Supplementary Material to "A Doubly Robust Method to Handle Missing Multilevel Outcome Data with Application to the China Health and Nutrition Survey"

Nicole M. Butera, Donglin Zeng, Annie Green Howard, Penny Gordon-Larsen, and Jianwen Cai

August 16, 2021

S1 Web Appendix A: Assumptions

Here we list the assumptions for the main results of this paper.

A1 There exist two independent latent vectors, $\mathbf{a_j}$ and $\mathbf{b_j}$, such that R_{ij} and Y_{ij} are conditionally independent given $\mathbf{Z_{ij}}$, $\mathbf{a_j}$, $\mathbf{b_j}$ (i.e., $R_{ij} \perp Y_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}, \mathbf{b_j}$), and R_{ij} depends on $\mathbf{Z_{ij}}$ and $\mathbf{a_j}$ only (i.e., $R_{ij} \perp \mathbf{b_j}|\mathbf{Z_{ij}}, \mathbf{a_j}$) and Y_{ij} depends on $\mathbf{Z_{ij}}$ and $\mathbf{b_j}$ only (i.e., $Y_{ij} \perp \mathbf{a_j}|\mathbf{Z_{ij}}, \mathbf{b_j}$); note that these assumptions imply that the data are missing at random (i.e., that the outcome variable is independent of missingness, conditional on the observed data: $R_{ij} \perp Y_{ij}|\mathbf{Z_{ij}}$)[1]. In addition, all parameters that define the joint distribution of R_{ij} and $\mathbf{a_j}$ given $\mathbf{Z_{ij}}$ are distinct from the parameters that define the joint distribution of Y_{ij} and $\mathbf{b_j}$ given $\mathbf{Z_{ij}}$ (i.e., the model for the joint distribution for $(R_j, \mathbf{a_j})$ conditional on $\mathbf{Z_j}$ and the model for the joint distribution for $(Y_j, \mathbf{b_j})$ conditional on $\mathbf{Z_j}$ do not share the parameters).

A2 The number of data records per cluster n_j is bounded for all $j = 1, ..., m$.

A3 $\underset{\boldsymbol{\theta}}{sup}|m^{-1}S_m(\boldsymbol{\beta}; \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\phi}) - E[g(\mathbf{R_j}, \mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\phi})]| \underset{p}{\rightarrow} 0$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\phi})$.

A4 $E[S_m(\boldsymbol{\beta}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)] = 0$ has a unique solution for $\boldsymbol{\beta}$.

A5 $S_m(\boldsymbol{\beta})$ is twice continuously differentiable in the support of $(\mathbf{R}, \mathbf{Y}, \mathbf{Z})$.

A6 There exists an integrable function $f(\cdot)$ such that $||\partial^2_{\boldsymbol{\beta}} S_m(\boldsymbol{\beta})|| \leq f(\mathbf{R}, \mathbf{Y}, \mathbf{Z})$ for any value of $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}^*$.

A7 $E||\partial_{\boldsymbol{\beta}} S_m(\boldsymbol{\beta})||^2 < \infty$.

A8 $E\left[\partial_{\boldsymbol{\beta}} S_m(\boldsymbol{\beta}^*)\right]$ exists and is non-singular.

S2 Web Appendix B: Proofs of Main Theoretical Results

S2.1 Proof of Lemma 1

The estimating functions can be re-written as the following, where $\pi^*_{ij}(\mathbf{a_j}) = \tilde{P}[R_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}; \boldsymbol{\alpha}^*]$ and $\nu^*_{ij}(\mathbf{b_j}) = \tilde{E}[Y_{ij}|\mathbf{Z_{ij}}, \mathbf{b_j}; \boldsymbol{\gamma}^*]$ based on the specified working models:

$$
\begin{aligned}
&S_m(\boldsymbol{\beta}^*; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) \\
&= \sum_{i,j} \int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})} (Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)) \partial_{\boldsymbol{\beta}} \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*) \tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) d\mathbf{a_j} \\
&\quad - \sum_{i,j} \int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})} \tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) d\mathbf{a_j} \int (\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)) \partial_{\boldsymbol{\beta}} \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*) \tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) d\mathbf{b_j} \\
&\quad + \sum_{i,j} \int (\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)) \partial_{\boldsymbol{\beta}} \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*) \tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) d\mathbf{b_j},
\end{aligned}
$$

where $\tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) \propto \tilde{p}(\mathbf{a_j}; \boldsymbol{\tau}^*) \prod_{i=1}^{n_j} \tilde{p}(R_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}; \boldsymbol{\alpha}^*)$ and $\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) \propto \tilde{p}(\mathbf{b_j}; \boldsymbol{\phi}^*) \prod_{i=1}^{n_j} \tilde{p}(Y_{ij}|\mathbf{Z_{ij}}, \mathbf{b_j}; \boldsymbol{\gamma}^*)^{R_{ij}}$. $\int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})} \tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) d\mathbf{a_j} = E\left[\frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]$ if working model $[\mathbf{R_j}, \mathbf{a_j}|\mathbf{Z_j}; \boldsymbol{\alpha}, \boldsymbol{\tau}]$ is specified correctly, and $\int (\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)) \partial_{\boldsymbol{\beta}} \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)$ $\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) d\mathbf{b_j} = E\left[(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)) \partial_{\boldsymbol{\beta}} \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]$ if working model $[\mathbf{Y_j}, \mathbf{b_j}|\mathbf{Z_j}; \boldsymbol{\gamma}, \boldsymbol{\phi}]$ is specified correctly.

If working model $[\mathbf{R_j}, \mathbf{a_j}|\mathbf{Z_j}; \boldsymbol{\alpha}, \boldsymbol{\tau}]$ is correct:

Then substituting $\int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}\tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)d\mathbf{a_j} = E\left[\frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]$ gives

$$
\begin{aligned}
&E[S_m(\boldsymbol{\beta}^*; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)|\mathbf{X}] \\
=&\sum_{i,j} E\left[E\left[\frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}(Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]|\mathbf{X_j}\right] \\
&- \sum_{i,j} E\left[E\left[\frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}\int(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)d\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]|\mathbf{X_j}\right] \\
&+ \sum_{i,j} E\left[\int(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)d\mathbf{b_j}|\mathbf{X_j}\right]
\end{aligned}
$$

Since $R_{ij} = 0, 1$ for all i, j, $R_{ij} \perp Y_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}, \mathbf{b_j}$, and $\pi^*_{ij}(\mathbf{a_j}) = E[R_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}] = E[R_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}, \mathbf{b_j}]$, it can be shown that

$$
\begin{aligned}
&E[S_m(\boldsymbol{\beta}^*; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)|\mathbf{X}] \\
&\quad= \sum_{i,j} E\left[(Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)|\mathbf{X_j}\right] \\
&\quad\quad- \sum_{i,j} E\left[E\left[\int(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)d\mathbf{b_j}|\mathbf{Z_j}, \mathbf{a_j}, \mathbf{b_j}\right]|\mathbf{X_j}\right] \\
&\quad\quad+ \sum_{i,j} E\left[\int(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)d\mathbf{b_j}|\mathbf{X_j}\right] \\
&\quad= \sum_{i,j} E\left[(Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)|\mathbf{X_j}\right] = 0
\end{aligned}
$$

If working model $[\mathbf{Y_j}, \mathbf{b_j}|\mathbf{Z_j}; \boldsymbol{\gamma}, \boldsymbol{\phi}]$ is correct:

Then substituting $\int(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta}^*)\tilde{p}(\mathbf{b_j}|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)d\mathbf{b_j}$

$= E\left[(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]$ gives

$$
\begin{aligned}
&E[S_m(\boldsymbol{\beta^*}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*}, \boldsymbol{\gamma^*}, \boldsymbol{\phi^*})|\mathbf{X}] \\
&\quad = \sum_{i,j} E\left[\int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}(Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})\tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*})d\mathbf{a_j}|\mathbf{X_j}\right] \\
&\qquad - \sum_{i,j} E\left[\int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}\tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*})d\mathbf{a_j}\; E\left[(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]|\mathbf{X_j}\right] \\
&\qquad + \sum_{i,j} E\left[E\left[(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})|\mathbf{R_j}, \mathbf{R_j}\mathbf{Y_j}, \mathbf{Z_j}\right]|\mathbf{X_j}\right]
\end{aligned}
$$

Since $R_{ij} \perp Y_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}, \mathbf{b_j}$ and $\nu^*_{ij}(\mathbf{b_j}) = E[Y_{ij}|\mathbf{Z_{ij}}, \mathbf{b_j}] = E[Y_{ij}|\mathbf{Z_{ij}}, \mathbf{a_j}, \mathbf{b_j}]$, it can be shown that

$$
\begin{aligned}
&E[S_m(\boldsymbol{\beta^*}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*}, \boldsymbol{\gamma^*}, \boldsymbol{\phi^*})|\mathbf{X}] \\
&\quad = \sum_{i,j} E\left[(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})E\left[R_{ij}\int \frac{1}{\pi^*_{ij}(\mathbf{a_j})}\tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*})d\mathbf{a_j}|\mathbf{Z_j}, \mathbf{a_j}, \mathbf{b_j}\right]|\mathbf{X_j}\right] \\
&\qquad - \sum_{i,j} E\left[(\nu^*_{ij}(\mathbf{b_j}) - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})\int \frac{R_{ij}}{\pi^*_{ij}(\mathbf{a_j})}\tilde{p}(\mathbf{a_j}|\mathbf{R_j}, \mathbf{Z_j}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*})d\mathbf{a_j}|\mathbf{X_j}\right] \\
&\qquad + \sum_{i,j} E\left[(Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})|\mathbf{X_j}\right] \\
&\quad = \sum_{i,j} E\left[(Y_{ij} - \mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*}))\partial_{\boldsymbol{\beta}}\mu(\mathbf{X_{ij}^T}\boldsymbol{\beta^*})|\mathbf{X_j}\right] = 0
\end{aligned}
$$

$$E[S_m(\boldsymbol{\beta^*}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*}, \boldsymbol{\gamma^*}, \boldsymbol{\phi^*})] = E[E[S_m(\boldsymbol{\beta^*}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*}, \boldsymbol{\gamma^*}, \boldsymbol{\phi^*})|\mathbf{X}]] = E[0] = 0$$

S2.2 Proof of Theorem 1

S2.2.1 Consistency of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}$

Let $S(\boldsymbol{\beta'}; \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\phi}) = E[g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\phi})]|_{(\boldsymbol{\beta'}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\phi})}$. If $\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}$ belongs to a compact set containing $\boldsymbol{\beta^*}$, then every subsequence has a further subsequence $\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_{k,l}}}$ that converges to some $\tilde{\boldsymbol{\beta}}_{\boldsymbol{l}}$ almost surely (by the Bolzano-Weierstrass Theorem). Therefore, $S(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_{k,l}}}; \widehat{\boldsymbol{\alpha}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\tau}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\phi}}_{\boldsymbol{m_{k,l}}}) \underset{p}{\rightarrow} S(\tilde{\boldsymbol{\beta}}_{\boldsymbol{l}}; \boldsymbol{\alpha^*}, \boldsymbol{\tau^*}, \boldsymbol{\gamma^*}, \boldsymbol{\phi^*})$. Since $S_{m_{k,l}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_{k,l}}}; \widehat{\boldsymbol{\alpha}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\tau}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\phi}}_{\boldsymbol{m_{k,l}}}) = 0$, and $|S(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_{k,l}}}; \widehat{\boldsymbol{\alpha}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\tau}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\phi}}_{\boldsymbol{m_{k,l}}})| = |m_{k,l}^{-1}S_{m_{k,l}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_{k,l}}}; \widehat{\boldsymbol{\alpha}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\tau}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\phi}}_{\boldsymbol{m_{k,l}}}) - S(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m_{k,l}}}; \widehat{\boldsymbol{\alpha}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\tau}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{m_{k,l}}}, \widehat{\boldsymbol{\phi}}_{\boldsymbol{m_{k,l}}})| \underset{p}{\rightarrow} 0$ uniformly, then $S(\tilde{\boldsymbol{\beta}}_{\boldsymbol{l}};\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\gamma^*,\boldsymbol{\phi}^*)=0$ for each $\tilde{\boldsymbol{\beta}}_{\boldsymbol{l}}$. Since we assume that $mS(\boldsymbol{\beta};\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\gamma^*,\boldsymbol{\phi}^*)=E[S_m(\boldsymbol{\beta};\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\boldsymbol{\gamma}^*,\boldsymbol{\phi}^*)]=0$ has a unique solution at $\boldsymbol{\beta}=\boldsymbol{\beta}^*$, then $\tilde{\boldsymbol{\beta}}_{\boldsymbol{l}}=\boldsymbol{\beta}^*$ for every convergent subsubsequence, and so $\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}} \underset{p}{\rightarrow} \boldsymbol{\beta}^*$.

S2.2.2 Asymptotic Distribution of $m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}-\boldsymbol{\beta}^*\right)$

Step 1: Asymptotic distribution of $m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}-\boldsymbol{\beta}^*\right)$ if $(\boldsymbol{\alpha},\boldsymbol{\tau},\gamma,\boldsymbol{\phi})$ are known constants

Note that $m^{-1/2}\sum_{j=1}^m g(\mathbf{R_j},\mathbf{Y_j},\mathbf{Z_j};\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\gamma^*,\boldsymbol{\phi}^*)$ converges to a normal distribution with mean zero as $m\to\infty$, and $m^{-1}\sum_{j=1}^m \partial_{\boldsymbol{\beta}} g(\mathbf{R},\mathbf{Y},\mathbf{Z};\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\gamma^*,\boldsymbol{\phi}^*) \underset{p}{\rightarrow} \mathbf{C}=E\left[\partial_{\boldsymbol{\beta}} g(\mathbf{R},\mathbf{Y},\mathbf{Z};\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\gamma^*,\boldsymbol{\phi}^*)\right]$ as $m\to\infty$. Therefore, using a Taylor series expansion of $S_m(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}};\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\boldsymbol{\gamma}^*,\boldsymbol{\phi}^*)$ around $\boldsymbol{\beta}^*$, we obtain $m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}-\boldsymbol{\beta}^*\right)=-\mathbf{C}^{-1}m^{-1/2}\sum_{j=1}^m g(\mathbf{R_j},\mathbf{Y_j},\mathbf{Z_j};\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\gamma^*,\boldsymbol{\phi}^*)+o_p(1)$, which converges to a normal distribution with mean zero.

Step 2: Asymptotic distribution for $m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}-\boldsymbol{\beta}^*\right)$ if $(\boldsymbol{\alpha},\boldsymbol{\tau},\gamma,\boldsymbol{\phi})$ are estimated

In this case, $m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}-\boldsymbol{\beta}^*\right)=-\mathbf{C}^{-1}m^{1/2}\left[\frac{1}{m}\sum_{j=1}^m g(\mathbf{R_j},\mathbf{Y_j},\mathbf{Z_j};\boldsymbol{\beta}^*,\widehat{\boldsymbol{\alpha}}_{\boldsymbol{m}},\widehat{\boldsymbol{\tau}}_{\boldsymbol{m}},\widehat{\gamma}_{\boldsymbol{m}},\widehat{\boldsymbol{\phi}}_{\boldsymbol{m}})\right]+o_p(1)$. Using a Taylor series expansion of this expression around $(\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*)$, we obtain

$$
\begin{aligned}
&m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}}-\boldsymbol{\beta}^*\right)\\
&\quad=-\mathbf{C}^{-1}m^{1/2}\left[\frac{1}{m}\sum_{j=1}^m g(\mathbf{R_j},\mathbf{Y_j},\mathbf{Z_j};\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\widehat{\gamma},\widehat{\boldsymbol{\phi}})\right]\\
&\qquad-\mathbf{C}^{-1}\left[\frac{1}{m}\sum_{j=1}^m \partial_{\boldsymbol{\alpha},\boldsymbol{\tau}} g(\mathbf{R_j},\mathbf{Y_j},\mathbf{Z_j};\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*,\widehat{\gamma},\widehat{\boldsymbol{\phi}})\right]\psi_{\boldsymbol{\alpha},\boldsymbol{\tau}}(\mathbf{R},\mathbf{Z};\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*)+o_p(1),
\end{aligned}
$$

where $\psi_{\boldsymbol{\alpha},\boldsymbol{\tau}}(\cdot)$ is defined in equation (5) in the main text. Further expanding the above expression using a Taylor series expansion around $(\boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)$, we obtain

$$
\begin{aligned}
& m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}} - \boldsymbol{\beta}^*\right) \\
& = -\mathbf{C}^{-1} m^{1/2}\left[\frac{1}{m}\sum_{j=1}^{m} g(\mathbf{R_j}, \mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] \\
& \quad -\mathbf{C}^{-1}\left[\frac{1}{m}\sum_{j=1}^{m} \partial_{\boldsymbol{\alpha},\boldsymbol{\tau}} g(\mathbf{R_j}, \mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] \psi_{\boldsymbol{\alpha},\boldsymbol{\tau}}(\mathbf{R}, \mathbf{Z}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) \\
& \quad -\mathbf{C}^{-1}\left[\frac{1}{m}\sum_{j=1}^{m} \partial_{\gamma,\boldsymbol{\phi}} g(\mathbf{R_j}, \mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] \psi_{\boldsymbol{\gamma},\boldsymbol{\phi}}(\mathbf{R}, \mathbf{RY}, \mathbf{Z}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) \\
& \quad -\mathbf{C}^{-1} m^{-1/2}\left[\frac{1}{m}\sum_{j=1}^{m} \partial_{\boldsymbol{\alpha},\boldsymbol{\tau},\gamma,\boldsymbol{\phi}} g(\mathbf{R_j}, \mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] \\
& \qquad \psi_{\boldsymbol{\alpha},\boldsymbol{\tau}}(\mathbf{R}, \mathbf{Z}; \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) \psi_{\boldsymbol{\gamma},\boldsymbol{\phi}}(\mathbf{R}, \mathbf{RY}, \mathbf{Z}; \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) + o_p(1),
\end{aligned}
$$

where $\psi_{\boldsymbol{\gamma},\boldsymbol{\phi}}(\cdot)$ is defined in equation (6) in the main text. By the Weak Law of Large Numbers and Slutsky's Theorem,

$$
\begin{aligned}
& m^{1/2}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{m}} - \boldsymbol{\beta}^*\right) \\
& = m^{-1/2}\{E\left[\partial_{\boldsymbol{\beta}} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right]\}^{-1} \sum_{j=1}^{m}\{-g(\mathbf{R_j}, \mathbf{Y_j}, \mathbf{Z_j}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \boldsymbol{\gamma}^*, \boldsymbol{\phi}^*) \\
& \quad + E\left[\partial_{\boldsymbol{\alpha},\boldsymbol{\tau}} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] E[\partial^2_{\boldsymbol{\alpha},\boldsymbol{\tau}} l(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)]^{-1} \partial_{\boldsymbol{\alpha},\boldsymbol{\tau}} l_j(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) \\
& \quad + E\left[\partial_{\gamma,\boldsymbol{\phi}} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] E[\partial^2_{\boldsymbol{\gamma},\boldsymbol{\phi}} l(\boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)]^{-1} \partial_{\boldsymbol{\gamma},\boldsymbol{\phi}} l_j(\boldsymbol{\gamma}^*, \boldsymbol{\phi}^*)\big\} + o_p(1),
\end{aligned}
$$

which converges to a normal distribution with mean zero and covariance matrix

$$
\begin{aligned}
& \{E\left[\partial_{\boldsymbol{\beta}} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right]\}^{-1} \\
& E\big[\{-g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*) \\
& \qquad + E\left[\partial_{\alpha,\tau} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] E[\partial^2_{\boldsymbol{\alpha},\boldsymbol{\tau}} l(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)]^{-1} \partial_{\boldsymbol{\alpha},\boldsymbol{\tau}} l(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) \\
& \qquad + E\left[\partial_{\gamma,\phi} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right] E[\partial^2_{\gamma,\boldsymbol{\phi}} l(\gamma^*, \boldsymbol{\phi}^*)]^{-1} \partial_{\gamma,\phi} l(\gamma^*, \boldsymbol{\phi}^*)\big\}^{\otimes 2}\big] \\
& \{E\left[\partial_{\boldsymbol{\beta}} g(\mathbf{R}, \mathbf{Y}, \mathbf{Z}; \boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \gamma^*, \boldsymbol{\phi}^*)\right]\}^{-1}.
\end{aligned}
$$

S3 Web Appendix C: Simulation Study with Binary Outcome

S3.1 General Set-Up

We now present additional simulation studies, considering a binary outcome variable. Data were simulated in the following way. One thousand datasets were simulated, each with 1000 clusters with 2 data records each (i.e., 1000 individuals with data for 2 time-points each). Let j indicate the individual and $i = 1, 2$ indicate the time-point. One time-varying predictor variable of interest, $\mathbf{X_j} = \binom{X_{1j}}{X_{2j}}$, was generated for each cluster from a multivariate normal distribution, $N_2\left(\left(\begin{smallmatrix} 1.5 \\ 1.5 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0.3 \\ 0.3 & 1 \end{smallmatrix}\right)\right)$, where the first element of the random vector $\mathbf{X_j}$ corresponded to the first time-point and the second element corresponded to the second time-point. Similarly, three time-varying auxiliary variables were generated for each cluster based on the value of $\mathbf{X_j}$: $\mathbf{Z_{1,j}} = \binom{Z_{1,1j}}{Z_{1,2j}} \sim N_2\left(\binom{0.2+0.2X_{1j}}{0.2+0.2X_{2j}}, \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)\right)$, $\mathbf{Z_{2,j}} = \binom{Z_{2,1j}}{Z_{2,2j}} \sim N_2\left(\binom{0.7+0.2X_{1j}}{0.7+0.2X_{2j}}, \left(\begin{smallmatrix} 1 & 0.1 \\ 0.1 & 1 \end{smallmatrix}\right)\right)$, and $Z_{3,ij} \sim Exp(mean = |0.7 + 0.2X_{ij}|)$. In addition, one time-invariant auxiliary variable was generated for each cluster: $Z_{4,1j} = Z_{4,2j} \sim Bernoulli(0.5)$. Two random intercepts, a_j (used to generate missingness R_{ij}) and b_j (used to generate the outcome Y_{ij}), were independently generated from a normal distribution with mean 0 and variance 1.

The proposed multilevel approach and comparison methods (available case analysis and the marginal approach of Scharfstein et al.[2]) were implemented in a similar manner to the simulation study presented in the main text, except that a logistic mixed effect model was fit as the working model for the outcome Y_{ij} using the proposed multilevel approach (i.e., $logit\{P(Y_{ij} = 1|\mathbf{Z_{ij}}, b_j)\} = \mathbf{Z_{ij}}\boldsymbol{\gamma} + b_j$), and an independent-data logistic regression model was fit as the working model for the outcome Y_{ij} using the marginal approach. Note that since both working models had a non-identity link function (e.g., logistic regression), the marginal working models were always misspecified, even when using the correct set of fixed effects, because R_{ij} and Y_{ij} were generated based on models conditional on a random intercept.

S3.2 Misspecification of Working Models by Omitting an Important Covariate

First, we considered the performance of the proposed method when either working model was misspecified by omitting an important covariate. The outcome variable Y_{ij} was generated from a Bernoulli distribution with probability $logit^{-1}(-1 + Z_{1,ij} - Z_{2,ij} + \gamma_3 * Z_{3,ij} + Z_{4,ij} - X_{ij} + b_j)$, where γ_3 equaled -0.2 (weak effect) or -1 (strong effect). An indicator that Y_{ij} was observed (R_{ij}) was generated from a Bernoulli distribution with probability $logit^{-1}(\alpha_0 - Z_{1,ij} + Z_{2,ij} + \alpha_3 * Z_{3,ij} - Z_{4,ij} + X_{ij} + a_j)$, where α_0 equaled 0.5 (20% missing) or -1 (35% missing), and α_3 equaled 0.2 (weak effect) or 1 (strong effect). For both the proposed multilevel approach and the marginal approach, each working model was either fit using the correct set of fixed effects, or by excluding $Z_{3,ij}$ from the model.

Table S1 presents bias, empirical standard deviation of the estimates (SDE), average estimated standard errors (ESE), mean square error (MSE), and coverage rates for 95% confidence intervals (CP) for the proposed multilevel approach. Table S2 presents ratios of the empirical variance and MSE for the multilevel approach to the available case and marginal approaches. The proposed multilevel approach exhibited essentially no bias when either the working model for $[\mathbf{R_j}, a_j|\mathbf{Z_j}]$ and/or the working model for $[\mathbf{Y_j}, b_j|\mathbf{Z_j}]$ were specified correctly, confirming the double robustness property. Bias for the multilevel approach tended to decrease as the percent missing decreased and as the magnitude of the omitted effect decreased. The 95% confidence interval coverage rates were also nearly at the nominal level when at least one working model was specified correctly. The proposed standard error estimator for the multilevel approach approximated the SDE well in most cases. Both the empirical variance and MSE were almost always smaller for the proposed multilevel approach than the marginal approach, although this improvement in the empirical variance and MSE for the proposed method compared to the marginal method was smaller than for the continuous outcome (results presented in the main text).

Table S1: Results from simulation study for the multilevel approach with a binary outcome where working models were misspecified by omitting an important covariate

				β_0					β_1				
Effect strength	% Miss.	R[a]	Y[a]	Bias	SDE	ESE	MSE	CP	Bias	SDE	ESE	MSE	CP
Weak	20	T	T	-0.008	0.112	0.113	0.013	95.1	0.005	0.070	0.069	0.005	94.7
		T	F	-0.009	0.112	0.113	0.013	95.5	0.005	0.070	0.069	0.005	94.8
		F	T	-0.008	0.112	0.113	0.013	95.2	0.005	0.070	0.069	0.005	94.9
		F	F	-0.013	0.112	0.113	0.013	95.5	0.005	0.070	0.069	0.005	94.9
	35	T	T	-0.008	0.162	0.152	0.026	94.6	0.003	0.097	0.091	0.009	94.6
		T	F	-0.009	0.162	0.153	0.026	94.6	0.003	0.097	0.091	0.009	94.8
		F	T	-0.008	0.162	0.152	0.026	95.0	0.003	0.097	0.090	0.009	94.6
		F	F	-0.016	0.163	0.153	0.027	95.0	0.002	0.097	0.091	0.009	94.4
Strong	20	T	T	-0.012	0.120	0.113	0.014	93.9	0.006	0.076	0.076	0.006	95.5
		T	F	-0.021	0.121	0.114	0.015	94.1	0.008	0.076	0.077	0.006	95.5
		F	T	-0.012	0.119	0.113	0.014	94.2	0.007	0.076	0.076	0.006	95.5
		F	F	-0.087	0.122	0.116	0.022	88.7	0.016	0.078	0.078	0.006	95.3
	35	T	T	-0.014	0.167	0.150	0.028	93.0	0.007	0.105	0.098	0.011	94.8
		T	F	-0.036	0.172	0.155	0.031	91.9	0.011	0.108	0.102	0.012	94.0
		F	T	-0.015	0.164	0.148	0.027	92.4	0.008	0.102	0.097	0.010	94.1
		F	F	-0.175	0.176	0.159	0.061	80.9	0.017	0.112	0.105	0.013	94.2

[a] T = Working model specified correctly. F = Working model misspecified by excluding the covariate $Z_{3,ij}$.

Table S2: Comparison of multilevel approach with the marginal approach and the available case approach from simulation study for binary outcome where working models were misspecified by omitting an important covariate

				β_0				β_1			
				Emp var ratio[b]		MSE ratio[b]		Emp var ratio[b]		MSE ratio[b]	
Effect strength	% Miss.	R[a]	Y[a]	Available case approach	Marginal approach	Available case approach	Marginal approach	Available case approach	Marginal approach	Available case approach	Marginal approach
Weak	20	T	T	1.043	0.984	0.088	0.983	1.148	0.979	0.423	0.978
		T	F	1.042	0.984	0.088	0.985	1.145	0.978	0.422	0.978
		F	T	1.048	0.986	0.088	0.986	1.152	0.980	0.424	0.980
		F	F	1.048	0.986	0.089	0.986	1.152	0.979	0.424	0.979
	35	T	T	1.313	0.977	0.067	0.976	1.390	0.942	0.484	0.941
		T	F	1.319	0.978	0.068	0.979	1.392	0.944	0.484	0.943
		F	T	1.318	0.969	0.067	0.969	1.396	0.939	0.485	0.938
		F	F	1.327	0.972	0.068	0.972	1.407	0.941	0.489	0.940
Strong	20	T	T	1.104	0.982	0.108	0.982	1.208	0.976	0.489	0.976
		T	F	1.118	0.994	0.111	1.017	1.216	0.983	0.494	0.989
		F	T	1.099	0.982	0.107	0.984	1.217	0.971	0.492	0.971
		F	F	1.154	0.988	0.168	1.002	1.292	0.978	0.540	0.983
	35	T	T	1.267	0.984	0.072	0.985	1.438	0.940	0.543	0.942
		T	F	1.344	1.000	0.079	1.040	1.521	0.956	0.578	0.963
		F	T	1.209	0.997	0.069	1.002	1.359	0.958	0.514	0.961
		F	F	1.396	1.008	0.157	1.042	1.636	0.969	0.629	0.980

[a] T = Working model specified correctly. F = Working model misspecified by excluding the covariate $Z_{3,ij}$.
[b] Ratio comparing multilevel approach to corresponding comparison method.

S3.3 Misspecification of Working Models by Omitting a Non-Linear Effect

We also considered the performance of the proposed method when either working model was misspecified by omitting a quadratic term. The outcome variable Y_{ij} was generated from a Bernoulli distribution with probability $logit^{-1}(-1 + Z_{1,ij} - Z_{2,ij} - Z_{3,ij} + Z_{4,ij} + \gamma_5 * Z_{2,ij}^2 - X_{ij} + b_j)$, where γ_5 equaled 0.1 (weak effect) or 0.5 (strong effect). An indicator that Y_{ij} was observed (R_{ij}) was generated from a Bernoulli distribution with probability $logit^{-1}(\alpha_0 - Z_{1,ij} + Z_{2,ij} + Z_{3,ij} - Z_{4,ij} + \alpha_5 * Z_{2,ij}^2 + X_{ij} + a_j)$, where α_0 equaled 0.5 (20% missing) or -1 (35% missing) and α_5 equaled -0.1 (weak effect) or -0.5 (strong effect). For both the proposed multilevel approach and the marginal approach, each working model was either fit using the correct set of fixed effects, or by excluding the quadratic term $Z_{2,ij}^2$ from the model.

Table S3 presents bias, SDE, ESE, MSE, and CP for the proposed multilevel approach. Table S4 presents ratios of the empirical variance and MSE for the multilevel approach to the available case and marginal approaches. These results were similar to the scenario based on omitting an important covariate from the working models (presented in Tables S1 and S2).

S4 Web Appendix D: Sample SAS and R Code

Here we present sample SAS and R code to implement the proposed method in statistical practice. Code is provided that corresponds to the situations explored in the simulation studies presented in this paper, where there are two-level data, each hierarchical working model includes a cluster-level random intercept, the working model for missingness is a logistic mixed effects model, the working model for the outcome of interest is either a linear or logistic mixed effects model, and all integrals are estimated using Gauss-Hermite quadrature. The code could be modified to accommodate more general situations (e.g., different types of working models, more than two levels of data). SAS code is provided to fit the hierarchical working models for the outcome variable and missingness using PROC GLIMMIX. An R function written by the authors is provided to solve the estimating equations for $\boldsymbol{\beta}$ (equation

Table S3: Results from simulation study for the multilevel approach with a binary outcome where working models were misspecified by omitting a quadratic term

				β_0					β_1				
Effect strength	% Miss.	R[a]	Y[a]	Bias	SDE	ESE	MSE	CP	Bias	SDE	ESE	MSE	CP
Weak	20	T	T	-0.013	0.116	0.114	0.014	95.2	0.006	0.075	0.075	0.006	94.4
		T	F	-0.014	0.117	0.114	0.014	94.9	0.007	0.075	0.075	0.006	94.2
		F	T	-0.013	0.115	0.114	0.013	94.9	0.006	0.074	0.075	0.006	94.4
		F	F	-0.016	0.116	0.114	0.014	94.9	0.007	0.074	0.075	0.006	94.1
	35	T	T	-0.006	0.149	0.152	0.022	96.5	-0.002	0.100	0.098	0.010	95.8
		T	F	-0.007	0.150	0.152	0.022	96.6	-0.002	0.101	0.098	0.010	95.6
		F	T	-0.006	0.146	0.150	0.021	96.6	-0.002	0.098	0.097	0.010	95.5
		F	F	-0.011	0.146	0.151	0.022	96.5	-0.001	0.099	0.097	0.010	95.6
Strong	20	T	T	-0.004	0.110	0.114	0.012	96.7	-0.001	0.066	0.068	0.004	95.7
		T	F	-0.017	0.115	0.120	0.013	95.0	0.001	0.069	0.072	0.005	96.0
		F	T	-0.004	0.108	0.113	0.012	96.6	-0.001	0.065	0.067	0.004	95.7
		F	F	-0.087	0.111	0.116	0.020	89.0	0.007	0.067	0.069	0.005	95.2
	35	T	T	0.006	0.159	0.155	0.025	94.8	-0.006	0.091	0.090	0.008	94.9
		T	F	-0.018	0.172	0.169	0.030	95.5	-0.003	0.101	0.098	0.010	94.9
		F	T	0.008	0.153	0.152	0.023	94.5	-0.008	0.087	0.088	0.008	95.2
		F	F	-0.110	0.160	0.157	0.038	90.8	-0.003	0.092	0.092	0.009	95.0

[a] T = Working model specified correctly. F = Working model misspecified by excluding the quadratic term $Z^2_{2,ij}$.

Table S4: Comparison of multilevel approach with the marginal approach and the available case approach from simulation study for binary outcome where working models were misspecified by omitting a quadratic term

				β_0				β_1			
				Emp var ratio[b]		MSE ratio[b]		Emp var ratio[b]		MSE ratio[b]	
Effect strength	% Miss.	R[a]	Y[a]	Available case approach	Marginal approach	Available case approach	Marginal approach	Available case approach	Marginal approach	Available case approach	Marginal approach
Weak	20	T	T	1.110	0.953	0.107	0.954	1.252	0.939	0.500	0.939
		T	F	1.118	0.957	0.108	0.959	1.258	0.941	0.503	0.942
		F	T	1.091	0.952	0.105	0.953	1.231	0.939	0.492	0.939
		F	F	1.097	0.955	0.106	0.957	1.236	0.940	0.494	0.941
	35	T	T	1.315	0.970	0.062	0.969	1.575	0.939	0.567	0.939
		T	F	1.319	0.966	0.062	0.966	1.578	0.938	0.568	0.938
		F	T	1.256	0.955	0.059	0.954	1.513	0.930	0.545	0.930
		F	F	1.262	0.953	0.060	0.953	1.522	0.930	0.548	0.930
Strong	20	T	T	1.032	0.972	0.080	0.972	1.093	0.956	0.479	0.956
		T	F	1.118	0.969	0.088	0.988	1.213	0.943	0.531	0.942
		F	T	0.993	0.976	0.077	0.976	1.061	0.963	0.464	0.963
		F	F	1.046	0.983	0.130	1.011	1.128	0.968	0.499	0.971
	35	T	T	1.231	0.985	0.063	0.985	1.288	0.975	0.609	0.975
		T	F	1.452	0.949	0.075	0.959	1.598	0.952	0.752	0.951
		F	T	1.145	0.982	0.059	0.982	1.185	0.965	0.562	0.965
		F	F	1.247	0.985	0.094	1.019	1.337	0.971	0.629	0.969

[a] T = Working model specified correctly. F = Working model misspecified by excluding the quadratic term $Z_{2,ij}^2$.

[b] Ratio comparing multilevel approach to corresponding comparison method.

(4) in the main text) and estimate the covariance matrix (based on equation (8) in the main text), using the estimated working model parameters.

SAS code for hierarchical working models. The following code fits the hierarchical working models for the outcome variable and missingness using PROC GLIMMIX. Let the dataset named `dat` have one row per data record (e.g., if there are two levels in the data, where $j = 1, ..., m$ indicates the cluster, and $i = 1, ..., n_j$ indicates the data record, then there is one record for each combination of i and j). The dataset `dat` has the following variables for each data record for each combination of subscripts i and j:

- `ID` is a unique identifier variable for each cluster (i.e., corresponding to subscript i)
- `Y` is the outcome of interest
- `R` is an indicator that `Y` is observed (i.e., non-missing)
- `X1,...,Xp` is the set of predictors of interest for `Y` (does not include an intercept)
- `Z1,...,Zq` is the set of covariates that will be included in the hierarchical working models (does not include an intercept, should include the predictor variables `X1,...,Xp`)

First, the following code fits a logistic mixed effects model for missingness, with a cluster-level random intercept.

```
proc glimmix data=dat method=quadrature(qpoints=25) noclprint;
        class ID;
        model R(event='1') = Z1 ... Zq / s dist=binary;
        random intercept / subject=ID;
        ods output ParameterEstimates=coefparm_r CovParms=covparm_r;
run;
```

Next, the following code fits a linear mixed effects model (e.g., for a continuous outcome) or a logistic mixed effects model (e.g., for a binary outcome) for the outcome of interest, with a cluster-level random intercept.

```
/* Linear mixed effects model (e.g., for a continuous outcome) */
proc glimmix data=dat method=quadrature(qpoints=25) noclprint;
        class ID;
        model Y = Z1 ... Zq / s dist=normal;
        random intercept / subject=ID;
        ods output ParameterEstimates=coefparm_y CovParms=covparm_y;
run;

/* Logistic mixed effects model (e.g., for a binary outcome) */
proc glimmix data=dat method=quadrature(qpoints=25) noclprint;
        class ID;
        model Y(event='1') = Z1 ... Zq / s dist=binary;
        random intercept / subject=ID;
        ods output ParameterEstimates=coefparm_y CovParms=covparm_y;
run;
```

Then, the following code combines the parameter estimates from both working models into the same dataset, prepares the variable names in the combined dataset to be used by the R function written by the authors to estimate $\hat{\boldsymbol{\beta}}_{\boldsymbol{m}}$ and the covariance matrix, and exports the parameter estimates dataset as a CSV data file that can be read into R.

```
/* define the correlation between the random effects from both */
/* working models as 0 (i.e., independent random effects) */
data rho;
        Parameter='rho'; Estimate=0; output;
run;

/* combine the parameter estimates from both working models, */
/* and adjust variable names to be used by R function */
data parms(keep=parameter estimate);
```

```
        length Parameter $ 50;
        set coefparm_r(keep=effect estimate in=coefr)
        coefparm_y(keep=effect estimate in=coefy)
        covparm_r(keep=covparm estimate in=covr)
        covparm_y(keep=covparm estimate in=covy)
        rho;
        if covparm='Residual' then parameter='sigma';
        if covparm='Intercept' then do;
                if covr then parameter='tau';
                if covy then parameter='phi';
        end;
        if coefr then parameter=cats('r_',strip(effect));
        if coefy then parameter=cats('y_',strip(effect));
run;

/* export parameter estimates dataset as a CSV data file */
proc export data=parms
            outfile="path/parms.csv"
            dbms=csv
            replace;
run;
```

R code to prepare data. The following code loads a CSV dataset containing the parameter estimates from the hierarchical working models that were estimated using SAS, and saves the variables that will be used to estimate the doubly robust regression coefficients and covariance matrix.

```
# LOAD PARAMETER ESTIMATES DATASET FROM SAS
parms=read.csv("path/parms.csv",header=TRUE)

# ASSIGN VARIABLES THAT WILL BE NEEDED FOR DOUBLY ROBUST ANALYSIS
```

```
n=nrow(dat)                    # number of data records
m=length(unique(dat$ID))       # number of clusters
nj=table(dat$ID)               # number of data records for each cluster
y=dat$Y
x=as.matrix(cbind(rep(1,times=n),dat[,c('X1',...,'Xp')]))
colnames(x)=c('intercept','X1',...,'Xp')
z.r=as.matrix(cbind(dat[,c('Z1',...,'Zq')]))
z.y=as.matrix(cbind(dat[,c('Z1',...,'Zq')]))

alpha=parms$Estimate[substr(parms$Parameter,1,2)=='r_']
gamma=parms$Estimate[substr(parms$Parameter,1,2)=='y_']
sigma=parms$Estimate[parms$Parameter=='sigma']
phi=parms$Estimate[parms$Parameter=='phi']
tau=parms$Estimate[parms$Parameter=='tau']
```

R function to solve the estimating equations for β and estimate the covariance matrix. The authors have written an R function that uses the methods described in this paper to estimate the doubly robust regression coefficients of interest, and estimate the covariance matrix for this doubly robust estimator. The R function, called `DoublyRobust_Multilevel.R`, is provided as supporting information for this paper. The following code calls the function `DoublyRobust_Multilevel` to obtain doubly robust regression coefficent and covariance matrix estimates, for the case where a linear mixed effects model is used for the working model for the outcome variable, and an identity link is used for $\mu(\cdot)$ in the estimating equations in (4) in the main text (e.g., for a continuous outcome).

```
DR_Results=DoublyRobust_Multilevel(n,m,nj,y,x,z.r,z.y,qpoints=25,
                                   alpha,gamma,sigma,tau,phi,
                                   dist='gaussian',link='identity',
                                   conv=.0001,maxiter=50,maxpiinv=100,
                                   se=TRUE,verbose=FALSE)
```

```
# Doubly robust regression estimates for beta
DR_Results$beta

# Standard errors for beta
sqrt(diag(DR_Results$var.beta))
```

In addition, the following code calls the function `DoublyRobust_Multilevel` for the case where a logistic mixed effects model is used for the working model for the outcome variable, and a logit link is used for $\mu(\cdot)$ in the estimating equations in (4) in the main text (e.g., for a binary outcome).

```
DR_Results=DoublyRobust_Multilevel(n,m,nj,y,x,z.r,z.y,qpoints=25,
                                   alpha,gamma,sigma=NULL,tau,phi,
                                   dist='binomial',link='logit',
                                   conv=.0001,maxiter=50,maxpiinv=100,
                                   se=TRUE,verbose=FALSE)

# Doubly robust regression estimates for beta
DR_Results$beta

# Standard errors for beta
sqrt(diag(DR_Results$var.beta))
```

References

[1] Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* Hoboken, NJ: John Wiley & Sons, Inc. 2nd ed. 2002.

[2] Scharfstein DO, Rotnitzky A, Robins JM. Rejoinder to adjusting for non-ignorable drop-out using semiparametric non-response models. *Journal of the American Statistical Association* 1999; 94(448): 1135-1146.