Submitted to the Annals of Applied Statistics

BIOMARKER CHANGE POINT ESTIMATION WITH RIGHT CENSORING IN LONGITUDINAL STUDIES*

By Xiaoying Tang†, Michael I Miller‡ and Laurent Younes‡

Sun Yat-sen University† and Johns Hopkins University‡

We consider in this paper a statistical two-phase regression model in which the change point of a disease biomarker is measured relative to another point in time, such as the manifestation of the disease, which is subject to right-censoring (i.e., possibly unobserved over the entire course of the study). We develop point estimation methods for this model, based on maximum likelihood, and bootstrap validation methods. The effectiveness of our approach is illustrated by numerical simulations, and by the estimation of a change point for amygdalar atrophy in the context of Alzheimer's disease, wherein it is related to the cognitive manifestation of the disease.

1. Introduction. The manifestation of an event, such as the onset of a disease, is not always immediate and often requires some time for its repercussions to become observable. Slowly progressing diseases, and in particular neuro-degenerative disorders such as Alzheimer's disease (AD) which is a focus of the current paper, fall into this category. The manifestation of such diseases is related to the onset of cognitive or functional impairment and, at the time when this occurs, the disease may have already had been affecting the brain anatomically and functionally for a considerable time. Such effects, however, are only observable through costly and sometimes even invasive medical examinations, which are not routinely performed on healthy, or apparently healthy populations.

It is however extremely important scientifically and clinically to determine how the disease evolves and the time when brain change begins, especially when the disease's pathology is (currently) irreversible like that of AD. The goal of this paper is to propose and analyze a statistical model

*This work was partially supported by the National Natural Science Foundation of China (81501546) , the SYSU-CMU Shunde International Joint Research Institute Start-up Grant (20150306), the National Institutes of Health (U01-AG03365, P50-AG005146, R01-EB017638, R01-NS084957, and P41-EB015909), the Office of Naval Research (ONR-90048203), and the National Science Foundation (ACI-1053575) for the Computational Anatomy Gateway via Extreme Science and Engineering Discovery Environment (XSEDE) and the Kavli founation via the Kavli Neuroscience Discovery Institute.

MSC 2010 subject classifications: Primary 62G10; secondary 62G09

Keywords and phrases: change point estimation, right censoring, medical imaging

that addresses this issue by determining a change point, at the population level, at which the evolution of a given biomarker develops a regime change. Assuming the measurements of these biomarkers are taken from a dataset including asymptomatic subjects, with a subsequent determination of the disease manifestation made at a later time, we will describe the associated estimation procedure and provide numerical experiments on both simulated and real data.

It is worthy of note that performing such disease-related observations on human beings is a very challenging process. It requires following subjects regularly over many years, starting at a time point when they have not manifested any sign of the disease yet, with an uncertainty over how many subjects will have converted to diseased status within the time frame of the study. While some genetic or family history information can be used to increase the likelihood of observing disease manifestation, the difficulty of this data acquisition task explains why datasets of this kind are still relatively rare today. One may expect, however, that systematic patient monitoring and computerized medical recording will lead to more such datasets emerging in the future.

The real data that we will use in this paper are provided by the BIOCARD study, which is a longitudinal study of AD in which subjects have been continuously followed for more than 20 years. BIOCARD has, compared to other longitudinal studies on AD such as ADNI (Mueller et al., 2005), the distinction of having only included individuals who had no sign of cognitive impairment at baseline. The BIOCARD study was initiated in 1995 with the subjects having received their most recent cognitive assessment in 2012, resulting in about one fourth of the group being diagnosed with mild cognitive impairment (MCI) or AD. This dataset motivates the development of the statistical model and parameter estimation method presented in this paper, and more detailed information about it will be provided in section 4.2.

We will describe the basic assumptions of our model and our notation in section 2. Section 3 will describe the parameter estimation procedure. Section 4 will provide experimental results, both on simulated data and on the BIOCARD dataset.

2. Statistical Model and Notation. We will let Y denote the dependent variable, which, in our application, will be associated to the measurement of a biomarker for AD. The value of this variable is assumed to depend on time (which, in this paper, will always be the subject's age), disease status, and possibly other covariates (e.g., gender, intracranial vol-

ume, etc.). The general model assumes that the disease status results in a change point in the evolution of the biomarker, at a time which is indirectly observable through the induced external manifestation (e.g., cognitive impairment) which happens after a delay, Δ, from the change point time. We will assume that Δ is a fixed parameter specific to the biomarker.

Let the random variable U denote the time of disease manifestation itself (and thus the biomarker's change point occurs at time $U-\Delta$). We will refer to U as the "manifest onset time". We will assume that it is always finite, in the sense that everyone in the considered population would eventually develop the disease if they were to live indefinitely. This manifest onset time is not always observable, since patients may still be healthy – or have an undetected onset – at the end of the study, resulting in right censoring. Some form of left censoring is also possible, depending on how subjects participating in the study were selected.

We let n denote the total number of subjects involved in the study, each with their own manifest onset times $U_1, \ldots, U_n$. We let $T_1, \ldots, T_n$ be the ages of the subjects at the end of the study, which are the right censoring times, so that U_k is observable only when it is less than the corresponding T_k. A value y_{kj} of the biomarker is measured for subject k at age t_{kj}, with $t_{k1} \leq \cdots \leq t_{kp_k} \leq T_k$ (p_k therefore denotes the total number of longitudinal observations for subject k). We will work with a model assuming a linear dependence of Y with respect to age and onset time, with a rate change at $t = U - \Delta$, in the form

$$Y(t) = a + b_1 t + b_2 U + c(t - U + \Delta)^+ + \eta + \varepsilon(t), \tag{1}$$

where $(x)^+ = \max(x, 0)$. Here, a, b_1, b_2, c and Δ are parameters, $\eta \sim \mathcal{N}(0, \rho^2)$ models a time-independent random effect, and $\varepsilon(t) \sim \mathcal{N}(0, \sigma^2)$ is a noise variable modeling the longitudinal intrasubject variation. In our model, η and all $\varepsilon(t)$ are mutually independent, and all are independent of U. Note that the well-posedness of the model requires some limits on the values of Δ; if $\Delta \ll 0$, then $(t-U+\Delta)^+$ is almost always zero, and thus c and Δ become barely identifiable; conversely, if $\Delta \gg 0$, then $(t-U+\Delta)^+ \simeq t-U+\Delta$ and the model becomes over-parametrized. We will return to these issues later.

Assuming n independent realizations of this model, the observations are: $y_{kj} = Y(t_{kj})$, for $k = 1, \ldots, n$ and $j = 1, \ldots p_k$; the end-of-study ages, T_k; and the censored manifest onset times, $Z_k = \min(U_k, T_k)$. We will assume that all T_k's are deterministic, or, equivalently, that they are independent of other variables, and work conditionally to them. The final piece of the model is the distribution of the manifest onset time, U. We will use either a Gaussian or an exponentially-modified Gaussian distribution (see Section

3.1).

Our model is therefore a two-phase (or segmented) regression model with right-censoring on the time variable. Parametric inference and testing for multi-phase regression were studied in Quandt (1958); Sprent (1961); Hudson (1966); Hinkley (1969); Farley and Hinich (1970); Hinkley (1971); Feder (1975); Gombay and Horváth (1994b,a). Change point models have also been introduced for survival analysis and hazard estimation, especially in the context of right censoring (Nguyen et al., 1984; Pons, 2003; Wu et al., 2003; Dupuy, 2006; Li et al., 2013) (see also Chen and Gupta (2012) and references cited there).

Note that, even though our data includes right censoring, we are using a linear regression model rather than a hazard- or Cox- regression model. The model in (1) provides the correct "causal" relationship in which the change point $U - \Delta$ triggers a change of regime in the dependent variable Y.

3. Parameter Estimation.

3.1. *Manifest Onset Time: Prior Model.* We here describe how the parameters of the manifest onset time distribution can be estimated from data under various hypotheses of right and left censoring. We consider it as a prior distribution, and therefore assume that it is estimated separately from the other model parameters. This is justified by the fact that, while it would be possible to estimate the joint distribution of (Y, U) for any given biomarker, it is generally preferable to work with a single model of U shared by all biomarker variables Y. Moreover, manifest onset time information for medical data is usually available for more diverse and larger datasets than those on which biomarkers are measured, and this information can naturally be used to estimate their distribution.

In typical studies, one can generally separate the subjects into three groups: the subjects who converted during the study, that we will denote J_0, the subjects who converted after study end (right censored), denoted J_1 and those who entered the study with the disease (left censored), denoted J_2. Some study designs (such as BIOCARD) focused on incident disease are restricted to disease-free cohorts, thus eliminating the last group.

Assuming right censoring only ($J_2 = \emptyset$), and letting as before T_k be the age at the end of the study, $J_1 \subset \{1, \dots, n\}$ is the subset of subject indices for which $U_k \geq T_k$ and $J_0 = \{1, \dots n\} \setminus J_1$. Denote by $f_U(u|\theta)$ the p.d.f. of the variable U (for a given parameter θ) and by $F_U(u|\theta)$ the corresponding c.d.f. The log-likelihood of the observed data, which is $(\min(U_k, T_k), k = 1, \dots, n)$,

is

$$\ell(\theta) = \sum_{k\in J_0} \log f_U(u_k|\theta) + \sum_{k\in J_1} \log(1 - F_U(T_k|\theta)).$$

For studies in which diseased subjects are not enrolled by design, the likelihood must be modified to account for this bias, which requires that $U_k \geq t_{k1}$ for all k. Taking the likelihood conditional on this event, we obtain

$$\ell(\theta) = \sum_{k\in J_0} \log f_U(u_k|\theta) + \sum_{k\in J_1} \log(1 - F_U(T_k|\theta)) - \sum_{k=1}^{n} \log(1 - F_U(t_{k1}|\theta)). \tag{2}$$

Finally, if diseased subjects can be included in the study, yielding a non-empty set J_2, the resulting likelihood is

$$\ell(\theta) = \sum_{k\in J_0} \log f_U(u_k|\theta) + \sum_{k\in J_1} \log(1 - F_U(T_k)|\theta) + \sum_{k\in J_2} \log F_U(t_{k1}|\theta) - \sum_{k\in J_0\cup J_1} \log(1 - F_U(t_{k1}|\theta)). \tag{3}$$

First consider the case in which f_U is a Gaussian distribution $\mathcal{N}(m_1, \sigma_1^2)$, which is probably the simplest choice. In this case, the log-likelihood is

$$\ell(m_1, \sigma_1^2) = -\frac{|J_0|}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k\in J_0}(u_k - m_2)^2 + \sum_{k\in J_1} \log\left(1 - \Phi\left(\frac{T_k - m_1}{\sigma}\right)\right) + \sum_{k\in J_2} \log\Phi\left(\frac{t_{k1} - m_1}{\sigma}\right) - \sum_{k\in J_0\cup J_1} \log\left(1 - \Phi\left(\frac{t_{k1} - m_1}{\sigma}\right)\right) \tag{4}$$

and its gradient can be easily computed.

One may however prefer using a distribution with heavier tails, allowing for large values of the variable to occasionally occur. Such a behavior may be important to allow for healthy controls to have a manifest onset time so far in the future that they do not enter the second phase of the regression model during the study time. We will present simulations using an exponentially-modified Gaussian, which can be written as $U = W + S$, where $W \sim \mathcal{N}(m_1, \sigma_1^2)$ and $S \sim \exp(\alpha)$ (an exponential variable with mean

α). Its p.d.f. is the convolution of the Gaussian and exponential densities, and is given by

$$\begin{aligned} f_U(u|m_1,\sigma_1^2,\alpha) &= \frac{1}{\alpha\sqrt{2\pi\sigma_1^2}}\int_{-\infty}^{u}\exp\left(-\frac{(w-m_1)^2}{2\sigma_1^2}-\frac{u-w}{\alpha}\right)dw \\ &= \frac{1}{\alpha}\exp\left(-\frac{u-m_1}{\alpha}+\frac{\sigma_1^2}{2\alpha^2}\right)\Phi\left(\frac{u-m_1}{\sigma_1}-\frac{\sigma_1}{\alpha}\right) \end{aligned} \tag{5}$$

where Φ is the cumulative distribution function (c.d.f.) of a standard Gaussian variable. Similarly, the c.d.f. of U is

$$F_U(u|m_1,\sigma_1^2,\alpha) = P(U\le u) = \Phi\left(\frac{u-m_1}{\sigma_1}\right) - \exp\left(-\frac{u-m_1}{\alpha}+\frac{\sigma_1^2}{2\alpha^2}\right)\Phi\left(\frac{u-m_1}{\sigma_1}-\frac{\sigma_1}{\alpha}\right).$$

These expressions can be plugged into (3), and the gradient of the resulting log-likelihood with respect to each of the three parameters m_1, σ_1 and α can also be computed analytically. Our implementation uses the Matlab optimization toolbox to compute the maximum likelihood estimators of those three parameters.

This exponentially-modified Gaussian model obviously reduces to the Gaussian one when $\alpha = 0$, and we used a log-likelihood ratio test in order to assess the validity of the hypothesis $\alpha > 0$.

Note that we should in principle have conditioned our prior distribution to take only positive values. However, the models estimated in our applications are such that the probability of taking a negative value is negligible (less than 10^{-10}) and such a modification was not necessary. If implemented, this modification would only have affected the case $J_2 \neq \emptyset$, since the likelihood for subjects in J_0 and J_1 is already left-censored by $U > t_{k,1}$, and would have resulted in the addition of the term

$$-\sum_{k\in J_2}\log(1-F_U(0|\theta))$$

to the log-likelihood in (3).

Illustration. We first provide an application using simulations based on population data relative to Alzheimer's disease. The prevalence of Alzheimer's disease over various age groups was published in Hebert et al. (2013). From this source, prevalence is about 3% in the 65-74 age group, 17.6% in the 75-84 age group and 30% in the 85-94 age group. Similarly, data in Alzheimer's

Association (2015) indicate that prevalence among people above 95 years may be as high as 50%. Based on results such as those provided in Larrieu et al. (2002), one may add about 5% to these numbers to also include the fraction of population with mild cognitive impairment, a precursor state to Alzheimer's. Using this information, it is easy to derive a logistic regression model that provides the conditional probability of disease conditional to age.

We can then use this information combined with census data to simulate a large-scale sample of population at various ages and their disease status. For such a sample, which is purely cross-sectional (i.e., does not contain any longitudinal information), one can estimate an exponentially modified Gaussian distribution with (using the previous notation) $u_k = T_k = t_{k1}$, $J_0 = \emptyset$, J_1 being the set of healthy subjects and J_2 the set of disease subjects. Doing so, the obtained parameters provide a Gaussian term with a very large variance ($m_1 \simeq 95$, $\sigma_1 \simeq 20$) and a small exponential term ($\alpha \simeq 0.1$) which is not significant (implying that a Gaussian model can be used). The values that we used in our experiments were slightly different ($m_1 = 93, \sigma_1 = 14$), because the BIOCARD dataset is slightly bisased, in the sense that it enrolled a majority of patients with a family history of AD. We estimated these parameters from another dataset, with enrollment procedures similar to BIOCARD.

As a further illustration, Tables 1 to 3 provide simulation results for various parameters of an exponentially-modified Gaussian model, with 1,000 subjects initially generated with ages normally distributed with mean 57 and standard deviation 10, and true model parameters m_1, σ_1 and α. We assumed both left and right censoring: only healthy subjects were kept at the beginning of the "study", with a study length being 15 years (so that $T - t_1 = 15$). We kept $\alpha = 2$ years and ran simulations with $m_1 = 55, 65, 75$ years and $\sigma_1 = 1, 2, 3, 5, 10$ and 15 years.

From these results, we note that α becomes almost impossible to separate from 0 when σ_1 increases, inducing biases in the combined estimation of m_1 and α. For smaller values of σ_1, the bias is small and the likelihood ratio test has high power; this is mostly independent of the proportion of left censored data. For large standard deviation, the null hypothesis is almost always accepted (about 98% of the time). The estimations of the mean $(m_1 + \alpha)$ and standard deviation $(\sqrt{\sigma_1^2 + \alpha^2})$ remain relatively accurate via the Gaussian submodel, even when the bias and variance of the full model parameter estimates increase.

True parameters					Bias			Standard dev.			Power
m_1	σ_1	α	n	$\lvert J_1\rvert$	m_1	σ_1	α	m_1	σ_1	α	(in %)
55	1	2	384	40	0.008	-0.005	-0.006	0.150	0.122	0.165	100.0
(57)	(2.2)	(0)			(-0.100)	(0.228)		(0.122)	(0.159)		
55	2	2	386	42	0.023	-0.001	-0.022	0.265	0.194	0.245	98.7
(57)	(2.8)	(0)			(-0.100)	(0.187)		(0.170)	(0.163)		
55	3	2	388	45	0.076	0.004	-0.078	0.474	0.286	0.444	67.2
(57)	(3.6)	(0)			(-0.102)	(0.149)		(0.241)	(0.184)		
55	5	2	395	57	0.400	-0.003	-0.392	1.076	0.457	1.049	11.1
(57)	(5.4)	(0)			(-0.094)	(0.094)		(0.428)	(0.274)		
55	10	2	417	104	0.090	-0.455	0.108	2.167	1.101	2.162	3.0
(57)	(10.2)	(0)			(-0.124)	(0.059)		(1.233)	(0.695)		
55	15	2	434	160	-0.942	-1.291	1.577	3.632	2.335	3.781	3.0
(57)	(15.1)	(0)			(-0.335)	(0.112)		(2.658)	(1.443)		

TABLE 1

Parametric estimation of the onset time model averaged over 10,000 simulations. Here n is the number of subjects after left censoring (out of 1,000), and $\lvert J_1\rvert$ is the average number of right-censored objects (healthy 15 years after enrollment). Numbers between parentheses provide information on the estimation of a Gaussian model with same mean and variance. The power is computed as the fraction of likelihood ratios relative to the submodel $\alpha = 0$ who were larger than 3.84.

True parameters					Bias			Standard dev.			Power
m_1	σ_1	α	n	$\lvert J_1\rvert$	m_1	σ_1	α	m_1	σ_1	α	(in %)
65	1	2	752	217	0.007	-0.001	-0.007	0.138	0.100	0.157	99.8
(67)	(2.2)	(0)			(-0.034)	(-0.016)		(0.093)	(0.105)		
65	2	2	749	220	0.018	-0.000	-0.017	0.218	0.142	0.221	99.6
(67)	(2.8)	(0)			(-0.034)	(-0.008)		(0.122)	(0.109)		
65	3	2	745	225	0.071	0.009	-0.069	0.421	0.211	0.416	72.0
(67)	(3.6)	(0)			(-0.033)	(-0.003)		(0.160)	(0.127)		
65	5	2	731	240	0.359	0.004	-0.353	1.070	0.362	1.073	11.5
(67)	(5.4)	(0)			(-0.032)	(0.004)		(0.259)	(0.198)		
65	10	2	688	287	-0.181	-0.371	0.279	2.289	0.884	2.370	3.0
(67)	(10.2)	(0)			(-0.052)	(0.017)		(0.646)	(0.527)		
65	15	2	650	329	-1.569	-1.107	1.980	3.888	1.929	4.310	2.8
(67)	(15.1)	(0)			(-0.129)	(0.057)		(1.392)	(1.122)		

TABLE 2

Same as Table 1, with $m_1 = 65$ years.

True parameters					Bias			Standard dev.			Power
m_1	σ_1	α	n	$\lvert J_1\rvert$	m_1	σ_1	α	m_1	σ_1	α	(in %)
75	1	2	952	576	0.005	-0.005	-0.005	0.138	0.096	0.178	100.0
(77)	(2.2)	(0)			(-0.033)	(-0.201)		(0.110)	(0.108)		
75	2	2	949	576	0.026	-0.002	-0.024	0.285	0.158	0.309	94.8
(77)	(2.8)	(0)			(-0.032)	(-0.151)		(0.137)	(0.112)		
75	3	2	945	574	0.120	0.010	-0.120	0.602	0.233	0.623	48.2
(77)	(3.6)	(0)			(-0.034)	(-0.112)		(0.172)	(0.134)		
75	5	2	933	569	0.336	-0.023	-0.322	1.252	0.368	1.284	8.0
(77)	(5.4)	(0)			(-0.027)	(-0.064)		(0.256)	(0.203)		
75	10	2	883	556	-0.639	-0.423	0.794	2.710	0.864	2.899	2.8
(77)	(10.2)	(0)			(-0.014)	(-0.016)		(0.511)	(0.496)		
75	15	2	826	544	-2.353	-1.133	2.857	4.665	1.846	5.326	2.8
(77)	(15.1)	(0)			(-0.037)	(0.015)		(0.886)	(1.026)		

TABLE 3
Same as Table 1, with $m_1 = 75$ years.

3.2. *Change point Onset Model.* We now describe the estimation of the parameters a, b_1, b_2, c, ρ^2 and σ^2 which affect the change point onset model (1), assuming that m_1, σ_1^2 and α are fixed. The joint p.d.f. of the model variables is $\prod_k f(y_k, u_k, \eta_k)$ where $y_k = (y_{k1}, \ldots, y_{kp_k})$, with

$$f(y_k, u_k, \eta_k) = \frac{f_U(u_k)}{(2\pi)^{(p_k+1)/2}\rho\sigma^{p_k}}$$
$$\exp\left(-\frac{\eta_k^2}{2\rho^2} - \frac{1}{2\sigma^2}\sum_{j=1}^{p_k}\left(y_{kj} - a - b_1 t_{kj} - b_2 u_k - c(t_{kj} + \Delta - u_k)^+ - \eta_k\right)^2\right),$$

and $f_U(u) = \frac{1}{\alpha}\exp(-\frac{u}{\alpha} + \frac{m_1}{\alpha} + \frac{\sigma_1^2}{2\alpha^2})\Phi(\frac{u-m_1}{\sigma_1} - \frac{\sigma_1}{\alpha})$, or simply $\varphi((u-m_1)/\sigma_1)/\sigma_1$ if $\alpha = 0$, where φ and Φ are the p.d.f. and c.d.f. of a standard Gaussian variable. Because we assume that m_1, σ_1 and α are fixed, we omit the left-censoring normalization which only depends on them and therefore will not impact the observation process.

The observations are either (y_k, u_k) for $k \in J_0$, or just y_k and the additional information that $u_k \geq T_k$ (resp. $u_k \leq t_{k1}$) for $k \in J_1$ (resp. $k \in J_2$). The likelihood of the observed data is therefore given by

$$L(\theta) = \prod_{k\in J_0} f_{Y,U}(y_k, u_k) \prod_{k\in J_1} f_Y(y_k | U_k \geq T_k) \prod_{k\in J_2} f_Y(y_k | U_k \leq t_{k,1})$$

where $f_{Y,U}$ is the marginal density of Y and U in the first product, and the p.d.f.'s in the next two products are conditional densities of Y given the relevant event for U. Note that

$$f_Y(y_k|U_k \geq T_k) = \frac{1}{P(U_k \geq T_k)} \int_{T_k}^{\infty} f_{Y,U}(y_k, u)du.$$

The denominator was computed in the previous section. It depends on only the fixed distribution of U, and can therefore be treated as a constant. Using the notation

$$f_{Y,U}(y; T, S) = \int_T^S f_{Y,U}(y, u)du,$$

we therefore need to maximize

$$\tilde{L}(\theta) = \prod_{k \in J_0} f_{Y,U}(y_k, u_k) \prod_{k \in J_1} f_{Y,U}(y_k; T_k, +\infty) \prod_{k \in J_2} f_{Y,U}(y_k; -\infty, t_{k1}).$$

We now provide the expression of the marginal densities in the product. For simplicity, we will drop the index k in the rest of the computation, therefore letting $y = (y_1, \ldots, y_p)$ and $t = (t_1, \ldots, t_p)$.

We start with

$$f_{Y,U}(y, u) = \int_{-\infty}^{\infty} f(y, u, \eta)d\eta.$$

Using $r_j = y_j - a - b_1 t_j - b_2 u - c(t_j + \Delta - u)^+$, we have

$$\begin{aligned} & f_{Y,U}(y, u) \\ &= \int_{-\infty}^{+\infty} \frac{f_U(u)}{\sqrt{2\pi}^{p+1}\rho\sigma^p} \exp\left(-\Big(\frac{1}{2\rho^2} + \frac{p}{2\sigma^2}\Big)\eta^2 + \frac{2}{2\sigma^2}\eta \sum_{j=1}^p r_j - \frac{1}{2\sigma^2}\sum_{j=1}^p r_j^2 \right) d\eta \\ &= \frac{\tau_p f_U(u)}{\sqrt{2\pi}^p \rho\sigma^{p-1}} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^p r_j^2 + \frac{\tau_p^2}{2\sigma^2}\left(\sum_{j=1}^p r_j\right)^2 \right). \end{aligned}$$

with

$$\tau_p^2 = \frac{\rho^2}{\sigma^2 + p\rho^2} = \frac{1}{p} - \frac{\sigma^2}{p(\sigma^2 + p\rho^2)}.$$

Given this, we have

$$f_{Y,U}(y, T, S) = \frac{\tau_p}{\sqrt{2\pi}^p \rho\sigma^{p-1}} \int_T^S f_U(u) \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^p r_j^2(u) + \frac{\tau_p^2}{2\sigma^2}\left(\sum_{j=1}^p r_j(u)\right)^2 \right) du. \tag{6}$$

in which we have made explicit the dependency of r_j in u. The resulting integral cannot be computed analytically and our implementation is based on numerical evaluations using Matlab. Let $\psi(\theta, u)$ denote the term in the exponential in (6). The gradient of (6) with respect to the parameters then takes the form

$$f_{Y,U}(y, T, S) = \frac{\tau_p}{\sqrt{2\pi}^p \rho \sigma^{p-1}} \int_T^S \nabla_\theta \psi(\theta, u) f_U(u) \exp\left(-\psi(\theta, u)\right) du,$$

which can also be computed numerically. We used Matlab's gradient-based optimization functions to maximize the log-likelihood. Because of the non-convexity of the likelihood function, we have found it a necessity to perform several runs with different initial conditions. Our numerical procedure is summarized below. When estimating the optimal value of Δ, we fix an interval $[\Delta_{\min}, \Delta_{\max}]$ and step δ for incrementing Δ. (To ensure that the likelihood is differentiable in Δ, we have replaced the positive part function $x \mapsto x^+$ by a smooth approximation, replacing x^+ with $(x+\varepsilon)^2/(4\varepsilon)$ for $x \in [-\varepsilon, \varepsilon]$ and ε small enough.)

Optimization Algorithm

Initialization: We first estimate a sub-model with $c = 0$, estimating initial values for a, b_1, b_2, ρ^2 and σ^2. To simplify this initialization step, we impute values for the missing observations of U (using the conditional expectation of U given $U \geq T$ in place of right-censored observations), therefore reducing the problem to a linear model with random effects.

Preliminary Step: Maximize the complete model log-likelihood initializing the gradient ascent algorithm with the parameters obtained at the previous step completed with $c = 0$ and $\Delta = \Delta_{\min}$. Assuming that $\Delta_{\min}$ is small enough for $(t - U + \Delta_{\min})^+$ to vanish most of the time, the parameters estimated for the reduced model can be expected to provide a reasonable initial guess.

Step m: If $\Delta_{\min} + m\delta > \Delta_{\max}$, go to the final step. Otherwise, maximize the likelihood using as an initial condition the parameters found in the previous iteration, but replacing Δ with $\Delta_{\min} + m\delta$.

Final Step: Keep the set of parameters that provided the largest value of the log-likelihood.

3.3. *Validation.* Since Δ is not identifiable when $c = 0$, it is important to reject this hypothesis to ensure that the estimated value of the change point is meaningful. We used a likelihood ratio test, therefore comparing the maximum log-likelihood obtained in the previous section to the one obtained in

the case of $c = 0$. The likelihood in this latter case (which can be computed using the same approach as for the general model) was also maximized using a gradient-based method, and in this case, we used the better result from two runs: the first one starts from the parameters estimated from the imputed model in the initialization step of the general algorithm. The second starts with $a + c\Delta$, $b_1 + c$, $b_2 - c$, ρ and σ, where $(a, b_1, b_2, c, \Delta, \rho, \sigma)$ are the maximum likelihood parameters estimated for the general model. This choice relies on the fact that for large values of Δ, ensuring $t - U + \Delta > 0$ with large probability, the general model becomes very close to the linear submodel with the chosen transformation of parameters.

A p-value for the likelihood ratio can be computed based on bootstrap estimates. We used for this purpose a semi-parametric approach in which each bootstrap sample was computed as follows, given the maximum likelihood parameters $(a, b_1, b_2, c, \Delta, \rho, \sigma)$.

(i) Impute random values u_k for the right-censored U's, drawn, for each k, according to the conditional distribution of U given the observed variables.
(ii) Compute "model residuals"

$$R_{kj} = y_{kj} - a - b_1 t_{kj} - b_2 u_k - c(t_{kj} - u_k + \Delta)^+.$$

(iii) Whiten residuals according to the random effect model as follows. For each k, with p_k observations, let $\tau_{p_k}^2 = \sigma^2/(\sigma^2 + p_k\rho^2)$. Set

$$W_{kj} = R_{kj} + (\tau_{p_k} - 1)\bar{R}_k$$

where $\bar{R}_k = (R_{k1} + \cdots + R_{kp_k})/p_k$.
(iv) Stack all W_{kj} in an $N = p_1 + \cdots + p_n$ vector, and sample with replacements new values W_{kj}^* from this vector.
(v) Reconstruct bootstrap residuals using

$$R_{kj}^* = W_{kj}^* - (1 - \tau_{p_k}^{-1})\bar{W}_k^*.$$

(vi) Define the complete bootstrap sample by

$$y_{kj}^* = a + b_1 t_{kj} + b_2 u_k + c(t_{kj} - u_k + \Delta)^+ + R_{kj}^*$$

and the null bootstrap sample by

$$y_{kj}^{0,*} = a + b_1 t_{kj} + b_2 u_k + R_{kj}^*.$$

A p-value for the hypothesis $c = 0$ can then be obtained by computing the fraction of times the likelihood ratio obtained on the true sample is smaller

than the ratios obtained on a large number of null bootstrap samples. When this p-value is small enough, standard deviations, or confidence intervals on the estimated parameters can be deduced from the distribution of the complete bootstrap samples.

Steps (iii) and (v) above are justified as follows (fixing k and letting $p = p_k$). The covariance matrix between the residuals is $S = (s_{ij}, 1 \leq i, j \leq p)$ with $s_{ij} = \sigma^2\delta_{ij} + \rho^2$. The vector $\mathbf{1}_p/\sqrt{p}$ (where $\mathbf{1}_p$ is p-dimensional and composed entirely of ones) is an eigenvector with eigenvalue $\sigma^2 + p\rho^2$ and the other $p-1$ eigenvectors span the space orthonormal to it, with eigenvalue σ^2. This implies that, for $x \in \mathbb{R}^p$, and letting $\bar{x} = (x_1 + \cdots + x_p)/p = x^T\mathbf{1}_p/p$,

$$\begin{aligned} S^{-1/2}x &= (\sigma^2 + p\rho^2)^{-1/2}(\bar{x})\mathbf{1}_p + \sigma^{-1}(x - (\bar{x})\mathbf{1}_p) \\ &= \sigma^{-1}\left(x - \left(1 - \sqrt{\frac{\sigma^2}{\sigma^2 + p\rho^2}}\right)\bar{x}\mathbf{1}_p\right) \\ &= \sigma^{-1}\left(x - (1 - \tau_p)\,\bar{x}\mathbf{1}_p\right). \end{aligned}$$

Similarly

$$\begin{aligned} S^{1/2}x &= (\sigma^2 + p\rho^2)^{1/2}(\bar{x})\mathbf{1}_p + \sigma(x - (\bar{x})\mathbf{1}_p) \\ &= \sigma\left(x + \left(\tau_p^{-1} - 1\right)\bar{x}\mathbf{1}_p\right), \end{aligned}$$

which justifies steps (iii) and (v), in which we have used the fact that the factors σ and σ^{-1} cancel in the overall computation.

3.4. *Remarks.*

1. Our change-point onset model is similar to the one proposed in Younes et al. (2014), which was applied to anatomical changes of brain structures based on the BIOCARD dataset. The approach in Younes et al. (2014), however, made the simplifying assumption that whenever $u_k \geq T_k$ (and therefore the manifest onset time is not observed), then $u_k - \Delta$ was also larger than the last measurement time, t_{kp_k}. This was justified by the rather large delay between this last measurement and the last cognitive assessment (about five years), and the belief that Δ would be less than or comparable to this delay. Some other biomarkers, however, which were not considered in Younes et al. (2014), appear to be associated to large values of Δ for which these assumptions would not be justified, motivating the more complex procedure proposed in this paper.

2. Even though we have limited our discussion to the model described in equation (1), it is easy to generalize it to more complex models, including for

example additional variables (covariates) or higher-order dependency with respect to age.

4. Experiments.

4.1. *Simulations.* In our first analysis, we conducted simulation experiments with synthetic data, which allowed us to evaluate the performance of our estimation algorithm with a known ground truth. We used two sets of parameter values for a, b_1, b_2, c, ρ and σ, and for each of these sets, simulated data with $\Delta = 5$, 10, 15 or 20 years (see Table 4). These values were based on two sets of observed values from real-data experiments. We also simulated samples with $c = 0$ to estimate a threshold for the likelihood ratio under the null hypothesis.

Our simulations try to follow a subject recruitment process. We started with 300 "subjects", with age at the beginning of the study (t_{k1}) simulated from a Gaussian distribution with mean 60 and standard deviation 10. The manifest onset time (u_k) was sampled according to an exponentially modified Gaussian distribution with $m_1 = 73$, $\sigma_1 = 17$ and $\alpha = 2$. Subjects that did not satisfy the right censoring condition $u_k \geq t_{k1}$ were automatically excluded from analysis (the average number of excluded subjects was 67.2). We also assumed that the number of biomarker observations for each subject followed a uniform distribution over $\{1, \ldots, 5\}$. The interval between every two consecutive longitudinal measurements was fixed, equal to 2 years. The length of the entire study was assumed to be 15 years, and thus $T_k = t_{k1} + 15$ for every k. About 64.5% of the selected subjects were right censored (i.e., had a manifest onset age posterior to the end of the study). An example of a simulated dataset is illustrated in Figure 1.

For each of these 8 estimation experiments (and for the two simulations made under the null hypothesis), a total of 1,000 independent simulations were performed. Table 4 provides 25, 50 and 75 percentiles of the observed distribution of the estimated coefficients. The estimated 95 percentile of the log-likelihood difference under the null hypothesis ($c = 0$) was found to equal 3.6 in both models. This allows us to estimate the power of the rejection test in the other cases. The fraction of simulations for which the log-likelihood difference was larger than this threshold was 87%, 99.2%, 99.5% and 99.5% for $\Delta = 5$, 10, 15 and 20 respectively under the first parameter set (for which $a = 20$) and 94%, 99.6%, 99.5% and 99.3% for the second parameter set ($a = 1.4$).

These results indicate a reasonable accuracy in the estimation of the change point. Small change point values are more accurately estimated. The estimation of a also seem to slightly degrade with large change points. The

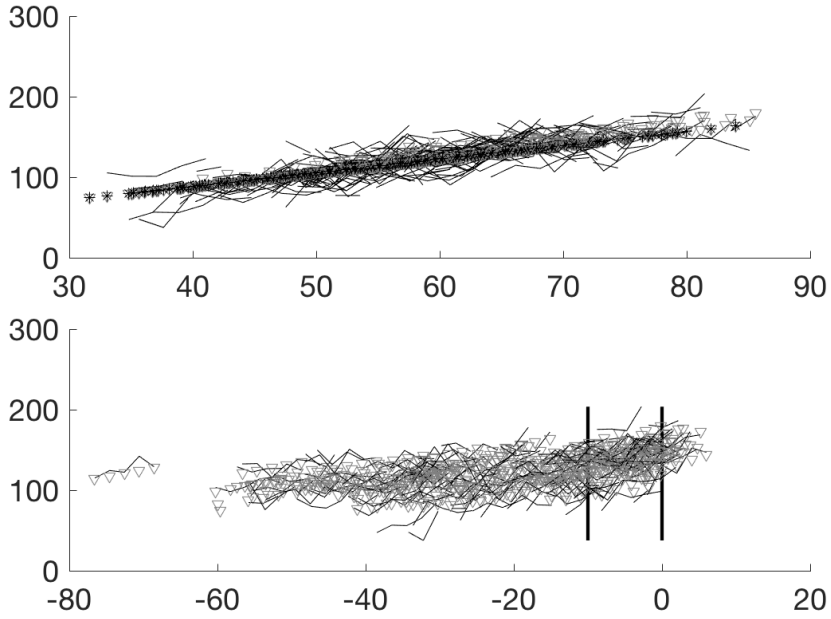

FIG 1. *Simulated dataset with 206 subjects and 599 total observations represented without right censoring. Lines represent observed data (one line per subject) while gray triangles are the model predictions. The true changepoint is 10 years before onset (represented by vertical lines in the second chart. The other parameters are those used in the four first simulations in Table 4*

.

estimation of the change point's slope is also quite accurate and the fixed and random effect standard deviations (ρ and σ) are consistently well estimated.

4.2. *Onset of Alzheimer's Disease.* We now provide a few results based on real data, focusing on biomarkers related to AD. Magnetic resonance imaging (MRI) measures are an indirect reflection of the neuronal injury that occurs in the brain as the AD pathophysiological process evolves. The volumetric measurements of medial temporal lobe structures, such as the hippocampus and the amygdala, have been shown to be important anatomical hallmarks for AD, exhibiting significant atrophy in patients with both AD and MCI as compared to their healthy counterparts (Jack et al., 1997, 1992). Those volumetric measurements have also been shown to be associated with time to progress from MCI to AD dementia (Atiya et al., 2003; Kantarci and Jack, 2003). In addition to volumetric measurements, shape-based biomarkers have also been found to be sensitive to the pathology of AD, revealing region-specific heterogeneous atrophy patterns in the hip-

	a	b_1	b_2	c	Δ	ρ	σ
True Values	20.00	1.70	0.01	1.40	5.00	10.00	10.00
25 percentile	9.485	1.619	-0.014	1.169	4.181	9.174	9.873
Median	18.15	1.668	0.047	1.489	5.438	9.539	10.125
75 percentile	21.318	1.722	0.171	1.816	7.011	10.132	10.243
True Values	20.00	1.70	0.01	1.40	10.00	10.00	10.00
25 percentile	7.969	1.578	-0.025	1.382	9.002	9.118	9.870
Median	14.59	1.671	0.113	1.524	10.160	9.540	10.100
75 percentile	20.906	1.738	0.207	1.774	11.797	10.120	10.207
True Values	20.00	1.70	0.01	1.40	15.00	10.00	10.00
25 percentile	6.784	1.547	-0.124	1.308	12.166	8.741	9.871
Median	16.68	1.693	0.051	1.556	13.717	9.533	10.103
75 percentile	26.955	1.796	0.250	1.772	17.708	10.012	10.200
True Values	20.00	1.70	0.01	1.40	20.00	10.00	10.00
25 percentile	4.713	1.471	-0.221	1.205	15.678	8.645	9.911
Median	17.96	1.746	-0.039	1.486	18.070	9.370	10.087
75 percentile	31.384	1.892	0.297	1.729	23.962	9.992	10.218
True Values	1.40	-0.10	0.00	-0.02	5.00	0.16	0.12
25 percentile	1.304	-0.101	-0.001	-0.025	3.702	0.150	0.118
Median	1.40	-0.100	0.000	-0.021	4.770	0.155	0.121
75 percentile	1.458	-0.099	0.001	-0.016	6.331	0.161	0.123
True Values	1.40	-0.10	0.00	-0.02	10.00	0.16	0.12
25 percentile	1.263	-0.101	-0.001	-0.024	8.747	0.150	0.118
Median	1.39	-0.100	0.000	-0.020	9.847	0.156	0.121
75 percentile	1.433	-0.099	0.002	-0.018	10.750	0.162	0.123
True Values	1.40	-0.10	0.00	-0.02	15.00	0.16	0.12
25 percentile	1.265	-0.101	-0.001	-0.022	12.436	0.149	0.118
Median	1.32	-0.100	0.001	-0.020	14.273	0.153	0.121
75 percentile	1.420	-0.098	0.003	-0.017	17.222	0.161	0.123
True Values	1.40	-0.10	0.00	-0.02	20.00	0.16	0.12
25 percentile	1.143	-0.103	-0.002	-0.022	16.107	0.145	0.119
Median	1.33	-0.101	0.002	-0.020	20.438	0.150	0.122
75 percentile	1.481	-0.097	0.005	-0.017	22.806	0.157	0.123

TABLE 4

Simulation results. Each group of results provides the true values of the model parameters, followed by the 25, 50 *and* 75 *percentiles of the estimated values based on 1,000 independent simulations.*

pocampus and the amygdala (Tang et al., 2014; Miller et al., 2015b). It has also been demonstrated that baseline morphometric measures, in terms of both volume and shape, of the hippocampus and the amygdala in healthy controls were capable of predicting subsequent development of MCI (den Heijer et al., 2006), with hippocampal shape differences detected among healthy controls who subsequently developed cognitive impairment (Csernansky et al., 2005; den Heijer et al., 2006; Kantarci and Jack, 2003; Rusinek et al., 2003; Thambisetty et al., 2010).

Extracting shape measurements of the structural biomarkers of AD, such as the hippocampus and the amygdala, from MRI datasets usually requires a complex processing pipeline before the statistical analysis described in this paper can be performed. This starts with the extraction of "regions of interest" (ROI) which are 3D volumes or 2D surfaces of specific anatomical structures of the human brain that are affected by the disease, such as the entorhinal cortex, the hippocampus or the amygdala in AD (Fischl, 2012; Pierson et al., 2011; Tang et al., 2013). This segmentation step, even if mostly automated, still require significant human intervention for quality control and small manual corrections. The 2D surfaces contouring the boundary of the segmentations provide the collection of "shapes" on which the second step, nonrigid alignment, will be performed.

Nonrigid alignment can be interpreted as an operation that places all shapes in a common "coordinate system" in an infinite-dimensional "shape space". While all of this can be mathematically formalized (Miller et al., 2014, 2015a; Younes, 2010; Bauer et al., 2014), from a computational point of view, the operation boils down to solving a collection of optimal control problems, each of which optimizes a deformation process in which an initial shape (called template) is mapped onto a subject shape, with the template being optimized at the same time (Ma et al., 2010). At the end of the process, each subject shape is represented as a diffeomorphic transformation of a common shape (the template), and the problem is reduced to the study of the resulting collection of diffeomorphisms (yielding the term"diffeomorphometry" (Miller et al., 2014)). A variety of mathematically refined descriptors of the diffeomorphisms can then be analyzed, with the simplest case usually being their Jacobian determinant indexed at each vertex of the template surface, where the diffeomorphism is either considered as a 3D transformation (interpreting the Jacobian as a volume ratio), or a 2D transformation between surfaces (interpreting the Jacobian as a surface area ratio). As a result, shape data are transformed so that each individual surface is represented as a large collection of variables, a random field supported by the template surface. In this paper, we take an additional step

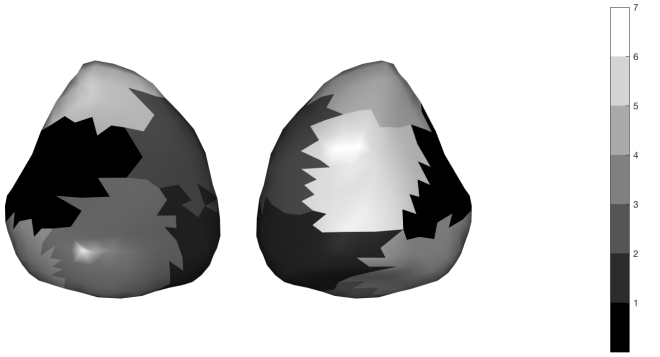

FIG 2. *Two views of the template surface for the amygdala. Gray-levels are associated with spectral segmentation labels.*

to reduce the dimension of the shape variables by averaging these variables over sub-regions of the template obtained via spectral segmentation (Reuter, 2010). The template surface for the amygdala, on which we will focus here, and the computed subregions are illustrated in Figure 2.

With seven segmented sub-regions, and a separate analysis of the amygdalar surfaces in both hemispheres, our final real data consist in a collection of fourteen variables that provide an average amount of atrophy or expansion measured for each subject relative to the template. The dataset we used included 292 subjects among which 70 were diagnosed as MCI before the end of the study (right-censoring therefore applying to 222 subjects).

Based on the discussion made at the end of section 3.1, we used a Gaussian prior with $m_1 = 93$ and $\sigma_1 = 14.5$. We applied the change point model separately to each surface, obtaining in this way fourteen regional estimates. Figure 3 provides likelihood plots (maximum log-likelihood as a function of Δ for the first four regions on the left hemisphere). We used 1,000 bootstrap replicates to estimate the 25th and 75th percentiles of the estimators' distribution, and 1,000 additional replicates to estimate p-values for testing the null hypothesis of no change point. These results are summarized in Table 5. Note that the p-values we tabulate in this table were computed separately for each variable, and were therefore not corrected for multiple comparisons. Subregions 4 and 5 (p-values: 0.088, 0.130) on the left amygdala and subregions 1, 2, 4, 5, and 6 (p-values: 0.231, 0.078, 0.240, 0.127, 0.182) on the right amygdala were not significant and are not reported in this table. Figures 4, 5 and 6 illustrate the results on regions 1, 2 and 3 of the left amygdala by

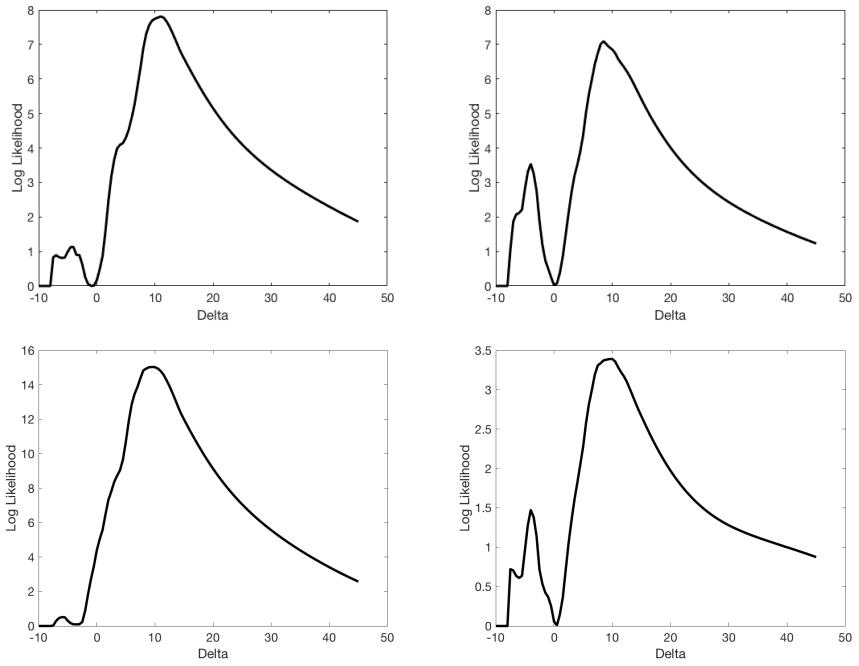

FIG 3. *Likelihood profiles (maximum log-likelihood with fixed Δ as a function of Δ) for the first four regions on the left amygdala (from left to right and top to bottom). Note that the change point model for the fourth region is not significant, with a likelihood exhibiting a smaller spread in value. These plots also illustrate the non-concavity of the likelihood function.*

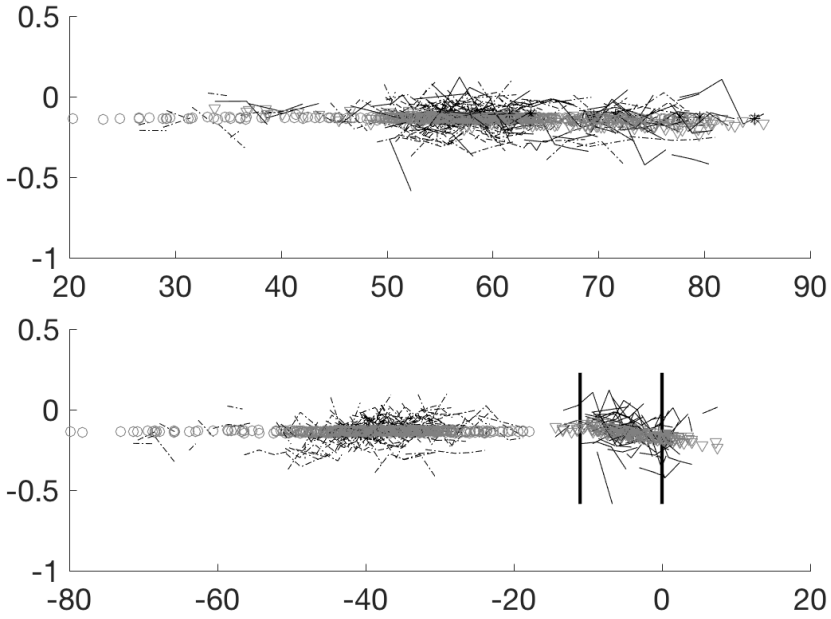

FIG 4. *Shape biomarket for the left amygdala (region 1) plotted as a function of age (up) and as a function of years before onset (down). Lines represent observed data (one line per subject) while gray circles and triangle are the model predictions respectively for right-censored (controls) and diseased subjects. On the second chart, the time before onset for right-censored subjects (circles) is repalced by its posterior mean. The distance between the two vertical lines in this chart in the estimated change-point time before onset.*

plotting the biomarker values and the model predictions as functions of age and of years before onset. Figure 7 provides a visual representation of the estimated change point results for each subregion of the bilateral amygdalas.

For regions on which p-values were not significant, we also tested the hypothesis $b_2 \neq 0$ within the submodel $c = \Delta = 0$, in order to evaluate the significance of the onset time U on a linear model without changepoint. None of these tests showed significance.

The amygdala is suggested to play a major role in enhancing the explicit memory related to emotional stimuli, by modulating the consolidation of memory (Hamann, 2001) which is greatly affected by the pathology of AD. Studies have reported amygdalar abnormalities induced by AD, such as loss of neurons as measured in neuropathological analyses (Tsuchiya and Kosaka, 1990; Scott et al., 1991, 1992); loss of volume measured from MRI (den Heijer et al., 2006; Poulin et al., 2011); local atrophy based on shape analysis

	a	b_1	b_2	c	Δ	ρ	σ
Region 1 (Left), $p = 0.002$							
Estimated Value	0.004	0.000	-0.001	-0.007	11.052	0.067	0.064
MAD	0.054	0.000	0.001	0.002	2.109	0.003	0.002
Region 2 (Left), $p = 0.002$							
Estimated Value	0.086	0.000	-0.002	-0.010	8.394	0.079	0.081
MAD	0.057	0.000	0.000	0.002	1.670	0.004	0.002
Region 3 (Left), $p = 0.001$							
Estimated Value	0.051	0.000	-0.001	-0.011	9.723	0.071	0.070
MAD	0.056	0.000	0.000	0.002	1.408	0.004	0.002
Region 7 (Left), $p = 0.006$							
Estimated Value	0.044	0.000	-0.001	-0.009	9.225	0.084	0.085
MAD	0.063	0.000	0.001	0.002	2.120	0.004	0.002
Region 2 (Right), $p = 0.021$							
Estimated Value	0.079	-0.000	-0.001	-0.007	11.199	0.079	0.084
MAD	0.071	0.000	0.001	0.002	3.032	0.004	0.003
Region 3 (Right), $p = 0.012$							
Estimated Value	-0.019	0.000	-0.001	-0.011	5.469	0.066	0.077
MAD	0.047	0.000	0.000	0.003	1.234	0.004	0.003
Region 5 (Right), $p = 0.036$							
Estimated Value	0.176	-0.001	-0.001	-0.007	11.348	0.093	0.091
MAD	0.088	0.000	0.001	0.002	4.027	0.005	0.003
Region 7 (Right), $p = 0.044$							
Estimated Value	0.103	0.000	-0.001	-0.006	10.834	0.078	0.083
MAD	0.067	0.000	0.001	0.002	3.260	0.004	0.002

TABLE 5

Results of change point estimation for the amygdala, based on the BIOCARD dataset. Each group of results provides the estimated values of the model parameters, followed by the median absolute deviation (MAD). Only results with uncorrected p-value $p < 0.05$ are provided.

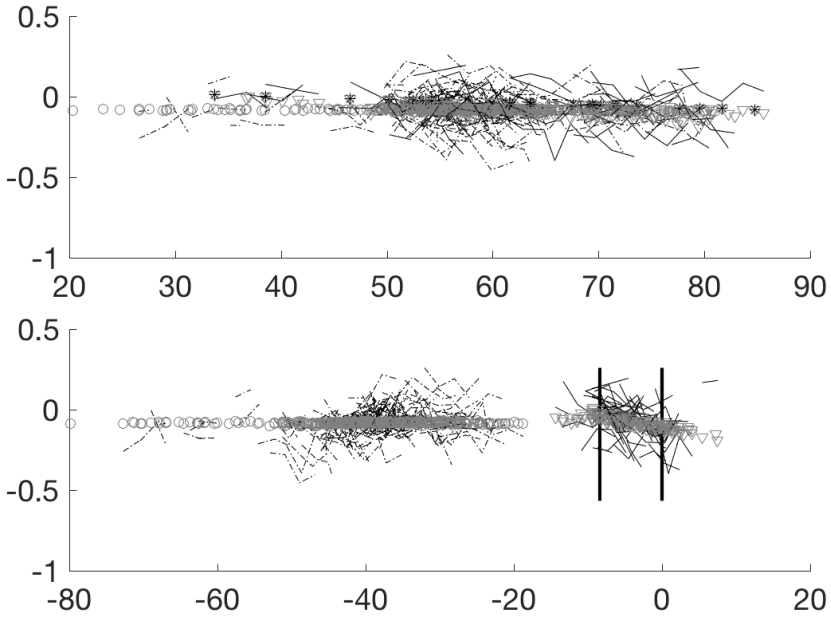

FIG 5. *Shape biomarker for the left amygdala (region 2) plotted as a function of age (up) and as a function of years before onset (down). See Figure 4 for more details.*

on MRI (Cavedo et al., 2011; Miller et al., 2015b). Our observations of amygdalar shape atrophy (see figure 7 and Table 5) are keeping in line with previous findings. In addition, results from our method revealed an accelerated amygdalar atrophy rate induced by AD relative to normal aging. This finding is also consistent with those reported in longitudinal studies of AD focusing on amygdalar shape (Tang et al., 2015). The localization of subregions found to be significantly affected by AD in our experiment is also roughly consistent with previous longitudinal results obtained from high-field subregion segmentations (Miller et al., 2015b; Tang et al., 2015), even though variation may occur due to the limited resolution of our MRI data. As shown in Table 5, mainly the amygdalar subregions 2, 3, and 7 were identified, which roughly corresponds to the basolateral and basomedial subregions of the amygdala, the core subregions of the amygdala as defined according to functional characteristics Price (2003); Sheline et al. (1998). With that being said, none of these previous research work has ever provided the information that seems to be emerging from our analyses, namely an acceleration or start of amygdalar atrophy about 10 years before the onset of AD. Our analysis is the first time, to the best of our knowledge, to have demonstrated that there are subregion-dependent amygdalar atrophy onsets

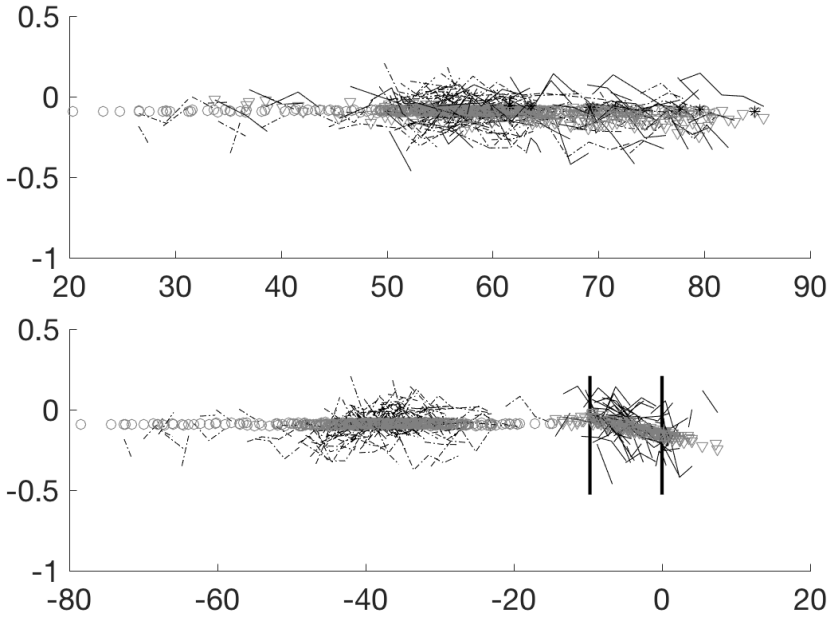

FIG 6. *Shape biomarker for the left amygdala (region 3) plotted as a function of age (up) and as a function of years before onset (down). See Figure 4 for more details.*

ranging from 8 to 11 years before the clinical onset of AD. This provides unique and important information that furthers our understanding of the pathology of AD, especially its influence on the amygdalar morphometry.

5. Discussion. We have described in this paper an approach to estimate a change point for a biomarker relative to the occurrence of an event (manifest onset) which may be only partially observed. We have described parameter estimation procedures for the prior model on the manifest onset time, and for the two-phase regression model on the biomarkers, with a bootstrap-based model validation scheme.

Our simulation study shows that the learning procedures perform satisfactorily in the ideal case (correct model class), with parameters akin to those estimated from some of the real world data we considered later. With roughly 230 observations, among which about 2/3 were right-censored, change point estimates for true values of 5, 10, 15 and 20 years showed little or even no bias with the gap between the 1/4 and 3/4 quantiles being about 1 and 2 years away. Most of the other coefficients were estimated with very good accuracy, except for the intercept (the variations of which are exacerbated by the fact that the sample ages were far away from 0). It is also important

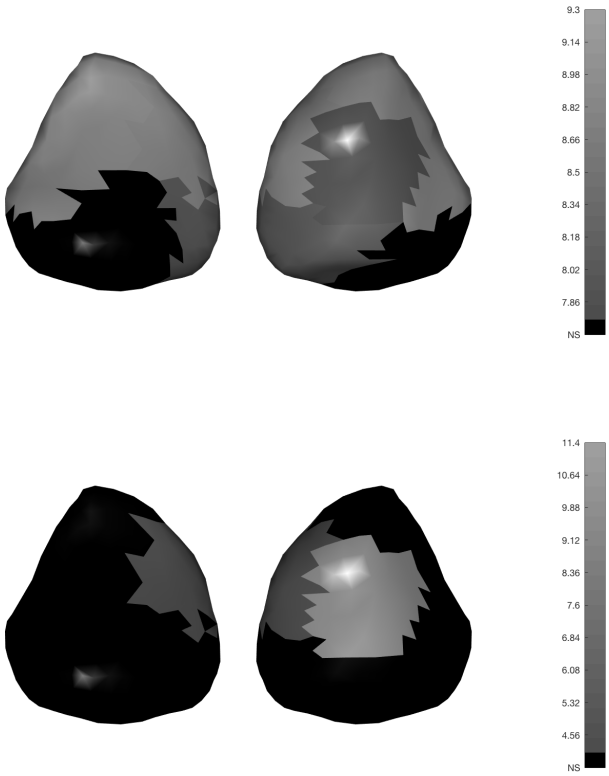

FIG 7. *Change point estimates mapped on segmented regions in the amygdala (two views of left-side results, followed by two views of right side). Black areas are not significant*

for the validation of the real-data study that we found a rather large power (close to or larger than 90%) for the likelihood-ratio test of the change point detection.

The bootstrap-estimated variations of the change point estimated from real shape data were consistent with those observed in the simulations. They indicated a disease effect around 10 years before the manifest (cognitive) onset. The likelihood profiles computed in Figure 3 illustrate the difficulty of the estimation of the change point, with a non-concave likelihood exhibiting several local maxima. This observation, which is typical of change point estimation problems, is reinforced in our case by the fact that we are working with a significant amount of right censored subjects.

On the theoretical side, important problems are raised by the presented approach, this paper being limited to experimental validations. The consistency of the maximum-likelihood estimate, and its asymptotic accuracy need to be studied. A rigorous justification of the bootstrap procedure also needs to be developed. These issues, which are left open in the present paper, will be the subject of future work in our group. Future work will also focus on extensions of the model, allowing for evolutions that are more complex than a two-phase linear regression: estimating more than one change point or allowing for nonlinear changes in each of the phases.

The results presented here are valid at the population level. Even though, using our model, we were able to compute an individual estimator of the time to onset (as used in Figures 4, 5 and 6), this estimator is very crude and does not provide a reliable individual prediction. Research in this direction is likely to intensify in the near future, however, and we can expect that several weak predictors, such as those derived here for the amygdala, will need to be combined for early diagnosis of AD.

References.

Alzheimer's Association
2015. 2015 Alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 11(3):332.

Atiya, M., B. T. Hyman, M. Albert, and R. Killiany
2003. Structural magnetic resonance imaging in established and prodromal Alzheimer disease: a review. *Alzheimer Disease & Associated Disorders*, 17(3):177–195.

Bauer, M., M. Bruveris, and P. W. Michor
2014. Overview of the geometries of shape spaces and diffeomorphism groups. *Journal of Mathematical Imaging and Vision*, 50(1-2):60–97.

Cavedo, E., M. Boccardi, R. Ganzola, E. Canu, A. Beltramello, C. Caltagirone, P. Thompson, and G. Frisoni
2011. Local amygdala structural differences with 3t mri in patients with Alzheimer disease. *Neurology*, 76(8):727–733.

Chen, J. and A. K. Gupta
2012. *Parametric Statistical Change Point Analysis*. Boston: Birkhäuser Boston.

Csernansky, J. G., L. Wang, J. Swank, J. P. Miller, M. Gado, D. McKeel, M. I. Miller, and J. C. Morris
2005. Preclinical detection of alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage*, 25(3):783–792.

den Heijer, T., M. I. Geerlings, F. E. Hoebeek, A. Hofman, P. J. Koudstaal, and M. M. Breteler
2006. Use of hippocampal and amygdalar volumes on magnetic resonance imaging to predict dementia in cognitively intact elderly people. *Archives of general psychiatry*, 63(1):57–62.

Dupuy, J.-F.
2006. Estimation in a change-point hazard regression model. *Statistics & Probability Letters*, 76(2):182–190.

Farley, J. U. and M. J. Hinich
1970. A test for a shifting slope coefficient in a linear model. *Journal of the American Statistical Association*, 65(331):1320–1329.

Feder, P. I.
1975. On Asymptotic Distribution Theory in Segmented Regression Problems– Identified Case. *The Annals of Statistics*, 3(1):49–83.

Fischl, B.
2012. Freesurfer. *Neuroimage*, 62(2):774–781.

Gombay, E. and L. Horváth
1994a. An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications*, 50(1):161–171.

Gombay, E. and L. Horváth
1994b. Limit theorems for change in linear regression. *Journal of Multivariate Analysis*, 48(1):43–69.

Hamann, S.
2001. Cognitive and neural mechanisms of emotional memory. *Trends in cognitive sciences*, 5(9):394–400.

Hebert, L. E., J. Weuve, P. A. Scherr, and D. A. Evans
2013. Alzheimer disease in the united states (2010–2050) estimated using the 2010 census. *Neurology*, 80(19):1778–1783.

Hinkley, D. V.
1969. Inference about the intersection in two-phase regression. *Biometrika*, 56(3):495–504.

Hinkley, D. V.
1971. Inference in Two-Phase Regression. *Journal of the American Statistical Association*, 66(336):736–743.

Hudson, D. J.
1966. Fitting segmented curves whose join points have to be estimated. *Journal of the american statistical association*, 61(316):1097–1129.

Jack, C. R., R. C. Petersen, P. C. O'Brien, and E. G. Tangalos
1992. Mr-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology*, 42(1):183.

Jack, C. R., R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen
1997. Medial temporal atrophy on mri in normal aging and very mild alzheimer's disease. *Neurology*, 49(3):786–794.

Kantarci, K. K. and C. R. Jack
2003. Neuroimaging in Alzheimer disease: an evidence-based review. *Neuroimaging clinics of North America*, 13(2):197–209.

Larrieu, S., L. Letenneur, J. Orgogozo, C. Fabrigoule, H. Amieva, N. Le Carret, P. Barberger-Gateau, and J. Dartigues
2002. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology*, 59(10):1594–1599.

Li, Y., L. Qian, and W. Zhang
2013. Estimation in a change-point hazard regression model with long-term survivors. *Statistics & Probability Letters*, 83(7):1683–1691.

Ma, J., M. I. Miller, and L. Younes
2010. A Bayesian generative model for surface template estimation. *Journal of Biomedical Imaging*, 2010:16.

Miller, M. I., A. Trouvé, and L. Younes
2015a. Hamiltonian systems in computational anatomy: 100 years since d'arcy thompson. *Annual Review of Biomedical Engineering*, 17(1).

Miller, M. I., L. Younes, J. T. Ratnanather, T. Brown, H. Trinh, D. S. Lee, D. Tward, P. B. Mahon, S. Mori, and M. Albert
2015b. Amygdalar atrophy in symptomatic alzheimer's disease based on diffeomorphometry: the {BIOCARD} cohort. *Neurobiology of Aging*, 36, Supplement 1:S3 – S10. Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD).

Miller, M. I., L. Younes, and A. Trouvé
2014. Diffeomorphometry and geodesic positioning systems for human anatomy. *Technology*, 2(01):36–43.

Mueller, S. G., M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett
2005. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (adni). *Alzheimer's & Dementia*, 1(1):55–66.

Nguyen, H. T., G. S. Rogers, and E. A. Walker
1984. Estimation in change-point hazard rate models. *Biometrika*, 71(2):299–304.

Pierson, R., H. Johnson, G. Harris, H. Keefe, J. S. Paulsen, N. C. Andreasen, and V. A. Magnotta
2011. Fully automated analysis using brains: Autoworkup. *NeuroImage*, 54(1):328–336.

Pons, O.
2003. Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics*, 31(2):442–463.

Poulin, S. P., R. Dautoff, J. C. Morris, L. F. Barrett, B. C. Dickerson, A. D. N. Initiative, et al.
2011. Amygdala atrophy is prominent in early alzheimer's disease and relates to symptom severity. *Psychiatry Research: Neuroimaging*, 194(1):7–13.

Price, J. L.
2003. Comparative aspects of amygdala connectivity. *Annals of the New York Academy of Sciences*, 985(1):50–58.

Quandt, R. E.
1958. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association*, 53(284):873–880.

Reuter, M.
2010. Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions. *International Journal of Computer Vision*, 89(2-3):287–308.

Rusinek, H., S. De Santi, D. Frid, W.-H. Tsui, C. Y. Tarshish, A. Convit, and M. J. de Leon
2003. Regional brain atrophy rate predicts future cognitive decline: 6-year longitudinal mr imaging study of normal aging 1. *Radiology*, 229(3):691–696.

Scott, S. A., S. T. DeKosky, and S. W. Scheff
1991. Volumetric atrophy of the amygdala in Alzheimer's disease quantitative serial reconstruction. *Neurology*, 41(3):351–351.

Scott, S. A., D. L. Sparks, S. W. Scheff, S. T. Dekosky, and C. A. Knox
1992. Amygdala cell loss and atrophy in Alzheimer's disease. *Annals of neurology*, 32(4):555–563.

Sheline, Y. I., M. H. Gado, and J. L. Price
1998. Amygdala core nuclei volumes are decreased in recurrent major depression. *Neuroreport*, 9(9):2023–2028.

Sprent, P.
1961. Some Hypotheses Concerning Two Phase Regression Lines. *Biometrics*, 17(4):634–645.

Tang, X., D. Holland, A. M. Dale, L. Younes, and M. I. Miller
2014. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer's disease: Detecting, quantifying, and predicting. *Human brain mapping*, 35(8):3701–3725.

Tang, X., D. Holland, A. M. Dale, L. Younes, and M. I. Miller
2015. The diffeomorphometry of regional shape change rates and its relevance to cognitive deterioration in mild cognitive impairment and alzheimer's disease. *Human brain mapping*, 36(6):2093–2117.

Tang, X., K. Oishi, A. V. Faria, A. E. Hillis, M. Albert, S. Mori, and M. I. Miller
2013. Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. *PLoS ONE*, 8(6).

Thambisetty, M., A. Simmons, L. Velayudhan, A. Hye, J. Campbell, Y. Zhang, L.-O. Wahlund, E. Westman, A. Kinsey, A. Güntert, et al.
2010. Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Archives of general psychiatry*, 67(7):739–748.

Tsuchiya, K. and K. Kosaka
1990. Neuropathological study of the amygdala in presenile Alzheimer's disease. *Journal of the neurological sciences*, 100(1):165–173.

Wu, C., L. Zhao, and Y. Wu
2003. Estimation in change-point hazard function models. *Statistics & Probability Letters*, 63(1):41–48.

Younes, L.
2010. *Shapes and Diffeomorphisms*. Springer Berlin Heidelberg.

Younes, L., M. Albert, M. I. Miller, B. R. Team, et al.
2014. Inferring changepoint times of medial temporal lobe morphometric change in preclinical Alzheimer's disease. *NeuroImage: Clinical*, 5:178–187.

SYSU-CMU Joint Institute of Engineering
Sun Yat-sen University
No. 132, East Waihuan Road
Guangzhou Higher Education Mega Center
Guangzhou, 510006, P.R. China
E-mail: tangxiaoy@mail.sysu.edu.cn

Center for Imaging Science
Johns Hopkins University
3400 N. Charles st.
Baltimore MD 21218, USA
E-mail: mim@cis.jhu.edu
E-mail: laurent.younes@jhu.edu