



Published in final edited form as:

Neuroimage. 2014 August 15; 97: 9–18. doi:10.1016/j.neuroimage.2014.04.015.

Event Time Analysis of Longitudinal Neuroimage Data

Mert R. Sabuncu^{1,2,*}, Jorge L. Bernal-Rusiel^{1,*}, Martin Reuter^{1,3}, Douglas N. Greve¹, and Bruce Fischl^{1,2} for the Alzheimer's Disease Neuroimaging Initiative**

¹Athinoula A. Martinos Center for Biomedical Imaging, Harvard Medical School/Massachusetts General Hospital, Charlestown, MA

²Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA

³Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA

Abstract

This paper presents a method for the statistical analysis of the associations between longitudinal neuroimaging measurements, e.g., of cortical thickness, and the timing of a clinical event of interest, e.g., disease onset. The proposed approach consists of two steps, the first of which employs a linear mixed effects (LME) model to capture temporal variation in serial imaging data. The second step utilizes the extended Cox regression model to examine the relationship between *time-dependent* imaging measurements and the timing of the event of interest. We demonstrate the proposed method both for the univariate analysis of image-derived biomarkers, e.g., the volume of a structure of interest, and the exploratory mass-univariate analysis of measurements contained in maps, such as cortical thickness or gray matter density. The mass-univariate method employs a recently developed spatial extension of the LME model. We applied our method to analyze structural measurements computed using FreeSurfer, a widely used brain Magnetic Resonance Image (MRI) analysis software package. We provide a quantitative and objective empirical evaluation of the statistical performance of the proposed method on longitudinal data from subjects suffering from Mild Cognitive Impairment (MCI) at baseline.

Keywords

Longitudinal Studies; Linear Mixed Effects Models; Event Time Analysis; Cox Regression; Survival Analysis

Corresponding Author: Mert R. Sabuncu, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Building 149, 13th Street, Room 2301, Charlestown, Massachusetts, USA 02129, Phone: 617 643-7460, Fax: 617 726-7422, msabuncu@nmr.mgh.harvard.edu.

*Mert R. Sabuncu and Jorge L. Bernal-Rusiel contributed equally.

**Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators is available at <http://tinyurl.com/ADNI-main>.

INTRODUCTION

Medical events, such as the onset of disease, represent major landmarks in the course of a patient's clinical history. A significant portion of biomedical research is dedicated to studying the risk factors associated with these events, aiming to predict, delay and ultimately prevent their occurrence.

In recent decades, neuroimaging has accelerated the study of brain-related clinical conditions. A classical neuroimaging approach has been to contrast measurements obtained from those who have experienced the event (i.e., cases) with measurements from those who have not (i.e., controls). This methodology has yielded reliable markers of disease, e.g., (Jack Jr et al., 2012), while providing insights about underlying biological mechanisms, e.g. (Buckner et al., 2005; Sabuncu et al., 2012)

Yet, the classical case-control approach treats the two groups as distinct entities and assumes a certain amount of within-group homogeneity. This approach can therefore be limited when the control group is a high-risk cohort, that is, when a significant proportion of subjects have not yet experienced the event of interest but are likely to do so in the not-too-distant future. Such "pre-event" cases, which, in the absence of other information will be treated as controls, typically fall in the gray area between a pure case and a pure control. Thus the within-group homogeneity assumption is violated, which will in turn impact statistical inference. Common examples for this are longitudinal studies of populations that are at high risk for disease, based on their genetic make-up (e.g., carriers of a faulty allele of the *Huntingtin* gene in a Huntington's study (Albin et al., 1990)), familial history (e.g., subjects who have a first-degree relative with schizophrenia (Whitfield-Gabrieli et al., 2009)) or clinical presentation (e.g., subjects with Mild Cognitive Impairment, or MCI, in an Alzheimer's study (Forsberg et al., 2008)). These examples are particularly relevant to drug trials focused on the pre-clinical or early phases of a disease and thus target high-risk populations. In such scenarios, an inappropriate statistical treatment of the group of subjects who have not been observed to experience the event (diagnosis or conversion to disease) during the follow-up period (sometimes referred to as "non-converters"), can introduce bias into the analysis and/or reduce efficiency.

An alternative strategy that addresses this issue, directly models the timing of the event of interest, while accounting for finite follow-up or censoring. This is the event time (or survival) analysis approach (Kleinbaum and Klein, 2012), which includes classical models such as Cox proportional hazards regression (Cox, 1972). Standard event time analysis models have been applied in prior neuroimaging studies (Desikan et al., 2009; Desikan et al., 2010; Devanand et al., 2007; Geerlings et al., 2008; Marcus et al., 2007; Sabuncu, 2013; Stoub et al., 2005; Tintore et al., 2008; Vemuri et al., 2011) and have yielded novel insights about various clinical conditions. Most of these prior studies have analyzed associations between imaging measurements from a *single* baseline visit and the timing of the event of interest identified via follow-up clinical assessments. These analyses typically rely on survival models (e.g., the standard Cox model) that assume the explanatory variables are independent of time (e.g., gender, genetic marker, birth place, etc.). The employed models are useful for constructing individualized survival curves and making predictions about the

timing of a future event. Furthermore, they offer insights about the relationships between independent variables and the event time. As such, survival models have been used to draw conclusions about associations between neuroimaging measurements (e.g., volume of a structure) and the clinical event (e.g., disease onset). This type of inference, however, suffers from two problems. Firstly, imaging measurements typically vary over time (e.g., due to anatomical changes). Yet, interpretation of the standard Cox model, for example, has to be done with respect to the baseline imaging measurements only and not with respect to the dynamically changing measurements. Secondly, in longitudinal designs that span an extended time period, imaging measurements are likely to vary substantially over time, making it harder to detect associations between baseline imaging markers and the clinical event.

Longitudinal neuroimaging (LNI) studies, where *multiple* serial images are acquired for each participant, provide a means to characterize the temporal trajectories of imaging measurements. Furthermore LNI studies can offer a substantial increase in statistical power for studying imaging markers (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b), while opening up the possibility of examining the relationship between the temporal dynamics of imaging markers and clinical variables (Sabuncu et al., 2011). Today, the standard strategy for analyzing the association between LNI data and the occurrence of a clinical event, such as disease onset, is to perform a group comparison based on dichotomizing the subjects into, for example, “converters” versus “non-converters” (Borgwardt et al., 2011; Chetelat et al., 2005; Jack Jr et al., 2008a; Morgan et al., 2011; Sun et al., 2009). However, as we discussed above, this approach can be suboptimal, since the non-converter group likely includes subjects who might convert beyond the study follow-up.

The core goal of this paper is to propose a powerful method for the statistical analysis of the associations between longitudinal neuroimaging measurements, e.g., of gray matter density or cortical thickness, and the timing of a clinical event of interest, such as disease onset. The proposed approach combines a linear mixed effects (LME) model that captures the spatiotemporal correlation pattern in serial imaging data (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b; Verbeke and Molenberghs, 2000) and an extended Cox regression model that allows the examination of associations between the *time-dependent* imaging measurements and the timing of a clinical event (Kleinbaum and Klein, 2012). Recent work showed that such a joint analysis can reduce bias and increase statistical efficiency by exploiting all available information (Tsiatis and Davidian, 2004).

We demonstrate the proposed method both for the univariate and mass-univariate analysis of imaging measurements automatically computed with FreeSurfer, a widely used brain Magnetic Resonance Image (MRI) data analysis software package (Dale et al., 1999; Fischl, 2012; Fischl and Dale, 2000; Fischl et al., 1999a; Fischl et al., 1999b). We include a quantitative and objective empirical evaluation of the statistical performance of the proposed method based on publicly available data (the Alzheimer’s disease neuroimaging initiative, ADNI¹) from a group of subjects with Mild Cognitive Impairment (MCI) (Gauthier et al., 2006), a clinically defined condition associated with high-risk incipient dementia. Our

¹<http://tinyurl.com/ADNI-main>

experiments revealed that the proposed method offers a substantial increase in statistical efficiency relative to a “two-sample” benchmark method that compares those who convert from MCI to clinical AD against those who remain MCI through follow-up; and a classical Cox regression analysis that employs only baseline scans.

The paper is organized as follows. Section 2.1 and 2.2 review the Cox proportional hazards and linear mixed effects models, respectively. Section 2.3 presents the proposed method that unifies these two frameworks. Section 2.4 describes the alternative analysis strategies that we will use to benchmark our experimental results. Section 2.5 offers a description of the data used in the experiments and section 2.6 details the statistical analyses conducted on these data. In Section 3, we present experimental results that illustrate the proposed joint modeling approach and compare it against benchmarks. Finally, Section 4 provides a discussion of the main experimental findings and Section 5 closes with concluding remarks.

2 MATERIAL AND METHODS

2.1 The Cox Proportional Hazards Model and Its Extension

In this section, we provide a brief overview of the classical Cox proportional hazards model (Cox, 1972) and its extension for time-varying explanatory (independent) variables. For a detailed treatment, the reader is referred to dedicated texts, such as (Kleinbaum and Klein, 2012).

A core component of event time models is the so-called hazard function $h(t)$, which is the instantaneous probability of experiencing the event of interest (e.g., disease onset), given no event up to time t . The hazard function is mathematically defined as:

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t+dt | T \geq t)}{dt},$$

where T is the random variable that represents the time of event and $p(\cdot|\cdot)$ denotes conditional probability. The classical Cox model assumes that the hazard function of a sample with p *time-independent* explanatory variables $X = (X_1, X_2, \dots, X_p)$ can be expressed as:

$$h(t, X) = h_0(t) \exp \left(\sum_{i=1}^p \alpha_i X_i \right), \quad (1)$$

where $h_0(t)$ is the so-called baseline hazard function and $a = (a_1, \dots, a_p)$ are the coefficients associated with the explanatory variables. This model assumes that the hazard function can be written as a product of two factors: one that varies with time but is independent of X , and another that is a function of the *time-independent* explanatory variables X and thus is fixed over time. The foundation of the classical Cox model is the proportional hazards assumption, i.e., the proportion of the hazard functions of two samples is constant over time:

$$\frac{h(t, X^1)}{h(t, X^2)} = \exp \left(\sum_{i=1}^p \alpha_i (X_i^1 - X_i^2) \right) = \text{const.},$$

where X^j is the independent variables of the j 'th sample.

A popular strategy to estimate the coefficients α is the so-called *partial* likelihood maximization method (Cox, 1972, 1975). In particular, the partial likelihood is expressed as a product of K terms, each corresponding to the likelihood of an observed event computed based on the time of occurrence (i.e., there are K events observed during the study and K N , where N is the total number of samples²). Note that, although only observed events are considered, their likelihood values depend on those samples that haven't experienced the event yet and still remain in the study (i.e., have not dropped out or are not "censored" before the particular event time³.)

For readability, let us index the samples such that the first K are those that we have event timing information on. All remaining samples are thus censored. That is, they either drop out of the study before experiencing the event or do not experience the event during their follow-up. Then, mathematically, the partial likelihood is computed as:

$$L = \prod_{k=1}^K L^k, \text{ where } L^k = \frac{h(t^k, X^k)}{\sum_{r \in R^k} h(t^k, X^r)}. \quad (2)$$

In Equation (2), t^k and X^k denote the event time and exploratory variables of the k 'th sample, respectively. R^k is the so-called risk set at t^k , i.e., the set of samples that are *known* to have not experienced the event at t^k . The partial likelihood function of Equation (2) is then maximized with respect to the unknown model parameters α via a numerical optimization strategy, such as Newton-Raphson.

As mentioned above, the original Cox model requires that the explanatory variables be constant over time (e.g., a genetic marker). Hence, this model is not appropriate for analyzing longitudinal imaging measurements that typically vary over time. We note that baseline imaging measurements, however, have been employed in prior neuroimaging studies with standard proportional hazard models (Desikan et al., 2009; Desikan et al., 2010; Devanand et al., 2007; Geerlings et al., 2008; Marcus et al., 2007; Tintore et al., 2008; Vemuri et al., 2011). Although these analyses offer predictive models, their use for inferring associations are restricted to baseline measurements.

The Cox model can easily be extended to handle time-varying variables (Kleinbaum and Klein, 2012). The hazard function is then expressed as:

$$h(t, X, Y(t)) = h_0(t) \exp \left(\sum_{i=1}^p \alpha_i X_i + \sum_{j=1}^q \gamma_j Y_j(t) \right), \quad (3)$$

²In this paper, we assume that each sample can experience the event of interest at most once. There are extended treatments that relax this assumption. We consider these outside the scope of this manuscript.

³We note that the Cox model does not assume that the event will definitely occur for each individual during his or her lifetime, since T can theoretically approach infinity.

where the second term in the exponential includes the effects of q time-varying variables $Y(t) = (Y_1(t), \dots, Y_q(t))$ with associated coefficients $\gamma = (\gamma_1, \dots, \gamma_q)$. Partial likelihood maximization can also be employed to solve the extended Cox model of Equation (3).

Here, we make a crucial observation. To evaluate the partial likelihood function of the extended Cox model, at each event time we need to be able to compute or observe the time-dependent variable for all samples in the risk set, that is those samples *under observation and have not experienced the event yet*. Note that, this means that the value of a time-dependent variable of a subject needs to be identifiable not only for the event time of that particular subject, but for other relevant event times that are prior to the subject's own event/censor time as well.

Longitudinal neuroimaging (LNI) studies offer us the opportunity for identifying the time-varying imaging measurements at different time points. However, in a typical LNI study, image data are acquired at certain intervals (e.g., every six months). Furthermore, these serial scans are usually unbalanced across subjects, i.e., their number and timing can vary between subjects. Finally, the visits for clinical assessments and image acquisitions might not coincide. *Hence, the fundamental challenge of the application of the (extended) Cox model to the analysis of LNI data is the computation (or estimation) of the image data at the times of the clinical events.* Note that, even if we have image data coinciding with a subject's own event time (e.g. disease onset), we typically will not have the corresponding imaging measurements for that subject for other relevant event times observed in the study. *We propose to use a linear mixed effects (LME) model, which captures the temporal trajectories of each individual's imaging measurements, to estimate image data at all observed event times.* The following section provides a brief description of the LME approach and its recently introduced spatial extension.

2.2 Linear Mixed Effects Models for Longitudinal Data

In two recent papers (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b), we illustrated the use of Linear Mixed Effects (LME) models for the analysis of longitudinal neuroimage data. The classical LME model can handle unbalanced data with high inter-subject variability in scan times and missing data points, while offering a parsimonious yet effective strategy to model the mean and covariance structure in longitudinal data (Fitzmaurice et al., 2011; Verbeke and Molenberghs, 2000). The central idea in LME is to allow a subset of the regression parameters to vary randomly across subjects. Hence, the mean trajectory is modeled as a combination of population-level “fixed” effects and subject-specific “random” effects (together, they are called mixed effects).

Formally, the LME model for longitudinal data can be expressed as:

$$Y(t) = \sum_{i=1}^f \beta_i F_i(t) + \sum_{j=1}^r b_j R_j(t) + e, \quad (4)$$

where Y is the time-dependent outcome measurement of a subject, $F(t)$ denotes the value of F at time t , $F = (F_1, \dots, F_f)$'s are the f so-called “fixed” effects that can include subject-level constant variables (e.g., gender, genotype) or time-varying variables (e.g., clinical status,

measurement time, etc.), $R = (R_1, \dots, R_r)$'s are the r "random" effects, which can include a constant bias term and/or a subset of the time-varying fixed effect variables (e.g., measurement time). $\beta = (\beta_1, \dots, \beta_p)$'s and $b = (b_1, \dots, b_r)$'s are the unknown fixed and random effect coefficients, respectively and e is independent and identically distributed zero-mean Gaussian measurement noise with an unknown variance σ^2 . We further assume the random effect vector b is sampled for each subject from a zero-mean Gaussian with an unknown, non-diagonal $r \times r$ covariance matrix D . The unknown model parameters are the measurement variance σ^2 and random effect covariance matrix D . The traditional LME approach solves for the model parameters via an iterative restricted maximum likelihood (ReML) procedure.

Given estimates \hat{D} and $\hat{\sigma}$, we have a closed-form solution for the maximum likelihood (ML) estimate of the fixed effect coefficients $\hat{\beta}$. We can further use these estimates to compute a prediction for the values of the subject-specific random effect coefficients \hat{b}_s , where the subscript s denotes subject index (Fitzmaurice et al., 2011). One can then easily compute an unbiased prediction for the temporal trajectory for subject s as:

$$\hat{Y}_s(t) = \sum_{i=1}^f \hat{\beta}_i F_{si}(t) + \sum_{j=1}^r \hat{b}_{sj} R_{sj}(t). \quad (5)$$

We recently extended the classical LME model to handle spatial data, such as image-wide measurements in a mass-univariate analysis (Bernal-Rusiel et al., 2013b). This approach, called ST-LME, essentially builds on the LME framework but modifies it to model spatial correlations via a parametric spatial covariance matrix. The parameters associated with the spatial matrix are added to the list of unknown model parameters and estimated via ReML. The prediction of subject-level trajectories can then be computed exactly the same way using the estimated model coefficients and Equation (5). Our prior experiments have demonstrated that the ST-LME model offers superior statistical efficiency since it exploits the spatial structure in image data.

2.3 Proposed Strategy for Joint Analysis of Event Time and LNI Data

As discussed above, the extended Cox model requires that the value for the time-dependent variables be specified for each subject in the risk set at each observed event time. In general, many of these imaging measurements are unavailable. However, one can estimate these data based on serial measurements available in a study. We propose to use the LME model to compute these estimates. The LME approach (and its mass-univariate, spatial extension) provides a way to parsimoniously model the spatial and temporal correlation structure in serial imaging data, while accounting for unbalanced longitudinal data collection.

The proposed strategy consists of two steps. In the first step, we fit an LME model to the longitudinal imaging data. The details of model selection, the determination of random effects, and parameter estimation are provided in (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b). Once the model parameters are estimated, one then computes and saves the population-level fixed-effect coefficients and subject-level random-effect coefficients.

In the second step, we fit an extended Cox regression model (Equation 3) to the event time data, where for each observed event time we compute the imaging measurements of each subject based on the LME model fit in the previous step (Equation 5). Statistical inference (i.e., hypothesis testing) on the association between imaging measurements and the event can then be conducted based on the Wald statistic of the approximately standard normal

distribution of $\frac{\hat{\gamma}_i}{\sqrt{\hat{V}\hat{a}r(\hat{\gamma}_i)}}$ (a Z score), which yields a p-value for the effect of interest (Kleinbaum and Klein, 2012). The estimate for $\hat{V}\hat{a}r(\hat{\gamma}_i)$ is computed as the negative of the inverse of the Hessian matrix associated with the ML solution of the extended Cox model.

Here, we would like to make several remarks. First, the LME model of the first step and the extended Cox model of the second step can contain different sets of time-independent variables, although in our experiments we chose to use the same set for both models. This decision was motivated by recent joint frameworks, where models for the event time distribution and longitudinal data are taken to depend on a common set of effects (Tsiatis and Davidian, 2004). We note that the selection of the explanatory variables to be included in a model is a general problem in regression and should be made based on domain knowledge and study constraints. Second, the time-dependent variables of the LME model have to be identifiable for arbitrary time-points, since we need to compute predictions for all relevant observed event times, which in general do not coincide with the imaging times. Common examples for such time-varying variables are time elapsed from baseline, subject age, and functions of these, e.g., time squared. Finally, the predicted image measurements computed at the exact imaging times, in general, will not be equal to the actual measurements themselves. Rather, they can be considered as “denoised” measurements, where the error in longitudinal observations is estimated and discounted via the LME model. This is similar in spirit to recent joint longitudinal and survival models, e.g. (Kleinbaum and Klein, 2012).

2.4 Alternative Methods

To date, the most common approach to perform an analysis between neuroimage data and a clinical event of interest, such as disease onset, relies on a two-group comparison (Borgwardt et al., 2011; Chetelat et al., 2005; Jack Jr et al., 2008a; Julkunen et al., 2009; Morgan et al., 2011; Risacher et al., 2009; Sun et al., 2009). In this method, the subjects are divided into two groups: “converters”, i.e., those who experience the event during follow-up, and “non-converters”, i.e. those who remain clinically stable for a certain amount of follow-up time. In our experiments, we used the two-group approach with a LME model that has been demonstrated to offer excellent statistical power for the analysis of longitudinal neuroimage data (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b). Here, the output variables are the imaging measurements and the analyses examine the differences between the intercepts *and* slope coefficients of the two groups.

As a second alternative, we consider employing the extended Cox model but with a simpler method to estimate the longitudinal trajectories of the imaging data. Recall that the proposed approach fits an LME-based statistical model to longitudinal neuroimage data, which is then utilized to estimate the imaging measurements for all observed event times in the second

step. The LME-based statistical model examines the entire data, in order to characterize and estimate the individual-level temporal trajectories. A more basic strategy would be to fit a linear function independently to each individual's serial imaging data. Note that this alternative approach, while computationally very efficient, ignores the spatial structure in the images. Furthermore, when estimating the individual temporal trajectories, it does not pool information across the population, the way the LME approach does. Hence, we expect the estimates of the temporal trajectories of the imaging measurements to be noisier and thus the inference of the extended Cox model to be statistically less efficient.

Finally, to our knowledge, all prior neuroimaging studies that conducted a Cox regression analysis, simply utilized the baseline scans of each subject. In our experiments, we considered this method as a benchmark as well. However, as we discuss below, this analysis tests a slightly different association.

2.5 The ADNI Data

2.5.1 MRI Processing—We analyzed serial brain MRI data (T1-weighted, 1.5 Tesla), which were acquired and made public by the Alzheimer Disease Neuroimaging Initiative (ADNI). We processed all MRI scans automatically using FreeSurfer (Fischl, 2012) (version 5.1.0, <http://surfer.nmr.mgh.harvard.edu>, specifically its longitudinal processing pipeline (<http://surfer.nmr.mgh.harvard.edu/fswiki/LongitudinalProcessing>) (Reuter and Fischl, 2011; Reuter et al., 2010; Reuter et al., 2012)).

FreeSurfer computes volume estimates for a wide range of brain structures such as the hippocampus, and estimates of the intra-cranial volume (ICV). In all subsequent analyses, we summed the volumes of the left and right hippocampi to obtain the total hippocampal volume (HV). Additionally, for each MRI scan, FreeSurfer automatically computes subject-specific thickness measurements across the entire cortical surface of each cerebral hemisphere. These measurements are further spatially re-sampled onto a standard surface-based template (*fsaverage*), which represents an average brain.

In our experiments we performed both univariate and mass-univariate analyses. Our goal was to detect the association between the longitudinal measurements of neuroimaging biomarkers of AD and clinical progression from MCI to AD. Total hippocampal volume (HV) was the imaging variable of interest in the univariate analyses. The mass-univariate analyses were conducted on cortical thickness data computed across the entire cortex. These two types of measurements were chosen since they have been shown to be strongly associated with progression from MCI to AD (Dickerson et al., 2009; Dickerson et al., 2001; Jack Jr et al., 1997; Lerch et al., 2005). Cortical thickness maps were smoothed by applying an iterative nearest neighbor averaging procedure that approximates Gaussian kernel smoothing on the high resolution surface of FreeSurfer's *fsaverage* template subject (full-width at half max = 15 mm) (Hagler Jr et al., 2006; Han et al., 2006). For computational efficiency, the mass-univariate analyses were conducted on the left hemisphere of *fsaverage6*, which is a lower resolution version of *fsaverage* (FreeSurfer's average template surface) and has about 35k vertices.

2.5.2 Longitudinal Data from MCI Subjects—In our experiments, we analyzed longitudinal imaging and clinical data from the ADNI subjects with MCI (N=374, 75.7 ± 6.7 years, 32.6% female). As can be appreciated from Table 1, there is substantial variation between the timing and number of longitudinal visits across these subjects. In many prior studies, the MCI subjects were sub-divided into two categories: *progressor MCIs* (or converters), i.e., those who convert to clinical AD during follow-up; and *stable MCIs* (or non-converters), i.e., those who do not progress. However, as we have emphasized above, many so-called stable MCIs, in fact, drop out from the study prematurely or might convert beyond the study follow-up. Hence, a better illustration of the MCI-to-AD conversion data is the non-parametric estimate of the cumulative AD diagnosis probability (see Figure 1).

In the analyzed data, there were 160 observed event times (i.e., diagnosis of clinical AD), which was on average 1.32 years from baseline with a standard deviation of 0.77 years. For the remaining 214 MCI subjects, the average censor time, i.e., the final follow-up visit, was on average 2.35 years from baseline with a standard deviation of 1.16 years.

2.6 Statistical Models

2.6.1 Proposed Method—The first step of the proposed method fits an LME model to the longitudinal neuroimage data. Here, two important design decisions need to be made: (1) the specification of time-dependent variables that model the mean temporal trajectory, and (2) the selection of the intercept and/or time-dependent variables that will determine the temporal covariance structure. For further detail, the reader is referred to (Bernal-Rusiel et al., 2013a). In the mass-univariate setting, these model specification/selection questions are particularly challenging due to the large number of tests that need to be conducted. In our previous analyses of the ADNI data (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b), we found that a clinical group-specific linear trajectory was an appropriate model for Alzheimer-associated hippocampal atrophy and cortical thinning during the 4.5-year follow-up period of the ADNI study.

In all reported analyses with the proposed method, the following variables were included as independent (fixed effect) variables: time from baseline, baseline age, sex, APOE genotype status (one if e4 carrier or zero otherwise), interaction between time and APOE genotype status (note that this variable was included based on the evidence that e4 accelerates atrophy during the prodromal phases of AD (Jack Jr et al., 2008b)), and education (in years). For hippocampal volume (HV), we further added ICV to normalize for the confounding effect of head size. Intercept and time from baseline were the only random effect variables in the LME models. For the mass-univariate analyses, we used the recently developed spatial extension of LME, namely ST-LME, with the parameter setting recommended in (Bernal-Rusiel et al., 2013b).

We report the results of the statistical tests that examine the association between LNI data and the timing of MCI-to-AD progression. This test employs an extended Cox model with the aforementioned explanatory variables and the time-dependent imaging measurements (variable of interest) estimated using the LME model fit in the first step. For all the mass-univariate analyses, multiple comparisons were corrected by employing a powerful two-stage adaptive False Discovery Rate (FDR) procedure at q -level=0.05 (Benjamini et al.,

2006). We note that the proposed method tests for the hypothesis that there is an association between the imaging measurements and concurrent AD onset.

2.6.2 Benchmark two-group method—Our two-sample LME-based analyses used the same independent fixed and random effect variables as the proposed method. In addition, these models included a binary clinical group membership variable (1 if the subject converted to from MCI to AD or 0 if subject was stable during follow-up⁴), and the interaction between group membership and time as additional fixed effects. Under the null hypothesis of the extended Cox analysis there is no association between the imaging measurements and MCI-to-AD conversion. The corresponding null hypothesis for a two-group analysis is that the coefficients associated with group membership are zero. In other words, under the null, the progressor and stable MCIs have the *same* intercept *and* slope coefficient. All the reported results for the benchmark two-group method employed an F-test that was based on this null hypothesis.

2.6.3 Alternative Cox Models—We conducted two alternative event time analyses using the Cox proportional hazards model. The first method replaces the first step of the proposed approach with a simple line-fitting scheme. So, instead of fitting an LME based (or ST-LME based) model to the longitudinal imaging data, we fit the best line (in the least squares sense) to the serial measurements of each individual. In the mass-univariate setting, each spatial location was treated independently. We then computed the imaging measurements for each subject and at each observed event time based on these estimated linear trajectories. Given these estimates, the extended Cox regression model and statistical test were identical to the proposed method. The tested hypothesis was identical to that of the proposed model.

In the second method, we ignored the temporal trajectories in the imaging data and simply treated *baseline* measurements as time-independent variables in a classical Cox regression analysis. All remaining explanatory variables were identical to the extended Cox models. The reported results were for the test of association between the baseline image variables and time of event. Here, we caution the reader that the tested hypothesis is in fact somewhat different from the one tested with the longitudinal neuroimage data. Relying on baseline measurements allow us only to test associations between the baseline measurement and the timing of a future event. In contrast, the extended Cox analysis we propose in this paper can be used to test associations between the (time-dependent) imaging marker and clinical event. We return to this issue in the Discussion, where we emphasize the distinction between the two approaches.

3 EXPERIMENTAL RESULTS

3.1 Univariate analysis of hippocampal volume

In our first experiment, we compared the statistical performance of the proposed method with the alternative benchmarks based on detecting the known association between MCI-to-

⁴Stable MCIs were those subjects who were categorized as MCI at baseline and remained so during a clinical follow-up of at least 1 year. Converter MCIs were those MCI subjects who were diagnosed with clinical AD at follow-up.

AD conversion and total hippocampal volume. To achieve this, we utilized an empirical strategy inspired by (Thirion et al., 2007).

For different sample size values ($N = 30-100$), we randomly selected two sets of independent MCI samples, (i.e., two independent samples of size N) from the MCI subjects in the ADNI sample (Total $N=374$). Note that there was no overlap between the two independent samples. We repeated this procedure 1,000 times to obtain 1,000 random pairs of independent MCI samples of a certain size (that is, for a given size we had 2,000 random MCI samples in total).

To assess sensitivity, we computed the detection (true positive) rate across the 2,000 samples for a range of p-value thresholds and sample sizes ($N=30-100$). Here we assumed that the underlying ground truth was that there is an association between hippocampal volume and MCI-to-AD conversion. Instances where the p-value was less than a threshold were considered a “detection” and remaining cases were treated as a false negative. The true positive rate (or sensitivity) was quantified as the fraction of detections.

Figure 2 shows empirical sensitivity as a function of the p-value threshold and sample size. Figure 3 plots repeatability, defined as the rate of detection in both of the independent samples. These results demonstrate that the proposed method offers the highest statistical sensitivity and repeatability for detecting associations between imaging measurements and the clinical event of interest. Yet, as we emphasized above, technically, each method is testing a slightly different hypothesis. In particular, the classical Cox analyses of the baseline measurements are testing associations between these values and the timing of the future diagnosis of AD. The two-group method is testing for differences in the trajectories of imaging measurements between those who convert to AD and those who remain MCI during follow-up. Finally, the extended Cox analyses directly test for the associations between the value of imaging measurements and concurrent risk of AD diagnosis.

The boost in statistical efficiency with respect to the common two-group method is substantial: for a given sample size (e.g. 50) and p-value threshold (e.g. 0.05), the increase in statistical power can be over 10%. These results further illustrate that employing longitudinal imaging data can improve the statistical efficiency of a Cox regression model. Finally, these univariate analyses revealed that the proposed LME-based two-step approach can offer a modest, yet consistent improvement over a more basic extended Cox strategy that uses a simple line-fitting scheme to interpolate the imaging data. This result suggests that the LME-based model, which pools data across subjects, provides improved estimates of the longitudinal trajectories of the imaging measurements.

Finally, to examine the type I error control offered by the Cox models, we conducted an additional analysis. Here, we randomly permuted the event time information within each sample (1,000 times), and repeated the analyses with the three Cox-based methods: (i) a classical proportional hazards model that uses the baseline scans only, extended Cox models, with (ii) a line-based interpolation scheme, and (iii) the proposed LME-based estimation strategy. The random permutations simulate a null hypothesis, where the imaging measurements and the event of MCI-to-AD conversion were statistically independent. Since

the random permutations broke down the relationship between the image data and the timing of MCI-to-AD conversion, we considered these data as samples from the null hypothesis (Good, 2000; Nichols and Holmes, 2002). Then, for each p-value threshold, we computed the detection rate of the association between the event time and imaging measurement, with each method. Under these random permutations, a method with good type I error control should achieve a detection rate that is close to the used theoretical threshold. Table 2 shows that this is indeed the case. All three Cox-based methods achieve type I error rates that are very close to each other and to the theoretical threshold.

3.2 Mass-univariate analysis of cortical thickness

In our second experiment, we exploited the known association between regional cortical thinning and MCI-to-AD conversion (Dickerson et al., 2009) and used an empirical strategy similar to the one in the first experiment. Out of 374 ADNI MCI subjects, we randomly drew independent pairs of samples of 80 subjects. Each independent pair shared no common subject. This way, we generated 1,000 pairs of independent samples (or 2,000 samples in total). For each sample, we used the aforementioned methods to compute significance maps for the association between cortical thickness values and MCI-to-AD conversion. We used the two-stage adaptive FDR procedure with an array of q-values (Benjamini et al., 2006) to control for multiple comparisons. Thus, for each sample, each method and each q-value threshold, we obtained a map of significant associations.

We computed the statistical power (sensitivity) at the sample-level as the fraction of instances (out of the 2,000) where a statistical method detected a significant association at a given FDR q-value (see Figure 4). Next, we assessed repeatability via the overlap area between the two independent MCI samples of size 80 (for FDR q-value = 0.05). Figure 5 shows the means and standard errors across the 1000 random draws for the four methods that were compared in this study. These results demonstrate that the benchmark two-group method offers the least repeatability, while the extended Cox strategy yields a dramatic improvement in the repeatability of the results. The proposed method of using an LME-based first step to capture the spatiotemporal patterns in the neuroimaging data provides a subtle increase in repeatability.

3.3 Analyzing the entire ADNI MCI sample

Finally, we employed the proposed strategy to analyze the association between longitudinal cortical thickness measurements and MCI-to-AD conversion in the entire ADNI MCI sample. Figure 6 presents the map of significant associations. We note that these maps, which include regions associated with the default-mode network, show a striking resemblance to previously reported regions where cortical atrophy was correlated with AD dementia (Buckner et al., 2005; Dickerson et al., 2009). The strong agreement between these results and prior studies increase our confidence in the validity of the proposed method.

4. DISCUSSION

As shown by recent studies, the event time analysis framework can provide a substantial increase in statistical efficiency when examining associations between imaging biomarkers

and a clinical event of interest in a longitudinal study design (Vemuri et al., 2011). In contrast with the more popular two-group comparison method that compares converters and non-converters (i.e., those who experience the event and those who do not), the event time analysis method exploits the variation in the event timing data and, crucially, accounts for finite follow-up. Hence non-converters, i.e., those who are not observed to experience the event of interest, are not treated as a uniform group, separate from the converters, since some of these subjects might eventually experience the event beyond the study follow-up.

The Cox proportional hazards model, which has been employed in neuroimage analysis before (Desikan et al., 2009; Desikan et al., 2010; Devanand et al., 2007; Geerlings et al., 2008; Marcus et al., 2007; Stoub et al., 2005; Tintore et al., 2008; Vemuri et al., 2011), is a very flexible and powerful method for conducting event time analyses. However, this method has only been employed to analyze baseline measurements with respect to the timing of a future event, which restricts the associations we can examine and detect.

An alternative approach is to model the temporal trajectory in image data and use an extended Cox model that handles time-dependent variables. Longitudinal neuroimaging studies offer us this opportunity. However, the fundamental challenge is that in the extended Cox modeling approach, time-varying variables have to be identifiable at all relevant observed event times, and not just the time of event for the corresponding sample. In this paper, we proposed to model the spatiotemporal patterns in neuroimaging measurements using an LME-based approach (Bernal-Rusiel et al., 2013a; Bernal-Rusiel et al., 2013b). The LME model is then used to estimate the image data at observed event times.

We conducted an empirical evaluation of the proposed method, along with three alternative strategies. Each comparison with a benchmark allowed us to quantify the effect of a different factor. The comparison with the popular two-group analysis, which simply contrasts the longitudinal data of converters versus non-converters, reveals the influence of explicit event time modeling achieved with the Cox approach. The comparison with the baseline Cox model, which ignores the longitudinal neuroimage data and simply uses the baseline scans (as done in prior neuroimaging studies), gives us information on the effect of exploiting longitudinal imaging. Finally, the comparison with an alternative extended Cox method that uses a simple line-fitting step to model the temporal trajectory in the imaging measurements, uncovers the impact of the LME-based first step in the proposed approach.

In agreement with prior studies, we found that the Cox proportional hazards approach offers a significant improvement in statistical power and repeatability. For example, the agreement between the maps of two independent MCI samples of 80 subjects was about 50 times greater for the proposed method compared to the two-group benchmark. Similar gains were consistently observed in other univariate and mass-univariate analyses.

Secondly, the proposed method was substantially more powerful and reliable than a classical Cox analysis of the baseline scans. In the mass-univariate experiment, this gain was slightly less than the improvement with respect to the two-group method, yet it was consistent across all analyses. These results highlight the advantage of utilizing and modeling longitudinal imaging data in a Cox analysis. Obviously, this might be impractical in experimental designs

with no serial imaging follow-up. However, to our knowledge, we offer the first discussion of this issue in neuroimaging and propose a strategy that might boost statistical efficiency for detecting associations of interest.

Finally, we can quantify the effect of the LME-based first step in the proposed approach. The LME model attempts to capture the spatiotemporal structure in the neuroimaging data by examining the entire longitudinal sample. In contrast, a simpler strategy would be to estimate the temporal trajectory of each imaging measurement in isolation, without considering other subjects or spatial locations in a mass-univariate analysis. From a theoretical standpoint, we expect the proposed method to yield a more accurate model for the longitudinal data, and hence improve the power of the Cox regression, which relies on the estimates of the imaging data at time-points with no observation. Empirically, we find this is indeed the case. Our experiments suggest that the improvement in statistical power and reliability due to the LME step is subtle, yet consistent. In the mass-univariate setting, the agreement between the maps of two independent samples can increase by over 5% as a result of the LME step.

The extended Cox analysis we advocate in this article has several drawbacks and technical subtleties, as discussed in prior work (Fisher and Lin, 1999). Firstly, the model parameters have no straightforward interpretation as in the classical Cox model, where under the proportional hazards assumption, the estimated coefficients can be interpreted as a constant hazard ratio. Therefore, we are largely constrained to testing the statistical strength of associations, rather than presenting *interpretable* hazard ratios. Secondly, we need to underscore the difference between *internal* versus *external* time-dependent variables, as distinguished by (Kalbfleisch and Prentice, 2011). Internal variables are those that are generated by the studied subject directly (e.g., blood pressure) and are directly related to the event (i.e., the event is defined via this variable or these variables cannot be measured after the event – e.g., blood pressure after death). When dealing with such variables and events, the relationship between the conditional hazard function and survival function breaks down. For example, a measurable value of blood pressure is indicative of the subject being still alive and hence survival at that time point is known. However, in the scenarios we considered in this article, longitudinal imaging measurements, which might be considered as *internal*, do not generally suffer from this technical problem. This is because scans can be and are acquired after the event of interest and the event is not directly defined via imaging measurements. Another issue with time-dependent variables is that usually they do not yield individual predictions of event risk curves. This is because these curves depend on the typically unknown temporal trajectories of the time-dependent variables. Finally, our analysis interrogated the relationship between imaging measurements and concurrent risk of event. Alternative functional forms for this relationship can also be used within the extended Cox analysis. For instance, one could consider the cumulative history of the variable (e.g., history of high blood pressure). One alternative promising approach in neuroimaging might be to consider the concurrent or past slope (rate of change) in imaging measurements as a potential biomarker of the event. This would require reliable estimates of derivatives of the longitudinal imaging trajectory, which could also be obtained from the LME step. We leave the exploration of this direction to future work.

5 CONCLUSIONS

We presented a statistical method for the event time analysis of longitudinal neuroimage data. We have implemented and validated the proposed method for mapping longitudinal brain atrophy effects within the FreeSurfer framework, yet its adaptation to other types of spatial data is straightforward. Our results suggest that the proposed method can offer excellent statistical power for detecting associations between longitudinal imaging measurements and a clinical event, such as disease onset.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the National Institutes of Health (NIH) (grant U01 AG024904). The ADNI is funded by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott Laboratories, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation Plc, Genentech Inc, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson Services Inc, Eli Lilly and Company, Medpace Inc, Merck and Co Inc, Novartis International AG, Pfizer Inc, F. Hoffman-La Roche Ltd, Schering-Plough Corporation, CCBR-SYNARC Inc, and Wyeth Pharmaceuticals, as well as nonprofit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the US Food and Drug Administration. Private sector contributions to the ADNI are facilitated by the Foundation for the NIH. The grantee organization is the Northern California Institute for Research and Education Inc, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. The ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California.

This research was carried out in whole or in part at the Athinoula A. Martinos Center for Biomedical Imaging at the Massachusetts General Hospital, using resources provided by the Center for Functional Neuroimaging Technologies, P41EB015896, a P41 Biotechnology Resource Grant supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), National Institutes of Health. Further support for this research was provided in part by the National Center for Research Resources (P41-RR14075), the National Institute for Biomedical Imaging and Bioengineering (R01EB006758), the National Institute on Aging (AG022381), the National Center for Alternative Medicine (RC1 AT005728-01), the National Institute for Neurological Disorders and Stroke (R01 NS052585-01, 1R21NS072652-01, 1R01NS070963, 2R01NS042861-06A1, 5P01NS058793-03), the National Institute of Child Health and Human Development (R01-HD071664), and was made possible by the resources provided by Shared Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043. Additional support was provided by The Autism & Dyslexia Project funded by the Ellison Medical Foundation, by the NIH 5R01AG008122-22 grant, and by the NIH Blueprint for Neuroscience Research (5U01-MH093765), part of the multi-institutional Human Connectome Project. Dr. Sabuncu received support from an AHAF (BrightFocus) Alzheimer's Disease pilot grant (AHAF A2012333), and an NIH K25 grant (NIBIB 1K25EB013649-01).

Finally, the authors would like to thank Nick Schmansky and Louis Vinke for their efforts in downloading and processing the ADNI MRI scans.

References

- Albin RL, Young AB, Penney JB, Handelin B, Balfour R, Anderson KD, Markel DS, Tourtellotte WW, Reiner A. Abnormalities of striatal projection neurons and N-methyl-D-aspartate receptors in presymptomatic Huntington's disease. *New England Journal of Medicine*. 1990; 322:1293–1298. [PubMed: 1691447]
- Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*. 2006; 93:491–507.
- Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. Statistical Analysis of Longitudinal Neuroimage Data with Linear Mixed Effects Models. *Neuroimage*. 2013a; 66:249–260.
- Bernal-Rusiel JL, Reuter M, Greve DN, Fischl B, Sabuncu MR. Spatiotemporal Linear Mixed Effects Modeling for the Mass-univariate Analysis of Longitudinal Neuroimage Data. *Neuroimage*. 2013b; 81.
- Borgwardt S, McGuire P, Fusar-Poli P. Gray matters: Mapping the transition to psychosis. *Schizophrenia research*. 2011; 133:63–67. [PubMed: 21943556]

- Buckner RL, Snyder AZ, Shannon BJ, LaRossa G, Sachs R, Fotenos AF, Sheline YI, Klunk WE, Mathis CA, Morris JC. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *The Journal of Neuroscience*. 2005; 25:7709–7717. [PubMed: 16120771]
- Chetelat G, Landeau B, Eustache F, Mezenge F, Viader F, de La Sayette V, Desgranges B, Baron JC. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. *Neuroimage*. 2005; 27:934–946. [PubMed: 15979341]
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972:187–220.
- Cox DR. Partial likelihood. *Biometrika*. 1975; 62:269–276.
- Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis* 1.: I. Segmentation and Surface Reconstruction. *Neuroimage*. 1999; 9:179–194. [PubMed: 9931268]
- Desikan RS, Cabral HJ, Fischl B, Guttman CR, Blacker D, Hyman BT, Albert MS, Killiany RJ. Temporoparietal MR imaging measures of atrophy in subjects with mild cognitive impairment that predict subsequent diagnosis of Alzheimer disease. *American Journal of Neuroradiology*. 2009; 30:532–538. [PubMed: 19112067]
- Desikan RS, Cabral HJ, Settecase F, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Salat DH, Fischl B. Automated MRI measures predict progression to Alzheimer's disease. *Neurobiology of aging*. 2010; 31:1364–1374. [PubMed: 20570399]
- Devanand D, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, Rusinek H, Pelton G, Honig L, Mayeux R. Hippocampal and entorhinal atrophy in mild cognitive impairment Prediction of Alzheimer disease. *Neurology*. 2007; 68:828–836. [PubMed: 17353470]
- Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, Grodstein F, Wright CI, Blacker D, Rosas HD. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex*. 2009; 19:497–510. [PubMed: 18632739]
- Dickerson BC, Goncharova I, Sullivan M, Forchetti C, Wilson R, Bennett D, Beckett L, deToledo-Morrell L. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiology of aging*. 2001; 22:747–754. [PubMed: 11705634]
- Fischl B. Freesurfer. *Neuroimage*. 2012
- Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*. 2000; 97:11050.
- Fischl B, Sereno MI, Dale AM. Cortical Surface-Based Analysis* 1.: II: Inflation, Flattening, and a Surface-Based Coordinate System. *Neuroimage*. 1999a; 9:195–207. [PubMed: 9931269]
- Fischl B, Sereno MI, Tootell RBH, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*. 1999b; 8:272–284. [PubMed: 10619420]
- Fisher LD, Lin D. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual review of public health*. 1999; 20:145–157.
- Fitzmaurice, GM.; Laird, NM.; Ware, JH. *Applied longitudinal analysis*. Wiley; 2011.
- Forsberg A, Engler H, Almkvist O, Blomquist G, Hagman G, Wall A, Ringheim A, Langstrom B, Nordberg A. PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiology of aging*. 2008; 29:1456–1465. [PubMed: 17499392]
- Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H. Mild cognitive impairment. *The Lancet*. 2006; 367:1262–1270.
- Geerlings M, Den Heijer T, Koudstaal P, Hofman A, Breteler M. History of depression, depressive symptoms, and medial temporal lobe atrophy and the risk of Alzheimer disease. *Neurology*. 2008; 70:1258–1264. [PubMed: 18391157]
- Good, PI. *Permutation tests*. Wiley Online Library; 2000.
- Hagler DJ Jr, Saygin AP, Sereno MI. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage*. 2006; 33:1093–1103. [PubMed: 17011792]
- Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R. Reliability of MRI-derived measurements of human cerebral cortical thickness: the

- effects of field strength, scanner upgrade and manufacturer. *Neuroimage*. 2006; 32:180–194. [PubMed: 16651008]
- Jack CR Jr, Petersen RC, Grundman M, Jin S, Gamst A, Ward CP, Sencakova D, Doody RS, Thal LJ. Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI. *Neurobiology of aging*. 2008a; 29:1285–1295. [PubMed: 17452062]
- Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, Smith GE, Ivnik RJ, Kokmen E. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*. 1997; 49:786–794. [PubMed: 9305341]
- Jack CR Jr, Vemuri P, Wiste HJ, Weigand SD, Lesnick TG, Lowe V, Kantarci K, Bernstein MA, Senjem ML, Gunter JL. Shapes of the Trajectories of 5 Major Biomarkers of Alzheimer Disease. *Archives of Neurology, archneurol*. 2012 2011.3405 v2011.
- Jack CR Jr, Weigand SD, Shiung MM, Przybelski SA, O'Brien PC, Gunter JL, Knopman DS, Boeve BF, Smith GE, Petersen RC. Atrophy rates accelerate in amnesic mild cognitive impairment. *Neurology*. 2008b; 70:1740–1752. [PubMed: 18032747]
- Julkunen V, Niskanen E, Muehlboeck S, Pihlajamaki M, Kononen M, Hallikainen M, Kivipelto M, Tervo S, Vanninen R, Evans A. Cortical thickness analysis to detect progressive mild cognitive impairment: a reference to Alzheimer's disease. *Dementia and geriatric cognitive disorders*. 2009; 28:404–412. [PubMed: 19907176]
- Kalbfleisch, JD.; Prentice, RL. *The statistical analysis of failure time data*. John Wiley & Sons; 2011.
- Kleinbaum, DG.; Klein, M. *Survival analysis*. Springer; 2012.
- Lerch JP, Pruessner JC, Zijdenbos A, Hampel H, Teipel SJ, Evans AC. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cereb Cortex*. 2005; 15:995–1001. [PubMed: 15537673]
- Marcus KJ, Astrakas LG, Zurakowski D, Zarifi MK, Mintzopoulos D, Poussaint TY, Anthony DC, De Girolami U, Black PM, Tarbell NJ. Predicting survival of children with CNS tumors using proton magnetic resonance spectroscopic imaging biomarkers. *International journal of oncology*. 2007; 30:651. [PubMed: 17273766]
- Morgan V, Riches S, Thomas K, Vanas N, Parker C, Giles S, Desouza N. Diffusion-weighted magnetic resonance imaging for monitoring prostate cancer progression in patients managed by active surveillance. *British Journal of radiology*. 2011; 84:31–37. [PubMed: 21172965]
- Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*. 2002; 15:1–25. [PubMed: 11747097]
- Reuter M, Fischl B. Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage*. 2011; 57:19–21. [PubMed: 21376812]
- Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: A robust approach. *Neuroimage*. 2010; 53:1181–1196. [PubMed: 20637289]
- Reuter, M.; Schmansky, NJ.; Rosas, HD.; Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*. 2012. <http://dx.doi.org/10.1016/j.neuroimage.2012.02.084>
- Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research*. 2009; 6:347. [PubMed: 19689234]
- Sabuncu, MR. A Bayesian Algorithm for Image-based Time-to-event Prediction. *MLMI Workshop, MICCAI*; 2013; 2013.
- Sabuncu MR, Buckner RL, Smoller JW, Lee PH, Fischl B, Sperling RA. The association between a polygenic alzheimer score and cortical thickness in clinically normal subjects. *Cerebral Cortex*. 2012; 22:2653–2661. [PubMed: 22169231]
- Sabuncu MR, Desikan RS, Sepulcre J, Yeo BTT, Liu H, Schmansky NJ, Reuter M, Weiner MW, Buckner RL, Sperling RA. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Archives of neurology*. 2011; 68:1040. [PubMed: 21825241]
- Stoub T, Bulgakova M, Leurgans S, Bennett D, Fleischman D, Turner D. MRI predictors of risk of incident Alzheimer disease A longitudinal study. *Neurology*. 2005; 64:1520–1524. [PubMed: 15883311]

- Sun D, Phillips L, Velakoulis D, Yung A, McGorry PD, Wood SJ, van Erp TG, Thompson PM, Toga AW, Cannon TD. Progressive brain structural changes mapped as psychosis develops in at risk individuals. *Schizophrenia research*. 2009; 108:85–92. [PubMed: 19138834]
- Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline JB. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*. 2007; 35:105–120. [PubMed: 17239619]
- Tintore M, Rovira A, Rio J, Tur C, Pelayo R, Nos C, Tellez N, Perkal H, Comabella M, Sastre-Garriga J. Do oligoclonal bands add information to MRI in first attacks of multiple sclerosis? *Neurology*. 2008; 70:1079–1083. [PubMed: 17881717]
- Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*. 2004; 14:809–834.
- Vemuri P, Weigand SD, Knopman DS, Kantarci K, Boeve BF, Petersen RC, Jack CR Jr. Time-to-event voxel-based techniques to assess regional atrophy associated with MCI risk of progression to AD. *Neuroimage*. 2011; 54:985–991. [PubMed: 20832487]
- Verbeke, G.; Molenberghs, G. *Linear mixed models for longitudinal data*. N.Y: Springer; 2000.
- Whitfield-Gabrieli S, Thermenos HW, Milanovic S, Tsuang MT, Faraone SV, McCarley RW, Shenton ME, Green AI, Nieto-Castanon A, LaViolette P. Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences*. 2009; 106:1279–1284.

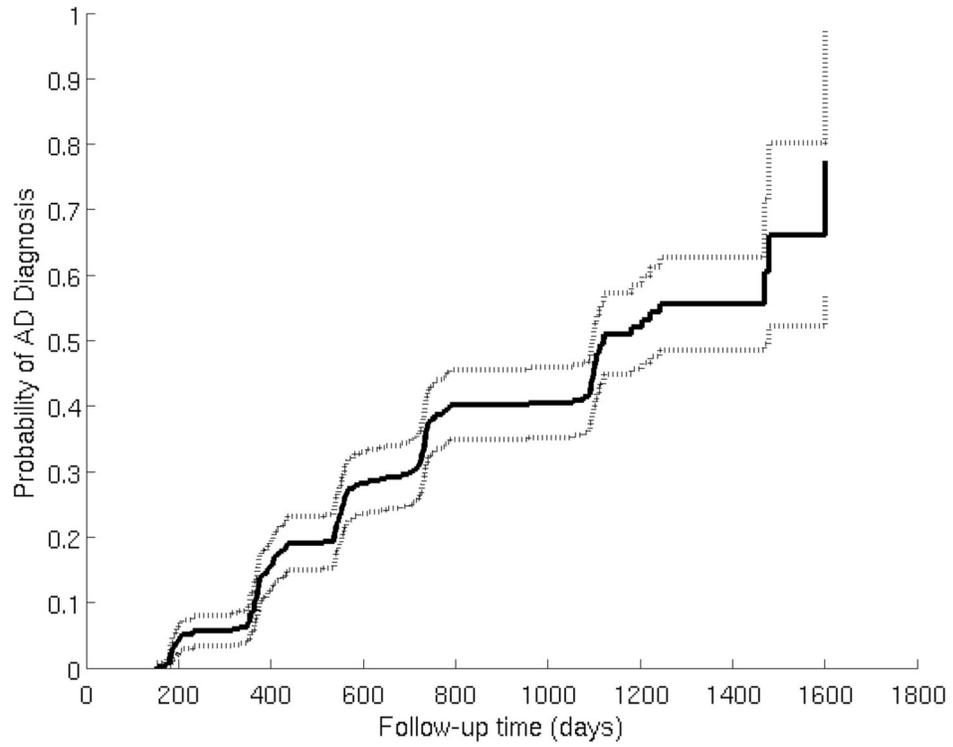
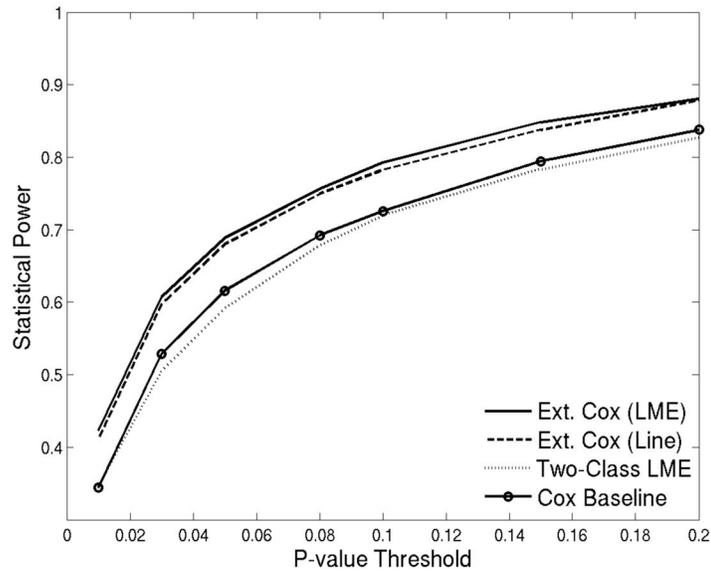
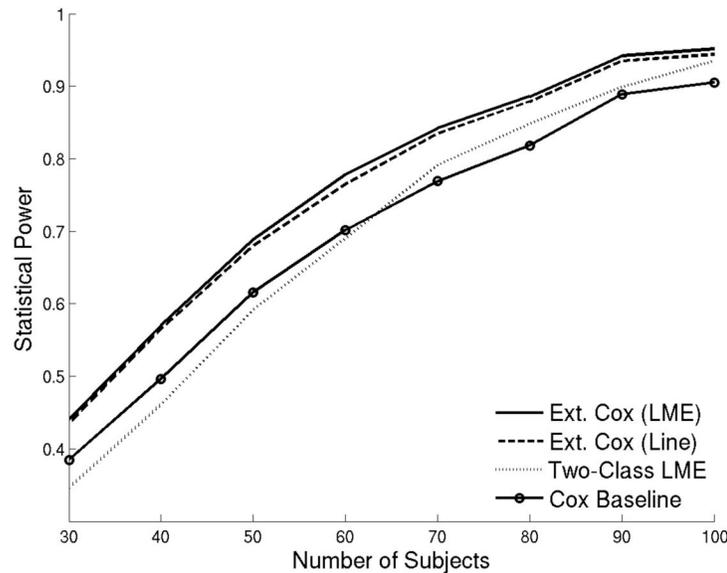


Figure 1. Non-parametric estimate of the AD diagnosis cumulative distribution function for the ADNI MCI subjects. The dashed lines show the 95% confidence interval.



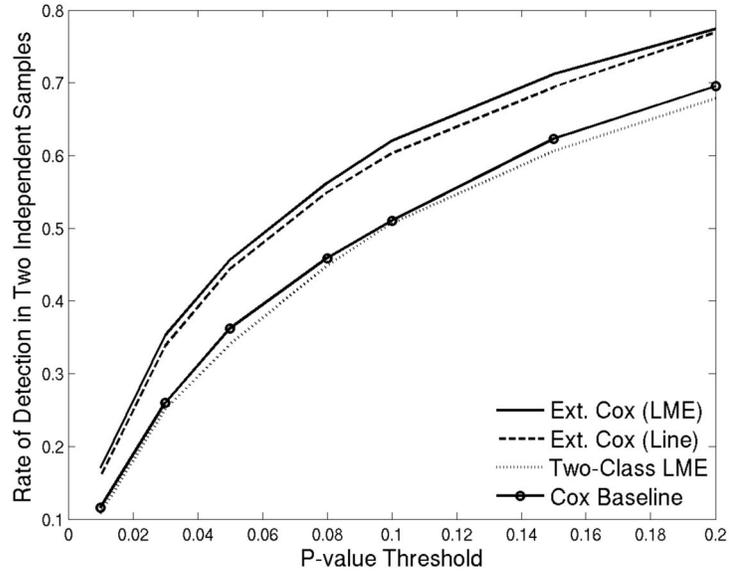
(a)



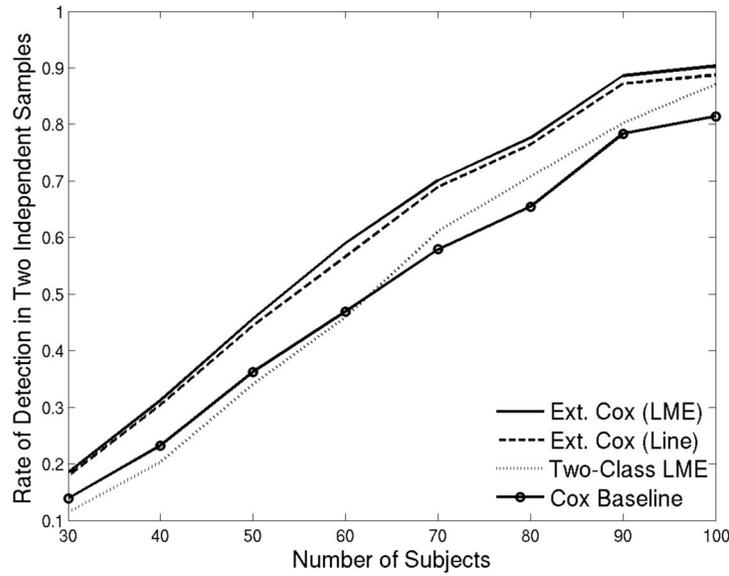
(b)

Figure 2.

Empirical sensitivity (statistical power) as a function of (a) the p-value threshold ($N=50$) and (b) sample size (p-value threshold = 0.05) for detecting the association between total hippocampal volume, a *univariate marker*, and MCI-to-AD conversion. The proposed method (Ext. Cox with LME) yields the most statistical power. Ext. Cox (Line) replaces the LME-based first step of the proposed method with a simple line fit. Two-Class LME is the popular approach of comparing converter MCIs with stable MCIs. Cox Baseline uses only the baseline scans and treats imaging measurements as time-independent exploratory variables in a classical Cox regression. See text for further details.



(a)



(b)

Figure 3. Repeatability (the frequency at which a method detects an association between total hippocampal volume, a *univariate* marker, and MCI-to-AD conversion in two independent samples) as a function of: (a) the p-value threshold (N=50), (b) sample size (p-value threshold=0.05). See caption of Figure 2 for further details.

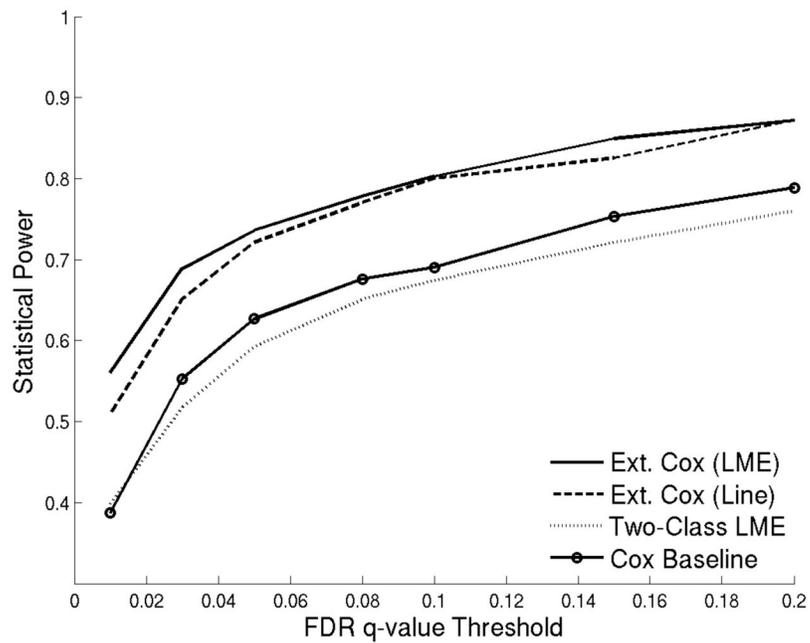


Figure 4.

Empirical statistical power as a function of FDR q-value for detecting the association between *mass-univariate* cortical thickness measurements and MCI-to-AD conversion in groups of 80 MCI subjects. The proposed method (Ext. Cox with LME) yields the most statistical power. Ext. Cox (Line) replaces the LME-based first step of the proposed method with a simple line fit. Two-Class LME is the popular approach of comparing converter MCIs with stable MCIs. Cox Baseline uses only the baseline scans and treats imaging measurements as time-independent exploratory variables in a classical Cox regression. See text for further details.

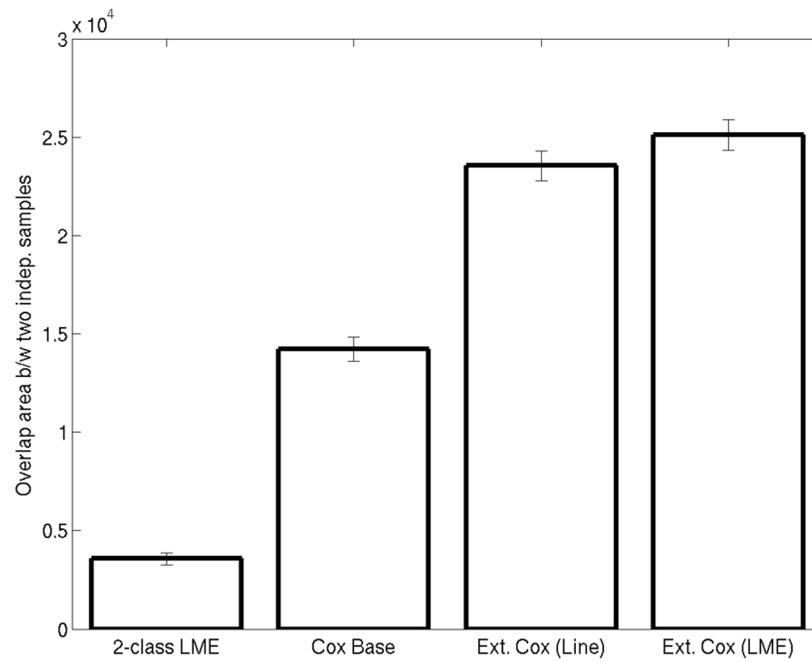


Figure 5.

Average overlap area (in mm²) between the detection maps obtained from two independent samples consisting of 80 MCI subjects. The detection maps were binary masks, where the FDR-corrected statistical association (q -value = 0.05) between cortical thickness and MCI-to-AD conversion was recorded. Vertices that exhibited a significant association in both independent samples were included in the overlap area. See caption of Figure 4 and text for further details. Errorbars show the standard error of the mean.

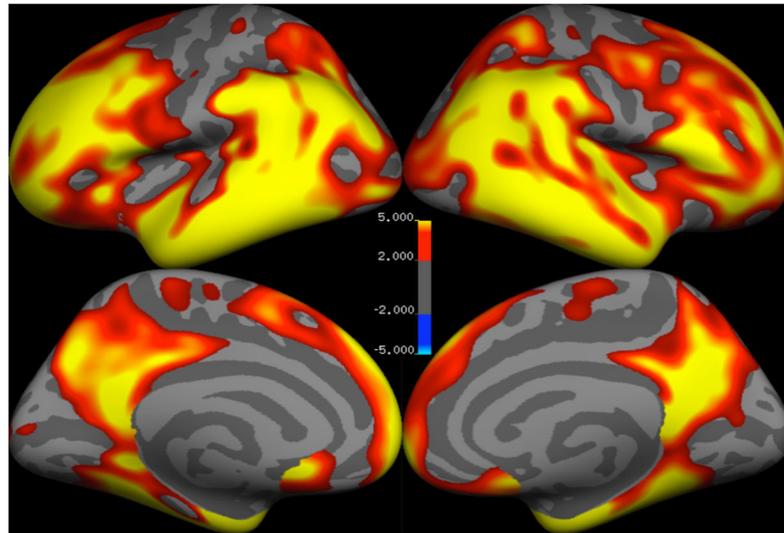


Figure 6.

Brain regions where there is an association between cortical thickness and MCI-to-AD conversion, as determined by the proposed method and based on analyzing the entire longitudinal ADNI MCI sample. Results are overlaid on an inflated representation of the cortex of an average human brain (fsaverage). The gray pattern reflects the underlying cortical folding. Colored regions are where there is a significant association (uncorrected $P < 0.01$). The color reflects the significance of the association, which was computed as $-\log_{10}(P\text{-value})$ (see color bar). The sign encodes the direction of association, where a positive value suggests that thinner cortex at that particular location is associated with an earlier date of onset of AD dementia. Top and bottom rows show views of the lateral and medial surfaces, respectively. Left and right panels correspond to the left and right hemispheres, respectively.

Table 1
The ADNI MCI Cohort

Number and timing of analyzed longitudinal MRI scans per visit

Visit	Number of Individual Scans	Time from baseline (years)
Baseline	374	0
year 0.5 (month 6)	354	0.58 ± 0.07 [0.32–0.94]
year 1	319	1.08 ± 0.07 [0.82–1.35]
year 1.5	281	1.59 ± 0.07 [1.26–1.92]
year 2	210	2.08 ± 0.08 [1.70–2.52]
year 3	100	3.09 ± 0.08 [2.98–3.45]
year 4	13	4.10 ± 0.10 [3.98–4.38]
Total	1651	

Time from baseline (in years) is in mean ± standard deviation; Ranges are listed in square brackets.

Table 2
Empirical type I error rate for the three Cox-based event time models and different p-value thresholds

The null hypothesis was simulated by randomly permuting the event time data.

Theoretical P-value Threshold	0.01	0.03	0.05	0.10
Cox with Baseline Scans	0.00	0.03	0.06	0.11
Extended Cox with Line-based Interp.	0.01	0.03	0.06	0.12
Extended Cox with LME	0.01	0.03	0.06	0.11