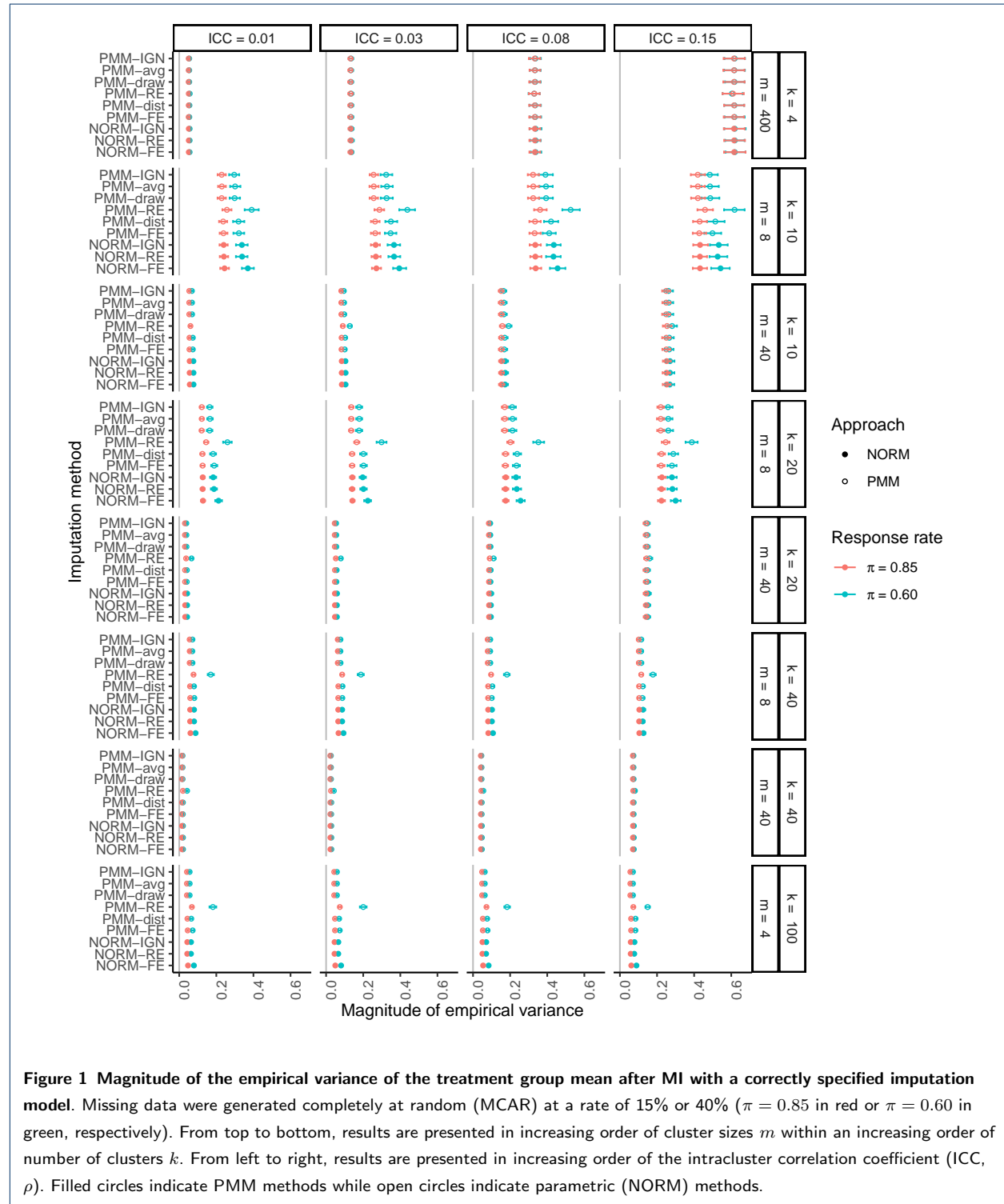


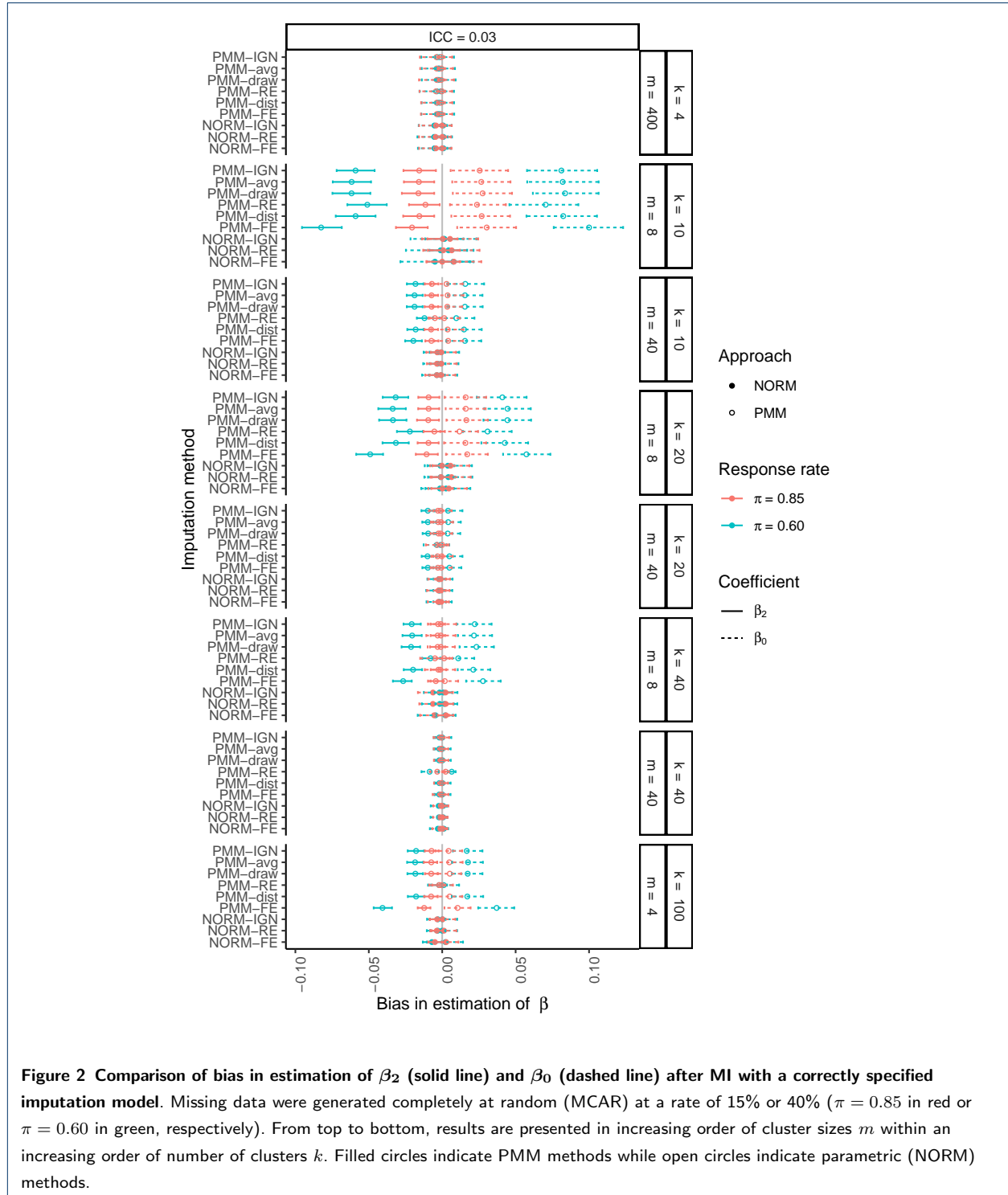
## Supplemental material

### Supplemental figures

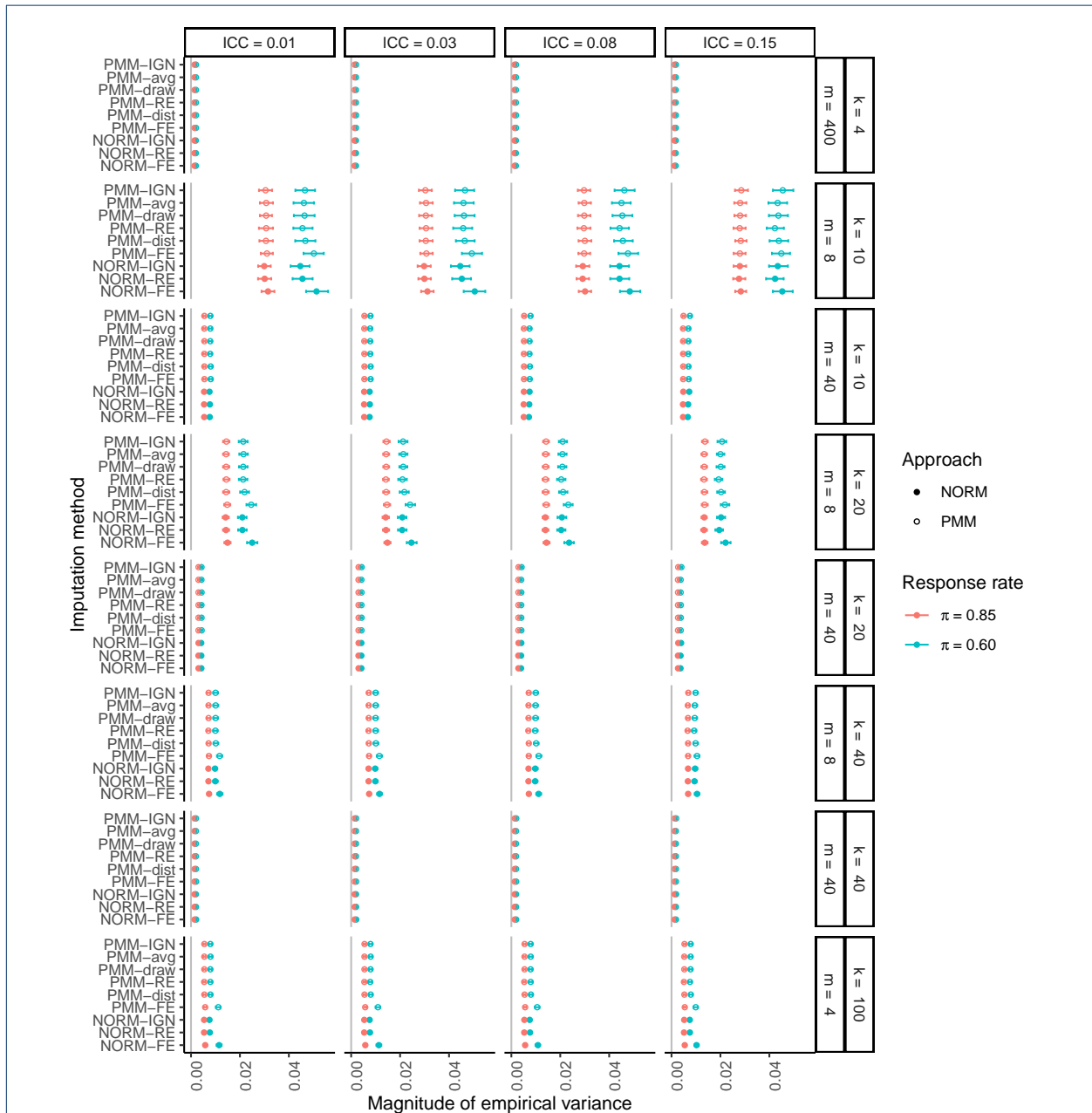
Model 1a: Estimation of treatment group mean with correctly specified mean model



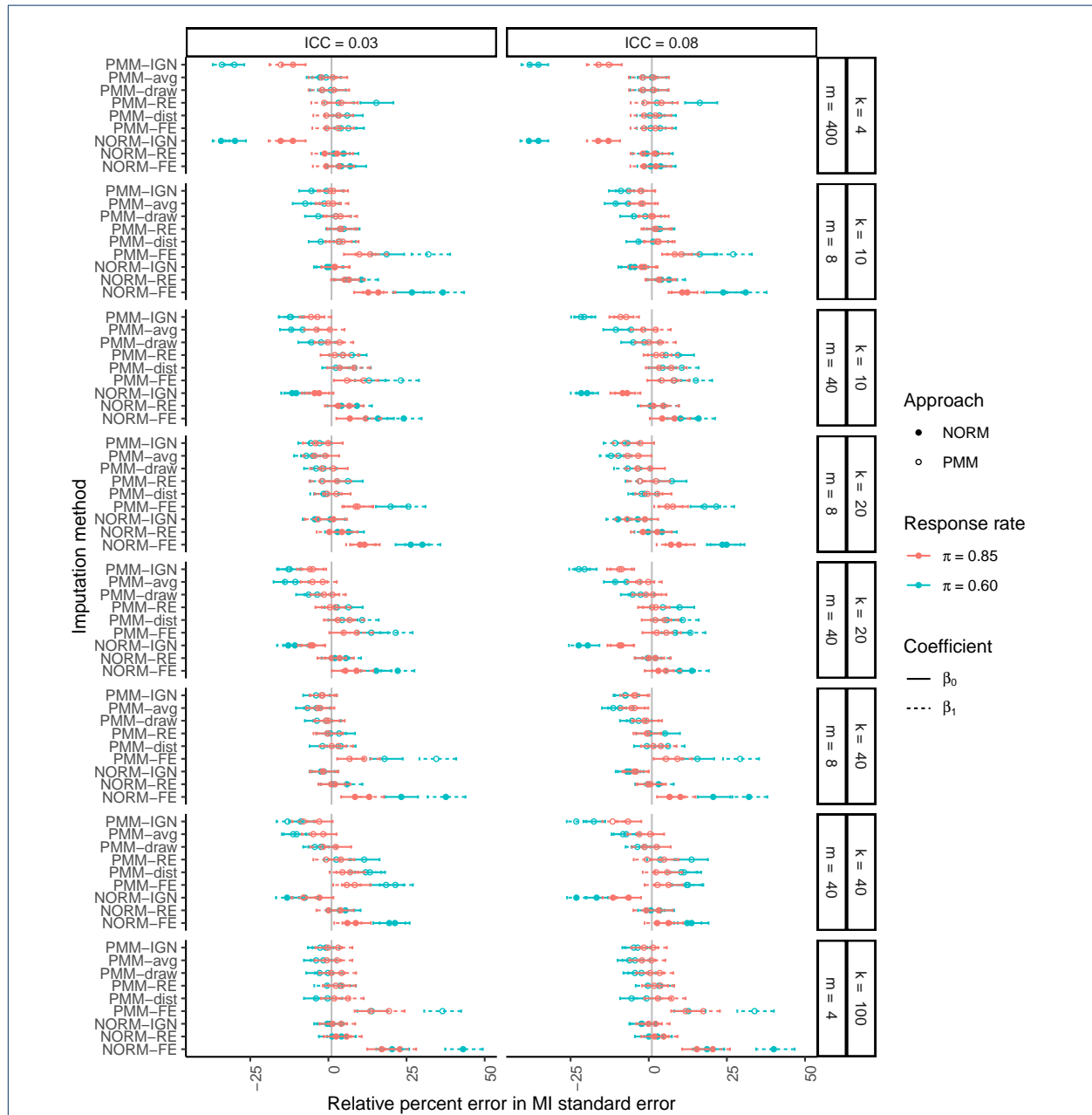
Model 1b: Estimation of regression coefficient with correctly specified mean model



**Figure 2 Comparison of bias in estimation of  $\beta_2$  (solid line) and  $\beta_0$  (dashed line) after MI with a correctly specified imputation model.** Missing data were generated completely at random (MCAR) at a rate of 15% or 40% ( $\pi = 0.85$  in red or  $\pi = 0.60$  in green, respectively). From top to bottom, results are presented in increasing order of cluster sizes  $m$  within an increasing order of number of clusters  $k$ . Filled circles indicate PMM methods while open circles indicate parametric (NORM) methods.

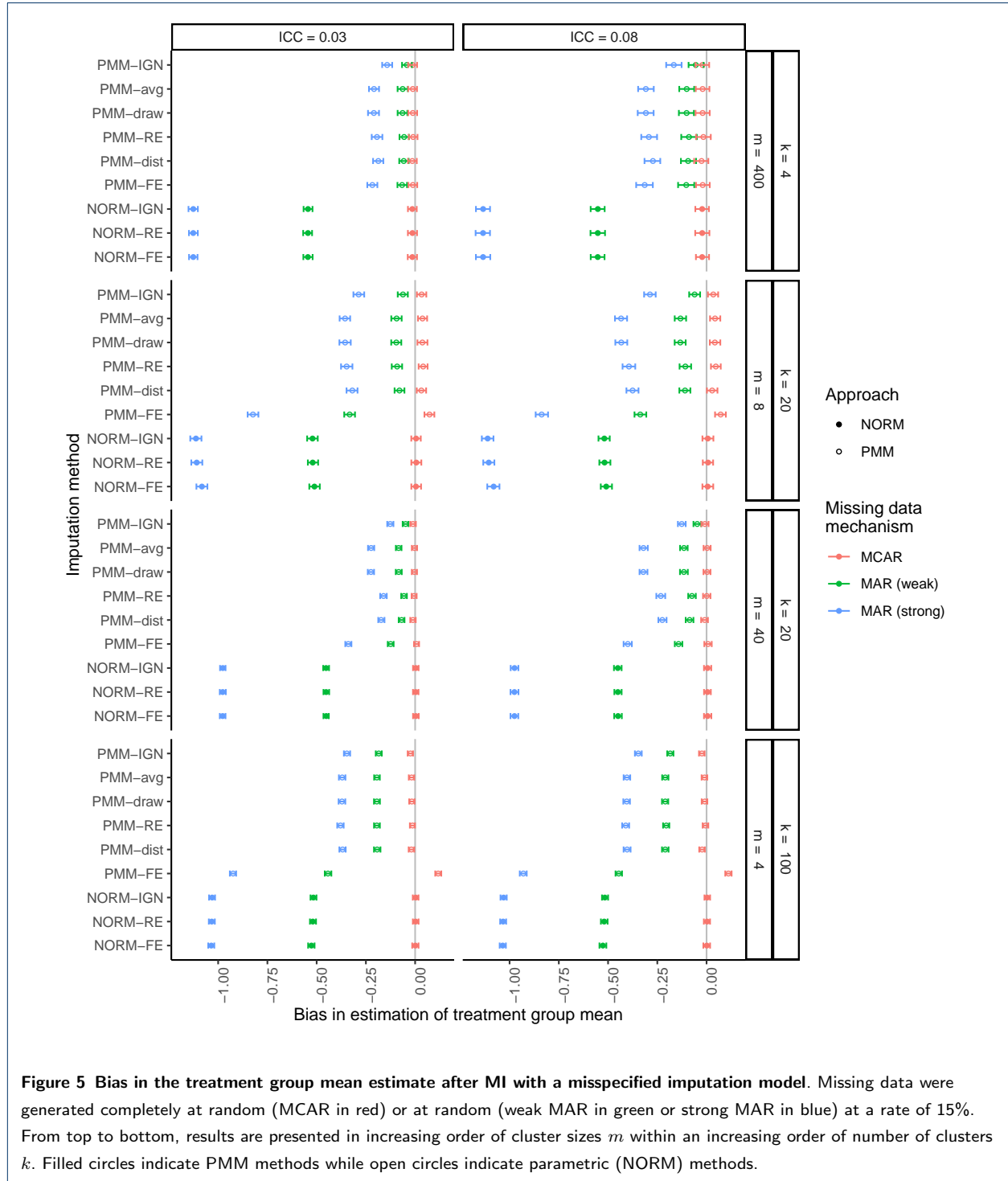


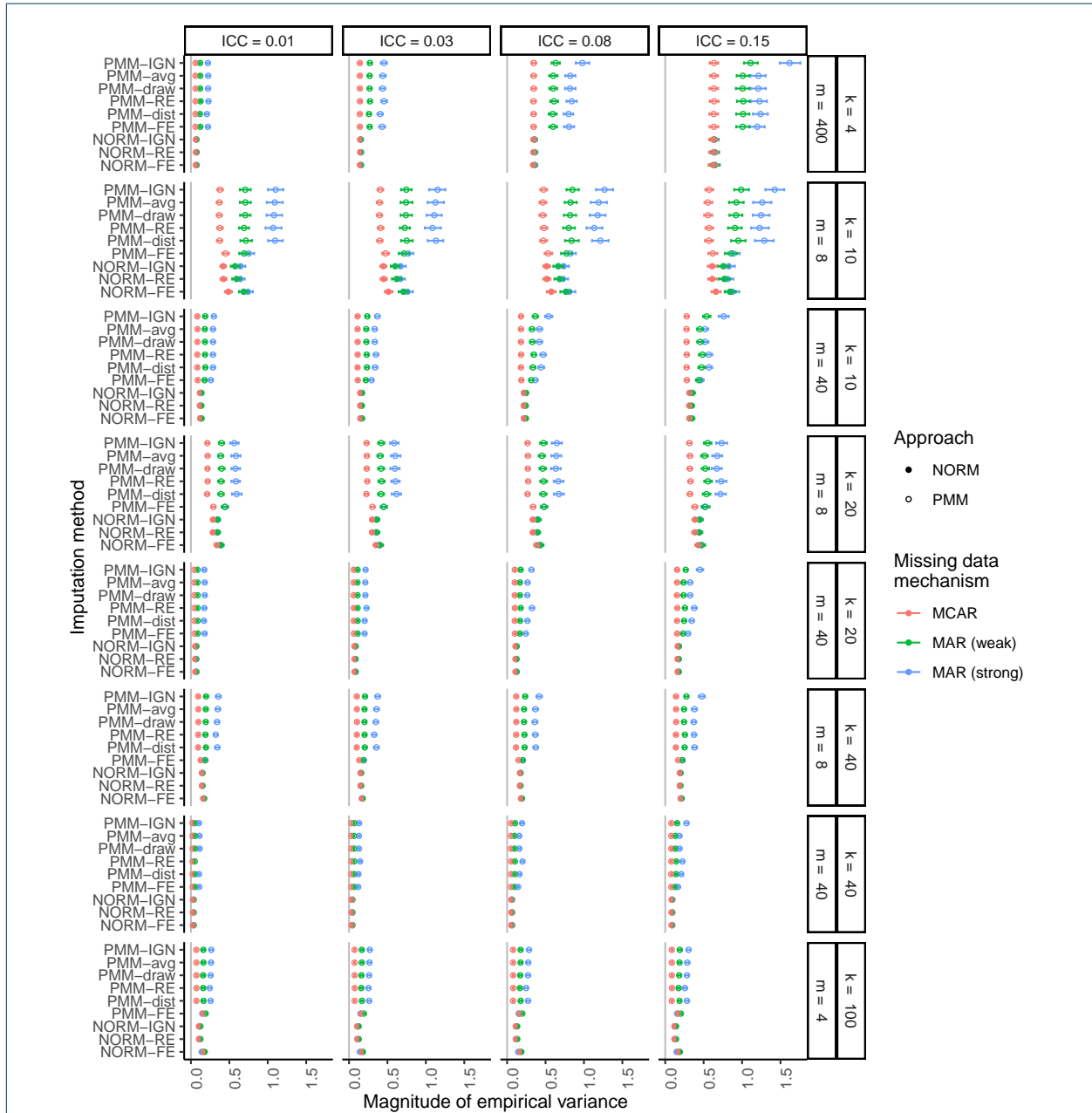
**Figure 3** Magnitude of the empirical variance of  $\beta_2$  after MI with a correctly specified imputation model. Missing data were generated completely at random (MCAR) at a rate of 15% or 40% ( $\pi = 0.85$  in red or  $\pi = 0.60$  in green, respectively). From top to bottom, results are presented in increasing order of cluster sizes  $m$  within an increasing order of number of clusters  $k$ . From left to right, results are presented in increasing order of the intracluster correlation coefficient (ICC,  $\rho$ ). Filled circles indicate PMM methods while open circles indicate parametric (NORM) methods.



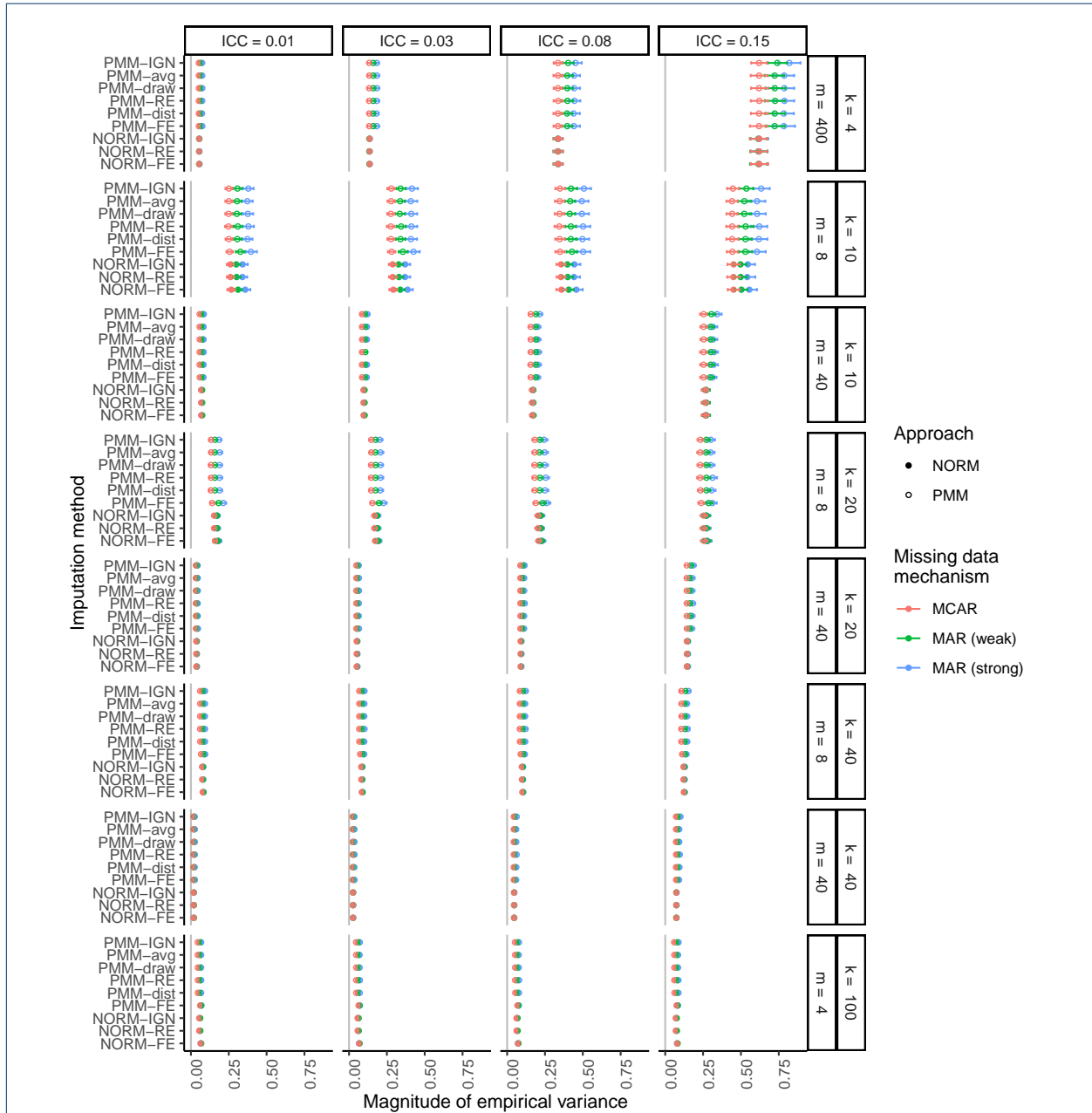
**Figure 4** Comparison of relative percent error in the MI standard error of  $\beta_0$  (solid line) and  $\beta_1$  (dashed line) after MI with a correctly specified imputation model. Missing data were generated completely at random (MCAR) at a rate of 15% or 40% ( $\pi = 0.85$  in red or  $\pi = 0.60$  in green, respectively). From top to bottom, results are presented in increasing order of cluster sizes  $m$  within an increasing order of number of clusters  $k$ . Filled circles indicate PMM methods while open circles indicate parametric (NORM) methods.

Model 2: Estimation of treatment group mean with misspecified mean model

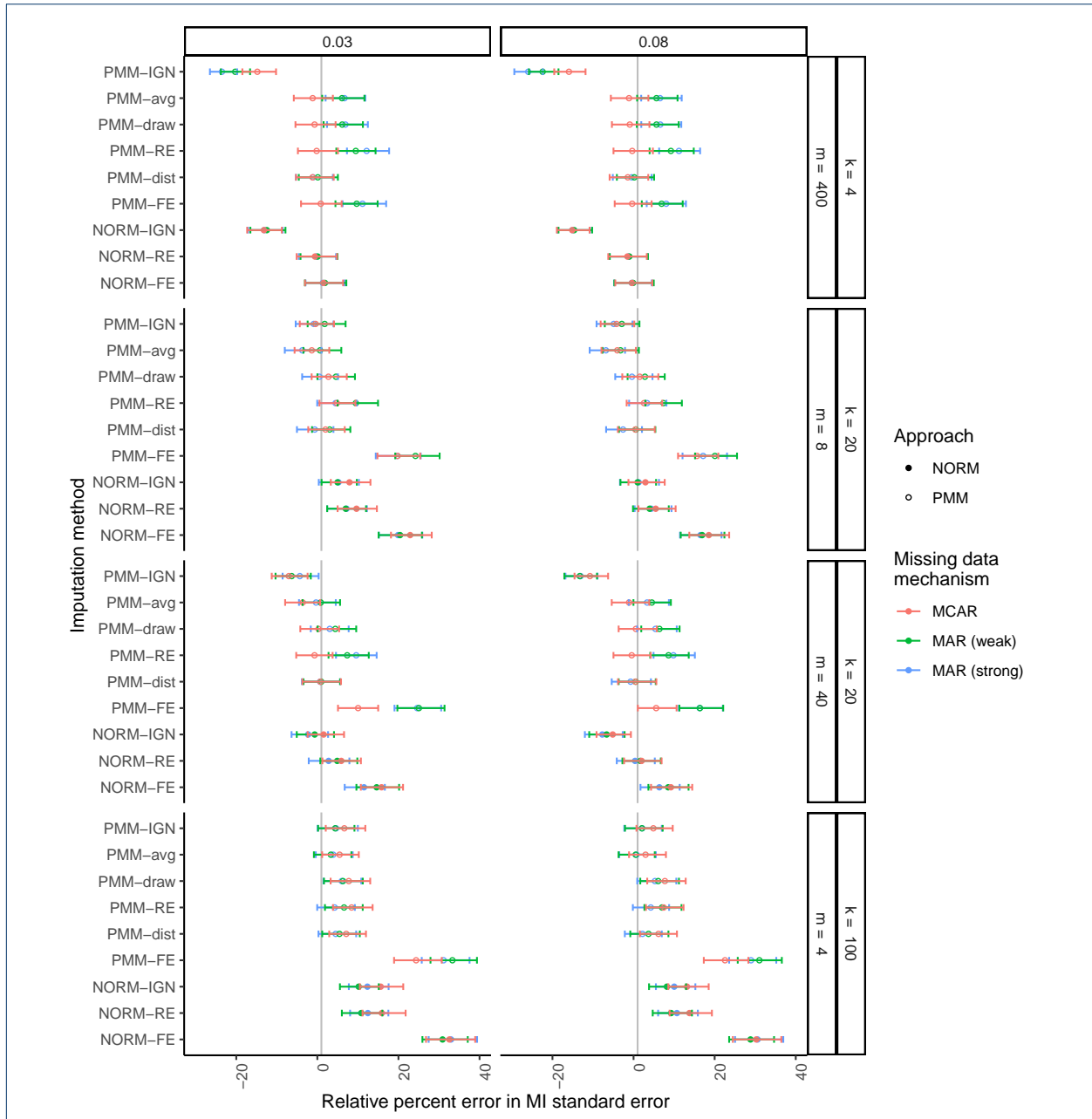




**Figure 6** Magnitude of the empirical variance of the treatment group mean after MI with a misspecified imputation model. Missing data were generated completely at random (MCAR in red) or at random (weak MAR in green or strong MAR in blue) at a rate of 40%. From top to bottom, results are presented in increasing order of cluster sizes  $m$  within an increasing order of number of clusters  $k$ . From left to right, results are presented in increasing order of the intracluster correlation coefficient (ICC,  $\rho$ ). Filled circles indicate PMM methods while open circles indicate parametric (NORM) methods.



**Figure 7** Magnitude of the empirical variance of the treatment group mean after MI with a misspecified imputation model. Missing data were generated completely at random (MCAR in red) or at random (weak MAR in green or strong MAR in blue) at a rate of 15%. From top to bottom, results are presented in increasing order of cluster sizes  $m$  within an increasing order of number of clusters  $k$ . From left to right, results are presented in increasing order of the intracluster correlation coefficient (ICC,  $\rho$ ). Filled circles indicate PMM methods while open circles indicate parametric (NORM) methods.



**Figure 8** Relative percent error in the MI standard error of the treatment group mean after MI with a misspecified imputation model. Missing data were generated completely at random (MCAR in red) or at random (weak MAR in green or strong MAR in blue) at a rate of 15%. From top to bottom, results are presented in increasing order of cluster sizes  $m$  within an increasing order of number of clusters  $k$ . Filled circles indicate PMM methods while open circles indicate parametric (NORM) methods.



## Supplemental tables

*Model 1a: Estimation of treatment group mean with correctly specified mean model***Table 1** Empirical coverage of 95% confidence intervals for the treatment group mean after multiple imputation by PMM.

ICC rho	Cluster size m	No. clusters k	PMM FE	PMM dist	PMM RE	PMM draw	PMM avg	PMM IGN
0.03	4	100	<i>98.4</i>	<i>98.1</i>	<i>96.6</i>	<i>96.9</i>	96.3	<i>96.4</i>
0.03	8	10	<i>98</i>	<i>97.6</i>	<i>97.5</i>	<i>97.3</i>	<i>97.1</i>	<i>97</i>
0.03	8	20	<i>97.1</i>	95.9	95.2	95.4	94.2	94.9
0.03	8	40	<i>97.1</i>	<i>96.8</i>	<i>95.7</i>	96	95.4	95.5
0.03	40	10	<i>97.9</i>	<i>97.7</i>	<i>97.2</i>	<i>96.8</i>	<i>96.4</i>	<i>96.1</i>
0.03	40	20	95.9	95.9	95.4	94.6	93.9	93.8
0.03	40	40	<i>97.1</i>	<i>96.8</i>	96	96	95.4	94.7
0.03	400	4	96.2	96.3	<i>96.4</i>	96	96	94.6
0.08	4	100	<i>98.5</i>	<i>98.3</i>	<i>97.1</i>	<i>96.6</i>	95.8	96.1
0.08	8	10	<i>97.4</i>	<i>97.2</i>	<i>97.2</i>	<i>96.6</i>	96.2	95.7
0.08	8	20	<i>96.6</i>	96.2	95	94.1	<b>93.4</b>	<b>93.6</b>
0.08	8	40	<i>96.9</i>	<i>96.7</i>	95.8	95.4	94.8	95
0.08	40	10	<i>97.2</i>	<i>97.1</i>	<i>97</i>	<i>96.8</i>	<i>96.4</i>	94.5
0.08	40	20	95.8	95.6	95.7	94.9	94.4	<b>93</b>
0.08	40	40	<i>96.4</i>	96.2	95.9	95.5	95.5	<b>93.6</b>
0.08	400	4	<i>96.4</i>	<i>96.5</i>	<i>96.5</i>	96.2	96.3	93.9

**Bolded** values indicate coverage rates below 95%.

*Italicized* values indicate coverage rates above 95%.

Missing data were generated completely at random with a response rate of 85%, and data were imputed using an imputation model that correctly specified the covariate effects.

**Table 2** Empirical coverage of 95% confidence intervals for the treatment group mean after parametric multiple imputation.

ICC	Cluster size	No. clusters	NORM	NORM	NORM
$\rho$	m	k	FE	RE	IGN
0.03	4	100	<i>98.6</i>	<i>97</i>	96.3
0.03	8	10	<i>98.5</i>	<i>97.2</i>	<i>97.6</i>
0.03	8	20	<i>97.7</i>	95.1	94.5
0.03	8	40	<i>97.8</i>	95.9	95.3
0.03	40	10	<i>97.8</i>	<i>97.1</i>	95.6
0.03	40	20	96	94.9	93.9
0.03	40	40	<i>97.3</i>	96	94.3
0.03	400	4	96.3	96.1	94.5
0.08	4	100	<i>98.5</i>	<i>96.9</i>	95.8
0.08	8	10	<i>98.1</i>	<i>96.8</i>	96
0.08	8	20	<i>97.1</i>	94.5	93.9
0.08	8	40	<i>97.1</i>	95.5	94.7
0.08	40	10	<i>97.1</i>	<i>97</i>	94.5
0.08	40	20	95.8	94.9	<b>93</b>
0.08	40	40	96.3	95.7	<b>93.4</b>
0.08	400	4	<i>96.5</i>	<i>96.4</i>	94

**Bolded** values indicate coverage rates below 95%.

*Italicized* values indicate coverage rates above 95%.

Missing data were generated completely at random with a response rate of 85%, and data were imputed using an imputation model that correctly specified the covariate effects.

*Model 1b: Estimation of regression coefficient with correctly specified mean model***Table 3** Empirical coverage of 95% confidence intervals for  $\beta_2$  after multiple imputation by PMM.

ICC	Cluster size	No. clusters	PMM	PMM	PMM	PMM	PMM	PMM
rho	m	k	FE	dist	RE	draw	avg	IGN
0.03	4	100	95.6	95.3	<i>96.8</i>	96.1	95.6	96.1
0.03	8	10	<i>97.5</i>	<i>97</i>	<i>96.5</i>	<i>97.3</i>	<i>96.5</i>	<i>97.2</i>
0.03	8	20	95.1	94.1	94.5	94.9	93.8	94.6
0.03	8	40	94.7	94.5	94.8	95.1	94.5	95
0.03	40	10	<i>97.7</i>	<i>97.5</i>	<i>97.6</i>	<i>97.9</i>	<i>96.8</i>	<i>97.8</i>
0.03	40	20	96.2	96.1	96.3	<i>96.5</i>	94.1	96
0.03	40	40	95.6	95.2	95.4	95.6	93.9	96.3
0.03	400	4	<i>99.6</i>	<i>99.6</i>	<i>99.6</i>	<i>99.6</i>	<i>99.3</i>	<i>99.6</i>
0.08	4	100	96	95.5	<i>96.4</i>	<i>96.4</i>	95.3	95.9
0.08	8	10	<i>97.2</i>	<i>96.7</i>	<i>97.1</i>	<i>97.6</i>	95.9	<i>97</i>
0.08	8	20	94.9	94.2	94.5	95	<b>93.3</b>	94.5
0.08	8	40	94.5	94.5	94.7	95	93.9	95.3
0.08	40	10	<i>97.7</i>	<i>97.7</i>	<i>97.7</i>	<i>97.9</i>	<i>97.2</i>	<i>97.7</i>
0.08	40	20	96.3	96.1	96.1	96.1	94.5	96.1
0.08	40	40	95.3	95.3	95.5	95.6	93.8	95.7
0.08	400	4	<i>99.7</i>	<i>99.8</i>	<i>99.8</i>	<i>99.8</i>	<i>99.7</i>	<i>99.7</i>

**Bolded** values indicate coverage rates below 95%.

*Italicized* values indicate coverage rates above 95%.

Missing data were generated completely at random (MCAR) with a response rate of 85%, and data were imputed using an imputation model that correctly specified the covariate effects.

**Table 4** Empirical coverage of 95% confidence intervals for  $\beta_2$  after parametric multiple imputation.

ICC $\rho$	Cluster size m	No. clusters k	NORM FE	NORM RE	NORM IGN
0.03	4	100	<i>97.3</i>	<i>96.6</i>	95.9
0.03	8	10	<i>97.6</i>	<i>97.6</i>	<i>97.3</i>
0.03	8	20	95.3	95.2	95.3
0.03	8	40	95.7	95.6	95.6
0.03	40	10	<i>98</i>	<i>98</i>	<i>97.9</i>
0.03	40	20	<i>96.4</i>	<i>96.5</i>	96.3
0.03	40	40	95.8	95.5	96.1
0.03	400	4	<i>99.8</i>	<i>99.9</i>	<i>99.8</i>
0.08	4	100	<i>97.1</i>	<i>96.6</i>	96.1
0.08	8	10	<i>97.5</i>	<i>97.5</i>	<i>97.4</i>
0.08	8	20	95.3	95.2	95.4
0.08	8	40	95.7	95.8	95.6
0.08	40	10	<i>98</i>	<i>97.9</i>	<i>98</i>
0.08	40	20	<i>96.3</i>	<i>96.5</i>	<i>96.6</i>
0.08	40	40	95.7	95.3	96.3
0.08	400	4	<i>99.8</i>	<i>99.9</i>	<i>99.8</i>

**Bolded** values indicate coverage rates below 95%.

*Italicized* values indicate coverage rates above 95%.

Missing data were generated completely at random (MCAR) with a response rate of 85%, and data were imputed using an imputation model that correctly specified the covariate effects.