eAppendix 1: Detailed simulation study metrics and additional results

Let S be the total number of simulation iterations, $s = 1, \dots, S$; in our case $S = 1{,}000$. For the model without batch indicator terms, empirical bias and MSE are defined as follows:

$$\widehat{Bias}\left(\hat{\beta}_E^{NBI}\right) = \frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_E^{NBI(s)} - \beta_E\right) \text{ and } \widehat{MSE}\left(\hat{\beta}_E^{NBI}\right) = \frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_E^{NBI(s)} - \beta_E\right)^2.$$

For the model with batch indicator terms, empirical bias and MSE are defined as follows:

$$\widehat{Bias}\left(\hat{\beta}_E^{WBI}\right) = \frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_E^{WBI(s)} - \beta_E\right) \text{ and } \widehat{MSE}\left(\hat{\beta}_E^{WBI}\right) = \frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_E^{WBI(s)} - \beta_E\right)^2.$$

To compare the relative magnitude of the bias to the true regression parameter β_E, we will present the relative bias, defined as:

$$\widehat{RBias}\left(\hat{\beta}_E^{NBI}\right) = 100 \times \frac{\widehat{Bias}\left(\hat{\beta}_E^{NBI}\right)}{\beta_E} \text{ and } \widehat{RBias}\left(\hat{\beta}_E^{WBI}\right) = 100 \times \frac{\widehat{Bias}\left(\hat{\beta}_E^{WBI}\right)}{\beta_E}.$$

To compute coverage probability, we first need to establish how confidence intervals for each method are constructed. For CCA and $LOD/\sqrt{2}$, the s-th 95% confidence interval for β_E is given by $\hat{\beta}_E^{NBI(s)} \pm 1.96 \times SE\left(\hat{\beta}_E^{NBI(s)}\right)$ or $\hat{\beta}_E^{WBI(s)} \pm 1.96 \times SE\left(\hat{\beta}_E^{WBI(s)}\right)$ depending on batch indicator inclusion. For MICE and CLMI the confidence interval for β_E is given by $\hat{\beta}_E^{NBI(s)} \pm t_{0.975,\tilde{\nu}_m} \times SE\left(\hat{\beta}_E^{NBI(s)}\right)$ or $\hat{\beta}_E^{WBI(s)} \pm t_{0.975,\tilde{\nu}_m} \times SE\left(\hat{\beta}_E^{WBI(s)}\right)$, where $\hat{\beta}_E^{NBI(s)}$, $SE\left(\hat{\beta}_E^{NBI(s)}\right)$, $\hat{\beta}_E^{WBI(s)}$, $SE\left(\hat{\beta}_E^{WBI(s)}\right)$, and $\tilde{\nu}_m$ are defined in Barnard et al.[25] and Rubin.[27] Let $CI\left(\hat{\beta}_E^{NBI(s)}\right)$ and $CI\left(\hat{\beta}_E^{WBI(s)}\right)$ denote the confidence interval for $\hat{\beta}_E^{NBI(s)}$ and $\hat{\beta}_E^{WBI(s)}$, respectively. Then the coverage probability is defined as:

$$\widehat{CP}\left(\hat{\beta}_E^{NBI}\right) = \frac{1}{S}\sum_{s=1}^{S} I\left(\beta_E \in CI\left(\hat{\beta}_E^{NBI(s)}\right)\right) \text{ and } \widehat{CP}\left(\hat{\beta}_E^{WBI}\right) = \frac{1}{S}\sum_{s=1}^{S} I\left(\beta_E \in CI\left(\hat{\beta}_E^{WBI(s)}\right)\right),$$

for the model without and with batch indicator terms, respectively.

eAppendix 1.1: Simulation results for single batch and multiple batches

eTable 1. Simulation Results for a Single Batch with True $\beta_E = \log(1.5)$.

Method	Simulation Metrics							
	Cohort (N = 1,000)				Cohort (N = 5,000)			
	$Bias(\hat{\beta}_E)$	$SE(\hat{\beta}_E)$	MSE	CP	$Bias(\hat{\beta}_E)$	$SE(\hat{\beta}_E)$	MSE	CP
15% Below the LOD								
CCA	0.007	0.110	0.012	0.960	0.006	0.049	0.002	0.955
$LOD/\sqrt{2}$	0.010	0.099	0.009	0.959	0.011	0.044	0.002	0.946
MICE	0.005	0.111	0.012	0.950	0.004	0.049	0.002	0.956
CLMI	0.001	0.098	0.009	0.957	0.003	0.043	0.002	0.954
30% Below the LOD								
CCA	0.004	0.125	0.015	0.964	0.004	0.055	0.003	0.958
$LOD/\sqrt{2}$	0.026	0.104	0.011	0.960	0.028	0.046	0.003	0.910
MICE	0.002	0.128	0.017	0.951	0.000	0.056	0.003	0.949
CLMI	0.002	0.100	0.009	0.963	0.003	0.044	0.002	0.950
60% Below the LOD								
CCA	0.013	0.179	0.030	0.964	0.011	0.078	0.006	0.955
$LOD/\sqrt{2}$	0.077	0.126	0.021	0.912	0.079	0.056	0.009	0.703
MICE	0.028	0.200	0.041	0.943	0.011	0.088	0.009	0.927
CLMI	0.003	0.109	0.012	0.953	0.004	0.048	0.002	0.958

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; CP, coverage probability; LOD, limit of detection; MICE, multiple imputation using chained equations; MSE, mean-squared error; SE, standard error; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

eTable 2. Empirical Coverage Probabilities of 95% Confidence Intervals with N = 5,000.

LOD Info	Empirical Coverage Probability							
	With Batch Indicator ($\hat{\beta}_E^{WBI}$)				Without Batch Indicator ($\hat{\beta}_E^{NBI}$)			
	CCA	$LOD/\sqrt{2}$	MICE	CLMI	CCA	$LOD/\sqrt{2}$	MICE	CLMI
Random [a]								
(15, 15) [b]	0.954	0.946	0.955	0.954	0.955	0.946	0.956	0.954
(15, 30)	0.956	0.931	0.952	0.954	0.959	0.932	0.954	0.955
(15, 60)	0.952	0.900	0.963	0.956	0.960	0.925	0.885	0.956
(30, 15)	0.957	0.930	0.959	0.951	0.960	0.935	0.956	0.951
(30, 30)	0.958	0.908	0.948	0.952	0.958	0.910	0.949	0.950
(30, 60)	0.957	0.859	0.953	0.949	0.955	0.880	0.910	0.950
(60, 15)	0.955	0.904	0.965	0.955	0.952	0.923	0.899	0.956
(60, 30)	0.953	0.860	0.950	0.956	0.952	0.884	0.924	0.954
(60, 60)	0.957	0.700	0.927	0.956	0.955	0.703	0.927	0.958
Outcome-Dependent [c]								
(15, 15)	0.961	0.944	0.959	0.948	0.953	0.946	0.955	0.953
(15, 30)	0.951	0.938	0.958	0.948	0.878	0.945	0.835	0.953
(15, 60)	0.962	0.932	0.964	0.951	0.239	0.817	0.014	0.954
(30, 15)	0.954	0.919	0.955	0.948	0.795	0.903	0.822	0.954
(30, 30)	0.957	0.909	0.956	0.954	0.949	0.907	0.950	0.952
(30, 60)	0.958	0.895	0.936	0.959	0.459	0.913	0.241	0.957
(60, 15)	0.956	0.830	0.977	0.962	0.054	0.382	0.524	0.962
(60, 30)	0.956	0.794	0.964	0.952	0.182	0.359	0.519	0.952
(60, 60)	0.956	0.704	0.960	0.953	0.927	0.611	0.937	0.955
Covariate-Dependent [d]								
(15, 15)	0.960	0.946	0.957	0.953	0.960	0.949	0.955	0.953
(15, 30)	0.958	0.931	0.963	0.947	0.963	0.934	0.968	0.946
(15, 60)	0.959	0.890	0.976	0.962	0.963	0.920	0.918	0.962
(30, 15)	0.965	0.928	0.959	0.957	0.963	0.929	0.956	0.956
(30, 30)	0.959	0.912	0.961	0.949	0.960	0.911	0.957	0.951
(30, 60)	0.963	0.841	0.973	0.958	0.962	0.881	0.939	0.957
(60, 15)	0.958	0.903	0.961	0.959	0.954	0.914	0.932	0.960
(60, 30)	0.943	0.869	0.944	0.949	0.944	0.880	0.948	0.950
(60, 60)	0.948	0.694	0.964	0.955	0.949	0.704	0.962	0.956

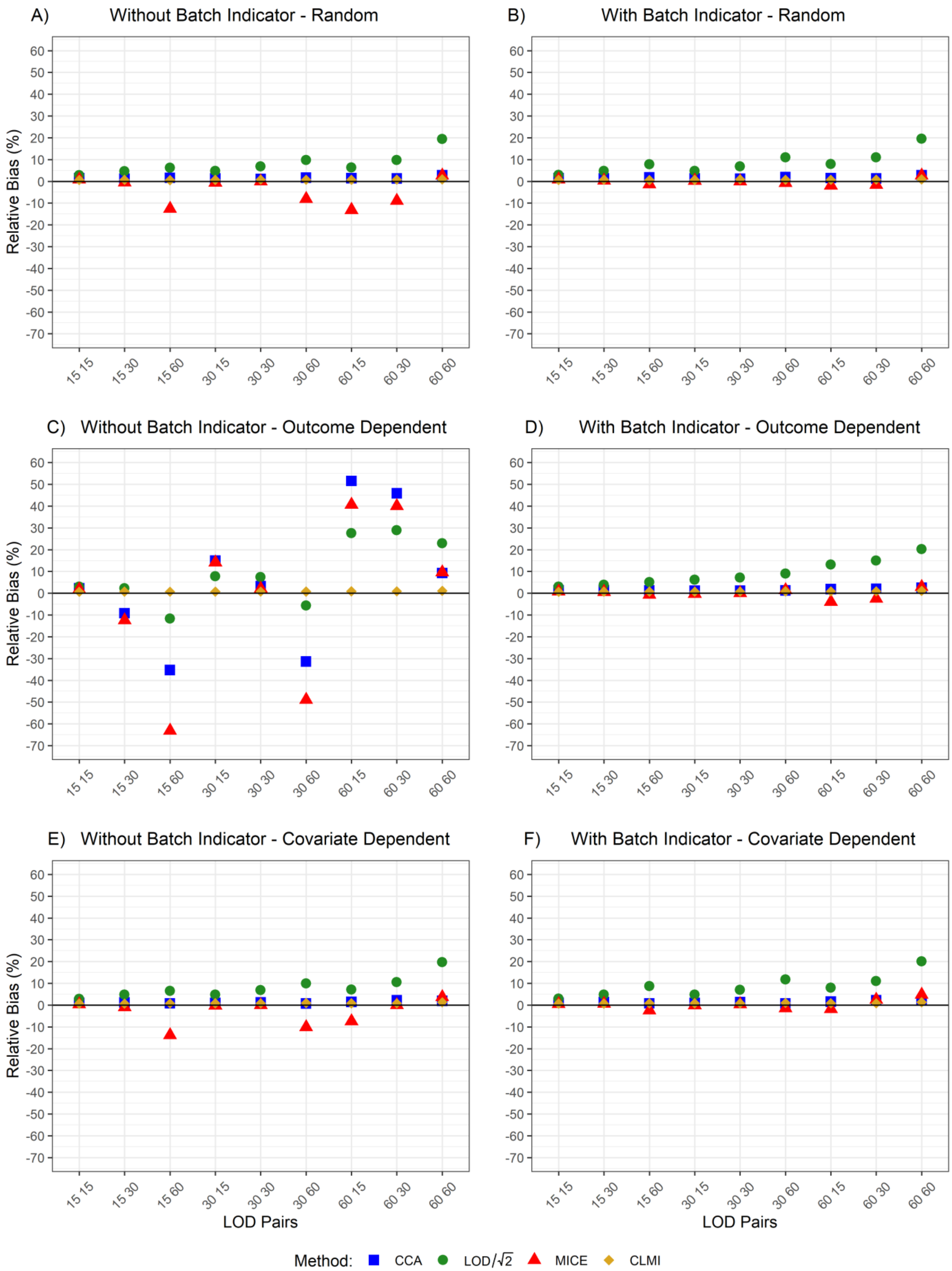
Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

[a] "Random" refers to random batch assignment.

[b] The notation (A, B) means that approximately A% of observations in batch 1 were below the batch 1 LOD and approximately B% of observations in batch 2 were below the batch 2 LOD.

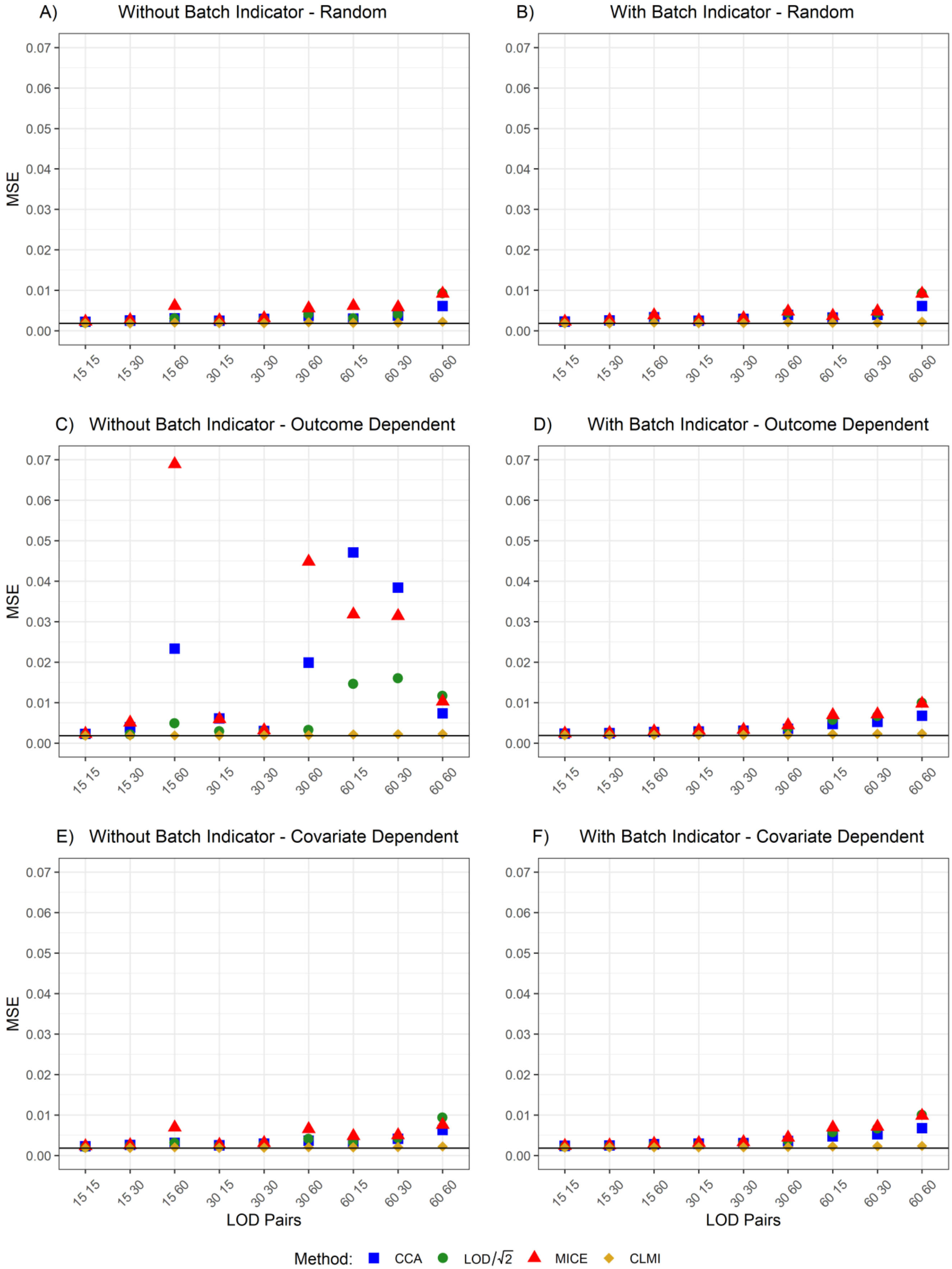
[c] "Outcome-Dependent" refers to batch assignment that depends on Y.

[d] "Covariate-Dependent" refers to batch assignment that depends on S and G.


A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
Relative Bias (%)
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

eFigure 1a. Relative bias at various LOD combinations in a simulated, large cohort study (N = 5,000). The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A is a scenario in which batch is randomly assigned and a batch indicator term is not included in the analysis model, panel B is a scenario in which batch is randomly assigned and a batch indicator term is included in the analysis model, panel C is a scenario in which batch depends on Y and a batch indicator term is not included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel D is a scenario in which batch depends on Y and a batch indicator term is included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel E is a scenario in which batch depends on S and G and a batch indicator term is not included in the analysis model, and panel F is a scenario in which batch depends on S and G and a batch indicator term is included in the analysis model. The bolded black line indicates a relative bias of 0% (true $\beta_E = \log(1.5)$).

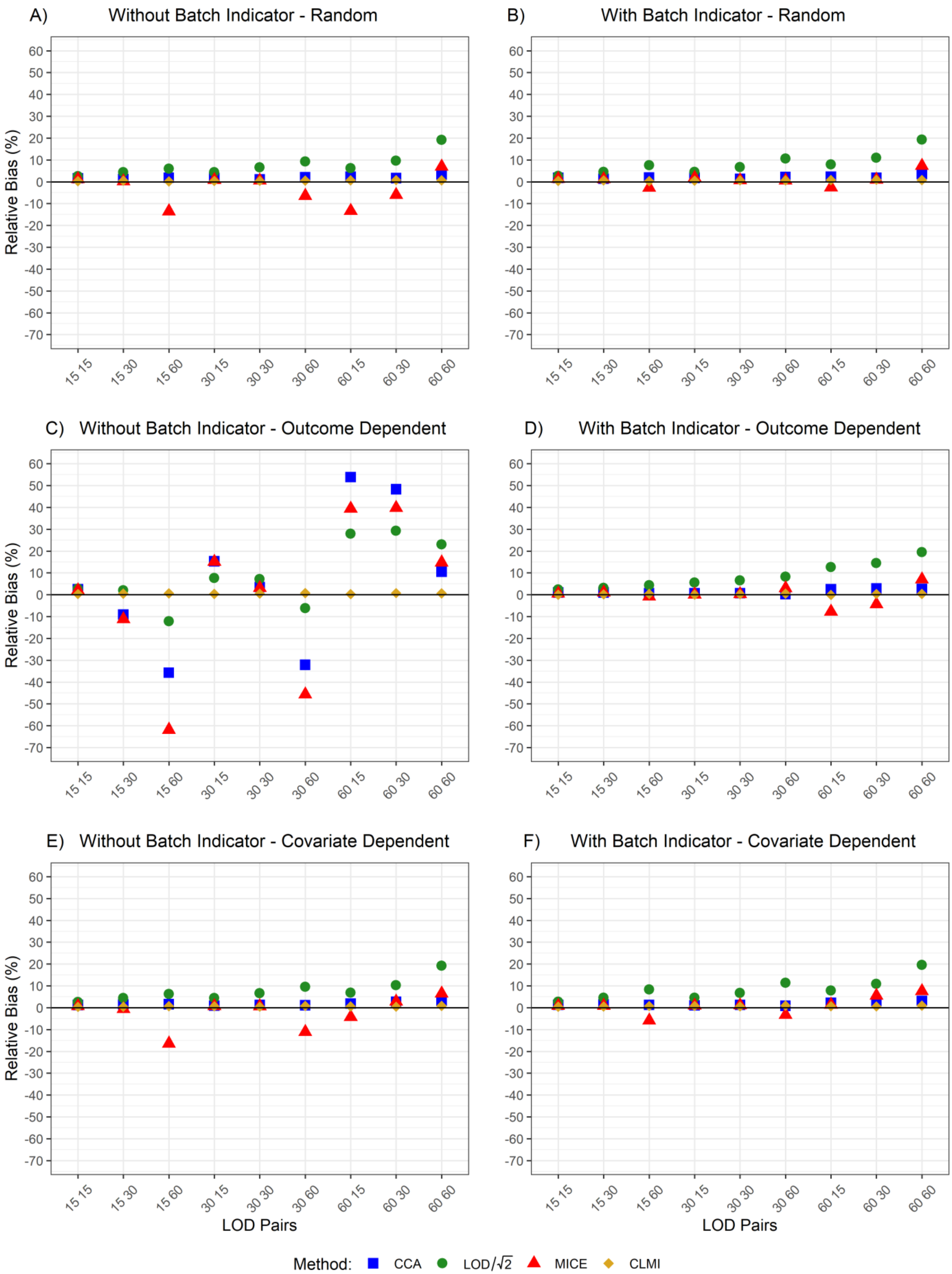
Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.



A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
MSE
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

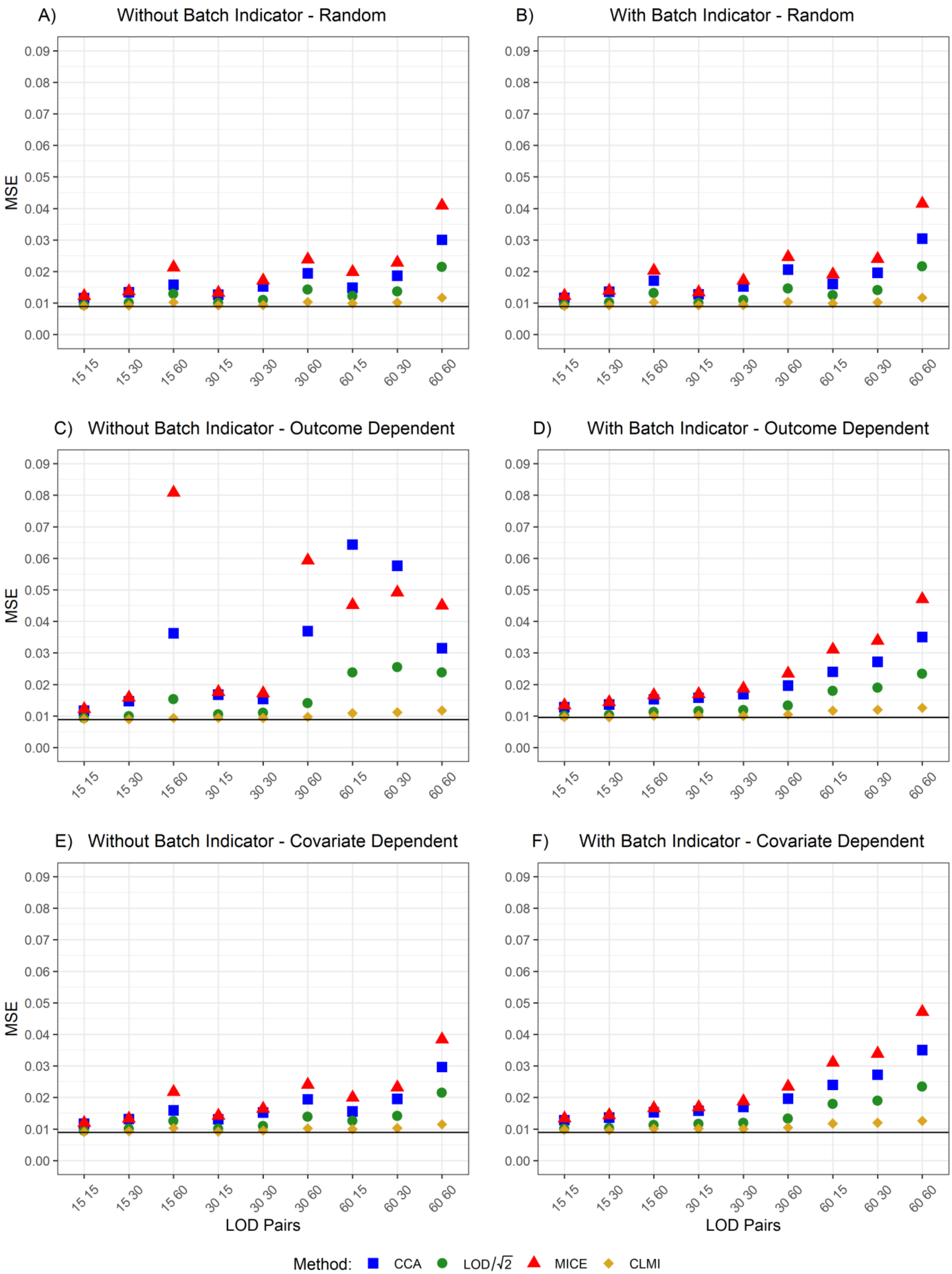
eFigure 1b. Mean-Squared Error (MSE) at various LOD combinations in a simulated, large cohort study (N = 5,000). The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A is a scenario in which batch is randomly assigned and a batch indicator term is not included in the analysis model, panel B is a scenario in which batch is randomly assigned and a batch indicator term is included in the analysis model, panel C is a scenario in which batch depends on Y and a batch indicator term is not included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel D is a scenario in which batch depends on Y and a batch indicator term is included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel E is a scenario in which batch depends on S and G and a batch indicator term is not included in the analysis model, and panel F is a scenario in which batch depends on S and G and a batch indicator term is included in the analysis model. The bolded black line indicates the gold standard MSE (no observations subject to censoring).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.


A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
Relative Bias (%)
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

eFigure 1c. Relative bias at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000). The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A is a scenario in which batch is randomly assigned and a batch indicator term is not included in the analysis model, panel B is a scenario in which batch is randomly assigned and a batch indicator term is included in the analysis model, panel C is a scenario in which batch depends on Y and a batch indicator term is not included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel D is a scenario in which batch depends on Y and a batch indicator term is included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel E is a scenario in which batch depends on S and G and a batch indicator term is not included in the analysis model, and panel F is a scenario in which batch depends on S and G and a batch indicator term is included in the analysis model. The bolded black line indicates a relative bias of 0% (true $\beta_E = \log(1.5)$).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

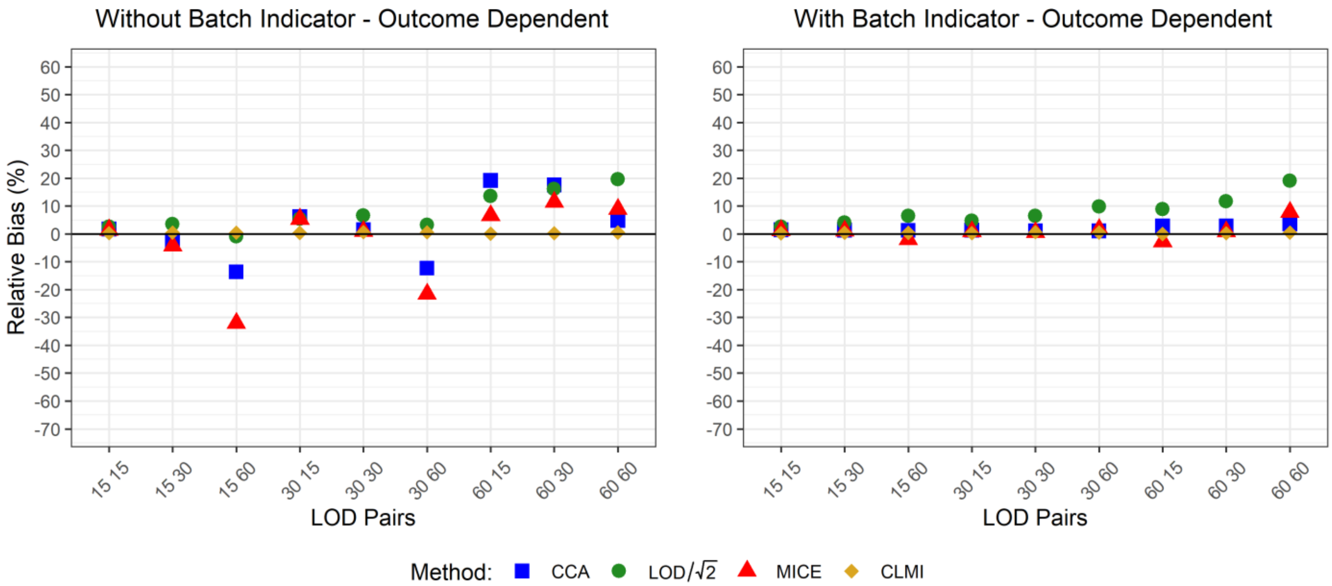


A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
MSE
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

eFigure 1d. Mean-Squared Error (MSE) at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000). The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A is a scenario in which batch is randomly assigned and a batch indicator term is not included in the analysis model, panel B is a scenario in which batch is randomly assigned and a batch indicator term is included in the analysis model, panel C is a scenario in which batch depends on Y and a batch indicator term is not included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel D is a scenario in which batch depends on Y and a batch indicator term is included in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), panel E is a scenario in which batch depends on S and G and a batch indicator term is not included in the analysis model, and panel F is a scenario in which batch depends on S and G and a batch indicator term is included in the analysis model. The bolded black line indicates the gold standard MSE (no observations subject to censoring).

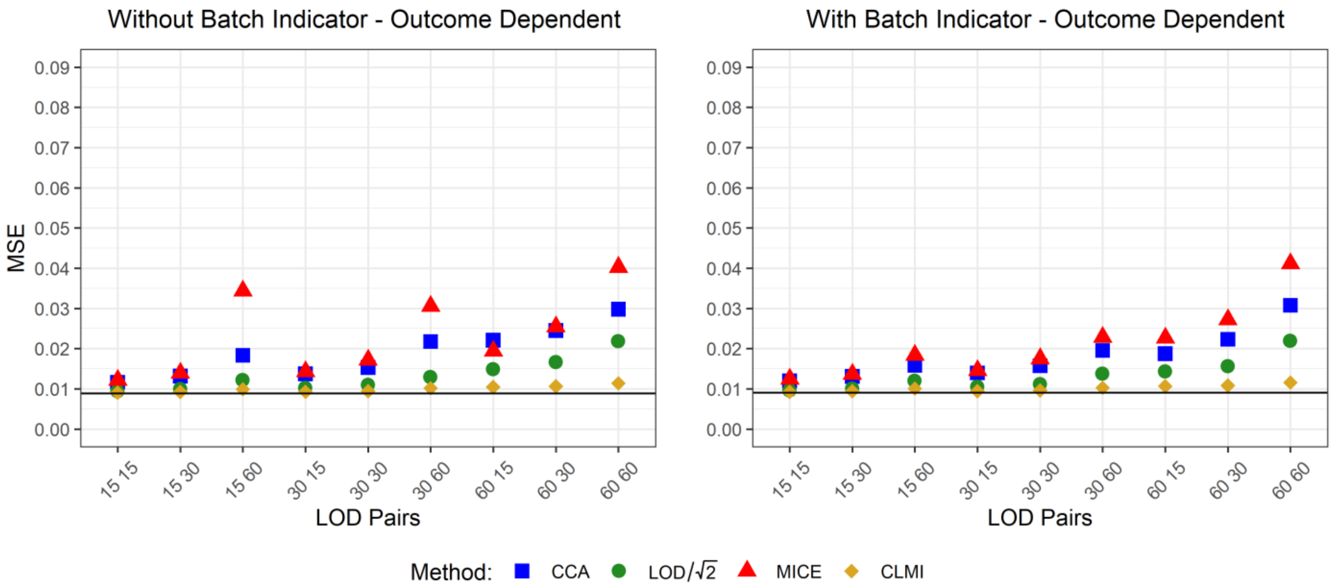
Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

eAppendix 1.2: Simulation results for the outcome-dependent batch assignment sensitivity check



eFigure 2a. Relative bias at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000) when $P(Batch\ 1 \mid Y_i = 1) = 0.6$. The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. The panel on the left is a scenario in which batch depends on Y and a batch indicator term is not included in the analysis model and the panel on the right is a scenario in which batch depends on Y and a batch indicator term is included in the analysis model. The bolded black line indicates a relative bias of 0% (true $\beta_E =$ log(1.5)).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.



eFigure 2b. Mean-Squared Error (MSE) at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000) when $P(Batch\ 1 \mid Y_i = 1) = 0.6$. The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. The panel on the left is a scenario in which batch depends on Y and a batch indicator term is not included in the analysis model and the panel on the right is a scenario in which batch depends on Y and a batch indicator term is included in the analysis model. The bolded black line indicates the gold standard MSE (no observations subject to censoring).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

eAppendix 1.3: PROTECT exploratory analysis results

eTable 3. PROTECT Summary Statistics Stratified by Spontaneous Preterm / Full-term Delivery at Visit 1 and Visit 2. Categorical variables are reported as N (%) and continuous variables are reported as mean (standard deviation).

Variable	Visit 1		Visit 2	
	SPD (N = 43)	FTD (N = 583)	SPD (N = 42)	FTD (N = 598)
Maternal Age (years)	27.7 (5.8)	26.7 (5.3)	28.2 (5.5)	26.9 (5.6)
Specific Gravity	1.020 (0.005)	1.019 (0.006)	1.018 (0.005)	1.019 (0.006)
Parity				
0	12 (27.9)	291 (49.9)	11 (26.2)	293 (49.0)
1	20 (46.5)	229 (39.3)	20 (47.6)	237 (39.6)
≥ 2	11 (25.6)	63 (10.8)	11 (26.2)	68 (11.4)
Employed				
Yes	20 (46.5)	370 (63.5)	19 (45.2)	372 (62.2)
No	23 (53.5)	213 (36.5)	23 (54.8)	226 (37.8)

Abbreviations: FTD, full-term delivery; SPD, spontaneous preterm delivery.

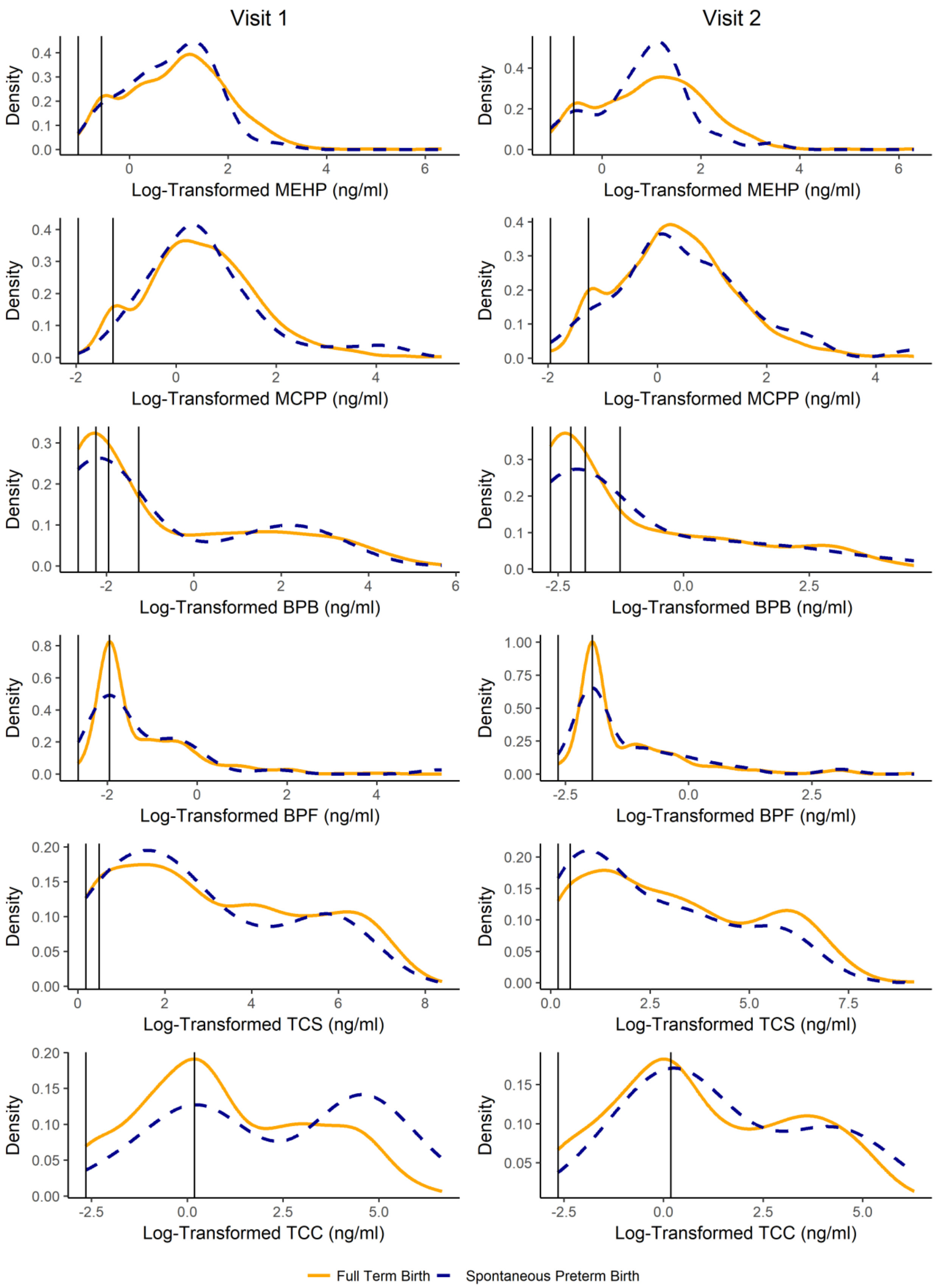
eTable 4. PROTECT LOD Summary Statistics at Visit 1 and Visit 2.

Contaminant	LOD [a]	Visit 1 N (%)	Visit 2 N (%)
MEHP	Above [b]	544 (86.9)	548 (85.6)
	<0.5	10 (1.6)	20 (3.1)
	<0.8	72 (11.5)	72 (11.3)
MCPP	Above	570 (91.1)	571 (89.2)
	<0.2	3 (0.5)	7 (1.1)
	<0.4	53 (8.5)	62 (9.7)
BPB	Above	392 (62.6)	363 (56.7)
	<0.1	146 (23.3)	173 (27.0)
	<0.15	4 (0.6)	5 (0.8)
	<0.2	83 (13.3)	98 (15.3)
	<0.4	1 (0.2)	1 (0.2)
BPF	Above	182 (47.2)	172 (45.0)
	<0.1	7 (1.8)	11 (2.9)
	<0.2	197 (51.0)	199 (52.1)
TCS	Above	545 (87.1)	555 (86.7)
	<1.7	52 (8.3)	62 (9.7)
	<2.3	29 (4.6)	23 (3.6)
TCC	Above	342 (88.6)	334 (87.4)
	<0.1	22 (5.7)	22 (5.8)
	<1.7	22 (5.7)	26 (6.8)

Abbreviations: BPB, butylparaben; BPF, bisphenol F; LOD, limit of detection; MCPP, mono-(3-carboxypropyl) phthalate; MEHP, mono-(2-ethylhexyl) phthalate; TCC, triclocarban; TCS, triclosan.

[a] Units are ng/ml.

[b] "Above" refers to above the limit of detection.


Visit 1
Visit 2
Density
Log-Transformed MEHP (ng/ml)
Log-Transformed MCPP (ng/ml)
Log-Transformed BPB (ng/ml)
Log-Transformed BPF (ng/ml)
Log-Transformed TCS (ng/ml)
Log-Transformed TCC (ng/ml)
Full Term Birth
Spontaneous Preterm Birth

eFigure 3. Kernel density estimates of log-transformed contaminant concentrations stratified by preterm and full-term delivery. The bolded vertical lines indicate the distinct LODs divided by $\sqrt{2}$. For the purposes of this figure, values below the LOD were substituted with $LOD_j/\sqrt{2}$ prior to log-transformation.

Abbreviations: BPB, butylparaben; BPF, bisphenol F; LOD, limit of detection; MCPP, mono-(3-carboxypropyl) phthalate; MEHP, mono-(2-ethylhexyl) phthalate; TCC, triclocarban; TCS, triclosan.

eAppendix 2: Simulation study to explore robustness to the number of batches

To test the effect of many batches we consider the large cohort simulation setting (N = 5,000) with random batch assignment where instead of having two batches we have five evenly sized batches. In one setting the percent below LOD for each batch is $P_1 = 10$, $P_2 = 15$, $P_3 = 20$, $P_4 = 25$, and $P_5 = 30$ and for the second setting we have that $P_1 = 10$, $P_2 = 15$, $P_3 = 35$, $P_4 = 45$, and $P_5 = 60$. eTable 5 contains the results for the five batch scenario compared with the two batch simulation setting (eFigure 1a, eFigure 1b, and eTable 2). The relative bias, MSE, and coverage probability are not directly comparable between the two batch and five batch simulations, however we see the same pattern. CLMI has the smallest relative bias and MSE, CCA has small relative bias regardless of the inclusion of a batch indicator, MICE has small relative bias only when a batch indicator is included in the analysis model, and $LOD/\sqrt{2}$ has the largest relative bias across most settings (although $LOD/\sqrt{2}$ usually has the second smallest MSE). In terms of coverage probability, CLMI and CCA have proper coverage, MICE has proper coverage only when a batch indicator is included in the analysis model, and $LOD/\sqrt{2}$ does not attain the nominal coverage.

One important point to note is that as the number of batches increase, so does the number of batch indicator regression parameters. Consequently, for a fixed sample size, methods like CCA become less efficient when there are more batch indicator regression parameters to estimate. However, CLMI does not require that we include batch indicators as covariates in the analysis model; batch information is used to construct the observed likelihood. Therefore, the impact of increasing the number of batches has less of an effect on the efficiency of CLMI.

eTable 5. Simulation Results for Large Cohort Simulation Setting (N = 5,000) with Five Batches and Two Batches* (Random Batch Assignment).

			With Batch Indicator			Without Batch Indicator		
LODs	Batches	Method	$\widehat{RBias}(\hat{\beta}_E^{WBI})$	$\widehat{MSE}(\hat{\beta}_E^{WBI})$	$\widehat{CP}(\hat{\beta}_E^{WBI})$	$\widehat{RBias}(\hat{\beta}_E^{NBI})$	$\widehat{MSE}(\hat{\beta}_E^{NBI})$	$\widehat{CP}(\hat{\beta}_E^{NBI})$
(10, 15, 20, 25, 30)	5	CCA	1.38	0.0024	0.957	1.27	0.0024	0.955
		$LOD/\sqrt{2}$	4.16	0.0023	0.934	4.03	0.0022	0.939
		MICE	0.59	0.0025	0.960	-0.18	0.0024	0.963
		CLMI	0.19	0.0019	0.952	0.09	0.0019	0954
(10, 15, 35, 45, 60)	5	CCA	1.42	0.0031	0.950	1.33	0.0029	0.949
		$LOD/\sqrt{2}$	7.12	0.0031	0.912	6.15	0.0028	0.923
		MICE	-0.41	0.0032	0.972	-7.31	0.0040	0.929
		CLMI	1.39	0.0019	0.955	1.28	0.0019	0.959
(15, 30)	2	CCA	1.27	0.0025	0.956	1.24	0.0025	0.959
		$LOD/\sqrt{2}$	4.68	0.0023	0.932	4.63	0.0023	0.932
		MICE	0.40	0.0027	0.952	-0.48	0.0027	0.954
		CLMI	0.75	0.0019	0.954	0.72	0.0019	0.955
(15, 60)	2	CCA	1.85	0.0033	0.952	1.62	0.0031	0.960
		$LOD/\sqrt{2}$	7.75	0.0034	0.900	6.19	0.0030	0.925
		MICE	-1.33	0.0039	0.963	-12.62	0.0062	0.885
		CLMI	0.57	0.0020	0.956	0.55	0.0020	0.956

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

* Relative biases for the two batch scenario are reported in eFigure 1a, mean-squared errors (MSE) are reported in eFigure 1b, and coverage probabilities (CP) are reported in eTable 2.

eAppendix 3: Simulation study to explore the impact of misspecified exposure distribution

For the misspecification simulations, most simulation parameters are identical to the simulations in the main paper. The only changes are related the distribution of the residual error term. In one scenario, we consider the residual errors following a shifted gamma distribution, i.e.,

$$\{X_i|S_i, G_i\} \sim -0.5 + 1.25 \cdot S_i + 1.25 \cdot G_i + \epsilon_i + c$$

where $\epsilon_i \sim Gamma(5.3, 2)$ and $c = -5.3/2$. In the other scenario the residual errors follow a two-component mixture of normal distributions, i.e.,

$$\{X_i|S_i, G_i\} \sim\ -0.5 + 1.25 \cdot S_i + 1.25 \cdot G_i + \epsilon_i^*$$

where $\epsilon_i^* \sim 0.3 \times\ N(1.5, 0.6^2) + 0.7N(-0.65, 0.6^2)$.

The parameters of the gamma and the mixture of normal distributions were selected so that they had approximately the same mean and variance as $N(0, 1.15^2)$ (to match the distribution of the residual error term in the correctly specified simulations). See eFigure 4 for a comparison between the residual error distributions and to visualize the nature of departure from normality. eFigure 5a, eFigure 5b, and eTable 6 present the relative bias, MSE, and coverage probability for the parameters of interest under the gamma specification. Similarly, eFigure 6a, eFigure 6b, and eTable 7 present the relative bias, MSE, and coverage probability under the mixture of normal misspecification. For both types of misspecification, CCA generally has the smallest relative bias when a batch indicator is included in the analysis model. However, the absolute relative bias for CLMI is less than 3% for gamma misspecification and less than 5% for mixture normal misspecification when $P_1, P_2 \in \{15{,}30\}$. CLMI still has the smallest MSE of the four methods across all LOD pairs and maintains good coverage probability when $P_1, P_2 \in \{15{,}30\}$. When $P_1 = 60$ or $P_2 = 60$, CLMI has absolute relative bias up to 8% and slightly underestimates the coverage probability (for example, when $P_1 = P_2 = 60$ with random batch assignment, $\widehat{CP}\left(\hat{\beta}_E^{WBI}\right) = 0.925$, $\widehat{CP}\left(\hat{\beta}_E^{NBI}\right) = 0.924$; see eTable 7). General trends for MICE and $LOD/\sqrt{2}$ methods are similar to the correctly specified results presented in the manuscript.

eTable 6. Empirical Coverage Probabilities of 95% Confidence Intervals with Misspecification of the Gamma-distributed Exposure (N = 1,000).

LOD Info	Empirical Coverage Probability							
	With Batch Indicator ($\hat{\beta}_E^{WBI}$)				Without Batch Indicator ($\hat{\beta}_E^{NBI}$)			
	CCA	$LOD/\sqrt{2}$	MICE	CLMI	CCA	$LOD/\sqrt{2}$	MICE	CLMI
Random [a]								
(15, 15) [b]	0.952	0.952	0.950	0.950	0.953	0.953	0.950	0.949
(15, 30)	0.955	0.950	0.957	0.946	0.960	0.949	0.959	0.945
(15, 60)	0.957	0.951	0.963	0.952	0.954	0.953	0.938	0.951
(30, 15)	0.948	0.951	0.950	0.948	0.947	0.950	0.948	0.945
(30, 30)	0.954	0.950	0.954	0.945	0.955	0.951	0.953	0.943
(30, 60)	0.949	0.948	0.966	0.950	0.958	0.949	0.956	0.951
(60, 15)	0.956	0.943	0.961	0.944	0.952	0.946	0.935	0.944
(60, 30)	0.951	0.943	0.966	0.943	0.948	0.941	0.949	0.943
(60, 60)	0.952	0.928	0.953	0.931	0.952	0.930	0.955	0.930
Outcome-Dependent [c]								
(15, 15)	0.946	0.951	0.942	0.945	0.952	0.953	0.945	0.949
(15, 30)	0.955	0.951	0.949	0.949	0.933	0.948	0.924	0.947
(15, 60)	0.956	0.949	0.954	0.950	0.780	0.918	0.579	0.947
(30, 15)	0.954	0.952	0.959	0.943	0.934	0.952	0.942	0.946
(30, 30)	0.952	0.950	0.954	0.946	0.958	0.950	0.953	0.944
(30, 60)	0.954	0.946	0.955	0.940	0.852	0.933	0.799	0.944
(60, 15)	0.955	0.936	0.957	0.947	0.704	0.893	0.897	0.945
(60, 30)	0.957	0.934	0.962	0.936	0.788	0.895	0.924	0.939
(60, 60)	0.951	0.924	0.947	0.945	0.958	0.919	0.951	0.941
Covariate-Dependent [d]								
(15, 15)	0.958	0.951	0.953	0.948	0.958	0.950	0.957	0.949
(15, 30)	0.954	0.948	0.957	0.942	0.953	0.950	0.960	0.943
(15, 60)	0.947	0.945	0.960	0.949	0.945	0.946	0.939	0.949
(30, 15)	0.953	0.949	0.957	0.948	0.955	0.949	0.956	0.947
(30, 30)	0.957	0.950	0.954	0.946	0.956	0.948	0.953	0.947
(30, 60)	0.955	0.941	0.963	0.940	0.951	0.943	0.951	0.941
(60, 15)	0.957	0.949	0.952	0.942	0.961	0.949	0.951	0.943
(60, 30)	0.957	0.945	0.966	0.943	0.954	0.950	0.964	0.946
(60, 60)	0.952	0.929	0.959	0.927	0.954	0.929	0.961	0.927

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

[a] "Random" refers to random batch assignment.

[b] The notation (A, B) means that approximately A% of observations in batch 1 were below the batch 1 LOD and approximately B% of observations in batch 2 were below the batch 2 LOD.

[c] "Outcome-Dependent" refers to batch assignment that depends on Y when $P(Batch\ 1 \mid Y_i = 1) = 0.8$.

[d] "Covariate-Dependent" refers to batch assignment that depends on S and G.

eTable 7. Empirical Coverage Probabilities of 95% Confidence Intervals with Misspecification of the Mixture Normal-distributed Exposure (N = 1,000).

LOD Info	Empirical Coverage Probability							
	With Batch Indicator ($\hat{\beta}_E^{WBI}$)				Without Batch Indicator ($\hat{\beta}_E^{NBI}$)			
	CCA	$LOD/\sqrt{2}$	MICE	CLMI	CCA	$LOD/\sqrt{2}$	MICE	CLMI
Random [a]								
(15, 15) [b]	0.948	0.956	0.944	0.946	0.947	0.952	0.942	0.944
(15, 30)	0.955	0.950	0.962	0.950	0.958	0.950	0.961	0.949
(15, 60)	0.955	0.941	0.940	0.946	0.963	0.942	0.897	0.947
(30, 15)	0.957	0.953	0.962	0.942	0.958	0.952	0.961	0.941
(30, 30)	0.964	0.951	0.959	0.940	0.964	0.952	0.956	0.942
(30, 60)	0.965	0.934	0.956	0.938	0.965	0.937	0.946	0.938
(60, 15)	0.954	0.950	0.949	0.949	0.957	0.949	0.902	0.947
(60, 30)	0.953	0.942	0.944	0.945	0.955	0.946	0.941	0.943
(60, 60)	0.953	0.915	0.947	0.925	0.955	0.919	0.946	0.924
Outcome-Dependent [c]								
(15, 15)	0.954	0.955	0.956	0.949	0.949	0.952	0.951	0.944
(15, 30)	0.947	0.955	0.949	0.945	0.934	0.949	0.917	0.947
(15, 60)	0.958	0.952	0.942	0.946	0.681	0.913	0.426	0.945
(30, 15)	0.953	0.952	0.966	0.939	0.938	0.949	0.949	0.940
(30, 30)	0.950	0.954	0.959	0.943	0.963	0.950	0.965	0.941
(30, 60)	0.959	0.949	0.951	0.940	0.765	0.934	0.692	0.935
(60, 15)	0.956	0.932	0.950	0.940	0.659	0.883	0.901	0.934
(60, 30)	0.956	0.934	0.966	0.939	0.737	0.879	0.906	0.933
(60, 60)	0.954	0.919	0.951	0.926	0.953	0.905	0.945	0.930
Covariate-Dependent [d]								
(15, 15)	0.949	0.952	0.946	0.948	0.948	0.953	0.948	0.945
(15, 30)	0.954	0.954	0.958	0.946	0.955	0.951	0.949	0.947
(15, 60)	0.952	0.945	0.950	0.943	0.946	0.949	0.921	0.944
(30, 15)	0.957	0.949	0.954	0.946	0.962	0.950	0.956	0.944
(30, 30)	0.952	0.947	0.953	0.942	0.954	0.950	0.954	0.943
(30, 60)	0.953	0.945	0.946	0.933	0.949	0.949	0.926	0.934
(60, 15)	0.947	0.948	0.948	0.949	0.951	0.947	0.931	0.949
(60, 30)	0.955	0.938	0.948	0.942	0.962	0.937	0.949	0.943
(60, 60)	0.957	0.917	0.949	0.927	0.960	0.919	0.951	0.925

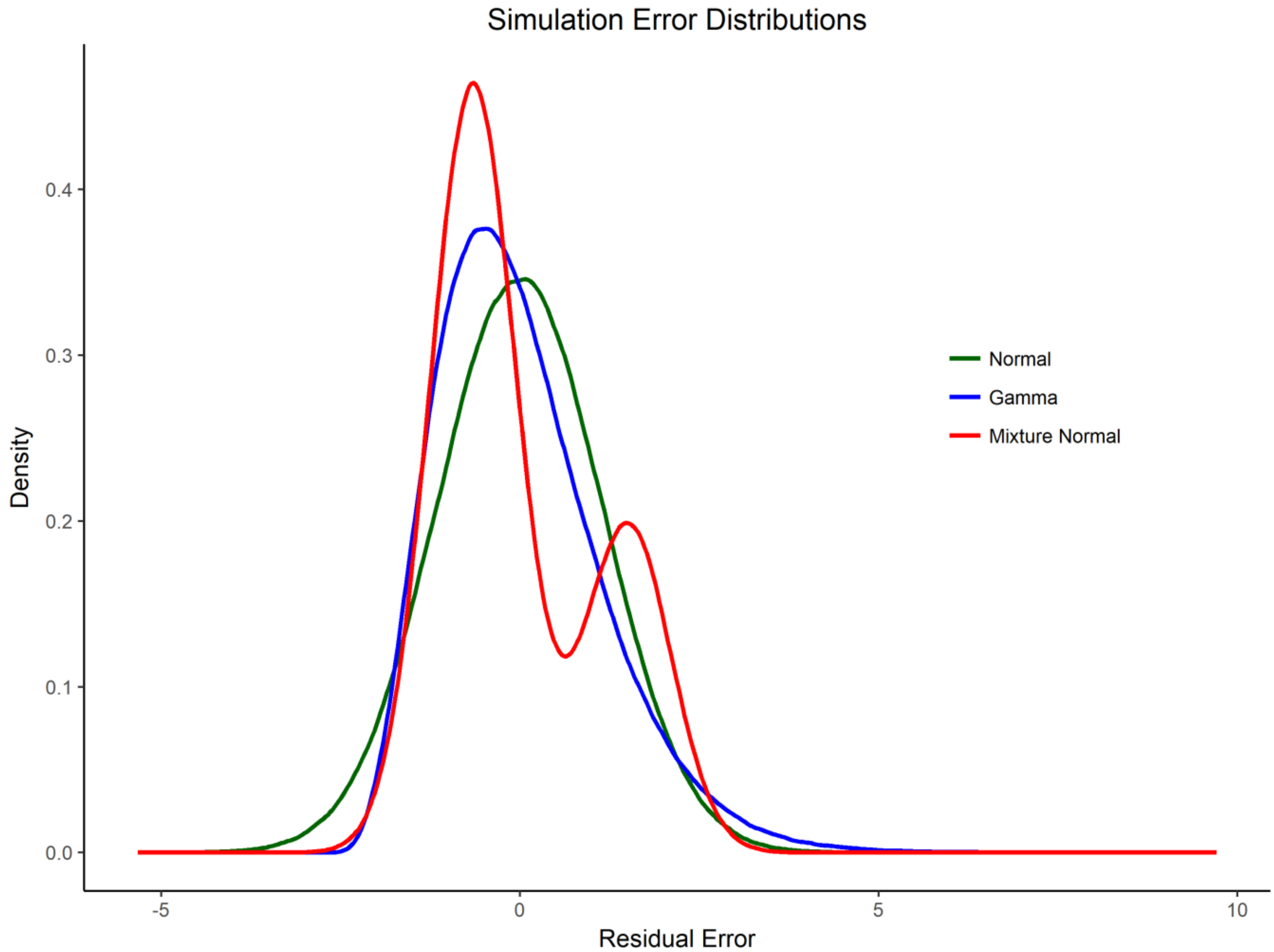
Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.

[a] "Random" refers to random batch assignment.

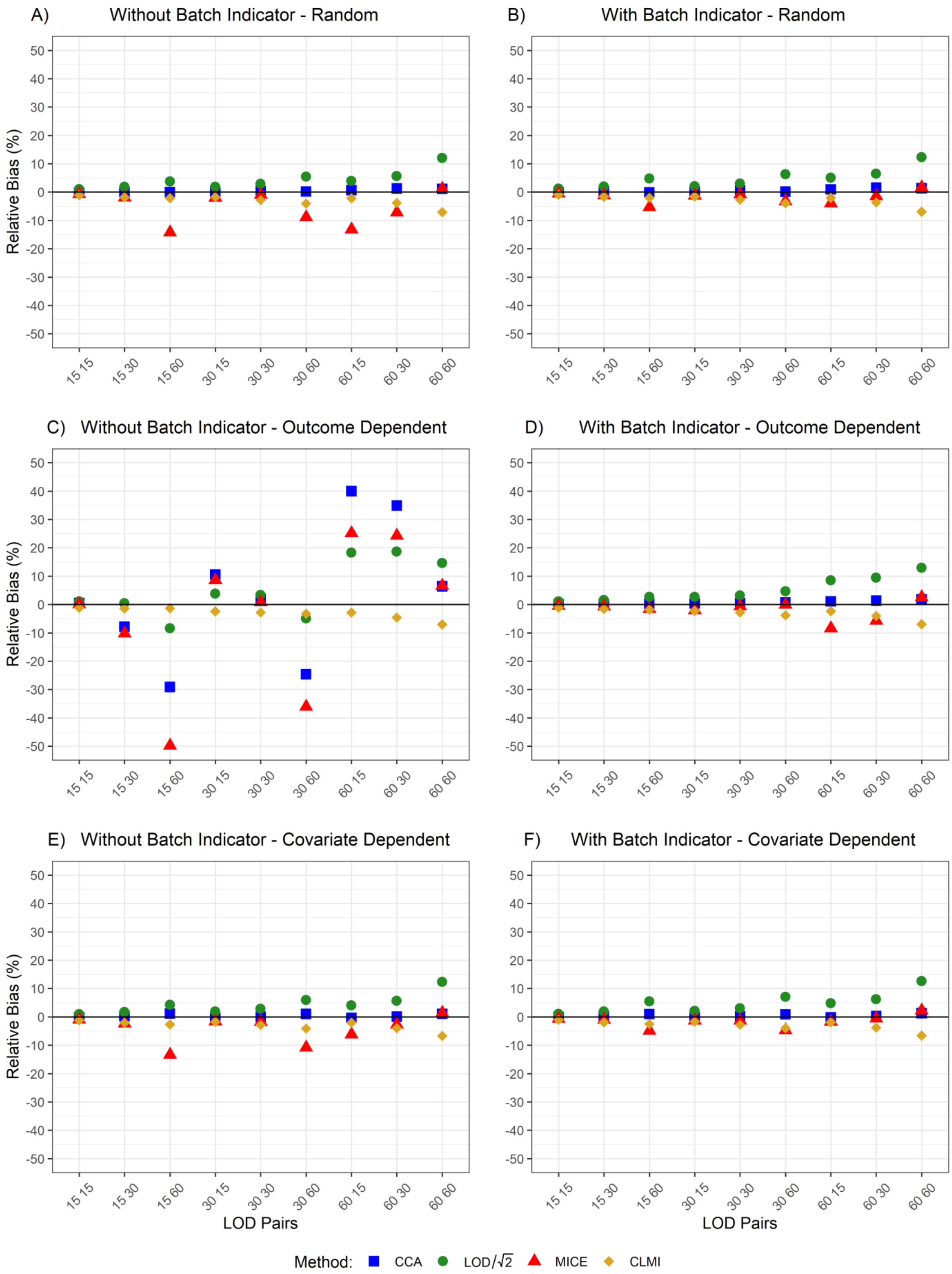
[b] The notation (A, B) means that approximately A% of observations in batch 1 were below the batch 1 LOD and approximately B% of observations in batch 2 were below the batch 2 LOD.

[c] "Outcome-Dependent" refers to batch assignment that depends on Y when $P(Batch\ 1 \mid Y_i = 1) = 0.8$.

[d] "Covariate-Dependent" refers to batch assignment that depends on S and G.



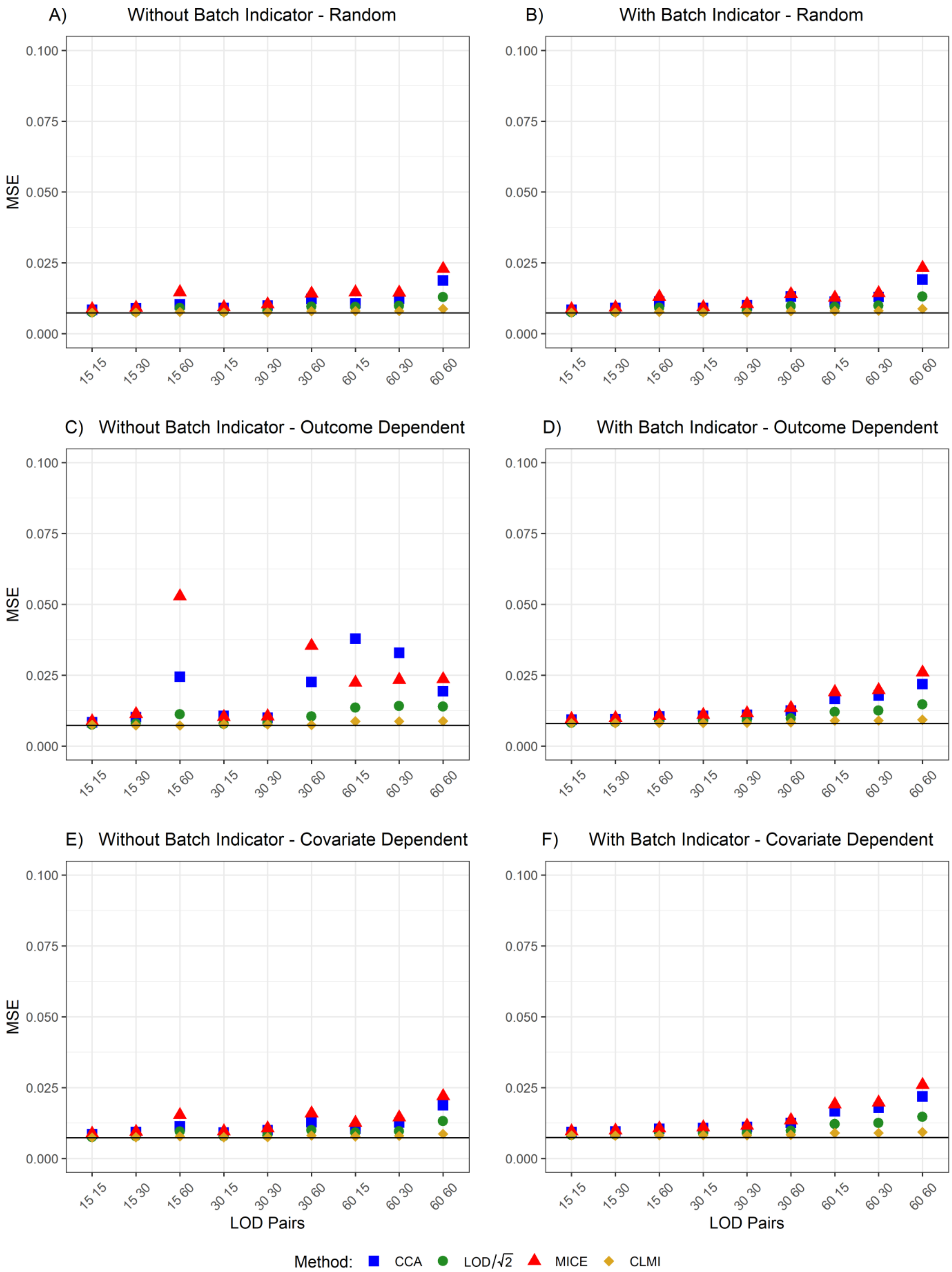
eFigure 4. Comparison of the densities of the residual error distribution for the correctly specified simulations (Normal) and the misspecified simulations (Gamma and Mixture of Normals).



A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
Relative Bias (%)
50
40
30
20
10
0
-10
-20
-30
-40
-50
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

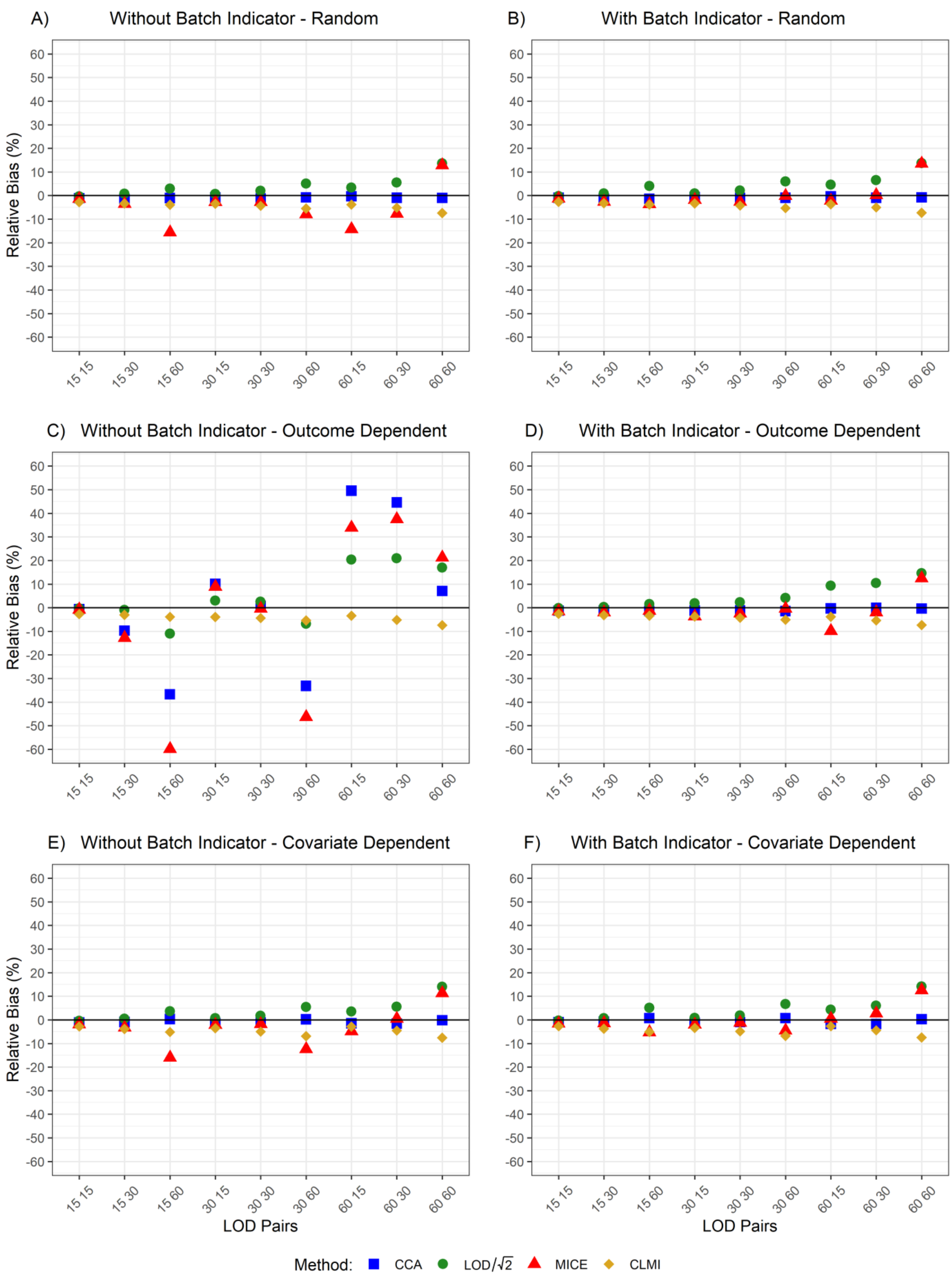
eFigure 5a. Relative bias at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000) with *misspecification of a gamma-distributed exposure*. The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A (B) are scenarios in which batch is randomly assigned and a batch indicator term is not included (included) in the analysis model, panel C (D) are scenarios in which batch depends on Y and a batch indicator term is not included (included) in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), and panel E (F) are scenarios in which batch depends on S and G and a batch indicator term is not included (included) in the analysis model. The bolded black line indicates a relative bias of 0% (true $\beta_E = \log(1.5)$).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.


A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
MSE
0.100
0.075
0.050
0.025
0.000
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

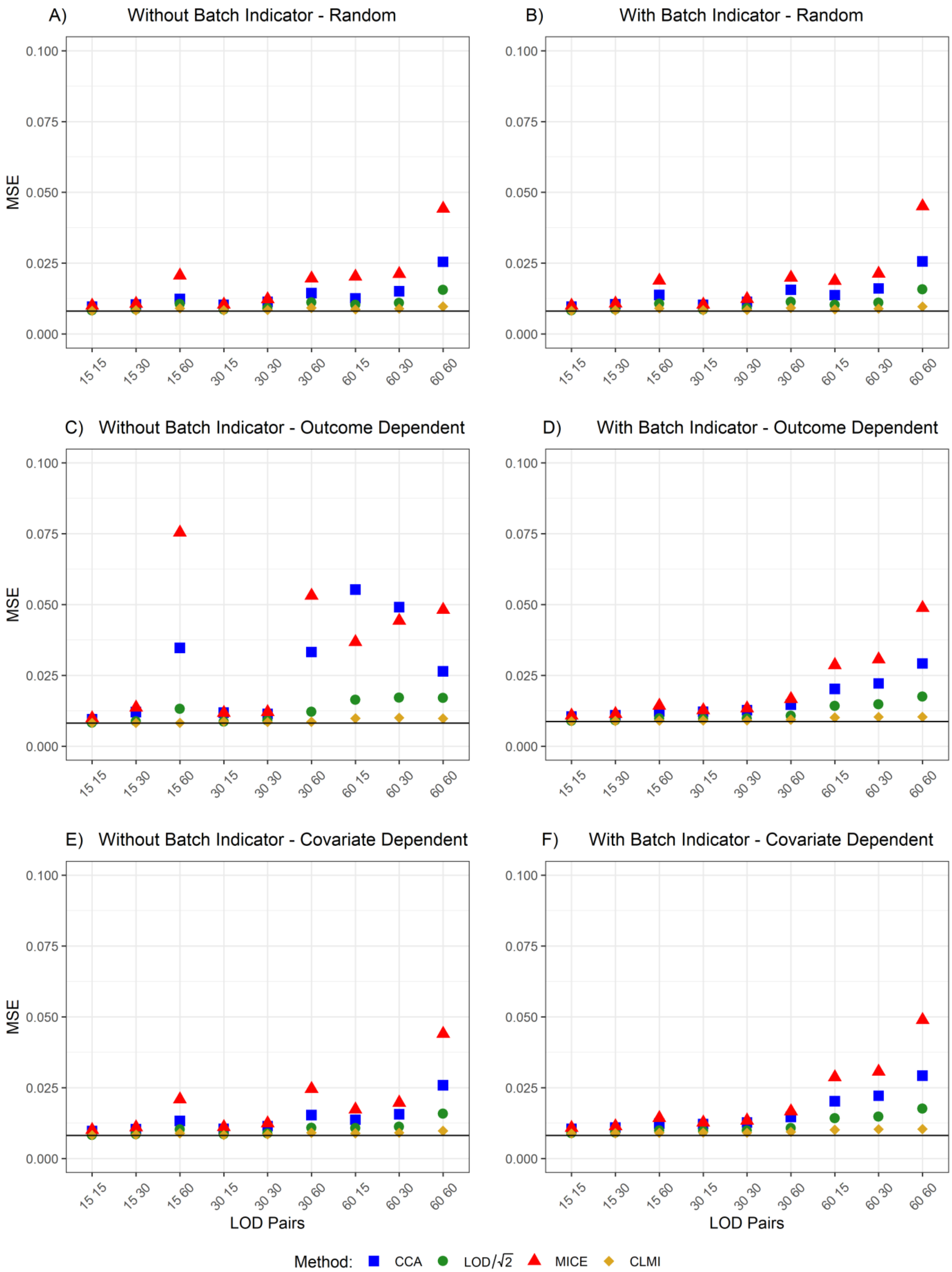
eFigure 5b. Mean-Squared Error (MSE) at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000) with *misspecification of a gamma-distributed exposure*. The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A (B) are scenarios in which batch is randomly assigned and a batch indicator term is not included (included) in the analysis model, panel C (D) are scenarios in which batch depends on Y and a batch indicator term is not included (included) in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), and panel E (F) are scenarios in which batch depends on S and G and a batch indicator term is not included (included) in the analysis model. The bolded black line indicates the gold standard MSE (no observations subject to censoring).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.


A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
Relative Bias (%)
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method:
CCA
LOD/√2
MICE
CLMI

eFigure 6a. Relative bias at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000) with *misspecification of a mixture normal-distributed exposure*. The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A (B) are scenarios in which batch is randomly assigned and a batch indicator term is not included (included) in the analysis model, panel C (D) are scenarios in which batch depends on Y and a batch indicator term is not included (included) in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), and panel E (F) are scenarios in which batch depends on S and G and a batch indicator term is not included (included) in the analysis model. The bolded black line indicates a relative bias of 0% (true $\beta_E = \log(1.5)$).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.


A) Without Batch Indicator - Random
B) With Batch Indicator - Random
C) Without Batch Indicator - Outcome Dependent
D) With Batch Indicator - Outcome Dependent
E) Without Batch Indicator - Covariate Dependent
F) With Batch Indicator - Covariate Dependent
MSE
0.100
0.075
0.050
0.025
0.000
15 15
15 30
15 60
30 15
30 30
30 60
60 15
60 30
60 60
LOD Pairs
Method: CCA LOD/√2 MICE CLMI

eFigure 6b. Mean-Squared Error (MSE) at various LOD combinations in a simulated, moderately-sized cohort study (N = 1,000) with *misspecification of a mixture normal-distributed exposure*. The first number in the LOD pair corresponds to the approximate percent of observations in batch 1 below the batch 1 LOD and the second number in the LOD pair corresponds to the approximate percent of observations in batch 2 below the batch 2 LOD. Panel A (B) are scenarios in which batch is randomly assigned and a batch indicator term is not included (included) in the analysis model, panel C (D) are scenarios in which batch depends on Y and a batch indicator term is not included (included) in the analysis model (when $P(Batch\ 1 \mid Y_i = 1) = 0.8$), and panel E (F) are scenarios in which batch depends on S and G and a batch indicator term is not included (included) in the analysis model. The bolded black line indicates the gold standard MSE (no observations subject to censoring).

Abbreviations: CCA, complete-case analysis; CLMI, censored likelihood multiple imputation; LOD, limit of detection; MICE, multiple imputation using chained equations; $LOD/\sqrt{2}$, constant imputation with $LOD/\sqrt{2}$.