 individuals with LMCI, and 117 individuals with AD, but the algorithm was blind to these diagnostic categorizations. Using the same approach as the simulation, the B-spline smooths were fitted with five equally spaced knots and fifth-degree polynomial splines.

Fig. 4A top, shows the estimated long-term trajectories among amyloid+ ADNI subjects. Time has been shifted so that time zero represents the time at which the mean CDRSB score reaches the 80th percentile. The resulting timescale can be interpreted as time until progression to the worst 20th percentile of CDRSB. To reduce clutter and because it was very similar to the ADAS13 trajectory, CDRSB is not shown

Table 2
Counts of subjects and observations by baseline diagnosis

| Outcome | CN, n/N* | EMCI, n/N* | LMCI, n/N* | AD, n/N* | Total |
|---|---|---|---|---|---|
| CSF tau, pg/mL | 36/132 | | 54/211 | 16/63 | 106/406 |
| CSF p-tau, pg/mL | 36/132 | | 54/211 | 16/63 | 106/406 |
| CSF amyloid-β, pg/mL | 36/132 | | 54/211 | 16/63 | 106/406 |
| PiB PET, SUVR | 19/49 | | 65/141 | 19/34 | 103/224 |
| Florbetapir PET, SUVR | 270/270 | 290/290 | 252/252 | 98/98 | 910/910 |
| FDG PET uptake | 345/763 | 299/346 | 409/1263 | 208/434 | 1261/2806 |
| Ventricles, % ICV | 413/1448 | 298/852 | 548/1927 | 267/637 | 1526/4864 |
| Hippocampus, % ICV | 413/1448 | 298/852 | 548/1927 | 267/637 | 1526/4864 |
| ADAS13 | 418/1883 | 305/825 | 560/2735 | 309/871 | 1592/6314 |
| MMSE | 608/2094 | 450/979 | 864/3074 | 474/1069 | 2396/7216 |
| FAQ | 416/1891 | 304/823 | 560/2781 | 310/907 | 1590/6402 |
| RAVLT | 419/1893 | 306/825 | 560/2739 | 312/884 | 1597/6341 |

Abbreviations: CN, cognitively normal; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, Alzheimer's disease; CSF, cerebrospinal fluid; p-tau, phosphorylated tau; PiB, Pittsburgh compound B; PET, positron emission tomography; SUVR, standardized uptake value ratio; FDG, fluorodeoxyglucose; ICV, intracranial volume; ADAS13, the 13-item Alzheimer's Disease Assessment Scale–Cognitive Subscale; MMSE, Mini-Mental State Examination; FAQ, Alzheimer's Disease Cooperative Study Functional Activities Questionnaire; RAVLT, Rey Auditory Visual Learning Test.

NOTE. The number of subjects and observations available in the ADNI data set varies by outcome.

*Total number of subjects (n) and observations (N).

in the middle panel of Fig. 4A. We also omit ventricular volume, which tracks closely with the hippocampus. Fig. 4A, bottom left, depicts the first derivatives of each curve divided by the standard deviation of its residuals. The shaded regions in the top panels of Fig. 4 depict bootstrap estimates of the confidence bands. We resampled the subjects with replacement and reapplied the algorithm 100 times. Each resampled population contained the same number of subjects as the observed population. For each time point, we then took the 2.5 percentile and the 97.5 percentile of the 100 curves as the lower and upper limits.

We also applied the algorithm to the subset of 570 ADNI participants who had at least one apolipoprotein E (*APOE*) ε4 allele (Fig. 4). This subgroup consisted of 92 CN subjects, 85 individuals with EMCI, 248 individuals with LMCI, and 145 individuals with AD, but again the algorithm was blind to the diagnoses. Note that this group contains many subjects who would be classified as amyloid–. Using MMSE trajectories from the PAQUID study, we applied a postprocessing step to transform time. The PAQUID timescale is time to onset of dementia. The ADNI timescale, with estimated subject-specific time shifts, lacks a pathological anchor. To transform the ADNI timescale to the PAQUID time to onset, we composed the ADNI MMSE trajectory with the inverse PAQUID trajectory. That is, if $f$ denotes the estimated MMSE curve from ADNI and $g$ denotes the same from PAQUID, with inverse $g^{-1}$, we transform ADNI time, $t$, via the composition $g^{-1}[f(t)]$. Because PAQUID lacks measures of amyloid burden, we could not do this transformation with our amyloid+ analysis.

For comparison, Fig. 5 depicts trajectories estimated from those without evidence of amyloid burden and those without an *APOE* ε4 allele. We used the same transformations of time as described earlier, in particular using estimated trajectories from PAQUID to calibrate the ADNI timescale. The amyloid– group consisted of 190 CN sub-

jects, 153 individuals with EMCI, 92 individuals with LMCI, and 13 individuals with AD. The *APOE* ε4 allele noncarrier group consisted of 263 CN subjects, 124 individuals with EMCI, 219 individuals with LMCI, and 75 individuals with AD.

## 6. Discussion

Bateman and colleagues [28] recently produced estimated progression curves for a 50-year span using data from the Dominantly Inherited Alzheimer's Network. A key feature of autosomal dominant AD is that the age of onset of symptoms is expected to be close to the age of onset of the parent. Bateman and colleagues [28] use the parents' ages of onset to estimate long-term disease progression from cross-sectional data from mutation carriers spanning 25 years before, to 10 years after, the parents' age of onset. In contrast, we have less confidence about the age of onset in the ADNI population of sporadic AD. Our SEMOR approach addresses this limitation of ADNI by estimating age of onset and the progression curves simultaneously.

Simulations suggest that our iterative algorithm can recover reasonable estimates of the long-term trajectories from short-term observations. Nonparametric estimation of the monotone curves allows different-shape curves to emerge without prespecifying parametric families. Additional simulation studies and analytical development of asymptotic convergence is warranted. Convergence of estimates of the time shifts will rely in part on the abundance of outcomes. With only three outcomes used in the simulation, the algorithm estimated the time shifts surprisingly well.

Jack and colleagues [5,6] proposed that all disease markers range from zero (absolutely normal) to one (absolutely abnormal) and follow sigmoidal shapes. Rather than assuming this to be true, and using a parametric approach, we opted to follow a nonparametric monotone

Table 3
Percentiles of key ADNI outcomes

| Outcome | n/N* | Percentiles | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 25th | 50th | 75th | 100th |
| CSF tau, pg/mL | 106/406 | 31 | 64 | 90 | 122 | 379 |
| CSF p-tau, pg/mL | 106/406 | 10.0 | 21.6 | 32.3 | 42.0 | 82.0 |
| CSF amyloid-β, pg/mL | 106/406 | 364 | 181 | 146 | 131 | 98 |
| PiB PET, SUVR | 103/224 | 1.09 | 1.36 | 1.85 | 2.09 | 2.93 |
| Florbetapir PET, SUVR | 910/910 | 0.83 | 1.01 | 1.21 | 1.41 | 2.01 |
| FDG PET uptake | 1261/2806 | 8.54 | 6.56 | 6.06 | 5.50 | 3.20 |
| Ventricles, % ICV | 1526/4864 | 0.45 | 1.67 | 2.44 | 3.40 | 9.03 |
| Hippocampus, % ICV | 1526/4864 | 0.79 | 0.50 | 0.43 | 0.36 | 0.20 |
| ADAS13 | 1592/6314 | 0.00 | 9.23 | 15.91 | 26.26 | 85.00 |
| MMSE | 2396/7216 | 30.0 | 29.2 | 27.5 | 24.4 | 0.0 |
| FAQ | 1590/6402 | 0.00 | 0.73 | 4.04 | 12.87 | 30.00 |
| RAVLT (Trial 5–Trial 1) | 1597/6341 | 14.00 | 6.37 | 4.08 | 2.18 | −5.00 |

Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, Alzheimer's disease; CSF, cerebrospinal fluid; p-tau, phosphorylated tau; PiB, Pittsburgh compound B; PET, positron emission tomography; SUVR, standardized uptake value ratio; FDG, fluorodeoxyglucose; ICV, intracranial volume; ADAS13, the 13-item Alzheimer's Disease Assessment Scale–Cognitive Subscale; MMSE, Mini-Mental State Examination; FAQ, Alzheimer's Disease Cooperative Study Functional Activities Questionnaire; RAVLT, Rey Auditory Visual Learning Test.

NOTE. Each of the ADNI outcomes was transformed to a common percentile scale. Percentiles were calculated using the empirical cumulative distribution function weighted according to the inverse of the proportion of observations from each diagnostic category (CN, EMCI, LMCI, and AD). Increasing percentile scores are intended to be associated with worsening of the disease. Here we provide the raw values associated with the given percentile values.

*Total number of subjects (n) and observations (N).

smoothing approach. We chose to apply our algorithm to relatively pathologically homogenous amyloid+ and *APOE* ε4+ subsets. Although restricting to amyloid+ is ostensibly assuming that amyloid is the precursor to the AD cascade, we feel that including many subjects with a low likelihood of AD pathology may lead to distorted trajectories. The *APOE* ε4 allele is the major genetic risk factor for sporadic AD, although roughly one-third of individuals with AD do not carry it.

Our nonparametric approach does not assume sigmoidal curves, but rather a very flexible class of monotone curves. Surprisingly, among amyloid+ subjects, we found mean CSF Aβ follows a linear trajectory, whereas tau, p-tau, and PiB PET follow sigmoidal shapes. However, the sigmoidal shapes are flatter than those proposed by Jack and colleagues [5] and remain within the 40th to 80th percentile range. Glucose metabolism (fluorodeoxyglucose PET), hippocampal volume, ventricular volume, learning, and cognition (ADAS13) all track very close to each other in near-linear trajectories. Function (Alzheimer's Disease Cooperative Study Functional Activities Questionnaire) was the final domain to

fail following a parabolic trajectory. It is quite possible that ADNI does not have enough data from later stage dementia, which might demonstrate the final plateau of a sigmoid. The relative paucity of available observations at the most severe stage of disease is a limitation that will be addressed as the model is expanded to include additional data sets.

The question of which markers become abnormal first is distinct from the question of which markers can be estimated efficiently in terms of the signal-to-noise ratio. To explore the latter question, we provide plots of the first derivatives of curves divided by the residual standard deviation. Hippocampal volume appears to dominate the other measures across the 15-year span in both analyses, with the possible exception of CSF markers. The CSF markers show some areas of relatively high standardized slopes, but these could be a result of scant data and spurious acceleration near the boundaries of observation (Fig. 5, bottom). In other cases, the CSF measures are relatively flat, which may cause spurious acceleration depicted in the bell shapes (Fig. 4A, bottom).

Our approach also does not assume that the mean should attain zero and one. Without this assumption, our algorithm demonstrates much pathological heterogeneity or measurement variability, even in the selected amyloid+ subset. For instance, 15 years before reaching the worst 20th percentile of CDRSB, CSF Aβ ranges between zero and the 80th percentile, with a mean at about the 20th percentile. Between-subject variability tends to flatten the mean trajectory, such that most estimated trajectories in Fig. 4A do not cover the full range from zero to one. A model which forces the mean value to attain zero may mask the heterogeneity of some markers over the course of the disease.

Perhaps some of the heterogeneity can be explained by diet, lifestyle, education, occupation, or other covariates related to cognitive reserve. Genetics or family history might also explain heterogeneity. We plan to investigate these hypotheses in the future by building covariates into the model, but more data on the earliest phases of the disease are necessary. Fortunately, the mixed-model framework we have adopted is well suited to pooling data sets for meta-analyses. Hierarchical random effects can be used to model within-study and within-subject correlation. Meta-analyses may also help address a key limitation of the ADNI data, which is that the age range of ADNI participants is restricted to 55 to 95 years at baseline. In fact, across our 15-year span of estimated long-term progression, the mean age of subjects represented remains in the 70- to 75-year range. Clearly, we need to incorporate data from younger cohorts.

The comparison groups depicted in Fig. 5 are difficult to interpret. Note that only 13 of 130 subjects with AD (10%) with known amyloid status are classified as amyloid−, and 75 of 220 subjects with AD (34%) have no *APOE* ε4 allele. Also, many of the subjects with mild or no impairment entered into this analysis may never progress. However, we might interpret the amyloid− trajectories as a
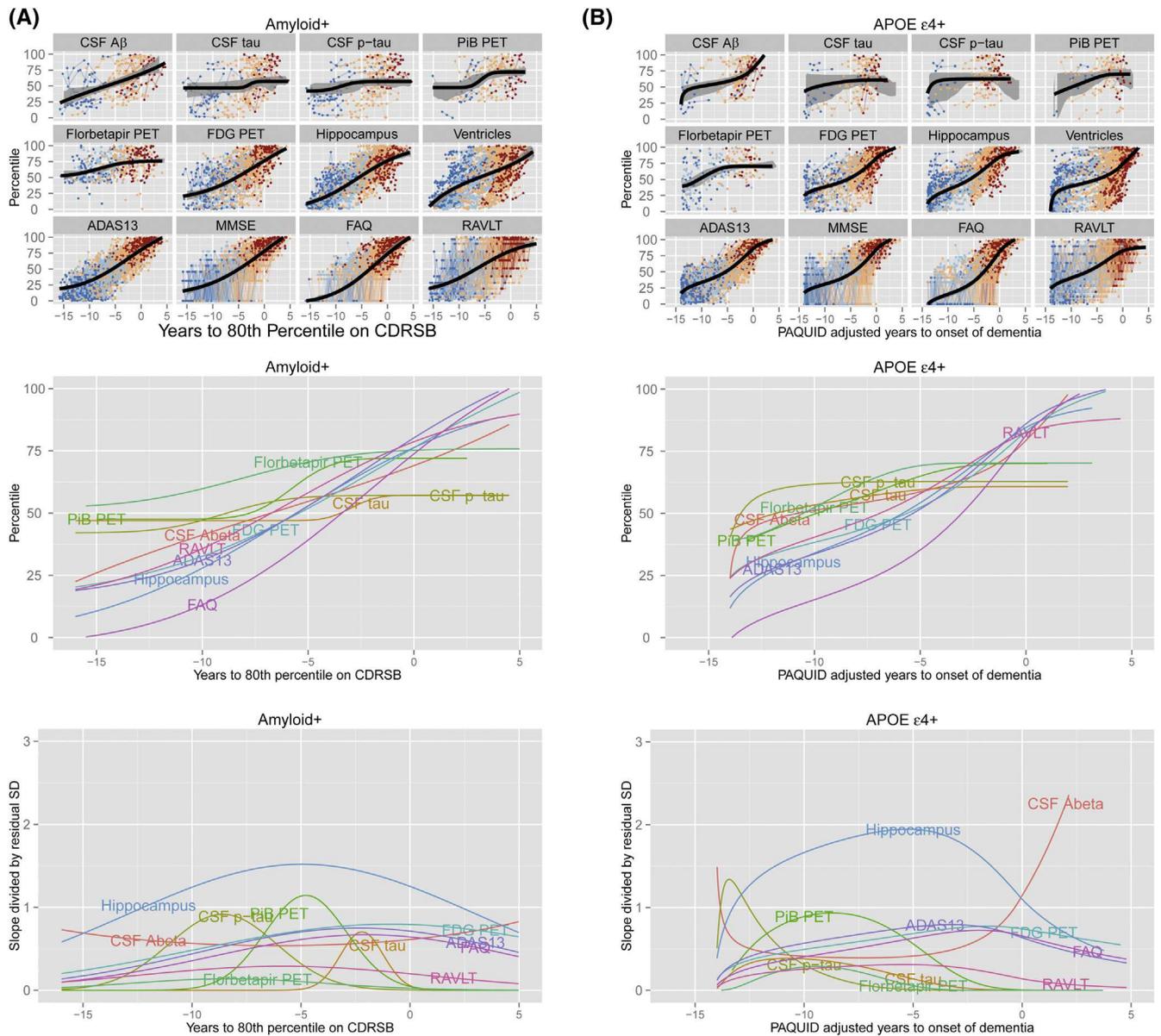
Fig. 4. (A) Alzheimer's Disease Neuroimaging Initiative (ADNI) amyloid+ subjects. ADNI apolipoprotein E (*APOE*) ε4 allele carriers. The top panels show each of the mean trajectories superimposed over the subject-level observations from 579 amyloid+ and 570 *APOE* ε4 individuals, colored by diagnosis. Colors in the top panel represent diagnosis at ADNI baseline— cognitively normal (CN) in dark blue, early mild cognitive impairment (EMCI) in light blue, late mild cognitive impairment (LMCI) in light red, and Alzheimer's disease (AD) in dark red. Shaded gray regions, where visible in the top panels, represent bootstrap 95% confidence bands. The middle panels show all the trajectories at once. On the left, time has been shifted so that time zero represents the time at which mean Clinical Dementia Rating Scale Sum of Boxes (CDRSB) trajectory (not shown) meets the 80th percentile. On the right, time has been adjusted using long-term "Personnes Agées Quid" (PAQUID) Mini-Mental State Examination trajectories so that time zero represents the estimated time to onset of dementia. The bottom panels show rates of change standardized by residual standard deviation (SD). CSF, cerebrospinal fluid; Aβ, amyloid-β; p-tau, phosphorylated tau; PiB, Pittsburgh compound B; PET, positron emission tomography; FDG, fluorodeoxyglucose; ADAS13, the 13-item Alzheimer's Disease Assessment Scale–Cognitive Subscale; MMSE, Mini-Mental State Examination; FAQ, Alzheimer's Disease Cooperative Study Functional Activities Questionnaire; RAVLT, Rey Auditory Visual Learning Test; CN, cognitively normal; LMCI, late mild cognitive impairment; AD, Alzheimer's disease.

representation of non-AD pathology, which is marked by a divergent biomarker signature, including normal CSF levels and less pronounced hippocampal atrophy and ventricular expansion. In contrast, the *APOE* ε4 noncarrier group appears to converge toward the *APOE* ε4 carrier group as symptoms progress. This apparent convergence is possibly a result of the concatenation of subjects without AD pathol-

ogy antecedent to those with AD pathology, rather than a true acceleration of pathology in *APOE* ε4 noncarriers.

Our analysis suggests that amyloid PET imaging with florbetapir or PiB may reach abnormal levels first, followed by CSF tau and p-tau. This is consistent with the view that PET imaging is the most direct measure of amyloid accumulation in the brain (generally considered to be the inciting
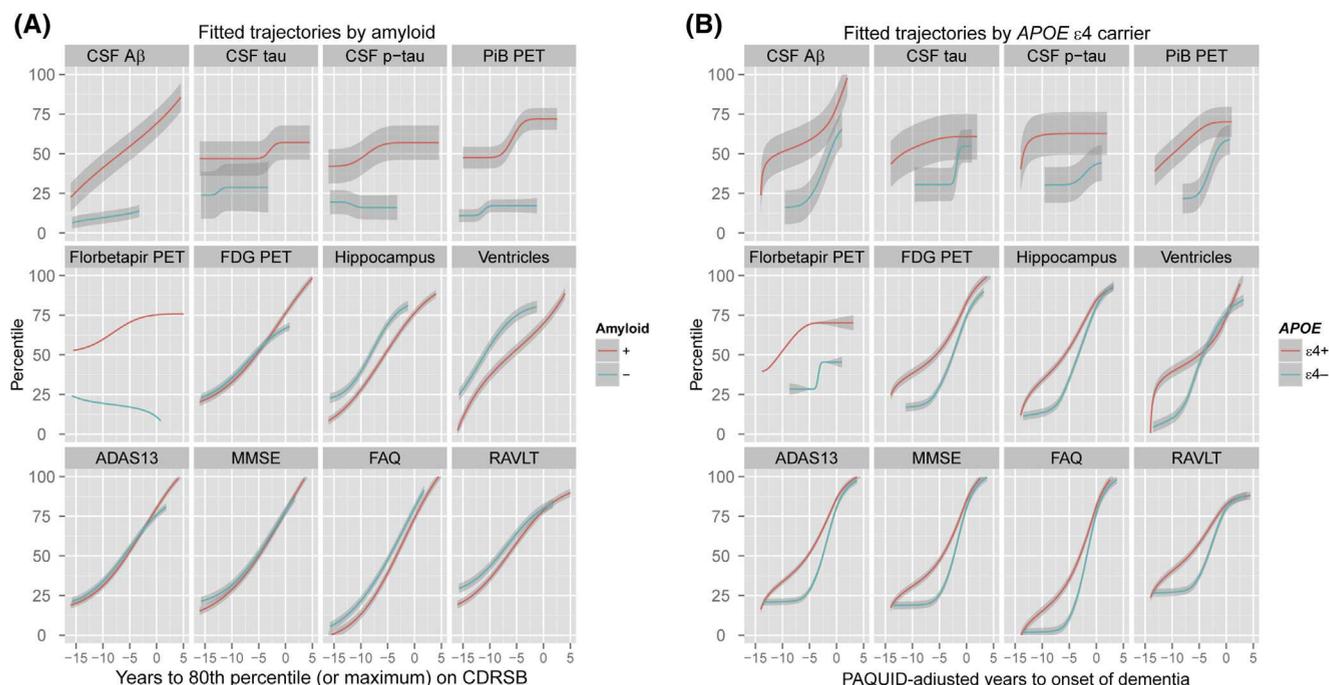
Fig. 5. polipoprotein E (*APOE*) ε4 allele carriers vs. noncarriers. Each panel repeats the same estimated trajectories in the amyloid+ and *APOE* ε4 carrier groups, as in Fig. 4, and includes the amyloid– and *APOE* ε4 noncarrier comparison groups. The amyloid– group consisted of 190 cognitively normal (CN) individuals, 153 individuals with early mild cognitive impairment (EMCI), 92 individuals with late mild cognitive impairment (LMCI), and 13 individuals with Alzheimer's disease (AD). The *APOE* ε4 allele noncarrier group consisted of 263 CN individuals, 124 individuals with EMCI, 219 individuals with LMCI, and 75 individuals with AD. Shaded gray regions represent 95% confidence bands derived analytically, rather than by the bootstrap as in Fig. 4. CSF, cerebrospinal fluid; Aβ, amyloid-β; p-tau, phosphorylated tau; PiB, Pittsburgh compound B; PET, positron emission tomography; FDG, fluorodeoxyglucose; ADAS13, the 13-item Alzheimer's Disease Assessment Scale–Cognitive Subscale; MMSE, Mini-Mental State Examination; FAQ, Alzheimer's Disease Cooperative Study Functional Activities Questionnaire; RAVLT, Rey Auditory Visual Learning Test; PAQUID, "Personnes Agées Quid."

event in AD), and suggests a delay before abnormalities are observed in "downstream" markers of neurodegeneration in the CSF. Learning, glucose metabolism, hippocampal atrophy, and cognition all follow in close succession. Function is the last domain to progress to abnormality, as expected. Plots of the adjusted slopes indicate that hippocampal volume assessed by structural MRI provides the most efficient measure of disease progression across the full span. These observations are consistent with our current understanding of the disease and paint a picture in general agreement with the model [5]. Our approach will facilitate analyses using diverse data sets with overlapping measures, providing a framework for validating models of disease progression.

Last, our framework provides an approach to assessing the growing body of outcome data, providing quantitative data to inform still-hypothetical biomarker models. As noted in an editorial accompanying the revised model [6,29], biomarker modeling will be facilitated by ongoing accrual of data that reduces the gaps in our observations; this applies to both hypothetical and data-driven efforts.

## Acknowledgments

## References

[1] Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. Biometrics 1990;46:673–87.

[2] Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: linear and nonlinear mixed effects models. R package version 3.1-106. Vienna: R Foundation for Statistical Computing; 2012.

[3] Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. New York: Springer Verlag; 2000.

[4] Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 2010;3:201–9.

[5] Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.

[6] Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner WM, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013;12:207–16.

[7] Kneip A, Gasser T. Convergence and consistency results for self-modeling nonlinear regression. Ann Stat 1988;16:82–112.

[8] Lawton WH, Sylvestre EA, Maggio MS. Self modeling nonlinear regression. Technometrics 1972;14:513–32.

[9] Ladd WM, Lindstrom MJ. Self-modeling for two-dimensional response curves. Biometrics 2000;56:89–97.

[10] Brumback LC, Lindstrom MJ. Self modeling with flexible, random time transformations. Biometrics 2004;60:461–70.

[11] Beath KJ. Infant growth modelling using a shape invariant model with random effects. Stat Med 2007;26:2547–64.

[12] Lindstrom MJ. Self-modeling with random shift and scale parameters and a free-knot spline shape function. Stat Med 1995;14:2009–21.

[13] Wang Y, Ke C, Brown MB. Shape-invariant modeling of circadian rhythms with random effects and smoothing spline ANOVA decompositions. Biometrics 2003;59:804–12.

[14] Altman N, Villarreal J. Self-modelling regression for longitudinal data with time-invariant covariates. Can J Stat 2004;32:251–68.

[15] Coull BA, Staudenmayer J. Self-modeling regression for multivariate curve data. Statistica Sinica 2004;14:695–711.

[16] Telesca D, Inoue LYT. Bayesian hierarchical curve registration. J Am Stat Assoc 2008;103:328–39.

[17] Dartigues JF, Gagnon M, Barberger-Gateau P, Letenneur L, Commenges D, Sauvel C, et al. The PAQUID epidemiological program on brain aging. Neuroepidemiology 1992;11:8.

[18] Amieva H, Le Goff M, Millet X, Orgogozo JM, Peres K, Barberger-Gateau P, et al. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. Ann Neurol 2008;64:492–8.

[19] Hastie T, Tibshirani JR. Generalized additive models. Stat Sci 1986; 1:297–310.

[20] Ramsay JO. Estimating smooth monotone functions. J R Stat Soc B Stat Method 1998;60:365–75.

[21] Ramsay JO, Wickham H, Graves S, Hooker G. fda: functional data analysis. R package version 2.2.8. Vienna: R Foundation for Statistical Computing; 2012.

[22] Bates D, Maechler M, Bolker B. lme4: linear mixed-effects models using S4 classes. R package version 0.999999-0. Vienna: R Foundation for Statistical Computing; 2012.

[23] Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2009.

[24] Donohue MC. Growth models by alternating conditional expectation (GRACE). R package version 1.0. 2013. San Diego. Available at: http://mdonohue.bitbucket.org/grace/. Accessed February 24, 2014.

[25] Shaw LM, Knapik-Czajka M, Lee VMY, Trojanowski JQ, Vander-Stichele H, Clark CM, et al. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. Ann Neurol 2009;65:403–13.

[26] Jack CR, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, et al. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 2008;131:665–80.

[27] Jagust W, Landau S, Shaw L, Trojanowski J, Koeppe R, Reiman E, et al. Relationships between biomarkers in aging and dementia. Neurology 2009;73:1193–9.

[28] Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 2012;367:795–804.

[29] Frisoni GB, Blennow K. Biomarkers for Alzheimer's: the sequel of an original model. Lancet Neurol 2013;12:126–8.