# Web-based Supplementary Materials for "Moment Adjusted Imputation for Multivariate Measurement Error Data with Applications to Logistic Regression"

BY

Laine Thomas, Leonard Stefanski, Marie Davidian

## A Special Case: Matching two moments and cross product

In the simple case where $\boldsymbol{V}_i = \boldsymbol{Y}_i$ and we are only interested in matching two moments of $\boldsymbol{X}_i$, the $\widehat{\boldsymbol{X}}_i$ can be obtained analytically. In order to simply derivation, in this section only, we minimize the unweighted distance

$$\sum_{i=1}^{n}(\boldsymbol{W}_i - \boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{X}_i) \tag{1}$$

rather than the weighted distance

$$\sum_{i=1}^{n}(\boldsymbol{W}_i - \boldsymbol{X}_i)^T\Sigma_{ui}^{-1}(\boldsymbol{W}_i - \boldsymbol{X}_i).$$

Based on Web Appendix Equation (1), the objective function is

$$\begin{aligned}&n^{-1}\textstyle\sum_{i=1}^{n}(\boldsymbol{W}_i - \boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{X}_i) + \boldsymbol{\lambda}_1^T(n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{W}_i)\\ &+\boldsymbol{\lambda}_2^T\text{vech}(n^{-1}\textstyle\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T - \boldsymbol{W}_i\boldsymbol{W}_i^T + \Sigma_{ui}) + \boldsymbol{\lambda}_3^T(n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_iY_i - \boldsymbol{W}_iY_i)\end{aligned}$$

where $\boldsymbol{\lambda}_1^T$ and $\boldsymbol{\lambda}_3^T$ are vectors of Lagrange multipliers and $\boldsymbol{\lambda}_2^T$ is a vector. Let $\mathrm{I}_G$ denote the identity matrix of dimension $G$. Taking the derivative with respect to $\boldsymbol{X}_i$ gives $(\boldsymbol{X}_i - \boldsymbol{W}_i) + \boldsymbol{\lambda}_1 + (\{\text{vech}^{-1}(\boldsymbol{\lambda}_2) + \mathrm{I}_G\}\boldsymbol{X}_i + \boldsymbol{\lambda}_3Y_i = 0$, and the solution for $\boldsymbol{X}_i$ is $\boldsymbol{X}_i = \boldsymbol{A}(\boldsymbol{W}_i - \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_3Y_i)$ where $\boldsymbol{A} = \{2\mathrm{I}_G\text{vech}^{-1}(\boldsymbol{\lambda}_2) + \mathrm{I}_G\}^{-1}$, where $\text{vech}^{-1}$ re-creates a symmetric matrix from its vech half so that if $\boldsymbol{A} = \text{vech}(\boldsymbol{B})$ for symmetric matrix $\boldsymbol{B}$, then $\text{vech}^{-1}(\boldsymbol{A}) = \boldsymbol{B}$. The solution for $\boldsymbol{X}_i$ depends on the unknown $\Lambda = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_3^T, \boldsymbol{\lambda}_2^T)^T$, which must be estimated to obtain $\widehat{\boldsymbol{X}}_i$. Taking the derivative with respect to $\Lambda$ provides additional equations that we can solve to obtain $\widehat{\Lambda}$.

Rather than solving for $\Lambda$ directly, it is easier to solve for "coefficients" in the equation for $\boldsymbol{X}_i$. Note that the $\boldsymbol{X}_i$ have the form $\boldsymbol{A}\boldsymbol{W}_i + \boldsymbol{B} + \boldsymbol{C}Y_i$, and we solve for the "coefficients" $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ as follows:

Constraint 1: $\overline{\widehat{\boldsymbol{X}}} = \overline{\boldsymbol{W}}$ for $\overline{\widehat{\boldsymbol{X}}} = n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{X}}_i$, and $\overline{\boldsymbol{W}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i$

So $\widehat{\boldsymbol{A}}\overline{\boldsymbol{W}} + \widehat{\boldsymbol{B}} + \widehat{\boldsymbol{C}}\overline{Y} = \overline{\boldsymbol{W}}$ and $\widehat{\boldsymbol{B}} = (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} - \widehat{\boldsymbol{C}}\overline{Y}$.

Substituting this for $\widehat{\boldsymbol{B}}$ we have $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{A}}\boldsymbol{W}_i + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} + \widehat{\boldsymbol{C}}(Y_i - \overline{Y})$.

Constraint 3: $n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{X}}_iY_i = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_iY_i$

Equivalently $\boldsymbol{S}_{\widehat{X}Y} = \boldsymbol{S}_{WY}$ for

$$\boldsymbol{S}_{\widehat{X}Y} = n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{X}}_i(Y_i - \overline{Y}) \text{ and } \boldsymbol{S}_{WY} = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i(Y_i - \overline{Y}).$$

Substituting $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{A}}\boldsymbol{W}_i + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} + \widehat{\boldsymbol{C}}(Y_i - \overline{Y})$ implies that

$$\widehat{\boldsymbol{A}}\boldsymbol{S}_{WY} + \widehat{\boldsymbol{C}}s_Y^2 = \boldsymbol{S}_{WY}, \text{ where } s_Y^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

This gives $\widehat{\boldsymbol{C}} = (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{S}_{WY}/s_Y^2$.

Constraint 2: $n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i\boldsymbol{W}_i^T - \Sigma_{ui}$

Equivalently $\boldsymbol{S}_{\widehat{X}\widehat{X}} = \boldsymbol{S}_{WW} - \Sigma_{ui}$.

$$\boldsymbol{S}_{\widehat{X}\widehat{X}} = n^{-1}\sum_{i=1}^{n}(\widehat{\boldsymbol{X}}_i - \overline{\widehat{\boldsymbol{X}}}_i)(\widehat{\boldsymbol{X}}_i - \overline{\widehat{\boldsymbol{X}}}_i)^T$$

$$\boldsymbol{S}_{WW} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{W}_i - \overline{\boldsymbol{W}})(\boldsymbol{W}_i - \overline{\boldsymbol{W}})^T$$

Substituting $(\widehat{\boldsymbol{X}}_i - \overline{\widehat{\boldsymbol{X}}}_i) = \widehat{\boldsymbol{A}}(\boldsymbol{W}_i - \overline{\boldsymbol{W}}) + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{S}_{WY}(Y_i - \overline{Y})/s_Y^2$ we have

$$\boldsymbol{S}_{\widehat{X}\widehat{X}} = \widehat{\boldsymbol{A}}\boldsymbol{S}_{WW}\widehat{\boldsymbol{A}}^T + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{M} + \widehat{\boldsymbol{A}}\boldsymbol{M}(\mathrm{I}_G - \widehat{\boldsymbol{A}})^T + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{M}\widehat{\boldsymbol{A}}^T = \boldsymbol{S}_{WW} - \Sigma_{ui},$$

where $\boldsymbol{M} = \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T/s_Y^2$.

Algebraic simplification leads to the equation

$$\widehat{\boldsymbol{A}}(\boldsymbol{S}_{WW} - \boldsymbol{M})\widehat{\boldsymbol{A}}^T = \boldsymbol{S}_{WW} - \Sigma_{ui} - \boldsymbol{M}, \text{ or}$$

$$\widehat{\boldsymbol{A}}(\boldsymbol{S}_{WW}s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T)\widehat{\boldsymbol{A}}^T = (\boldsymbol{S}_{WW} - \Sigma_{ui})s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T.$$

The matrix $\boldsymbol{V}_1 = (\boldsymbol{S}_{WW}s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T)$ is positive definite by the Cauchy Schwartz inequality. If $\boldsymbol{V}_2 = (\boldsymbol{S}_{WW} - \Sigma_{ui})s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T$ is non-negative definite, then the equation can be solved to yield $\widehat{\boldsymbol{A}} = \boldsymbol{V}_1^{-1/2}(\boldsymbol{V}_1^{1/2}\boldsymbol{V}_2\boldsymbol{V}_1^{1/2})^{1/2}\boldsymbol{V}_1^{-1/2}$. The second matrix $\boldsymbol{V}_2$ is usually non-negative definite, unless the measurement error is very large and/or the correlation between $\boldsymbol{X}_i$ and $Y_i$ is extremely high.

This defines our adjusted data $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{A}}\boldsymbol{W}_i + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} + \widehat{\boldsymbol{C}}(Y_i - \overline{Y})$.

# B Comparison of Sequential and Joint Minimization for Adjusting Data

Here, we consider additional cases that support the conclusions in Section 4.1, and we assess the importance of order in adjusting data sequentially. We compare two implementation strategies. Joint weighted minimization (JW) refers to to the method developed in Section 2.1, and sequential adjustment was proposed in Section 3. In Section 3, we propose that sequential adjustment be performed in the order of smallest to largest measurement error. The adjusted data obtained this way will be denoted $\widehat{\boldsymbol{X}}_{Seq-C}$, indicating "correct" adjustment order. When the order is switched, the adjusted data are denoted $\widehat{\boldsymbol{X}}_{Seq-W}$, for "wrong" order. Our focus is on comparing alternative implementations of MAI, and in call versions we match four moments.

## B.1 Comparison of Adjustment: Scenario 1

Here we compare the adjusted data, $\widehat{\boldsymbol{X}}$, and estimation of logistic regression model coefficients, based on a single data set. The data generation is similar to Section 4.1. However, the measurement error in $W_1$ and $W_2$ is extremely different. $W_1$ has large measurement error, with a reliability ratio of 0.5, and $W_2$ has virtually no measurement error, with a reliability ratio greater than 0.99. The scenario is the following:

- Model: $P(Y=1|X_1,X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- $n = 1000$ observations of $(Y, X_1, X_2)$
- $\boldsymbol{X} = (X_1, X_2)^T \sim MVN(0, \Sigma_X)$
- $\Sigma_X = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
- $(\beta_0, \beta_1, \beta_2) = (1, .5, .5)$ (moderately significant for $n = 1000$, 70% event rate)
- $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$
- $\Sigma_U = \begin{pmatrix} 1 & \rho_u(.01) \\ \rho_u(.01) & .0001 \end{pmatrix}$

- $\rho_u = 0.8$ indicates a correlation in the measurement errors of 0.8.

- $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$

Here, we really have univariate mis-measured data, $W_1$, which should be adjusted, and $W_2$ should not. If an adjustment procedure is implemented for $\boldsymbol{W}$, including both $W_1$ and $W_2$, we would hope that the impact on $W_2$ would be negligible and that parameter estimation would be equivalent to a univariate adjustment of $W_1$. In Web Appendix Figure 1 we compare the adjusted $\widehat{X}_2$ to $W_2$. The correctly ordered sequential adjustment and joint adjustment do not meaningfully alter $W_2$. The wrongly ordered sequential does alter $W_2$, even though it was error free. If the adjusted data are of particular interest, this is undesirable.

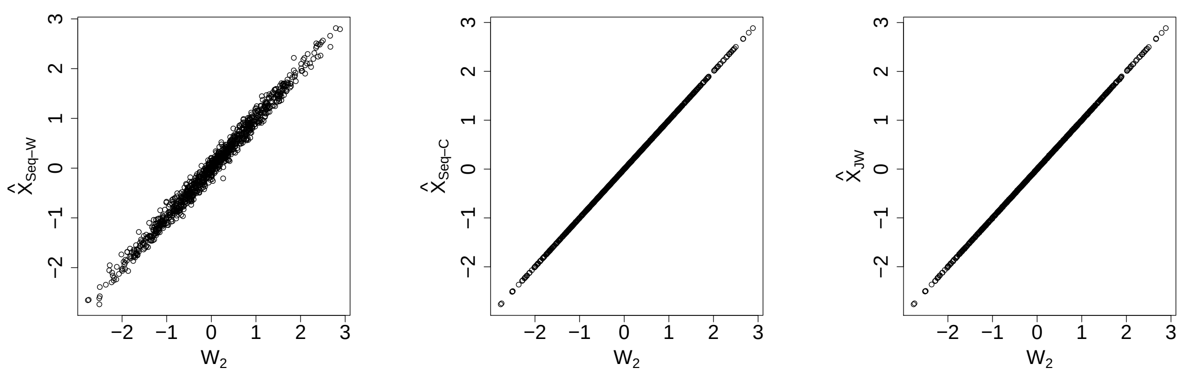


Figure 1: *Direct Comparison of* $\widehat{X}_2$ *to* $W_2$ *for a single data set.* $\widehat{X}_2$*: Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; JW, joint weighted minimization*

Despite differences in the adjusted data themselves, these various methods produce virtually identical logistic regression parameter estimates, even under such extreme circumstances (Web Appendix Table 1). This is consistent with the simulation results in Section 4.1.

Table 1: *Coefficient estimates. Methods: X, true covariates; W, mis-measured covariates; Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; JW, joint weighted minimization*

| Method | $\widehat{\beta}_0$ | $\widehat{\beta}_1$ | $\widehat{\beta}_2$ |
|---|---|---|---|
| X | 1.048 | 0.557 | 0.454 |
| W | 1.032 | 0.238 | 0.592 |
| Seq-W | 1.043 | 0.541 | 0.475 |
| Seq-C | 1.043 | 0.540 | 0.474 |
| JW | 1.043 | 0.540 | 0.474 |

## B.2 Comparison of Adjustment: Scenario 2

Here we make comparisons for the case where both variables are measured with error and the measurement error is highly correlated. The data are generated under the following scenario:

- Model: $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- $n = 1000$ observations of $(Y, X_1, X_2)$
- $\boldsymbol{X} = (X_1, X_2)^T \sim MVN(0, \Sigma_X)$
- $\Sigma_X = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
- $(\beta_0, \beta_1, \beta_2) = (1.5, .5, .5)$ (moderately significant for $n = 1000$, 70% event rate)
- $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$
- $\Sigma_U = \begin{pmatrix} 1 & \rho_u(.6) \\ \rho_u(.6) & .36 \end{pmatrix}$
- $\rho_u = 0.8$ indicates a correlation in the measurement errors of 0.8.
- $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$

The sequential adjustment accounts for correlated measurement error only in the moment estimation, but not in the minimization. Thus, it will differ from joint adjustment. Here, we evaluate the extent of that difference. In Web Appendix Figure 2 we see that the sequentially adjusted data are more similar to the joint minimization when the adjustment occurs in the "wrong" order. In terms of re-creating the joint adjustment, the "wrong" order is preferable under this circumstance. However, in Web Appendix Table 2 we see that none of the methods differ in terms of logistic regression parameter estimation and the difference in density estimation is minimal (Web Appendix Figure 3).

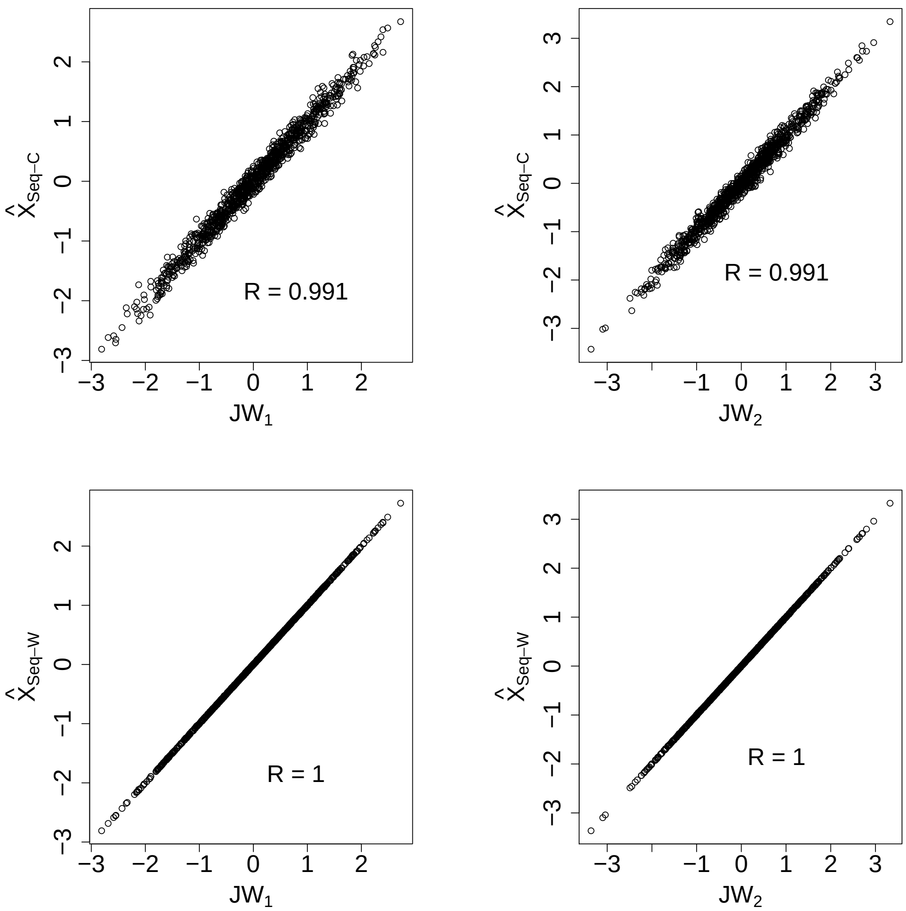


Figure 2: *Direct Comparison of* $\widehat{\boldsymbol{X}}$ *for a single data set.* $\widehat{X}$*: Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; JW, joint weighted minimization*

Table 2: *Coefficient estimates. Methods: X, true covariates; W, mis-measured covariates; Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; JW, joint weighted minimization*

| Method | $\widehat{\beta}_0$ | $\widehat{\beta}_1$ | $\widehat{\beta}_2$ |
|---|---|---|---|
| X | 1.570 | 0.549 | 0.466 |
| W | 1.471 | 0.205 | 0.408 |
| Seq-W | 1.575 | 0.649 | 0.493 |
| Seq-C | 1.577 | 0.658 | 0.490 |
| JW | 1.577 | 0.658 | 0.490 |

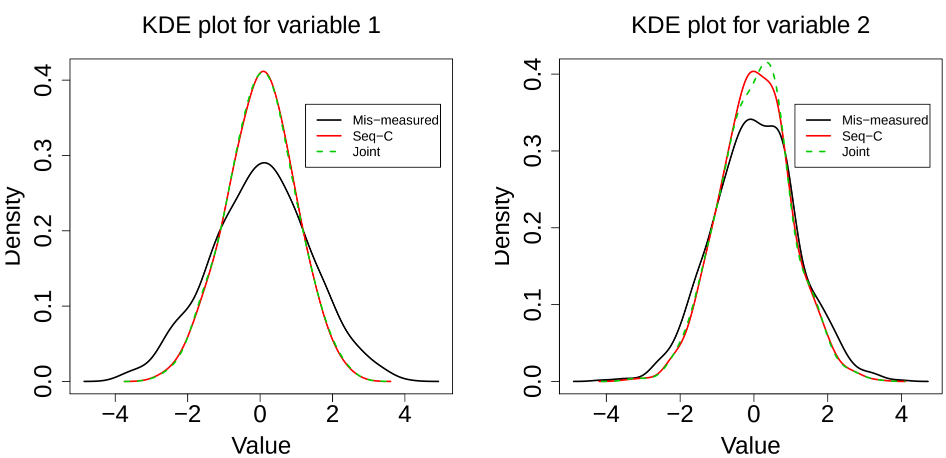


Figure 3: *Kernel density estimation for the marginal densities of* $X_1$ *and* $X_2$*: KDE of* $W$*, solid-dark line: KDE of* $\widehat{X}$ *for sequential MAI, solid-light line: KDE of* $\widehat{X}$ *for JW, joint weighted minimization;* $n = 1000$*.*

# C Simulations in Density Estimation

Table 3: Simulation results for density estimation. Statistics reported: (a) $MSE(W)/MSE(\widehat{X})$, where $MSE(\widehat{X}) = B^{-1}\sum_{b=1}^{B} n^{-1}\sum_{i=1}^{n}(\widehat{X}_{i,b} - X_{i,b})^2$ (coefficient of variation $\approx$ 0.001), and (b) $ISE(G_W)/ISE(G_{\widehat{X}})$, where $ISE(G_{\widehat{X}}) = B^{-1}\sum_{b=1}^{B}\int\{G_{\widehat{X},b}(t) - G_{X,b}(t)\}^2 dt$, for $G_X(t) = n^{-1}\sum_{i=1}^{n} I_{(X_i \le t)}$, $-\infty < t < \infty$ (coefficient of variation $\approx$ 0.02). Adjusted data $\widehat{X}$: M2 and M4, multivariate MAI with M=2 and 4, respectively; U2 and U4, univariate MAI with M=2 and 4, respectively; RC, regression calibration; MR, moment reconstruction;

| | | | | | | | (a) $\frac{MSE(W)}{MSE(\widehat{X})}$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_i$ | $\rho$ | $n$ | RC | MR | M2 | M4 | U2 | U4 | RC | MR | M2 | M4 | U2 | U4 |
| 1 | 0.5 | 1000 | 2.04 | 1.79 | 1.74 | 1.94 | 1.71 | 1.85 | 1.33 | 1.26 | 1.23 | 1.36 | 1.22 | 1.35 |
| | | 2000 | 2.04 | 1.79 | 1.74 | 1.89 | 1.71 | 1.82 | 1.33 | 1.27 | 1.23 | 1.36 | 1.22 | 1.35 |
| | 0.7 | 1000 | 2.43 | 2.01 | 2.04 | 2.24 | 1.71 | 1.88 | 1.30 | 1.20 | 1.23 | 1.37 | 1.22 | 1.35 |
| | | 2000 | 2.45 | 2.01 | 2.04 | 2.21 | 1.71 | 1.86 | 1.30 | 1.23 | 1.23 | 1.37 | 1.22 | 1.35 |
| 1-5 | 0.5 | 1000 | 1.52 | 1.42 | 1.30 | 1.44 | 1.31 | 1.45 | 1.18 | 1.15 | 1.09 | 1.16 | 1.09 | 1.17 |
| | | 2000 | 1.52 | 1.42 | 1.30 | 1.45 | 1.31 | 1.45 | 1.18 | 1.15 | 1.09 | 1.16 | 1.09 | 1.17 |
| | 0.7 | 1000 | 1.72 | 1.55 | 1.46 | 1.59 | 1.31 | 1.45 | 1.17 | 1.13 | 1.09 | 1.17 | 1.09 | 1.17 |
| | | 2000 | 1.73 | 1.55 | 1.45 | 1.59 | 1.31 | 1.45 | 1.17 | 1.13 | 1.09 | 1.17 | 1.09 | 1.17 |

| | | | | | | | (b) $\frac{ISE(W)}{ISE(\widehat{X})}$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_i$ | $\rho$ | $n$ | RC | MR | M2 | M4 | U2 | U4 | RC | MR | M2 | M4 | U2 | U4 |
| 1 | 0.5 | 1000 | 1.86 | 5.49 | 4.35 | 10.51 | 4.34 | 10.48 | 1.36 | 2.73 | 2.27 | 6.48 | 2.27 | 6.43 |
| | | 2000 | 1.88 | 5.85 | 4.66 | 18.25 | 4.65 | 18.13 | 1.44 | 2.93 | 2.42 | 8.48 | 2.43 | 8.36 |
| | 0.7 | 1000 | 2.65 | 6.55 | 5.30 | 10.57 | 4.34 | 10.40 | 1.37 | 2.78 | 2.29 | 7.10 | 2.32 | 7.05 |
| | | 2000 | 2.79 | 7.07 | 5.67 | 16.97 | 4.61 | 16.61 | 1.37 | 2.98 | 2.38 | 8.73 | 2.41 | 8.62 |
| 1-5 | 0.5 | 1000 | 1.98 | 3.29 | 2.60 | 5.40 | 2.68 | 6.39 | 1.36 | 2.01 | 1.70 | 2.97 | 1.73 | 3.25 |
| | | 2000 | 2.07 | 3.44 | 2.70 | 7.68 | 2.79 | 9.54 | 1.32 | 2.09 | 1.73 | 3.52 | 1.78 | 3.99 |
| | 0.7 | 1000 | 2.68 | 3.90 | 3.18 | 6.24 | 2.72 | 6.80 | 1.26 | 1.98 | 1.66 | 2.86 | 1.69 | 3.10 |
| | | 2000 | 2.79 | 4.05 | 3.25 | 8.22 | 2.78 | 9.08 | 1.32 | 2.14 | 1.74 | 3.71 | 1.80 | 4.16 |

# D Simulations in Logistic Regression

Table 4: *Estimation of* $\boldsymbol{\beta}$ *in* $B = 250$ *simulated data sets with* $n = 1000$ *and* $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$, *where* $\boldsymbol{X}$ *are standardized chi-square df=4 with four possible correlations* $\rho$, $\boldsymbol{W}_i = \boldsymbol{X}_i + (1/r_i)\sum_{j=1}^{r_i} \boldsymbol{U}_j$, $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$ $\Sigma_U = vech\{1, 0.5(0.3)^{1/2}, 0.3\}$, *and two levels of replication:* $r_i = r = 1$ *or* $r_i = 1, 2, 3, 4$ *or* 5 *with equal probability. True value of* $\boldsymbol{\beta} = (1.5, .5, .5)$; *B, bias; SD, standard deviation; M,* $MSE_W/MSE_{\widehat{\boldsymbol{X}}}$. *Adjusted data* $\widehat{X}$: *M2, MAI with M=2, M4, MAI with M=4, RC, regression calibration; MR, moment reconstruction.*

| | | | $\beta_1$ | | | | | $\beta_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_i$ | $\rho$ | Stat. | W | M2 | M4 | RC | MR | W | M2 | M4 | RC | MR |
| 1 | 0.5 | Bias | −0.35 | −0.10 | 0.03 | −0.11 | −0.08 | −0.11 | −0.05 | 0.00 | −0.07 | −0.04 |
| | | SD | 0.06 | 0.15 | 0.24 | 0.14 | 0.16 | 0.09 | 0.13 | 0.16 | 0.12 | 0.13 |
| | | MSE-R | 1.00 | 4.00 | 2.12 | 4.00 | 4.01 | 1.00 | 1.11 | 0.81 | 1.06 | 1.10 |
| | -0.5 | Bias | −0.35 | −0.10 | 0.03 | −0.11 | −0.08 | −0.11 | −0.05 | 0.00 | −0.07 | −0.04 |
| | | SD | 0.06 | 0.14 | 0.24 | 0.13 | 0.15 | 0.09 | 0.13 | 0.16 | 0.12 | 0.13 |
| | | MSE-R | 1.00 | 4.20 | 2.27 | 4.14 | 4.24 | 1.00 | 1.06 | 0.80 | 1.01 | 1.06 |
| | 0.7 | Bias | −0.39 | −0.09 | 0.09 | −0.10 | −0.08 | −0.03 | −0.06 | −0.03 | −0.08 | −0.05 |
| | | SD | 0.08 | 0.26 | 0.40 | 0.25 | 0.28 | 0.09 | 0.22 | 0.27 | 0.21 | 0.23 |
| | | MSE-R | 1.00 | 2.08 | 0.95 | 2.22 | 1.94 | 1.00 | 0.19 | 0.13 | 0.18 | 0.18 |
| | -0.7 | Bias | −0.40 | −0.11 | 0.07 | −0.12 | −0.10 | −0.02 | −0.04 | −0.01 | −0.06 | −0.03 |
| | | SD | 0.08 | 0.25 | 0.37 | 0.24 | 0.26 | 0.10 | 0.21 | 0.26 | 0.21 | 0.22 |
| | | MSE-R | 1.00 | 2.22 | 1.17 | 2.31 | 2.12 | 1.00 | 0.22 | 0.15 | 0.22 | 0.21 |
| 1-5 | 0.5 | Bias | −0.25 | −0.08 | 0.00 | −0.11 | −0.06 | −0.04 | −0.01 | 0.02 | −0.02 | 0.00 |
| | | SD | 0.08 | 0.12 | 0.16 | 0.11 | 0.13 | 0.10 | 0.12 | 0.14 | 0.12 | 0.12 |
| | | MSE-R | 1.00 | 3.22 | 2.67 | 2.81 | 3.39 | 1.00 | 0.81 | 0.63 | 0.86 | 0.78 |
| | -0.5 | Bias | −0.24 | −0.07 | 0.01 | −0.10 | −0.06 | −0.05 | −0.02 | 0.00 | −0.03 | −0.01 |
| | | SD | 0.09 | 0.14 | 0.18 | 0.12 | 0.14 | 0.10 | 0.12 | 0.13 | 0.11 | 0.12 |
| | | MSE-R | 1.00 | 2.84 | 2.07 | 2.61 | 2.92 | 1.00 | 0.86 | 0.70 | 0.90 | 0.84 |
| | 0.7 | Bias | −0.30 | −0.09 | 0.03 | −0.14 | −0.07 | 0.05 | 0.01 | 0.01 | 0.02 | 0.01 |
| | | SD | 0.10 | 0.19 | 0.24 | 0.16 | 0.19 | 0.14 | 0.20 | 0.22 | 0.18 | 0.20 |
| | | MSE-R | 1.00 | 2.29 | 1.62 | 2.17 | 2.33 | 1.00 | 0.55 | 0.44 | 0.67 | 0.54 |
| | -0.7 | Bias | −0.30 | −0.09 | 0.02 | −0.15 | −0.08 | 0.05 | 0.00 | 0.01 | 0.02 | 0.00 |
| | | SD | 0.10 | 0.20 | 0.26 | 0.16 | 0.20 | 0.13 | 0.19 | 0.22 | 0.18 | 0.20 |
| | | MSE-R | 1.00 | 2.19 | 1.54 | 2.12 | 2.19 | 1.00 | 0.53 | 0.41 | 0.63 | 0.50 |

Table 5: *Estimation of* $\boldsymbol{\beta}$ *in* $B = 250$ *simulated data sets for* $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$*, where* $\boldsymbol{X}$ *are standardized chi-square df=4 with three possible correlations* $\rho$*,* $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$*,* $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$ $\Sigma_U = vech\{1, 0.5(0.3)^{1/2}, 0.3\}$*, and three sample sizes (n). True value of* $\boldsymbol{\beta} = (1.5, .5, .5)$*; B, bias; SD, standard deviation; M,* $MSE_W/MSE_{\widehat{\boldsymbol{X}}}$*. Adjusted data* $\widehat{X}$*: M2, MAI with M=2, M4, MAI with M=4 RC, regression calibration; MR, moment reconstruction.*

| | | | $\beta_1$ | | | | | $\beta_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $n$ | Stat. | W | M2 | M4 | RC | MR | W | M2 | M4 | RC | MR |
| 0.30 | 1000 | Bias | −0.34 | −0.08 | 0.03 | −0.10 | −0.07 | −0.15 | −0.05 | −0.01 | −0.07 | −0.04 |
| | | SD | 0.06 | 0.14 | 0.30 | 0.12 | 0.14 | 0.08 | 0.10 | 0.12 | 0.10 | 0.11 |
| | | MSE-R | 1.00 | 4.72 | 1.32 | 4.71 | 4.73 | 1.00 | 2.16 | 1.80 | 2.00 | 2.21 |
| | 2000 | Bias | −0.34 | −0.09 | −0.01 | −0.11 | −0.08 | −0.16 | −0.07 | −0.03 | −0.08 | −0.05 |
| | | SD | 0.04 | 0.09 | 0.13 | 0.08 | 0.09 | 0.06 | 0.08 | 0.10 | 0.08 | 0.08 |
| | | MSE-R | 1.00 | 6.94 | 6.80 | 6.22 | 7.44 | 1.00 | 2.68 | 2.90 | 2.32 | 2.91 |
| | 9000 | Bias | −0.34 | −0.10 | −0.01 | −0.11 | −0.09 | −0.15 | −0.06 | −0.01 | −0.07 | −0.05 |
| | | SD | 0.02 | 0.05 | 0.07 | 0.04 | 0.05 | 0.03 | 0.04 | 0.05 | 0.04 | 0.04 |
| | | MSE-R | 1.00 | 10.20 | 25.19 | 8.18 | 11.90 | 1.00 | 5.07 | 9.03 | 3.60 | 6.58 |
| 0.50 | 1000 | Bias | −0.35 | −0.10 | 0.02 | −0.12 | −0.09 | −0.12 | −0.07 | −0.02 | −0.09 | −0.06 |
| | | SD | 0.07 | 0.15 | 0.24 | 0.14 | 0.15 | 0.09 | 0.13 | 0.16 | 0.12 | 0.13 |
| | | MSE-R | 1.00 | 3.98 | 2.19 | 3.98 | 4.06 | 1.00 | 1.16 | 0.94 | 1.08 | 1.16 |
| | 2000 | Bias | −0.35 | −0.10 | 0.01 | −0.11 | −0.09 | −0.12 | −0.05 | −0.01 | −0.07 | −0.04 |
| | | SD | 0.04 | 0.10 | 0.15 | 0.09 | 0.11 | 0.06 | 0.09 | 0.11 | 0.08 | 0.09 |
| | | MSE-R | 1.00 | 6.26 | 5.46 | 5.78 | 6.76 | 1.00 | 1.66 | 1.50 | 1.41 | 1.74 |
| | 9000 | Bias | −0.35 | −0.11 | −0.01 | −0.12 | −0.10 | −0.12 | −0.06 | −0.01 | −0.08 | −0.05 |
| | | SD | 0.02 | 0.05 | 0.07 | 0.05 | 0.05 | 0.03 | 0.04 | 0.05 | 0.04 | 0.04 |
| | | MSE-R | 1.00 | 8.66 | 23.44 | 7.30 | 10.28 | 1.00 | 2.97 | 5.40 | 1.99 | 3.56 |
| 0.70 | 1000 | Bias | −0.39 | −0.10 | 0.07 | −0.11 | −0.09 | −0.03 | −0.05 | −0.01 | −0.07 | −0.04 |
| | | SD | 0.08 | 0.25 | 0.39 | 0.24 | 0.26 | 0.10 | 0.22 | 0.27 | 0.21 | 0.22 |
| | | MSE-R | 1.00 | 2.20 | 1.03 | 2.32 | 2.11 | 1.00 | 0.23 | 0.16 | 0.23 | 0.22 |
| | 2000 | Bias | −0.39 | −0.11 | 0.04 | −0.12 | −0.10 | −0.03 | −0.05 | −0.01 | −0.07 | −0.04 |
| | | SD | 0.06 | 0.18 | 0.26 | 0.17 | 0.19 | 0.08 | 0.17 | 0.20 | 0.16 | 0.17 |
| | | MSE-R | 1.00 | 3.48 | 2.29 | 3.53 | 3.45 | 1.00 | 0.23 | 0.16 | 0.22 | 0.22 |
| | 9000 | Bias | −0.40 | −0.13 | 0.00 | −0.14 | −0.13 | −0.02 | −0.02 | 0.02 | −0.05 | −0.02 |
| | | SD | 0.02 | 0.08 | 0.11 | 0.08 | 0.08 | 0.03 | 0.07 | 0.09 | 0.07 | 0.07 |
| | | MSE-R | 1.00 | 6.61 | 12.99 | 6.15 | 7.13 | 1.00 | 0.25 | 0.18 | 0.19 | 0.25 |

Table 6: *Estimation of* $\boldsymbol{\beta}$ *in* $B = 250$ *simulated data sets for* $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$*, where* $\boldsymbol{X} = (X_1, X_2)^T \sim MVN(0, \Sigma_X)$*,* $\Sigma_X = vech\{1, \rho, 1\}$ *and three possible correlations* $\rho$*,* $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$*,* $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$ $\Sigma_U = vech\{1, 0.5(0.3)^{1/2}, 0.3\}$*, and three sample sizes (n). True value of* $\boldsymbol{\beta} = (1.5, .5, .5)$*; B, bias; SD, standard deviation; M,* $MSE_W/MSE_{\widehat{\boldsymbol{X}}}$*. Adjusted data* $\widehat{X}$*: M2, MAI with M=2, M4, MAI with M=4 RC, regression calibration; MR, moment reconstruction.*

| | | | $\beta_1$ | | | | | $\beta_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $n$ | Stat. | W | M2 | M4 | RC | MR | W | M2 | M4 | RC | MR |
| 0.3 | 1000 | Bias | −0.30 | 0.00 | 0.00 | −0.02 | 0.00 | −0.10 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.06 | 0.13 | 0.14 | 0.12 | 0.14 | 0.07 | 0.10 | 0.10 | 0.10 | 0.10 |
| | | MSE-R | 1.00 | 5.26 | 4.92 | 6.04 | 5.21 | 1.00 | 1.60 | 1.57 | 1.74 | 1.58 |
| | 2000 | Bias | −0.30 | 0.00 | 0.00 | −0.02 | 0.00 | −0.11 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.05 | 0.10 | 0.10 | 0.09 | 0.10 | 0.06 | 0.07 | 0.08 | 0.07 | 0.07 |
| | | MSE-R | 1.00 | 9.28 | 8.90 | 10.36 | 9.20 | 1.00 | 2.60 | 2.55 | 2.70 | 2.60 |
| | 9000 | Bias | −0.30 | 0.00 | 0.00 | −0.02 | 0.00 | −0.11 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.02 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| | | MSE-R | 1.00 | 55.87 | 55.43 | 50.45 | 55.42 | 1.00 | 10.44 | 10.38 | 8.86 | 10.40 |
| 0.5 | 1000 | Bias | −0.31 | 0.01 | 0.01 | −0.02 | 0.01 | −0.06 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.07 | 0.15 | 0.15 | 0.14 | 0.15 | 0.08 | 0.12 | 0.12 | 0.12 | 0.12 |
| | | MSE-R | 1.00 | 4.35 | 4.25 | 4.98 | 4.28 | 1.00 | 0.68 | 0.67 | 0.72 | 0.68 |
| | 2000 | Bias | −0.31 | 0.00 | 0.00 | −0.02 | 0.00 | −0.06 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.04 | 0.10 | 0.11 | 0.10 | 0.11 | 0.06 | 0.08 | 0.08 | 0.08 | 0.08 |
| | | MSE-R | 1.00 | 9.12 | 8.45 | 10.20 | 8.95 | 1.00 | 1.06 | 1.04 | 1.10 | 1.05 |
| | 9000 | Bias | −0.31 | 0.00 | 0.00 | −0.02 | 0.00 | −0.06 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.02 | 0.05 | 0.05 | 0.05 | 0.05 | 0.03 | 0.04 | 0.04 | 0.04 | 0.04 |
| | | MSE-R | 1.00 | 38.37 | 37.23 | 37.43 | 38.31 | 1.00 | 2.74 | 2.71 | 2.37 | 2.74 |
| 0.7 | 1000 | Bias | −0.34 | 0.00 | 0.01 | −0.02 | 0.00 | 0.01 | 0.01 | 0.01 | −0.01 | 0.01 |
| | | SD | 0.07 | 0.19 | 0.21 | 0.18 | 0.19 | 0.10 | 0.16 | 0.17 | 0.16 | 0.16 |
| | | MSE-R | 1.00 | 3.21 | 2.85 | 3.66 | 3.20 | 1.00 | 0.36 | 0.34 | 0.38 | 0.35 |
| | 2000 | Bias | −0.35 | −0.02 | −0.01 | −0.04 | −0.02 | 0.01 | 0.01 | 0.01 | −0.01 | 0.01 |
| | | SD | 0.05 | 0.14 | 0.15 | 0.13 | 0.14 | 0.07 | 0.12 | 0.12 | 0.11 | 0.12 |
| | | MSE-R | 1.00 | 6.08 | 5.74 | 6.52 | 6.02 | 1.00 | 0.33 | 0.33 | 0.36 | 0.33 |
| | 9000 | Bias | −0.34 | 0.00 | 0.00 | −0.02 | 0.00 | 0.01 | 0.00 | 0.00 | −0.02 | 0.00 |
| | | SD | 0.02 | 0.06 | 0.07 | 0.06 | 0.06 | 0.03 | 0.06 | 0.06 | 0.05 | 0.06 |
| | | MSE-R | 1.00 | 27.92 | 26.24 | 27.89 | 27.89 | 1.00 | 0.34 | 0.33 | 0.32 | 0.34 |

# E OPTIMIZE-HF supplement

Table 7: *Covariates included in Model 2 and regarded as error-free*

| |
|---|
| age |
| black race |
| heart rate |
| sodium level |
| serum creatinine level |
| hemoglobin level |
| primary cause of hospital admission |
| prior cerebrovascular accident or transient ischemic attack |
| hyperlipidemia |
| hypertension |
| liver disease |
| smoker within past year |
| chronic obstructive pulmonary disease |
| peripheral vascular disease |
| known heart failure prior to this admission |
| rales |
| LVSD |

We have implemented an analysis that regards these covariates as error free. We focus on the error in systolic and diastolic blood pressure that is known to be substantial. Other variables in this list could also have measurement error, including lab values like sodium, serum creatinine and hemoglobin. One could easily incorporate these variables into multivariate MAI if information about the magnitude of error were available. This may not be necessary if the amount of error is small.