IISE Transactions on Healthcare Systems Engineering

ISSN: 2472-5579 (Print) 2472-5587 (Online) Journal homepage: http://www.tandfonline.com/loi/uhse21

A sequential tree-based classifier for personalized biomarker testing of Alzheimer's disease risk

Bing Si, Igor Yakushev & Jing Li

To cite this article: Bing Si, Igor Yakushev & Jing Li (2017) A sequential tree-based classifier for personalized biomarker testing of Alzheimer's disease risk, IISE Transactions on Healthcare Systems Engineering, 7:4, 248-260, DOI: 10.1080/24725579.2017.1367979

To link to this article: https://doi.org/10.1080/24725579.2017.1367979

Accepted author version posted online: 24 Aug 2017.
Published online: 04 Oct 2017.

IISE TRANSACTIONS ON HEALTHCARE SYSTEMS ENGINEERING
2017, VOL. 7, NO. 4, 248–260
https://doi.org/10.1080/24725579.2017.1367979

A sequential tree-based classifier for personalized biomarker testing of Alzheimer's disease risk

Bing Si[a], Igor Yakushev[b], and Jing Li[a] for the Alzheimer's Disease Neuroimaging Initiative[1]

[a]Industrial Engineering, School of Computing Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA; [b]Nuclear Medicine Department, Technische Universität, München, Germany

ABSTRACT

Using baseline biomarkers to predict the conversion of mild cognitive impairment (MCI) to Alzheimer's disease (AD) has considerable clinical interest in recent years. The existing studies have several limitations, including unsatisfactory accuracy due to MCI heterogeneity, use of conventional classification models that require biomarkers to be measured all at once instead of sequentially and as needed, and use of raw numerical measurement of the biomarkers instead of discretized levels that are more robust to measurement errors and provide convenience for clinical utilization. To tackle these limitations, we propose a novel sequence tree-based classifier (STC) for predicting the conversion of MCI to AD. Different from conventional classification models, STC achieves a sequential, as-needed use of biomarkers and a three-category classification (high-risk converter, low-risk converter, and inconclusive diagnosis) by finding an optimal sequence of biomarkers and two-sided cutoffs of each biomarker that satisfy pre-specified accuracy requirements while minimizing the proportion of inconclusive diagnosis. STC is also a personalized approach, as it allows patient characteristic variables to be included to help identify patient-specific cutoffs for each biomarker. We apply STC to two important clinical applications using the data from the worldwide Alzheimer's Disease Neuroimaging Initiative project: prediction of MCI conversion and patient selection for AD-related clinical trials.

ARTICLE HISTORY
Received May 2016
Accepted August 2017

KEYWORDS
Classification; sequential decision making; biomarker; Alzheimer's disease; personalized medicine

1. Introduction

Alzheimer's disease (AD) is an irreversible neurodegenerative disease of the brain characterized by debilitating impairment in daily activities and cognitive decline. More than five million people in the US currently have AD, and the number is expected to increase to 16 million by 2050. The direct healthcare cost is over $200 billion per year and projected to reach $1.2 trillion by 2050. Recent clinical trials designed to treat AD at the mild- to moderate-dementia phase have been largely unsuccessful. There is a growing consensus that treatment should target the disease in its early phases before irreversible brain damage occurs. Mild cognitive impairment (MCI) is a prodromal phase of AD at which patients experience cognitive decline but have not developed dementia. Treatment at the MCI phase could potentially delay the progression to AD or even prevent the patient from developing AD, and therefore has considerable interest.

Important to early detection and prevention of AD is the use of biomarkers to precisely predict the conversion of MCI to AD within a clinical time of interest. According to the new diagnostic guidelines recommended by the National Institute on Aging and the Alzheimer's Association (Albert *et al.*, 2011), the important biomarkers include those measuring Aβ deposition in plagues and those linked to downstream neuronal degeneration or injury processes, such as the phosphorylated tau (p-tau) level in cerebrospinal fluid (CSF), mean cerebral metabolism on ^{18}F fludeoxyglucose positron emission tomography (FDG-PET), and hippocampal volume on structural magnetic resonance imaging (MRI).

There has been a vast amount of studies aiming at using biomarker data to predict the conversion of MCI patients to AD (Barnes *et al.*, 2014; Cui *et al.*, 2011; Heister *et al.*, 2011; Hinrichs *et al.*, 2011; Jack *et al.*, 2010 a; Misra *et al.*, 2009; Risacher *et al.*, 2009; Wee *et al.*, 2013; Ye *et al.*, 2012; Yu *et al.*, 2012; Zhang *et al.*, 2012a, 2012b). A particular area of study with clear clinical relevance is to achieve this prediction using *baseline* biomarker measurements (Barnes *et al.*, 2014; Cui *et al.*, 20122; Heister *et al.*, 2011; Jack *et al.*, 2010 a; Risacher *et al.*, 2009; Wee *et al.*, 2013; Ye *et al.*, 2012; Yu *et al.*, 2012; Zhang *et al.*, 2012b). Although using longitudinal repeated measurements of the same biomarkers has a potential to improve the prediction accuracy, this prolongs the diagnostic time span and makes clinical trials more time-consuming and costly. In using baseline biomarkers to predict MCI conversion, most of the existing studies built statistical classification models that assign each MCI patient to be a converter or non-converter using a pre-trained model. The accuracy on large public datasets like the Alzheimer's Disease Neuroimaging Initiative (ADNI) has been reported to be

CONTACT Jing Li jing.li.8@asu.edu Industrial Engineering, Arizona State University, 699 S Mill Ave., Tempe, AZ 85281, USA.

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

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhse.

between 60% and 72%. However, the existing research has a few limitations.

First, the prediction accuracy is unsatisfactory. This can be attributed to the heterogeneity of MCI patients. That is, there may be subgroups across which different biomarkers or different combinations of biomarkers are useful for predicting conversion to AD. MCI heterogeneity is a known challenge in AD studies and has been reported in many papers (Cerami *et al.*, 2015; Yu *et al.*, 2012). A recent study using the comprehensive dataset collected through the worldwide ADNI project revealed that there is little agreement in using different biomarkers for predicting the conversion of MCI to AD, such as the p-tau level in CSF, mean cerebral metabolism on FDG-PET, and hippocampal volume on MRI. Conflicting predictions by the different biomarkers happen in roughly every third MCI patient (Alexopouos *et al.*, 2014). This provides strong evidence that MCI is a heterogeneous group and that the existing research of "one-model-fits-all" (OMFA) is unlikely to work well. Here, OMFA means building one classification model, which assumes the same multivariate association of biomarkers with conversion/non-conversion, across all of the MCI patients.

Second, the existing research is bounded by an inherent limitation of conventional classification models that require the biomarkers to be measured *all at once*. This is because a conventional classification model takes the form of $Y = f(X_1, \ldots, X_p)$, where $X_1, \ldots, X_p$ are biomarkers and Y is a binary variable of conversion or non-conversion. When using this model to predict an MCI patient, data on all of the biomarkers included in the model—i.e., $X_1, \ldots, X_p$—must be available. Otherwise, the model cannot be applied. Almost all of the commonly used classification models have this limitation, such as logistic regression, discriminate analysis, support vector machine, and artificial neural network. However, requiring biomarkers to be available all at once at the time of making a prediction/diagnosis does not reflect the reality of clinical practices in which biomarkers are typically measured *sequentially*. That is, the most predictive biomarker is first tested for a patient. If the result is conclusive—e.g., the patient is predicted to be a converter or non-converter with a high confidence—no other biomarkers need to be tested. Otherwise, if the result from the first biomarker is inconclusive, an additional biomarker may be tested. More biomarkers may be added until a conclusive diagnosis is reached. It is also possible that no conclusive diagnosis can be reached, even with all of the biomarkers having been tested, which is common for early stages of a disease. If this happens, the patient will be asked to come back to re-test during a follow-up visit.

Lastly, in most existing research that uses biomarkers to predict MCI conversion, biomarkers are treated as numerical variables. Although the raw biomarker measurement is on a numerical scale, clinical interpretation is typically based on a cutoff that dichotomizes the biomarker into "positive" and "negative." For example, 1.21, 3260 mm^3, and 23 pg/mL are the currently used clinical cutoffs for the mean cerebral metabolism on FDG-PET, hippocampal volume on MRI, and p-tau level in CSF, respectively (Jack *et al.*, 2008; Kim *et al.*, 2011). Both approaches have limitations. Using the raw, numerical measurement of biomarkers is clinically inconvenient. Also, there may be measurement errors associated with the testing instrument and bias due to patient's health condition and exposure to environmental factors that potentially confound with the target disease. This makes the use of raw biomarker measurement a less robust approach. On the other hand, using a single cutoff, as in the current clinical practice, is an oversimplification by ignoring the quantitative relationship between biomarker values and disease risks. Between using the numerical measurement and a single cutoff, a "middle" approach that uses more discretized levels of a biomarker may be more appropriate.

To overcome the aforementioned limitations of the existing research, we propose a sequential tree-based classifier (STC) for predicting MCI patients' risks of converting to AD in this article. Compared with conventional classification models, STC does not require all of the biomarkers to be available for every patient at the time of the prediction, but sequentially adds biomarkers only when necessary. Another difference is that, unlike conventional classification models that enforce a binary decision (conversion vs. non-conversion) for each patient, STC classifies patients into three categories: a clinically defined high-risk (HR) category, a clinically defined low-risk (LR) category, and an inconclusive category. The HR and LR categories include MCI patients who will convert to AD within a clinical time of interest with a high and a low probability; e.g., 80% and 20%, respectively. HR patients need immediate medical attention. LR patients can be cleared of the disease or put on long-term observation. Patients falling into the inconclusive category at the baseline may be asked for a re-test in a follow-up visit. In essence, STC achieves the sequential, as-needed use of biomarkers and the three-category classification by finding an optimal sequence of biomarkers and two-sided cutoffs of each biomarker that satisfy the HR and LR requirement while minimizing the proportion of MCI patients classified as inconclusive. Also, STC is personalized because it allows patient-specific information such as age, gender, education level, and genotyping to be included to help identify patient-specific cutoffs for each biomarker. Additionally, STC is flexible in the sense that it can be developed depending on the available biomarkers in a clinic. Each clinic has a different level of resources, which may limit its biomarker testing capability. A model has limited use if it has to assume the same biomarkers to be tested across different clinics. Finally, we would like to stress that STC approaches the challenge of low accuracy in predicting MCI conversion, which is faced by the existing research, from a different angle. That is, a target prediction accuracy is first defined, which is reflected by HR and LR, and it is then used by STC for identifying groups of patients for which this accuracy can be reached. This capability has tremendous value for disease management and patient selection in clinical trials.

We apply STC to two important clinical applications using the ADNI data. One is to predict/classify MCI patients into HR, LR, or inconclusive categories so that appropriate medical decisions can be made for each patient. The other application is to help patient selection in clinical trials—i.e., identify a subcohort of MCI patients with a HR of converting to AD—as these patients are more likely to benefit from the intervention being tested. The remainder of this article is organized as follows: Section 2 provides a literature review of the statistical methods used for prediction of MCI conversion to AD. Section 3 presents the formulation, estimation, and algorithm of the proposed STC model. Section 4 presents the application. Section 5 concludes the article.

2. Literature review

One of the most prominent findings on AD is that AD patients have significant hippocampal atrophy that can be seen on an MRI scan. Because of this, abundant research has been devoted to using MRI imaging data for prediction of MCI conversion to AD. Risacher *et al.* (2009) analyzed MRI data using voxel-based morphometry and automated parcellation methods, and identified the degree of neurodegeneration in medial temporal structures as the best antecedent MRI marker of imminent conversion, with decreased hippocampal volume being the most robust. Zhang *et al.* (2012b) applied a logistic regression model on MRI imaging, and found that combining a medical temporal lobe atrophy scale (MTAS) and a brain atrophy and lesion index (BALI) results in an improved predictive accuracy for MCI conversion. Wee *et al.* (2013) proposed a novel approach to extract correlative morphological information from MRI, and demonstrated that combining this information with the conventional ROI-based information via multi-kernel support vector machines improves the prediction of MCI conversion.

Due to the complicated nature of MCI, it has been acknowledged that using MRI data alone may not suffice. As a result, abundant research has been done to integrate MRI with other data sources such as CSF measurement, cognitive test scores, and functional imaging like FDG-PET. Barnes *et al.* (2014) proposed a point-based risk score for prediction of MCI conversion, which combines MRI hippocampal subcortical volume and middle temporal cortical thinning together with the scores from several cognitive test instruments. Heister *et al.* (2011) used a cox proportional hazard model to predict MCI conversion, which integrated medial temporal atrophy measured by MRI, CSF biomarker levels, and the degree of learning impairment measured by the Rey Auditory Verbal Learning Test. Jack *et al.* (2010a) proposed to integrate hippocampal volumes on MRI with CSF Aβ42 levels and Pittsburgh compound B PET measures in prediction of time-to-conversion from MCI to AD. Ye *et al.* (2012) proposed a sparse learning model that integrated 15 features from MRI scans, cognitive measures, and APOE genotype.

Because multi-source data have been used in prediction of MCI conversion, there is a growing interest in evaluating which source carries the most weight. Toward this end, a number of comparative studies have been performed. Landau *et al.* (2010) compared APOE ϵ4 allele frequency, CSF measurement, FDG-PET, hippocampal volume on MRI, and episodic memory performance at baseline. Their result showed that FDG-PET and episodic memory best predicted MCI conversion to AD. Cui *et al.* (2011) compared MRI morphometry features, CSF measurement, and neuropsychological and functional measures (NMs). Their result showed that NMs outperformed CSF and MRI features. Yu *et al.* (2012) compared MRI, FDG-PET, and CSF measurement, and found that MRI measures had the best predictive power. Overall, the existing comparative studies reached inconsistent conclusions regarding the relative importance of different data sources. The inconsistency might be caused by the difference in the subject pools included in each study and in the statistical methods used for the data analysis. Another possible reason may be the inherent heterogeneity of the MCI population. However, almost all of the studies reached the same conclusion that integrating multi-source information yields a significantly better accuracy than using a single data source alone.

In addition to the aforementioned studies using baseline data, longitudinal data have also been used for MCI prediction. Zhang *et al.* (2012a) developed a longitudinal feature selection method to jointly select brain regions across multiple time points and proposed a multi-kernel support vector machine for MCI prediction based on MRI, FDG-PET, and cognitive scores. Misra *et al.* (2009) investigated baseline and longitudinal patterns of brain atrophy in MCI patients, and found MCI converters displayed significantly lower volume in a number of white matter and grey matter regions. Hinrichs *et al.* (2011) developed predictive markers for MCI conversion using a multi-kernel learning (MKL) framework.

3. Proposed method: A sequential tree-based classifier (STC)

3.1. Formulation of STC

Suppose there are p biomarkers $\mathbf{X} = \{X_1, \ldots, X_p\}$, q patient characteristic variables/risk factors $\mathbf{Z} = \{Z_1, \ldots, Z_q\}$, and a binary diagnostic outcome Y. For example, in diagnosing/predicting the conversion of MCI to AD, commonly used biomarkers include the p-tau level in CSF, mean cerebral metabolism on FDG-PET, and hippocampal volume on MRI, referred to as P-tau, FDG-PET, and MRI hereafter. Risk factors may include age, education level, and status of APOE e4 gene (Liu *et al.*, 2013). $Y = 1$ if an MCI patient converts to AD within a clinical time of interest and $Y = 0$ otherwise. Our objective is to find a testing sequence for the biomarkers as well as a lower and an upper cutoff value for each biomarker adjusted for patient difference in terms of the risk factors, in order to classify patients into a HR, a LR, or an inconclusive category.

First, we focus on a less complicated problem in which the sequence of biomarkers is given. Without loss of generality, assume the sequence to be $X_1 \to X_2 \to \cdots \to X_p$. Also assume a positive correlation between each biomarker and the disease risk; i.e., a higher value of a biomarker means a higher risk of the disease. Although negative correlations exist for some biomarkers, we can always turn the correlations into positive by transforming the biomarkers. This assumption was made for simplicity of the subsequent discussion. We would like to sequentially find two cutoffs for each biomarker. That is, we would like to first find a lower and an upper cutoff for X_1, $l_1(\mathbf{Z})$ and $u_1(\mathbf{Z})$, which are functions of the risk factors $\mathbf{Z}$, such that a patient will have a HR of having the disease if $X_1 \geq u_1(\mathbf{Z})$, a LR if $X_1 \leq l_1(\mathbf{Z})$, and be inconclusive otherwise. HR and LR patients will need no more biomarker testing. Inconclusive patients will be further tested for the second biomarker X_2. Therefore, we will need to find a lower and an upper cutoff for X_2, $l_2(\mathbf{Z})$ and $u_2(\mathbf{Z})$, such that an inconclusive patient from the previous biomarker testing will have a HR of having the disease if $X_2 \geq u_2(\mathbf{Z})$, a LR if $X_2 \leq l_2(\mathbf{Z})$, and continuously be inconclusive otherwise. The inconclusive patients at the current step will be further tested for X_3. This process will continue until all of the biomarkers have been tested.

In a mathematically rigorous way, we can formulate the i-th step of the previous process as follows: Let D_{i-1} be the cohort

of inconclusive patients from the previous step. The goal of the i-th step is to find $l_i(\mathbf{Z})$ and $u_i(\mathbf{Z})$ for X_i that:

$$\min_{l_i(\mathbf{Z}),u_i(\mathbf{Z})} \quad p\left(l_i(\mathbf{Z}) \le X_i \le u_i(\mathbf{Z}) \,|D_{i-1}\right) \tag{1}$$

$$s.t. \quad p\left(Y=1|X_i \ge u_i(\mathbf{Z}), \mathbf{Z}, D_{i-1}\right) \ge r_h$$
$$p\left(Y=1|X_i \le l_i(\mathbf{Z}), \mathbf{Z}, D_{i-1}\right) \le r_l. \tag{2}$$

The objective function is to minimize the proportion of inconclusive patients. This is important to patients by reducing the need and the associated cost and waiting time for another biomarker testing before a conclusive diagnosis can be made. It is also important to the clinic by reducing the overall cost, including the labor and resource spent on the diagnosis. r_h and r_l are clinically defined HR and LR thresholds, respectively, and are typically known for specific applications. For example, in diagnosis, r_h is typically 80–85% and r_l 10–20%. r_h is not necessarily equal to $1-r_l$. Proposition 1 shows that the optimization problem in (Eq. (1)) is equivalent to two simpler sub-optimization problems.

Proposition 1: Let $\tilde{l}_i(\mathbf{Z})$ and $\tilde{u}_i(\mathbf{Z})$ denote the optimal solutions to (1). Let $l_i^*(\mathbf{Z})$ and $u_i^*(\mathbf{Z})$ be the optimal solutions to the optimization problems in Eqs. (2) and (3), respectively.

$$l_i^*(\mathbf{Z}) = \begin{cases} \max\limits_{l_i(\mathbf{Z})} & l_i(\mathbf{Z}) \\ s.t. & p\left(Y=1|X_i \le l_i(\mathbf{Z}), \mathbf{Z}, D_{i-1}\right) \le r_l \end{cases}. \tag{3}$$

$$u_i^*(\mathbf{Z}) = \begin{cases} \min\limits_{u_i(\mathbf{Z})} & u_i(\mathbf{Z}) \\ s.t. & p\left(Y=1|X_i \ge u_i(\mathbf{Z}), \mathbf{Z}, D_{i-1}\right) \ge r_h \end{cases}. \tag{4}$$

Then, $l_i^*(\mathbf{Z}) = \tilde{l}_i(\mathbf{Z})$ and $u_i^*(\mathbf{Z}) = \tilde{u}_i(\mathbf{Z})$. (Proof skipped.)

Proposition 1 implies that $l_i^*(\mathbf{Z})$ can be obtained by first finding the feasible region of $l_i(\mathbf{Z})$, following which the $l_i^*(\mathbf{Z})$ can be naturally obtained by using the maximum value in that region. The same implication applies to $u_i^*(\mathbf{Z})$.

Furthermore, to facilitate identification of the feasible region for $u_i(\mathbf{Z})$, we apply Bayes' rule to the constraints in Eqs. (4) and (3) and get

$$\frac{1-\varphi_{X_i|Y=1,\mathbf{Z}=\mathbf{z}}\left(u_i(\mathbf{Z})\right)}{1-\varphi_{X_i|Y=0,\mathbf{Z}=\mathbf{z}}\left(u_i(\mathbf{Z})\right)} \ge \frac{r_h}{1-r_h} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}, \text{ and} \tag{5}$$

$$\frac{\varphi_{X_i|Y=1,\mathbf{Z}=\mathbf{z}}\left(l_i(\mathbf{Z})\right)}{\varphi_{X_i|Y=0,\mathbf{Z}=\mathbf{z}}\left(l_i(\mathbf{Z})\right)} \le \frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}, \tag{6}$$

respectively. D_{i-1} was dropped for notation simplicity. $\varphi_{X_i|Y,\mathbf{Z}}(x)$ denotes the cumulative distribution function (CDF) of X_i given Y and $\mathbf{Z}$. $\pi(\mathbf{Z}) = p(Y=1|\mathbf{Z})$ is the prior of Y before the biomarker X_i is tested. In Eqs. (5) and (6), r_h and r_l are given constants. $\pi(\mathbf{Z})$ can be known from population statistics; i.e., the probability for people with a certain demographic profile (e.g., female, older than 65, and APOE e4 carrier) to have the disease. Therefore, the key to identifying the feasible regions of $l_i(\mathbf{Z})$ and $u_i(\mathbf{Z})$ is to know the distribution of $X_i|Y,\mathbf{Z}$. Because biomarkers are typically measured on a continuous scale, we assume a Gaussian distribution for $X_i|Y,\mathbf{Z}$. Note that even though the distribution of a biomarker may not be strictly Gaussian, we can apply Box-Cox transformation (Barnes *et al.*, 2014) to make it approximately Gaussian. Under the Gaussian distribution, we can further link X_i and Y, $\mathbf{Z}$ by a linear model; i.e.,

$$X_i = \beta_{0,i} + \beta_{y,i}Y + \beta_{z,i}^T\mathbf{Z} + \varepsilon_i, \tag{7}$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$. Then, $\varphi_{X_i|Y,\mathbf{Z}}(x)$ becomes $\Phi\left(\frac{x-(\beta_{0,i}+\beta_{y,i}Y+\beta_{z,i}^T\mathbf{Z})}{\sigma_i}\right)$, where $\Phi(\cdot)$ is the CDF of $N(0, 1)$. Inserting this into Eqs. (5) and (6) and further into the optimization problems in Eqs. (2) and (3), we get:

$$l_i^*(\mathbf{Z}) = \begin{cases} \max\limits_{l_i(\mathbf{Z})} & l_i(\mathbf{Z}) \\ s.t. & \dfrac{\Phi\left(\frac{l_i(\mathbf{Z})-\left(\beta_{0,i}+\beta_{y,i}+\beta_{z,i}^T\mathbf{Z}\right)}{\sigma_i}\right)}{\Phi\left(\frac{l_i(\mathbf{Z})-\left(\beta_{0,i}+\beta_{z,i}^T\mathbf{Z}\right)}{\sigma_i}\right)} \le \frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \end{cases}. \tag{8}$$

$$u_i^*(\mathbf{Z}) = \begin{cases} \min\limits_{u_i(\mathbf{Z})} & u_i(\mathbf{Z}) \\ s.t. & \dfrac{1-\Phi\left(\frac{u_i(\mathbf{Z})-\left(\beta_{0,i}+\beta_{y,i}+\beta_{z,i}^T\mathbf{Z}\right)}{\sigma_i}\right)}{1-\Phi\left(\frac{u_i(\mathbf{Z})-\left(\beta_{0,i}+\beta_{z,i}^T\mathbf{Z}\right)}{\sigma_i}\right)} \ge \frac{r_h}{1-r_h} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \end{cases}. \tag{9}$$

Next, we present an important property of the solutions to the optimization problems in Eqs. (8) and (9) in Propositions 2 and 3, respectively. The proof for Proposition 2 is given in the Appendix. The proof for Proposition 3 is similar and therefore not provided.

Proposition 2: The solution to Eq. (8) exists and is unique. When $\frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \in (0, 1)$, $l_i^*(\mathbf{Z})$ is the feasible solution at which the equality of the constraint is achieved. When $\frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \ge 1$, $l_i^*(\mathbf{Z}) = \infty$.

Proposition 3: The solution to Eq. (9) exists and is unique. When $\frac{r_h}{1-r_h} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} > 1$, $u_i^*(\mathbf{Z})$ is the feasible solution at which the equality of the constraint is achieved. When $\frac{r_h}{1-r_h} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \in (0,1]$, $u_i^*(\mathbf{Z}) = -\infty$.

3.2. Model estimation for STC

Proposition 2 sheds some light on how to find the lower cutoff of the biomarker, $l_i^*(\mathbf{Z})$. Before the patient takes the biomarker testing, his/her $\frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}$ will be computed. If it is greater than or equal to one, the lower cutoff of the biomarker for this patient is infinity. This means that the patient can be considered LR regardless of the biomarker value. In other words, this patient does not need to be tested for the biomarker. Such situations rarely happen in practice, except for people with extremely high resistance to a certain disease; e.g., people carrying some genes that are disease-protective. In most cases, people coming to a clinic for diagnosis of a disease usually bear a fairly extensive amount of suspicion or risk for the disease. Therefore, we focus on the condition when $\frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \in (0, 1)$. Then, the problem becomes finding the $l_i(\mathbf{Z})$ satisfying the equality of

$$\frac{\Phi\left(\frac{l_i(\mathbf{Z})-\left(\beta_{0,i}+\beta_{y,i}+\beta_{z,i}^T\mathbf{Z}\right)}{\sigma_i}\right)}{\Phi\left(\frac{l_i(\mathbf{Z})-\left(\beta_{0,i}+\beta_{z,i}^T\mathbf{Z}\right)}{\sigma_i}\right)} = \frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}. \tag{10}$$

Unfortunately, this problem does not have an analytical solution. To solve it, we may adopt one of two approaches: a numerical approach that finds the $l_i(\mathbf{Z})$ satisfying Eq. (10) for any given $\mathbf{Z}$. This approach can achieve any required precision for the solution, but is computationally intensive. An alternative approach is to use an approximation for $\Phi(x)$ proposed by Bowling *et al.*

(2009); i.e.,

$$\Phi(x) \approx \frac{1}{1+\exp(-1.702x)}. \tag{11}$$

By substituting Eq. (11) into Eq. (10), $l_i^*(\mathbf{Z})$ can be solved analytically as

$$l_i^*(\mathbf{Z}) = -\frac{\sigma_i}{1.702}\ln \times \left(\frac{1-\frac{r_l}{1-r_l}\times\frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}}{\frac{r_l}{1-r_l}\times\frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}\exp\left(1.702\frac{(\beta_{0,i}+\beta_{y,i}+\beta_{z,i}^T\mathbf{Z})}{\sigma_i}\right)-\exp\left(1.702\frac{(\beta_{0,i}+\beta_{z,i}^T\mathbf{Z})}{\sigma_i}\right)} \right). \tag{12}$$

Likewise, Proposition 3 sheds some light on how to find the higher cutoff of the biomarker; i.e., $u_i^*(\mathbf{Z})$. Following similar reasoning and using the approximation in Eq. (11), $u_i^*(\mathbf{Z})$ can be solved analytically as

$$u_i^*(\mathbf{Z}) = \frac{\sigma_i}{1.702}\ln \times \left(\frac{1-\frac{r_h}{1-r_h}\times\frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}}{\frac{r_h}{1-r_h}\times\frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}\exp\left(1.702\frac{(\beta_{0,i}+\beta_{y,i}+\beta_{z,i}^T\mathbf{Z})}{\sigma_i}\right)-\exp\left(1.702\frac{(\beta_{0,i}+\beta_{z,i}^T\mathbf{Z})}{\sigma_i}\right)} \right). \tag{13}$$

Finally, we would like to point out that the $\beta_{0,i}$, $\beta_{y,i}$, $\beta_{z,i}$, and σ_i in Eqs. (12) and (13) are unknown but can be estimated from a training dataset. For example, under the linear model in Eq. (7), $\beta_{0,i}$, $\beta_{y,i}$, $\beta_{z,i}$, and σ_i can be estimated by an maximum likelihood estimation (MLE). If $\mathbf{Z}$ is high-dimensional, variable selection techniques may be adopted to select a small subset of $\mathbf{Z}$ that have non-zero coefficients, such as the well-known lasso model (Wee *et al.*, 2013), followed by an MLE on the non-zero coefficients. However, regardless of the estimation method, there is sampling uncertainty in the estimated $\beta_{0,i}$, $\beta_{y,i}$, $\beta_{z,i}$, and σ_i due to the finite sample size of the training dataset, which will further introduce uncertainty into $u_i^*(\mathbf{Z})$ and $l_i^*(\mathbf{Z})$. To better account for the sampling uncertainty, we use Monte Carlo simulation to generate an empirical sampling distribution for $\hat{u}_i^*(\mathbf{Z})$ and $\hat{l}_i^*(\mathbf{Z})$, respectively, and then use the empirical means as the solutions to Eqs. (13) and (12). This approach is found to be more robust to sampling uncertainty and has better accuracy in our case studies. Specifically, the empirical sampling distribution for $\hat{u}_i^*(\mathbf{Z})$ is generated as follows (a similar procedure can be used for $\hat{l}_i^*(\mathbf{Z})$): Let $\tilde{\beta}_{0,i}$, $\tilde{\beta}_{y,i}$, $\tilde{\beta}_{z,i}$, and $\tilde{\sigma}_i$ be the estimated model parameters from the training dataset through MLE. We use Monte Carlo simulation to generate N samples from the following empirical distributions:

$$\begin{pmatrix}\hat{\beta}_{0,i}^{(t)}\\ \hat{\beta}_{y,i}^{(t)}\\ \hat{\boldsymbol{\beta}}_{z,i}^{(t)}\end{pmatrix} \sim N\left(\begin{pmatrix}\tilde{\beta}_{0,i}\\ \tilde{\beta}_{y,i}\\ \tilde{\beta}_{z,i}\end{pmatrix}, \tilde{\sigma}_i^2\left(\begin{pmatrix}1 & \mathbf{y} & \mathbf{z}\end{pmatrix}^T\begin{pmatrix}1 & \mathbf{y} & \mathbf{z}\end{pmatrix}\right)^{-1}\begin{pmatrix}1 & \mathbf{y} & \mathbf{z}\end{pmatrix}^T\mathbf{x}_i\right), \tag{14}$$

$$\hat{\sigma}_i^{2(t)} \sim \frac{\left(\mathbf{x}_i-\left(\tilde{\beta}_{0,i}+\tilde{\beta}_{y,i}\mathbf{y}+\tilde{\beta}_{z,i}^T\mathbf{z}\right)\right)^T\left(\mathbf{x}_i-\left(\tilde{\beta}_{0,i}+\tilde{\beta}_{y,i}\mathbf{y}+\tilde{\beta}_{z,i}^T\mathbf{z}\right)\right)}{\chi_{n-p}^2}, \tag{15}$$

$t = 1, \ldots, N$. $\mathbf{x}_i$, $\mathbf{y}$, and $\mathbf{z}$ are training data for n patients. 1 is a $n \times 1$ vector of ones. p is the column dimension of the predictor matrix $(1\ \mathbf{y}\ \mathbf{z})$. Then, each sample generated from Eqs. (14) and (15)—i.e., $\hat{\beta}_{0,i}^{(t)}$, $\hat{\beta}_{y,i}^{(t)}$, $\hat{\boldsymbol{\beta}}_{z,i}^{(t)}$, and $\hat{\sigma}_i^{2(t)}$—is inserted into Eq. (13) to obtain $\hat{u}_i^*(\mathbf{Z})^{(t)}$. The average, $\overline{\hat{u}}_i^*(\mathbf{Z}) = \frac{\sum_{t=1}^N \hat{u}_i^*(\mathbf{Z})^{(t)}}{N}$, is used as the final solution to Eq. (13).

3.3. Algorithm for STC

Section 3.1 and 3.2 assumed that the biomarker sequence is known and the discussion was focused on the i-th step (i.e., the i-th biomarker) of the modeling building process of the STC. In this section, we present the full algorithm. The input to the algorithm includes a specification on the biomarkers that are allowed to be used in a clinic. This may be clinic-specific depending on availability and resource constraints. The input also includes a training and a validation set on the biomarkers $\mathbf{X}$, patient characteristic variables/risk factors $\mathbf{Z}$, and the diagnostic outcome Y, the HR and LR thresholds, r_h and r_l, and the prior, $\pi(\mathbf{Z})$. Suppose p biomarkers are available. Then, the objective or output of the algorithm is to find an optimal sequence of the biomarkers with cutoffs for each biomarker, $u_i^*(\mathbf{Z})$ and $l_i^*(\mathbf{Z})$, $i = 1, \ldots, p$. Since the number of biomarkers for a particular disease is usually small, we will perform an exhaustive search over all possible sequences. We will report three metrics computed on the validation set for comparing the sequences: positive prediction value (PPV), negative prediction value (NPV), and the percentage of patients classified as inconclusive. The first two metrics reflect the accuracy, where PPV measures the proportion of patients classified as HR that are true converters and NPV measures the proportion of patients classified as LR that are true non-converters. The last metric reflects the efficiency: the lower the inconclusive percentage, the more efficient the biomarker sequence.

Specifically, given that p biomarkers are available in a clinic, our algorithm performs three major steps for each of $p!$ possible biomarker sequences. Without loss of generality, denote each sequence by $X_1 \to X_2 \to \cdots \to X_p$.

Step 1 (initialization): Initialize the algorithm by having $i \leftarrow 1$ and putting the entire training set into D_{i-1}.

Step 2 (sequential estimation):

- **Sub-step 2.1 (identification of the cutoffs for X_i)**: Fit a linear model as (7) for X_i using the training data in D_{i-1}, and obtain estimates for the model coefficients, $\tilde{\beta}_{0,i}$, $\tilde{\beta}_{y,i}$, $\tilde{\beta}_{z,i}$, and $\tilde{\sigma}_i$. Check the normality assumption of the model and apply box-cox transformation to X_i if needed. Use the estimated model coefficients to obtain N Monte Carlo samples $\hat{\beta}_{0,i}^{(t)}$, $\hat{\beta}_{y,i}^{(t)}$, $\hat{\boldsymbol{\beta}}_{z,i}^{(t)}$, and $\hat{\sigma}_i^{2(t)}$, $t = 1, \ldots, N$. Insert each sample into Eqs. (13) and (12) and obtain sample realizations for the cutoffs, i.e., $\hat{u}_i^*(\mathbf{Z})^{(t)}$ and $\hat{l}_i^*(\mathbf{Z})^{(t)}$, $t = 1, \ldots, N$. Use the sample averages, $\overline{\hat{u}}_i^*(\mathbf{Z})$ and $\overline{\hat{l}}_i^*(\mathbf{Z})$, as the estimated cutoffs for X_i.
- **Sub-step 2.2 (subsetting of the training set)**: Apply the estimated cutoffs in sub-step 2.1 to the patients in D_{i-1} and only keep patients with $\overline{\hat{l}}_i^*(\mathbf{Z}) < X_i < \overline{\hat{u}}_i^*(\mathbf{Z})$ in the training set. Denote the current training set by D_i.

Sub-step 2.3 (continuation or stopping): Move onto the next biomarker by having $i \leftarrow i+1$ and going to sub-step 2.1, until $i+1=p$.

Step 3 (evaluation): Apply the estimated cutoffs for each biomarker; i.e., $\overline{\hat{u}}_i^*(\mathbf{Z})$ and $\overline{\hat{l}}_i^*(\mathbf{Z})$, $i=1,\ \ldots, p$, to the validation set and compute PPV, NPV, and the percentage of patients classified as inconclusive.

This three-step algorithm will be applied to each of the $p!$ possible biomarker sequences. These sequences will then be compared in terms of the diagnostic accuracy (PPV and NPV) and efficiency (percentage of inconclusive patients) evaluated on the validation set. Because multiple metrics are used in the comparison, an integrated metric may be used to help select the optimal sequence. Alternatively, a Pareto optimal frontier may be provided to practitioners to show the tradeoffs between multiple Pareto optimal solutions/sequences.

3.4. Extension to non-Gaussian biomarkers

When the biomarkers do not follow Gaussian distributions, one approach is to apply Box-Cox transformation to make them approximately Gaussian, which was mentioned in Section 3.1. An alternative approach is to deal with the non-Gaussian distributions directly. Specifically, instead of linking the biomarker X_i with Y, $\mathbf{Z}$ by a linear model as in Eq. (7), we can use a Generalized Linear Model (GLM); i.e.,

$$E(X_i) = g^{-1}\left(\beta_{0,i} + \beta_{y,i}Y + \beta_{z,i}^T\mathbf{Z}\right), \tag{16}$$

where $g(\cdot)$ is an appropriate link function depending on the distribution of the biomarker. Consequently, Eqs. (8) and (9) change to

$$l_i^*(\mathbf{Z}) = \begin{cases} \max\limits_{l_i(\mathbf{Z})} & l_i(\mathbf{Z}) \\ s.t. & \frac{\varphi_{X_i|Y=1,\mathbf{Z}=\mathbf{z}}(l_i(\mathbf{Z}))}{\varphi_{X_i|Y=0,\mathbf{Z}=\mathbf{z}}(l_i(\mathbf{Z}))} \le \frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \end{cases}, \tag{17}$$

$$u_i^*(\mathbf{Z}) = \begin{cases} \min\limits_{u_i(\mathbf{Z})} & u_i(\mathbf{Z}) \\ s.t. & \frac{1-\varphi_{X_i|Y=1,\mathbf{Z}=\mathbf{z}}(u_i(\mathbf{Z}))}{1-\varphi_{X_i|Y=0,\mathbf{Z}=\mathbf{z}}(u_i(\mathbf{Z}))} \ge \frac{r_h}{1-r_h} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})} \end{cases}. \tag{18}$$

$\varphi_{X_i|Y,\mathbf{Z}}(x)$ is the CDF of X_i given Y and $\mathbf{Z}$, which can be specificed according to the GLM in Eq. (16). $\varphi_{X_i|Y,\mathbf{Z}}(x)$ is not Gaussian, so the approximation in Eq. (11) cannot be used. Consequently, Eqs. (17) and (18) cannot be solved analytically but by a numerical search, which is computationally more intensive. The modified STC algorithm is the following:

Step 1 (initialization): Initialize the algorithm by having $i \leftarrow 1$ and putting the entire training set into D_{i-1}.

Step 2 (sequential estimation):

Sub-step 2.1 (identification of the cutoffs for $g(E(\boldsymbol{X}_i))$): Fit a GLM as in Eq. (16) using the training data in D_{i-1}, and obtain estimates for the model coefficients, $\tilde{\beta}_{0,i}$, $\tilde{\beta}_{y,i}$, and $\tilde{\beta}_{z,i}$. In order to solve the optimization problems in Eq. (17), we can start from a small $l_i(\mathbf{Z})$ for which the constraint holds, and increase $l_i(\mathbf{Z})$ in small steps until the constraint is violated. The last value of $l_i(\mathbf{Z})$ before the constrain is violated is the optimal solution $l_i^*(\mathbf{Z})$. Likewise, we can obtain the optimal solution $u_i^*(\mathbf{Z})$ in Eq. (18).

Sub-step 2.2 (subsetting of the training set): Apply the estimated cutoffs in sub-step 2.1 to the patients in D_{i-1} and only keep patients with $l_i^*(\mathbf{Z}) < g(E(\boldsymbol{X}_i)) < u_i^*(\mathbf{Z})$ in the training set. Denote the current training set by D_i.

Sub-step 2.3 (continuation or stopping): Move onto the next biomarker by having $i \leftarrow i+1$ and going to sub-step 2.1, until $i+1=p$.

Step 3 (evaluation): Apply the estimated cutoffs for each biomarker (i.e., $u_i^*(\mathbf{Z})$ and $l_i^*(\mathbf{Z})$, $i=1,\ \ldots, p$) to the validation set and compute PPV, NPV, and the percentage of patients classified as inconclusive.

4. Case studies in prediction of MCI conversion to AD

In this section, we present two clinical applications using the proposed STC: Sub-Section 4.1 presents an application in clinical diagnosis; i.e., prediction/classification of MCI patients into HR, LR, or inconclusive categories so that appropriate medical decisions can be made for each patient. Sub-Section 4.1 presents another application of using STC to help patient selection in clinical trials. As mentioned in the Introduction, there has been a growing consensus in the medical society that treatment of AD should target its early phases before irreversible brain damage occurs. MCI is such an early phase and therefore has been targeted by drug companies to develop treatment for slowing down or even stopping the progression to AD. However, it is well-known that MCI patients are heterogeneous and not all of them will eventually develop AD. To be able to appropriately assess the efficacy of an AD-defeating drug, it is important to identify a sub-cohort of MCI patients with a HR of converting to AD and enter these patients into the drug trial. This important task is known as patient selection in clinical trials and can be accomplished with the help of STC.

The data used in this section were obtained from the ADNI database (http:// adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations, as a $60 million, five-year public-private partnership. The primary goal of ADNI has been to test whether MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the US and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for three years, 400 people with MCI to be followed for at least three years, and 200 people with early AD to be followed for two years. For up-to-date information, see http://www.adni-info.org/.

Table 1. Description of the data.

Variable	Non-Converters	Converters
Sample size	87	100
Gender: female %	59.8	58
Age: ave. (std.)	73.1 (7.3)	73.6 (7.6)
Education years: ave. (std.)	16.6 (2.6)	16.3 (2.7)
APOE e4 status: carriers %	41.4	68
Mini-mental State Examination score: ave. (std.)	27.9 (1.6)	26.7 (1.7)
P-tau, *pg/mL*: ave. (std.)	35.8 (22.8)	50.6 (25.1)
FDG-PET, relative counts: ave. (std.)	1.25 (0.13)	1.16 (0.11)
Hippo, mm^3: ave. (std.)	3449 (551)	3067 (497)

Specifically, our study contains 187 MCI patients included in the ADNI database who have complete data on three biomarkers—P-tau, FDG-PET, and Hippo—at their baseline visits, patient-specific variables/risk factors such as age, gender, education level, APOE e4 status, and cognitive test scores, as well as conversion vs. non-conversion to AD at the end of their clinical follow-up time periods. A detailed description of the data is shown in Table 1.

Standardized biomarker acquisition and performance methods of ADNI are described at www.loni.ucla.edu/ADNI. Protocols of image and CSF analyses are reported in detail elsewhere (Jack *et al.*, 2010b; Jagust *et al.*, 2009; Kim *et al.*, 2011; Landau *et al.*, 2010). In brief, the mean FDG count per subject (i.e., biomarker "FDG-PET") was extracted from a composite region of interest on the basis of the AD-typical hypometabolic pattern (Jack *et al.*, 2008; Kim *et al.*, 2011). Hippocampal volumes (i.e., biomarker "Hippo") were extracted from structural MRI scans (1.5 T) using the FreeSurfer software http://surfer.nmr.mgh.harvard.edu (Kim *et al.*, 2011). Peptide concentrations (i.e., biomarker "P-tau") were measured in CSF using aliquots obtained from the same vial at the same thaw (Jagust *et al.*, 2010).

4.1. Clinical diagnosis of MCI conversion to AD

4.1.1. Diagnosis of MCI conversion to AD with three biomarkers

We first focus on a scenario that all three biomarkers—P-tau, FDG-PET, and Hippo—are available in the clinic. Then, the goal is to find an optimal sequence of the biomarkers with cutoffs for each biomarker; i.e., $u_i^*(\mathbf{Z})$ and $l_i^*(\mathbf{Z})$, $i = 1, 2, 3$. Among known risk factors such as age, gender, education level, and APOE e4 status, only age is found to be significant in this dataset. Therefore, $\mathbf{Z}$ includes age. The HR and LR thresholds are set to be $r_h = 0.8$ and $r_l = 0.2$, which are common choices in clinical diagnosis. Also, a uniform prior is adopted; i.e., $\pi(\mathbf{Z}) = 0.5$. Three biomarkers compose $3! = 6$ possible sequences. For each sequence, we apply the algorithm in Section 3.3 with a minor modification of using cross-validation (CV) instead of arbitrarily splitting the entire dataset into a training and a validation set. The CV-based PPV, NPV, and percentage of inconclusive patients for each sequence are summarized in Table 2. Box-Cox transformation on the biomarkers is used and the transformed biomarkers in each sequence follow Gaussian distributions. For example, Fig. 1 shows the QQ plot of each transformed biomarker in the sequence "P-tau->FDG-PET->Hippo," which demonstrates clear normality.

Table 2. CV-based PPV, NPV, and percentage of inconclusive patients for all possible sequences of three biomarkers using STC.

Sequence of biomarkers	PPV	NPV	% inconclusive patients
P-tau->FDG-PET->Hippo	74%	81%	59%
FDG-PET->P-tau->Hippo	70%	74%	52%
P-tau->Hippo->FDG-PET	71%	78%	56%
Hippo->P-tau->FDG-PET	73%	78%	52%
FDG-PET->Hippo->P-tau	72%	77%	58%
Hippo->FDG-PET->P-tau	72%	77%	63%

A clear trend of the results in Table 2 is that the NPVs are higher than PPVs regardless of the sequence of biomarkers. This suggests that the three biomarkers have a better capability for identifying non-converters than converters. Another observation is that the PPVs are lower than the HR threshold $r_h = 0.8$. This is reasonable because $r_h = 0.8$ is set for model training and the PPVs are computed based on CV, which reflect the accuracy of the trained model applied to unseen data. The fact that the PPVs are only slightly lower than 0.8 implies that STC has good generalization capability. Likewise, the NPVs are only slightly lower than or almost equal to $1 - r_l = 0.8$, which also indicates good generalization capability of STC. Last but not least, we observe that over half of the MCI patients in the dataset are found to be inconclusive, no matter which sequence of the biomarkers is used. This is expected because this study only uses *baseline* biomarker measurements to predict the conversion of MCI to AD. Use of baseline biomarkers for the prediction has clear clinical benefits, as it enables early decision making for patients classified as HR and LR converters. On the other hand,

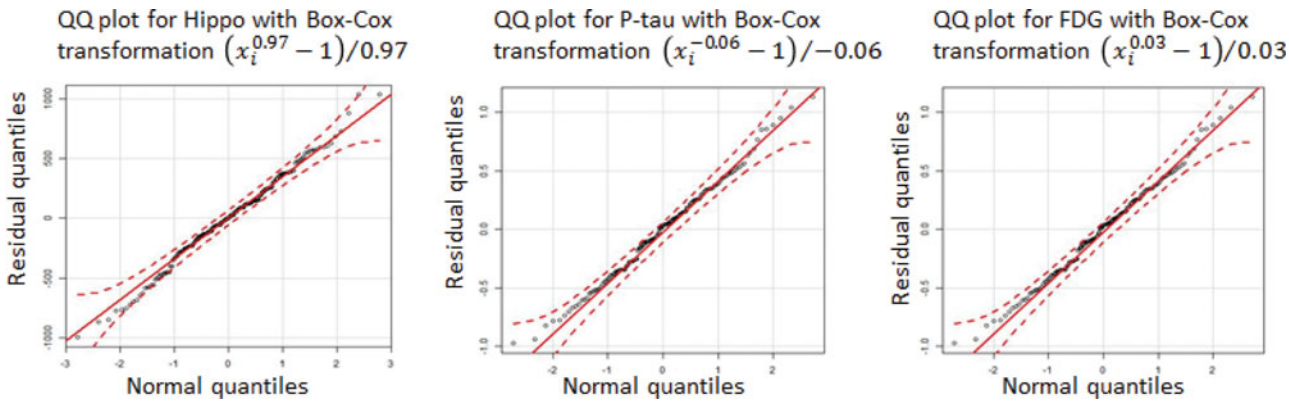


Figure 1. QQ plots for biomarkers after Box-Cox transformation in the sequence "P-tau->FDG-PET->Hippo."

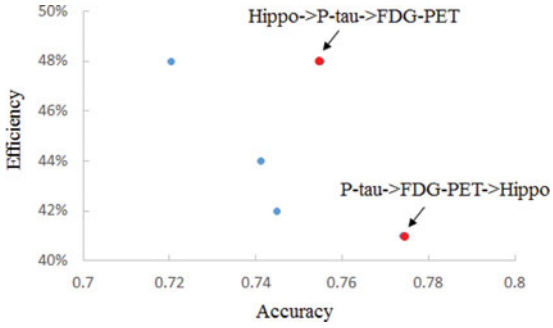


Figure 2. Efficiency vs. accuracy of six possible sequences given all three biomarkers. Sequences in red are on the Pareto optimal frontier.

it is highly likely that a conclusive classification is not possible for some MCI patients using baseline biomarker measurements alone. These patients need to be followed up and kept tracked of for the changes in their biomarker measurements over time before a conclusive prediction can be reached.

To select an optimal biomarker sequence among the six possible ones in Table 2, we need to make a tradeoff between accuracy (measured by PPV and NPV) and efficiency (measured by the percentage of inconclusive patients), because no sequence optimizes the two criteria simultaneously. If accuracy is the primary consideration, the sequence "P-tau->FDG-PET->Hippo" should be selected because it has the highest PPV (74%) and NPV (81%). This sequence, on the other hand, classifies 59% of MCI patients as inconclusive, which makes it the second least efficient sequence among the six. If efficiency is the primary consideration, the sequence "Hippo->P-tau->FDG-PET" should be selected, as it has the lowest percentage of inconclusive patients (52%), although its accuracy is sub-optimal.

A commonly used approach in optimization when multiple criteria need to be optimized is to examine the Pareto optimal frontier. Figure 2 shows the Pareto optimal frontier for the six biomarker sequences. The vertical axis "efficiency" is defined as 1 − percentage of inconclusive patients or the percentage of patients classified as HR or LR by STC. The horizontal axis "accuracy" is defined as a weighted average of PPV and NPV, where the weights are proportions of samples classified as HR and LR, respectively. Each sequence is represented by a dot. Two dots in red are sequences on the Pareto optimal frontier. In particular, the sequence "P-tau->FDG-PET->Hippo" optimizes the accuracy criterion while "Hippo->P-tau->FDG-PET" optimizes the efficiency.

Next, we would like to show the cutoffs of each biomarker found by STC. We choose to show these for the sequence "P-tau->FDG-PET->Hippo" as an example using a tree-like plot in Fig. 3. Specifically, In Fig. 3(a), branches in green/red represent HR/LR converters classified by STC. The branch in grey represents the inconclusive category. Sizes of the branches/circles are in proportion to the sample sizes of the branches. A clear observation is that fewer samples are classified as HR and LR as the tree goes deeper. This is a result of the sequential nature of STC; i.e., a later biomarker needs to classify samples that are failed to be classified (the inconclusive samples) by a previous biomarker so it has a "tougher" mission to accomplish. Figure 3(b) shows the cutoffs as functions of "age" using the approximations in Eqs. (13) and (15). Values of the cutoffs at the median age of the dataset are also provided for illustration purposes.

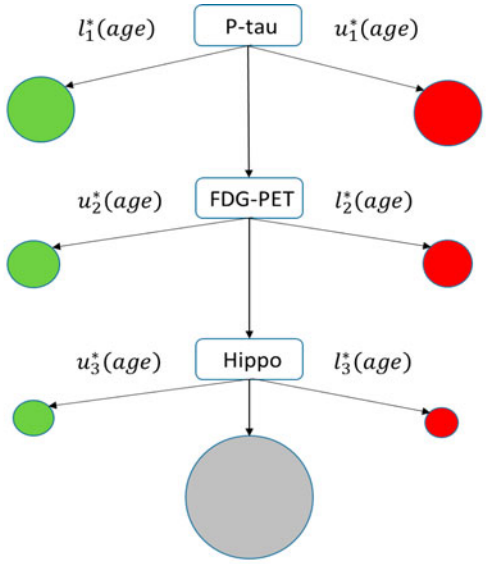
Moreover, we would like to point out that the findings from STC can help not only clinical diagnosis, but also knowledge discovery, such as discovering disease sub-types. Using the tree in Fig. 3 as an example, there seem to exist three distinct sub-types of HR converters; i.e., the sub-types of P-tau-abnormality (P-tau $\geq u_1^*(age)$), FDG-PET-abnormality (P-tau $< u_1^*(age)$ & FDG-PET $\leq l_2^*(age)$), and Hippo-abnormality (P-tau $< u_1^*(age)$ & FDG-PET $> l_2^*(age)$ & Hippo $\leq l_3^*(age)$). Indeed, there has been medical evidence that the three biomarkers track distinct aspects of the AD pathophysiological process (Jack *et al.*, 2010a). That is, FDG-PET, as a measure for AD-related glucose hypometabolism, reflects reduction in synaptic density/activity and phenomena of diaschisis, Hippos, as a measure for hippocampal atrophy, reflects neural loss, while P-tau reflects intracellular hyperphosphorylation of tau. STC allows for a finer distinction of HR converters into different sub-types according to specific biomarker abnormality, which may lead to more targeted and effective treatment. Likewise, STC can help discover sub-types of LR converters. This would facilitate the study of different pathophysiological mechanisms that lead to disease protection or resistance.

4.1.2. Diagnosis of MCI conversion to AD with two biomarkers: A limited-resource scenario

Next, we present the results of STC in a "limited-resource" scenario; e.g., when only two out of the three biomarkers are available. This situation is common in many clinics. We use the same setting as the previously presented three-biomarker scenario; i.e., $r_h = 0.8$, $r_l = 0.2$, $\pi(\mathbf{Z}) = 0.5$, and $\mathbf{Z} = \{age\}$. Two biomarkers compose six possible sequences. For each sequence, we apply the algorithm in Section 3.3 and compute the CV-based PPV, NPV, and percentage of inconclusive patients. Figure 4 shows the Pareto optimal frontier for the six sequences, in which efficiency and accuracy are defined in the same way as Fig. 2. The sequence "Hippo->P-tau" optimizes the accuracy criterion, while "FDG-PET->P-tau" optimizes the efficiency. Furthermore, each ellipse includes two sequences with the same pair of biomarkers but in different orders. If a clinic only has the resource for testing two specific biomarkers, we can compare the two dots/sequences within the same ellipse and select an order of the two biomarkers that is more appropriate in terms of efficiency or/and accuracy. For example, if a clinic only has FDG-PET and Hippos, we can compare the two dots within the middle ellipse. The dot at the upper-right corner corresponds to the sequence "Hippo->FDG-PET" and is clearly better because it has better efficiency and accuracy.

4.1.3. Comparison between STC and decision tree

Finally, we compare STC with the conventional decision tree. Specifically, we apply the C4.5 algorithm in the Weka software (Hall *et al.*, 2009) to the same dataset as that used by STC. Because STC uses age in addition to three biomarkers, we include the same variables in C4.5 for a fair comparison. Parameters of C4.5 are tuned to optimize the CV accuracy. Figure 5 shows the decision tree generated by C4.5. Compared with the tree generated by STC in Fig. 3, we can obtain the



(a)

Cutoffs as functions of age	Cutoffs at median age
$l_1^*(age) = exp\left(\frac{log\left(-0.01 \times ln\left(\frac{1}{3} \times e^{(16.83-0.04\times age)} - \frac{4}{3} \times e^{(15.43-0.04\times age)}\right)+1\right)}{-0.04}\right)$	19.4
$u_1^*(age) = exp\left(\frac{log\left(0.01 \times ln\left(-\frac{4}{3} \times e^{(16.83-0.04\times age)} + \frac{1}{3} \times e^{(15.43-0.04\times age)}\right)+1\right)}{-0.04}\right)$	68.4
$l_2^*(age) = exp\left(\frac{log\left(0.01 \times ln\left(-\frac{4}{3} \times e^{(-3.87+0.02\times age)} + \frac{1}{3} \times e^{(-5.24+0.02\times age)}\right)+1\right)}{0.15}\right)$	1.054
$u_2^*(age) = exp\left(\frac{log\left(-0.01 \times ln\left(\frac{1}{3} \times e^{(-3.87+0.02\times age)} - \frac{4}{3} \times e^{(-5.24+0.02\times age)}\right)+1\right)}{0.15}\right)$	1.360
$l_3^*(age) = exp\left(\frac{log\left(80.17 \times ln\left(-\frac{4}{3} \times e^{(-20.42+0.11\times age)} + \frac{1}{3} \times e^{(-21.70+0.11\times age)}\right)+1\right)}{0.86}\right)$	2471.8
$u_3^*(age) = exp\left(\frac{log\left(-80.17 \times ln\left(\frac{1}{3} \times e^{(-20.42+0.11\times age)} - \frac{4}{3} \times e^{(-21.70+0.11\times age)}\right)+1\right)}{0.86}\right)$	3976.2

(b)

Figure 3. Cutoffs found by STC for biomarker sequence "P-tau->FDG-PET->Hippo" represented by (a) a tree-like plot in which green/red/grey circles represent LR/HR/inconclusive categories and sizes of the circles are in proportion to the sample size of each branch; (b) cutoffs of each biomarker as functions of "age."

following observations: both methods find P-tau as the first biomarker to be used for the classification. This suggests that P-tau may be more informative than the other two biomarkers. The differences between the two methods are summarized as follows: (1) The CV-based PPV and NPV of the decision tree are 68% and 69%, respectively, which are significantly lower than the PPV (74%) and NPV (81%) of the optimal sequence found by STC. This is because the decision tree, by design, must assign a class membership to every sample, even when a sample does not have a significantly higher probability of belonging to one class than the other. This leads to potentially large classification errors. In contrast, STC only classifies samples with a HR or LR of conversion, while putting samples with only a mild risk in either direction (i.e., a risk between LR and HR) in an inconclusive category. From a disease management point of view, STC is more appropriate by allowing HR patients to receive immediate medical attention, LR patients to be put on long-term observation, and patients in between to be followed up to track the changes in their disease risks. (2) According to the decision tree in Fig. 5, no patients can be classified using a single biomarker. In contrast, according to the tree in Fig. 3 produced by STC, 52.3% of the patients classified as HR and LR only need to be tested by P-tau. In this sense, STC means fewer diagnostic costs, less patient waiting time, and more timely medical

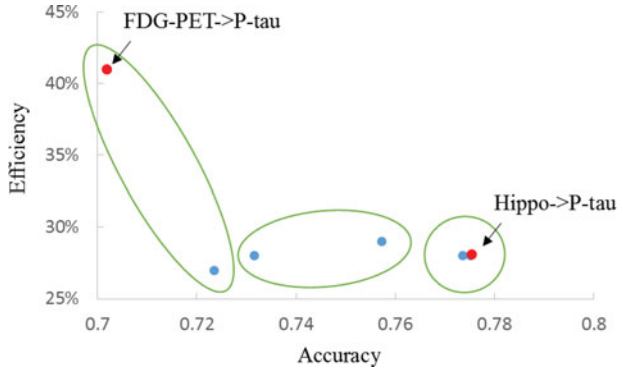


Figure 4. Efficiency vs. accuracy of six possible sequences given two out of three biomarkers. Sequences in red are on the Pareto optimal frontier. Each ellipse highlights two sequences with the same pair of biomarkers but in different orders.

decision making. (3) The decision tree in Fig. 5 is somewhat counterintuitive. Biomarkers are expected to have a monotonic relationship with the risk of disease. For example, a higher P-tau, lower FDG-PET, or lower Hippo indicates a higher risk of AD pathology. However, there are several branches in Fig. 5 whose biomarker ranges are contrary to this expectation. For instance, the top-right green circle represents non-converters whose classification rule is P-tau $>$ 28.5 and FDG-PET $>$ 1.19. This higher value range for P-tau is expected to indicate a higher risk of AD pathology. From a clinical utilization's point of view, clinicians would be very reluctant to adopt such a model as the decision tree in Fig. 5, regardless of the accuracy, because the model is against their medical knowledge and thus is difficult to understand and trust. In essence, the decision tree is a pure data-driven model that does not integrate medical knowledge and biological principles into its model-building process. In contrast, STC, by its unique design, honors the monotonic relationship between a biomarker and disease risk, and therefore is able to provide a model with good interpretability and clinical utility.

4.2. Selection of HR converters for clinical trials

Here, our objective is to identify a sub-cohort of MCI patients with a HR of converting to AD and enter these patients into a drug trial. This objective is different from clinical diagnosis as presented in sub-Section 4.1 in the sense that we only care about maximizing PPV, as opposed to accuracy that includes both PPV and NPV, and maximizing the number/proportion of patients

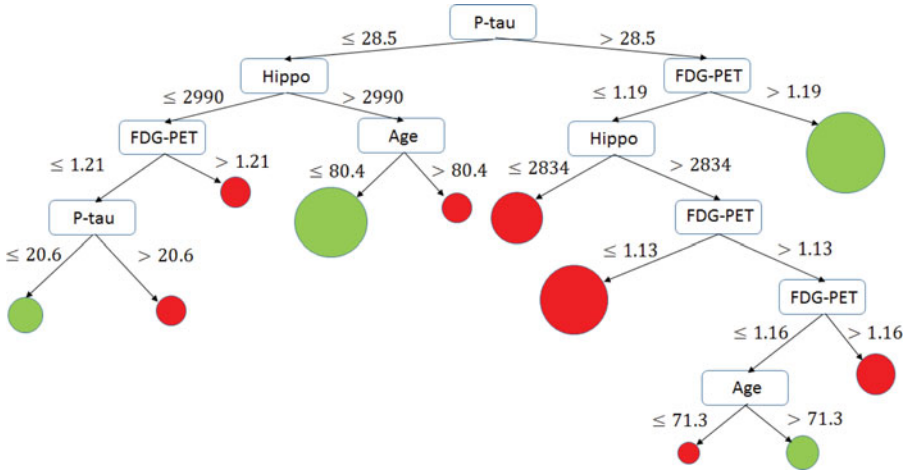


Figure 5. Decision tree generated by C4.5. Circles (green/red) represent non-convert/converter categories and sizes of the circles are in proportion to the sample size of each branch. quence "Hippo->P-tau->FDG-PET" at $r_l = 0.35$ found by STC.

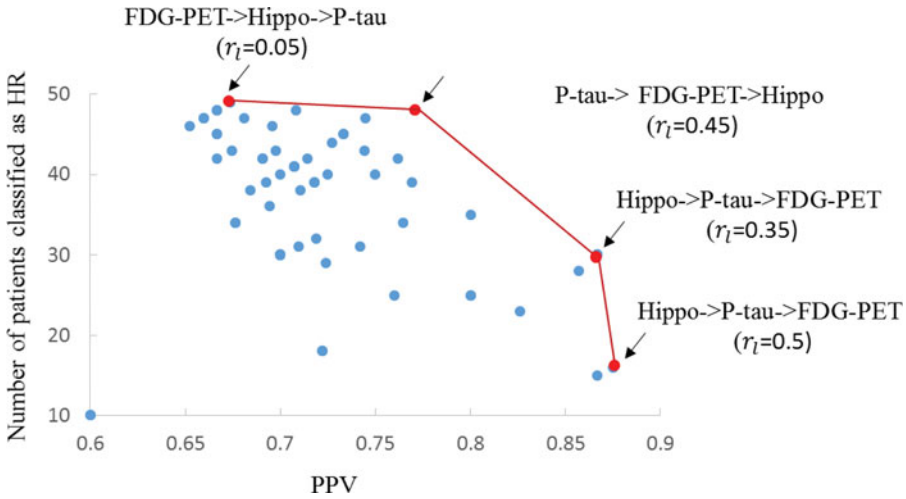
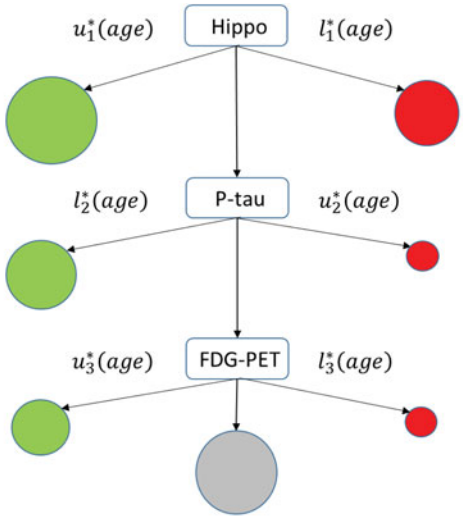


Figure 6. Number of HR patients vs. PPV of six possible sequences given all three biomarkers at r_l ranging from 0.05 to 0.5 in increment of 0.05. Connected circles (red) are on the Pareto optimal frontier.



(a)

Cutoffs as functions of age	Cutoffs at median age
$l_1^*(age) = exp\left(\frac{log\left(209.97 \times ln\left(-\frac{4}{3} \times e^{(-19.80+0.11\times age)} + \frac{1}{3} \times e^{(-21.12+0.11\times age)}\right) + 1\right.}{0.97}\right.$	2613.8
$u_1^*(age) = exp\left(\frac{log\left(-209.97 \times ln\left(\frac{7}{5} \times e^{(-19.80+0.11\times age)} - \frac{13}{5} \times e^{(-21.12+0.11\times age)}\right) +\right.}{0.97}\right.$	3470.1
$l_2^*(age) = exp\left(\frac{log\left(-0.01 \times ln\left(\frac{7}{5} \times e^{(16.28-0.03\times age)} - \frac{13}{5} \times e^{(15.13-0.03\times age)}\right) + 1\right)}{-0.04}\right)$	29.7
$u_2^*(age) = exp\left(\frac{log\left(0.01 \times ln\left(-\frac{4}{3} \times e^{(16.28-0.03\times age)} + \frac{1}{3} \times e^{(15.13-0.03\times age)}\right) + 1\right)}{-0.04}\right)$	91.5
$l_3^*(age) = exp\left(\frac{log\left(0.01 \times ln\left(-\frac{4}{3} \times e^{(-0.85-0.03\times age)} + \frac{1}{3} \times e^{(-1.86-0.03\times age)}\right) + 1\right)}{0.19}\right)$	0.992
$u_3^*(age) = exp\left(\frac{log\left(-0.01 \times ln\left(\frac{7}{5} \times e^{(-0.85-0.03\times age)} - \frac{13}{5} \times e^{(-1.86-0.03\times age)}\right) + 1\right)}{0.19}\right)$	1.261

(b)

Figure 7. Cutoffs found by STC for biomarker sequence "Hippo->P-tau->FDG-PET" represented by (a) a tree-like plot in which green/red/grey circles represent LR/HR/inconclusive categories and sizes of the circles are in proportion to the sample size of each branch; (b) cutoffs of each biomarker as functions of "age."

classified as HR, as opposed to efficiency that includes patients classified as HR or LR. To serve this objective, we modify the STC algorithm by treating r_l as a tuning parameter ranging from 0.05 to 0.5 in increment of 0.05. We adopt the same setting as that in sub-Section 4.1; i.e., $r_h = 0.8$, $\pi(\mathbf{Z}) = 0.5$, and $\mathbf{Z} = \{age\}$. Figure 6 shows the Pareto optimal frontier for the biomarker sequences, in which each dot represents a sequence at a specific r_h (a total of 6 sequences $\times$ 10 r_h values = 60 dots). On the frontier, the sequence "Hippo->P-tau->FDG-PET" at $r_l = 0.35$ is probably the one achieving the best tradeoff between the CV-based PPV (87%) and number of HR patients (30), and therefore recommended as the biomarker testing sequence used for HR converter patient selection in AD-related clinical trials. Finally, Fig. 7 shows the cutoffs of each biomarker for the sequence "Hippo->P-tau->FDG-PET" at $r_l = 0.35$ found by STC.

5. Conclusion

In this article, we developed a STC for predicting the conversion of MCI to AD. The uniqueness of the STC is to find an optimal testing sequence of the biomarkers and two-sided cutoffs of each biomarker that satisfy pre-specified accuracy requirements while minimizing the proportion of inconclusive diagnosis. The cutoffs can be customized for each individual patient by taking into account patient demographic and genetic variables that are potential risk factors for AD. We formulated

STC into an optimization problem and performed theoretical analysis to prove the existence and uniqueness of the solution to STC. Then, we proposed two approaches for estimating the cutoffs of the biomarkers, including a numerical approach and an approximate-analytical approach, with consideration of sampling uncertainty. Next, we presented the full algorithm integrating the estimation approaches for the cutoffs with a search of the optimal sequence. Finally, we presented two applications of STC using the ADNI data. In the first application, we used STC to identify an optimal sequence of three and two biomarkers (as an example of a resource-limited situation) and the associated cutoffs for classifying MCI patients into HR converters, LR converters, or the inconclusive category. The CV-based PPV and NPV of the optimal sequence are close to the pre-specified HR and LR thresholds that reflected the expected accuracy. STC also allowed multiple criteria—e.g., accuracy and efficiency—to be optimized using a Pareto optimal frontier. The results also helped identify subtypes within HR converters. Compared with the conventional decision tree classifier, STC achieved higher PPV and NPV, saved biomarker testing costs and patient waiting time, facilitated timely medical decision making, and produced a model that is consistent with medical knowledge and biological principles and thus is clinically more trustworthy. In the other application, we used STC to identify a sub-cohort of MCI patients with a HR of converting to AD. With a slight modification of the STC algorithm, we were able to identify such a sub-cohort with a high CV-based PPV (87%) and a reasonable size appropriate for clinical trials.

Finally, we would like to point out that STC is applicable to other disease diagnosis for which multiple biomarkers need to be tested, such as Parkinson's disease and cancer. The key benefit of STC is to allow physicians to test the biomarkers sequentially with a known sequence optimized for each patient's demographic profile, and on an as-needed basis. This would save diagnostic time—a benefit to the patient—and resources—a benefit to the healthcare provider. We plan to explore the application of STC to other diseases in future work.

Funding

This work was partially funded by NSF CMMI grant #1149602. In addition, data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … and Snyder, P. J. (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, **7**(3), 270–279.

Alexopoulos, P., Kriett, L., Haller, B., Klupp, E., Gray, K., Grimmer, T., … and Drzezga, A. (2014) Limited agreement between biomarkers of neuronal injury at different stages of Alzheimer's disease. *Alzheimer's & Dementia*, **10**(6), 684–689.

Barnes, D. E., Cenzer, I. S., Yaffe, K., Ritchie, C. S., Lee, S. J., and Alzheimer's Disease Neuroimaging Initiative. (2014) A point-based tool to predict conversion from mild cognitive impairment to probable Alzheimer's disease. *Alzheimer's & Dementia*, **10**(6), 646–655.

Bowling, S. R., Khasawneh, M. T., Kaewkuekool, S., and Cho, B. R. (2009) A logistic approximation to the cumulative normal distribution. *Journal of Industrial Engineering and Management*, **2**(1), 114–127.

Box, G. E., and Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological)*, **26**, 211–252.

Cerami, C., Della Rosa, P. A., Magnani, G., Santangelo, R., Marcone, A., Cappa, S. F., and Perani, D. (2015) Brain metabolic maps in mild cognitive impairment predict heterogeneity of progression to dementia. *NeuroImage: Clinical*, **7**, 187–194.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., … and Alzheimer's Disease Neuroimaging Initiative. (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PloS One*, **6**(7), e21896.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009) The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, **11**(1), 10–18.

Heister, D., Brewer, J. B., Magda, S., Blennow, K., McEvoy, L. K., and Alzheimer's Disease Neuroimaging Initiative. (2011) Predicting MCI outcome with clinically available MRI and CSF biomarkers. *Neurology*, **77**(17), 1619–1628.

Hinrichs, C., Singh, V., Xu, G., Johnson, S. C., and Alzheimers Disease Neuroimaging Initiative. (2011) Predictive markers for AD in a multimodality framework: An analysis of MCI progression in the ADNI population. *Neuroimage*, **55**(2), 574–589.

Jack, C. R., Wiste, H. J., Vemuri, P., Weigand, S. D., Senjem, M. L., Zeng, G., … and Weiner, M. W. (2010a) Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain*, **133**(11), 3336–3348.

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

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., … and Dale, A. M. (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, **27**(4), 685–691.

Jagust, W. J., Landau, S. M., Shaw, L. M., Trojanowski, J. Q., Koeppe, R. A., Reiman, E. M., … and Mathis, C. A. (2009) Relationships between biomarkers in aging and dementia. *Neurology*, **73**(15), 1193–1199.

Jagust, W. J., Bandy, D., Chen, K., Foster, N. L., Landau, S. M., Mathis, C. A., … and Alzheimer's Disease Neuroimaging Initiative. (2010) The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer's & Dementia*, **6**(3), 221–229.

Kim, S., Swaminathan, S., Shen, L., Risacher, S. L., Nho, K., Foroud, T., … and Craig, D. W. (2011) Genome-wide association study of CSF biomarkers Aβ1-42, t-tau, and p-tau181p in the ADNI cohort. *Neurology*, **76**(1), 69–79.

Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M., Foster, N. L., Aisen, P. S., … and Weiner, M. W. (2010) Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*, **75**(3), 230–238.

Landau, S. M., Mintun, M. A., Joshi, A. D., Koeppe, R. A., Petersen, R. C., Aisen, P. S., … and Jagust, W. J. (2012) Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of Neurology*, **72**(4), 578–586.

Liu, C. C., Kanekiyo, T., Xu, H., and Bu, G. (2013) Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy. *Nature Reviews Neurology*, **9**(2), 106–118.

Misra, C., Fan, Y., and Davatzikos, C. (2009) Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *Neuroimage*, **44**(4), 1415–1422.

Risacher, S. L., Saykin, A. J., Wes, J. D., Shen, L., Firpi, H. A., and McDonald, B. C. (2009) Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research*, **6**(4), 347–361.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, **58**, 267–288.

Trittschuh, E., Larson, E., Crane, P., Cholerton, B., McCurry, S., McCormick, W., … and Craft, S. (2010) Heterogeneity of MCI characterization across two time points in a community-based population. *Alzheimer's & Dementia*, **6**(4), S78–S79.

Wee, C. Y., Yap, P. T., and Shen, D. (2013) Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping*, **34**(12), 3411–3425.

Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., … and Narayan, V. A. (2012) Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurology*, **12**(1), 1.

Yu, P., Dean, R. A., Hall, S. D., Qi, Y., Sethuraman, G., Willis, B. A., … and Schwarz, A. J. (2012) Enriching amnestic mild cognitive impairment populations for clinical trials: Optimal combination of biomarkers to predict conversion to dementia. *Journal of Alzheimer's Disease*, **32**(2), 373–385.

Zhang, D., Shen, D., and Alzheimer's Disease Neuroimaging Initiative. (2012a) Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PloS One*, **7**(3), e33182.

Zhang, N., Song, X., and Zhang, Y. (2012b) Combining structural brain changes improves the prediction of Alzheimer's disease and mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, **33**(5), 318–326.

Appendix: Proof of Proposition 2

Proof. Let $= \frac{l_i(\mathbf{Z})-(\beta_{0,i}+\beta_{z,i}^T\mathbf{Z})}{\sigma_i}$, $\delta = \frac{\beta_{y,i}}{\sigma_i}$, $r(x) = \frac{\Phi(x-\delta)}{\Phi(x)}$, and $r_0 = \frac{r_l}{1-r_l} \times \frac{1-\pi(\mathbf{Z})}{\pi(\mathbf{Z})}$. Then, the constraint in Eq. (8) becomes $r(x) \le r_0$. Here, $\delta > 0$ because $\beta_{y,i}$ represents the increase in the biomarker value as Y changes from 0 (non-diseased) to 1 (diseased). Recall that we made an assumption earlier on that there is a positive correlation between each biomarker and the disease risk, which suggests that $\beta_{y,i} > 0$. Also, $r_0 > 0$ by definition.

Next, we will show that $r(x)$ is strictly monotonically increasing from 0 to 1 as x increases from $-\infty$ to $+\infty$. When $x \to +\infty$, we have

$$\lim_{x\to+\infty} r(x) = \frac{\lim_{x\to+\infty} \Phi(x-\delta)}{\lim_{x\to+\infty} \Phi(x)} = \frac{1}{1} = 1.$$

When $x \to -\infty$, using L'Hospital's Rule, we have

$$\lim_{x\to-\infty} r(x) = \lim_{x\to-\infty} \frac{\Phi(x-\delta)}{\Phi(x)} = \lim_{x\to-\infty} \frac{\int_{-\infty}^{x-\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}{\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}$$

$$= \lim_{x\to-\infty} \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\delta)^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} = \lim_{x\to-\infty} e^{\delta x - \frac{\delta^2}{2}} = 0.$$

For finite x, $r(x)$ is strictly monotonically increasing because

$$\frac{d\,r(x)}{dx} = d\left(\frac{\int_{-\infty}^{x-\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}{\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}\right) / dx$$

$$= \frac{e^{-\frac{(x-\delta)^2}{2}} \times \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt - e^{-\frac{x^2}{2}} \times \int_{-\infty}^{x-\delta} e^{-\frac{t^2}{2}} dt}{\left(\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\right)^2}$$

$$= \frac{e^{-\frac{x^2}{2}} \times \int_{x-\delta}^{x} e^{-\frac{t^2}{2}} dt}{\left(\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\right)^2} + \frac{e^{\left(-\frac{\delta^2}{2}+x\delta\right)} \times \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt}{\left(\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\right)^2} > 0.$$

Hence, when $0 < r_0 < 1$, the feasible region of x is $(-\infty,\ x_0]$, where x_0 satisfies $r(x_0) = r_0$. Because $r(x)$ strictly monotonically increases with respect to x, the maximum $r(x)$ is achieved at x_0 and this solution is unique. When $r_0 \ge 1$, the feasible region of x is $[-\infty,\ +\infty]$. The maximum $r(x)$ is achieved at $+\infty$ and this solution is unique.