Inference for Semiparametric AUC Regression Models with Discrete Covariates

Lin Zhang[1], Yan D. Zhao[2]* and Jack D. Tubbs[3]

[1]*Quintiles Inc.,* [2]*University of Texas Southwestern Medical Center and*
[3]*Baylor University*

Abstract: In this paper we consider clinical trials with two treatments and a non-normally distributed response variable. In addition, we focus on applications which include only discrete covariates and their interactions. For such applications, the semi-parametric Area Under the ROC Curve (AUC) regression model proposed by Dodd and Pepe (2003) can be used. However, because a logistic regression procedure is used to obtain parameter estimates and a bootstrapping method is needed for computing parameter standard errors, their method may be cumbersome to implement. In this paper we propose to use a set of AUC estimates to obtain parameter estimates and combine DeLong's method and the delta method for computing parameter standard errors. Our new method avoids heavy computation associated with the Dodd and Pepe's method and hence is easy to implement. We conduct simulation studies to show that the two methods yield similar results. Finally, we illustrate our new method using data from urinary incontinence clinical trials.

Key words: AUC, discrete covariates, interaction, NAOV, ROC curve, semi-parametric regression.

1. Introduction

The Wilcoxon-Mann-Whitney test is a widely used nonparametric method for comparing two treatments in clinical trials. In the presence of a discrete confounding stratum effect, the van Elteren (vE) test (van Elteren, 1960) is used to adjust for the stratum effect. However, the vE test does not handle the interaction between treatment and the stratum effect. In this case, Dodd and Pepe (2003) proposed an area under the curve (AUC) regression model which can test the interaction. Their method can also be applied to models with both discrete and continuous covariates.

---

*Corresponding author.

The AUC regression model utilizes concepts relevant to the Receiver Operating Characteristic (ROC) curve. The ROC curve is a widely used statistical tool for assessing the performance of a binary classifier with continuous or ordinal variables. Its use has gained increased attention in various biostatistics areas such as evaluating diagnostic tests, finding potential biomarkers, or analyzing controlled clinical trials. The use of this statistical method has been extensively developed since 1990's. Pepe (2003) and Zhou *et al.* (2002) provide excellent review on the ROC and its use.

A useful application of ROC curve is in diagnostic testing when one has a continuous variable Y, which will be used to classify the subjects into either diseased (D) or non-diseased ($\bar{D}$) groups according to some classification rule: $Y > c$ for the threshold c. The $ROC(c)$ curve is the function given by plotting $P(Y > c \mid D)$ vs $P(Y > c \mid \bar{D})$ in a square from vertices $(0,0)$ to $(1,1)$ for all possible thresholds, which displays how the true positive rate (TPR) is changed as false positive rate (FPR) and provides a visible inspection of the accuracy of the diagnostic test to make decision about the optimal threshold for the output relying on the requirement for the relative importance of sensitivity and specificity in the application (Dodd and Pepe, 2003). In the two most extreme situations, if the distribution of Y in the diseased group is exactly overlapped with the distribution of Y in the non-diseased group, the ROC curve will be a diagonal line from vertices $(0,0)$ to $(1,1)$, which means that the test with output Y is same for FPR and TPR for all thresholds and is useless in classification; if the distributions of Y in the two groups are totally separated from each other, the ROC curve will be a curve passing through the vertex $(0,1)$, which means that the test can easily select the threshold to guarantee the best classification in sensitivity and specificity.

The commonly used summary index for ROC curve is the area under the ROC curve (AUC). It can be shown that $AUC = P(Y_D > Y_{\bar{D}})$. Therefore, in the worst case, the AUC is 0.5, that is, the ability to classify a subject into a right group for this test is no more than by chance. While in the perfect case, the AUC equals to 1, which represents that the probability of correct classification for a proper threshold is 1.

The estimated AUC can be derived from the Mann-Whitney U statistic for testing the equality of two distributions. Based on this property Dodd and Pepe (2003) proposed an AUC regression model for data with a non-normally distributed response variable which can adjust for continuous and discrete covariates. In the model, the response variable is a cross-correlated bernoulli variable. Because the usual standard errors derived from a logistic regression model are incorrect, they proposed to use bootstrapping method to estimate the variance of non-parametric AUC and the model parameters. Because a logistic regression

procedure is used to obtain parameter estimates and a bootstrapping method is needed for computing parameter standard errors, their method may be cumbersome to implement.

In practice, there are many applications which involves only discrete covariates and their interactions. Such applications include testing the interaction between treatment and subgroup variable in typical subgroup analysis in clinical trials. For such applications, we aim to alleviate the computation burden associated with Dodd and Pepe. We propose to use a set of AUC estimates to obtain parameter estimates and combine Delong's method and the delta method for computing parameter standard errors.

The remainder of this paper is organized as follows. The AUC regression model by Dodd and Pepe is introduced in Section 2. DeLong's method (DeLong, DeLong and Clarke, 1998) for standard error estimation of the unadjusted AUC is described in Section 3. Our newly proposed method for computing parameter estimates and standard errors is developed in Section 4. Simulation studies comparing Dodd and Pepe's method and our method are presented in Section 5. A real data example is shown in Section 6 to illustrate our new method. Some discussions can be found in Section 7.

2. Semi-Parametric AUC Regression Model

In this section we review the semi-parametric AUC regression model proposed by Dodd and Pepe. Assume that one needs to adjust the AUC for a covariate X, the covariate-specific AUC can be expressed as

$$AUC_{ij} = P\big(Y_i^D > Y_j^{\bar{D}} \mid X_i, X_j\big),$$

where Y_i^D is the ith response in diseased (or treatment) group with covariate value X_i $(i = 1, \cdots, N_D)$ and $Y_j^{\bar{D}}$ is the jth response in non-diseased (or control) group with covariate value X_j $(j = 1, \cdots, N_{\bar{D}})$. Often one is interested in estimating the AUC at a specified covariate level, i.e. $P\big(Y_i^D > Y_j^{\bar{D}} \mid X_i = X_j = X\big)$.

Dodd and Pepe applied this model to the Generalized Linear Model (GLM) framework which allows one to model the AUC with covariates, in which case their model can be written as,

$$g(AUC_{ij}) = X_{ij}^T \beta, \tag{2.1}$$

where g is a monotone link function such as the probit or logit link, X_{ij} is a vector function of X_i and X_j, and β is a vector fixed and unknown parameters to be estimated.

Note that

$$E\big(I\big(Y_i^D > Y_j^{\bar{D}}\big) \mid X_{ij}\big) = AUC_{ij}.$$

Thus, for estimating the parameters in the model, Dodd and Pepe proposed the use of the logistic regression model where the response variable is a Bernoulli variable $I(Y_i^D > Y_j^{\bar{D}})$. Dodd and Pepe demonstrated that the estimates of parameters are found as solution to the usual score equations given by

$$\sum_i^{N_D} \sum_j^{N_{\bar{D}}} \frac{(I_{ij} - AUC_{ij})}{V(I_{ij})} \frac{\partial AUC_{ij}}{\partial \beta}, \tag{2.2}$$

where $I_{ij} = I(Y_i^D > Y_j^{\bar{D}})$. Therefore, one obtains this estimate using standard statistical software, such as SAS PROC GENMOD or PROC LOGISTIC. However, the usual standard errors of the estimates can not be used since the binary variables I_{ij} in equation (2.2) are not independent. Dodd and Pepe recommended the bootstrap for obtaining the needed standard errors. Their procedure can be summarized in the following steps in the presence of covariates:

1. Stratify the range of the covariate variable as S strata. If the covariate is discrete, each level of the covariate becomes a stratum. While for continuous variable, it is impossible to make each covariate value a stratum. Cluster the adjacent values into a stratum to ensure enough fitting data in each stratum.

2. For discrete covariate, within each stratum s ($s = 1, \cdots, S$), generate all of the 0 or 1 indicator data as $I(Y_{is}^D > Y_{js}^{\bar{D}})(i = 1, \cdots, n_D^s \; ; \; j = 1, \cdots, n_{\bar{D}}^s)$. In this case, the model is $g(AUC_{ij}) = \beta_0 + \beta_s$.

3. If there is a continuous covariate in addition to the discrete covariate, other parameters should be included in the model in order to fit the data obtained by comparing two responses from different covariate values, such as $I(Y_{is}^D > Y_{js}^{\bar{D}})$ (the ith and j^{th} outputs are from different covariate value but in the same stratum s). The model can be expressed as $g(AUC_{ij}) = \beta_0 + \beta_1 X_i + \beta_2(X_i - X_j)$.

4. Use the standard logistic regression procedure to fit the data with strata as covariate to obtain parameter estimates.

5. Bootstrap the original data within each stratum to compute the parameter standard errors.

The above procedure involves bootstrap so it is difficult to implement. For models with only discrete covariates and their interactions, we aim to simplify the above model fitting procedure. We propose a new algorithm which involves computing variance of a non-parametric AUC estimate which was first proposed by the

DeLong, DeLong and Clarke (1998). This algorithm is described in detail in the next section.

3. DeLong's Method for Computing the Variance of Unadjusted AUC

Several approaches have been proposed to compute the variance of the unadjusted AUC. See Hanley and Hajian-Tilakin (1997) for a review. Among them, the method provided by DeLong, DeLong, and Clarke (1988) is most widely used and has the plain analytical structure, which completely relies on the Mann-Whitney statistic.

Bamber (1975) provided the equivalence between Mann-Whitney two-sample rank sum statistic and the empirical estimate of AUC. When the outputs in disease group and non-diseased group have ties, the nonparametric AUC can be expressed as:

$$\widehat{AUC} = \frac{\sum_{i=1}^{N_D} \sum_{j=1}^{N_{\bar{D}}} I_{ij}}{N_D N_{\bar{D}}}, \tag{3.1}$$

where

$$I_{ij} = \begin{cases} 1, & Y_i^D > Y_j^{\bar{D}}, \\ \frac{1}{2}, & Y_i^D = Y_j^{\bar{D}}, \\ 0, & Y_i^D < Y_j^{\bar{D}}. \end{cases}$$

The variance for (3.1) by DeLong's method involves in two components which are defined as

$$V_i^D = \frac{1}{N_{\bar{D}}} \sum_{j=1}^{N_{\bar{D}}} I_{ij}, \quad i = 1, \cdots, N_D,$$

and

$$V_j^{\bar{D}} = \frac{1}{N_D} \sum_{i=1}^{N_D} I_{ij}, \quad j = 1, \cdots, N_{\bar{D}}.$$

V_i^D is the percentage of $Y^{\bar{D}}$'s that Y_i^D is bigger or equal to. It measures the relative rank of the ith output of diseased group in the non-diseased group (i.e. its relative percentile when Y_i^D is put into the non-diseased group). The explanation for $V_j^{\bar{D}}$ is similar.

An estimate of the variance of the nonparametric AUC is

$$\hat{V}(\widehat{AUC}) = \frac{s_D^2}{N_D} + \frac{s_{\bar{D}}^2}{N_{\bar{D}}}, \tag{3.2}$$

where s_D^2 and $s_{\bar{D}}^2$ are the sample variances of $\left\{V_i^D, i = 1, \cdots, N_D\right\}$ and $\left\{V_j^{\bar{D}}, j = 1, \cdots, N_{\bar{D}}\right\}$, respectively.

4. A New Algorithm for Estimating Parameters and Standard Errors for the AUC Regression Model

Our new algorithm is best described using an example. Let the covariate be X with 2 levels specified as 0 and 1. The logistic regression model with link function g can be expressed as

$$g(AUC|X) = \beta_0 + \beta_1 X.$$

When $X = 0$,

$$g(\widehat{AUC} \mid X = 0) = g(\widehat{AUC_0}) = \hat{\beta}_0,$$

and $X = 1$,

$$g(\widehat{AUC} \mid X = 1) = g(\widehat{AUC_1}) = \hat{\beta}_0 + \hat{\beta}_1.$$

Thus,

$$\hat{\beta}_1 = g(\widehat{AUC_1}) - g(\widehat{AUC_0}),$$

where $\widehat{AUC_0}$ and $\widehat{AUC_1}$ are computed using (3.1) and subseting observations with $X = 0$ and $X = 1$, respectively. We see that the parameter estimates are explicit functions of the AUC estimates at each stratum. Therefore, our new method avoids the logistic regression procedure necessary by Dodd and Pepe's method.

In the following we describe how to compute the standard errors for these parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. Because the observations from two strata are independent, the variance of $\hat{\beta}_0$ and $\hat{\beta}_1$ are, respectively,

$$V(\hat{\beta}_0) = V(g(\widehat{AUC_0})),$$

and

$$V(\hat{\beta}_1) = V(g(\widehat{AUC_1})) + V(g(\widehat{AUC_0})).$$

The above variance can be estimated by combining the delta method and (3.2). In what follows we provide the variance estimates (standard errors) of the parameter estimates for logit and probit links. When g is logit function,

$$\hat{V}(g(\widehat{AUC_i})) = \frac{\hat{V}(\widehat{AUC_i})}{\widehat{AUC_i}^2(1 - \widehat{AUC_i}^2)}.$$

Let $f(\cdot)$ and $\Phi(\cdot)$ be the PDF and the CDF of the standard normal distribution, respectively. When g is probit function,

$$\hat{V}(g(\widehat{AUC_i})) = \frac{\hat{V}(\widehat{AUC_i})}{f^2(\Phi^{-1}(\widehat{AUC_i}))}.$$

Therefore, consistent estimators of $V(\hat{\beta}_0)$ and $V(\hat{\beta}_1)$ are

$$\hat{V}(\hat{\beta}_0) = \hat{V}(g(\widehat{AUC}_0)),$$
$$\hat{V}(\hat{\beta}_1) = \hat{V}(g(\widehat{AUC}_0)) + \hat{V}(g(\widehat{AUC}_1)),$$

respectively. Finally, Wald-type $100(1-\alpha)\%$ confidence intervals for β_i, $i = 1, 2$, can be constructed as

$$\hat{\beta}_i \pm Z_{\alpha/2}\sqrt{\hat{V}(\hat{\beta}_i)},$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$th quantile of the standard normal distribution.

The above procedure can be readily generalized to models with more than one discrete covariates and their interactions. First, computing AUC estimates using (3.1) for each stratum resulting from all possible combinations of the covariates. Then, by equating the model parameters to these AUC estimates through the link function, we can solve for parameter estimates. Finally, by combining (3.2) and the delta method we obtain standard error estimates. In spirit, this approach is very similar to the analysis of variance model in the normal-theory linear models. Therefore, we term our method as nonparametric analysis of variance method (NAOV).

5. Simulation Study

We conduct simulation studies to compare the Dodd and Pepe method and our new NAOV method for estimating model parameters and their standard errors. In addition, we compare the two methods in terms of coverage probabilities of the 95% confidence intervals for each parameter. Data are generated from models with probit link and logit link, respectively. For each link function, we illustrate the model using a discrete covariate with 3 strata.

5.1 Probit Link

When the link function is probit, data is generated such that $Y_{ij}^{\bar{D}} \sim N(\mu_{1i}, \sigma_1^2)$, $Y_{ij}^D \sim N(\mu_{2i}, \sigma_2^2)$, $i = 1, \cdots, S; j = 1, \cdots, n$, where $\mu_{1i} = \delta_0 + \delta_{2i}$ and $\mu_{2i} = (\delta_0 + \delta_1) + (\delta_{2i} + \delta_{3i})$. The parameters of the model with probit link can be derived based on:

$$AUC_i = \Phi\left(\frac{\delta_1 + \delta_{3i}}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right). \tag{5.1}$$

When $i = 1$, let $\delta_{3i} = 0$, then

$$AUC_1 = \Phi(\beta_0)$$

and when $i = 2, \cdots, S$,
$$AUC_i = \Phi\left(\beta_0 + \beta_{i-1}\right),$$

where $\beta_0 = \frac{\delta_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ and $\beta_j = \frac{\delta_{3(j+1)}}{\sqrt{\sigma_1^2 + \sigma_2^2}} - \beta_0$ $(j = 1, \cdots, S - 1)$.

We choose $\delta_0 = 0$, $\delta_1 = 0.15$, $\delta_{2i} = 0$, $\delta_{32} = 0.5$, $\delta_{33} = 1$, $\sigma_1 = 1$ and $\sigma_2 = 1.2$ to compare the two methods. The simulation size is 1,000 and the number of bootstrap samples is 200. Table 1 gives the comparison results for $n = 30$ and $n = 100$. We see that both methods produce almost identical parameter estimates and very similar standard error estimates. In addition, the coverage probabilities of the 95% CIs are close to the nominal levels for both methods.

Table 1: Comparison of parameter estimates, standard errors, and 95% CIs for the model with probit link with 30 or 100 samples each group each level. Results represent 1000 realizations of the model and 200 bootstrap samples each

				Dodd and Pepe		NAOV	
n	Parameter	True Value	Estimate	Standard Error	Coverage 95% CI	Standard Error	Coverage 95% CI
30	β_0	0.15	0.16	0.20	0.965	0.19	0.958
	β_1	0.50	0.52	0.29	0.951	0.28	0.945
	β_2	1.00	1.03	0.32	0.970	0.30	0.951
100	β_0	0.15	0.15	0.10	0.952	0.10	0.950
	β_1	0.50	0.50	0.15	0.937	0.15	0.939
	β_2	1.00	1.00	0.16	0.954	0.16	0.954

5.2 Logit Link

When logit link is used, we generate data $Y_i^{\bar{D}} = -\log(u_1) + \delta_{1i}$ and $Y_i^{D} = -\log(u_2) + \delta_0 + (\delta_{1i} + \delta_{2i})$, where $u_1 \sim exponential(1)$ and $u_2 \sim exponential(1)$. The parameters in the model can be derived based on:

$$AUC_i = F(\delta_0 + \delta_{2i}), \tag{5.2}$$

where $F(\cdot)$ is a CDF of a logistic r.v. $\sim Lo(0, \pi^2/3)$ and $F(x) = \frac{1}{1+e^{-x}}$ (see reference from Balakrishnan and Nevzorov, 2003).

When $i = 1$, let $\delta_{2i} = 0$, then

$$AUC_1 = F(\beta_0)$$

and when $i = 2, \cdots, S$,
$$AUC_i = F\left(\beta_0 + \beta_{i-1}\right),$$
where $\beta_0 = \delta_0$ and $\beta_j = \delta_{2(j+1)}$ $(j = 1, \cdots, S - 1)$.

We choose $\delta_0 = 0.15$, $\delta_{1i} = 0$, $\delta_{22} = 0.5$ and $\delta_{23} = 1$. The number of bootstrap samples is 200. Table 2 gives the comparison results for $n = 30$ and $n = 100$ in each group in each level, respectively. We see that both methods produce almost identical parameter estimates and very similar standard error estimates. In addition, the coverage probabilities of the 95% CIs are close to the nominal levels for both methods.

Table 2: Comparison of parameter estimates, standard errors, and 95% CIs for the model with logit link with 30 or 100 samples each group each level. Results represent 1000 realizations of the model and 200 bootstrap samples each

n	Parameter	True Value	Estimate	Dodd and Pepe		NAOV	
				Standard Error	Coverage 95% CI	Standard Error	Coverage 95% CI
30	β_0	0.15	0.14	0.32	0.963	0.31	0.959
	β_1	0.50	0.53	0.46	0.958	0.45	0.947
	β_2	1.00	1.03	0.49	0.971	0.47	0.955
100	β_0	0.15	0.15	0.17	0.956	0.17	0.952
	β_1	0.50	0.51	0.24	0.952	0.24	0.950
	β_2	1.00	1.00	0.25	0.957	0.25	0.951

6. Real Data Example

In this section we illustrate our new NAOV method using real data from clinical trials. The purpose of the clinical trials is to investigate the efficacy of an active drug to treat stress urinary incontinence in women by comparing with placebo. The data are provided by a pharmaceutical company and not publicly available. The response variable is the per cent (relative) reduction in incontinence episode frequency (PIEF) from baseline to the last postbaseline visit. There is a variable reflecting the disease severity at the baseline (BIEF) with 4 strata. The variable BIEF takes values from 1 (mild) to 4 (severe). Also, the consistency of treatment effect across another covariate of interest HORM50 needs to be considered. The variable HORM50 indicates whether a woman has hormone replacement (yes or no). In summary, we are interested in analyzing the joint predictive and prognostic effects of BIEF and HORM50. For this analysis a total of 4940 subjects with nonmissing PIEF, BIEF, and HORM50 are included.

The distribution of PIEF variable is very skewed in each group. In a natural way, we can use AUC regression model to adjust the covariate effect to treatment effect. The overall model is logit(AUC(HORM50, BIEF)) $= \beta_0 + \beta_1 I(\text{HORM50} = 0) + \beta_2 I(\text{BIEF} = 1) + \beta_3 I(\text{BIEF} = 2) + \beta_4 I(\text{BIEF} = 3) + \beta_5 I(\text{BIEF} = 1 \ \& \ \text{HORM50} = 0) + \beta_6 I(\text{BIEF} = 2 \ \& \ \text{HORM50} = 0) + \beta_7 I(\text{BIEF} = 3 \ \& \ \text{HORM50} = 0).$

Table 3 shows the AUC estimates, 95% CI of the AUC, and the expression of AUC in terms of the model parameters by the combination levels of the two covariates. By equating the third column and the fifth column of Table 3, we can solve for estimates for the model parameters $\beta_0, \cdots, \beta_7$. These parameter estimates are presented in the third column of Table 4. The standard errors and 95% CIs for the model parameters are computed using the procedure described in Section 4 and displayed in the fourth and fifth columns of Table 4. By examining these CIs we see that the main effect of HORM50 and the interaction effects are not significant at .05 level, indicating there are no interactions between BIEF and HORM 50 and HORM50 is not predictive of treatment effect. The CI for β_3 does not include 0, indicating BIEF is predictive of treatment effect.

We fit another model with only BIEF main effect and the results are shown in Table 5. The CIs indicate that β_1 and β_2 are significantly different from 0 at .05 level, which means BIEF is a predictive of treatment effect. Note that β_3 is not significantly different from 0 at .05 level, possibly due to thet fact that only a small percentage of patients are in that very severe stratum. Parameters β_1 to β_3 have meaningful interpretations. For example, e^{β_1} is the odds ratio that the active drug is better than placebo in the second stratum of BIEF compared with that in the first stratum of BIEF. In this case, that odds ratio is estimated as $e^{\hat{\beta}_1} = 1.4$.

Table 3: Estimates of AUC for the example

BIEF	HORM50	AUC Estimate	95% CI by DeLong	AUC in terms of model parameters
1	0	.467	(.306, .629)	$1/(1 + \exp(-\beta_0 - \beta_4 - \beta_1 - \beta_5))$
	1	.531	(.393, .670)	$1/(1 + \exp(-\beta_0 - \beta_1))$
2	0	.583	(.510, .657)	$1/(1 + \exp(-\beta_0 - \beta_4 - \beta_2 - \beta_6))$
	1	.519	(.447, .592)	$1/(1 + \exp(-\beta_0 - \beta_2))$
3	0	.589	(.507, .671)	$1/(1 + \exp(-\beta_0 - \beta_4 - \beta_3 - \beta_7))$
	1	.713	(.628, .798)	$1/(1 + \exp(-\beta_0 - \beta_3))$
4	0	.606	(.534, .679)	$1/(1 + \exp(-\beta_0 - \beta_4))$
	1	.667	(.604, .731)	$1/(1 + \exp(-\beta_0))$

Table 4: Parameter estimates and 95% CI by NOAV with interactions for the example

Parameter	Level	Estimate	SE	95% CI
β_0	Intercept	-0.70	0.15	$(-0.99, -0.41)$
β_1	HORM50 = 0	0.27	0.21	$(-0.15, 0.68)$
β_2	BIEF = 1	0.57	0.32	$(-0.05, 1.20)$
β_3	BIEF = 2	0.62	0.21	$(0.21, 1.03)$
β_4	BIEF = 3	-0.21	0.26	$(-0.72, 0.29)$
β_5	BIEF = 1 and HORM50 = 0	-0.01	0.53	$(-1.04, 1.02)$
β_6	BIEF = 2 and HORM50 = 0	-0.52	0.37	$(-1.24, 0.19)$
β_7	BIEF = 3 and HORM50 = 0	0.29	0.40	$(-0.50, 1.08)$

Table 5: Parameter estimates and 95% CI by NOAV with only BIEF main effect for the example

Parameter	Level	Estimate	SE	95% CI
β_0	Intercept	-0.56	0.06	$(-0.67, -0.45)$
β_1	BIEF = 1	0.37	0.15	$(0.07, 0.67)$
β_2	BIEF = 2	0.19	0.08	$(0.03, 0.35)$
β_3	BIEF = 3	0.09	0.09	$(-0.08, 0.27)$

7. Discussion

In this article, we developed an analytical NAOV method to computing parameter estimates and standard errors for the semi-parametric AUC regression model with only discrete covariates. The NAOV method involves only straightforward computations and is much easier to implement than the Dodd and Pepe method. Simulation studies have shown that both methods yield similar results.

The NAOV method involves computing AUC estimates and standard errors at each cell of combination levels of all the covariates. Therefore, it requires a reasonable amount of observations at each cell. When there are empty cells, a saturated model can not be fitted and so some terms need to be dropped from the model. This is the same problem suffered by linear models and the Dodd and Pepe's method.

Note that in Tables 1 and 2, we set the true parameters values to be different than zero. However, even when the sample size is 100, some of the 95% confidence intervals of these parameters contain zero. This may yield a conclusion that the parameters are not significantly away from zero. We remark that this is a phenomenon we often encounter when the sample size is not sufficiently large. So if we increase the sample size, then these confidence intervals will become narrower and exclude zero.

Note that Tables 1 and 2 are based on different models, so it may not be appropriate to compare results between Table 1 and 2. In our experience, for the same model, using probit or logit links gives very similar results. This is also true from the literature for generalized linear models. In general, researchers tend to use logit links because it is easier to interpret the parameters.

Although Dodd and Pepe's method can be used for models with both discrete and continuous covariates, their method involves somewhat arbitrary grouping of observations in the presence of continuous covariates. For future research, we intend to generalize the NAOV method to models with both discrete and continuous covariates.

Acknowledgements

References

Balakrishnan, N. and Nevzorov, V. B. (2003). *A Primer on Statistical Distributions.* Wiley, Hoboken, New Jersey.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* **12**, 387-415.

Brumback, L. C., Pepe, M. S. and Alonzo, T. A. (2006). Using the ROC curve for gauging treatment effect in clinical trials. *Statistics in Medicine* **25**, 575-590.

DeLong, E. R., DeLong, D. and Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837-845.

Dodd, L. E. and Pepe, M. S. (2003). Semiparametric regression for the area under the receiver operating characteristic curve. *Journal of the American Statistical Association* **98**, 409-417.

Hanley, J. A. and Hajian-Tilaki, K. O. (1997). Sampling variability of nonparametric estimate of the areas under receiver operating characteristic curves: an update. *Academic Radiology* **4**, 49-58.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford Statistical Science Series 28, Oxford University Press, New York.

Van Elteren P. H. (1960). On the combination of independent two sample tests of Wilcoxon. *Bulletin of the Institute of International Statistics* **37**, 351-361.

Zhou, X. H., McClish, D. K. and Obuchowski, N. A. (2002). *Statistical Methods in Diagnostic Medicine.* Wiley, New York.

Lin Zhang
Quintiles Inc.
Overland Park, KS 66213, USA
lin.zhang@quintiles.com.

Yan D. Zhao
Department of Clinical Sciences and Simmons Cancer Center
University of Texas Southwestern Medical Center
Dallas, TX 75390, USA
yandzhao@gmail.com

Jack D. Tubbs
Department of Statistical Science
Baylor University
Waco, TX 76798, USA
Jack_Tubbs@baylor.edu