Published in final edited form as:

Stat Methods Med Res. 2018 March ; 27(3): 701–714. doi:10.1177/0962280217742539.

Estimating diagnostic accuracy for clustered ordinal diagnostic groups in the three-class case--application to the early diagnosis of Alzheimer disease

Chengjie Xiong^{1,2,#}, Jingqin Luo^{3,4,#}, Ling Chen¹, Feng Gao^{3,4}, Jingxia Liu^{3,4}, Guoqiao Wang^{1,2}, Randall Bateman², and John C. Morris^{2,5}

¹Division of Biostatistics, Washington University in St. Louis, St. Louis, MO, U.S.A

²Department of Neurology, Washington University in St. Louis, St. Louis, MO, U.S.A

³Division of Public health, Department of Surgery, Washington University in St. Louis, St. Louis, MO, U.S.A

⁴Biostatistics Core, Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO, U.S.A

⁵Departments of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO, U.S.A

Abstract

Many medical diagnostic studies involve three ordinal diagnostic populations in which the diagnostic accuracy can be summarized by the volume or partial volume under the Receiver Operating Characteristic (ROC) surface for a diagnostic marker. When the diagnostic populations are clustered, e.g., by families, we propose to model the diagnostic marker by a general linear mixed model that takes into account of the correlation on the diagnostic marker from members of the same clusters. This model then facilitates the maximum likelihood estimation and statistical inferences of the diagnostic accuracy for the diagnostic marker. This approach naturally allows the incorporation of covariates as well as missing data when some clusters do not have subjects on all diagnostic groups in the estimation of, and the subsequent inferences on the diagnostic accuracy. We further study the performance of the proposed methods in a large simulation study with clustered data. Finally we apply the proposed methodology to the data of several biomarkers collected by the Dominantly Inherited Alzheimer Network (DIAN), an international family-clustered registry to study autosomal dominant Alzheimer disease (ADAD) which is a rare form of Alzheimer disease (AD) caused by mutations in any of the three genes including the amyloid precursor protein, presenilin 1 and presenilin 2. We estimate the accuracy of several cerebrospinal fluid (CSF) and neuroimaging biomarkers in differentiating three diagnostic and genetic populations: normal non-mutation carriers, asymptomatic mutation carriers, and symptomatic mutation carriers.

Corresponding Author: Chengjie Xiong, Ph.D. Division of Biostatistics, Campus Box 8067, Washington University in St. Louis, St. Louis, MO 63110, U.S.A, chengjie@wubios.wustl.edu.

[#]**Co-first author.** Both CX and JL contributed equally to the manuscript.

Keywords

Alzheimer's Disease (AD); General Linear Mixed Models; Maximum Likelihood Estimate (MLE); Receiver Operating Characteristic (ROC) Surface; Sensitivity; Specificity; Volume under ROC Surface (VUS); Clustered study

1 Introduction

When the gold standard of a disease diagnosis involves more than a simple dichotomization such as healthy and diseased, ordinal rating scales of three or more options are typically used. This occurs particularly frequently when the multiple diagnostic alternatives are associated with the severity of a disease. For example, reading mammograms might be conceptualized as a task in which a radiologist has to decide whether a film depicts cancerous nodules, benign nodules, or no nodules. Another example comes from the clinical diagnosis of Alzheimer's disease (AD) in which the mild cognitive impairment (MCI) and/or very early stage of AD serves as a transitional stage between the cognitive changes of normal aging and the more serious problems caused by AD^{1,2}. The diagnostic accuracy of a diagnostic marker when the gold standard diagnosis is an ordinal rating scale of more than two options is summarized by a Receiver operating characteristic (ROC) surface or manifold from which a summary index of diagnostic accuracy, such as the volume under the surface (VUS) or the partial volume under the surface, can be derived³⁻⁷.

Most of the research in diagnostic accuracy of diagnostic markers when there are more than two ordinal diagnostic groups has focused on the data types in which subjects from multiple diagnostic groups are statistically independent. In many biomedical studies, however, subjects are enrolled in clusters, resulting in potential correlations on the diagnostic markers from the same members of the clusters. Whereas statistical methods to estimate diagnostic accuracy when the gold standard diagnosis is binary with a clustered sampling scheme have been studied^{8,9}, it remains an open question as to how to estimate the diagnostic accuracy of markers when the gold standard diagnosis is an ordinal scale of more than two options and is rendered on individuals in families or other clusters. The question is more complicated if important covariates, such as age in the study of AD, are to be incorporated in the estimation of diagnostic accuracy, especially if the unequal and imbalanced cluster sizes and missing data (i.e., not all clusters have subjects in all diagnostic groups) are also allowed in the statistical inferential process.

We develop a statistical approach to address this very question by relying on the standard general linear mixed models for clustered data. Using standard inferential procedures from the general linear mixed models, we propose the maximum likelihood estimates (MLEs) to the summary index of diagnostic accuracy, and derive the standard errors of the estimated diagnostic accuracy with clustered data. Because of the flexibility of general linear mixed model in allowing cluster-level and subject-level covariates, our method can incorporate potential covariates as well as missing data when some clusters do not have subjects on all diagnostic groups in the estimation of, and the subsequent inferences on the diagnostic accuracy, hence allowing an assessment of how covariates and potential missing data contribute to the estimated diagnostic accuracy. We further study the performance of the

proposed methods in a large simulation study with clustered data. Finally, we apply the proposed methodology to the data of several biomarkers collected by the Dominantly Inherited Alzheimer Network (DIAN)², an international family-clustered registry to study autosomal dominant Alzheimer disease (ADAD) which is a rare form of Alzheimer disease (AD) caused by mutations in any of the three genes including the amyloid precursor protein, presenilin 1 and presenilin 2. We estimate the accuracy of cerebrospinal fluid (CSF) and neuroimaging biomarkers in differentiating three diagnostic and genetic populations: normal non-mutation carriers, asymptomatic mutation carriers, and symptomatic mutation carriers.

The paper is organized as follows. The volume under a ROC surface is briefly introduced in Section 2. Section 3 elaborates on the linear mixed effects model framework to derive the estimate of diagnostic accuracy in the presence of clustered three-class ordinal diagnostic groups. Results from simulations are shown in Section 4 to examine the statistical properties of the proposed estimator. Section 5 presents the analysis results on several cerebrospinal fluid (CSF) and imaging biomarkers in the Dominantly Inherited Alzheimer Network (DIAN). We finally conclude the paper with a discussion in Section 6.

2 A summary measure of diagnostic accuracy with three ordinal diagnostic groups

Similar to Scurfield¹⁰ and Nakas and Yiannoutsos³, Xiong et al.⁴ proposed a summary measure of diagnostic accuracy when there are three ordinal diagnostic groups. Their approach was based on a Receiver Operating Characteristic (ROC) surface that was used to describe the probabilities of correct classifications into three diagnostic groups based on various sets of diagnostic thresholds of a test. Let D^+ , D^0 , and D^- denote the diseased (i.e., the positive condition) group, the early stage/very mildly diseased group, and the healthy (i.e., the negative condition) group, respectively. We assume that a biomarker or a diagnostic test in a continuous scale is used in all three groups of subjects and follows the convention that higher values of test result are associated with the higher severity of the disease. Let U be the test score for subjects in group D^- , V be the test score for subjects in group D^0 , and W be the test score for subjects in group D^+ . Let f_d , $d = +, 0, -$, be the probability density of the test in group D^d . Let F_d , $d = +, 0, -$, be the corresponding absolutely continuous and strictly increasing distribution function of the test in group D^d . Let $G_d = 1 - F_d$ for $d = +, 0, -$. Intuitively, in order for the diagnostic test to differentiate subjects from three ordinal diagnostic groups, a pair of thresholds t_- and t_+ , $t_- < t_+$, is needed. An intuitive decision rule is then to diagnose the subjects whose test results fall below t_- into the healthy subject group D^- and those with test results above t_+ to the diseased group D^+ . Therefore, subjects whose test results fall between t_- and t_+ will be diagnosed into the group D^0 . Let $x = F_-(t_-)$ be the probability that a randomly selected subject from group D^- has a test result below t_- , and let $y = G_+(t_+)$ be the probability that a randomly selected subject from group D^+ has a test result above t_+ . In the binary diagnosis between group D^- and D^+ , x and y are the specificity and sensitivity associated with threshold t_- and t_+ , respectively. The probability that a randomly selected subject from group D^0 has a test result between t_- and t_+ is given by

$$z = F_0(t_+) - F_0(t_-) = F_0(G_+^{-1}(y)) - F_0(F_-^{-1}(x)),$$

where G_+^{-1} and F_-^{-1} are the inverse function of G_+ and F_- , respectively. As a function of (x, y) , $z = z(x, y)$ defines the ROC surface in the 3-dimensional space (x, y, z) . Notice that point $(x, y, z) = (1, 1, 1)$ indicates the perfect discrimination of the diagnostic test where there exist two cutoff points $t_- < t_+$ such that the entire distribution of group D^+ would be above t_+ and the entire distribution of group D^- below t_- , while the entire distribution of group D^0 would be between t_- and t_+ . The closer a ROC surface is to this ideal point, the more accurately a pair of thresholds t_- and t_+ can be used to diagnose individuals into three ordinal diagnostic groups. The volume under the ROC surface (VUS) is used as a generalization of the AUC for a three-class classification problem. The proof for the following Equation is in Xiong et. al⁴,

$$\theta = \int \int z dx dy = \int_{-\infty}^{\infty} F_-(\eta) G_+(\eta) f_0(\eta) d\eta = \Pr(U < V < W). \quad (1)$$

As indicated by the last equality, θ is equal to the probability of correctly ranking the results of the diagnostic test from a randomly selected triple with one individual from each of the diagnostic groups.

3 Methods

We now assume a clustered design by conceptualizing that a population of subjects are clustered in a population of clusters. The sampling of the subject population typically involves two stages: a sample of clusters is first drawn, and then within each cluster, a sample of subjects was drawn. Let c be the total number of sampled clusters. Within each clusters s ($s = 1, 2, \dots, c$), let n_{sd} be the sample size of subjects in group D^d , $d = +, 0, -$. Let $n_s = n_{s-} + n_{s0} + n_{s+}$ be the total sample size from cluster s . Notice that n_s may be less than 3, and that some of the n_{sd} may be 0 within a given cluster. For $d = +, 0, -$, let

$n_d = \sum_{s=1}^c n_{sd}$. Let $n = n_- + n_0 + n_+$ be the total sample size across all sampled clusters.

Let (U, V, W) be a triplet of test scores from three randomly sampled subjects from the three diagnostic group D^- , D^0 , and D^+ , respectively. Notice that the triplet is associated with another triplet of cluster membership (I, J, K) indicating the clusters from which U, V, W , are observed (maybe from the same cluster). We propose to model (U, V, W) and (I, J, K) , with a standard general linear mixed effects model:

$$\begin{aligned} U &= C_I + U^- \\ V &= C_J + V^0 \\ W &= C_K + W^+, \end{aligned} \quad (2)$$

where C 's represent the random effects of clusters (I, J, K) to the biomarker (or diagnostic test), and U^-, V^0, W^+ account for the remaining effects in U, V, W that are beyond the random effects of clusters and, as explained below, can be modeled to incorporate the fixed effects attributable to important covariates and subject-level random errors. Assuming that (C_I, C_J, C_K) is independent of (U^-, V^0, W^+) , the marginal distribution function of U, V , and W can be obtained by convoluting the density function of C 's, $\pi(c)$, with the density function of U^-, V^0, W^+ , respectively. When there are covariates that are associated with the diagnostic test, we can further model (U^-, V^0, W^+) by a regression model of covariates (X^d representing relevant covariates of subjects in group D^d , $d = +, 0, -$) with subject-level random error, i.e.,

$$\begin{aligned}U^- &= \beta^- X^- + e^- \\V^0 &= \beta^0 X^0 + e^0 \\W^+ &= \beta^+ X^+ + e^+, \end{aligned}$$

resulting in the traditional general linear mixed model¹¹ (note here β 's may contain a group-specific intercept, as in the case of Section 4: Simulation studies below) with a cluster random effect and fixed effects from the covariates and the diagnostic groups as

$$\begin{aligned}U &= C_I + \beta^- X^- + e^- \\V &= C_J + \beta^0 X^0 + e^0 \\W &= C_K + \beta^+ X^+ + e^+, \quad (3)\end{aligned}$$

where C 's again represent the random effects of clusters (I, J, K) to the biomarker and are assumed to follow a normal distribution with mean 0 and variance σ_c^2 , and for the simplicity of presentation, X 's are observed values of a single subject-level covariate (the proposed methodology can seamlessly include more than one covariates), and e 's are subjects level random errors. Given the cluster effect C 's, e^d , $d = -, 0, +$, are assumed independent of each other, and have a normal distribution of mean 0 and a group-specific error variance σ_d^2 , $d = -, 0, +$, to accommodate potential heteroscedasticity. The independence and normal distribution assumptions of subject level random errors are the standard ones in general linear mixed effects models, and can be examined graphically and tested on marginal residuals as usual for linear mixed effects models. We further assume that C 's and e 's are all independent. Next, we can write the diagnostic accuracy index as

$$\begin{aligned}\theta &= \Pr(U < V < W) \\&= \Pr(U < V < W, I = J = K) + \Pr(U < V < W, I = J \neq K) \\&\quad + \Pr(U < V < W, I = K \neq J) + \Pr(U < V < W, I \neq J = K) \\&\quad + \Pr(U < V < W, I \neq J \neq K).\end{aligned}$$

The likelihood functions of the linear mixed model in Equation (3) can be readily found in McCulloch and Searle¹¹. Let $B=(\beta^-, \beta^0, \beta^+, \sigma_c^2, \sigma_-^2, \sigma_0^2, \sigma_+^2)$ be the entire set of parameters in Model (3), and let $\hat{B}=(\hat{\beta}^-, \hat{\beta}^0, \hat{\beta}^+, \hat{\sigma}_c^2, \hat{\sigma}_-^2, \hat{\sigma}_0^2, \hat{\sigma}_+^2)$ be its MLE. The diagnostic accuracy index θ is the summation of the following five components,

$$\begin{aligned}\theta_1 &= \Pr(U < V < W, I = J = K) \\ &= \sum_{s=1}^{\infty} \Pr(U < V < W | I = J = K = s) P(I = J = K = s) \\ \theta_2 &= \Pr(U < V < W, I = J \neq K) \\ &= \sum_{s,k=1, s \neq k}^c \Pr(U < V < W | I = J = s, K = k) \Pr(I = J = s, K = k) \\ \theta_3 &= \Pr(U < V < W, I = K \neq J) \\ &= \sum_{s,k=1, s \neq k}^c \Pr(U < V < W | I = K = s, J = k) \Pr(I = K = s, J = k) \\ \theta_4 &= \Pr(U < V < W, I \neq J = K) \\ &= \sum_{s,k=1, s \neq k}^c \Pr(U < V < W | I = s, J = K = k) \Pr(I = s, J = K = k) \\ \theta_5 &= \Pr(U < V < W, I \neq J \neq K) \\ &= \sum_{l,j,k=1, l \neq j \neq k}^c \Pr(U < V < W | I = l, J = j, K = k) \Pr(I = l, J = j, K = k)\end{aligned}$$

Estimation to the five components of the diagnostic index θ is elaborated in Appendix A1.

Whereas the linear mixed model can handle unbalanced cluster sizes across clusters well in estimating the regression parameters and variance components, i.e., $B=(\beta^-, \beta^0, \beta^+, \sigma_{c_t}^2, \sigma_e^2)$, the estimate of the diagnostic accuracy also requires estimates of cluster probabilities, i.e., $P(I = i, J = j, K = k)$. The cluster probabilities can be estimated based on cluster-specific group sample size,

$$\hat{P}(I=i, J=j, K=k, i \neq j \neq k) = \frac{n_{i-}n_{j0}n_{k+}}{n_{+}n_{0}n_{-}}$$

However, when cluster size is small, especially if the cluster size is less than 3, for some of the proposed estimators above, $\hat{P}(I=i, J=j, K=k) = 0$ because some counts in n_{i+} , n_{j0} , n_{k-} are 0. When this happens, we propose another estimator to $P(I=i, J=j, K=k)$ as

$$\tilde{P}(I=i, J=j, K=k) = \frac{n_i n_j n_k}{n^3}.$$

This new estimator is justified if clusters share the same probability for subjects being in each of the three disease groups, D^d , $d = +, 0, -$. In our simulations, we found that the use of $\hat{P}(I=i, J=j, K=k)$ resulted in biased estimation to θ , especially when there are a relatively large number of clusters with a zero count of subjects in one of the three diagnostic groups. For the rest of the paper including the simulations study and the real world application, our

focus will be on the inferences of the diagnostic accuracy θ that are based on the estimates of $B=(\beta^-, \beta^0, \beta^+, \sigma_{c-}^2, \sigma_{c0}^2, \sigma_{c+}^2, \sigma_e^2)$ from the linear mixed model and $\tilde{P}(I=i, J=j, K=k)$ as the estimators to the cluster probabilities. A comprehensive assessment on the estimates of diagnostic accuracy θ when $P(I=i, J=j, K=k)$ is estimated differentially (i.e., by $\tilde{P}(I=i, J=j, K=k)$ vs. $\tilde{P}(I=i, J=j, K=k)$ or others) will be studied in a subsequent report.

With MLEs for the fixed effects and all variance components, i.e.,

$B=(\beta^-, \beta^0, \beta^+, \sigma_c^2, \sigma_-^2, \sigma_0^2, \sigma_+^2)$, obtained from fitting the general linear mixed model through standard statistical software packages such as PROC MIXED/SAS¹² and the cluster probability estimates, the MLE for θ can be readily obtained as $\hat{\theta} = \hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3 + \hat{\theta}_4 + \hat{\theta}_5$, and its variance can be approximated through the Delta method by

$$\sigma_{\hat{\theta}}^2 = \sum_{m=1}^5 \sigma_{\hat{\theta}_m}^2 + \sum_{m_1, m_2=1, m_1 \neq m_2}^5 \sigma_{\hat{\theta}_{m_1}, \hat{\theta}_{m_2}}$$

where

$$\sigma_{\hat{\theta}_m}^2 = \frac{\partial \theta_m}{\partial B} \sum \left(\frac{\partial \theta_m}{\partial B} \right)',$$

$$\sigma_{\hat{\theta}_{m_1}, \hat{\theta}_{m_2}}^2 = \frac{\partial \theta_{m_1}}{\partial B} \sum \left(\frac{\partial \theta_{m_2}}{\partial B} \right)', m_1, m_2=1, 2, \dots, 5$$

And $\frac{\partial \theta_m}{\partial B}$ is the row vector of the derivatives of θ_m with respect to

$B=(\beta^-, \beta^0, \beta^+, \sigma_c^2, \sigma_-^2, \sigma_0^2, \sigma_+^2)$, which can be obtained by passing the derivatives to the integrand.

Whereas the optimal thresholds are not the focus of the current manuscript, they are important in real world diagnostic applications, and can be achieved by maximizing certain measures of the diagnostic accuracy, such as the generalized Youden index. The generalized Youden index can be modified to take into account weights for the various true class fractions in order to take the costs of misclassification into account if these are known¹³.

4 Simulation studies

In order to examine the performance of the proposed estimator to the diagnostic accuracy θ of a biomarker, we performed simulation studies. Because the proposed estimator is based on the estimates of regression parameters from the fixed effects and the variance components from the random effects in a general linear mixed model, it is a mixture of both finite-sample inferences (for fixed effects parameters) and asymptotic inferences (for variance components) as typically output by standard software packages such as PROC MIXED/SAS¹². It is hence of interest to also assess how the performance of the estimator to the diagnostic accuracy θ changes as the cluster size and the sample size within each cluster increase.

In the simulation studies, we considered the total number of clusters $c \in \{30, 60, 100\}$, and assumed that each cluster has the same size $n_s \in \{3, 6, 12\}$. The diagnostic groups of subjects in each family were independently simulated from a multinomial distribution

assuming equal group probabilities, $\text{multinomial}(n, (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}))$. We considered a covariate X simulated from either a uniform distribution $\text{uniform}(-3, 3)$ or a standard normal distribution $N(0, 1)$ in order to examine how the inferential procedures are influenced by the underlying distribution. We considered two parameter settings in Equation (3) for the group-specific intercept, slope and error variance parameters in the three diagnostic groups:

Setting 1:

$$\{\alpha^- = 0.2, \alpha^0 = 1, \alpha^+ = 4.8, \beta^- = -0.07, \beta^0 = 0.17, \beta^+ = 0.3, \sigma_- = 0.6, \sigma_0 = 1, \sigma_+ = 2\}$$

and

Setting 2:

$$\{\alpha^- = 3.3, \alpha^0 = 3.8, \alpha^+ = 4.3, \beta^- = 0.09, \beta^0 = 0.08, \beta^+ = 0.04, \sigma_- = 0.28, \sigma_0 = 0.43, \sigma_+ = 0.51\}$$

where α^d and β^d , $d = -, 0, +$ denote the group-specific intercept and the regression coefficient corresponding to covariate X , respectively. We chose these parameters to maintain monotonically increasing marginal means as the disease severity increases, while allowing the covariate to affect the diagnostic marker in a differential manner. We assumed an inhomogeneous error variance across the three diagnostic groups but allowed different choices of the intra-class correlation coefficient (ICC) in the simulations, i.e., $ICC = \{0.1, 0.3, 0.5\}$. As ICC is conventionally defined as

$$ICC = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_\varepsilon^2} \text{ in linear mixed effects models with a homogeneous error variance } \sigma_\varepsilon^2,$$

we allowed heteroscedasticity with group-specific error variances, σ_d^2 , $d = -, 0, +$ and

used $\sigma_\varepsilon = \frac{\sigma_- + \sigma_0 + \sigma_+}{3}$, which then implied that $\sigma_c = \sqrt{\frac{ICC}{1-ICC}} \times \sigma_\varepsilon$. This led to $\sigma_c = \{0.4, 0.7856, 1.2\}$ and $\sigma_c = \{0.1356, 0.2662, 0.4067\}$, corresponding to the specified $ICC = \{0.1, 0.3, 0.5\}$ under parameter Setting 1 and Setting 2, respectively. There were a total of 108 (=3 total numbers of cluster $\times$ 3 cluster sizes $\times$ 2 covariate distributions $\times$ 3 ICCs $\times$ 2 parameter settings) simulation scenarios. The diagnostic marker was generated from a multivariate normal distribution according to Model (3) based on the pre-specified parameters under each simulation scenario.

Under each simulation scenario, 200 independent data sets were simulated. Then, the true covariate-specific diagnostic accuracy index θ was calculated by plugging in the true parameters as specified above, and the estimate to θ was estimated by plugging in the parameter estimates from the fitted linear mixed effects models. The associated variances and the 95% nominal Wald confidence intervals (CI) were derived. The covariate-specific θ estimates were finally evaluated by biases (difference between estimate and truth), root

mean square error (RMSE: square root of the summation of squared bias and variance), and coverage of the 95% CI (the proportion among all 200 simulated data whose 95% CIs include the true θ) across all simulations under each simulation scenario. *PROC MIXED* in SAS[®] version 9.4 was used to obtain parameter estimation from the general linear mixed effects model (see Appendix A2). All other computations on estimating the covariate-specific θ and associated variance were implemented in R version 3.1.1 (<http://cran.r-project.org>) (see Appendix A3).

Figure 1, 2 and 3 presents, respectively, the bias, RMSE and empirical coverage of 95% CI of the estimated θ as a function of the covariate values across all simulated scenarios for parameter Setting 1. Supplemental Figure 1, 2, 3 in Appendix A4 are the same as Figure 1, 2, and 3, but for parameter Setting 2. Overall, the results from the two parameter settings are similar with small differences. Bias is dependent on the total sample size as determined by both the total number of clusters and within-cluster sample size. Biases decreases with increasing number of clusters. At $c=100$, biases under all combinations were close to 0 regardless cluster sizes, although smaller cluster sizes still corresponded to slightly larger biases. When data were simulated from a smaller number of clusters ($c=30$ and 60), greater variabilities in bias were observed. Larger ICCs (leading to larger cluster variance) were also associated with larger biases. The covariate-specific θ was slightly over-estimated for all combinations at $c=30$. Even with $c=30$, biases diminished to around 0 at a sample size of 12 within each cluster. In terms of RMSE, similar decreasing trends were seen with increasing total number of clusters and increasing sample size within clusters. ICC exerted much greater impact on RMSE than on bias, indicating that larger cluster variances were associated with larger RMSE, especially in presence of smaller sample sizes within clusters. The empirical coverage of 95% CI seems to be affected mainly by the sample size within clusters: with 30 clusters, nominal coverages ranged between 75% to 85% at a within-cluster size of 3 but increased quickly to around 90% with an increasing within-cluster size of 6 and 12. A larger total number of clusters boosted the coverage when within-cluster sizes are small of 3 and 6 but did not affect much when within-cluster size is as large as 12. Impact of the distribution of the covariate is obvious, especially when the total number of clusters was small. The overall decreasing trends of bias and RMSE with increasing total number of clusters and within-cluster size when the covariate was simulated from the normal distribution are similar to those from the uniform distribution. However, because covariate values simulated from the normal distribution were concentrated more around the mean, parameter estimation is more accurate around the mean than away from the mean, implying smaller bias and RMSE and better coverage around the mean and worse performance along the two extreme side. In contrast, results when the covariate was simulated from the uniform distribution are relatively more lateral. The impact of the covariate distribution was mitigated with larger total number of clusters and/or larger within-cluster sizes.

For comparison, we also implemented the naïve estimator to the diagnostic accuracy index, θ , when all subjects are assumed to be independent by ignoring the clustered design. The simulation results of the naïve estimator using the same simulated data sets as described earlier for our proposed estimator were presented in Supplemental Figure 4–9 in Appendix A5 for bias, RMSE and coverage of 95% CI under the two parameter Settings. These results clearly indicated that failure to account for clustering yielded much larger biases and RMSE,

as well as extremely poor coverage of 95% CI. Depending on the parameter settings, the bias and RMSE either monotonically increases (in parameter Setting 1) or decreases (in parameter Setting2) with covariate values. The coverage of 95% CI was all low, but was better under parameter Setting 2 than Setting 1. Similar to what were observed for our proposed estimator, larger ICCs were associated with larger biases and RMSE. Unlike in the results for our proposed estimator, cluster size mostly showed little effect on bias and RMSE.

5 Applications

Alzheimer Disease (AD) is a progressive dementing illness that typically affects adults over the age of 65. AD is the most common cause of dementia in older adults and currently affects 5.2 million people in the US. In contrast to the late onset AD, autosomal dominant Alzheimer Disease (ADAD) is a rare disorder that typically affects individuals under the age of 60¹⁴. Even though ADAD is rare, the study of the genes that cause ADAD (presenilin 1, presenilin 2, and amyloid precursor protein) has provided insights in the pathophysiology of AD, and enabled the development of animal and cell culture models of the disease¹⁵. In families with ADAD, individuals who have the causative mutation develop dementia at about the similar age as their parents, making it possible to predict the age of dementia onset with some degree of confidence. Individuals at-risk for ADAD are thus a unique population for studying the natural history of AD and for testing treatments at the early symptomatic or even asymptomatic stages. The discovery of ADAD-causing mutations has already led to a significant increase of knowledge about AD through molecular biology studies and animal models, many based on the mutations that were discovered in ADAD pedigrees. The study of the underlying pathophysiology of ADAD in human, through metabolism studies, biomarkers and imaging techniques may well further the understanding of late onset AD.

The Dominantly Inherited Alzheimer Network (DIAN)² was launched in 2008 with funding from NIH to establish an international, multicenter registry of individuals from families whose parents have a known causative mutation of AD in the amyloid precursor protein (APP), presenilin 1 (PS1), or presenilin 2 (PS2) genes. Multiple family members of age 18 or older can be enrolled in the DIAN. At the enrollment, most subjects can be classified as three ordinal types: non-mutation carriers (Non-MC), asymptomatic or preclinical mutation carriers (Asy-MC), and symptomatic mutation carriers (Sym-MC). In addition to clinical and cognitive measures, a broad spectrum of biomarkers for AD were also assessed at enrollment, including magnetic resonance imaging (MRI)-based regional brain volumes¹⁶, cerebrospinal fluid (CSF) analytes¹⁷, and molecular imaging of cerebral fibrillar amyloid with positron emission tomography (PET) using the [¹¹C] benzothiazole tracer, Pittsburgh Compound-B (PIB)¹⁸. An important scientific objective of the DIAN is to study the utility of these biomarkers in differentiating the three ordinal subjects groups: Non-MC, Asy-MC, and Sym-MC. The answer to this question has the potential to better understand the transitional stage of the more prevalent late onset AD, and given the irreversible disease processes of AD, can also provide potential targets for developing pharmaceutical treatments or preventions. A good diagnostic marker requires good discrimination among all three ordinal diagnostic groups. Here we apply our proposed methodology to assess the diagnostic accuracy of several biomarkers of AD in differentiating subjects who are Non-MC, Asy-MC,

or Sym-MC. Because age is a critical covariate of AD, in addressing the diagnostic accuracy of biomarkers, we also included age as a covariate in the analyses.

We used the same data set from an earlier DIAN data freeze we recently analyzed¹⁹ which contained baseline cognitive and biomarker data for a total of 248 DIAN subjects from 103 families. The presence or absence of dementia and, when present, its severity were operationalized with the Clinical Dementia Rating (CDR)²⁰. We focused on subjects classified into three ordinal diagnostic and genetic groups: 96 cognitively normal (i.e., CDR of 0) non-mutation carriers of any of the three genes (Non-MC), 87 cognitively normal mutation carriers (Asy-MC), and 65 subjects (Sym-MC) who were mutation carriers and already symptomatic (i.e., with CDR>0). Four biomarkers were analyzed for illustration purpose here, including two CSF markers, total tau (labeled as “tau”) and phosphorylated tau (“ptau”), and two PET PIB biomarkers: the PIB uptake in two brain regions (“prefrontal” and “precuneus”).

In the analysis, the imaging markers were first log transformed. The diagnostic accuracy index, θ , was estimated as a function of age. Because of the familial clustering, a linear mixed effect model including group specific intercepts, group-specific slopes on age, a family random effect and group-specific error variances was fitted to data from each marker. The intercept, slope, and variance parameters estimated were subsequently used to estimate age-specific θ along with a selected range of ages. Figure 4 displays, for each biomarker, the age-specific θ with 95% pointwise confidence band. Figure 5 presents the age-specific θ of all four biomarkers together. The four markers exhibit similar diagnostic ability based on estimated θ ranging between 0.17~0.28. The age-specific θ of tau and ptau decreases with age almost linearly and the two lines are almost parallel of each other, implicating the similar underlying biological process that tau and phosphorylated tau accumulate in CSF in the development of AD. Unlike the two CSF biomarkers, the two curves of the estimated θ for the PET PIB markers show similar parabolic-shaped patterns where the diagnostic ability of both markers peaks around age of 42. These estimates also imply the diagnostic utility of all four biomarkers in differentiating three ordinal groups, Non-MC, Asy-MC, and Sym-MC, as all the 95% CIs are above 1/6 (the diagnostic accuracy by chance) except for extreme ages in the graph. The very different patterns of the estimated diagnostic accuracy (as a function of age) between CSF markers and neuroimaging markers is important to note, and will be the subject of a subsequent report. Further, these results also demonstrate the importance of including important covariates in observational studies when diagnostic accuracy is assessed.

6 Discussions

Accurately diagnosing individuals into multiple and often times ordinal categories other than a simple binary classification is very important in medical practice, especially in chronic diseases such as AD. Whereas the statistical estimation and inferences of diagnostic accuracy in the literature have been largely focused on the case when subjects are categorized in a binary fashion i.e., healthy and diseased, through the sensitivities and specificities^{21–23}, the use of simple pairs of sensitivity and specificity to describe diagnostic accuracy requires one to assume---many times incorrectly---that the severity of a disease

does not matter when it comes to diagnostic accuracy, i.e., the diagnostic accuracy of a test remains constant regardless of whether subjects are at the very early or even preclinical stage or a very late stage of the disease. When the early stage of a disease is critical to detect such as in the case of AD, three ordinal diagnostic groups must be involved in the assessment of diagnostic accuracy: healthy individuals, the individuals at the early stage of the disease, and the individuals with full development of the disease. The diagnostic accuracy must then measure the correct classification of subjects into three different groups. A diagnostic test with excellent discriminating power between only two of the three diagnostic groups may not be useful in the discrimination of the three diagnostic groups. Our approach is based on the well-established index of diagnostic accuracy^{3, 4} when three ordinal diagnostic groups are used, i.e., the Receiver Operating Characteristic (ROC) surface and the associated volume under the ROC surface. Many other authors have also studied the estimation and statistical inference to the volume under ROC surface. Scurfield¹⁰ and Mossman²⁴ described the measure of discriminability based on the information theory. Heckerling²⁵ developed a parametric approach and provided a computer program to estimate the discriminatory accuracy of two diagnostic tests for three disease groups. Nakas and Yiannoutsos³ and Dreiseitl, Ohno-Machado, and Binder²⁶ proposed nonparametric methods to compare the diagnostic accuracy with ordered multiple-class ROC analysis. As proved by Xiong et al.⁴, this volume under the ROC surface is equal to the probability that the results of the diagnostic test from a randomly selected triple with one individual from each diagnostic group will be ranked in the correct order, the latter of which is what we focused on in this report.

We considered a very common design in many biomedical studies, especially in those focusing on the genetic contribution of disease phenotypes. Such studies often enroll subjects in clusters, such as familial clusters or geographical locations. An important statistical feature of such studies is the fact that data collected from the individuals within the same clusters tend to be correlated, which implies that existing statistical methods in estimating the diagnostic accuracy of markers and testing relevant hypotheses about the diagnostic accuracy do not apply. We developed a statistical method to address this very question in biomedical research by focusing on the estimation of diagnostic accuracy of diagnostic tests in a clustered design when subjects can be diagnosed in three ordinal groups. Our proposed method is based on the standard general linear mixed model on the diagnostic tests, and hence easy to incorporate covariates as fixed effects for the assessment of covariates in diagnostic accuracy. Our proposed method also maximizes the use of available data in a clustered design, allowing clusters with very small sample size (even with no subjects in some of the diagnostic groups) to contribute to the estimation and inference of the diagnostic accuracy. The MLEs for model parameters, including regression parameters from the fixed effects and variance components from the random effects are used to derive the MLEs for the diagnostic accuracy, as well as the associated variation. Whereas many of the inferences on the regression parameters from the fixed effects of the model are based on approximate *t* or *F* tests which are valid for moderate sample sizes, the inferences for variance components are mostly asymptotic, and hence requiring large sample sizes. Our simulation results indicate that the empirical 95% CI coverage seems to be affected mainly by the sample size within clusters: with 30 clusters, nominal coverages ranged between 75%

to 85% at a within-cluster size of 3 but increased quickly to around 90% with an increasing within-cluster size of 6 and 12. Our method can be easily extended to more than three ordinal diagnostic groups though the diagnostic accuracy index θ will consist of many more components (e.g., in terms of 4 ordinal groups, θ consists of 11 components). Our simulations thus provided important information about the proposed estimator of diagnostic accuracy as a function of the size of the clusters and the within-cluster sample size. Finally, we successfully implemented our proposed methodology to estimate the power of several CSF and imaging biomarkers in discriminating cognitively normal non-mutation carriers, cognitively normal mutation carriers, and symptomatic mutation carriers from the DIAN, and confirmed the utility of CSF and imaging biomarkers in diagnosing early AD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by National Institute on Aging (NIA) grant UF1 AG032438 (Dr. Bateman) and P50 AG005681 (Dr. Morris). Additional support was provided by NIA R01 AG034119 and R01 AG053550 (Dr. Xiong). The authors thank the Genetics Core (Alison Goate, DPhil, Core Leader) of the DIAN for the genetic data, Biomarker Core (Anne Fagan, PhD, Core leader) for the CSF data, and Imaging Core (Tammie Benzinger, MD and PhD, Core leader) for the imaging data. The corresponding author (C Xiong) had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Morris JC, Storandt M, Miller JP, et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol*. 2001; 58:397–405. [PubMed: 11255443]
2. Bateman RJ, Xiong C, Benzinger T, et al. Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. *The New England Journal of Medicine*. 2012; 367(9):795–804. [PubMed: 22784036]
3. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*. 2004; 23:3437–3449. [PubMed: 15505886]
4. Xiong C, van Belle G, Miller JP, et al. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*. 2006; 25:1251–1273. [PubMed: 16345029]
5. Nze Ossima AD, Daurès J-P, Bessaoud F, et al. The generalized Lehmann ROC curves: Lehmann family of ROC surfaces. *Journal of Statistical Computation and Simulation*. 2015; 85:596–607.
6. Li J, Zhou X. Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface. *J Stat Plan Inference*. 2009; 139:4133–4142.
7. Cléménçon S, Robbiano S. Building confidence regions for the ROC surface. *Pattern Recognition Letters*. 2014; 46:67–74.
8. Obuchowski NA. Nonparametric analysis of clustered ROC curve data. *Biometrics*. 1997; 53(2): 567–78. [PubMed: 9192452]
9. van Oirbeek R, Lesaffre E. Assessing the predictive ability of a multilevel binary regression model. *Computational Statistics & Data Analysis*. 2012; 56:1966–1980.
10. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology*. 1996; 40:253–269. [PubMed: 8979976]
11. McCulloch, CE., Searle, SR. Generalized, Linear, and Mixed Models. Wiley; New York: 2001.
12. Littell, R., Milliken, GA., Stroup, W., et al. SAS system for mixed models. SAS Institute Inc; 1996.
13. Bantis LE, Nakas CT, Reiser B, Myall D, Dalrymple-Alford JC. Construction of joint confidence regions for the optimal true class fractions of Receiver Operating Characteristic (ROC) surfaces and manifolds. *Statistical Methods in Medical Research*. First published date: April-24-2015.

14. Ryman DC, Acosta-Baena N, Aisen PS, et al. Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. *Neurology*. 2014; 83:253–260. [PubMed: 24928124]
15. Bateman RJ, Aisen PS, Strooper BD, et al. Autosomal-dominant Alzheimer's disease: A review and proposal for the prevention of Alzheimer's disease. *Alzheimer's Research & Therapy*. 2011; 3(1):1.doi: 10.1186/alzrt59
16. Benzinger TL, Blazey T, Jack CR, et al. Regional Variability of Imaging Biomarkers in Autosomal Dominant Alzheimer's Disease. *PNAS*. 2013; 110(47):E4502–4509. [PubMed: 24194552]
17. Fagan AM, Xiong C, Jasielec MS, et al. Longitudinal Change in CSF Biomarkers in Autosomal-Dominant Alzheimer's Disease. *Sci Transl Med*. 2014; 6:226ra30.
18. Mintun MA, LaRossa GN, Sheline YI, et al. [¹¹C] PIB in a nondemented population: Potential antecedent marker of Alzheimer disease. *Neurology*. 2006; 67:446–452. [PubMed: 16894106]
19. Luo J, D'Angela G, Gao F, et al. Bivariate correlation coefficients in family-type clustered studies. *Biometrical Journal*. 2015; 57:1084–1109. [PubMed: 26360805]
20. Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*. 1993; 43:2412–2414.
21. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
22. Wieand S, Gail MH, James BR, et al. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*. 1989; 76(3):585–592.
23. Zhou, XH., Obuchowski, NA., McClish, DK. *Statistical Methods in Diagnostic Medicine*. Wiley-Interscience; 2002.
24. Mossman D. Three-way ROCs. *Medical Decision Making*. 1999; 19:78–89. [PubMed: 9917023]
25. Heckerling PS. Parametric three-way receiver operating characteristic surface analysis using Mathematica. *Medical Decision Making*. 2001; 21:409–417. [PubMed: 11575490]
26. Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making*. 2000; 20:323–331. [PubMed: 10929855]

Appendix. A1. Estimation to the five components of θ

$$\theta_1 = \Pr(U < V < W, I = J = K) = \sum_{s=1}^{\infty} \Pr(U < V < W | I = J = K = s) P(I = J = K = s).$$

Let $\theta_{sss} = P(U < V < W | I = J = K = s)$. With a total of c clusters sampled from the entire population, θ_1 can be estimated by

$$\hat{\theta}_1 = \sum_{s=1}^c \hat{\theta}_{sss} \hat{P}(I = J = K = s), \quad (4)$$

where the sum is over all clusters with

$$\hat{P}(I = J = K = s) = \frac{n_{s+} n_{s0} n_{s-}}{n_+ n_{0+} n_{-}},$$

and

$$\hat{\theta}_{ss} = \hat{P}(U < V < W | I = J = K = s) = P(U^- < V^0 < W^+) = \int_{-\infty}^{\infty} \hat{F}^-(y) \hat{G}^+(y) \hat{f}^0(y) dy, \quad (5)$$

where $\hat{F}^-$ is the estimated distribution function of U^- , $\hat{G}^+$ is the estimated complementary cumulative distribution function of W^+ , and $\hat{f}^0$ is the estimated density function of V^0 . With the three subjects all drawn from the same cluster, the second equality of Equation (5) indicates that $\hat{\theta}_{ss}$ is equivalent to estimating VUS for independent subjects when restricted to the subjects in the same cluster only. The third equality of Equation (5) is proved in Xiong et al.⁴. In the general linear mixed effect model that assumes the normal random cluster effect and normal error distributions, let φ and Φ be the density and distribution function of the standard normal distribution, respectively. Then

$$\begin{aligned} \hat{f}^0(y) &= \varphi\left(\frac{y - \hat{\beta}^0 x^0}{\hat{\sigma}_0}\right), \\ \hat{F}^-(y) &= \Phi\left(\frac{y - \hat{\beta}^- x^-}{\hat{\sigma}_-}\right) \end{aligned}$$

and

$$\hat{G}^+(y) = 1 - \Phi\left(\frac{y - \hat{\beta}^+ x^+}{\hat{\sigma}_+}\right).$$

Similarly, θ_2 can be estimated by

$$\hat{\theta}_2 = \sum_{s,k=1, s \neq k}^c \hat{\theta}_{ssk} \hat{P}(I = J = s, K = k),$$

where

$$\hat{P}(I = J = s, K = k) = \frac{n_{s-} n_{s0} n_{k+}}{n_+ n_0 n_-},$$

and

$\theta_{ssk} = \Pr(U < V < W | I = J = s, K = k, s \neq k) = \Pr(U^- < V^0, V^0 < W^+ + C_k - C_s)$ can be estimated by

$$\hat{\theta}_{ssk} = \int_{-\infty}^{\infty} \hat{F}^{-}(y) \hat{G}_2^{+}(y) f^0(y) dy,$$

where $\hat{G}_2^{+}$ is the estimated complementary cumulative distribution function of $W^{+} + C_k - C_s$,

$$G_2^{+}(y) = 1 - \Phi \left(\frac{y - \hat{\beta}^{+} x^{+}}{\sqrt{2\hat{\sigma}_c^2 + \hat{\sigma}_+^2}} \right).$$

In a similar fashion, $\theta_3 = \Pr(U < V < W, I = K = J)$ can be estimated by

$$\hat{\theta}_3 = \sum_{s,k=1, s \neq k}^c \hat{\theta}_{sks} \hat{P}(I=K=s, J=k),$$

where

$$\hat{P}(I=K=s, J=k, s \neq k) = \frac{n_{s+}n_{k0}n_{s-}}{n_{+}n_0n_{-}}.$$

$\theta_{skk} = \Pr(U < V < W | I = K = s, J = k, s = k) = \Pr(U^{+} < V^0 - C_s + C_k < W^{+})$ can be estimated by

$$\hat{\theta}_{sks} = \int_{-\infty}^{\infty} \hat{F}^{-}(y) \hat{G}^{+}(y) \hat{f}_3^0(y) dy,$$

where $\hat{f}_3^0$ is the estimated density function of $V^0 - C_s + C_k$ i.e.,

$$\hat{f}_3^0(y) = \varphi \left(\frac{y - \hat{\beta}^0 x^0}{\sqrt{2\hat{\sigma}_c^2 + \hat{\sigma}_0^2}} \right).$$

$\theta_4 = \Pr(U < V < W, I \neq J = K)$ can be estimated by

$$\hat{\theta}_4 = \sum_{s,k=1, s \neq k}^c \hat{\theta}_{skk} \hat{P}(I=s, J=K=k),$$

where

$$\hat{P}(I=s, J=K=k, s \neq k) = \frac{n_{s-}n_{k0}n_{k+}}{n_+n_0n_-}.$$

$\theta_{skk} = \Pr(U < V < W \mid I = s, J = K = k, s \neq k) = \Pr(U^- + C_s - C_k < V^0, V^0 < W^+)$ can be estimated by

$$\hat{\theta}_{skk} = \int_{-\infty}^{\infty} \hat{F}_4^-(y) \hat{G}^+(y) \hat{f}^0(y) dy,$$

where $\hat{F}_4^-$ is the estimated cumulative distribution function of $U^- + C_s - C_k$,

$$F_4^-(y) = \Phi \left(\frac{y - \hat{\beta}^- x^-}{\sqrt{2\hat{\sigma}_c^2 + \hat{\sigma}_-^2}} \right).$$

Finally, $\theta_5 = \Pr(U < V < W, I \neq J \neq K) = \Pr(U^- + C_I < V^0 + C_J < W^+ + C_K)$ can be estimated by

$$\hat{\theta}_5 = \sum_{i,j,k=1, i \neq j \neq k}^c \hat{\theta}_{ijk} \hat{P}(I=i, J=j, K=k),$$

where

$$\hat{P}(I=i, J=j, K=k, i \neq j \neq k) = \frac{n_{i-}n_{j0}n_{k+}}{n_+n_0n_-}$$

and

$$\hat{\theta}_{ijk} = \int_{-\infty}^{\infty} \hat{F}_5^-(y) \hat{G}_5^+(y) \hat{f}_5^0(y) dy,$$

$$\hat{f}_5^0(y) = \varphi \left(\frac{y - \hat{\beta}^0 x^0}{\sqrt{\hat{\sigma}_c^2 + \hat{\sigma}_0^2}} \right),$$

$$\hat{F}_5^-(y) = \Phi \left(\frac{y - \hat{\beta}^- x^-}{\sqrt{\hat{\sigma}_c^2 + \hat{\sigma}_-^2}} \right),$$

and

$$\hat{G}_5^+(y) = 1 - \Phi \left(\frac{y - \hat{\beta}^+ x^+}{\sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_+^2}} \right).$$

A2. SAS codes to fit the linear mixed effects model in Equation (3)

The input data set should contain the scores of the diagnostic test (“Test”), diagnostic group membership (“Group”, i.e., −/0/+), a unique cluster identification (“Cluster”), and a unique subject identification (“Subject”), and a subject-level continuous covariate (“X”).

```
Proc Mixed data= method=reml asycorr asycov;
    class Group Subject Cluster;
    model Test= Group Group*X/NOINT e
s residual CovB;
    random intercept/solution
subject=Cluster g gcorr;
Run;
```

A3. Collection of R functions to estimate covariate-specific θ and associated variance using parameters estimated from the linear mixed effects model using SAS codes above

The R functions are stored in a stand-alone “.R” file.

A4. Supplemental figures on bias, RMSE and coverage of the estimated covariate-specific θ by a range of covariate value across simulation scenarios under Setting 2

A5. Supplemental figures to summarize the simulation results for the naïve estimator (i.e., ignoring the clustered design) using the same simulated data sets as described in Section 4 (Simulation studies)

Figure 1.

Under parameter **Setting 1**, bias of estimated covariate-specific θ by a range of covariate value across simulation scenarios of total number of clusters, cluster size, ICC and covariate distribution.

Figure 2. Under parameter **Setting 1**, RMSE of estimated covariate-specific θ by a range of covariate value across simulation scenarios of total number of clusters, cluster size, ICC and covariate distribution.

Figure 3.

Under parameter **Setting 1**, coverage of the 95% CIs of the estimated covariate-specific θ by a range of covariate value across simulation scenarios of total number of clusters, cluster size, ICC and covariate distribution.

Figure 4.

Age-specific estimate of θ with 95% pointwise confidence band, each of the four biomarkers of AD in a separate panel.

Figure 5.
Overlaid age-specific estimates of θ for all four biomarkers of AD.