

Estimation of the volume under a ROC surface in presence of covariates^{*}

Duc Khanh To^{a,*}, Gianfranco Adimari^a and Monica Chiogna^b

^aDepartment of Statistical Sciences, University of Padova, Via C. Battisti, 241, Padova, 35121, Italy

^aDepartment of Statistical Sciences "Paolo Fortunati", University of Bologna, Via Belle Arti, 41, Bologna, 40126, Italy

ARTICLE INFO

Keywords:

Diagnostic test
ROC analysis
Volume under ROC surface
Covariate adjustment
Location-scale models

ABSTRACT

A new method to adjust for covariate effects in the estimation of volume under a ROC surface (VUS) is presented. The method is based on the induced-regression methodology, which uses location-scale regression models to explain the relation between the test results and the covariate(s). For the estimation of the models, it is proposed to use a semiparametric generalized estimating equations (GEE) approach if the parametric forms of the mean and variance functions are specified. Alternatively, a nonparametric method is proposed, based on local linear regression (LL). In order to estimate the covariate-specific VUS, a covariate-specific Mann-Whitney representation of VUS is used, and working samples constructed after fitting the location-scale models by the GEE or LL approach. This leads to new MW-GEE and MW-LL covariate-specific VUS estimators. The asymptotic behaviour of the new estimators is investigated. More precisely, their mean squared consistency is proved. Moreover, the performance of the estimators in finite samples is explored through several simulation experiments, and an illustration, based on data from the Alzheimer's Disease Neuroimaging Initiative, is provided.

1. Introduction

Receiver operating characteristics (ROC) analysis is widely used to evaluate the ability of a diagnostic test (or biomarker) to distinguish among different states of a disease. In particular, if the disease status has two categories (e.g., diseased and non-diseased), the ROC curve and the area under the ROC curve (AUC) are classical tools that serve as measurements of test accuracy. In case of a three-class disease status (e.g., non-diseased, intermediate, diseased), the ROC surface and the volume under the ROC surface (VUS) represent natural generalizations, and several statistical methods have been developed for the estimation of these accuracy indexes; see, Zhou et al. (2011) and Pepe (2003) for comprehensive reviews.

Often, in biomedical studies, the researchers collect not only results of potential diagnostic tests but also additional information, about the subjects under study, as covariates (e.g., age, sex, comorbidity profiles). Such covariates may affect the accuracy measures for the tests.

As reviewed in Pardo Fernández et al. (2014), three methodologies are available capable to address possible covariates effects on the ROC curve and the AUC, which are based on conditional distributions, induced-regression and direct-regression, respectively. In the first case, the estimation of a covariate-specific ROC curve and a covariate-specific AUC are based on the two conditional distribution functions of the test result given the covariates values for diseased and non-diseased subjects (López-de Ullibarri et al., 2008; de Carvalho et al., 2013). As for the induced-regression methodology, two separate location-scale regression models are used to explain the covariates effect on the test results in the non-diseased class and the diseased class. Then, the covariate-specific ROC curve and the area underneath are derived by using the induced form (Pepe, 2003; Faraggi, 2003; Yao et al., 2010; González-Manteiga et al., 2011). Thus, in the induced-regression methodology, the covariates effect on the ROC curve and the AUC are evaluated indirectly. In contrast, the idea in the direct-regression methodology is to use a regression model for describing directly the effect of covariates on the ROC curve and the AUC (Alonzo and Pepe, 2002; Dodd and Pepe, 2003; Pepe, 2003; Rodríguez-Álvarez et al., 2011).

^{*} Supplemental material for this article is available online.

^{*}Corresponding author

✉ duckkhanh.to@unipd.it (D.K. To); gianfranco.adimari@unipd.it (G. Adimari); monica.chiogna2@unibo.it (M. Chiogna)
ORCID(s): 0000-0002-4641-0764 (D.K. To); 0000-0002-7811-912X (G. Adimari); 0000-0002-7238-3739 (M. Chiogna)

Studies of methods capable to account for possible effects of covariates on the ROC surface and the VUS are rarely presented in the statistical literature. To the best of our knowledge, only Li et al. (2012) discuss a method for estimating covariate-specific ROC surfaces and VUS, which is an extension of the induced-regression methodology for the two-class setting. In particular, the authors assume that the covariates effect on the test results in each class of disease can be explained by a transformation-location-scale model, with a linear form for the mean function and with constant variance. As for the parameter estimation, Li et al. (2012) propose three different approaches addressing the following three specific settings: (i) the transformation function and the density functions of error terms are known; (ii) the transformation function is unknown and the density functions of error terms are known; (iii) the transformation function has a parametric form and the density functions are unknown. The covariate-specific ROC surface is estimated by using the specific forms or the empirical estimates of the cumulative distribution functions of the error terms. Then, the covariate-specific VUS is estimated by integrating the estimate of the covariate-specific ROC surface.

In this paper, we propose an alternative, more general, approach to estimate the covariate-specific VUS, that relies on the induced-regression methodology and the working sample technique (Yao et al., 2010). More precisely, to describe the relation between test result and covariates, in each disease class we consider a location-scale regression model with heteroscedastic error. To estimate the mean and variance functions, we employ parametric and nonparametric estimation techniques, namely, generalized estimating equations and local linear regressions. Then, to estimate the covariate-specific VUS, we consider a nonparametric covariate-specific Mann-Whitney representation of VUS (see Nakas and Yiannoutsos (2004) for the simple setting with no covariates) and apply the plug-in principle. This step relies on working samples, constructed at the specific set of values for the covariates. We also discuss how to estimate the covariate-adjusted VUS (Li et al., 2012).

The asymptotic behaviour of our proposed estimators is investigated. More precisely, we prove that the proposed covariate-specific VUS estimators are mean squared (MS) consistent. Then, consistency of the covariate-adjusted VUS estimator is easily derived by the continuous mapping theorem. In order to estimate the variances of the estimators and to construct confidence intervals, we propose to use a bootstrap procedure. The performance of our proposed estimators in finite samples is explored through several simulation experiments.

Compared with the model in Li et al. (2012), our solution appears to be more flexible, as it allows any parametric or nonparametric forms for the mean and variance functions. Furthermore, our approach does not require any assumption about the distributions of the error terms. Note that our covariate-specific VUS estimators do not require estimation of the covariate-specific ROC surface; as a consequence, our approach is computationally advantageous when compared with the one in Li et al. (2012).

The paper is organized as follows. In Section 2, we discuss about the model settings and the proposed estimators for the mean and variance functions and for the covariate-specific VUS. In Section 3, we present our theoretical results about asymptotic behaviour of our proposed estimators. The simulation results are presented in Section 4, and an illustration, based on data from the Alzheimer's Disease Neuroimaging Initiative, is provided in Section 5. Concluding remarks are left to Section 6.

2. Model and estimation

2.1. Notation and model setting

Let Y_1, Y_2, Y_3 be the test results for subjects with respect to three classes of disease status, say 1, 2 and 3, respectively. Let X denote a covariate (or a set of covariates) which could affect the accuracy of the diagnostic test. In what follows, to ease presentation, we will refer to a single covariate, deferring to Section 6 a discussion about the extension to the multidimensional case.

To model the relationship between X and the test results, we specify the following location-scale regression models:

$$\begin{aligned} Y_1 &= \mu_1(X) + \sigma_1(X)\varepsilon_1, \\ Y_2 &= \mu_2(X) + \sigma_2(X)\varepsilon_2, \\ Y_3 &= \mu_3(X) + \sigma_3(X)\varepsilon_3, \end{aligned} \tag{1}$$

where $\mu_j(\cdot) = \mathbb{E}(Y_j|X = \cdot)$ and $\sigma_j^2(\cdot) = \mathbb{V}\text{ar}(Y_j|X = \cdot)$ are the conditional mean and conditional variance of the test result Y_j given the covariate X , respectively; the errors ε_j are independent of each other and have zero means and unit variances. Let $F_j(y|x) = \Pr(Y_j \leq y|X = x)$ be the conditional cumulative distribution functions (c.d.f.) of Y_j given X , and $G_j(\cdot)$ be the cumulative distribution functions of error ε_j .

Hereafter, we assume that higher values of the test are associated to higher severity of the disease at a given value x of covariate X , i.e., the intrinsic natural stochastic ordering is $Y_1 < Y_2 < Y_3$, for each fixed $X = x$ (see Li et al., 2014 for a discussion about situations in which the ordering is not a priori known). The covariate-specific volume under the ROC surface (VUS) is then defined as

$$\theta(x) = \Pr(Y_1 < Y_2 < Y_3 | X = x). \quad (2)$$

Let n_1 , n_2 and n_3 denote the sample sizes for the three classes of disease. Consider the following covariate-specific quantity:

$$\theta_{\text{MW}}(x) = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} \sum_{k=1}^{n_3} I(Y_{1i}(x) < Y_{2r}(x) < Y_{3k}(x)), \quad (3)$$

where $I(\cdot)$ stands for the indicator function,

$$Y_{1i}(x) = \mu_1(x) + \sigma_1(x)\varepsilon_{1i}, \quad Y_{2r}(x) = \mu_2(x) + \sigma_2(x)\varepsilon_{2r}, \quad Y_{3k}(x) = \mu_3(x) + \sigma_3(x)\varepsilon_{3k}, \quad (4)$$

and ε_{1i} , ε_{2r} and ε_{3k} are the standardized residuals:

$$\varepsilon_{1i} = \frac{Y_{1i} - \mu_1(X_{1i})}{\sigma_1(X_{1i})}, \quad \varepsilon_{2r} = \frac{Y_{2r} - \mu_2(X_{2r})}{\sigma_2(X_{2r})}, \quad \varepsilon_{3k} = \frac{Y_{3k} - \mu_3(X_{3k})}{\sigma_3(X_{3k})}. \quad (5)$$

In (5), for a subject in the j -th class, $j = 1, 2, 3$, Y_{ji} is a test value and X_{ji} is the associated observed covariate value.

When $Y_{ji}(x) = Y_{ji}$, that is, in absence of covariates, quantity (3) reduces to the nonparametric Mann-Whitney (MW) estimator of the VUS (Nakas and Yiannoutsos, 2004). In presence of covariates, (3) could still be used to estimate the covariate-specific VUS provided that both functions $\mu_j(\cdot)$, $\sigma_j^2(\cdot)$ and all the standardized residuals were known, a condition that allows to construct the working samples $\{Y_{1i}(x)\}_{i=1, \dots, n_1}$, $\{Y_{2r}(x)\}_{r=1, \dots, n_2}$ and $\{Y_{3k}(x)\}_{k=1, \dots, n_3}$, that pretend that all units are observed at the value $X = x$. In practice, however, such knowledge is not available, so that $\mu_j(\cdot)$ and $\sigma_j(\cdot)$ have to be estimated, along with the standardized residuals and the working samples. In the next subsections, we discuss two approaches for estimating such quantities starting from the location-scale regressions (1). The first approach is a parametric approach; the second, a non-parametric one.

With a slight abuse of terminology, we name quantity (3) *Mann-Whitney representation* of VUS (MW VUS) in the paper. Moreover, to account for possible ties, the definition of $\theta_{\text{MW}}(x)$ in (3) can be rewritten as

$$\begin{aligned} \theta_{\text{MW}}(x) = & \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} \sum_{k=1}^{n_3} \left\{ I(Y_{1i}(x) < Y_{2r}(x) < Y_{3k}(x)) + \frac{1}{2} I(Y_{1i}(x) < Y_{2r}(x) = Y_{3k}(x)) \right. \\ & \left. + \frac{1}{2} I(Y_{1i}(x) = Y_{2r}(x) < Y_{3k}(x)) + \frac{1}{6} I(Y_{1i}(x) = Y_{2r}(x) = Y_{3k}(x)) \right\}. \end{aligned} \quad (6)$$

Finally, note that, the distributions of error terms ε_j , $j = 1, 2, 3$, do not depend on the covariate X . Thus, terms ε_{ji} in (4) are realizations of independent and identically distributed (i.i.d.) random variables, with c.d.f. G_j for $i = 1, \dots, n_j$, $j=1,2,3$. As a consequence, $Y_{1i}(x)$, $Y_{2r}(x)$ or $Y_{3k}(x)$ in (3), (4) and (6) are realizations of i.i.d. random variables.

2.2. GEE estimation

Assume that the conditional mean and variance functions can be written in some parametric form, such as $\mu_j(x) = h_j(\beta_j; x)$ and $\sigma_j^2(x) = g_j(\alpha_j; x) > 0$, where h_j and g_j are known smooth functions. Then, the estimators of parameters β_1 , β_2 , β_3 and α_1 , α_2 , α_3 can be obtained through the method of generalized estimating equations (GEE), i.e., by solving the following equations (Liu and Zhou, 2011):

$$\begin{aligned} \mathbf{0} &= U_{1j}(\beta_j; \alpha_j) = \sum_{i=1}^{n_j} \left(\frac{\partial h_j(\beta_j; X_{ji})}{\partial \beta_j} \right)^\top \frac{Y_{ji} - h_j(\beta_j; X_{ji})}{g_j(\alpha_j; X_{ji})}, \\ \mathbf{0} &= U_{2j}(\alpha_j; \beta_j) = \sum_{i=1}^{n_j} \left(\frac{\partial g_j(\alpha_j; X_{ji})}{\partial \alpha_j} \right)^\top \frac{\{Y_{ji} - h_j(\beta_j; X_{ji})\}^2 - g_j(\alpha_j; X_{ji})}{g_j^2(\alpha_j; X_{ji})}, \end{aligned}$$

$j = 1, 2, 3$. The solutions $\hat{\beta}_j$ and $\hat{\alpha}_j$ can be obtained by the modified Fisher scoring method. The iterative procedure is described as follows.

1. Choose starting values $\beta_j^{(0)}$ and $\alpha_j^{(0)}$.
2. For $t = 0$, obtain

$$\beta_j^{(t+1)} = \beta_j^{(t)} + \left\{ \sum_{i=1}^{n_j} \frac{D_{ji}^{(t)\top} D_{ji}^{(t)}}{g_j(\alpha_j^{(t)}; X_{ji})} \right\}^{-1} \left[\sum_{i=1}^{n_j} \frac{D_{ji}^{(t)\top} \{Y_{ji} - h_j(\beta_j^{(t)}; X_{ji})\}}{g_j(\alpha_j^{(t)}; X_{ji})} \right],$$

$$\alpha_j^{(t+1)} = \alpha_j^{(t)} + \left\{ \sum_{i=1}^{n_j} \frac{E_{ji}^{(t)\top} E_{ji}^{(t)}}{g_j^2(\alpha_j^{(t)}; X_{ji})} \right\}^{-1} \left[\sum_{i=1}^{n_j} E_{ji}^{(t)\top} \frac{\{Y_{ji} - h_j(\beta_j^{(t)}; X_{ji})\}^2 - g_j(\alpha_j^{(t)}; X_{ji})}{g_j^2(\alpha_j^{(t)}; X_{ji})} \right],$$

where

$$D_{ji}^{(t)} = \left. \frac{\partial h_j(\beta_j; X_{ji})}{\partial \beta_j} \right|_{\beta_j = \beta_j^{(t)}}, \quad E_{ji}^{(t)} = \left. \frac{\partial g_j(\alpha_j; X_{ji})}{\partial \alpha_j} \right|_{\alpha_j = \alpha_j^{(t)}}.$$

3. Set $t = t + 1$ and repeat step 2 until $\|\beta_j^{(t+1)} - \beta_j^{(t)}\|_2 < \delta$ and $\|\alpha_j^{(t+1)} - \alpha_j^{(t)}\|_2 < \delta$, where δ is a small arbitrary number, say $\delta = 10^{-9}$, and $\|\cdot\|_2$ is a Euclidean norm.
4. Finally, let $\hat{\beta}_j = \beta_j^{(t+1)}$ and $\hat{\alpha}_j = \alpha_j^{(t+1)}$.

Plugging the GEE estimates $\hat{\beta}_j$ and $\hat{\alpha}_j$ into the functions $h_j(\cdot; x)$ and $g_j(\cdot; x)$, gives rise to the GEE estimates $\hat{\mu}_{j,\text{GEE}}(x)$ and $\hat{\sigma}_{j,\text{GEE}}^2(x)$ of the conditional means and conditional variances, respectively. In practice, in order to guarantee positiveness of the estimated variances, we suggest to use the exponential transformation $\exp(g_j(\alpha_j; x))$ before starting the algorithm.

2.3. Local linear regression estimation

In a non-parametric framework, the conditional means $\mu_j(x)$ and the conditional variances $\sigma_j^2(x)$ appearing in the location-scale regression models (1) could be estimated by using the classical Nadaraya-Watson kernel estimator; however, as such estimator suffers from boundary and design bias, in this paper we propose the use of the so-called local linear estimator.

Let $K(\cdot)$ be a symmetric kernel function and

$$\mathbf{Z}_{jx} = \begin{pmatrix} 1 & (X_{j1} - x) \\ 1 & (X_{j2} - x) \\ \vdots & \vdots \\ 1 & (X_{jn_j} - x) \end{pmatrix},$$

with $j = 1, 2, 3$, the design matrices. Let

$$\mathbf{W}_{x,u_j} = \text{diag} \left\{ K_{u_j}(X_{j1} - x), \dots, K_{u_j}(X_{jn_j} - x) \right\},$$

where $K_{u_j}(\cdot) = K(\cdot/u_j)$ and $u_j \equiv u_j(n_j)$ is a bandwidth sequence. The local linear regression estimator of the conditional mean $\mu_j(x)$ is given by (Wasserman, 2006)

$$\hat{\mu}_{j,\text{LL}}(x) = \mathbf{e}_1^\top \left(\mathbf{Z}_{jx}^\top \mathbf{W}_{x,u_j} \mathbf{Z}_{jx} \right)^{-1} \mathbf{Z}_{jx}^\top \mathbf{W}_{x,u_j} \mathbf{Y}_j, \quad (7)$$

where $\mathbf{e}_1 = (1, 0)^\top$. Similarly, we could use the local linear regression estimation for the conditional variances $\sigma_j^2(x)$. However, this approach does not guarantee that $\hat{\sigma}_j^2(x) > 0$. Thus, we employ the log-transform based method proposed by Chen et al. (2009). More precisely, we rewrite the location-scale regression models (1) as

$$\log(r_j) = \gamma_j(X) + \log(\varepsilon_j^2/d_j), \quad (8)$$

where $r_j = \{Y_j - \mu_j(X)\}^2$, $\gamma_j(\cdot) = \log(d_j \sigma_j^2(\cdot))$, and d_j that satisfies $\mathbb{E} \left\{ \log \left(\varepsilon_j^2 / d_j \right) \right\} = 0$. A local linear estimator $\hat{\gamma}_j(x)$ is obtained as \hat{a}_{j0} , where

$$(\hat{a}_{j0}, \hat{a}_{j1}) = \arg \min_{(a_{j0}, a_{j1})} \sum_{i=1}^{n_j} \left\{ \log \left(\hat{r}_{ji} + n_j^{-1} \right) - a_{j0} - a_{j1} (X_{ji} - x) \right\}^2 K_{v_j} (X_{ji} - x),$$

and $v_j \equiv v_j(n_j)$ is a bandwidth sequence. Here, $\log(\hat{r}_{ji} + n_j^{-1})$ is used to avoid $\log(0)$. Then, the estimator $\hat{\sigma}_{j,LL}^2(x)$ is obtained as $\exp \{ \hat{\gamma}_j(x) \} / \hat{d}_j$ with

$$\hat{d}_j = n_j \left[\sum_{i=1}^{n_j} \hat{r}_{ji} \exp \{ -\hat{\gamma}_j(X_{ji}) \} \right]^{-1}.$$

The estimator \hat{d}_j is derived from the fact that $\mathbb{E}(\varepsilon_{ji}^2 | X_{ji}) = 1$ and $r_{ji} = \exp \{ \gamma_j(X_{ji}) \} \varepsilon_{ji}^2 / d_j$ (Chen et al., 2009).

2.4. VUS estimation

By plugging the parametric or non-parametric estimators $\hat{\mu}_{j,*}(X_{ji})$ and $\hat{\sigma}_{j,*}^2(X_{ji})$ into (5), one obtains the estimated standardized residuals

$$\hat{\varepsilon}_{1i,*} = \frac{Y_{1i} - \hat{\mu}_{1,*}(X_{1i})}{\hat{\sigma}_{1,*}(X_{1i})}, \quad \hat{\varepsilon}_{2r,*} = \frac{Y_{2r} - \hat{\mu}_{2,*}(X_{2r})}{\hat{\sigma}_{2,*}(X_{2r})}, \quad \hat{\varepsilon}_{3k,*} = \frac{Y_{3k} - \hat{\mu}_{3,*}(X_{3k})}{\hat{\sigma}_{3,*}(X_{3k})} \quad (9)$$

where the star * stands for the type of estimator, i.e., GEE or LL. Therefore, the estimated working samples at $X = x$ are obtained as

$$\begin{aligned} \hat{Y}_{1i,*}(x) &= \hat{\mu}_{1,*}(x) + \hat{\sigma}_{1,*}(x) \hat{\varepsilon}_{1i,*}, \\ \hat{Y}_{2r,*}(x) &= \hat{\mu}_{2,*}(x) + \hat{\sigma}_{2,*}(x) \hat{\varepsilon}_{2r,*}, \\ \hat{Y}_{3k,*}(x) &= \hat{\mu}_{3,*}(x) + \hat{\sigma}_{3,*}(x) \hat{\varepsilon}_{3k,*}, \end{aligned}$$

and the covariate-specific VUS estimators are

$$\hat{\theta}_{MW-*}(x) = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} \sum_{k=1}^{n_3} \mathbb{I} \left(\hat{Y}_{1i,*}(x) < \hat{Y}_{2r,*}(x) < \hat{Y}_{3k,*}(x) \right), \quad (10)$$

where, again, the star * stands for GEE or LL. When ties are present in the data, we use

$$\begin{aligned} \hat{\theta}_{MW-*}(x) &= \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} \sum_{k=1}^{n_3} \left\{ \mathbb{I} \left(\hat{Y}_{1i,*}(x) < \hat{Y}_{2r,*}(x) < \hat{Y}_{3k,*}(x) \right) \right. \\ &\quad + \frac{1}{2} \mathbb{I} \left(\hat{Y}_{1i,*}(x) < \hat{Y}_{2r,*}(x) = \hat{Y}_{3k,*}(x) \right) + \frac{1}{2} \mathbb{I} \left(\hat{Y}_{1i,*}(x) = \hat{Y}_{2r,*}(x) < \hat{Y}_{3k,*}(x) \right) \\ &\quad \left. + \frac{1}{6} \mathbb{I} \left(\hat{Y}_{1i,*}(x) = \hat{Y}_{2r,*}(x) = \hat{Y}_{3k,*}(x) \right) \right\}. \quad (11) \end{aligned}$$

Observe that the proposed MW-GEE estimator is a semi-parametric estimator, because it requires the specification of parametric forms to estimate the conditional means $\mu_j(\cdot)$ and the conditional variances $\sigma_j^2(\cdot)$, but it relies on a nonparametric approach to estimate the covariate-specific VUS. Meanwhile, the MW-LL estimator is fully non-parametric.

In addition to evaluate covariate-specific VUS at relevant covariate values, it may be useful to estimate a synthetic index that provides an average measure of the VUS, across the distribution of X . Such an operation is easy if X is univariate. For X continuous, as in Li et al. (2012), we can define a covariate-adjusted VUS as

$$AVUS = \int_{\Omega} \theta(x) f_X(x) dx, \quad (12)$$

where Ω is the domain of X . In practice, one can focus only on an interval of covariate values, say $[a, b]$, instead of the entire domain Ω . In this case, we can use a truncated covariate-adjusted VUS defined as follows

$$\text{tAVUS} = \int_a^b \theta(x) \frac{f_X(x) I_{[a,b]}(x)}{F_X(b) - F_X(a)} dx. \quad (13)$$

The (marginal) density function $f_X(\cdot)$ of X can be estimated by using the histogram estimator, i.e.,

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n \mathbf{I} \left(\frac{|X_i - x|}{h} \leq \frac{1}{2} \right),$$

where h is a bandwidth, set as $h = 2\text{IQR}n^{-1/3}$ (Freedman and Diaconis, 1981), and IQR denotes the interquartile range of X . Moreover, $n = n_1 + n_2 + n_3$. The cumulative probability function $F_X(\cdot)$ can be estimated by using the empirical distribution $\hat{F}_X(\cdot)$. Substituting the estimates $\hat{\theta}_{\text{MW-}*}(x)$, $\hat{f}_X(x)$, $\hat{F}_X(a)$ and $\hat{F}_X(b)$ into equation (12) or (13), and using the trapezoidal rule leads to the estimators $\widehat{\text{AVUS}}_{\text{MW-}*}$ or $\widehat{\text{tAVUS}}_{\text{MW-}*}$.

3. Asymptotic properties

In this section, we establish the asymptotic properties of the covariate-specific MW-GEE and MW-LL estimators, at $X = x$. First, we show that the covariate-specific MW VUS $\theta_{\text{MW}}(x)$ in (3), which contains the true conditional mean and variance functions, is unbiased, consistent and asymptotically normal, when these functions are assumed to be known. Second, we prove that the MW-GEE and MW-LL estimators $\hat{\theta}_{\text{MW-GEE}}(x)$ and $\hat{\theta}_{\text{MW-LL}}(x)$ in (10) are mean squared (MS) consistent under some regularity conditions. The theory easily extends to the covariate-specific VUS estimators (6) and (11), in presence of ties.

Recall that n is the total sample size, i.e., $n = n_1 + n_2 + n_3$. Let

$$\phi(Y_1, Y_2, Y_3; x) = \mathbf{I}(Y_1 < Y_2 < Y_3 | X = x).$$

It is straightforward to show that $\mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) \} = \theta(x)$. Let

$$\begin{aligned} \phi_{000} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) \} - \theta(x) = 0, & \phi_{100} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_1 \} - \theta(x) \\ \phi_{010} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_2 \} - \theta(x), & \phi_{001} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_3 \} - \theta(x) \\ \phi_{110} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_1, Y_2 \} - \theta(x), & \phi_{101} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_1, Y_3 \} - \theta(x) \\ \phi_{011} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_2, Y_3 \} - \theta(x), & \phi_{111} &= \mathbb{E} \{ \phi(Y_1, Y_2, Y_3; x) | Y_1, Y_2, Y_3 \} - \theta(x). \end{aligned}$$

Let λ_{ijk}^2 be $\text{Var}(\phi_{ijk})$ with $i, j, k \in \{0, 1\}$, so that $\lambda_{000}^2 = 0$ and $\lambda_{111}^2 = \theta(x)(1 - \theta(x))$. For given $X = x$, it is worth noting that the MW VUS $\theta_{\text{MW}}(x)$ in (2) is actually a three-sample U-statistic. The theory of U-statistics (Lehmann, 1998) allows us to prove the following theorem.

Theorem 1. *Assume that the conditional means $\mu_j(\cdot)$ and the conditional variances $\sigma_j^2(\cdot)$ in the location-scale regression models (1) are known. For a fixed x , we have $\mathbb{E}(\theta_{\text{MW}}(x)) = \theta(x)$ and $\text{Var}(\theta_{\text{MW}}(x)) = O\left(\frac{1}{n_1+n_2+n_3}\right)$. Furthermore, we have*

$$\sqrt{n} (\theta_{\text{MW}}(x) - \theta(x)) \xrightarrow{d} \mathcal{N} \left(0, \frac{\lambda_{100}^2}{n_1} + \frac{\lambda_{010}^2}{n_2} + \frac{\lambda_{001}^2}{n_3} \right) \quad (14)$$

when $n \rightarrow \infty$, provided that $n_j/n \rightarrow \rho_j$ with $0 < \rho_j < 1$ ($j = 1, 2, 3$), and λ_{100}^2 , λ_{010}^2 and λ_{001}^2 are strictly positive.

Next, we establish the MS consistency of the MW-GEE estimator $\hat{\theta}_{\text{MW-GEE}}(x)$ and of the MW-LL estimator $\hat{\theta}_{\text{MW-LL}}(x)$, for a given covariate value $X = x$. We start by stating consistency and asymptotic normality of the estimators $\hat{\beta}_j$ and $\hat{\alpha}_j$. Then, we show that the estimators of mean and variance functions (both for GEE and LL approaches) are uniformly consistent. All the proofs can be found in Section A, Supplementary Material.

Lemma 2. Suppose that parametric forms of $h_j(\beta_j; \cdot)$ and $g_j(\alpha_j; \cdot)$ are specified, with $j = 1, 2, 3$. Then, under the regularity conditions (P1)-(P4) given in Section A.2, Supplementary Material, the estimators $\hat{\beta}_j$ and $\hat{\alpha}_j$ are consistent and

$$\sqrt{n} \left(\begin{pmatrix} \hat{\beta}_j \\ \hat{\alpha}_j \end{pmatrix} - \begin{pmatrix} \beta_j \\ \alpha_j \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, \Omega_j),$$

for suitable matrices Ω_j . Furthermore, the GEE estimators of the mean $\hat{\mu}_{j,\text{GEE}}(x)$ and the variance $\hat{\sigma}_{j,\text{GEE}}^2(x)$ functions are uniformly consistent.

Lemma 3. Suppose that the regularity conditions (N1)-(N9) stated in Section A.4, Supplementary Material, hold. Then, the LL estimators of the mean $\hat{\mu}_{j,\text{LL}}(x)$ and the variance $\hat{\sigma}_{j,\text{LL}}^2(x)$ functions are uniformly consistent.

Theorem 4. Under the regularity conditions (P1)-(P5) for the GEE approach or (N1)-(N10) for the LL approach, for given $X = x$, we have

$$\mathbb{E} \left\{ \hat{\theta}_{\text{MW-}*}(x) - \theta(x) \right\}^2 \longrightarrow 0, \quad (15)$$

(where the star * stands for GEE or LL) as $n \rightarrow \infty$ and $n_j/n \rightarrow \rho_j$, with $0 < \rho_j < 1$ ($j = 1, 2, 3$).

The MS consistency of the covariate-adjusted VUS estimator $\widehat{\text{AVUS}}_{\text{MW-}*}$ or the truncated covariate-adjusted VUS estimator $\widehat{\text{tAVUS}}_{\text{MW-}*}$ can be proved by using the continuous-mapping theorem and the Slutsky theorem, together with the MS consistency of $\hat{\theta}_{\text{MW-}*}(x)$ and of the histogram estimator $\hat{f}_X(\cdot)$ (Freedman and Diaconis, 1981), and the uniform convergence of the empirical estimator $\hat{F}_X(\cdot)$.

Although simulation results in Section 4 (see Figures 2-7 and Table 1) empirically support the conjecture of asymptotic normality of the covariate-specific estimators $\hat{\theta}_{\text{MW-}*}(x)$ and of the covariate-adjusted VUS estimators $\widehat{\text{AVUS}}_{\text{MW-}*}$ (also $\widehat{\text{tAVUS}}_{\text{MW-}*}$), a rigorous proof is not available to us. For this reason, we do not provide a closed-form for variance estimation. We propose to use the bootstrap procedure (Efron and Tibshirani, 1993) in order to: (i) estimate the variance of the estimators $\hat{\theta}_{\text{MW-}*}(x)$, $\widehat{\text{AVUS}}_{\text{MW-}*}$ and $\widehat{\text{tAVUS}}_{\text{MW-}*}$; (ii) obtain approximate pointwise confidence intervals for $\theta(x)$, AVUS or tAVUS through the percentile method.

4. Simulation study

In this section, we present the results of some simulation experiments carried out to evaluate the performance of the proposed MW-GEE and MW-LL methods for estimating the covariate-specific VUS, when the forms of the mean and variance functions are correctly specified or misspecified. In particular, we examine the Monte Carlo mean square errors (MSEs) of the estimators, and the empirical coverages of pointwise confidence intervals (and the behaviour of variance estimators) obtained by the bootstrap procedure.

In the simulation study, we set the following sample sizes: $n_1 = n_2 = n_3 = 50, 100$ and 200 , and perform 1000 Monte Carlo replications for each experiment. As for the bootstrap procedure, we fix $B = 500$ bootstrap samples. We use the standard nonparametric bootstrap, and resample, at each Monte Carlo replication, B times from the generated data, in order to obtain samples of sizes n_1, n_2 and n_3 for each class of disease, respectively.

We consider the following three different scenarios.

- I. The covariate X is generated from a uniform distribution $\mathcal{U}(0, 1)$. The error terms $\varepsilon_1, \varepsilon_2$ and ε_3 are generated from standard normal distributions. The test results Y_1, Y_2 and Y_3 are simulated from the location-scale models (1), with the following mean and variance functions:

$$\begin{aligned} \mu_1(x) &= 1 - 0.5x + x^2, \\ \mu_2(x) &= 1.5 + 0.5x + 2x^3, \\ \mu_3(x) &= 2 + 3x, \\ \sigma_j^2(x) &= (1 - x + x^2)^2, \end{aligned}$$

$$j = 1, 2, 3.$$

II. The covariate X and the test results Y_1, Y_2 and Y_3 are simulated as in the scenario I. The error terms $\varepsilon_1, \varepsilon_2$ and ε_3 are generated from t -distributions with degree of freedoms 5, 3 and 3, respectively, and then re-scaled to have unit variance.

III. The mean and variance functions are set as follows:

$$\begin{aligned}\mu_1(x) &= -0.3 + \sin(2\pi x), \\ \mu_2(x) &= 1.5 + \sin(2\pi x), \\ \mu_3(x) &= 2 + \sin(1.5x), \\ \sigma_j^2(x) &= (0.5 + 1.2x)^2,\end{aligned}$$

$j = 1, 2, 3$. The covariate X is uniformly distributed on $(0.5, 1.5)$, and the error terms $\varepsilon_1, \varepsilon_2$ and ε_3 have standard normal distribution.

In the first two scenarios, we apply the MW-GEE approach with correctly specified mean and variance functions, whereas the linear form $\mu_j(x) = \beta_{j0} + \beta_{j1}x$ for the mean function and constant variances $\sigma_j^2(x) = \sigma_j^2$ are implemented for the MW-GEE estimator in the last scenario. As for the MW-LL approach, we resort to the Epanechnikov kernel $K(u) = 0.75(1 - u)^2I(|u| < 1)$. The optimal bandwidths are obtained by using leave-one-out cross-validation method, given by R (R Core Team, 2018) package `locpol` (Ojeda Cabrera, 2018).

MSE of $\hat{\theta}_{\text{MW-GEE}}(x)$ and of $\hat{\theta}_{\text{MW-LL}}(x)$, at different values x for the covariate, are presented in Figure 1. In the first scenario, where we have correct model specifications and normal errors, the MW-GEE VUS estimator $\hat{\theta}_{\text{MW-GEE}}(x)$ appears to outperform the MW-LL estimator $\hat{\theta}_{\text{MW-LL}}(x)$. When the errors have heavy-tailed (re-scaled) t -distributions (Scenario II), $\hat{\theta}_{\text{MW-GEE}}(x)$ and $\hat{\theta}_{\text{MW-LL}}(x)$ yield comparable results. In the third scenario, where the (parametric) working models are misspecified, as expected, the non-parametric estimator $\hat{\theta}_{\text{MW-LL}}(x)$ is more effective than the semi-parametric one, namely $\hat{\theta}_{\text{MW-GEE}}(x)$. Finally, when sample sizes increase, the MSE of both covariate-specific VUS estimators tends to zero, except for $\hat{\theta}_{\text{MW-GEE}}(x)$ in the misspecification case.

The curves representing the averages of the Monte Carlo estimates (MCM) and of the Monte Carlo endpoints of the 95% pointwise bootstrap confidence intervals for $\theta(x)$ are showed in Figures 2-4. For the sake of comparison, we also plot the true covariate-specific VUS, and the 95% pointwise normal-based confidence intervals for $\theta(x)$, i.e., the averages of the Monte Carlo endpoints of the intervals obtained assuming the asymptotic normality of the estimators and using the bootstrap variances.

From Figures 2-4 we can see that the MCM curves for $\hat{\theta}_{\text{MW-LL}}(x)$ are close to the true ones in all scenarios, and the accuracy improves as the sample sizes increase. The MCM curves for $\hat{\theta}_{\text{MW-GEE}}(x)$ are close to the true ones in the two first scenarios, where mean and variance functions are correctly specified, and totally different to the true one under misspecification. We can see that confidence intervals obtained by the percentile method are (averagely) close to the ones built by using normal approximation. The Monte Carlo coverage probabilities (CP) for such intervals are showed in Figures 5-7. The good performances of the bootstrap estimator for the variance of $\hat{\theta}_{\text{MW-*}}(x)$, are documented in Figure S1, Section B of Supplementary Material, which shows the Monte Carlo standard deviations together with the averages of bootstrap standard deviations, in all scenarios with all considered sample sizes.

Finally, Table 1 presents the simulation results for the covariate-adjusted VUS estimators, for all scenarios I-III. Again, for each scenario and three different sample sizes, the table gives Monte Carlo mean (MCM), mean square error (MSE) multiplied by 100, Monte Carlo standard deviations (MCSD) and averages of estimated standard deviations (ASD) based on bootstrap procedure, for the covariate-adjusted VUS estimators. The table also provides the empirical coverages (CP) of the 95% confidence intervals for the covariate-adjusted VUS, obtained both, by using the bootstrap percentile method and assuming the validity of the normal approximation. As showed in the table, the estimator $\widehat{\text{AVUS}}_{\text{MW-LL}}$ performs well in terms of MSE and CP in all considered scenarios. The estimator $\widehat{\text{AVUS}}_{\text{MW-GEE}}$ performs well only when the working models for mean and variance functions are properly specified. In general, the large sample sizes helps to improve the accuracy of covariate-adjusted estimates, of course, for the $\widehat{\text{AVUS}}_{\text{MW-GEE}}$ we need to specify the mean and variance functions. We can also see the good performance of the bootstrap variance estimation, in all scenarios.

Table 1

Simulation results for the covariate-adjusted VUS estimators: Monte Carlo mean (MCM), mean square error (MSE) multiplied by 100, Monte Carlo standard deviations (MCSD), average of bootstrap standard deviations (ASD), and empirical coverages (CP) of the 95% confidence intervals obtained by using the percentile method (CP.Perc) and by using normal approximation (CP.Norm).

		MCM	MSE×100	MCSD	ASD	CP.Perc	CP.Norm
<i>Scenario I, True AVUS = 0.691</i>							
<i>n</i> = 50	MW-GEE	0.682	0.219	0.046	0.046	0.954	0.947
	MW-LL	0.694	0.222	0.047	0.048	0.949	0.951
<i>n</i> = 100	MW-GEE	0.683	0.109	0.032	0.033	0.956	0.949
	MW-LL	0.690	0.105	0.032	0.033	0.955	0.957
<i>n</i> = 200	MW-GEE	0.684	0.058	0.023	0.023	0.943	0.942
	MW-LL	0.689	0.054	0.023	0.023	0.950	0.948
<i>Scenario II, True AVUS = 0.765</i>							
<i>n</i> = 50	MW-GEE	0.742	0.338	0.053	0.051	0.919	0.928
	MW-LL	0.760	0.269	0.052	0.050	0.944	0.936
<i>n</i> = 100	MW-GEE	0.751	0.137	0.034	0.035	0.935	0.950
	MW-LL	0.762	0.115	0.034	0.034	0.956	0.949
<i>n</i> = 200	MW-GEE	0.755	0.074	0.025	0.025	0.930	0.939
	MW-LL	0.761	0.063	0.025	0.024	0.948	0.946
<i>Scenario III, True AVUS = 0.531</i>							
<i>n</i> = 50	MW-GEE	0.496	0.374	0.051	0.052	0.899	0.894
	MW-LL	0.534	0.270	0.052	0.053	0.954	0.958
<i>n</i> = 100	MW-GEE	0.499	0.226	0.035	0.036	0.850	0.863
	MW-LL	0.532	0.126	0.035	0.036	0.958	0.951
<i>n</i> = 200	MW-GEE	0.500	0.161	0.026	0.026	0.752	0.760
	MW-LL	0.531	0.064	0.025	0.025	0.946	0.947

5. Data application

In this section, we apply our proposed methods to the study of the effect of age on the ability of some cerebrospinal fluid (CSF) biomarkers to distinguish the stages of Alzheimer's disease. The data were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI, 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 (MRI), 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 Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

We consider 1215 subjects who registered in three ADNI projects: ADNI 1, ADNI 2 and ADNI GO. The true disease status of subjects were diagnosed by the means of neuropsychological tests, and consist of three categories: cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD). Among 1215 subjects, we have 369 CN, 617 MCI and 229 AD. We denote the first, the second and the third class as CN, MCI and AD, respectively. Among the CSF biomarkers measured in the three ADNI projects, we consider the three following biomarkers: $A\beta 1-42$ (amyloid- $\beta 1-42$), Tau (total tau) protein and pTau (phosphorylated tau) protein, as measured at baseline. For convenience, in our analysis we consider the log-transformation of biomarkers. Moreover, as the values of $A\beta 1-42$ are negatively related to the disease status, we take the minus of log-values of $A\beta 1-42$, so as to respect the ordering assumption.

Table 2

Estimated Age-adjusted VUS for $A\beta 1-42$, Tau and pTau, after log-transformation. The terms ‘‘Perc.CI’’ and ‘‘Normal.CI’’ stand for percentile confidence interval and Normal-approach confidence interval, respectively. Age from 55 to 85.

	Estimate	Standard error	95% Perc. CI	95% Normal.CI
$A\beta 1-42$				
MW-GEE	0.432	0.022	(0.377, 0.465)	(0.388, 0.475)
MW-LL	0.443	0.023	(0.390, 0.481)	(0.399, 0.488)
Tau-protein				
MW-GEE	0.359	0.020	(0.311, 0.391)	(0.319, 0.399)
MW-LL	0.360	0.020	(0.315, 0.395)	(0.320, 0.400)
pTau-protein				
MW-GEE	0.370	0.021	(0.326, 0.407)	(0.329, 0.410)
MW-LL	0.373	0.020	(0.329, 0.409)	(0.333, 0.413)

In our analysis, we estimate Age-specific VUS as well as Age-adjusted VUS for the log-transformed $A\beta 1-42$, Tau and pTau. For the GEE approach, we consider the following forms for the mean and variance functions:

$$\mu_j(\text{Age}) = \beta_{j0} + \beta_{j1} \text{Age}, \quad \sigma_j^2(\text{Age}) = (\gamma_{j0} + \gamma_{j1} \text{Age})^2,$$

with $j = 1, 2, 3$. For the LL approach, the bandwidths used to estimate mean and variance functions are obtained by the leave-one-out cross validation method. In order to construct the 95% confidence intervals for the Age-specific VUS, as well as to estimate the standard errors and get the 95% confidence intervals for the Age-adjusted VUS, we apply the bootstrap procedure, with 1,000 bootstrap replications. The results for the Age-specific VUS estimates and corresponding the 95% confidence intervals for the log-transformed $A\beta 1-42$, Tau and pTau are plotted in Figure 8. In addition, we report the GEE estimates of coefficients of the mean functions in Table S1 in Section C, Supplementary Material. The results show a significant linear relationship between the transformed biomarkers and the Age. We also present the GEE and LL estimates of the mean functions, for each biomarker and each class of disease, in Figure S2 (Section C, Supplementary Material). The shapes of the LL estimated functions again point towards linearity of the relation between the log-transformed biomarker and the Age in the three classes.

As showed in Figure 8, the estimated Age-specific VUS for all biomarkers appear to decrease with age, almost linearly. The results indicate a quite good diagnostic ability of the CSF biomarkers $A\beta 1-42$, Tau and pTau when applied for patients from 55 to 85 years old. For the older patients, the Age-specific VUS estimates are low, indicating that the biomarkers have not enough ability to distinguish between AD and the other two categories (CN and MCI). Overall, $A\beta 1-42$ seems to show the best behaviour.

Table 2 presents the estimated Age-adjusted VUS for the three CFS biomarkers. The results are quite different with respect to the unadjusted VUS estimates for $A\beta 1-42$ (0.409, standard error 0.018, 95% confidence interval (0.374, 0.445)), for Tau (0.335, standard error 0.017, 95% confidence interval (0.302, 0.369)) and for pTau (0.343, standard error 0.017, 95% confidence interval (0.310, 0.377)). Overall, again, they indicate that the diagnostic accuracy of $A\beta 1-42$ is the highest.

6. Conclusions

We present a new method to adjust for covariate effects in the estimation of VUS. Our method is based on the induced-regression methodology, which uses location-scale regression models to explain the relation between the test results and the covariate(s). For the estimation of the models, we propose to use a semiparametric GEE approach if the parametric forms of the mean and variance functions are specified. Alternatively, we propose a nonparametric method based on local linear regression. In order to estimate the covariate-specific VUS, we use a covariate-specific MW representation of VUS, and working samples constructed after fitting the location-scale models by the GEE or LL approach. This leads to new MW-GEE and MW-LL estimators. The asymptotic behaviour of the new covariate-specific estimators has been investigated. As expected, our simulation study shows that MW-GEE estimator is generally more accurate (with respect the MW-LL one) when the parametric models for the mean and variance functions are

correctly specified. However, the MW-LL estimator is robust with respect model misspecifications, while the MW-GEE estimator is even not consistent.

In the paper, we just consider the case of only one covariate. Nevertheless, the proposed method can be naturally extended to the case of multiple covariates. Such extension is simple for the semiparametric approach. In fact, for the MW-GEE estimator, we have only to specify suitable parametric models for the mean and variance functions, which accomodate for multidimensional covariates. For the MW-LL estimator, the extension can be more challenging, as multivariate local linear regression estimators are influenced by the curse of dimensionality, and their statistical accuracy generally decreases when the dimension of X increases, and the estimation process requires extensive computation. It is also worth mentioning that in the multidimensional case, from a practical point of view, a standardization of covariates is often greatly beneficial. An alternative way to extend our proposed approach is to use (when possible) additive models (Fan and Gijbels, 1996; Härdle et al., 2012) for mean and variance functions. Clearly, this last topic deserves further investigation.

When approaching statistical evaluation of a diagnostic test or biomarker, one typically has an idea about the possible association between the test and the disease status, so that elicitation of a monotone ordering for the classes may not represent a major criticality. However, in our paper methods are designed under the assumption that the monotone ordering hypothesis holds for every value of the covariates. This assumption may not be satisfied in practice and its check might add, from a practical point of view, some complexity to the analysis. A rather pragmatic solution could be based on the selection of the values for the covariates that are considered most interesting in terms of the diagnostic task, and then resort to the use of some method of order identification (such as those discussed in Li et al., 2014) before applying our covariate-specific VUS estimators.

Our proposed methodology and theory can be straightforwardly extended to M -class problems where $M > 3$. Let Y_1, Y_2, \dots, Y_M be the test results for subjects with respect to M ordered categories of disease status, say $1, 2, \dots, M$, respectively. In such case, one can model the relationship between the covariate X and test results by the M location-scale regression models as (1). Then the estimation procedure can be made consequently. In fact, one can apply the GEE or LL approaches to estimate the mean and variance functions, and then construct the working samples $\hat{Y}_{ji,*}(x) = \hat{\mu}_{j,*}(x) + \hat{\sigma}_{j,*}(x)\hat{\varepsilon}_{ji,*}$, at the value $X = x$, with $j = 1, 2, \dots, M$ and $i = 1, \dots, n_j$. An estimator of the covariate-specific hypervolume under the ROC manifold (HUM) can be obtained by extending expression (10) as follows:

$$\widehat{HUM}_*(x) = \frac{1}{n_1 n_2 \dots n_M} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} \dots \sum_{k=1}^{n_M} \mathbf{I} \left(\hat{Y}_{1i,*}(x) < \hat{Y}_{2r,*}(x) < \dots < \hat{Y}_{Mk,*}(x) \right).$$

The asymptotic properties of $\widehat{HUM}_*(x)$ can be proved by resorting to the classic limit theorem for M -sample U -statistics (Lehmann, 1998) and Lemma 2 and 3. Clearly, the GEE approach needs to specify the functional forms for the M mean and variance functions. Moreover, a potential challenging issue could be the requirement of the correct setting for the order of the M classes.

Beside VUS estimation, the search for an optimal pair of thresholds, based on some selection criteria, is a relevant issue in practice. Recently, several methods have been developed to this aim (we can cite the generalized Youden index method Nakas et al. (2010); the closest to perfection and the max volume approaches Attwood et al. (2014); and others Hua and Tian (2020)). Like the VUS, the optimal thresholds for a diagnostic test may be influenced by the covariates. Thus, the evaluation of the covariates effect on the optimal thresholds is of great interest, and, unfortunately, no results are available on this topic. As an optimal thresholds selection procedure requires the estimation of the three true class fractions, which actually can be estimated by using the working samples under the location-scale regression models, a possible solution to the problem could be based on the method proposed in this paper. Theoretical development for a new covariate-specific optimal thresholds selection method will be the focus of future research.

Finally, as suggested by a Reviewer, another interesting issue is the estimation of covariate-specific weighted VUS (see Huang and Li, 2019 when covariates are not present). We hope to deal with this topic in the near future.

Supplementary data

Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this 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

Acknowledgments

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

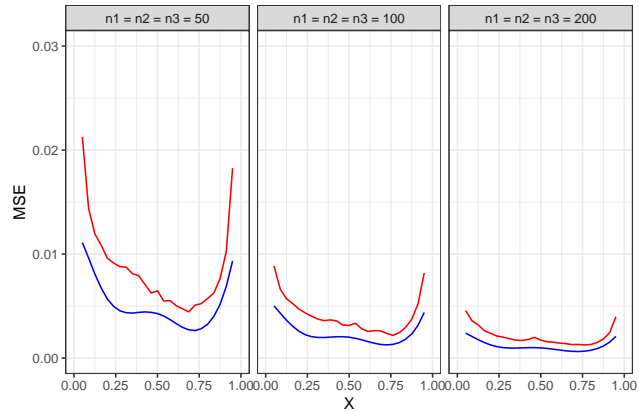
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Ministero dell'Istruzione, dell'Università e della Ricerca-Italy (grant number DIFO_ECCELLENZA18_01).

References

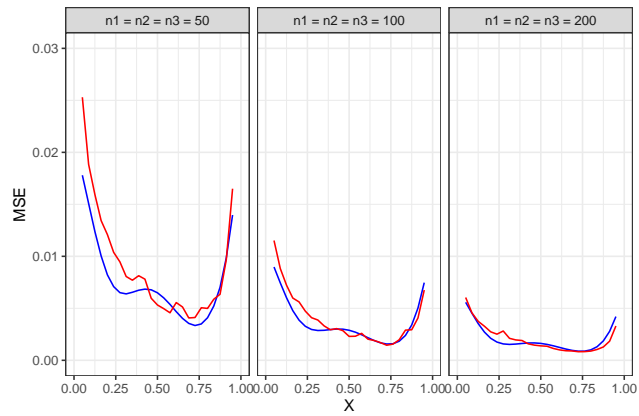
- Alonzo, T.A., Pepe, M.S., 2002. Distribution-free ROC analysis using binary regression techniques. *Biostatistics* 3, 421–432.
- Attwood, K., Tian, L., Xiong, C., 2014. Diagnostic thresholds with three ordinal groups. *J. Biopharm. Stat* 24, 608–633.
- de Carvalho, V.I., Jara, A., Hanson, T.E., de Carvalho, M., et al., 2013. Bayesian nonparametric ROC regression modeling. *Bayesian. Anal* 8, 623–646.
- Chen, L.H., Cheng, M.Y., Peng, L., 2009. Conditional variance estimation in heteroscedastic regression models. *J. Stat. Plan. Infer* 139, 236–245.
- Dodd, L.E., Pepe, M.S., 2003. Semiparametric regression for the area under the receiver operating characteristic curve. *J. Am. Stat. Assoc* 98, 409–417.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. CRC press.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.
- Faraggi, D., 2003. Adjusting receiver operating characteristic curves and related indices for covariates. *J. Roy. Stat. Soc. D-Sta* 52, 179–192.
- Freedman, D., Diaconis, P., 1981. On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57, 453–476.
- González-Manteiga, W., Pardo-Fernández, J.C., Keilegom, I.V., 2011. ROC curves in non-parametric location-scale regression models. *Scand. J. Stat* 38, 169–184.
- Härdle, W.K., Müller, M., Sperlich, S., Werwatz, A., 2012. *Nonparametric and semiparametric models*. Springer Science & Business Media.
- Hua, J., Tian, L., 2020. A comprehensive and comparative review of optimal cut-points selection methods for diseases with multiple ordinal stages. *J. Biopharm. Stat* 30, 46–68.
- Huang, L., Li, J., 2019. Weighted volume under the three-way receiver operating characteristic surface. *Stat. Methods Med. Res.* 28, 3627–3648.
- Lehmann, E.L., 1998. *Elements of Large-Sample Theory*. Springer Science & Business Media.
- Li, J., Chow, Y., Wong, W.K., Wong, T.Y., 2014. Sorting multiple classes in multi-dimensional ROC analysis: parametric and nonparametric approaches. *Biomarkers* 19, 1–8.
- Li, J., Zhou, X., Fine, J.P., 2012. A regression approach to ROC surface, with applications to Alzheimer's disease. *Sci. China Math.* 55, 1583–1595.
- Liu, D., Zhou, X.H., 2011. Semiparametric estimation of the covariate-specific ROC curve in presence of ignorable verification bias. *Biometrics* 67, 906–916.
- Nakas, C.T., Alonzo, T.A., Yiannoutsos, C.T., 2010. Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Stat. Med* 29, 2946–2955.
- Nakas, C.T., Yiannoutsos, C.T., 2004. Ordered multiple-class ROC analysis with continuous measurements. *Stat. Med* 23, 3437–3449.
- Ojeda Cabrera, J.L., 2018. **locpol**: Kernel local polynomial regression. R package version 0.7-0 URL: <https://CRAN.R-project.org/package=locpol>.
- Pardo Fernández, J.C., Rodríguez Álvarez, M.X., Keilegom, I.V., 2014. A review on ROC curves in the presence of covariates. *Revstat-Stat. J* 12, 21–41.
- Pepe, M.S., 2003. *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press.

- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Rodríguez-Álvarez, M.X., Roca-Pardiñas, J., Cadarso-Suárez, C., 2011. A new flexible direct ROC regression model: Application to the detection of cardiovascular risk factors by anthropometric measures. *Comput. Stat. Data. An* 55, 3257–3270.
- López-de Ullibarri, I., Cao, R., Cadarso-Suárez, C., Lado, M.J., 2008. Nonparametric estimation of conditional ROC curves: Application to discrimination tasks in computerized detection of early breast cancer. *Comput. Stat. Data. An* 52, 2623–2631.
- Wasserman, L., 2006. *All of nonparametric statistics*. Springer Science & Business Media.
- Yao, F., Craiu, R.V., Reiser, B., 2010. Nonparametric covariate adjustment for receiver operating characteristic curves. *Can. J. Stat* 38, 27–46.
- Zhou, X.H., McClish, D.K., Obuchowski, N.A., 2011. *Statistical methods in diagnostic medicine*. John Wiley & Sons.

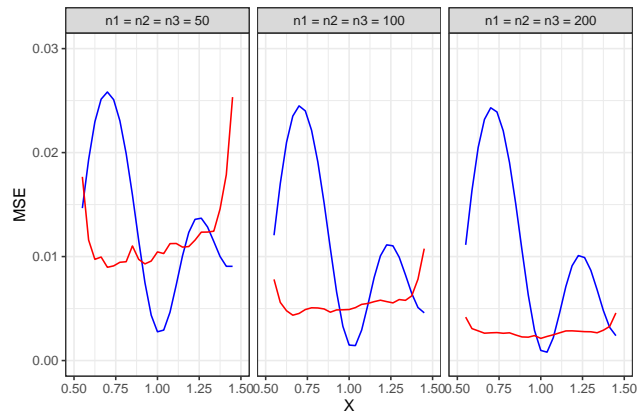
Covariate-specific VUS



(a) Scenario I



(b) Scenario II



(c) Scenario III

Figure 1: Mean square error (MSE) of the MW-GEE (blue solid line) and MW-LL (red solid line) covariate-specific VUS estimators in scenario I, II and III.

Covariate-specific VUS

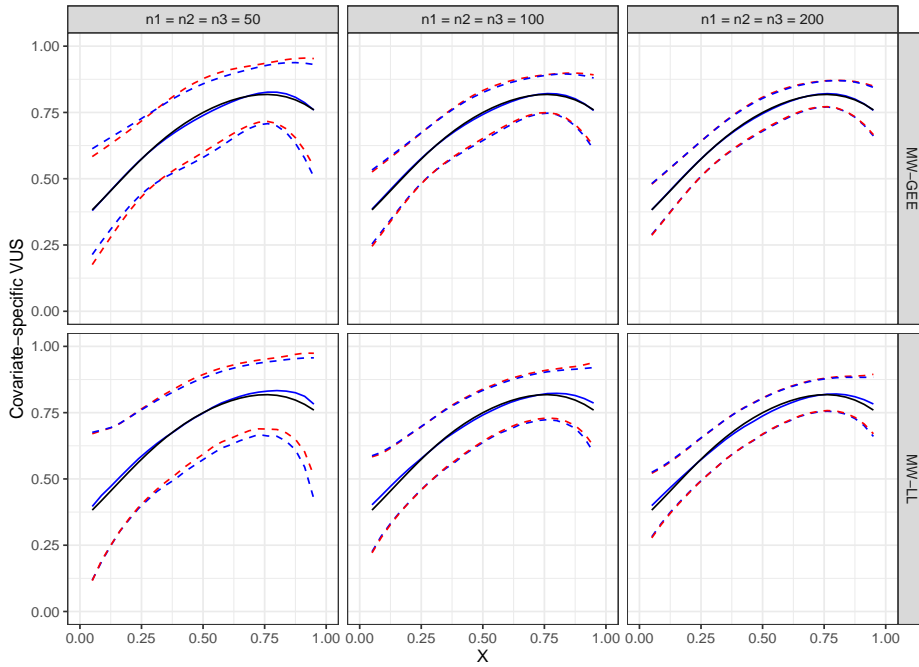


Figure 2: Monte Carlo mean (MCM) curves for the covariate-specific VUS estimators (blue solid line); average of the 95% pointwise bootstrap confidence intervals (blue dashed lines); average of the 95% pointwise normal-based confidence intervals (red dashed lines); true covariate-specific VUS (black solid line). Scenario I.

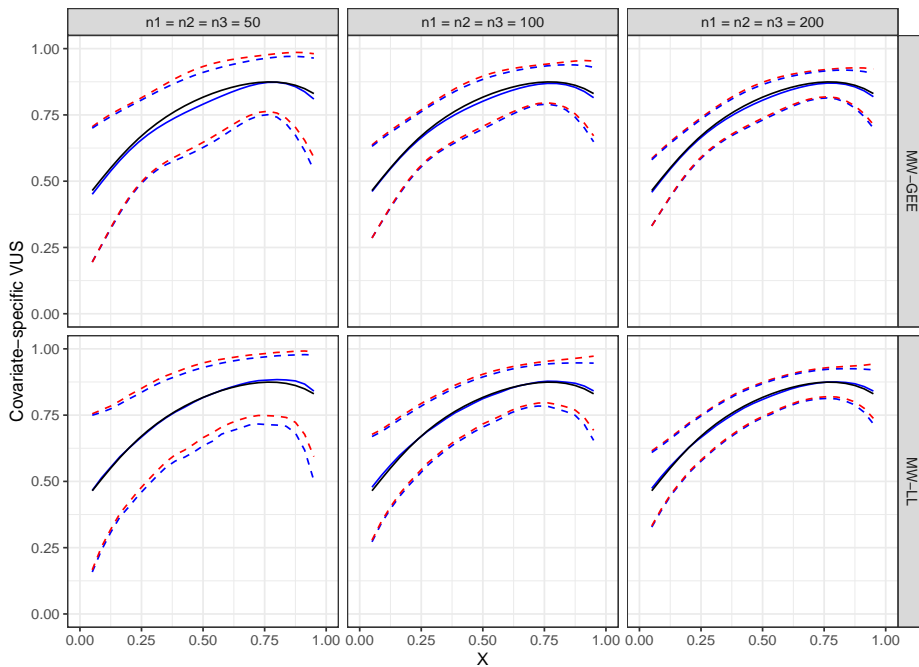


Figure 3: Monte Carlo mean (MCM) curves for the covariate-specific VUS estimators (blue solid line); average of the 95% pointwise bootstrap confidence intervals (blue dashed lines); average of the 95% pointwise normal-based confidence intervals (red dashed lines); true covariate-specific VUS (black solid line). Scenario II.

Covariate-specific VUS

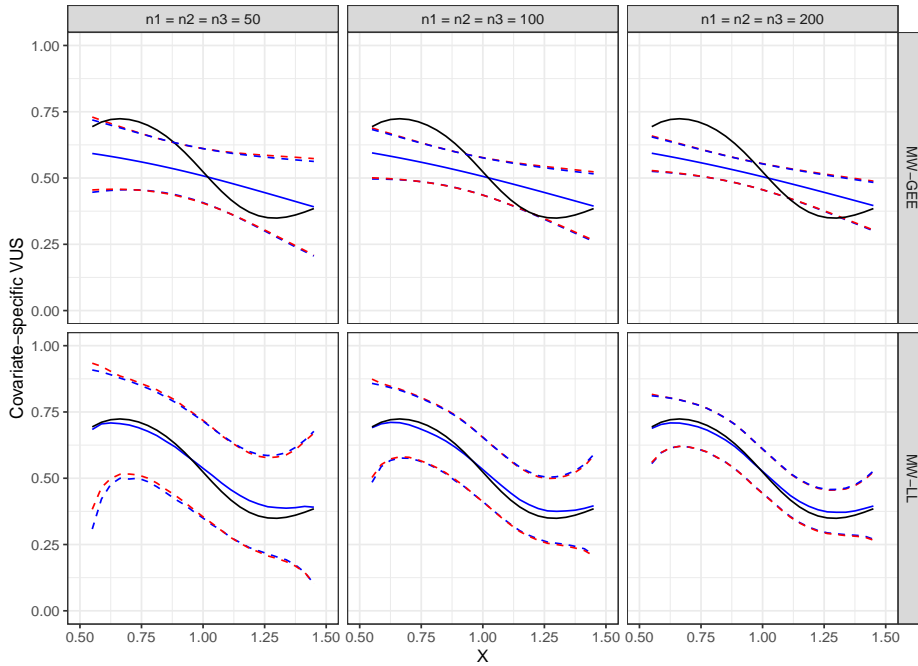


Figure 4: Monte Carlo mean (MCM) curves for the covariate-specific VUS estimators (blue solid line); average of the 95% pointwise bootstrap confidence intervals (blue dashed lines); average of the 95% pointwise normal-based confidence intervals (red dashed lines); true covariate-specific VUS (black solid line). Scenario III.

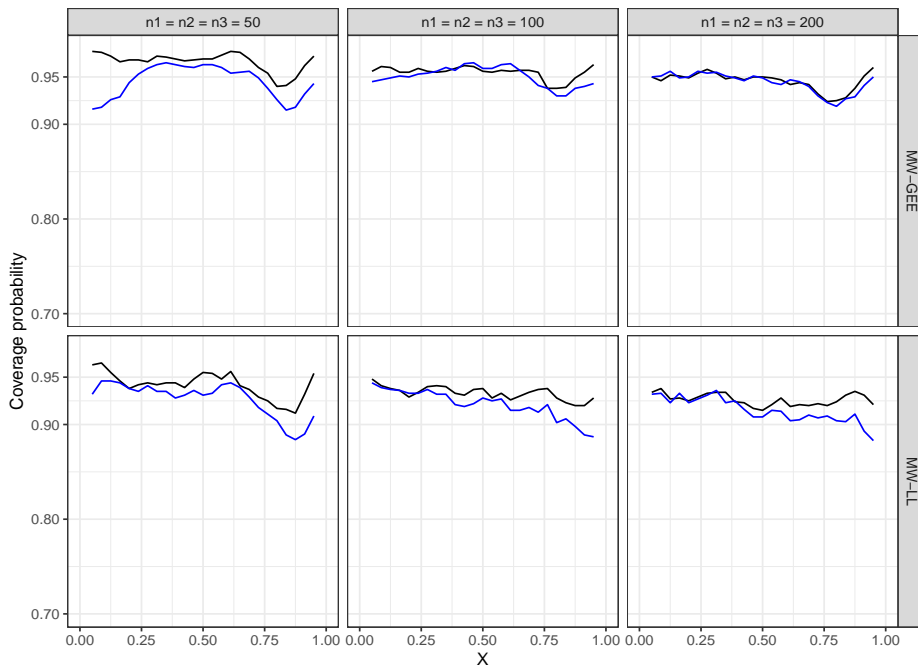


Figure 5: Empirical coverage of the pointwise bootstrap confidence intervals (black dashed lines) and of the pointwise normal-based confidence intervals (blue dashed lines), in scenario I. Nominal level 0.95.

Covariate-specific VUS

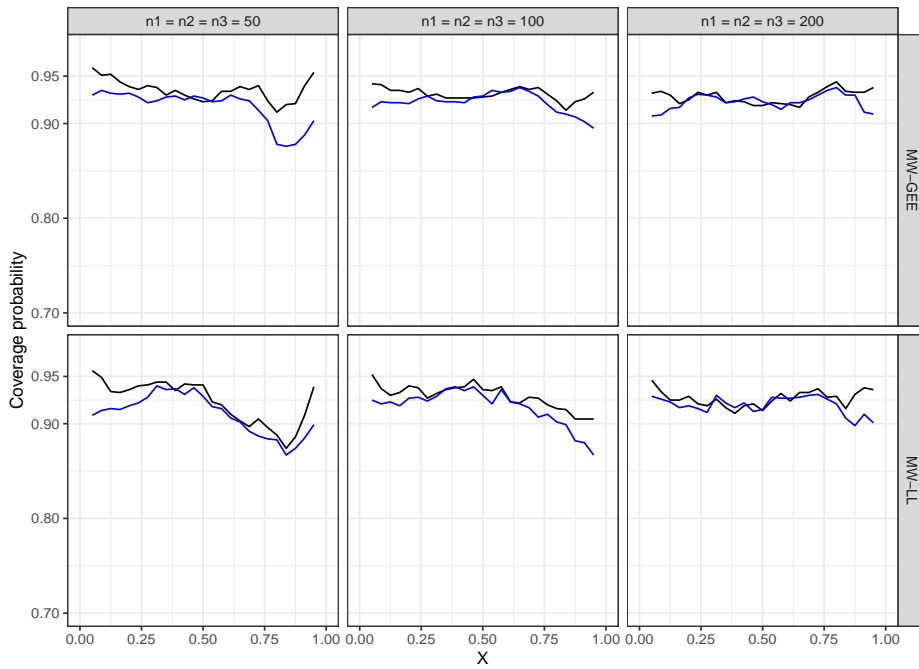


Figure 6: Empirical coverage of the pointwise bootstrap confidence intervals (black dashed lines) and of the pointwise normal-based confidence intervals (blue dashed lines), in scenario II. Nominal level 0.95.

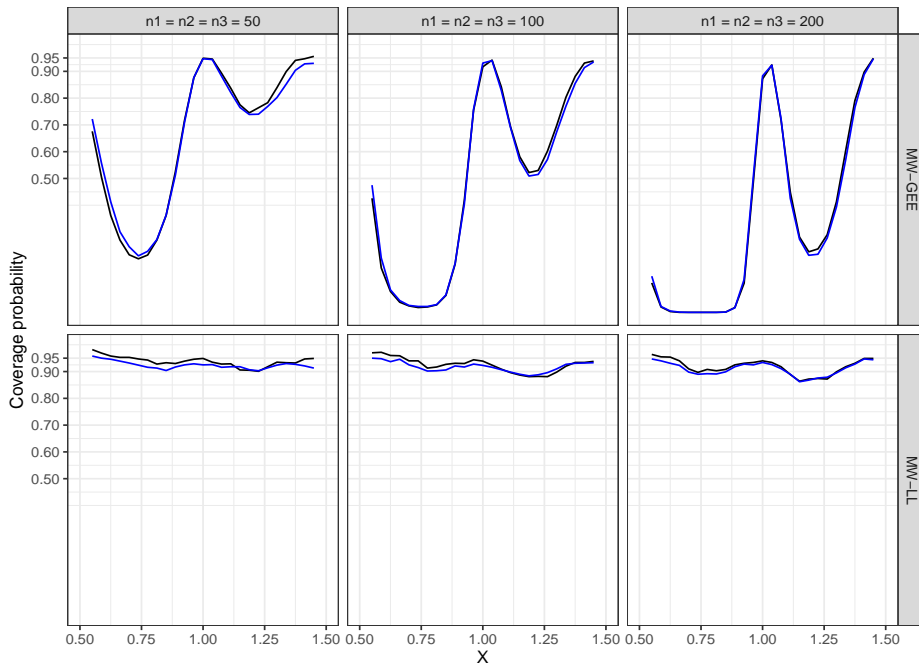


Figure 7: Empirical coverage of the pointwise bootstrap confidence intervals (black dashed lines) and of the pointwise normal-based confidence intervals (blue dashed lines), in scenario III. Nominal level 0.95.

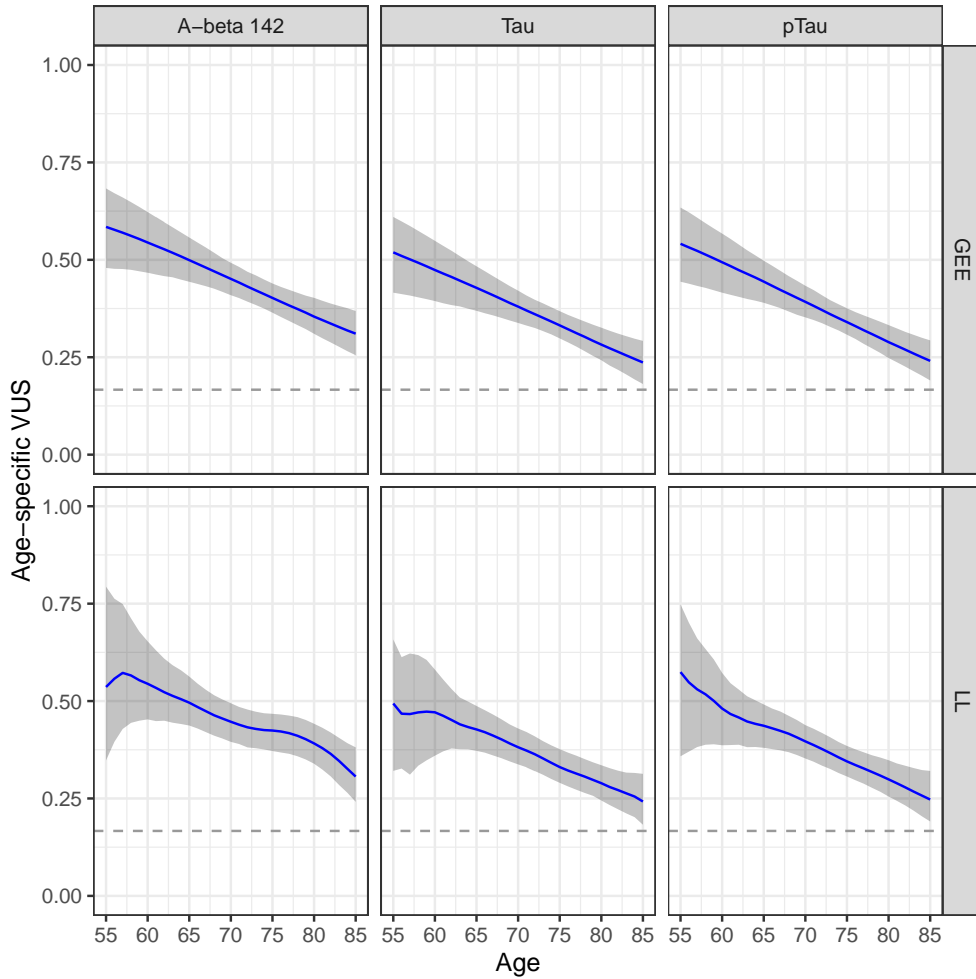


Figure 8: Age-specific VUS estimates with 95% pointwise confidence intervals (percentile) for the CSF biomarkers, after log-transformation. Age from 55 to 85.