

Analyzing heterogeneity in biomarker discriminative performance through partial time-dependent receiver operating characteristic curve modeling

Statistical Methods in Medical Research

2024, Vol. 0(0) 1–13

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802241262521

journals.sagepub.com/home/smm

Xinyang Jiang¹, Wen Li² , Kang Wang³, Ruosha Li¹
and Jing Ning, for the Alzheimer's Disease
Neuroimaging Initiative³

Abstract

This study investigates the heterogeneity of a biomarker's discriminative performance for predicting subsequent time-to-event outcomes across different patient subgroups. While the area under the curve (AUC) for the time-dependent receiver operating characteristic curve is commonly used to assess biomarker performance, the partial time-dependent AUC (PAUC) provides insights that are often more pertinent for population screening and diagnostic testing. To achieve this objective, we propose a regression model tailored for PAUC and develop two distinct estimation procedures for discrete and continuous covariates, employing a pseudo-partial likelihood method. Simulation studies are conducted to assess the performance of these procedures across various scenarios. We apply our model and inference procedure to the Alzheimer's Disease Neuroimaging Initiative data set to evaluate potential heterogeneities in the discriminative performance of biomarkers for early Alzheimer's disease diagnosis based on patients' characteristics.

Keywords

Alzheimer, discriminative performance, partial AUC, pseudo partial-likelihood, time-dependent AUC

1 Introduction

Biomarkers have gained widespread use in biomedical research, encompassing various applications such as disease diagnosis, monitoring, and drug development.¹ The discriminative ability of prognostic biomarkers, which are designed for identifying the likelihood of clinical events, disease recurrence, or progression in patients with a certain medical condition or disease, must undergo rigorous evaluation before their use in practice. In cases where research outcomes are time-to-events such as time-to-death, disease progression, or relapse,² the conventional measures used for binary outcomes, including sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, have been extended to account for the dynamic nature of biomarker performance over time.^{3–7} The concept of time-dependent ROC and AUC in the context of both longitudinal and time-to-event outcomes has been extensively reviewed in Chapter 6 of “ROC Analysis for Classification and Prediction in Practice.”⁸ Additionally, joint modeling settings are also discussed in Chapter 7 of “Joint models for longitudinal and time-to-event data: With applications in R.”⁹

¹ Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston, Houston, TX, USA

² Department of Internal Medicine, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX, USA

³ Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Corresponding author:

Jing Ning, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX, 77030, USA.
jning@mdanderson.org

AUC and its time-dependent variants are commonly used as global measures to characterize the discriminative performance of biomarkers. One of their practical drawbacks is that it summarizes the entire ROC curve, although clinical and practical interests often focus on specific regions under the ROC curve rather than the full range of specificity. This is particularly relevant in cancer population screening, where high specificity is a top priority.¹⁰ Given the low incidence of cancer, it is preferable to prioritize high specificity to minimize the number of subjects who undergo unnecessary medical procedures and experience significant psychological stress.¹¹ To address this clinical need, the partial area under the ROC curve (partial time-dependent area under the curve [PAUC]) has been recommended as a suitable metric for evaluating and comparing the performance of biomarkers, aligning with the clinical interest in targeted areas of the ROC curve.^{12,13}

With the rapid advancements in biomarker research, it has become increasingly clear that identifying a single optimal biomarker with consistent discriminative performance across the entire population is not a realistic expectation. By analyzing the heterogeneity in biomarker discriminative performance, researchers can gain valuable insights into which specific groups or subpopulations exhibit favorable performance with the investigated biomarker, and conversely, which groups may not benefit as much. Understanding the limitations of biomarker performance in specific subgroups can guide further research and the development of improved biomarkers that may better serve those populations. For example, it has been reported that levels of prostate-specific antigen (PSA), a biomarker widely used to screen men for prostate cancer, tend to increase with age.¹⁴ With this finding, age-adjusted PSA would provide a more accurate screening tool for prostate cancer, particularly in older populations.

In the context of evaluating biomarker performance for binary outcomes, such as disease status, researchers have dedicated considerable attention to regression analysis to directly assess the discriminative performance using regression techniques. Particularly, several methods have been proposed for directly evaluating the performance of partial AUC using regression techniques with an emphasis on the control of higher specificity.^{13,15,16} When biomarkers are employed for risk stratification and monitoring disease progression, the outcomes of interest often involve time-to-event outcomes. In such circumstances, both nonparametric and regression approaches have been proposed to estimate the covariate-specific time-dependent ROC curve.^{17–19} To the best of our knowledge, no existing methods have been developed to model the time-dependent PAUC and investigate the heterogeneity in biomarker discriminative performance within the region of the clinically acceptable specificity over time. Therefore, our objective is to fill this research gap and gain insights into the variations in biomarker performance and its discriminatory abilities over different time periods.

One motivation study is to understand how the discriminative ability of biomarkers for Alzheimer's disease (AD) is influenced by subjects' characteristics. Data used in the preparation of this article were obtained from the AD Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.^{20,21} Among the factors being investigated, three cerebrospinal fluid (CSF) biomarkers, namely A β 42, total-tau (t-tau), and phosphorylated-tau (p-tau), have been identified as having the highest diagnostic potential.^{22,23} Furthermore, the existing literature suggests that the discriminative ability of biomarkers for the progression from MCI to AD may vary across different age groups.¹⁸ In light of these considerations, we aim to develop a robust statistical tool that rigorously evaluates the variability of a biomarker's discriminative performance using the clinically meaningful measure known as PAUC.

The remaining sections of this article are organized as follows. In Section 2, we begin by providing a definition of the covariate-specific PAUC. We then introduce the regression model for the PAUC and propose a pseudo-partial likelihood approach to estimate the model parameters. Initially, we focus on categorical covariates and subsequently extend the estimation procedure to handle continuous covariates. Furthermore, we establish the asymptotic behavior of the proposed approach. In Section 3, we conduct comprehensive simulation studies to assess the finite sample performance of the proposed methods under various scenarios. In Section 4, we apply the proposed method to the ADNI data set and evaluate how the discriminative performance of biomarkers, as characterized by the PAUC, changes with different subjects' characteristics. We conclude the article with a discussion of our findings in Section 5.

2 Methods

2.1 Notations and time-dependent measures

Denote the event time as $\tilde{T}$ and the censoring time as C , where C is assumed to be independent of $\tilde{T}$. The observed survival time is defined as $T = \min(\tilde{T}, C)$, and the event indicator is $\delta = \mathbb{I}(\tilde{T} \leq C)$, where $\mathbb{I}(\cdot)$ denotes the indicator function. The time-invariant biomarker is denoted as Y , with higher values indicating a higher risk of disease. The vector of covariates is represented by $\mathbf{X}$.

Building upon the framework proposed by Heagerty and Zheng,³ we adopt the incident/dynamic definition of time-dependent accuracy summaries to incorporate covariate information. Specifically, the covariate-specific incident sensitivity, dynamic false positive (i.e. 1-specificity), and the corresponding time-dependent ROC are defined as follows:

$$\begin{aligned} \text{TP}_t(y|\mathbf{x}) &= P(Y > y \mid \tilde{T} = t, \mathbf{X} = \mathbf{x}) \\ \text{FP}_t(y|\mathbf{x}) &= P(Y > y \mid \tilde{T} > t, \mathbf{X} = \mathbf{x}) \end{aligned}$$

and

$$\text{ROC}_t(p|\mathbf{x}) = \text{TP}_t\{\text{FP}_t^{-1}(p|\mathbf{X} = \mathbf{x})\}, p \in [0, 1]$$

To focus on regions of p that hold particular relevance for practical applications and exclude regions that are frequently not of interest, we consider the PAUC, which summarizes the discriminative performance of the biomarker in a more targeted and meaningful manner. Specifically, by incorporating time and covariate information for the partial AUC definition in Dodd and Pepe,¹³ we define the covariate-specific and time-varying PAUC as follows:

$$\begin{aligned} \text{PAUC}(t, u; \mathbf{x}) &= \int_0^u \text{ROC}_t(p|\mathbf{x}) dp \\ &= P(Y_1 > Y_2, Y_2 \in \{\text{FP}_t^{-1}(u|\mathbf{x}), \text{FP}_t^{-1}(0|\mathbf{x})\} \mid \tilde{T}_1 = t, \tilde{T}_2 > t, \mathbf{X}_1 = \mathbf{X}_2 = \mathbf{x}) \end{aligned}$$

In this definition, $\mathbf{X}_1$ represents the covariate information for the case (i.e. $\tilde{T}_1 = t$), while $\mathbf{X}_2$ represents the covariate information for the control (i.e. $\tilde{T}_2 > t$). Note that we focus on subjects with the same covariate information, following the recommendation in the literature,¹⁵ to generalize covariate-specific measures. To account for the influence of u on the scale of the $\text{PAUC}(t; u, \mathbf{x}, \boldsymbol{\theta})$, we consider a standardized PAUC^{13} as follows:

$$\text{PAUC}^*(t; u, \mathbf{x}) = \text{PAUC}(t; u, \mathbf{x})/u \quad (1)$$

We denote the range of biomarker values based on the constraint of the dynamic false positive rate as $[\text{FP}_t^{-1}(u|\mathbf{x}), \text{FP}_t^{-1}(0|\mathbf{x})]$, represented as $B_t(u|\mathbf{x})$.

2.2 Model and estimation procedures

2.2.1 Model for the covariate-specific PAUC

Motivated by the need to comprehend the heterogeneity of biomarker discriminative performance, we propose a regression model for the covariate-specific PAUC. Using a link function $\eta(\cdot)$, fractional polynomials of t with coefficients $\boldsymbol{\alpha}$, and coefficients $\boldsymbol{\beta}$ for $\mathbf{x}$, we formulate the covariate-specific PAUC as follows:

$$\eta\{\text{PAUC}^*(t; u, \mathbf{x}, \boldsymbol{\theta})\} = \sum_{k=0}^K \alpha_k t^{(p_k)} + \mathbf{x}^T \boldsymbol{\beta} \quad (2)$$

where $\boldsymbol{\theta}^T$ represents the vector of coefficients $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)$, and

$$\{t^{(p_k)}, k = 0, \dots, 7\} = \{1, t^{-2}, t^{-1}, t^{-1/2}, \log(t), t^{1/2}, t, t^2\}$$

The model in equation (2) implies the baseline PAUC when all covariates are set to 0 is

$$\text{PAUC}_0^*(t; u, \boldsymbol{\alpha}) = \eta^{-1} \{ \alpha_0 + \alpha_1 t^{-2} + \alpha_2 t^{-1} + \alpha_3 t^{-1/2} + \alpha_4 \log(t) + \alpha_5 t^{1/2} + \alpha_6 t + \alpha_7 t^2 \}$$

When employing the logit link function $\eta(\cdot)$, the exponentiation of the regression coefficient represents the odds ratio of the corresponding covariate on the partial AUC. In simpler terms, the coefficient can be interpreted as the expected change in log odds of the partial AUC per unit change in its associated covariate for continuous covariates. To exemplify, following Dodd and Pepe's work,¹³ we consider a simple model featuring one covariate, where the log odds of covariate-specific standardized PAUC are defined as follows:

$$\log \left\{ \frac{\text{PAUC}^*(t; x)}{1 - \text{PAUC}^*(t; x)} \right\} = \log \left\{ \frac{\text{PAUC}(t; x)}{u - \text{PAUC}(t; x)} \right\} = \sum_{k=0}^K \alpha_k t^{(p_k)} + \beta_1 x$$

Within this model, if X is a continuous variable, then for each one-unit increase in X , the term β_1 describes the corresponding change in the odds of the covariate-specific PAUC* at the log scale. If X is a binary variable such as gender (with female coded as 0 and male as 1), e^{β_1} indicates the odds ratio of the covariate-specific PAUC* when comparing males to females. Furthermore, if $e^{\beta_1} > 1$, it suggests that the biomarker shows better discrimination capability for males.

Acknowledging clinical interest in the discriminative accuracy evaluation of a biomarker, which often lies in the discrimination between cases and controls within the same subgroup,¹⁵ we have enforced the use of identical covariate information for both cases and controls in our regression model. Although this model constraint is clinically meaningful, it presents a new challenge in estimating regression coefficients, particularly for continuous covariates due to the inherent curse of dimensionality.²⁴ In the following sections, we present separate estimation procedures tailored for handling discrete and continuous covariates. These procedures are designed to effectively tackle the challenges introduced by the constraint of using identical covariate information for cases and controls in our regression model and also incorporate covariate distance between cases and controls to enable information borrowing.

2.2.2 Estimation procedure for discrete covariates

For a given subject i with event time $T_i = t_i$ and covariate vector $\mathbf{X}_i = \mathbf{x}_i$, we define a covariate-specific risk set as follows:

$$R(t_i, \mathbf{x}_i) = \{j : T_j > t_i, \mathbf{X}_j = \mathbf{x}_i\}$$

which is a generalization of the conventional risk set. The covariate-specific risk set incorporates specific covariate information into the risk set definition, by restricting subjects with the same covariate information to be grouped together for the classification of cases and controls. Denote the number of individuals at risk at time t with covariate $\mathbf{x}$ as $n(t, \mathbf{x})$, where $n(t, \mathbf{x}) = \sum_{j=1}^n \mathbb{I}(T_j > t, \mathbf{X}_j = \mathbf{x})$. For subject i with an observed failure time $T_i = t_i$ and covariate $\mathbf{X}_i = \mathbf{x}_i$, we focus on a conditional event within the risk set $R(t_i, \mathbf{x}_i)$ and the predefined range for the biomarker, introduced by the dynamic false positive rate constraint:

$$e_{ij} = \mathbb{I}\{Y_i > Y_j \mid \delta_i = 1, j \in R(t_i, \mathbf{x}_i), Y_j \in B_{t_i}(u|\mathbf{x}_i)\} \quad (3)$$

Given a cutoff value c , the covariate-specific dynamic false positive rate within the aforementioned risk set can be calculated non-parametrically as:

$$\text{FP}_{t_i}(c|\mathbf{x}_i) = \frac{\sum_{j \in R(t_i, \mathbf{x}_i)} \mathbb{I}(Y_j > c)}{n(t_i, \mathbf{x}_i)}$$

which allows for straightforward verification of the condition $Y_j \in B_{t_i}(u|\mathbf{x}_i)$. Under the assumption that a higher value of the biomarker indicates an increased risk of disease, we refer to the event e_{ij} as a concordant event, as subject i experienced the event earlier and had a higher value of the biomarker compared to subject j . The event e_{ij} is closely connected to the standardized PAUC and, in fact, follows a Bernoulli distribution with a probability of $\text{PAUC}^*(t_i; u, \mathbf{x}_i, \theta)$. This connection serves as the foundation for building the likelihood to estimate parameter θ . Specifically, we multiply the probabilities of all such events across all risk sets constructed from the observed data. This gives rise to the following pseudo-partial likelihood:

$$L(\theta) \propto \prod_{i=1}^n \left(\prod_{j \in R(t_i, \mathbf{x}_i)} \left[\text{PAUC}^*(t_i; u, \mathbf{x}_i, \theta)^{e_{ij}} \{1 - \text{PAUC}^*(t_i; u, \mathbf{x}_i, \theta)\}^{(1-e_{ij})} \right]^{\mathbb{I}\{Y_j \in B_{t_i}(u|\mathbf{x}_i)\}} \right)^{\delta_i} \quad (4)$$

2.2.3 Estimation procedure for continuous covariates

When dealing with continuous covariates, the strategy of grouping subjects with the same covariate information to define the at-risk set is not feasible. Alternatively, we propose to assign weights to the controls, who are event-free at the event time of a case, based on the distance in covariates between the case and its controls, using a kernel smoothing technique. For illustrative purposes, we consider the scenario with one continuous covariate. Under this approach, we utilize the conventional risk set, defined as $R(t_i) = \{j : T_j > t_i\}$. However, we incorporate covariate information to define a weighted number of subjects in the risk set: $n_w(t_i, x_i) = \sum_{j=1}^n \mathbb{I}(T_j > t_i) \mathbb{K}_h(x_i, X_j)$, where $\mathbb{K}_h(x_i, X_j) = \mathbb{K}\{(x_i - X_j)/h\}/h$ represents a kernel function with a bandwidth of h .²⁴ Under the subsequent simulations and data application, we use the Epanechnikov kernel function $\mathbb{K}(x) = 0.75(1 - x^2)I(|x| < 1)$.²⁵ We select the bandwidth as $h = \text{range}(x) \times (n \times u)^{-1/3}$, recommended by Hu et al.²⁴

Accordingly, the false positive rate would be estimated by weighting the controls based on their covariate information in the risk set:

$$FP_{t_i}^w(c|x_i) = \frac{\sum_{j \in R(t_i)} \mathbb{I}(Y_j > c) \mathbb{K}_h(x_i, X_j)}{n_w(t_i, x_i)} \quad (5)$$

We next consider two conditional events in the risk set and the predefined range for the biomarker by the dynamic false positive rate constraint, termed as kernel-based concordant and discordant events

$$\begin{aligned} e_{ij}^c &= \mathbb{K}_h(x_i, X_j) \mathbb{I} \left\{ Y_i > Y_j \mid T_i = t_i, \delta_i = 1, j \in R(t_i), Y_j \in B_{t_i}^w(u|x_i) \right\}, \text{ and} \\ e_{ij}^d &= \mathbb{K}_h(x_i, X_j) \mathbb{I} \left\{ Y_i \leq Y_j \mid T_i = t_i, \delta_i = 1, j \in R(t_i), Y_j \in B_{t_i}^w(u|x_i) \right\} \end{aligned} \quad (6)$$

where $B_{t_i}^w(u|x_i)$ denotes the biomarker range based on the weighted dynamic false positive rate. Notably, the concordant or discordant counts for pairs with similar covariate information have larger weights compared to pairs with more different covariate values. Furthermore, we construct a weighted pseudo-partial likelihood for continuous covariates as follows:

$$L(\theta) \propto \prod_{i=1}^n \left(\prod_{j \in R(t_i)} \left[\text{PAUC}^*(t_i; u, x_i, \theta)^{e_{ij}^c} \{1 - \text{PAUC}^*(t_i; u, x_i, \theta)\}^{e_{ij}^d} \right]^{\mathbb{I}\{Y_j \in B_{t_i}^w(u|x_i)\}} \right)^{\delta_i} \quad (7)$$

Note that the univariate kernel can be extended to a multivariate form, given by $\mathbb{K}_h(\mathbf{x}_i, \mathbf{X}_j) = \mathbb{K}\{\|\mathbf{x}_i - \mathbf{X}_j\|/h\}/h$. This generalization allows for the handling of multiple continuous covariates, where $\|\cdot\|$ represents the Euclidean norm. In scenarios involving both discrete and continuous covariates, we can integrate the aforementioned estimation procedures. The discrete covariates play a role in determining the composition of risk sets, while the continuous covariates influence the weights of individual members within these risk sets.

Our proposed pseudo-partial likelihood is inspired by the pseudo-likelihood, which has been demonstrated as the most effective approach for analyzing censored data within the framework of the Cox proportional hazards model. To illustrate our method, we consider a scenario with discrete covariates. At each observed failure time point, denoted as t , we construct a covariate-specific risk set. This risk set comprises both censored and uncensored observations, encompassing individuals with times exceeding t and sharing the same covariate value. We then compare the biomarker values between the subject failing at time t and all subjects within the risk set, including censored subjects. With this information, we derive the pseudo partial function as defined in equation (4). By adopting this approach, censored subjects can contribute to multiple risk sets, thus enhancing the overall likelihood estimation.

2.3 Implementation

In terms of implementation, while developing computational codes to maximize our likelihood functions should not be complex, leveraging existing software would facilitate broader adoption of our proposed methods. Recognizing the similarity between the pseudo-partial likelihood functions presented in (4) and (7) and the likelihood function commonly associated with binary outcomes, we reshape the data and leverage the existing software for fitting generalized linear model for estimation.²⁶ Algorithm 1 summarizes the implementation procedure with one discrete (e.g. X_d) and one continuous covariate (e.g. X_c), which can be easily extended to cases with multiple discrete and continuous covariates. A pivotal step involves contracting $n(t, x)$ binary outcomes at each risk set based on concordant or discordant statuses between the subject failing at time t and all subjects within the risk set, as outlined in Algorithm 1. Once more, censored subjects are included in risk sets, and their information is leveraged for constructing these binary outcomes.

2.4 Asymptotic properties

In this section, we delve into the asymptotic properties of the proposed method. Note that the concordant events or the kernel-based concordant events do not exhibit mutual independence. As a result, the likelihood formulated using these events cannot be readily treated as a conventional likelihood for inference purposes, and overcoming this challenge requires additional considerations. Without loss of generality, we assume the model includes one discrete covariate X_d and one

Algorithm 1: Estimation procedure.**Data:** Data $D = \{T_i, \delta_i, Y_i, X_{c_i}, X_{d_i}; i = 1, \dots, n\}$ **Result:** Estimator of θ **Step 1 : Reshape the data D to D^s** **begin**Rank data set by T in ascending order to obtain $D^r = \{T_k, \delta_k, Y_k, X_{c_k}, X_{d_k}, k = 1, \dots, n\}$ Initialize $D^s \leftarrow \{\}$ $r \leftarrow 1$ **for** $k = 1$ to n **do** **if** $\delta_k = 1$ **then** Obtain j in $R(t_k, x_k)$ by filtering $T_k < T_j$ and $X_{d_k} = X_{d_j}$ Rank subjects in $R(t_k, x_k)$ by Y in ascending order $FP \leftarrow 1 - \sum_{l=k}^j K_h(X_{c_k}, X_{c_l}) / \sum_{l=k}^n K_h(X_{c_k}, X_{c_l})$ **for** j in $R(t_k, x_k)$ **do** **if** $FP \leq u$ **then** $e_r \leftarrow I(Y_k > Y_j);$ $W_r \leftarrow K_h(X_{c_k}, X_{c_j})$ Append $\{e_r, T_k, X_{d_k}, X_{c_k}, W_r\}$ to D^s $r = r + 1$ **return** D^s ;// $D^s = \{e_r, T_r, X_{d_r}, X_{c_r}, W_r; r = 1, 2, \dots, R\}$ **Step 2: Estimate θ using D^s** **begin** $YS \leftarrow e$ $XS \leftarrow \{T^{-2}, T^{-1}, T^{-1/2}, \log(T), T^{1/2}, T, T^2, X_d, X_c\}$ $WT \leftarrow W$ Run fastglm(XS, YS, weights = WT, family = binomial(link = "logit")) to estimate θ

continuous covariate X_c . The log-likelihood for this scenario can be expressed as follows:

$$l(\theta) = \sum_{i=1}^n \sum_{j=1}^n \delta_i \mathbb{I}\{T_j > T_i, X_{di} = X_{dj}, Y_j \in B_{T_i}^w(u|\mathbf{X}_i)\} \mathbb{K}_h(X_{ci}, X_{cj}) \\ \times [\mathbb{I}(Y_i > Y_j) \log \{\text{PAUC}^*(T_i, \mathbf{X}_i; \theta)\} + \mathbb{I}(Y_i \leq Y_j) \log \{(1 - \text{PAUC}^*(T_i, \mathbf{X}_i; \theta))\}] \quad (8)$$

Taking the first derivative of equation (8), we have the following score equation:

$$S(\theta) = \sum_{i=1}^n \sum_{j=1}^n \delta_i \mathbb{I}\{T_j > T_i, X_{di} = X_{dj}, Y_j \in B_{T_i}^w(u|\mathbf{X}_i)\} \mathbb{K}_h(X_{ci}, X_{cj}) \\ \times \left\{ \mathbb{I}(Y_i > Y_j) \frac{\nabla_{\theta} \text{PAUC}^*(T_i, \mathbf{X}_i; \theta)}{\text{PAUC}^*(T_i, \mathbf{X}_i; \theta)} - \mathbb{I}(Y_i \leq Y_j) \frac{\nabla_{\theta} \text{PAUC}^*(T_i, \mathbf{X}_i; \theta)}{1 - \text{PAUC}^*(T_i, \mathbf{X}_i; \theta)} \right\} \quad (9)$$

where $\nabla_{\theta} f(\theta)$ is the first derivative of the function $f(\theta)$ with respect to θ . For simplification, denote $S(\theta) = \sum_{i=1}^n \sum_{j=1}^n S_{ij}(\theta)$. Let $\hat{\theta}$ be the solution of $S(\theta) = 0$, and θ_0 be the true value of θ . To establish the link between the score equation and a U-statistics with a degree of 2, we construct a symmetric kernel function by $G_{ij}(\theta) = \{S_{ij}(\theta) + S_{ji}(\theta)\}/2$. Subject to regularity conditions outlined in the supplemental materials and employing the Taylor expansion and projection theorem for U-statistics, we can establish the asymptotic properties of the estimator $\hat{\theta}$, which are summarized in the following theorem.

Theorem 1. Under the regularity conditions (A1–A6) in the supplemental materials, the estimator $\hat{\theta}$ is consistent and asymptotically follows a normal distribution with a mean of θ_0 and a covariance matrix of $V = \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}$, where $\Sigma_1 = E\{-\partial S_{12}(\theta)/\partial \theta\} |_{\theta=\theta_0}$ and $\Sigma_2 = 4\text{cov}\{G_{12}(\theta), G_{13}(\theta)\} |_{\theta=\theta_0}$.

The covariance matrix V can be consistently estimated by plugging in $\hat{\theta}$ and empirical counterparts of Σ_1 and Σ_2 . Leveraging the asymptotic normality of $\hat{\theta}$, we apply the multivariate delta method to derive the asymptotic distribution of $PAUC^*(T_i, \mathbf{X}; \hat{\theta})$ for any fixed $\mathbf{x}$ and t , which are summarized in the following corollary.

Corollary 1.1. *Under the regularity conditions (A1–A6) in the supplemental materials, $PAUC^*(t; u, \mathbf{x}, \hat{\theta})$ is a consistent estimator of $PAUC^*(t; u, \mathbf{x}, \theta_0)$, and $\sqrt{n}(PAUC^*(t; u, \mathbf{x}, \hat{\theta}) - PAUC^*(t; u, \mathbf{x}, \theta_0))$ converges to a normal distribution with a mean of 0 and a variance of $\{\nabla_{\theta}\eta(\theta_0^T \mathbf{c})\}^{-1} V \{\nabla_{\theta}\eta(\theta_0^T \mathbf{c})\}^{-1}$, where $\mathbf{c}$ denotes the vector $(1, t^{-2}, t^{-1}, t^{-1/2}, \log(t), t^{1/2}, t, t^2, \mathbf{x}^T)^T$.*

3 Simulation

For the purpose of evaluating the performance of our proposed methods, we conducted a comprehensive simulation study under various data-generating scenarios.

3.1 Simulation settings

The biomarker Y and the log of survival time $\tilde{T}$ were generated from a bivariate normal distribution with a variance of 1 and a correlation of $\rho = -0.7$. The choice of a negative correlation reflects the assumption that higher biomarker values correspond to a greater indication of disease risk. The censoring time was generated independently from a uniform distribution $U(0, \tau)$, where τ was chosen to provide the desirable censoring rate (15% or 30%). The false positive rate was controlled at $u = 0.2, 0.4, 0.6, 0.8, 1$. For each setting, sample sizes of $n = 500$ and $n = 1000$ were used. The true values of all regression coefficients were approximated by using a large data set with a sample size of $n = 30,000$.

We considered three scenarios to evaluate the performance of the proposed methods when there are different types of covariates. In Scenario 1, we evaluated the performance of the proposed methods in the presence of a discrete covariate X_d , which was generated using a Bernoulli distribution with a probability of 0.5. Subsequently, the means of the biomarker and survival time were $\mu_Y = -X_d - 1$ and $\mu_{\log(\tilde{T})} = X_d + 1$. The model used for $PAUC^*(t; x_c, \alpha, \beta)$ under this scenario was assumed to take the form:

$$\text{logit}\{PAUC^*(t; x_d)\} = \alpha_0 + \alpha_1 t^{-2} + \alpha_2 t^{-1} + \alpha_3 t^{-1/2} + \alpha_4 \log(t) + \alpha_5 t^{1/2} + \alpha_6 t + \alpha_7 t^2 + \beta_d x_d$$

The focus of Scenario 2 is to evaluate the method's performance with a continuous covariate X_c , where X_c was generated from a uniform distribution within the range of 0 to 1. The mean of the biomarker was set to $\mu_Y = -X_c - 1$, while the mean of the log of survival time was set to $\mu_{\log(\tilde{T})} = X_c + 1$. The model used for this scenario was formulated as follows:

$$\text{logit}\{PAUC^*(t; x_c)\} = \alpha_0 + \alpha_1 t^{-2} + \alpha_2 t^{-1} + \alpha_3 t^{-1/2} + \alpha_4 \log(t) + \alpha_5 t^{1/2} + \alpha_6 t + \alpha_7 t^2 + \beta_c x_c$$

In Scenario 3, two covariates were generated independently: $X_d \sim \text{Bernoulli}(0.5)$ and $X_c \sim \text{Uniform}(0, 1)$. The mean of the biomarker and the time were set to $\mu_Y = -X_d - X_c - 1$ and $\mu_{\log(\tilde{T})} = X_d + X_c + 1$. Accordingly, the model was specified as follows:

$$\text{logit}\{PAUC^*(t; x_d, x_c)\} = \alpha_0 + \alpha_1 t^{-2} + \alpha_2 t^{-1} + \alpha_3 t^{-1/2} + \alpha_4 \log(t) + \alpha_5 t^{1/2} + \alpha_6 t + \alpha_7 t^2 + \beta_d x_d + \beta_c x_c$$

3.2 Simulation results

All results were derived from 1000 simulated data sets. Due to space constraints, we have included simulation results for Scenario 3 within the main body of the paper and left simulation results for Scenarios 1 and 2 in the supplemental materials.

For Scenarios 1 and 2, empirical biases (Bias) and standard deviations (SDs) of the proposed estimators under various sample sizes (500 or 1000), censoring rates (0%, 15%, or 30%), and threshold values (u) for the PAUC are displaced in Figures S1 and S2 of the supplemental materials. These figures showcase the satisfactory performance of the estimators, revealing only a minor increase in bias under high-censoring rates or with small-threshold values. It is noteworthy that the effective sample sizes significantly depend on the PAUC's threshold value, given a specific sample size. For smaller threshold values, larger sample sizes are needed to achieve reliable estimator performance.

The empirical coverage probabilities (CPs) of 95% confidence intervals under Scenarios 1 and 2 are summarized in Figures S1 and S2 of the supplemental materials. Despite occasional underestimation of standard error (SE) when compared to the empirical SD, the CPs consistently remained in close proximity to the nominal level. Graphical representations

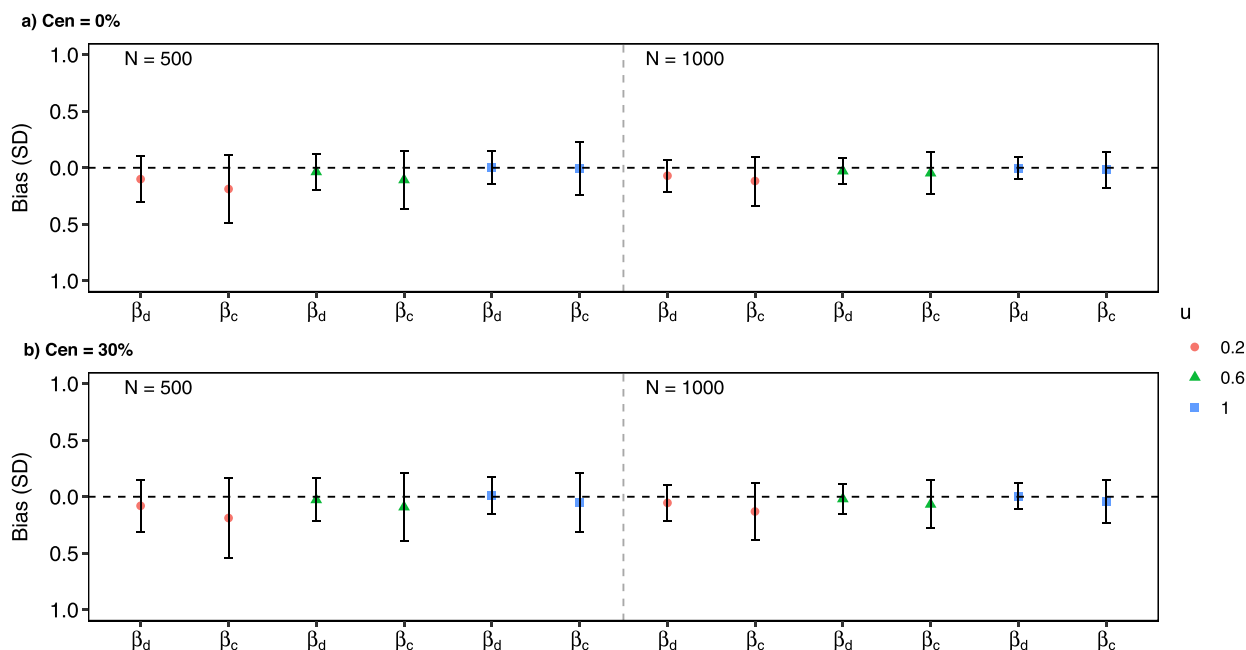


Figure 1. Simulation results: biases and empirical standard deviations in scenario 3.

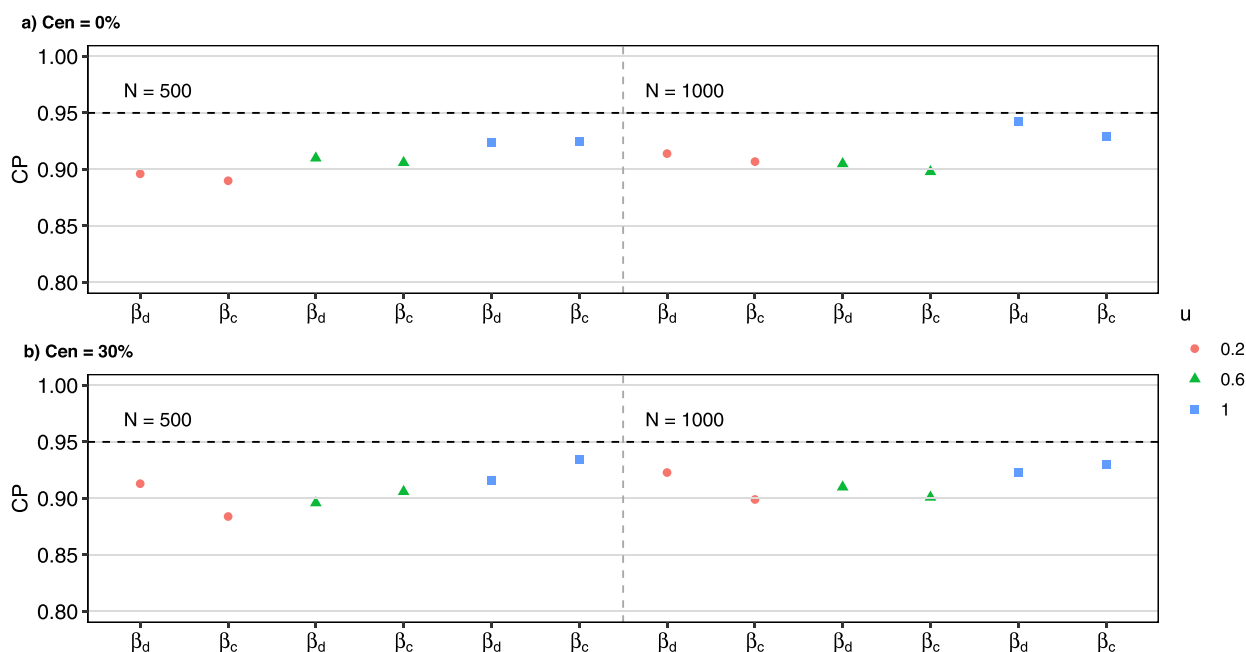


Figure 2. Simulation results: empirical coverage probabilities in scenario 3.

of estimation and confidence intervals for baseline PAUC are displaced in Figures S3 and S4. These visuals effectively illustrated the small biases of estimators to the true value, particularly with larger sample sizes. Variance escalation over time, attributed to the diminishing risk set size, was also evident. As expected, the overall performance notably improves with an increased sample size of 800, manifesting in reduced bias, variance, and coverage probabilities approaching the nominal level.

Detailed simulation results under Scenario 3 are presented in Figures 1 to 3. To elaborate, Figure 1 provides an overview of biases and SDs for estimated regression coefficients. Figure 2 illustrates the CPs of these estimates. Additionally, Figure

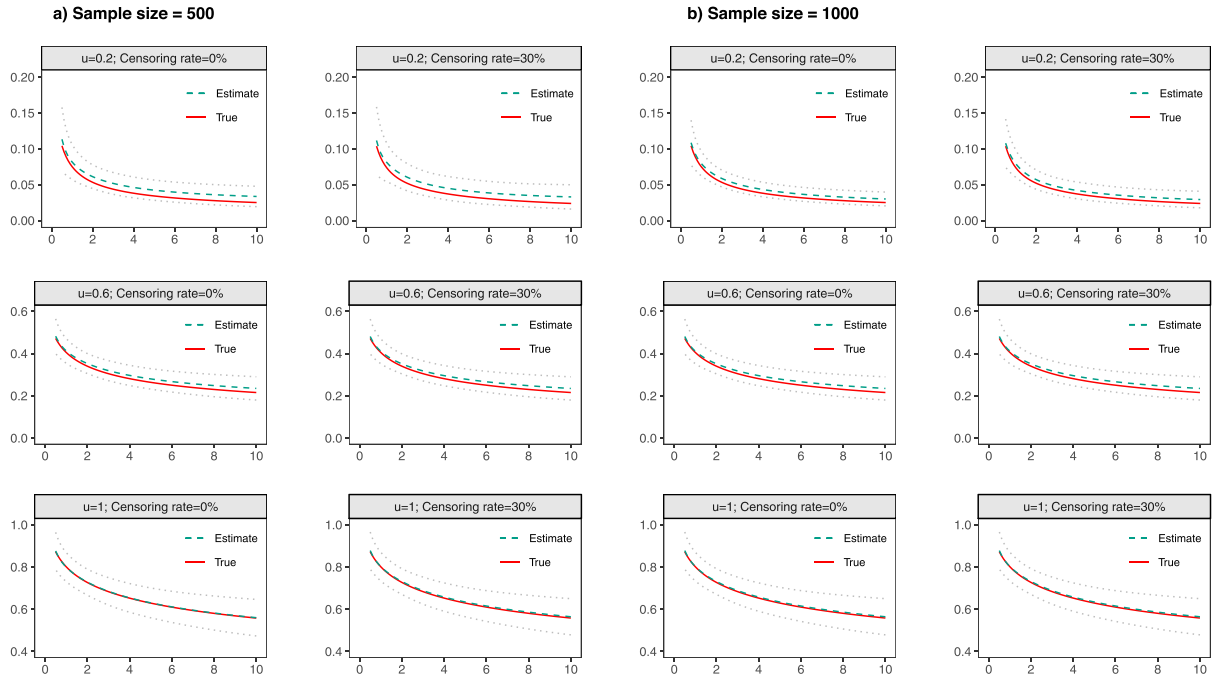


Figure 3. Estimated baseline $PAUC_0(t, \alpha)$ with 95% confidence intervals in scenario 3.

3 displays the average of the estimated baseline time-dependent partial AUC, along with their corresponding 95% empirical confidence intervals.

With both continuous and discrete covariates involved in the model, the simulation results had a minimal bias, with slightly larger biases observed under smaller effective sample sizes (i.e. smaller values of u and n , or larger censoring rates) for estimation and inference. This occasionally resulted in the underestimation of variance by using the asymptotic formula. Consequently, the CPs were marginally lower than the nominal value, although still within a reasonable range. The variance estimation bias diminished with larger sample sizes, reduced censoring rates, or larger threshold values for the PAUC, thereby yielding satisfactory CPs. Moreover, the estimation results for baseline partial AUC ($PAUC_0$) remain virtually unbiased. Consistent with previous scenarios, we observed slightly increased variance under higher censoring rates or smaller PAUC threshold values. It is worth noting that in our simulation settings, $PAUC_0(t)$ values exhibit higher initial values followed by a decreasing trend over time, indicative of the biomarker's strong short-term discriminative capability and under-optimal discriminative performance over time. For a detailed breakdown of numerical values associated with the figures, please refer to Tables S1–S3 in the supplemental materials.

Additional simulation studies have been conducted to further assess the performance of our proposed method. These studies include evaluations of models with varying numbers of polynomials included, models incorporating interaction terms, and additional settings with censoring times derived from the ADNI data. Detailed descriptions of these additional simulations and their results can be found in Section S2 of the supplemental materials.

4 Real data application

In recent studies, three CSF biomarkers— $A\beta 42$, t-tau, and p-tau—have gained prominence as effective indicators for the early detection of AD. In this context, we applied the proposed model and inference procedure to the ADNI database to investigate the impact of subjects' characteristics on the performance of these biomarkers. This analysis aims to evaluate potential heterogeneity in biomarker performance and identify subpopulations that may exhibit unfavorable performance with the investigated biomarkers, thereby highlighting the need for improved biomarkers.

During the ADNI study, participants were enrolled with varying AD statuses, including normal aging, early mild cognitive impairment, and late mild cognitive impairment or AD. These participants were prospectively followed and reassessed over time to track the progression of the disease. The primary outcome of our investigation was the time to diagnosis of AD. Our analytic cohort comprised 957 subjects after excluding prevalent cases (subjects with AD before study enrollment). The median follow-up time was 48 months, with a range from 6 to 180 months. Among these subjects, 261 (27.3%) were diagnosed with AD or died during the study period. The analytical cohort comprised 434 female patients (45.4%) and 523

Table 1. Estimated covariate effects on the PAUC of biomarker t -tau using the ADNI data.

u	Covariate	Estimate	SE	Wald	P value
0.2	Gender (female)	0.461	0.241	1.914	0.056
	Age	0.038	0.226	0.170	0.865
	Age ²	-0.237	0.249	-0.952	0.341
0.4	Gender (female)	0.391	0.210	1.862	0.063
	Age	-0.116	0.191	-0.605	0.545
	Age ²	-0.210	0.199	-1.058	0.290
0.6	Gender (female)	0.380	0.198	1.923	0.055
	Age	-0.258	0.184	-1.401	0.161
	Age ²	-0.346	0.189	-1.830	0.067
0.8	Gender (female)	0.447	0.187	2.388	0.017
	Age	-0.228	0.172	-1.328	0.184
	Age ²	-0.381	0.176	-2.165	0.030
1	Gender (female)	0.459	0.175	2.626	0.009
	Age	-0.280	0.159	-1.764	0.078
	Age ²	-0.352	0.161	-2.189	0.029

PAUC: partial time-dependent area under the curve; ADNI: Alzheimer's disease neuroimaging initiative; SE: standard error.

male patients (54.6%). The median patient age was 73 years, with an interquartile range of 66 to 78 years (Table S7 in supplemental materials).

We fitted three separate regression models to evaluate whether the discriminative ability of the three aforementioned biomarkers for AD progression would be influenced by subjects' age and gender. In these regression models, age was treated as a continuous variable centered at 75 years old, and the reference group for gender was set as male. Previous research by Pettigrew et al.²⁷ suggests that age may exhibit a nonlinear effect on AD progression, prompting the inclusion of a quadratic term in the regression model. Following this precedent, we incorporated the quadratic term in our model. Fitting results affirmed that a straightforward linear effect of age might not sufficiently capture the nuances of age-related effects. Let X_d represent gender (female = 1), and X_c represent age, where $X_c - 75$ represents age centered at 75 years. The model is specified as follows:

$$\text{logit}\{\text{PAUC}^*(t; x_d, x_c)\} = \alpha_0 + \alpha_1 t^{-2} + \alpha_2 t^{-1} + \alpha_3 t^{-1/2} + \alpha_4 \log(t) + \alpha_5 t^{1/2} + \alpha_6 t + \alpha_7 t^2 + \beta_1 x_d + \beta_2 (x_c - 75)/10 + \beta_3 \{(x_c - 75)/10\}^2$$

For $u = 0.2$, we have simplified the model specifications by employing three fractional polynomial terms $t^{1/2}$, t , and t^2 .

Table 1 summarizes the estimated regression coefficients with different threshold values on the PAUC for the biomarker t -tau, while Tables S4 and S5 of the supplemental materials list the estimated regression coefficients for the other two biomarkers. For the t -tau biomarker, the influence of subjects' age on their performance was contingent on the chosen threshold values. For values of u below or equal to 0.6, age did not yield a statistically significant impact. However, as the threshold increased, age began to play a discernible role in affecting the t -tau biomarker's performance. For instance, when $u = 1$, each additional year of age in patients aged 75 was associated with an odds ratio of 0.969. This suggests that the t -tau biomarker exhibits superior discriminatory capacity in younger patients. To determine the odds ratio for a one-year increase in age at the age of 75 after controlling for gender, we inserted the values into our model as follows:

$$\frac{\exp\{\beta_2(76 - 75)/10 + \beta_3\{(76 - 75)/10\}^2\}}{\exp\{\beta_2(75 - 75)/10 + \beta_3\{(75 - 75)/10\}^2\}} = \exp\{0.1\beta_2 + 0.1^2\beta_3\}$$

Meanwhile, the discriminatory power of t -tau was found to be more pronounced in female patients compared to male patients. Specifically, when comparing female to male subjects, the PAUC odds ratios were 1.586 (p -value = 0.056), 1.564 (p -value = 0.017), and 1.582 (p -value = 0.009) for $u = 0.2, 0.8$, and 1, respectively. We observed similar findings for the other two biomarkers and omitted detailed results here.

For a visual representation, Figure 4 plots the PAUC across different subgroups determined by subjects' gender and age. The estimated $\text{PAUC}^*(t; u, \mathbf{x}, \hat{\theta})$ function exhibited an overall descending trend over time, although the patterns were not strictly monotonic.

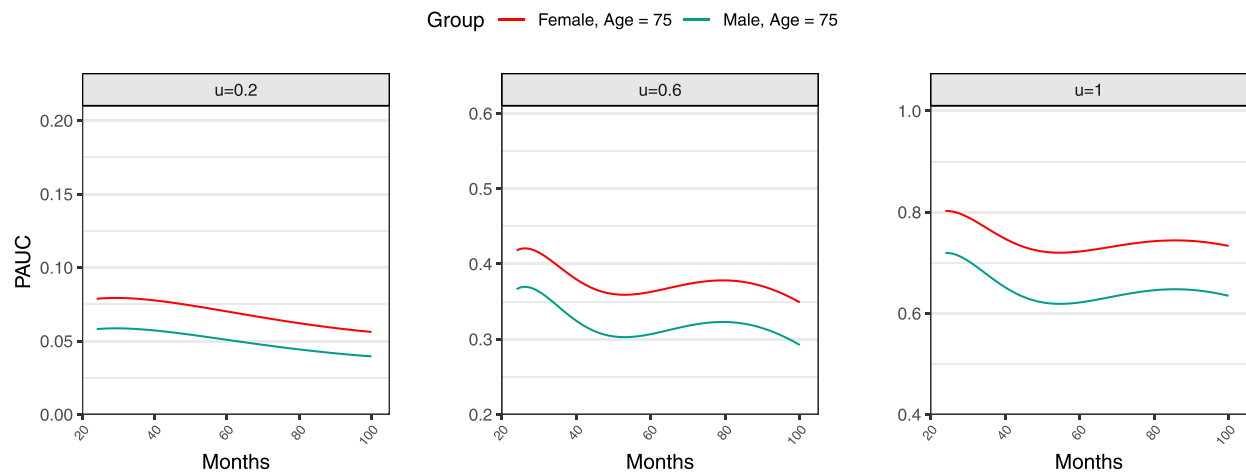


Figure 4. Estimated time-dependent time-dependent area under the curve (PAUC) curves by sex.

5 Discussion

Recognizing the clinical and practical interest in specific regions under the ROC curve, we introduce the PAUC regression model and inference procedure to comprehend the heterogeneity in the discriminative performance of biomarkers with time-to-event outcomes. One notable advantage of our proposed estimation approach lies in its simplicity for handling right censoring, obviating the need for an inverse weighting approach that involves estimating the censoring distribution. Furthermore, the analytical formula provided for variance estimation alleviates computational burden in practical applications.

While our primary focus in this paper is the assessment of heterogeneity in the discriminative performance of individual biomarkers, the proposed model and inference procedure can be extended to evaluate the performance of biomarker combinations or risk stratification scores. This extension would require the incorporation of cross-validation procedures to address additional variability when coefficients for biomarker combinations or risk stratification scores are unknown and need to be estimated. While our work does not focus specifically on formal heterogeneity testing, our proposed regression models can indeed serve as valuable tools for assessing the heterogeneity of a biomarker's effectiveness across various subgroups. By evaluating regression coefficients associated with covariates, such as age and gender, our models enable us to evaluate whether there are significant differences in the discriminative performance of the biomarker across diverse subgroups. Specifically, we can employ Wald-type tests to determine whether regression coefficients associated with these covariates are significantly different from zero. These tests allow us to investigate whether there are notable variations in the biomarker's discriminative performance among different subgroups. Identifying such disparities can be instrumental in pinpointing specific subgroups where further improvement in biomarker effectiveness may be necessary.

One assumption underlying our proposed inference procedure is the assumption of independent censoring. It is important to acknowledge that in the presence of competing risks, this assumption might be violated. However, the proposed model and inference procedure can be generalized to accommodate competing risks by drawing on the existing literature on time-dependent predictive measures.²⁸ This, however, falls beyond the scope of this present paper. Another challenge in our method is assessing model fitting for the PAUC. Standard tools, such as residual plots, are not directly applicable to evaluating the adequacy of the model. Developing rigorous statistical tools for this purpose is worthy of future research.

Acknowledgements

The authors acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing high performance computing resources that have contributed to the research results reported within this paper. Data used in the 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 the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article: This work was partially supported by awards from the National Institutes of Health (R01DK117209, U24CA230144, R01CA269696, and P50CA281701) and the Cancer Prevention and Research Institute of Texas (RP200633). Data collection and sharing for this project was funded by the 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.

ORCID iD

Jing Ning  <https://orcid.org/0000-0002-5289-331X>

Supplemental material

Supplemental material for this article is available online.

References

1. Stern AD, Alexander BM and Chandra A. Innovation incentives and biomarkers. *Clin Pharmacol Ther* 2018; **103**: 34–36.
2. Rulli E, Ghilotti F, Biagioli E, et al. Assessment of proportional hazard assumption in aggregate data: A systematic review on statistical methodology in clinical trials using time-to-event endpoint. *Br J Cancer* 2018; **119**: 1456–1463.
3. Heagerty PJ and Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**: 92–105.
4. Kamarudin AN, Cox T and Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: Current methods and applications. *BMC Med Res Methodol* 2017; **17**: 53.
5. Saha-Chaudhuri P and Heagerty PJ. Non-parametric estimation of a time-dependent predictive accuracy curve. *Biostatistics* 2013; **14**: 42–59.
6. Shen W, Ning J and Yuan Y. A direct method to evaluate the time-dependent predictive accuracy for biomarkers. *Biometrics* 2015; **71**: 439–449.
7. van Geloven N, He Y, Zwinderman AH, et al. Estimation of incident dynamic AUC in practice. *Comput Stat Data Anal* 2021; **154**: 107095.
8. Nakas C, Bantis L and Gatsonis C. *ROC analysis for classification and prediction in practice*. New York: CRC Press, 2023.
9. Rizopoulos D. *Joint models for longitudinal and time-to-event data: With applications in R*. New York: CRC Press, 2012.
10. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001; **93**: 1054–1061.
11. Lyu H, Xu T, Brotman D, et al. Overtreatment in the United States. *PLoS ONE* 2017; **12**: e0181970.
12. Ma H, Bandos AI, Rockette HE, et al. On use of partial area under the ROC curve for evaluation of diagnostic performance. *Stat Med* 2013; **32**: 3449–3458.
13. Dodd LE and Pepe MS. Partial AUC estimation and regression. *Biometrics* 2003; **59**: 614–623.
14. Janes H and Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. *Am J Epidemiol* 2008; **168**: 89–97.
15. Cai T and Dodd LE. Regression analysis for the partial area under the ROC curve. *Stat Sin* 2008; **18**: 817–836.
16. de Carvalho VI, de Carvalho M, Alonzo TA, et al. Functional covariate-adjusted partial area under the specificity-ROC curve with an application to metabolic syndrome diagnosis. *Ann Appl Stat* 2016; **10**: 1472–1495.
17. Song X and Zhou X-H. A semiparametric approach for the covariate specific ROC curve with survival outcome. *Stat Sin* 2008; **18**: 947–965.
18. Li S and Ning Y. Estimation of covariate-specific time-dependent ROC curves in the presence of missing biomarkers. *Biometrics* 2015; **71**: 666–676.
19. Rodriguez-Alvarez MX, Meira-Machado L, Abu-Assi E, et al. Nonparametric estimation of time-dependent ROC curves conditional on a continuous covariate. *Stat Med* 2016; **35**: 1090–1102.

20. Weiner MW, Aisen PS, Jack CR, Jr., et al. The Alzheimer's disease neuroimaging initiative: Progress report and future plans. *Alzheimers Dement* 2010; **6**: 202–211.e207.
21. Zhang D, Wang Y, Zhou L, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 2011; **55**: 856–867.
22. Anoop A, Singh PK, Jacob RS, et al. CSF biomarkers for Alzheimer's disease diagnosis. *Int J Alzheimers Dis* 2010; **2010**: 606802.
23. Hansson O, Seibyl J, Stomrud E, et al. CSF biomarkers of Alzheimer's disease concord with amyloid-beta PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement* 2018; **14**: 1470–1481.
24. Hu T, Nan B and Lin X. Proportional cross-ratio model. *Lifetime Data Anal* 2019; **25**: 480–506.
25. Sun Y, McCulloch CE, Marr KA, et al. Recurrent events analysis with data collected at informative clinical visits in electronic health records. *J Am Stat Assoc* 2021; **116**: 594–604.
26. Huling J. fastglm: Fast and stable fitting of generalized linear models using 'RcppEigen'. R package version 0.0. 2019;1.
27. Pettigrew C, Soldan A, Wang J, et al. Longitudinal CSF Alzheimer's disease biomarker changes from middle age to late adulthood. *Alzheimer's Dement: Diagn Assess Dis Monit* 2022; **14**: e12374.
28. Saha P and Heagerty PJ. Time-dependent predictive accuracy in the presence of competing risks. *Biometrics* 2010; **66**: 999–1011.