

**Web-based Supplementary Materials for “Robust Estimation of Area Under
ROC Curve Using Auxiliary Variables In the Presence of Missing Biomarker
Values”**

Qi Long^{1,*}, Xiaoxi Zhang², and Brent A. Johnson¹

¹Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, U.S.A.

² Pfizer Inc., New York, NY 11017, U.S.A.

**email*: qlong@emory.edu

Web Appendix A: Proof of Theorem 1 and 2

We first prove Theorem 1. Following the notation in Section 2, let

$$\begin{aligned} \mathcal{V}_{i,j}(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) &\equiv \theta \frac{\delta_i \delta_j}{\pi_i \pi_j} D_i (1 - D_j) - \frac{\delta_i \delta_j}{\pi_i \pi_j} D_i (1 - D_j) I_{ij} \\ &\quad + \frac{\delta_i \delta_j - \pi_i \pi_j}{\pi_i \pi_j} D_i (1 - D_j) E \{I(X_i > X_j) \mid \mathbf{Z}_i, \mathbf{Z}_j, D_i, D_j\}, \end{aligned}$$

where π_i depends on $\boldsymbol{\alpha}$ and $E \{I(X_i > X_j) \mid \mathbf{Z}_i, \mathbf{Z}_j, D_i, D_j\}$ depends on $\boldsymbol{\beta}$. It follows that $\mathcal{V} = \sum_{i \neq j} \mathcal{V}_{i,j}(\theta, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ are the set of estimating equations that are used to obtain $\hat{\theta}_{DR}$. Let $\mathcal{V}_n(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0.5n^{-2} \sum_{i,j} \{\mathcal{V}_{i,j}(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \mathcal{V}_{j,i}(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta})\}$, and it is straightforward to show that $\hat{\theta}_{DR}$ is the solution of $\mathcal{V}_n(\theta, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$.

Let $\mathcal{U}_n = \mathcal{V}_n(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathcal{V}_E(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta})$, where $\mathcal{V}_E = E(\mathcal{V}_n) = 0.5E[\mathcal{V}_{i,j}(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \mathcal{V}_{j,i}(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta})]$.

Suppose that the following conditions hold:

(A1) The U -process \mathcal{U}_n is stochastically equicontinuous.

(A2) \mathcal{V}_E is differentiable in $(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta})$.

(A3) \mathcal{V}_n and $\partial \mathcal{V}_n / \partial(\boldsymbol{\alpha}, \boldsymbol{\beta})$ converge uniformly to \mathcal{V}_E and $\partial \mathcal{V}_E / \partial(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Let $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ be the probability limits of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using the two working models. Consistency of $\hat{\theta}_{DR}$ follows by the uniform convergence of \mathcal{V}_n to \mathcal{V}_E , because it is straightforward to verify that $\mathcal{V}_E(\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$ when either working model is correctly specified, that is, $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_T$ and/or $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_T$. We now derive the asymptotic distribution of $\hat{\theta}_{DR}$.

$$\begin{aligned} 0 &= \sqrt{n} \mathcal{V}_n(\hat{\theta}_{DR}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}), \\ &= \sqrt{n} \left\{ \mathcal{V}_n(\hat{\theta}_{DR}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathcal{V}_E(\hat{\theta}_{DR}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \right\} - \sqrt{n} \left\{ \mathcal{V}_n(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \mathcal{V}_E(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \right\} \\ &\quad + \sqrt{n} \left\{ \mathcal{V}_E(\hat{\theta}_{DR}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathcal{V}_E(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \right\} + \sqrt{n} \mathcal{V}_n(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0), \\ &= o_p(1) + \sqrt{n} \left\{ \mathcal{V}_E(\hat{\theta}_{DR}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathcal{V}_E(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \right\} + \sqrt{n} \mathcal{V}_n(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0), \\ &= o_p(1) + \frac{\partial \mathcal{V}_E}{\partial \theta}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \sqrt{n}(\hat{\theta}_{DR} - \theta) + \frac{\partial \mathcal{V}_E}{\partial \boldsymbol{\alpha}}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ &\quad + \frac{\partial \mathcal{V}_E}{\partial \boldsymbol{\beta}}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \sqrt{n} \mathcal{V}_n(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0), \end{aligned}$$

where $(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1)$ lies between $(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and $(\hat{\theta}_{DR}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. The first identity follows our previous argument, the third identity follows from Condition (A1), and the fourth identity follows from a Taylor expansion and the regularity conditions. After we replace $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ by their respective influence function, that are due to the two working models for $(\mathcal{M}1)$ and $(\mathcal{M}2)$, namely, ψ_i^α and ψ_i^β , and rearrange terms, we arrive at

$$\begin{aligned} \sqrt{n}(\hat{\theta}_{DR} - \theta) = & \left\{ \frac{\partial \mathcal{V}_E}{\partial \theta}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \right\}^{-1} \left\{ -\frac{\partial \mathcal{V}_E}{\partial \boldsymbol{\alpha}}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \sqrt{n} \times n^{-1} \sum_i \psi_i^\alpha \right. \\ & \left. - \frac{\partial \mathcal{V}_E}{\partial \boldsymbol{\beta}}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \sqrt{n} \times n^{-1} \sum_i \psi_i^\beta - \sqrt{n} \mathcal{V}_n(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \right\} + o_p(1) \end{aligned}$$

It is straightforward to verify that 1) $\frac{\partial \mathcal{V}_E}{\partial \theta}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) \xrightarrow{p} E \left\{ \frac{\delta_i \delta_j}{\pi_i \pi_j} D_i (1 - D_j) \right\}$, which reduces to $\frac{n_1}{n} \frac{n_0}{n}$ when the working model for δ is correctly specified; 2) the remaining partial derivative terms all converge in probability to the expectation of the respective term evaluated at $(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, e.g., $\frac{\partial \mathcal{V}_E}{\partial \boldsymbol{\alpha}}(\theta^1, \boldsymbol{\alpha}^1, \boldsymbol{\beta}^1)$ converges to $E \left\{ \frac{\partial \mathcal{V}_{ij}}{\partial \boldsymbol{\alpha}}(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \right\}$; and 3) the last term $\sqrt{n} \mathcal{V}_n(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = \sqrt{n} \times n^{-1} \sum_i E \{ \mathcal{V}_{i,j}(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + \mathcal{V}_{j,i}(\theta, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \mid O_i \} + o_p(1)$, which follows using standard arguments on the limiting distribution of U -statistics (van der Vaart (1998), Ch. 12). Due to the uniform convergence of the terms in the expression of $\hat{\Omega}$, it follows that $\hat{\Omega}$ is a consistent estimator for Ω . The proof of Theorem 1 is now complete.

Theorem 2 can be proved along the same lines under regularity conditions that parallel (A1)-(A3).

Web Appendix B: Simulation Studies Using Original Weights

We repeated the simulations in Section 3.1 with the original weights (i.e., $\frac{1}{\pi_i}$) and the results are summarized in Web Tables 1 and 2, which parallel Tables 1 and 2 in the paper. Compared to Tables 1 and 2 in the paper, the original weights lead to more noise in the estimation of $\hat{\theta}_{IW}$, $\hat{\theta}_{DR}$ and $\hat{\theta}_{DR-N}$, i.e., large SD. In addition, the bootstrap SE of $\hat{\theta}_{DR-N}$ using the original weights tends to overestimate its SD when $(\mathcal{M}1)$ is correctly specified, which is likely due to

small sample sizes and some high missing probabilities. Extreme missing probabilities in a few bootstrap samples can lead to unstable estimates in these bootstrap samples and hence bootstrap SE that is greater than SD. In our additional simulations, this problem becomes less pronounced as the sample size increases and/or the missing probabilities become more moderate.

[Table 1 about here.]

[Table 2 about here.]

Web Appendix C: Data Analysis Using Original Weights

We repeated the data analysis in Section 4 using the original weights. The results are summarized in Web Table 3. Compared to the results using the modified weights (Table 4 in the paper), most results remain close except that the bootstrap SE of $\hat{\theta}_{DR-N}$ is substantially greater, which is primarily due to the high percentage of missing data.

[Table 3 about here.]

References

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Press.

Table 1

Results of simulation study under MAR: comparison of $\hat{\theta}_0$, $\hat{\theta}_{IW}$, $\hat{\theta}_{DR}$ and $\hat{\theta}_{DR-N}$ using the original weights, when \mathbf{Z}_1 and \mathbf{Z}_2 are identical. ε is Gaussian (i.e., $\varepsilon \sim N(0, 1)$) or non-Gaussian (i.e., $\varepsilon = 20\{\eta - E(\eta)\}$ with $\eta \sim \text{Beta}(5, 1)$). True θ is 0.722 for Gaussian ε and 0.675 for non-Gaussian ε . $\hat{\theta}_{GS}$ and $\hat{\theta}_{IMP}$ under the correctly specified ($\mathcal{M}2$) provide optimal benchmarks for bias and efficiency, respectively. RB, relative bias as the percentage of the true θ ; SD, Monte Carlo standard deviation of parameter estimates; SMSE, square root of mean squared errors; SE, mean of the standard error estimates; CR, coverage rate of 95% Wald's confidence interval.

	Gaussian ε					non-Gaussian ε				
	RB (%)	SE	SD	SMSE	CR (%)	RB (%)	SE	SD	SMSE	CR (%)
$\hat{\theta}_{GS}$	-0.2	0.036	0.037	0.037	94.0	0.0	0.038	0.038	0.038	95.8
$\hat{\theta}_0$	11.6	0.050	0.054	0.099	70.0	10.8	0.057	0.056	0.092	80.4
Both mean models correctly specified ¹										
$\hat{\theta}_{IMP}$	-0.1	0.040	0.042	0.042	95.0	-0.8	0.054	0.054	0.054	94.8
$\hat{\theta}_{IW}$	0.5	0.048	0.052	0.052	93.2	1.0	0.056	0.058	0.059	95.0
$\hat{\theta}_{DR}$	0.5	0.046	0.048	0.048	94.2	1.0	0.057	0.059	0.059	96.4
$\hat{\theta}_{DR-N}$	0.5	0.054	0.050	0.050	98.0	0.9	0.065	0.059	0.059	98.0
Mean model for ($\mathcal{M}1$) misspecified ²										
$\hat{\theta}_{IW}$	8.4	0.050	0.054	0.081	78.6	8.0	0.056	0.058	0.079	84.8
$\hat{\theta}_{DR}$	0.0	0.040	0.044	0.044	93.8	0.7	0.052	0.055	0.056	94.2
$\hat{\theta}_{DR-N}$	0.0	0.044	0.044	0.044	95.8	0.5	0.057	0.056	0.056	96.4
Mean model for ($\mathcal{M}2$) misspecified ³										
$\hat{\theta}_{DR}$	1.1	0.058	0.062	0.062	96.4	1.4	0.063	0.065	0.065	95.8
$\hat{\theta}_{DR-N}$	1.1	0.071	0.062	0.062	99.4	1.4	0.074	0.065	0.066	97.6
Both mean models misspecified ⁴										
$\hat{\theta}_{DR}$	8.4	0.050	0.054	0.081	79.2	8.0	0.057	0.059	0.080	85.4
$\hat{\theta}_{DR-N}$	8.4	0.053	0.053	0.081	82.6	8.0	0.059	0.059	0.080	88.2

1: The correct model includes \mathbf{Z}_1 and D for ($\mathcal{M}1$) and \mathbf{Z}_2 and D for ($\mathcal{M}2$)

2: The misspecified ($\mathcal{M}1$) includes only $Z_1^{(1)}$ and D .

3: The misspecified ($\mathcal{M}2$) includes only $Z_2^{(1)}$ and D .

4: Both mean working models are misspecified as in 2 and 3.

Table 2

Results of simulation study under MAR: comparison of $\hat{\theta}_0$, $\hat{\theta}_{IW}$, $\hat{\theta}_{DR}$ and $\hat{\theta}_{DR-N}$ using the original weights, when \mathbf{Z}_1 and \mathbf{Z}_2 are independent. ε is Gaussian (i.e., $\varepsilon \sim N(0, 1)$) or non-Gaussian (i.e., $\varepsilon = 20\{\eta - E(\eta)\}$ with $\eta \sim \text{Beta}(5, 1)$). True θ is 0.722 for Gaussian ε and 0.675 for non-Gaussian ε . $\hat{\theta}_{GS}$ and $\hat{\theta}_{IMP}$ under the correctly specified ($\mathcal{M}2$) provide optimal benchmarks for bias and efficiency, respectively. RB, relative bias as the percentage of the true θ ; SD, Monte Carlo standard deviation of parameter estimates; SMSE, square root of mean squared errors; SE, mean of the standard error estimates; CR, coverage rate of 95% Wald's confidence interval.

	Gaussian ε					non-Gaussian ε				
	RB (%)	SE	SD	SMSE	CR (%)	RB (%)	SE	SD	SMSE	CR (%)
$\hat{\theta}_{GS}$	0.2	0.036	0.035	0.035	96.0	0.2	0.038	0.036	0.036	96.0
$\hat{\theta}_0$	-0.1	0.059	0.057	0.057	95.8	0.4	0.063	0.061	0.061	96.4
Both mean models correctly specified ¹										
$\hat{\theta}_{IMP}$	0.1	0.039	0.040	0.040	94.2	-0.6	0.051	0.050	0.050	95.8
$\hat{\theta}_{IW}$	-0.1	0.059	0.060	0.059	94.8	0.4	0.062	0.065	0.065	94.8
$\hat{\theta}_{DR}$	0.8	0.047	0.046	0.046	96.8	0.9	0.057	0.056	0.056	95.4
$\hat{\theta}_{DR-N}$	0.8	0.055	0.046	0.046	97.6	0.8	0.062	0.056	0.056	98.4
Mean model for ($\mathcal{M}1$) misspecified ²										
$\hat{\theta}_{IW}$	-0.2	0.058	0.059	0.059	95.0	0.4	0.062	0.062	0.062	95.2
$\hat{\theta}_{DR}$	0.3	0.041	0.041	0.041	94.6	0.4	0.053	0.052	0.052	95.4
$\hat{\theta}_{DR-N}$	0.3	0.043	0.041	0.041	95.6	0.3	0.056	0.052	0.052	96.8
Mean model for ($\mathcal{M}2$) misspecified ³										
$\hat{\theta}_{DR}$	0.3	0.059	0.058	0.058	96.2	1.1	0.063	0.065	0.065	95.2
$\hat{\theta}_{DR-N}$	0.3	0.068	0.058	0.058	97.6	1.1	0.072	0.065	0.065	97.6
Both Mean Models misspecified ⁴										
$\hat{\theta}_{DR}$	-0.1	0.054	0.055	0.055	95.4	0.6	0.060	0.061	0.061	95.4
$\hat{\theta}_{DR-N}$	-0.1	0.057	0.055	0.055	97.0	0.5	0.063	0.060	0.061	96.6

1: The correct model includes \mathbf{Z}_1 and D for ($\mathcal{M}1$) and \mathbf{Z}_2 and D for ($\mathcal{M}2$)

2: The misspecified ($\mathcal{M}1$) includes only $Z_1^{(1)}$ and D .

3: The misspecified ($\mathcal{M}2$) includes only $Z_2^{(1)}$ and D .

4: Both mean working models are misspecified as in 2 and 3.

Table 3

Sensitivity analysis using the original weights for estimation of the ROC AUC (θ) in the maternal depression study

	$\alpha_X = -1$		$\alpha_X = 0$		$\alpha_X = 1$	
	Estimate	SE	Estimate	SE	Estimate	SE
$\hat{\theta}_{IW}$	0.864	0.037	0.851	0.040	0.849	0.042
$\hat{\theta}_{DR}$	0.874	0.028	0.853	0.030	0.842	0.032
$\hat{\theta}_{DR-N}$	0.874	0.058	0.853	0.056	0.842	0.057