Web Appendix for
Sample size calculation for randomized trials via inverse probability of response weighting when outcome data are missing at random

Linda J Harrison and Rui Wang

WEB APPENDIX A: DERIVATION OF EQUATIONS (1), (2), (5), (10), (11) AND (12), AND DEFINITION OF $\tau_{\text{C-KNOWN}}$ AND $\tau_{\text{C-APPROX}}$

Derivation of Equation (1)

Let Y_i represent the outcome for participant i and Z_i indicate the randomized intervention, such that,

$$Z_i = \begin{cases} 1 & \text{if participant } i \text{ is randomized to the intervention group} \\ 0 & \text{if participant } i \text{ is randomized to the control group} \end{cases}$$

Suppose that $g\left\{E(Y_i|Z_i=1)\right\} = g(\mu_1) = \lambda_1$ and $g\left\{E(Y_i|Z_i=0)\right\} = g(\mu_0) = \lambda_0$, where μ_1 is the population mean of the outcome Y_i under intervention and μ_0 is the population mean of the outcome Y_i under control. Define $\boldsymbol{\theta} = \left(\lambda_1, \lambda_0\right)^{\intercal}$ and consider g to be the identity or logit function. An M-estimator $\hat{\boldsymbol{\theta}}$[1] solves,

$$\sum_{i=1}^{n} \boldsymbol{u}_i(Y_i, Z_i; \boldsymbol{\theta}) = \mathbf{0}$$

Taking,

$$\boldsymbol{u}_i(Y_i, Z_i; \boldsymbol{\theta}) = \begin{pmatrix} Z_i(Y_i-\mu_1) \\ (1-Z_i)(Y_i-\mu_0) \end{pmatrix}$$

we will obtain the following estimates of the mean of the outcome in the two groups of the randomized trial:

$$\hat{\mu}_1 = \left\{\sum_{i=1}^{n} Z_i\right\}^{-1}\left\{\sum_{i=1}^{n} Z_iY_i\right\},\ \hat{\mu}_0 = \left\{\sum_{i=1}^{n}(1-Z_i)\right\}^{-1}\left\{\sum_{i=1}^{n}(1-Z_i)Y_i\right\}$$

Assuming our primary interest is the contrast between randomized intervention groups defined by $g(\mu_1)-g(\mu_0) = \lambda_1-\lambda_0$, from M-estimation theory we know,

$$\text{var}(\hat{\boldsymbol{\theta}}) = n^{-1}\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^{\intercal}$$

where

$$\mathcal{A} = \mathbb{E}\left(-\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\theta}^{\intercal}}\right) = \mathbb{E}\begin{pmatrix} Z_i\frac{\partial\mu_1}{\partial\lambda_1} & 0 \\ 0 & (1-Z_i)\frac{\partial\mu_0}{\partial\lambda_0} \end{pmatrix}$$

$$\mathcal{B} = \mathbb{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i^{\intercal}\right) = \mathbb{E}\begin{pmatrix} Z_i(Y_i-\mu_1)^2 & 0 \\ 0 & (1-Z_i)(Y_i-\mu_0)^2 \end{pmatrix}$$

Therefore,

$$\text{var}\{(\hat{\lambda}_1, \hat{\lambda}_0)\} = \frac{1}{n}\begin{pmatrix} \kappa^{-1}\mathbb{E}\{(Y_i-\mu_1)^2|Z_i=1\}\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2} & 0 \\ 0 & (1-\kappa)^{-1}\mathbb{E}\{(Y_i-\mu_0)^2|Z_i=0\}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2} \end{pmatrix}$$

leading to

$$\text{var}(\hat{\lambda}_1-\hat{\lambda}_0) = n^{-1}\left[\kappa^{-1}\mathbb{E}\{(Y_i-\mu_1)^2|Z_i=1\}\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2} + (1-\kappa)^{-1}\mathbb{E}\{(Y_i-\mu_0)^2|Z_i=0\}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\right]$$

where $\kappa = E(Z_i) = P(Z_i=1)$.
When Y_i is a continuous outcome with $\mathbb{E}\{(Y_i-\mu_1)^2|Z_i=1\} = \mathbb{E}\{(Y_i-\mu_0)^2|Z_i=0\} = \sigma_y^2$ and g is the identity function,

$$\text{var}(\hat{\mu}_1-\hat{\mu}_0) = \text{var}(\hat{\lambda}_1-\hat{\lambda}_0) = n^{-1}\left\{\kappa(1-\kappa)\right\}^{-1}\sigma_y^2$$

When Y_i is a binary outcome and g is the identity function,

$$\text{var}(\hat{\mu}_1-\hat{\mu}_0) = \text{var}(\hat{\lambda}_1-\hat{\lambda}_0) = n^{-1}\left\{\kappa^{-1}\mu_1(1-\mu_1) + (1-\kappa)^{-1}\mu_0(1-\mu_0)\right\}$$

When Y_i is a binary outcome and g is the logit function,

$$\operatorname{var}\{\operatorname{logit}(\hat{\mu}_1) - \operatorname{logit}(\hat{\mu}_0)\} = \operatorname{var}(\hat{\lambda}_1 - \hat{\lambda}_0) = n^{-1}\left[\{\kappa\mu_1(1-\mu_1)\}^{-1} + \{(1-\kappa)\mu_0(1-\mu_0)\}^{-1}\right]$$

The power of the Wald test for the contrast of primary interest is approximated by,

$$1 - \beta = \Phi\left[\frac{g(\mu_1) - g(\mu_0)}{\sqrt{\operatorname{var}\{g(\hat{\mu}_1) - g(\hat{\mu}_0)\}}} - z_{1-\alpha/2}\right]$$

Therefore, the number of participants needed to be observed is,

$$n_{\text{complete}} = \frac{\tau\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}{\left\{g(\mu_1) - g(\mu_0)\right\}^2}$$

where

$$\tau = \begin{cases} \{\kappa(1-\kappa)\}^{-1}\sigma_y^2 & \text{if } Y_i \text{ continuous and } g \text{ identity} \\ \kappa^{-1}\mu_1(1-\mu_1) + (1-\kappa)^{-1}\mu_0(1-\mu_0) & \text{if } Y_i \text{ binary and } g \text{ identity} \\ \{\kappa\mu_1(1-\mu_1)\}^{-1} + \{(1-\kappa)\mu_0(1-\mu_0)\}^{-1} & \text{if } Y_i \text{ binary and } g \text{ logit} \end{cases}$$

These three sample size formulas have previously been derived; see for example Chow et al[2] Sections 3.2, 4.2 and 4.6. The formulas are rederived in this Web Appendix to unify notation within the M-estimation framework.

Derivation of Equation (2)

Define R_i to be an indicator of whether the outcome Y_i is observed, as follows,

$$R_i = \begin{cases} 1 & \text{if } Y_i \text{ is observed} \\ 0 & \text{if } Y_i \text{ is missing} \end{cases}$$

The naive complete-case analysis leads to the following estimating equations,

$$\sum_{i=1}^{n} \boldsymbol{u}_i(Y_i, R_i, Z_i; \boldsymbol{\theta}) = \boldsymbol{0}$$

$$\boldsymbol{u}_i(Y_i, R_i, Z_i; \boldsymbol{\theta}) = \begin{pmatrix} R_i Z_i (Y_i - \mu_1) \\ R_i(1 - Z_i)(Y_i - \mu_0) \end{pmatrix}$$

and estimators of the mean of the outcome in the two groups of the randomized trial, as follows,

$$\hat{\mu}_1 = \left\{\sum_{i=1}^{n} R_i Z_i\right\}^{-1}\left\{\sum_{i=1}^{n} R_i Z_i Y_i\right\},\ \hat{\mu}_0 = \left\{\sum_{i=1}^{n} R_i(1 - Z_i)\right\}^{-1}\left\{\sum_{i=1}^{n} R_i(1 - Z_i)Y_i\right\}$$

Again, from M-estimation theory,

$$\operatorname{var}(\hat{\boldsymbol{\theta}}) = n^{-1}\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^{\intercal}$$

where

$$\mathcal{A} = \mathbb{E}\left(-\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\theta}^{\intercal}}\right) = \mathbb{E}\begin{pmatrix} R_i Z_i \frac{\partial \mu_1}{\partial \lambda_1} & 0 \\ 0 & R_i(1 - Z_i)\frac{\partial \mu_0}{\partial \lambda_0} \end{pmatrix}$$

$$\mathcal{B} = \mathbb{E}\left(\boldsymbol{u}_i \boldsymbol{u}_i^{\intercal}\right) = \mathbb{E}\begin{pmatrix} R_i Z_i (Y_i - \mu_1)^2 & 0 \\ 0 & R_i(1 - Z_i)(Y_i - \mu_0)^2 \end{pmatrix}$$

Assuming the probability of the outcome being observed is the same in the intervention and control group, we have $\mathbb{E}(R_i Z_i) = \mathbb{E}(R_i|Z_i = 1)P(Z_i = 1) = \phi\kappa$ and $\mathbb{E}\{R_i(1 - Z_i)\} = \mathbb{E}(R_i|Z_i = 0)P(Z_i = 1) = \phi(1 - \kappa)$ where $\phi = P(R_i = 1) = \mathbb{E}(R_i|Z_i = 1) = \mathbb{E}(R_i|Z_i = 0)$. Therefore,

$$\operatorname{var}(\hat{\lambda}_1 - \hat{\lambda}_0) = n^{-1}\left[\phi^{-2}\kappa^{-1}\mathbb{E}\{R_i(Y_i - \mu_1)^2|Z_i = 1\}\left(\frac{\partial \mu_1}{\partial \lambda_1}\right)^{-2} + \phi^{-2}(1 - \kappa)^{-1}\mathbb{E}\{R_i(Y_i - \mu_0)^2|Z_i = 0\}\left(\frac{\partial \mu_0}{\partial \lambda_0}\right)^{-2}\right]$$

Further assuming $R_i \perp\!\!\!\perp Y_i|Z_i$, we have $\mathbb{E}\{R_i(Y_i - \mu_1)^2|Z_i = 1\} = \mathbb{E}(R_i|Z_i = 1)\mathbb{E}\{(Y_i - \mu_1)^2|Z_i = 1\} = \phi\mathbb{E}\{(Y_i - \mu_1)^2|Z_i = 1\}$ and $\mathbb{E}\{R_i(Y_i - \mu_0)^2|Z_i = 0\} = \mathbb{E}(R_i|Z_i = 0)\mathbb{E}\{(Y_i - \mu_0)^2|Z_i = 0\} = \phi\mathbb{E}\{(Y_i - \mu_0)^2|Z_i = 0\}$, so

$$\operatorname{var}(\hat{\lambda}_1 - \hat{\lambda}_0) = (n\phi)^{-1}\left[\kappa^{-1}\mathbb{E}\{(Y_i - \mu_1)^2|Z_i = 1\}\left(\frac{\partial \mu_1}{\partial \lambda_1}\right)^{-2} + (1 - \kappa)^{-1}\mathbb{E}\{(Y_i - \mu_0)^2|Z_i = 0\}\left(\frac{\partial \mu_0}{\partial \lambda_0}\right)^{-2}\right]$$

The two assumptions imply $R_i \perp\!\!\!\perp Z_i$ and $R_i \perp\!\!\!\perp Y_i$, so outcome data are missing completely at random (MCAR)[3]. Therefore, the number of participants needed to be recruited is[4,5],

$$n_{\text{standard}} = \frac{\tau\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}{\phi\left\{g(\mu_1) - g(\mu_0)\right\}^2} = \frac{\tau_{\text{standard}}\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}{\left\{g(\mu_1) - g(\mu_0)\right\}^2}$$

where

$$\tau_{\text{standard}} = \begin{cases} \phi^{-1}\left\{\kappa(1-\kappa)\right\}^{-1}\sigma_y^2 & \text{if } Y_i \text{ continuous and } g \text{ identity} \\ \phi^{-1}\left\{\kappa^{-1}\mu_1(1-\mu_1) + (1-\kappa)^{-1}\mu_0(1-\mu_0)\right\} & \text{if } Y_i \text{ binary and } g \text{ identity} \\ \phi^{-1}\left[\{\kappa\mu_1(1-\mu_1)\}^{-1} + \{(1-\kappa)\mu_0(1-\mu_0)\}^{-1}\right] & \text{if } Y_i \text{ binary and } g \text{ logit} \end{cases}$$

Derivation of Equation (5)

Let $\boldsymbol{X}_i$ represent a vector of fully observed baseline covariates. Suppose that $\text{logit}\{P(R_i = 1|\boldsymbol{X}_i, Z_i = 1)\} = \text{logit}(e_{1i}) = \boldsymbol{X}_i^\intercal \boldsymbol{\beta}_1$ and $\text{logit}\{P(R_i = 1|\boldsymbol{X}_i, Z_i = 0)\} = \text{logit}(e_{0i}) = \boldsymbol{X}_i^\intercal \boldsymbol{\beta}_0$. Now defining $\boldsymbol{\theta} = \left(\lambda_1, \lambda_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_0\right)^\intercal$ and assuming the outcome Y_i is missing at random, i.e. $R_i \perp\!\!\!\perp Y_i|(\boldsymbol{X}_i, Z_i)$, leads to the following estimating equations,

$$\sum_{i=1}^{n} \boldsymbol{u}_i(Y_i, R_i, Z_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = \boldsymbol{0}$$

$$\boldsymbol{u}_i(Y_i, R_i, Z_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = \begin{pmatrix} R_i Z_i e_{1i}^{-1}(Y_i - \mu_1) \\ R_i(1 - Z_i)e_{0i}^{-1}(Y_i - \mu_0) \\ Z_i \boldsymbol{X}_i (R_i - e_{1i}) \\ (1 - Z_i)\boldsymbol{X}_i(R_i - e_{0i}) \end{pmatrix}$$

The latter two estimating equations follow from the score equations, $\sum_{i=1}^n \boldsymbol{D}_{1i}^\intercal V_{1i}^{-1}(R_i - e_{1i})$ and $\sum_{i=1}^n \boldsymbol{D}_{0i}^\intercal V_{0i}^{-1}(R_i - e_{0i})$ from logistic regression models where, $\boldsymbol{D}_{1i} = \frac{\partial e_{1i}}{\partial \boldsymbol{\beta}_1} = e_{1i}(1 - e_{1i})Z_i\boldsymbol{X}_i^\intercal$, $V_{1i} = e_{1i}(1 - e_{1i})$, $\boldsymbol{D}_{0i} = \frac{\partial e_{0i}}{\partial \boldsymbol{\beta}_0} = e_{0i}(1 - e_{0i})(1 - Z_i)\boldsymbol{X}_i^\intercal$, $V_{0i} = e_{0i}(1 - e_{0i})$. The Hájek ratio estimators of the mean of the outcome in the two groups of the randomized trial are, as follows,

$$\hat{\mu}_1 = \left\{\sum_{i=1}^n R_i Z_i \hat{e}_{1i}^{-1}\right\}^{-1}\left\{\sum_{i=1}^n R_i Z_i \hat{e}_{1i}^{-1} Y_i\right\}, \; \hat{\mu}_0 = \left\{\sum_{i=1}^n R_i (1 - Z_i) \hat{e}_{0i}^{-1}\right\}^{-1}\left\{\sum_{i=1}^n R_i (1 - Z_i) \hat{e}_{0i}^{-1} Y_i\right\}$$

Again, from M-estimation theory,

$$\text{var}(\hat{\boldsymbol{\theta}}) = n^{-1}\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^\intercal$$

where $\mathcal{A} = \mathbb{E}\left(-\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\theta}^\intercal}\right)$ and $\mathcal{B} = \mathbb{E}\left(\boldsymbol{u}_i \boldsymbol{u}_i^\intercal\right)$.

The expression for $\mathcal{A}$ is:

$$\mathbb{E}\begin{pmatrix} R_i Z_i e_{1i}^{-1}\frac{\partial \mu_1}{\partial \lambda_1} & 0 & R_i Z_i (Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})\boldsymbol{X}_i^\intercal & \boldsymbol{0}^\intercal \\ 0 & R_i(1 - Z_i)e_{0i}^{-1}\frac{\partial \mu_0}{\partial \lambda_0} & \boldsymbol{0}^\intercal & R_i(1 - Z_i)(Y_i - \mu_0)e_{0i}^{-1}(1 - e_{0i})\boldsymbol{X}_i^\intercal \\ \boldsymbol{0} & \boldsymbol{0} & Z_i e_{1i}(1 - e_{1i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal & 0 \\ \boldsymbol{0} & \boldsymbol{0} & 0 & (1 - Z_i)e_{0i}(1 - e_{0i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal \end{pmatrix}$$

The expression for $\mathcal{B}$ is:

$$\mathbb{E}\begin{pmatrix} R_i Z_i e_{1i}^{-2}(Y_i - \mu_1)^2 & 0 & R_i Z_i (Y_i - \mu_1)e_{1i}^{-1}(R_i - e_{1i})\boldsymbol{X}_i^\intercal & \boldsymbol{0}^\intercal \\ 0 & R_i(1 - Z_i)e_{0i}^{-2}(Y_i - \mu_0)^2 & \boldsymbol{0}^\intercal & R_i(1 - Z_i)(Y_i - \mu_0)e_{0i}^{-1}(R_i - e_{0i})\boldsymbol{X}_i^\intercal \\ R_i Z_i (Y_i - \mu_1)e_{1i}^{-1}(R_i - e_{1i})\boldsymbol{X}_i & \boldsymbol{0} & Z_i(R_i - e_{1i})^2\boldsymbol{X}_i\boldsymbol{X}_i^\intercal & 0 \\ \boldsymbol{0} & R_i(1 - Z_i)(Y_i - \mu_0)e_{0i}^{-1}(R_i - e_{0i})\boldsymbol{X}_i & 0 & (1 - Z_i)(R_i - e_{0i})^2\boldsymbol{X}_i\boldsymbol{X}_i^\intercal \end{pmatrix}$$

Considering the parameters of interest, $\text{var}\{(\hat{\lambda}_1, \hat{\lambda}_0)\}$ can be simplified to,

$$\frac{1}{n}\begin{pmatrix} \kappa^{-1}\left[\mathbb{E}\{R_i e_{1i}^{-2}(Y_i - \mu_1)^2|Z_i = 1\} - A\right]\left(\frac{\partial \mu_1}{\partial \lambda_1}\right)^{-2} & 0 \\ 0 & (1 - \kappa)^{-1}\left[\mathbb{E}\{R_i e_{0i}^{-2}(Y_i - \mu_0)^2|Z_i = 0\} - B\right]\left(\frac{\partial \mu_0}{\partial \lambda_0}\right)^{-2} \end{pmatrix}$$

where

$$\begin{aligned} A =& \mathbb{E}\{R_i(Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})\boldsymbol{X}_i^\intercal|Z_i = 1\}\left[\mathbb{E}\{e_{1i}(1 - e_{1i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i = 1\}\right]^{-1}\mathbb{E}\{R_i(Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})\boldsymbol{X}_i|Z_i = 1\} \\ B =& \mathbb{E}\{R_i(Y_i - \mu_0)e_{0i}^{-1}(1 - e_{0i})\boldsymbol{X}_i^\intercal|Z_i = 0\}\left[\mathbb{E}\{e_{0i}(1 - e_{0i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i = 0\}\right]^{-1}\mathbb{E}\{R_i(Y_i - \mu_0)e_{0i}^{-1}(1 - e_{0i})\boldsymbol{X}_i|Z_i = 0\} \end{aligned}$$

Therefore,

$$\begin{aligned}\operatorname{var}\{g(\hat{\mu}_1)-g(\hat{\mu}_0)\} &= n^{-1}\left(\kappa^{-1}\left[\mathbb{E}\{R_i e_{1i}^{-2}(Y_i-\mu_1)^2|Z_i=1\}-A\right]\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2}\right.\\ &\qquad\left.+(1-\kappa)^{-1}\left[\mathbb{E}\{R_i e_{0i}^{-2}(Y_i-\mu_0)^2|Z_i=0\}-B\right]\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\right)\\ &= n^{-1}\left(\kappa^{-1}\left[\mathbb{E}\{e_{1i}^{-1}(Y_i-\mu_1)^2|Z_i=1\}-A\right]\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2}\right.\\ &\qquad\left.+(1-\kappa)^{-1}\left[\mathbb{E}\{e_{0i}^{-1}(Y_i-\mu_0)^2|Z_i=0\}-B\right]\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\right)\end{aligned}$$

Derivation of Equation (10)

Let there be $k=1,\dots,K$ independent clusters each with $i=1,\dots,m$ participants. Let Y_{ki} represent the outcome in cluster k for participant i and Z_k indicate the randomized intervention for cluster k, such that,

$$Z_k=\begin{cases}1 & \text{if cluster } k \text{ is randomized to the intervention group}\\ 0 & \text{if cluster } k \text{ is randomized to the control group}\end{cases}$$

Suppose that $g\left\{E(Y_{ki}|Z_k=1)\right\}=g(\mu_1)=\lambda_1$ and $g\left\{E(Y_{ki}|Z_k=0)\right\}=g(\mu_0)=\lambda_0$, where μ_1 is the population mean of the outcome Y_{ki} under intervention and μ_0 is the population mean of the outcome Y_{ki} under control. Define $\boldsymbol{\theta}=\left(\lambda_1,\lambda_0\right)^{\intercal}$ and consider g to be the identity or logit function. Adopting a working independence correlation structure for estimating the means in each randomized group, an M-estimator $\hat{\boldsymbol{\theta}}^1$ solves,

$$\sum_{k=1}^{K}\boldsymbol{u}_k(Y_{ki},Z_k;\boldsymbol{\theta})=\boldsymbol{0}$$

with,

$$\boldsymbol{u}_k(Y_{ki},Z_k;\boldsymbol{\theta})=\begin{pmatrix}\boldsymbol{Z}_k^{\intercal}(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)\\ (1-\boldsymbol{Z}_k^{\intercal})(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^{m}Z_k(Y_{ki}-\mu_1)\\ \sum_{i=1}^{m}(1-Z_k)(Y_{ki}-\mu_0)\end{pmatrix}$$

where $\boldsymbol{Z}_k^{\intercal}=Z_k 1_m^{\intercal}$ with 1_m a vector of ones of length m, $\boldsymbol{Y}_k=(Y_{k1},\dots,Y_{km})^{\intercal}$ the outcome vector for cluster k, and $\boldsymbol{\mu}_k=\mu_1 1_m$ if cluster k is randomized to intervention and $\boldsymbol{\mu}_k=\mu_0 1_m$ if cluster k is randomized to control.

Assuming the true underlying correlation structure is exchangeable, from M-estimation theory we can derive,

$$\operatorname{var}(\hat{\boldsymbol{\theta}})=K^{-1}\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^{\intercal}$$

where

$$\mathcal{A}=\mathbb{E}\left(-\frac{\partial\boldsymbol{u}_k}{\partial\boldsymbol{\theta}^{\intercal}}\right)=m\mathbb{E}\begin{pmatrix}Z_k\frac{\partial\mu_1}{\partial\lambda_1} & 0\\ 0 & (1-Z_k)\frac{\partial\mu_0}{\partial\lambda_0}\end{pmatrix}$$

$$\begin{aligned}\mathcal{B}=\mathbb{E}\left(\boldsymbol{u}_k\boldsymbol{u}_k^{\intercal}\right) &=\mathbb{E}\begin{pmatrix}\left\{\sum_{i=1}^{m}Z_k(Y_{ki}-\mu_1)\right\}^2 & 0\\ 0 & \left\{\sum_{i=1}^{m}(1-Z_k)(Y_{ki}-\mu_0)\right\}^2\end{pmatrix}\\ &=m\{1+(m-1)\delta\}\begin{pmatrix}\mathbb{E}\{Z_k(Y_{ki}-\mu_1)^2\} & 0\\ 0 & \mathbb{E}\{(1-Z_k)(Y_{ki}-\mu_0)^2\}\end{pmatrix}\end{aligned}$$

with $\delta=\operatorname{corr}(Y_{ki},Y_{kj}|Z_k=1)=\operatorname{corr}(Y_{ki},Y_{kj}|Z_k=0)$ for $i\neq j$.

Therefore,

$$\operatorname{var}\{(\hat{\lambda}_1,\hat{\lambda}_0)\}=(mK)^{-1}\{1+(m-1)\delta\}\begin{pmatrix}\kappa^{-1}\mathbb{E}\{(Y_{ki}-\mu_1)^2|Z_k=1\}\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2} & 0\\ 0 & (1-\kappa)^{-1}\mathbb{E}\{(Y_{ki}-\mu_0)^2|Z_k=0\}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\end{pmatrix}$$

leading to

$$\mathrm{var}(\hat{\lambda}_1 - \hat{\lambda}_0) = (mK)^{-1}\{1 + (m-1)\delta\} \left[\kappa^{-1}\mathbb{E}\{(Y_{ki} - \mu_1)^2 | Z_k = 1\} \left(\frac{\partial \mu_1}{\partial \lambda_1}\right)^{-2} + (1-\kappa)^{-1}\mathbb{E}\{(Y_{ki} - \mu_0)^2 | Z_k = 0\} \left(\frac{\partial \mu_0}{\partial \lambda_0}\right)^{-2} \right]$$

where $\kappa = E(Z_k) = P(Z_k = 1)$.
So, the number of participants needed to be observed is[6],

$$n_{\mathrm{C}} = \frac{\tau_{\mathrm{C}} \left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}{\left\{g(\mu_1) - g(\mu_0)\right\}^2}$$

where

$$\tau_{\mathrm{C}} = \begin{cases} \{1 + (m-1)\delta\}\,\{\kappa(1-\kappa)\}^{-1}\,\sigma_y^2 & \text{if } Y_{ki} \text{ continuous and } g \text{ identity} \\ \{1 + (m-1)\delta\}\{\kappa^{-1}\mu_1(1-\mu_1) + (1-\kappa)^{-1}\mu_0(1-\mu_0)\} & \text{if } Y_{ki} \text{ binary and } g \text{ identity} \\ \{1 + (m-1)\delta\}\left[\kappa\{\mu_1(1-\mu_1)\}^{-1} + \{(1-\kappa)\mu_0(1-\mu_0)\}^{-1}\right] & \text{if } Y_{ki} \text{ binary and } g \text{ logit} \end{cases}$$

and for Y_{ki} continuous $\sigma_y^2 = \mathbb{E}\left\{(Y_{ki} - \mu_1)^2 | Z_k = 1\right\} = \mathbb{E}\left\{(Y_{ki} - \mu_0)^2 | Z_k = 0\right\}$.

Derivation of Equation (11)

Define R_{ki} to be an indicator of whether the outcome Y_{ki} is observed, as follows,

$$R_{ki} = \begin{cases} 1 & \text{if } Y_{ki} \text{ is observed} \\ 0 & \text{if } Y_{ki} \text{ is missing} \end{cases}$$

The naive complete-case analysis leads to the following estimating equations,

$$\sum_{k=1}^{K} \boldsymbol{u}_k(Y_{ki}, R_{ki}, Z_k; \boldsymbol{\theta}) = \mathbf{0}$$

$$\boldsymbol{u}_k(Y_{ki}, R_{ki}, Z_k; \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{Z}_k^{\intercal} \boldsymbol{R}_k(\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \\ (1 - \boldsymbol{Z}_k^{\intercal})\boldsymbol{R}_k(\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} R_{ki} Z_k (Y_{ki} - \mu_1) \\ \sum_{i=1}^{m} R_{ki}(1 - Z_k)(Y_{ki} - \mu_0) \end{pmatrix}$$

where $\boldsymbol{R}_k = \mathrm{diag}(R_{ki})$.
Again, from M-estimation theory,

$$\mathrm{var}(\hat{\boldsymbol{\theta}}) = K^{-1}\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^{\intercal}$$

where

$$\mathcal{A} = \mathbb{E}\left(-\frac{\partial \boldsymbol{u}_k}{\partial \boldsymbol{\theta}^{\intercal}}\right) = m\mathbb{E}\begin{pmatrix} R_{ki} Z_k \frac{\partial \mu_1}{\partial \lambda_1} & 0 \\ 0 & R_{ki}(1 - Z_k)\frac{\partial \mu_0}{\partial \lambda_0} \end{pmatrix} = m\phi \begin{pmatrix} \kappa \frac{\partial \mu_1}{\partial \lambda_1} & 0 \\ 0 & (1-\kappa)\frac{\partial \mu_0}{\partial \lambda_0} \end{pmatrix}$$

$$\begin{aligned} \mathcal{B} = \mathbb{E}\left(\boldsymbol{u}_k \boldsymbol{u}_k^{\intercal}\right) &= \mathbb{E}\begin{pmatrix} \left\{\sum_{i=1}^{m} R_{ki} Z_k (Y_{ki} - \mu_1)\right\}^2 & 0 \\ 0 & \left\{\sum_{i=1}^{m} R_{ki}(1 - Z_k)(Y_{ki} - \mu_0)\right\}^2 \end{pmatrix} \\ &= m\phi\{1 + (m-1)\phi\delta\} \begin{pmatrix} \kappa\mathbb{E}\{(Y_{ki} - \mu_1)^2 | Z_k = 1\} & 0 \\ 0 & (1-\kappa)\mathbb{E}\{(Y_{ki} - \mu_0)^2 | Z_k = 0\} \end{pmatrix} \end{aligned}$$

with $\kappa = E(Z_i)$, $\delta = \mathrm{corr}(Y_{ki}, Y_{kj} | Z_k = 1) = \mathrm{corr}(Y_{ki}, Y_{kj} | Z_k = 0)$, and assuming outcome data are MCAR, i.e. $\phi = P(R_{ki} = 1) = P(R_{ki} = 1 | Z_k = 1) = P(R_{ki} = 1 | Z_k = 0)$, $R_{ki} \perp\!\!\!\perp Y_{ki}$ and $R_{ki} \perp\!\!\!\perp R_{kj}$ for $i \neq j$.
Therefore,

$$\mathrm{var}(\hat{\lambda}_1 - \hat{\lambda}_0) = (mK\phi)^{-1}\{1 + (m-1)\phi\delta\} \left[\kappa^{-1}\mathbb{E}\{(Y_{ki} - \mu_1)^2 | Z_k = 1\} \left(\frac{\partial \mu_1}{\partial \lambda_1}\right)^{-2} \right.$$

$$+ (1-\kappa)^{-1}\mathbb{E}\{(Y_{ki}-\mu_0)^2|Z_k=0\}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\Bigg]$$

and, the number of participants needed to be recruited is,

$$n_{\text{C-standard}} = \frac{\tau_{\text{C-standard}}\left(z_{1-\beta}+z_{1-\alpha/2}\right)^2}{\left\{g(\mu_1)-g(\mu_0)\right\}^2}$$

where

$$\tau_{\text{C-standard}} = \begin{cases} \{\phi^{-1}+(m-1)\delta\}\left\{\kappa(1-\kappa)\right\}^{-1}\sigma_y^2 & \text{if } Y_{ki} \text{ continuous and } g \text{ identity} \\ \{\phi^{-1}+(m-1)\delta\}\{\kappa^{-1}\mu_1(1-\mu_1)+(1-\kappa)^{-1}\mu_0(1-\mu_0)\} & \text{if } Y_{ki} \text{ binary and } g \text{ identity} \\ \{\phi^{-1}+(m-1)\delta\}\left[\{\kappa\mu_1(1-\mu_1)\}^{-1}+\{(1-\kappa)\mu_0(1-\mu_0)\}^{-1}\right] & \text{if } Y_{ki} \text{ binary and } g \text{ logit} \end{cases}$$

Derivation of Equation (12)

The estimating equations for the population means are,

$$\sum_{k=1}^{K}\begin{pmatrix} \boldsymbol{Z}_k^{\intercal}R_kW_{1k}(\boldsymbol{Y}_k-\boldsymbol{\mu}_k) \\ (1-\boldsymbol{Z}_k^{\intercal})R_kW_{0k}(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)\end{pmatrix} = \sum_{k=1}^{K}\begin{pmatrix} \sum_{i=1}^{m} Z_kR_{ki}e_{1ki}^{-1}(Y_{ki}-\mu_1) \\ \sum_{i=1}^{m}(1-Z_k)R_{ki}e_{0ki}^{-1}(Y_{ki}-\mu_0)\end{pmatrix} = \boldsymbol{0}$$

where $W_{1k} = \text{diag}(e_{1ki}^{-1})$ with $e_{1ki} = P(R_{ki}=1|\boldsymbol{X}_{ki}, Z_k=1)$, $W_{0k} = \text{diag}(e_{0ki}^{-1})$ with $e_{0ki} = P(R_{ki}=1|\boldsymbol{X}_{ki}, Z_k=0)$, and $\boldsymbol{X}_{ki}$ is a vector of fully observed baseline covariates.

If we assume there is no clustering of missingness, e_{1ki} and e_{0ki} can be estimated by the following equations,

$$\sum_{k=1}^{K}\sum_{i=1}^{m}\begin{pmatrix} Z_k\boldsymbol{X}_{ki}(R_{ki}-e_{1ki}) \\ (1-Z_k)\boldsymbol{X}_{ki}(R_{ki}-e_{0ki})\end{pmatrix} = \boldsymbol{0}$$

this leads to,

$$\sum_{k=1}^{K}\boldsymbol{u}_k(Y_{ki}, R_{ki}, Z_k, \boldsymbol{X}_{ki};\boldsymbol{\theta}) = \sum_{k=1}^{K}\begin{pmatrix} \sum_{i=1}^{m} Z_kR_{ki}e_{1ki}^{-1}(Y_{ki}-\mu_1) \\ \sum_{i=1}^{m}(1-Z_k)R_{ki}e_{0ki}^{-1}(Y_{ki}-\mu_0) \\ \sum_{i=1}^{m} Z_k\boldsymbol{X}_{ki}(R_{ki}-e_{1ki}) \\ \sum_{i=1}^{m}(1-Z_k)\boldsymbol{X}_{ki}(R_{ki}-e_{0ki})\end{pmatrix} = \boldsymbol{0}$$

From M-estimation theory,

$$\text{var}(\hat{\boldsymbol{\theta}}) = K^{-1}\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^{\intercal}$$

where $\mathcal{A} = \mathbb{E}\left(-\frac{\partial\boldsymbol{u}_k}{\partial\theta^{\intercal}}\right)$ and $\mathcal{B} = \mathbb{E}\left(\boldsymbol{u}_k\boldsymbol{u}_k^{\intercal}\right)$.

The expression for $\mathcal{A}$ is:

$$m\mathbb{E}\begin{pmatrix} R_{ki}Z_k e_{1ki}^{-1}\frac{\partial\mu_1}{\partial\lambda_1} & 0 & R_{ki}Z_k(Y_{ki}-\mu_1)e_{1ki}^{-1}(1-e_{1ki})\boldsymbol{X}_{ki}^{\intercal} & \boldsymbol{0}^{\intercal} \\ 0 & R_{ki}(1-Z_k)e_{0ki}^{-1}\frac{\partial\mu_0}{\partial\lambda_0} & \boldsymbol{0}^{\intercal} & R_{ki}(1-Z_k)(Y_{ki}-\mu_0)e_{0ki}^{-1}(1-e_{0ki})\boldsymbol{X}_{ki}^{\intercal} \\ \boldsymbol{0} & \boldsymbol{0} & Z_k e_{1ki}(1-e_{1ki})\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal} & 0 \\ \boldsymbol{0} & \boldsymbol{0} & 0 & (1-Z_k)e_{0ki}(1-e_{0ki})\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal} \end{pmatrix}$$

The expression for $\mathcal{B}$ is:

$$m\mathbb{E}\begin{pmatrix} R_{ki}Z_k e_{1ki}^{-2}(Y_{ki}-\mu_1)^2+(m-1)\delta Z_k(Y_{ki}-\mu_1)^2 & 0 & R_{ki}Z_k(Y_{ki}-\mu_1)e_{1ki}^{-1}(R_i-e_{1ki})\boldsymbol{X}_{ki}^{\intercal} & \boldsymbol{0}^{\intercal} \\ 0 & R_{ki}(1-Z_k)e_{0ki}^{-2}(Y_{ki}-\mu_0)^2+(m-1)\delta(1-Z_k)(Y_{ki}-\mu_0)^2 & \boldsymbol{0}^{\intercal} & R_{ki}(1-Z_k)(Y_{ki}-\mu_0)e_{0ki}^{-1}(R_{ki}-e_{0ki})\boldsymbol{X}_i^{\intercal} \\ R_{ki}Z_k(Y_{ki}-\mu_1)e_{1ki}^{-1}(R_{ki}-e_{1ki})\boldsymbol{X}_{ki} & \boldsymbol{0} & Z_k(R_{ki}-e_{1ki})^2\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal} & 0 \\ \boldsymbol{0} & R_{ki}(1-Z_k)(Y_{ki}-\mu_0)e_{0ki}^{-1}(R_{ki}-e_{0ki})\boldsymbol{X}_{ki} & 0 & (1-Z_k)(R_{ki}-e_{0ki})^2\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal} \end{pmatrix}$$

Considering the parameters of interest, $\text{var}[(\hat{\lambda}_1,\hat{\lambda}_0)]$ can be simplified to,

$$\frac{1}{Km}\begin{pmatrix} \kappa^{-1}\left[\mathbb{E}\{R_{ki}e_{1ki}^{-2}(Y_{ki}-\mu_1)^2|Z_k=1\}+(m-1)\delta\mathbb{E}\{(Y_{ki}-\mu_1)^2|Z_k=1\}-A\right]\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2} & 0 \\ 0 & (1-\kappa)^{-1}\left[\mathbb{E}\{R_{ki}e_{0ki}^{-2}(Y_{ki}-\mu_0)^2|Z_k=0\}+(m-1)\delta\mathbb{E}\{(Y_{ki}-\mu_0)^2|Z_k=0\}-B\right]\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2} \end{pmatrix}$$

where

$$A=\mathbb{E}\{R_{ki}(Y_{ki}-\mu_1)e_{1ki}^{-1}(1-e_{1ki})\boldsymbol{X}_{ki}^{\intercal}|Z_k=1\}\left[\mathbb{E}\{e_{1ki}(1-e_{1ki})\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal}|Z_k=1\}\right]^{-1}\mathbb{E}\{R_{ki}(Y_{ki}-\mu_1)e_{1ki}^{-1}(1-e_{1ki})\boldsymbol{X}_{ki}|Z_k=1\}$$

$$B=\mathbb{E}\{R_{ki}(Y_{ki}-\mu_0)e_{0ki}^{-1}(1-e_{0ki})\boldsymbol{X}_{ki}^{\intercal}|Z_k=0\}\left[\mathbb{E}\{e_{0ki}(1-e_{0ki})\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal}|Z_k=0\}\right]^{-1}\mathbb{E}\{R_{ki}(Y_{ki}-\mu_0)e_{0ki}^{-1}(1-e_{0ki})\boldsymbol{X}_{ki}|Z_k=0\}$$

Therefore,

$$
\begin{aligned}
&\operatorname{var}\{g(\hat\mu_1)-g(\hat\mu_0)\}\\
=&(Km)^{-1}\left(\kappa^{-1}\left[\mathbb{E}\{R_{ki}e_{1ki}^{-2}(Y_{ki}-\mu_1)^2|Z_k=1\}+(m-1)\delta\mathbb{E}\{(Y_{ki}-\mu_1)^2|Z_k=1\}-A\right]\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2}\right.\\
&\left.+(1-\kappa)^{-1}\left[\mathbb{E}\{R_{ki}e_{0ki}^{-2}(Y_{ki}-\mu_0)^2|Z_k=0\}+(m-1)\delta\mathbb{E}\{(Y_{ki}-\mu_0)^2|Z_k=0\}-B\right]\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\right)\\
=&(Km)^{-1}\left(\kappa^{-1}\left[\mathbb{E}\{e_{1ki}^{-1}(Y_{ki}-\mu_1)^2|Z_k=1\}+(m-1)\delta\mathbb{E}\{(Y_{ki}-\mu_1)^2|Z_k=1\}-A\right]\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2}\right.\\
&\left.+(1-\kappa)^{-1}\left[\mathbb{E}\{e_{0ki}^{-1}(Y_{ki}-\mu_0)^2|Z_k=0\}+(m-1)\delta\mathbb{E}\{(Y_{ki}-\mu_0)^2|Z_k=0\}-B\right]\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\right)
\end{aligned}
$$

Definition of $\tau_{\text{C-known}}$ and $\tau_{\text{C-approx}}$

If the variance reduction from estimating the IPRWs from the data is ignored for CRTs, we arrive at the following definition for $\tau_{\text{C-known}}$

$$
\tau_{\text{C-known}}=\begin{cases}
\kappa^{-1}\mathbb{E}\{e_{1ki}^{-1}(Y_{ki}-\mu_1)^2|Z_k=1\} & \\
+(1-\kappa)^{-1}\mathbb{E}\{e_{0ki}^{-1}(Y_{ki}-\mu_0)^2|Z_k=0\} & \\
+(m-1)\delta\{\kappa(1-\kappa)\}^{-1}\sigma_y^2 & \text{if } Y_{ki} \text{ continuous and } g \text{ identity}\\
\kappa^{-1}\mathbb{E}\{e_{1ki}^{-1}(Y_{ki}-\mu_1)^2|Z_k=1\} & \\
+(1-\kappa)^{-1}\mathbb{E}\{e_{0ki}^{-1}(Y_{ki}-\mu_0)^2|Z_k=0\} & \\
+(m-1)\delta\left\{\kappa^{-1}\mu_1(1-\mu_1)+(1-\kappa)^{-1}\mu_0(1-\mu_0)\right\} & \text{if } Y_{ki} \text{ binary and } g \text{ identity}\\
\kappa^{-1}\mathbb{E}\{e_{1ki}^{-1}(Y_{ki}-\mu_1)^2|Z_k=1\}\left\{\mu_1(1-\mu_1)\right\}^{-2} & \\
+(1-\kappa)^{-1}\mathbb{E}\{e_{0ki}^{-1}(Y_{ki}-\mu_0)^2|Z_k=0\}\left\{\mu_0(1-\mu_0)\right\}^{-2} & \\
+(m-1)\delta\left[\left\{\kappa\mu_1(1-\mu_1)\right\}^{-1}+\left\{(1-\kappa)^{-1}\mu_0(1-\mu_0)\right\}^{-1}\right] & \text{if } Y_{ki} \text{ binary and } g \text{ logit}
\end{cases}
$$

Based on a similar approach to Shook-Sa and Hudgens[7] in the context of confounding adjustment by weighting, $\tau_{\text{C-approx}}$ is defined as follows:

$$
\tau_{\text{C-approx}}=\begin{cases}
\sigma_y^2\left[\kappa^{-1}\mathbb{E}(e_{1i}^{-1})+(1-\kappa)^{-1}\mathbb{E}(e_{0ki}^{-1})+(m-1)\delta\{\kappa(1-\kappa)\}^{-1}\right] & \text{if } Y_i \text{ continuous and } g \text{ identity}\\
\kappa^{-1}\mu_1(1-\mu_1)\left\{\mathbb{E}(e_{1ki}^{-1})+(m-1)\delta\right\} & \\
+(1-\kappa)^{-1}\mu_0(1-\mu_0)\left\{\mathbb{E}(e_{0ki}^{-1})+(m-1)\delta\right\} & \text{if } Y_i \text{ binary and } g \text{ identity}\\
\{\kappa\mu_1(1-\mu_1)\}^{-1}\left\{\mathbb{E}(e_{1ki}^{-1})+(m-1)\delta\right\} & \\
+\{(1-\kappa)\mu_0(1-\mu_0)\}^{-1}\left\{\mathbb{E}(e_{0ki}^{-1})+(m-1)\delta\right\} & \text{if } Y_i \text{ binary and } g \text{ logit}
\end{cases}
$$

WEB APPENDIX B: DERIVATION OF EQUATION (14) AND (18), APPROXIMATION OF var($\hat{\mu}_1 - \hat{\mu}_0$) UNDER EQUATION (5) AND DERIVATION OF EQUATION (19) AND (21), AND DEFINITIONS OF COMPONENTS IN EQUATIONS (22) AND (23)

Derivation of Equation (14)

In the scenario where the covariates consist of one baseline categorical variable with $c = 1, \dots, C$ categories, we can write

$$\boldsymbol{X}_i^\intercal = \left(I(X_i = 1), \dots, I(X_i = C)\right),\ \boldsymbol{\beta}_1 = \left(\beta_{11}, \dots, \beta_{C1}\right)^\intercal,\ \boldsymbol{\beta}_0 = \left(\beta_{10}, \dots, \beta_{C0}\right)^\intercal$$

$$\begin{aligned}
\mathbb{E}\{e_{1i}^{-1}(Y_i - \mu_1)^2|Z_i = 1\} &= \mathbb{E}_{X_i}\left[\frac{\mathbb{E}_{Y_i}\{(Y_i - \mu_1)^2|X_i = c, Z_i = 1\}}{\text{expit}(\beta_{c1})}\right] = \sum_{c=1}^{C}\frac{\pi_c\{\sigma_{c1}^2 + (\mu_{c1} - \mu_1)^2\}}{\text{expit}(\beta_{c1})} \\
\mathbb{E}\{e_{0i}^{-1}(Y_i - \mu_0)^2|Z_i = 0\} &= \mathbb{E}_{X_i}\left[\frac{\mathbb{E}_{Y_i}\{(Y_i - \mu_0)^2|X_i = c, Z_i = 0\}}{\text{expit}(\beta_{c0})}\right] = \sum_{c=1}^{C}\frac{\pi_c\{\sigma_{c0}^2 + (\mu_{c0} - \mu_0)^2\}}{\text{expit}(\beta_{c0})}
\end{aligned}$$

where $\pi_c = P(X_i = c) = P(X_i = c|Z_i = 1) = P(X_i = c|Z_i = 0)$, $\sigma_{c1}^2 = \text{var}(Y_i|X_i = c, Z_i = 1)$, $\sigma_{c0}^2 = \text{var}(Y_i|X_i = c, Z_i = 0)$, $\mu_{c1} = \mathbb{E}(Y_i|X_i = c, Z_i = 1)$ and $\mu_{c0} = \mathbb{E}(Y_i|X_i = c, Z_i = 0)$. Note that, if Y_i is binary, $\sigma_{c1}^2 = \mu_{c1}(1-\mu_{c1})$ and $\sigma_{c0}^2 = \mu_{c0}(1-\mu_{c0})$. We have assumed X_i is a fully observed covariate that is measured at baseline.

Furthermore,

$$\begin{aligned}
&\mathbb{E}\left\{R_i(Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})I(X_i = c)|Z_i = 1\right\} = \pi_c\{1 - \text{expit}(\beta_{c1})\}(\mu_{c1} - \mu_1) \\
&\mathbb{E}\left\{R_i(Y_i - \mu_1)e_{0i}^{-1}(1 - e_{0i})I(X_i = c)|Z_i = 0\right\} = \pi_c\{1 - \text{expit}(\beta_{c0})\}(\mu_{c0} - \mu_0) \\
&\mathbb{E}\{e_{1i}(1 - e_{1i})I(X_i = c)|Z_i = 1\} = \pi_c\text{expit}(\beta_{c1})\{1 - \text{expit}(\beta_{c1})\} \\
&\mathbb{E}\{e_{0i}(1 - e_{0i})I(X_i = c)|Z_i = 0\} = \pi_c\text{expit}(\beta_{c0})\{1 - \text{expit}(\beta_{c0})\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\{e_{1i}^{-1}(Y_i - \mu_1)^2|Z_i = 1\} - A \\
&\quad= \sum_{c=1}^{C}\frac{\pi_c\{\sigma_{c1}^2 + (\mu_{c1} - \mu_1)^2\}}{\text{expit}(\beta_{c1})} - \sum_{c=1}^{C}\frac{\pi_c(\mu_{c1} - \mu_1)^2\{1 - \text{expit}(\beta_{c1})\}}{\text{expit}(\beta_{c1})} \\
&\quad= \sum_{c=1}^{C}\pi_c\left\{\frac{\sigma_{c1}^2}{\text{expit}(\beta_{c1})} + (\mu_{c1} - \mu_1)^2\right\}
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\{e_{0i}^{-1}(Y_i - \mu_0)^2|Z_i = 0\} - B \\
&\quad= \sum_{c=1}^{C}\frac{\pi_c\{\sigma_{c0}^2 + (\mu_{c0} - \mu_0)^2\}}{\text{expit}(\beta_{c0})} - \sum_{c=1}^{C}\frac{\pi_c(\mu_{c0} - \mu_0)^2\{1 - \text{expit}(\beta_{c0})\}}{\text{expit}(\beta_{c0})} \\
&\quad= \sum_{c=1}^{C}\pi_c\left\{\frac{\sigma_{c0}^2}{\text{expit}(\beta_{c0})} + (\mu_{c0} - \mu_0)^2\right\}
\end{aligned}$$

So,

$$\begin{aligned}
\text{var}\{g(\hat{\mu}_1) - g(\hat{\mu}_0)\} = n^{-1}\sum_{c=1}^{C}\Bigg[&\kappa^{-1}\pi_c\left\{\frac{\sigma_{c1}^2}{\text{expit}(\beta_{c1})} + (\mu_{c1} - \mu_1)^2\right\}\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2} \\
&+(1-\kappa)^{-1}\pi_c\left\{\frac{\sigma_{c0}^2}{\text{expit}(\beta_{c0})} + (\mu_{c0} - \mu_0)^2\right\}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\Bigg]
\end{aligned}$$

Derivation of Equation (18)

For a single categorical variable, $\mathbb{E}(e_{1i}^{-1}) = \sum_c \pi_c \text{expit}(\beta_{c1})^{-1}$ and $\mathbb{E}(e_{0i}^{-1}) = \sum_c \pi_c \text{expit}(\beta_{c0})^{-1}$. Plugging these expressions into τ_{approx} in Section 2.2 gives the result in Equation (18).

Approximation of var($\hat{\mu}_1 - \hat{\mu}_0$) under Equation (5) and derivation of Equation (19)

In the scenario where $\boldsymbol{X}_i = (1, X_i)^\intercal$ and $(Y_i, X_i)|Z_i$ has a bivariate normal distribution, we can write

$$\boldsymbol{X}_i^\intercal = (1, X_i),\ \boldsymbol{\beta}_1 = (\beta_{01}, \beta_{11})^\intercal,\ \boldsymbol{\beta}_0 = (\beta_{00}, \beta_{10})^\intercal$$

$$\{(X_i, Y_i)|Z_i = 1\} \sim \text{N}\left(\begin{pmatrix}\mu_x \\ \mu_1\end{pmatrix}, \begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}\right)$$

$$\{(X_i, Y_i)|Z_i = 0\} \sim \text{N}\left(\begin{pmatrix}\mu_x \\ \mu_0\end{pmatrix}, \begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}\right)$$

$$\mathbb{E}(Y_i|X_i = x_i, Z_i = 1) = \mu_1 + \frac{\rho\sigma_y}{\sigma_x}(x_i - \mu_x),\ \mathbb{E}(Y_i|X_i = x_i, Z_i = 0) = \mu_0 + \frac{\rho\sigma_y}{\sigma_x}(x_i - \mu_x)$$

$$\text{var}(Y_i|X_i = x_i, Z_i = 1) = \text{var}(Y_i|X_i = x_i, Z_i = 0) = \sigma_y^2(1 - \rho^2)$$

We have assumed X_i is a fully observed covariate that is measured at baseline, so $\mu_x = \mathbb{E}(X_i) = \mathbb{E}(X_i|Z_i = 1) = \mathbb{E}(X_i|Z_i = 0)$. Using moment generating functions,

$$\begin{aligned}\mathbb{E}\{e_{1i}^{-1}(Y_i - \mu_1)^2|Z_i = 1\} &= \mathbb{E}_{X_i}\left[\frac{\mathbb{E}_{Y_i}\{(Y_i - \mu_1)^2|X_i = x_i, Z_i = 1\}}{\text{expit}(\beta_{01} + \beta_{11}x_i)}\right] \\ &= \sigma_y^2\left\{1 + \exp\left(-\beta_{01} - \mu_x\beta_{11} + \beta_{11}^2\sigma_x^2/2\right)\left(1 + \sigma_x^2\rho^2\beta_{11}^2\right)\right\}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\{e_{0i}^{-1}(Y_i - \mu_0)^2|Z_i = 0\} &= \mathbb{E}_{X_i}\left[\frac{\mathbb{E}_{Y_i}\{(Y_i - \mu_0)^2|X_i = x_i, Z_i = 0\}}{\text{expit}(\beta_{00} + \beta_{10}x_i)}\right] \\ &= \sigma_y^2\left\{1 + \exp\left(-\beta_{00} - \mu_x\beta_{10} + \beta_{10}^2\sigma_x^2/2\right)\left(1 + \sigma_x^2\rho^2\beta_{10}^2\right)\right\}\end{aligned}$$

$A = \mathbb{E}\{R_i(Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})\boldsymbol{X}_i^\intercal|Z_i = 1\}\left[\mathbb{E}\{e_{1i}(1 - e_{1i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i = 1\}\right]^{-1}\mathbb{E}\{R_i(Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})\boldsymbol{X}_i|Z_i = 1\}$ and $B = \mathbb{E}\{R_i(Y_i - \mu_0)e_{0i}^{-1}(1 - e_{0i})\boldsymbol{X}_i^\intercal|Z_i = 0\}\left[\mathbb{E}\{e_{0i}(1 - e_{0i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i = 0\}\right]^{-1}\mathbb{E}\{R_i(Y_i - \mu_0)e_{0i}^{-1}(1 - e_{0i})\boldsymbol{X}_i|Z_i = 0\}$ represent the reduction in the variance associated with estimating the inverse probability of response weights from the data, and are not available in closed-form as they involve integration over the expit function. However, they can be approximated by Gauss-Hermite quadrature, as follows,

$$\begin{aligned}&\mathbb{E}\{R_i(Y_i - \mu_1)e_{1i}^{-1}(1 - e_{1i})\boldsymbol{X}_i^\intercal|Z_i = 1\} \\ &= \mathbb{E}_{\boldsymbol{X}_i}\left[E_{Y_i}\{(Y_i - \mu_1)|\boldsymbol{X}_i = \boldsymbol{x}_i, Z_i = 1\}\{1 - \text{expit}(\boldsymbol{x}_i^\intercal\boldsymbol{\beta}_1)\}\boldsymbol{x}_i^\intercal\right] \\ &= \mathbb{E}_{\boldsymbol{X}_i}\left\{\frac{\sigma_y\rho}{\sigma_x}(X_i - \mu_x)\boldsymbol{X}_i^\intercal|Z_i = 1\right\} - \mathbb{E}_{\boldsymbol{X}_i}\left\{\frac{\sigma_y\rho}{\sigma_x}(X_i - \mu_x)\text{expit}(\boldsymbol{X}_i^\intercal\boldsymbol{\beta}_1)\boldsymbol{X}_i^\intercal|Z_i = 1\right\} \\ &= \left(0, \sigma_y\sigma_x\rho\right) \\ &- \frac{\sigma_y\rho}{\sigma_x}\left(\mathbb{E}_{\boldsymbol{X}_i}\left\{(X_i - \mu_x)\text{expit}(\beta_{01} + \beta_{11}X_i)|Z_i = 1\right\}, \mathbb{E}_{\boldsymbol{X}_i}\left\{(X_i - \mu_x)X_i\text{expit}(\beta_{01} + \beta_{11}X_i)|Z_i = 1\right\}\right) \\ &\approx \underbrace{\sigma_y\rho\left[(0, \sigma_x) - \sqrt{\frac{2}{\pi}}\sum_{j=1}^{J}\left(x_j\text{expit}\left\{\beta_{01} + \beta_{11}(\sqrt{2}\sigma_x x_j + \mu_x)\right\}w_j, x_j(\sqrt{2}\sigma_x x_j + \mu_x)\text{expit}\left\{\beta_{01} + \beta_{11}(\sqrt{2}\sigma_x x_j + \mu_x)\right\}w_j\right)\right]}_{C_1}\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\{e_{1i}(1 - e_{1i})\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i = 1\} \\ &\approx \underbrace{\frac{1}{\sqrt{\pi}}\sum_{j=1}^{J}\text{expit}\left\{\beta_{01} + \beta_{11}(\sqrt{2}\sigma_x x_j + \mu_x)\right\}\left[1 - \text{expit}\left\{\beta_{01} + \beta_{11}(\sqrt{2}\sigma_x x_j + \mu_x)\right\}\right]\begin{pmatrix}1 & \sqrt{2}\sigma_x x_j + \mu_x \\ \sqrt{2}\sigma_x x_j + \mu_x & (\sqrt{2}\sigma_x x_j + \mu_x)^2\end{pmatrix}w_j}_{D_1}\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\{R_i(Y_i - \mu_0)e_{0i}^{-1}(1 - e_{0i})\boldsymbol{X}_i^\intercal|Z_i = 0\} \\ &\approx \underbrace{\sigma_y\rho\left[(0, \sigma_x) - \sqrt{\frac{2}{\pi}}\sum_{j=1}^{J}\left(x_j\text{expit}\left\{\beta_{00} + \beta_{10}(\sqrt{2}\sigma_x x_j + \mu_x)\right\}w_j, x_j(\sqrt{2}\sigma_x x_j + \mu_x)\text{expit}\left\{\beta_{00} + \beta_{10}(\sqrt{2}\sigma_x x_j + \mu_x)\right\}w_j\right)\right]}_{C_0}\end{aligned}$$

$$\mathbb{E}\{e_{0i}(1-e_{0i})\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}|Z_i=0\}$$
$$\approx \underbrace{\frac{1}{\sqrt{\pi}}\sum_{j=1}^{J}\text{expit}\left\{\beta_{00}+\beta_{10}(\sqrt{2}\sigma_x x_j+\mu_x)\right\}\left[1-\text{expit}\left\{\beta_{00}+\beta_{10}(\sqrt{2}\sigma_x x_j+\mu_x)\right\}\right]\begin{pmatrix}1 & \sqrt{2}\sigma_x x_j+\mu_x\\ \sqrt{2}\sigma_x x_j+\mu_x & (\sqrt{2}\sigma_x x_j+\mu_x)^2\end{pmatrix}w_j}_{D_0}$$

where x_j are the quadrature points on the Hermite polynomial and w_j the associated weights for J quadrature points. $\text{var}(\hat{\mu}_1-\hat{\mu}_0)$ is approximated by,

$$\begin{aligned}n^{-1}\sigma_y^2\Big(\kappa^{-1}\text{expit}\left(\beta_{01}+\mu_x\beta_{11}-\beta_{11}^2\sigma_x^2/2\right)^{-1}+(1-\kappa)^{-1}\text{expit}\left(\beta_{00}+\mu_x\beta_{10}-\beta_{10}^2\sigma_x^2/2\right)^{-1}\\ +\rho^2\left[\kappa^{-1}\sigma_x^2\beta_{11}^2\exp\left\{-\left(\beta_{01}+\mu_x\beta_{11}-\beta_{11}^2\sigma_x^2/2\right)\right\}\right.\\ +(1-\kappa)^{-1}\sigma_x^2\beta_{10}^2\exp\left\{-\left(\beta_{00}+\mu_x\beta_{10}-\beta_{10}^2\sigma_x^2/2\right)\right\}\\ \left.-\kappa^{-1}C_1D_1^{-1}C_1-(1-\kappa)^{-1}C_0D_0^{-1}C_0\right]\Big)\end{aligned}$$

If you ignore the reduction in the variance from estimation of the inverse probability of response weights from the data, $\text{var}(\hat{\mu}_1-\hat{\mu}_0)$ simplifies to Equation (19),

$$\begin{aligned}n^{-1}\sigma_y^2\Big(\kappa^{-1}\text{expit}\left(\beta_{01}+\mu_x\beta_{11}-\beta_{11}^2\sigma_x^2/2\right)^{-1}+(1-\kappa)^{-1}\text{expit}\left(\beta_{00}+\mu_x\beta_{10}-\beta_{10}^2\sigma_x^2/2\right)^{-1}\\ +\rho^2\left[\kappa^{-1}\sigma_x^2\beta_{11}^2\exp\left\{-\left(\beta_{01}+\mu_x\beta_{11}-\beta_{11}^2\sigma_x^2/2\right)\right\}\right.\\ \left.+(1-\kappa)^{-1}\sigma_x^2\beta_{10}^2\exp\left\{-\left(\beta_{00}+\mu_x\beta_{10}-\beta_{10}^2\sigma_x^2/2\right)\right\}\right]\Big)\end{aligned}$$

Derivation of Equation (21)

For a single continuous normally distributed variable, $\mathbb{E}(e_{1i}^{-1})=1+\exp\left(-\beta_{01}-\mu_x\beta_{11}+\beta_{11}^2\sigma_x^2/2\right)$ and $\mathbb{E}(e_{0i}^{-1})=1+\exp\left(-\beta_{00}-\mu_x\beta_{10}+\beta_{10}^2\sigma_x^2/2\right)$. Plugging these expressions into τ_{approx} in Section 2.2 gives the result in Equation (21).

Definition of Components in Equation (22)

For IRTs, $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ are defined, as follows:

$$\hat{\mathcal{A}}=n_{\text{pilot}}^{-1}\sum_{i=1}^{n_{\text{pilot}}}\begin{pmatrix}R_iZ_i\hat{e}_{1i}^{-1}\frac{\partial\mu_1}{\partial\lambda_1}\Big|_{\hat{\mu}_1} & 0 & R_iZ_i(Y_i-\hat{\mu}_1)\hat{e}_{1i}^{-1}(1-\hat{e}_{1i})\boldsymbol{X}_i^{\intercal} & \boldsymbol{0}^{\intercal}\\ 0 & R_i(1-Z_i)\hat{e}_{0i}^{-1}\frac{\partial\mu_0}{\partial\lambda_0}\Big|_{\hat{\mu}_0} & \boldsymbol{0}^{\intercal} & R_i(1-Z_i)(Y_i-\hat{\mu}_0)\hat{e}_{0i}^{-1}(1-\hat{e}_{0i})\boldsymbol{X}_i^{\intercal}\\ \boldsymbol{0} & \boldsymbol{0} & Z_i\hat{e}_{1i}(1-\hat{e}_{1i})\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal} & 0\\ \boldsymbol{0} & \boldsymbol{0} & 0 & (1-Z_i)\hat{e}_{0i}(1-\hat{e}_{0i})\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}\end{pmatrix}$$

$$\hat{\mathcal{B}}=n_{\text{pilot}}^{-1}\sum_{i=1}^{n_{\text{pilot}}}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_i^{\intercal},\ \hat{\boldsymbol{u}}_i=\begin{pmatrix}R_iZ_i\hat{e}_{1i}^{-1}(Y_i-\hat{\mu}_1)\\ R_i(1-Z_i)\hat{e}_{0i}^{-1}(Y_i-\hat{\mu}_0)\\ Z_i\boldsymbol{X}_i(R_i-\hat{e}_{1i})\\ (1-Z_i)\boldsymbol{X}_i(R_i-\hat{e}_{0i})\end{pmatrix}$$

Definition of Components in Equation (23)

For CRTs, $\hat{\mathcal{A}}_{\text{C}}$ and $\hat{\mathcal{B}}_{\text{C}}$ are defined, as follows:

$$\hat{\mathcal{A}}_{\text{C}}=K_{\text{pilot}}^{-1}\sum_{k=1}^{K_{\text{pilot}}}\sum_{i=1}^{m}\begin{pmatrix}R_{ki}Z_k\hat{e}_{1ki}^{-1}\frac{\partial\mu_1}{\partial\lambda_1}\Big|_{\hat{\mu}_1} & 0 & R_{ki}Z_k(Y_{ki}-\hat{\mu}_1)\hat{e}_{1ki}^{-1}(1-\hat{e}_{1ki})\boldsymbol{X}_{ki}^{\intercal} & \boldsymbol{0}^{\intercal}\\ 0 & R_{ki}(1-Z_k)\hat{e}_{0ki}^{-1}\frac{\partial\mu_0}{\partial\lambda_0}\Big|_{\hat{\mu}_0} & \boldsymbol{0}^{\intercal} & R_{ki}(1-Z_k)(Y_{ki}-\hat{\mu}_0)\hat{e}_{0ki}^{-1}(1-\hat{e}_{0ki})\boldsymbol{X}_{ki}^{\intercal}\\ \boldsymbol{0} & \boldsymbol{0} & Z_k\hat{e}_{1ki}(1-\hat{e}_{1ki})\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal} & 0\\ \boldsymbol{0} & \boldsymbol{0} & 0 & (1-Z_k)\hat{e}_{0ki}(1-\hat{e}_{0ki})\boldsymbol{X}_{ki}\boldsymbol{X}_{ki}^{\intercal}\end{pmatrix}$$

$$\hat{\mathcal{B}}_{\text{C}}=K_{\text{pilot}}^{-1}\sum_{k=1}^{K_{\text{pilot}}}\hat{\boldsymbol{u}}_k\hat{\boldsymbol{u}}_k^{\intercal},\ \hat{\boldsymbol{u}}_k=\sum_{i=1}^{m}\begin{pmatrix}R_{ki}Z_k\hat{e}_{1ki}^{-1}(Y_{ki}-\hat{\mu}_1)\\ R_{ki}(1-Z_k)\hat{e}_{0ki}^{-1}(Y_{ki}-\hat{\mu}_0)\\ Z_k\boldsymbol{X}_{ki}(R_{ki}-\hat{e}_{1ki})\\ (1-Z_k)\boldsymbol{X}_{ki}(R_{ki}-\hat{e}_{0ki})\end{pmatrix}$$

WEB APPENDIX C: ADDITIONAL SIMULATION SCENARIOS AND RESULTS

Dataset Generation

Additional Parameters Varied for the Single Baseline Binary Covariate Scenarios

Six of the scenarios in Table 1, labeled as scenario 1 and 2 for each combination of outcome type and function (g), were altered to observe simulation results when missingness per intervention group, the association between the covariate and the probability of missingness, the proportion randomized to the intervention group and the sample size was varied, as follows:

- The amount of missingness by intervention group was varied by varying $P(R_i = 1|Z_i = 1)$ from 0.7 to 0.9 with the constraint $0.5 * \{P(R_i = 1|Z_i = 1) + P(R_i = 1|Z_i = 0)\} = \phi = 0.8$, so $P(R_i = 1|Z_i = 0) = 1.6 - P(R_i = 1|Z_i = 1)$, along with $\text{expit}(\beta_{11}) = P(R_i = 1|Z_i = 1) - 0.1$, $\text{expit}(\beta_{21}) = P(R_i = 1|Z_i = 1) + 0.1$, $\text{expit}(\beta_{10}) = P(R_i = 1|Z_i = 0) - 0.05$ and $\text{expit}(\beta_{20}) = P(R_i = 1|Z_i = 0) + 0.05$. Two figures with six panels for the six scenarios with different $P(R_i = 1|Z_i = 1)$ were produced. In each panel of the first figure, $P(R_i = 1|Z_i = 1)$ was varied on x-axis, and on the y-axis the sample size calculated by each method (n_{standard}, n_{IPRW}, n_{known} and n_{approx}) was displayed. In each panel of the second figure, the empirical power for each analysis method [standard (IPRW estimator), IPRW (IPRW estimator), known (IPRW estimator), approx (IPRW estimator)] was displayed.

- This approach was repeated for varying the association between the covariate and the probability of missingness. On the x-axis of each of the six panels, $P(R_i = 1|X_i = 1, Z_i = 1) = \text{expit}(\beta_{11})$ was varied from 0.7 to between 0.6 and 0.9. As the probability of the outcome being observed in the intervention group should remain at $P(R_i = 1|Z_i = 1) = 0.8$ and $P(Z_i = 1) = \kappa$ held at 0.5, $P(R_i = 1|X_i = 2, Z_i = 1) = \text{expit}(\beta_{21})$ was varied between 1 and 0.6 as $P(R_i = 1|X_i = 1, Z_i = 1) = \text{expit}(\beta_{11})$ varied between 0.6 and 1.

- This approach was repeated for varying the probability of being randomized to the intervention group $\kappa = P(Z_i = 1)$ from 0.5 to be between 0.2 and 0.8.

- This approach was repeated for smaller n. This was achieved by varying either μ_{21} or μ_{11} to change the intervention effect to be larger so a lower sample size was required. In scenario 1, μ_{11} was fixed at 0.9 and μ_{21} varied from 0.3 to 0.6, so n_{standard} varied from 1314 to 184. In scenario 2, μ_{21} was fixed at 0.95 and μ_{11} varied from 0.62 to 0.77 so n_{standard} varied from 1314 to 290.

Single Baseline Continuous Covariate Scenarios

For the scenarios with a continuous outcome where missingness depends on a single baseline continuous variable and the randomized intervention group, τ_{IPRW} was approximated by Gauss-Hermite quadrature with 100 quadrature points and $\mu_1 - \mu_0$ was set as 0.1. The baseline continuous covariate X_i was generated by random draws from a standard normal distribution. $R_i|(X_i, Z_i)$ was generated by random draws from a Bernoulli distribution, such that $R_i|(X_i = x_i, Z_i = 1) \sim \text{Ber}\{\text{expit}(\beta_{01} + \beta_{11}x_i)\} = \text{Ber}\{\text{expit}(1.4 + 0.21x_i)\}$ and $R_i|(X_i = x_i, Z_i = 0) \sim \text{Ber}\{\text{expit}(\beta_{00} + \beta_{10}x_i)\} = \text{Ber}\{\text{expit}(2 + 1.64x_i)\}$. The β coefficients were chosen so the overall probability of the outcome being observed was $\phi = P(R_i = 1) = 0.8$. Y_i was generated by random draws from a normal distribution, such that $Y_i|(X_i = x_i, Z_i = 1) \sim \text{N}(\mu_1 + \rho\sigma_y x_i, \sigma_y^2(1-\rho^2))$ $=\text{N}(0.475 - 0.75\sqrt{0}.245x_i, 0.245(1 - 0.75^2))$ and $Y_i|(X_i = x_i, Z_i = 0) \sim \text{N}(\mu_0 + \rho\sigma_y x_i, \sigma_y^2(1-\rho^2))$ $=\text{N}(0.375 - 0.75\sqrt{0}.245x_i, 0.245(1 - 0.75^2))$.

This scenario was altered to observe simulation results when the amount of missingness, missingness per intervention group, the association between the covariate and the probability of missingness, the proportion randomized to the intervention group and the sample size was varied, as follows:

- The amount of missingness $\phi = P(R_i = 1)$ was varied from 0.8 to be between 0.75 and 0.8. This was done by varying β_{11} and β_{10}, so that $P(R_i = 1|Z_i = 1) = \int \text{expit}(1.4 + \beta_{11}x_i)\frac{e^{-x_i^2/2}}{\sqrt{2\pi}}dx_i = \phi$ and $P(R_i = 1|Z_i = 0) = \int \text{expit}(2 + \beta_{10}x_i)\frac{e^{-x_i^2/2}}{\sqrt{2\pi}}dx_i = \phi$. Two figures for the scenario were produced, where ϕ was varied on the x-axis, and on the y-axis of the first figure the sample size calculated by each method (n_{standard}, n_{IPRW}, n_{known} and n_{approx}) was displayed. On the y-axis of the second figure, the empirical power for each analysis method [standard (IPRW estimator), IPRW (IPRW estimator), known (IPRW estimator), approx (IPRW estimator)] was displayed.

- This approach was repeated for varying the amount of missingness per intervention group. On the x-axis, $P(R_i = 1|Z_i = 1)$ was varied from 0.75 to 0.8 with the constraint $0.5 * \{P(R_i = 1|Z_i = 1) + P(R_i = 1|Z_i = 0)\} = \phi = 0.8$, so $P(R_i = 1|Z_i = 0) = 1.6 - P(R_i = 1|Z_i = 1)$. Along with $\beta_{01} = 1.4$ and $\beta_{00} = 2$ and varying β_{11} and β_{10}.

- This approach was repeated for varying the association between the covariate and the probability of missingness. On the x-axis, β_{11} was varied from 0.21 to 2.21 and β_{01} was set so that $P(R_i = 1|Z_i = 1) = \int \text{expit}(\beta_{01} + \beta_{11}x_i)\frac{e^{-x_i^2/2}}{\sqrt{2\pi}}dx_i = 0.8$.

- This approach was repeated for varying the probability of being randomized to the intervention group $\kappa = P(Z_i = 1)$ to be between 0.2 and 0.8.

- This approach was repeated for smaller n. This was achieved by varying μ_1 so the intervention effect was larger so a lower sample size was required. μ_1 was varied from 0.475 to 0.675, so n_{standard} varied from 1288 to 144.

CRT Scenarios

For CRTs, the intercluster correlation δ was set at 0.05 and the cluster size m was set at 5. The number of clusters was calculated as $2\lceil(n_{\text{C-standard}}/2)/m\rceil$, $2\lceil(n_{\text{C-IPRW}}/2)/m\rceil$, $2\lceil(n_{\text{C-known}}/2)/m\rceil$ and $2\lceil(n_{\text{C-approx}}/2)/m\rceil$, respectively, where $\lceil\ \rceil$ is the ceiling function. Half the clusters were assigned to the intervention group. The overall probability of the outcome being observed was set at $\phi = P(R_{ki} = 1) = 0.8$.

For a single baseline binary covariate ($X_{ki} = 1, 2$), the probability of the outcome being observed was generated from Bernoulli distributions as follows: $R_{ki}|(X_{ki} = 1, Z_k = 1) \sim \text{Ber}(0.7)$, $R_{ki}|(X_{ki} = 2, Z_k = 1) \sim \text{Ber}(0.9)$, $R_{ki}|(X_{ki} = 1, Z_k = 0) \sim \text{Ber}(0.75)$ and $R_{ki}|(X_{ki} = 2, Z_k = 0) \sim \text{Ber}(0.85)$. In the scenarios where Y_{ki} is a continuous outcome, Y_{ki} was generated by $Y_{ki}|(X_{ki} = c, Z_k = 1) = \mu_{c1} + \zeta_k + \epsilon_{ki}^{c1}$ and $Y_{ki}|(X_{ki} = c, Z_k = 0) = \mu_{c0} + \zeta_k + \epsilon_{ki}^{c0}$ with $\zeta_k \sim N(0, \delta\sigma_y^2)$, $\epsilon_{ki}^{c1} \sim N(0, \sigma_{c1}^2 - \delta\sigma_y^2)$ and $\epsilon_{ki}^{c0} \sim N(0, \sigma_{c0}^2 - \delta\sigma_y^2)$ for $c = 1, 2$, where $\sigma_y^2 = \sum_c \pi_c\{\sigma_{c1}^2 + (\mu_{c1} - \mu_1)^2\} = \sum_c \pi_c\{\sigma_{c0}^2 + (\mu_{c0} - \mu_0)^2\}$. When Y_{ki} is a binary outcome, Y_{ki} was generated using the method of Qaqish[8], such that the mean outcome in each category in the intervention group was μ_{c1} and each category in the control group was μ_{c0}, and the correlation between outcomes in each cluster was δ. These scenarios are displayed in Table 2(a).

For a single baseline continuous covariate, X_{ki} was generated by random draws from a standard normal distribution, and R_{ki} from a Bernoulli distribution such that $R_{ki}|(X_{ki} = x_{ki}, Z_k = 1) \sim \text{Ber}(\text{expit}(1.4 + 0.21x_{ki}))$ and $R_{ki}|(X_{ki} = x_{ki}, Z_k = 0) \sim \text{Ber}(\text{expit}(2 + 1.64x_{ki}))$. Y_{ki} was generated by $Y_{ki}|(X_{ki} = x_{ki}, Z_k = 1) = \mu_1 + \rho\sigma_y x_{ki} + \zeta_k + \epsilon_{ki} = 0.475 - 0.75\sqrt{0}.245x_{ki} + \zeta_k + \epsilon_{ki}$ and $Y_{ki}|(X_{ki} = x_{ki}, Z_k = 0) = \mu_0 + \rho\sigma_y x_{ki} + \zeta_k + \epsilon_{ki} = 0.375 - 0.75\sqrt{0}.245 + \zeta_k + \epsilon_{ki}$ with $\zeta_k \sim N(0, \delta\sigma_y^2) = N(0, 0.05 \times 0.245)$ and $\epsilon_{ki} \sim N(0, \sigma_y^2(1 - \delta - \rho^2)) = N(0, 0.245(1 - 0.05 - 0.75^2))$.

Each of the six scenarios in Table 2 and the one scenario for a baseline continuous covariate were altered to observe simulation results when the intercluster correlation (δ) and cluster size (m) was varied, as follows:

- The intercluster correlation (δ) was varied from 0.03 to 0.07 on the x-axis.

- The cluster size (m) was varied from 2 to 10 on the x-axis.

Dataset Analysis

The IPRW empirical sandwich variance estimator for IRTs was:

$$n_{\text{IPRW}}^{-1}\left\{\hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}\left(\hat{\mathcal{A}}^{-1}\right)^{\intercal}_{[1,1]} + \hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}\left(\hat{\mathcal{A}}^{-1}\right)^{\intercal}_{[2,2]}\right\}$$

where $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ are defined as in Equation (22) with n_{pilot} replaced with n_{IPRW}.
The IPRW empirical sandwich variance estimator for CRTs was:

$$K_{\text{C-IPRW}}^{-1}\left\{\hat{\mathcal{A}}_{\text{C}}^{-1}\hat{\mathcal{B}}_{\text{C}}\left(\hat{\mathcal{A}}_{\text{C}}^{-1}\right)^{\intercal}_{[1,1]} + \hat{\mathcal{A}}_{\text{C}}^{-1}\hat{\mathcal{B}}_{\text{C}}\left(\hat{\mathcal{A}}_{\text{C}}^{-1}\right)^{\intercal}_{[2,2]}\right\},\ K_{\text{C-IPRW}} = n_{\text{C-IPRW}}/m$$

where $\hat{\mathcal{A}}_{\text{C}}$ and $\hat{\mathcal{B}}_{\text{C}}$ are defined as in Equation (23) with K_{pilot} replaced with $K_{\text{C-IPRW}}$.

Additional Results

Additional Parameters Varied for a Single Baseline Binary Covariate Scenarios

The simulation results when missingness per intervention group was varied are in Web Figure 1, when the association between the covariate and the probability of missingness was varied are in Web Figure 2, when the proportion randomized to the intervention group was varied are in Web Figure 3 and when the sample size was varied are in Web Figure 4.

Single Baseline Continuous Covariate Scenarios

For IRTs with weighting based on the single continuous covariate, n_{standard} was calculated to be 1288, n_{IPRW} to be 1480, n_{known} to be 1836 and n_{approx} to be 1428. $\hat{\mu}_1 - \hat{\mu}_0$ based on n_{standard}, n_{IPRW}, n_{known} or n_{approx} participants by Equation (4) was correctly estimated as 0.10. The empirical power with n_{IPRW} using the IPRW estimator in Equation (4) was at the target of 90%. The empirical power with n_{standard}, n_{known} and n_{approx} using the IPRW estimator in Equation (4) was 87, 94 and 90%, respectively. For CRTs, $n_{\text{C-standard}}$ was calculated to be 1494 ($K_{\text{C-standard}} = 300$), $n_{\text{C-IPRW}}$ to be 1686 ($K_{\text{C-IPRW}} = 338$), $n_{\text{C-known}}$ to be 2042 ($K_{\text{C-known}} = 410$) and $n_{\text{C-approx}}$ to be 1634 ($K_{\text{C-approx}} = 328$). Results were similar to IRTs, $\hat{\mu}_1 - \hat{\mu}_0$ based on $n_{\text{C-standard}}$, $n_{\text{C-IPRW}}$ $n_{\text{C-known}}$ or $n_{\text{C-approx}}$ participants and the IPRW estimator was correctly estimated as 0.10. The empirical power with $n_{\text{C-IPRW}}$ using the IPRW estimator was at the target of 90%. The empirical power with $n_{\text{C-standard}}$, $n_{\text{C-known}}$ and $n_{\text{C-approx}}$ using the IPRW estimator was 88, 94 and 90%, respectively.

The simulation results when the amount of missingness was varied are in Web Figure 5, when missingness per intervention group was varied are in Web Figure 6, when the association between the covariate and the probability of missingness was varied are in Web Figure 7, when the proportion randomized to the intervention group was varied are in Web Figure 8 and when the sample size was varied are in Web Figure 9.

CRT Scenarios

The simulation results when the intercluster correlation was varied are in Web Figure 10 and 11 for a single baseline binary and continuous covariate, respectively. The results when the cluster size was varied are in Web Figure 12 and 13 for a single baseline binary and continuous covariate, respectively.

WEB APPENDIX D: CLUSTER RANDOMIZED TRIAL (CRT) TUTORIAL AND INDIVIDUALLY RANDOMIZED TRIAL (IRT) CASE STUDY

Estimating the Intercluster Correlation from Pilot Data for the Cluster Randomized Trial (CRT) Tutorial

An IPRW estimator of the intercluster correlation assuming an exchangeable correlation structure was constructed by considering the following estimating equation,

$$\sum_{k=1}^{K}\sum_{i\neq j}\left[Z_k R_{ki} R_{kj} e_{1ki}^{-1} e_{1kj}^{-1}\left\{\frac{(Y_{ki}-\mu_1)(Y_{kj}-\mu_1)}{\mu_1(1-\mu_1)}-\delta\right\}+(1-Z_k)R_{ki}R_{kj}e_{0ki}^{-1}e_{0kj}^{-1}\left\{\frac{(Y_{ki}-\mu_0)(Y_{kj}-\mu_0)}{\mu_0(1-\mu_0)}-\delta\right\}\right]=0$$

So,

$$\hat{\delta}=\frac{\sum_{k=1}^{K}\sum_{i\neq j}R_{ki}R_{kj}\left\{\frac{Z_k e_{1ki}^{-1}e_{1kj}^{-1}(Y_{ki}-\hat{\mu}_1)(Y_{kj}-\hat{\mu}_1)}{\hat{\mu}_1(1-\hat{\mu}_1)}+\frac{(1-Z_k)e_{0ki}^{-1}e_{0kj}^{-1}(Y_{ki}-\hat{\mu}_0)(Y_{kj}-\hat{\mu}_0)}{\hat{\mu}_0(1-\hat{\mu}_0)}\right\}}{\sum_{k=1}^{K}\sum_{i\neq j}R_{ki}R_{kj}\left\{Z_k e_{1ki}^{-1}e_{1kj}^{-1}+(1-Z_k)e_{0ki}^{-1}e_{0kj}^{-1}\right\}}$$

Individually Randomized Trial (IRT) Case Study of Sample Size for the IPRW versus the Standard Approach

To illustrate how different patterns of missing outcome data can influence the required variance inflation or reduction under an IPRW estimator compared to the standard approach, we explore outcome measures from the AIDS Clinical Trials Group (ACTG) A5273 IRT[9]. ACTG A5273 evaluated second-line antiretroviral therapy (ART) for treatment of HIV in resource-limited settings. It randomized half the participants to second-line ART consisting of lopinavir/ritonavir plus raltegravir (the intervention group) and half to second-line ART consisting of lopinavir/ritonavir plus nucleos(t)ide reverse transcriptase inhibitors (the control group). In this case study, we consider CD4 count and triglyceride levels measured at 96 weeks post-initiation of second-line ART as potential outcome measures. We estimate the mean and variance components relevant to each sample size formula for each outcome measure from the trial data, set the proportion of participants with an observed outcome at 90% in the intervention group and 70% in the control group, and explore the impact of the association between the fully observed baseline covariate and the probability of the outcome being observed by varying the β coefficients in the missingness model.

For CD4 count, we considered a binary outcome of CD4 $>$ 400 cells/mm^3 and a continuous outcome of log(CD4). We considered the primary outcome to be MAR given the randomized intervention group and the baseline CD4 count measurement; either dichotomized as baseline CD4 $>$ 100 versus $\leq$ 100 cells/mm^3 or as a continuous measure of log(baseline CD4). The relative efficiency of the IPRW versus the standard approach, defined as $\tau_{\text{IPRW}}/\tau_{\text{standard}}$, was evaluated. When weighting by dichotomized baseline CD4, the difference in relative efficiency was small; $\tau_{\text{IPRW}}/\tau_{\text{standard}}$ ranged from a minimum of 0.976 to a maximum of 1.02 (Web Figure 16A, B and C). This small difference is because the within-category variability of CD4 count was similar for all categories as shown in the τ_{IPRW} calculation underneath Web Figure 16. For the log(CD4) outcome weighting by log(baseline CD4), $\tau_{\text{IPRW}}/\tau_{\text{standard}}$ had a large range from a minimum of 0.8 to a maximum of 1.6 as the odds ratio for the association between a one unit increase in log(baseline CD4) and the outcome being observed ranged from 0.25 to 4 (Web Figure 16D). The IPRW estimator is less efficient compared to the standard approach when the strength of the association between log(baseline CD4) and the outcome being observed is weaker. This calculation assumed log(baseline CD4) and log(CD4) are both normally distributed.

For triglyceride levels, we considered a binary outcome of triglyceride $>$ 200 mg/dL and a continuous outcome of log(triglyceride). We considered the primary outcome to be MAR given the randomized intervention group and the baseline triglyceride level; either dichotomized as baseline triglyceride $>$ 150 versus $\leq$ 150 mg/dL or as a continuous measure of log(baseline triglyceride). When weighting by dichotomized baseline triglyceride, the difference in the relative efficiency of the IPRW versus standard approach was larger than for the CD4 count outcome; $\tau_{\text{IPRW}}/\tau_{\text{standard}}$ ranged from a minimum of 0.96 to a maximum of 1.14 (Web Figure 17A, B and C). The difference in magnitude compared to the CD4 count outcome is explained by the fact that in the control group the within-category variability of triglycerides is higher for participants with a baseline triglyceride level $\geq$ 150 mg/dL than for participants with a baseline triglyceride level $<$ 150 mg/dL. Indeed, the relative efficiency would differ even more for an outcome measure with a greater discrepancy in within-category variances. For the log(triglyceride) outcome weighting by log(baseline triglyceride), $\tau_{\text{IPRW}}/\tau_{\text{standard}}$ had a large range from a minimum of 0.8 to a maximum of 1.5 (Web Figure 17D), with a similar pattern as for the CD4 count outcome.

WEB APPENDIX E: SIMPLE G-COMPUTATION

Let Y_i represent the outcome for participant i and Z_i indicate the randomized intervention, such that,

$$Z_i = \begin{cases} 1 & \text{if participant } i \text{ is randomized to the intervention group} \\ 0 & \text{if participant } i \text{ is randomized to the control group} \end{cases}$$

Suppose that $g\left\{E(Y_i|Z_i=1)\right\} = g(\mu_1) = \lambda_1$ and $g\left\{E(Y_i|Z_i=0)\right\} = g(\mu_0) = \lambda_0$, where μ_1 is the population mean of the outcome Y_i under intervention and μ_0 is the population mean of the outcome Y_i under control. Define R_i to be an indicator of whether the outcome Y_i is observed, as follows,

$$R_i = \begin{cases} 1 & \text{if } Y_i \text{ is observed} \\ 0 & \text{if } Y_i \text{ is missing} \end{cases}$$

Let $\boldsymbol{X}_i$ represent a vector of fully observed baseline covariates. Suppose that $\mathbb{E}(Y_i|\boldsymbol{X}_i, Z_i=1) = m_{1i} = \boldsymbol{X}_i^\intercal \boldsymbol{\zeta}_1$ and $\mathbb{E}(Y_i|\boldsymbol{X}_i, Z_i=0) = m_{0i} = \boldsymbol{X}_i^\intercal \boldsymbol{\zeta}_0$. Defining $\boldsymbol{\theta} = \left(\lambda_1, \lambda_0, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_0\right)^\intercal$ and assuming the outcome Y_i is missing at random, i.e. $R_i \perp\!\!\!\perp Y_i|(\boldsymbol{X}_i, Z_i)$, leads to the following simple g-computation estimating equations,

$$\sum_{i=1}^{n} \boldsymbol{u}_i(Y_i, R_i, Z_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = \boldsymbol{0}$$

$$\boldsymbol{u}_i(Y_i, R_i, Z_i, \boldsymbol{X}_i; \boldsymbol{\theta}) = \begin{pmatrix} Z_i(m_{1i} - \mu_1) \\ (1-Z_i)(m_{0i} - \mu_0) \\ Z_i R_i \boldsymbol{X}_i (Y_i - m_{1i}) \\ (1-Z_i) R_i \boldsymbol{X}_i (Y_i - m_{0i}) \end{pmatrix}$$

and single regression imputation estimators of the mean of the outcome in the two groups of the randomized trial, as follows,

$$\hat{\mu}_1 = \left\{\sum_{i=1}^{n} Z_i\right\}^{-1} \left\{\sum_{i=1}^{n} Z_i \hat{m}_{1i}\right\}, \quad \hat{\mu}_0 = \left\{\sum_{i=1}^{n} (1-Z_i)\right\}^{-1} \left\{\sum_{i=1}^{n} (1-Z_i)\hat{m}_{0i}\right\}$$

From M-estimation theory,

$$\text{var}(\hat{\boldsymbol{\theta}}) = n^{-1} \mathcal{A}^{-1} \mathcal{B} (\mathcal{A}^{-1})^\intercal$$

where $\mathcal{A} = \mathbb{E}\left(-\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\theta}^\intercal}\right)$ and $\mathcal{B} = \mathbb{E}\left(\boldsymbol{u}_i \boldsymbol{u}_i^\intercal\right)$.
The expression for $\mathcal{A}$ is:

$$\mathbb{E}\begin{pmatrix} Z_i \frac{\partial \mu_1}{\partial \lambda_1} & 0 & -Z_i \boldsymbol{X}_i^\intercal & 0 \\ 0 & (1-Z_i)\frac{\partial \mu_0}{\partial \lambda_0} & 0 & -(1-Z_i)\boldsymbol{X}_i^\intercal \\ 0 & 0 & Z_i R_i \boldsymbol{X}_i \boldsymbol{X}_i^\intercal & 0 \\ 0 & 0 & 0 & (1-Z_i) R_i \boldsymbol{X}_i \boldsymbol{X}_i^\intercal \end{pmatrix}$$

The expression for $\mathcal{B}$ is:

$$\mathbb{E}\begin{pmatrix} Z_i(m_{1i}-\mu_1)^2 & 0 & 0 & 0 \\ 0 & (1-Z_i)(m_{0i}-\mu_0)^2 & 0 & 0 \\ 0 & 0 & R_i Z_i (Y_i - m_{1i})^2 \boldsymbol{X}_i \boldsymbol{X}_i^\intercal & 0 \\ 0 & 0 & 0 & R_i(1-Z_i)(Y_i - m_{0i})^2 \boldsymbol{X}_i \boldsymbol{X}_i^\intercal \end{pmatrix}$$

Considering the parameters of interest, $\text{var}\{(\hat{\lambda}_1, \hat{\lambda}_0)\}$ can be simplified to,

$$\frac{1}{n}\begin{pmatrix} \kappa^{-1}\left(\frac{\partial \mu_1}{\partial \lambda_1}\right)^{-2}\left[E\{(m_{1i}-\mu_1)^2|Z_i=1\} + A\right] & 0 \\ 0 & (1-\kappa)^{-1}\left(\frac{\partial \mu_0}{\partial \lambda_0}\right)^{-2}\left[E\{(m_{0i}-\mu_0)^2|Z_i=0\} + B\right] \end{pmatrix}$$

where

$$A = E(\boldsymbol{X}_i^\intercal|Z_i=1) E(e_{1i}\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i=1)^{-1} E\{e_{1i}(Y_i-m_{1i})^2 \boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i=1\} E(e_{1i}\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i=1)^{-1} E(\boldsymbol{X}_i|Z_i=1)$$
$$B = E(\boldsymbol{X}_i^\intercal|Z_i=0) E(e_{0i}\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i=0)^{-1} E\{e_{0i}(Y_i-m_{0i})^2 \boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i=0\} E(e_{0i}\boldsymbol{X}_i\boldsymbol{X}_i^\intercal|Z_i=0)^{-1} E(\boldsymbol{X}_i|Z_i=0)$$
$$e_{1i} = P(R_i=1|\boldsymbol{X}_i, Z_i=1), \; e_{0i} = P(R_i=1|\boldsymbol{X}_i, Z_i=0)$$

Therefore,

$$\operatorname{var}\{g(\hat{\mu}_1) - g(\hat{\mu}_0)\} = n^{-1}\left(\kappa^{-1}\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2}\left[E\{(m_{1i}-\mu_1)^2|Z_i=1\}+A\right] + (1-\kappa)^{-1}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\left[E\{(m_{0i}-\mu_0)^2|Z_i=0\}+B\right]\right)$$

In the scenario where the covariates consist of one baseline categorical variable with $c = 1, \dots, C$ categories, we can write

$$\boldsymbol{X}_i^{\intercal} = \left(I(X_i=1), \dots, I(X_i=C)\right),\ \boldsymbol{\zeta}_1 = \left(\zeta_{11}, \dots, \zeta_{C1}\right)^{\intercal},\ \boldsymbol{\zeta}_0 = \left(\zeta_{10}, \dots, \zeta_{C0}\right)^{\intercal}$$

$$\mathbb{E}\{(m_{1i}-\mu_1)^2|Z_i=1\} = \sum_{c=1}^{C}\pi_c(\mu_{c1}-\mu_1)^2,\ E\{(m_{0i}-\mu_0)^2|Z_i=0\} = \sum_{c=1}^{C}\pi_c(\mu_{c0}-\mu_0)^2$$

$$\mathbb{E}\{I(X_i=c)|Z_i=1\} = \mathbb{E}\{I(X_i=c)|Z_i=0\} = \mathbb{E}\{I(X_i=c)\} = \pi_c$$

$$\mathbb{E}\{e_{1i}I(X_i=c)|Z_i=1\} = P(R_i=1|X_i=c,Z_i=1)\pi_c,\ \mathbb{E}\{e_{0i}I(X_i=c)|Z_i=0\} = P(R_i=1|X_i=c,Z_i=0)\pi_c$$

$$\mathbb{E}\{e_{1i}(Y_i-m_{1i})^2I(X_i=c)|Z_i=1\} = P(R_i=1|X_i=c,Z_i=1)\sigma_{c1}^2\pi_c,$$

$$\mathbb{E}\{e_{0i}(Y_i-m_{0i})^2I(X_i=c)|Z_i=0\} = P(R_i=1|X_i=c,Z_i=0)\sigma_{c0}^2\pi_c$$

where $\pi_c = P(X_i=c) = P(X_i=c|Z_i=1) = P(X_i=c|Z_i=0)$, $\sigma_{c1}^2 = \operatorname{var}(Y_i|X_i=c,Z_i=1)$, $\sigma_{c0}^2 = \operatorname{var}(Y_i|X_i=c,Z_i=0)$, $\mu_{c1} = \mathbb{E}(Y_i|X_i=c,Z_i=1)$ and $\mu_{c0} = \mathbb{E}(Y_i|X_i=c,Z_i=0)$.

$$A = \sum_{c=1}^{C}\frac{\pi_c\sigma_{c1}^2}{P(R_i=1|X_i=c,Z_i=1)},\ B = \sum_{c=1}^{C}\frac{\pi_c\sigma_{c0}^2}{P(R_i=1|X_i=c,Z_i=0)}$$

Therefore,

$$E\{(m_{1i}-\mu_1)^2|Z_i=1\}+A = \sum_{c=1}^{C}\pi_c(\mu_{c1}-\mu_1)^2 + \sum_{c=1}^{C}\frac{\pi_c\sigma_{c1}^2}{P(R_i=1|X_i=c,Z_i=1)}$$

Similarly,

$$E\{(m_{0i}-\mu_0)^2|Z_i=0\}+B = \sum_{c=1}^{C}\pi_c(\mu_{c0}-\mu_0)^2 + \sum_{c=1}^{C}\frac{\pi_c\sigma_{c0}^2}{P(R_i=1|X_i=c,Z_i=0)}$$

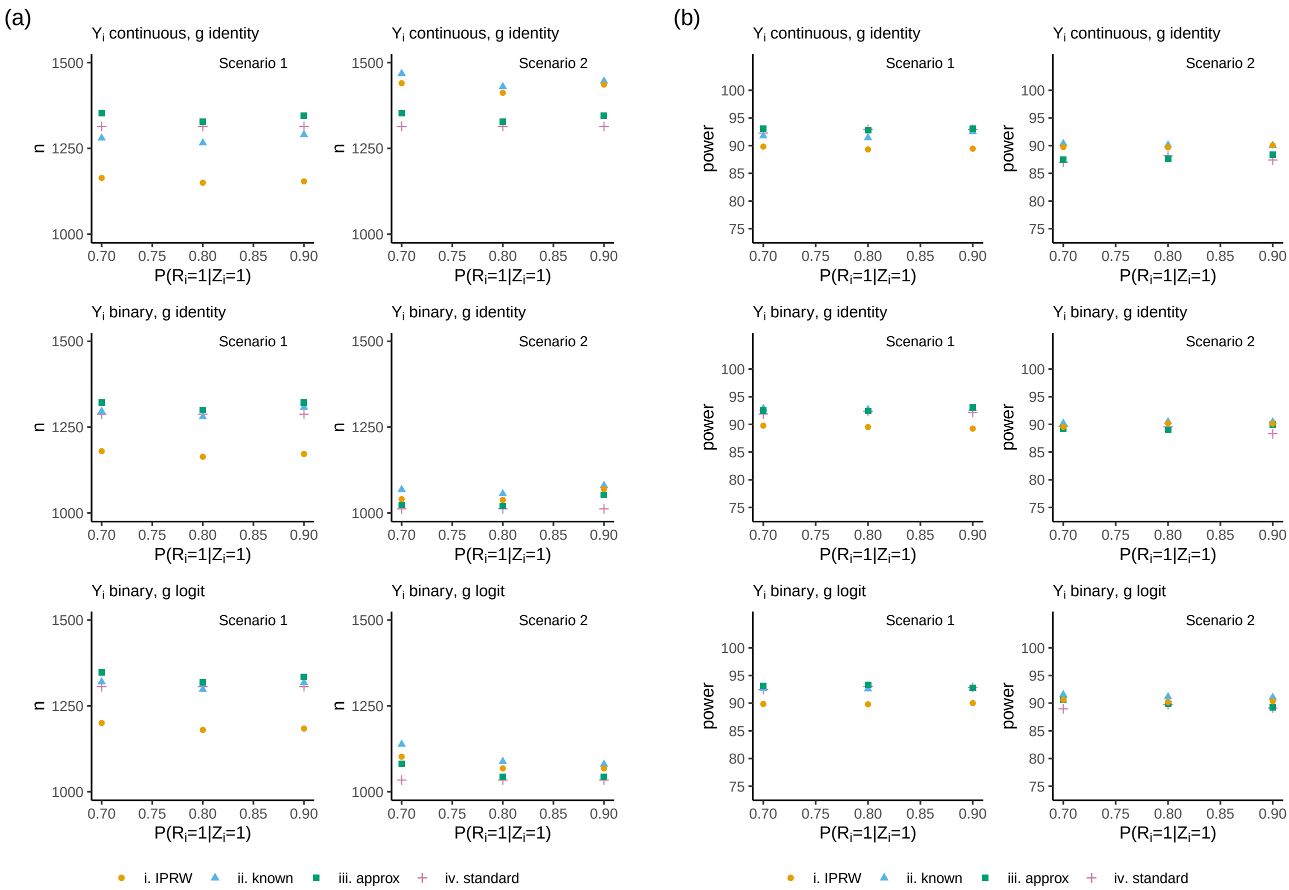
So,

$$\operatorname{var}\{g(\hat{\mu}_1)-g(\hat{\mu}_0)\} = n^{-1}\sum_{c=1}^{C}\left[\kappa^{-1}\pi_c\left\{\frac{\sigma_{c1}^2}{P(R_i=1|X_i=c,Z_i=1)}+(\mu_{c1}-\mu_1)^2\right\}\left(\frac{\partial\mu_1}{\partial\lambda_1}\right)^{-2} + (1-\kappa)^{-1}\pi_c\left\{\frac{\sigma_{c0}^2}{P(R_i=1|X_i=c,Z_i=0)}+(\mu_{c0}-\mu_0)^2\right\}\left(\frac{\partial\mu_0}{\partial\lambda_0}\right)^{-2}\right]$$

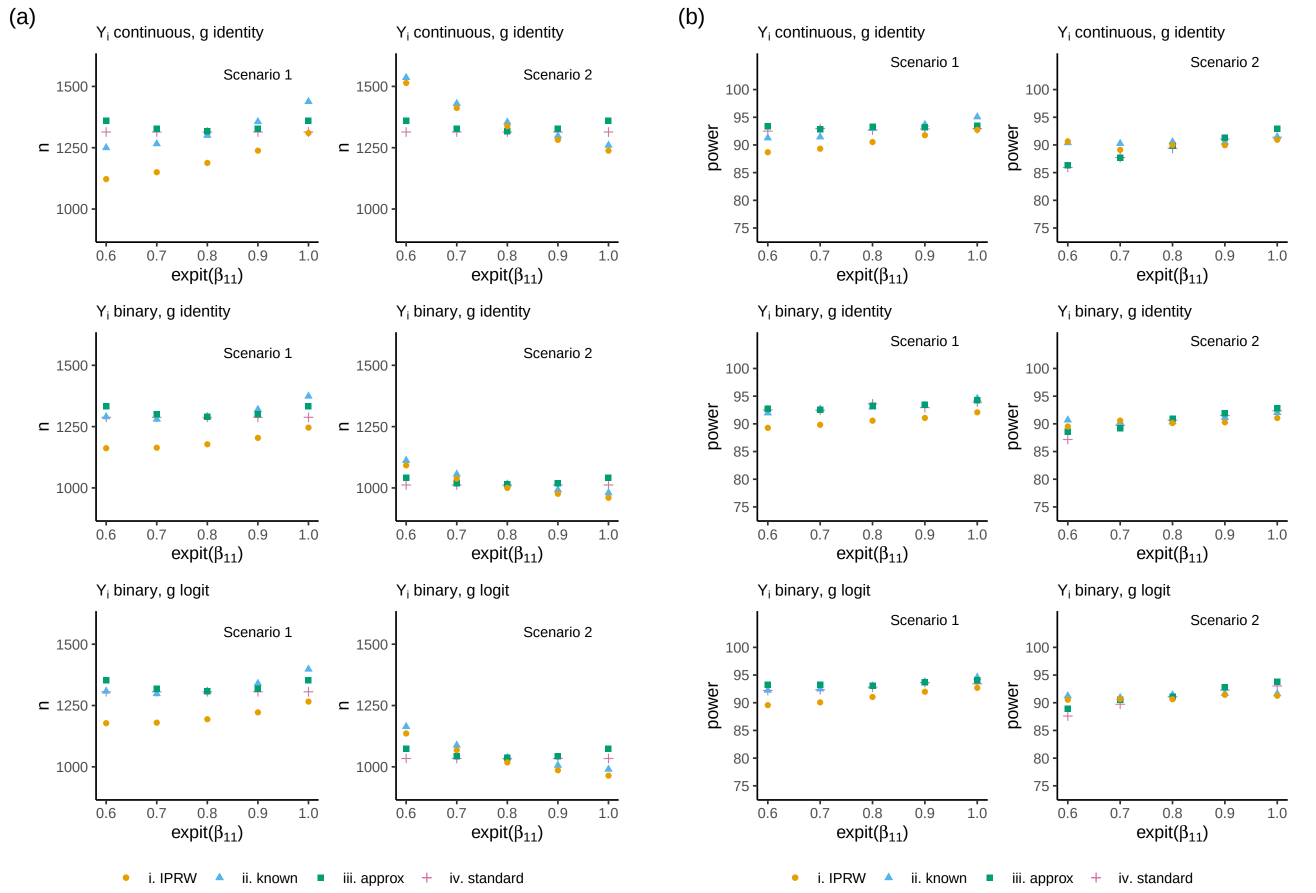
References

1. Stefanski LA, Boos DD. The calculus of M-estimation. *The American Statistician* 2002; 56(1): 29–38.
2. Chow SC, Shao J, Wang H, Lokhnygina Y. *Sample size calculations in clinical research*. CRC Press. second ed. 2008.
3. Little RJ, Rubin DB. *Statistical analysis with missing data*. John Wiley & Sons. second ed. 2002.
4. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine* 2012; 367(14): 1355–1360.
5. Donner A. Approaches to sample size estimation in the design of clinical trials—a review. *Statistics in Medicine* 1984; 3(3): 199–214.
6. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology* 2015; 44(3): 1051–1067. doi: 10.1093/ije/dyv113
7. Shook-Sa BE, Hudgens MG. Power and sample size for observational studies of point exposure effects. *Biometrics* 2020: doi:10.1111/biom.13405.
8. Qaqish BF. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 2003; 90(2): 455–463.
9. La Rosa AM, Harrison LJ, Taiwo B, et al. Raltegravir in second-line antiretroviral therapy in resource-limited settings (SELECT): a randomised, phase 3, non-inferiority study. *The Lancet HIV* 2016; 3(6): e247–e258.

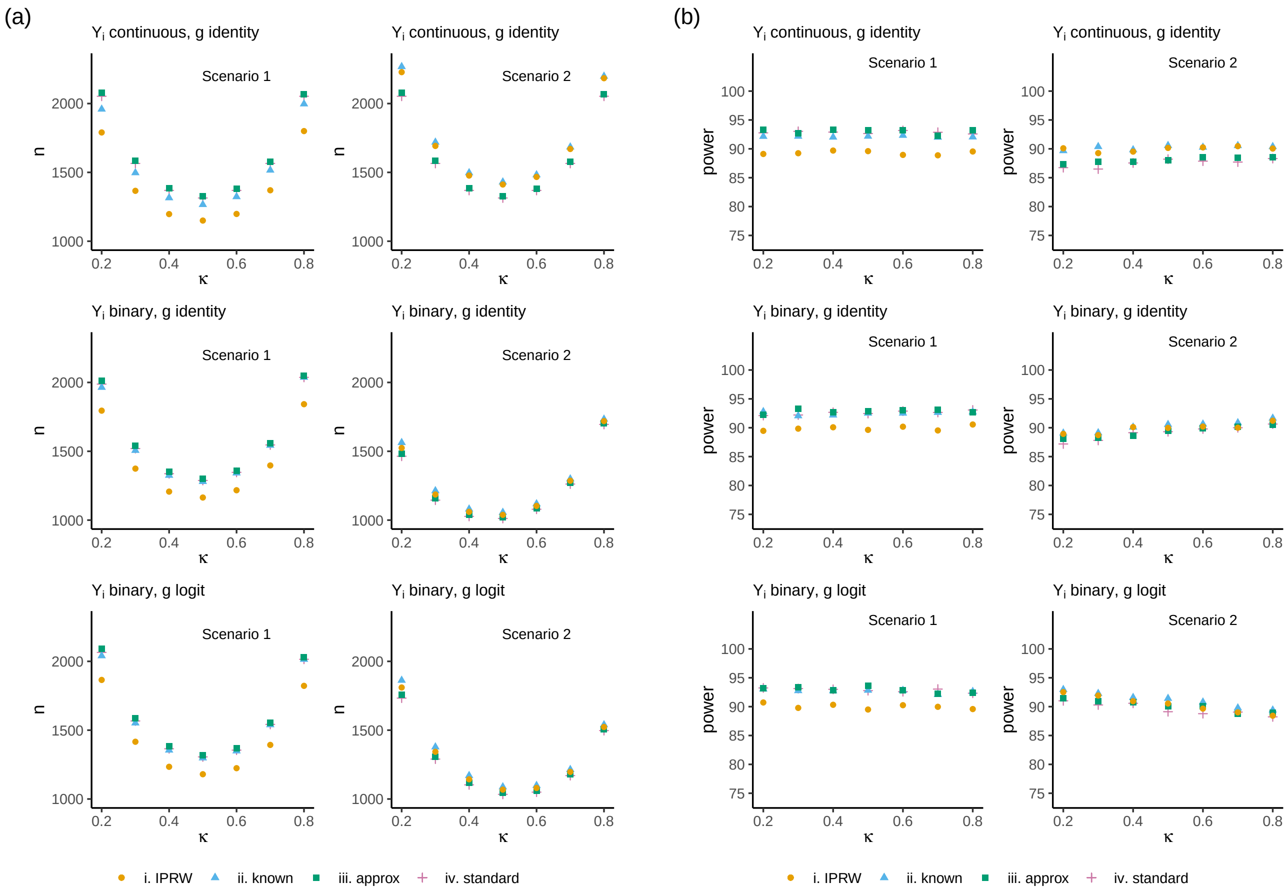
WEB TABLES AND WEB FIGURES



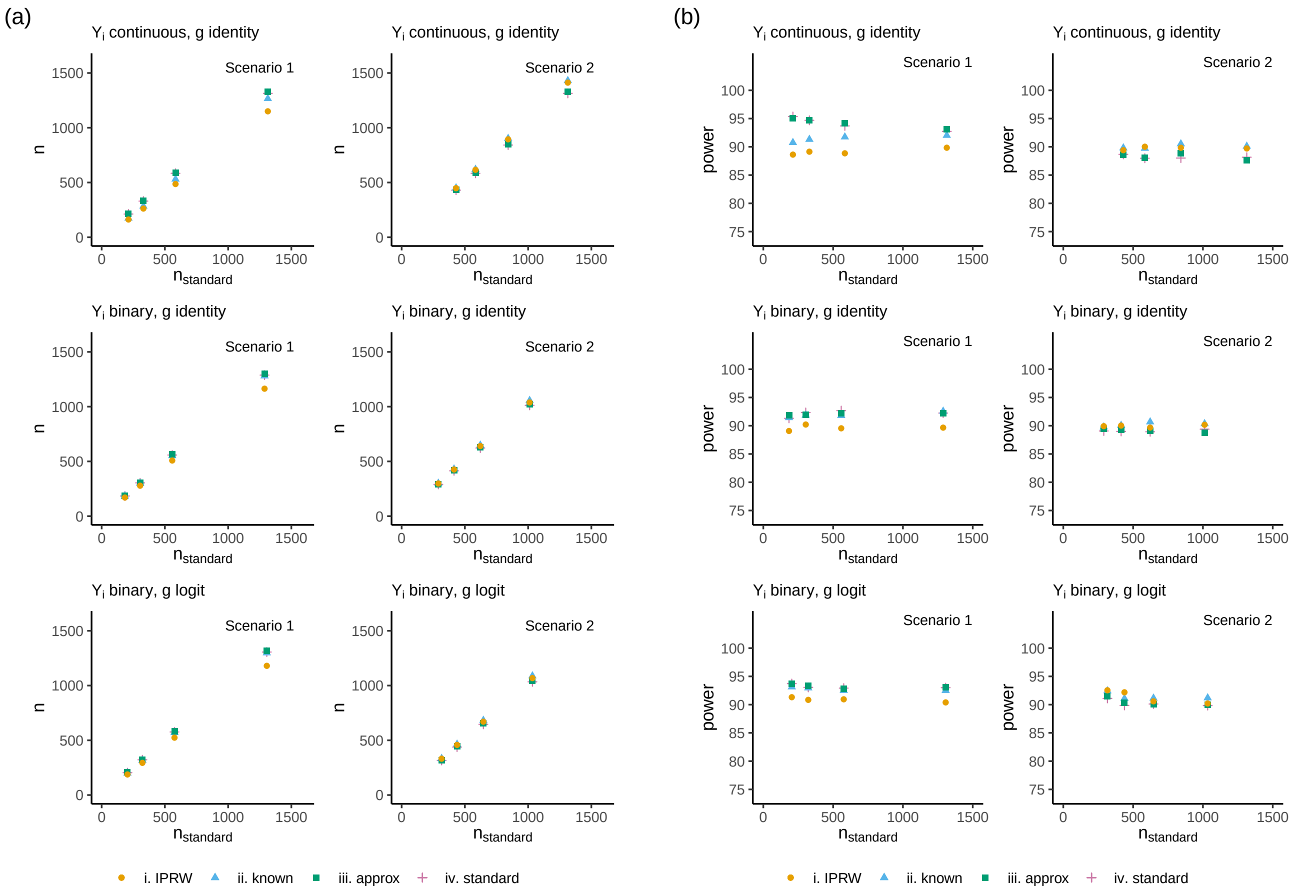
Web Figure 1. a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline binary covariate in an IRT, where the probability of the outcome being observed in the intervention group $P(R_i = 1|Z_i = 1)$ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



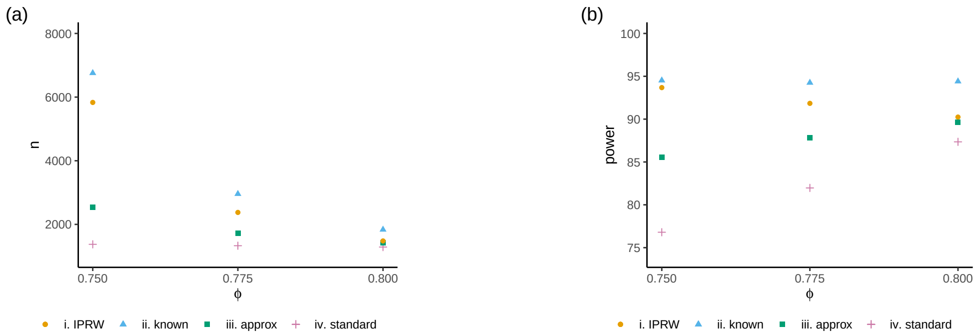
Web Figure 2. a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline binary covariate in an IRT, where the probability of being observed in the intervention group when $X_i = 1$[i.e. $P(R_i = 1|X_i = 1, Z_i = 1) = \text{expit}(\beta_{11})$] is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



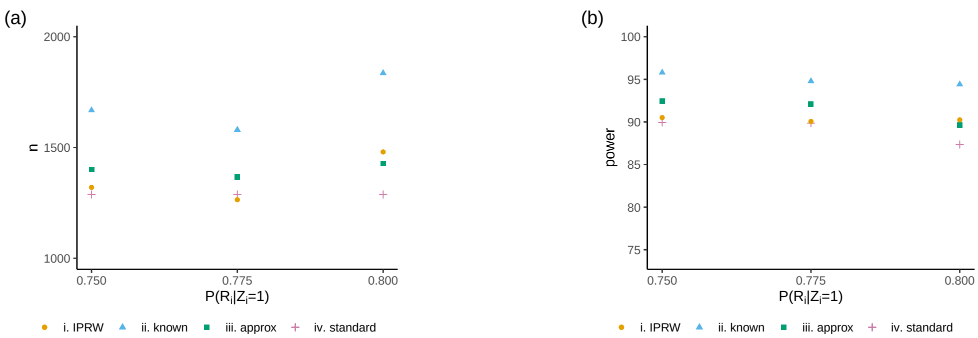
Web Figure 3. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline binary covariate in an IRT, where the probability of being randomized to the intervention group $\kappa = P(Z_i = 1)$ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



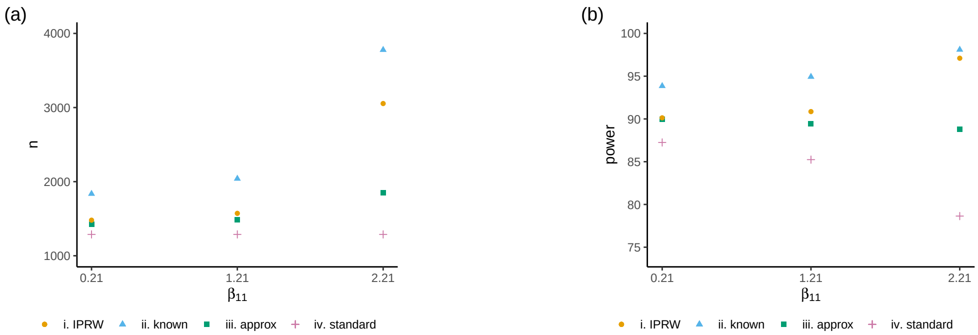
Web Figure 4. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline binary covariate in an IRT, where the sample size is varied on the x-axis by varying μ_{21} or μ_{11}. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



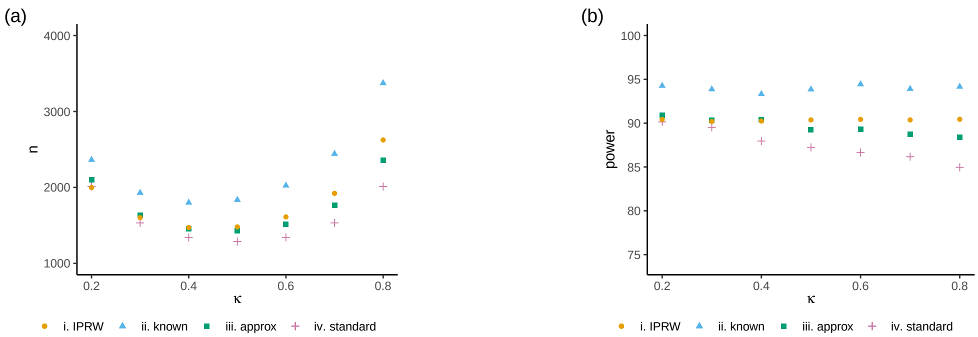
Web Figure 5. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline continuous covariate in an IRT, where the probability of the outcome being observed $\phi = P(R_i = 1)$ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



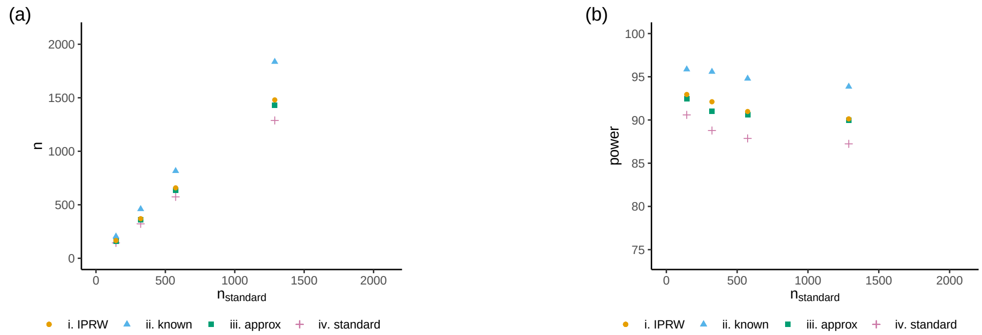
Web Figure 6. a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline continuous covariate in an IRT, where the probability of the outcome being observed in the intervention group $P(R_i = 1 | Z_i = 1)$ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



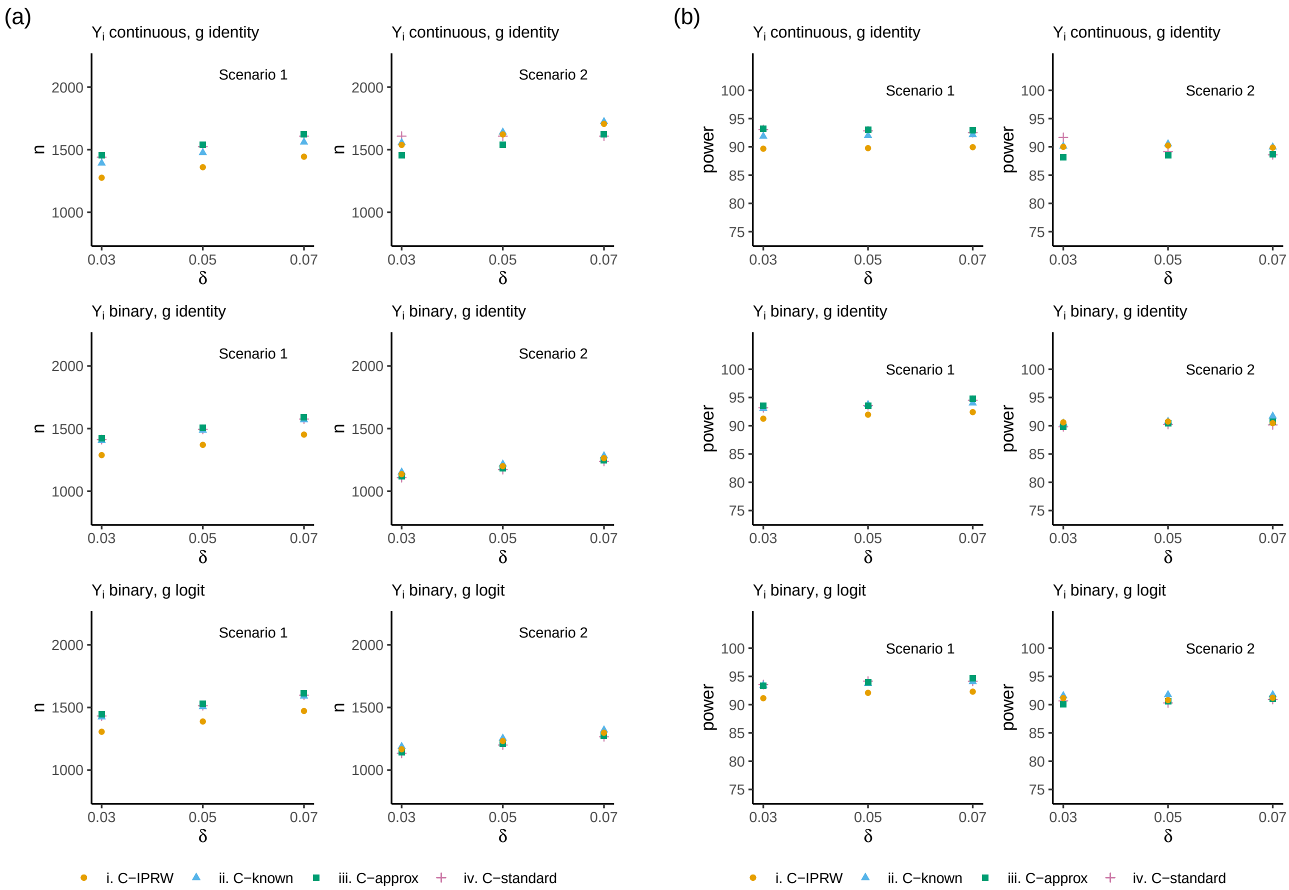
Web Figure 7. a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline continuous covariate in an IRT, where β_{11} is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



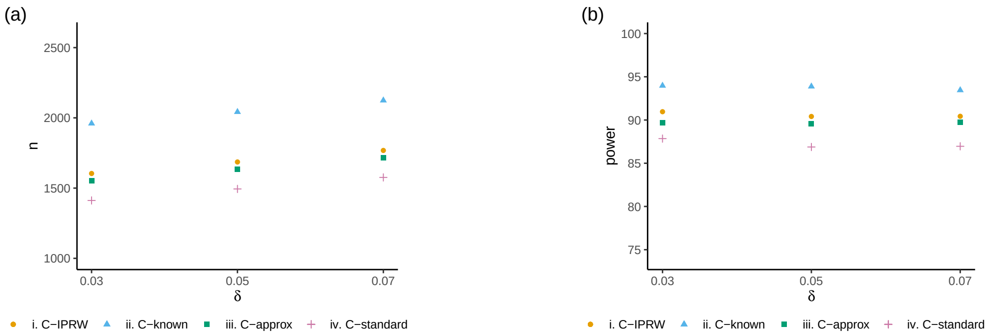
Web Figure 8. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline continuous covariate in an IRT, where the probability of being randomized to the intervention group $\kappa = P(Z_i = 1)$ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



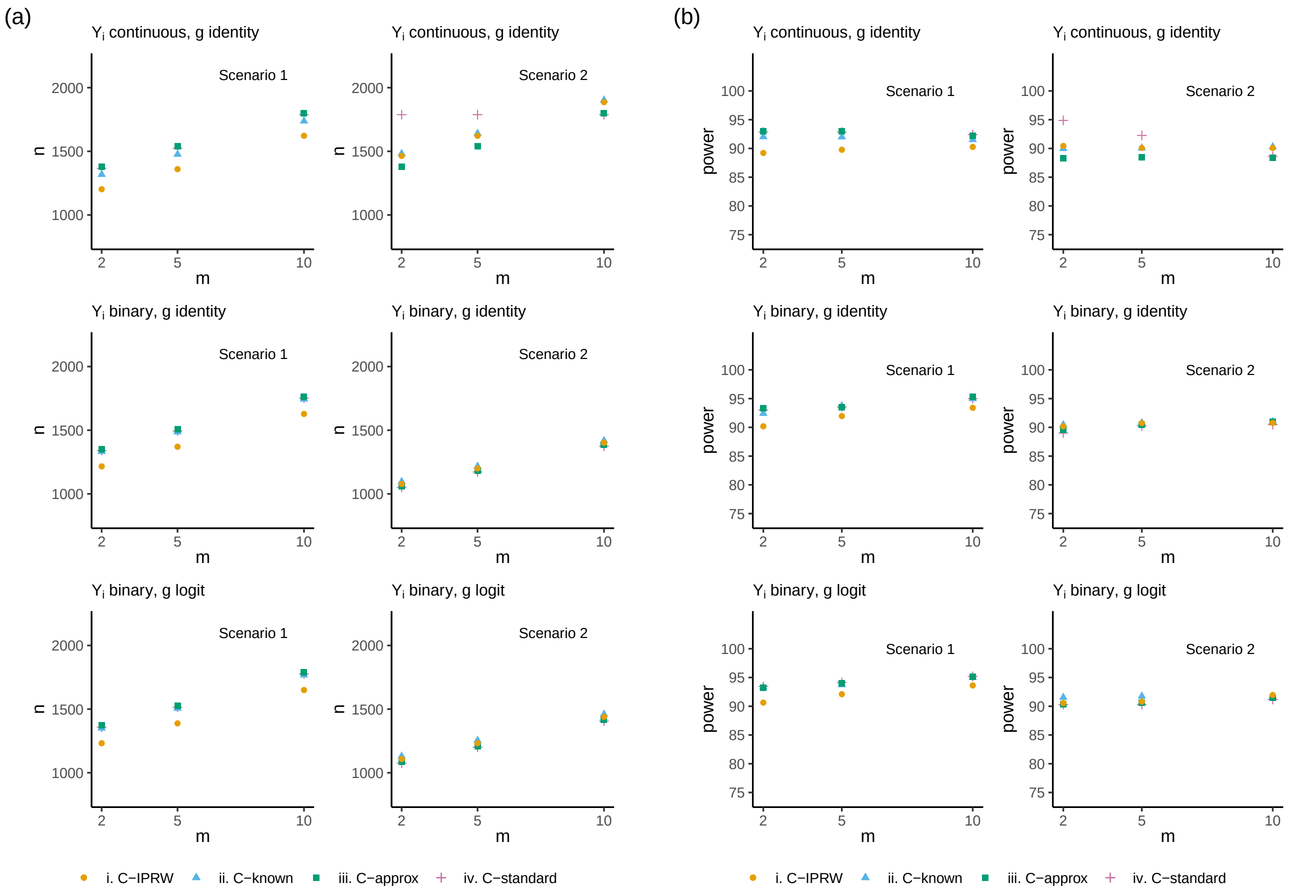
Web Figure 9. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas when weighting by a baseline continuous covariate in an IRT, where the sample size is varied on the x-axis by varying μ_1. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



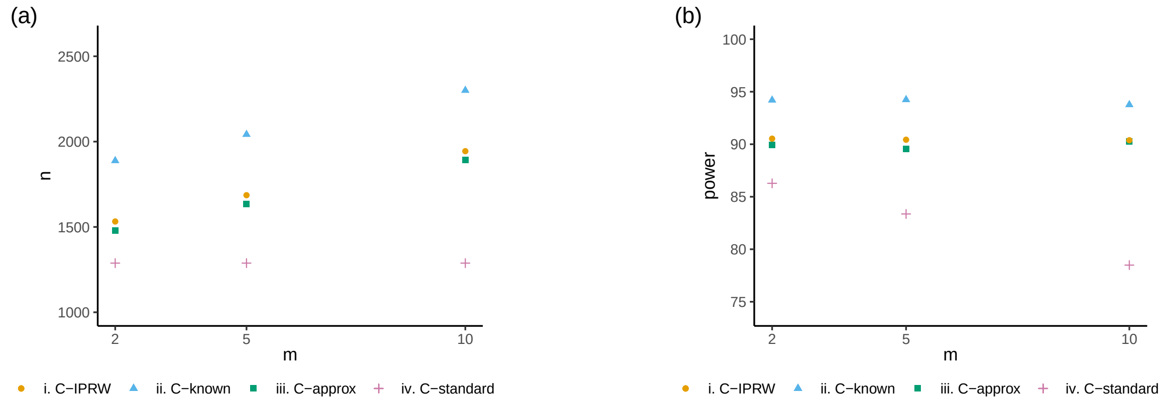
Web Figure 10. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the $n_{\text{C-IPRW}}$, $n_{\text{C-known}}$, $n_{\text{C-approx}}$ and $n_{\text{C-standard}}$ formulas when weighting by a baseline binary covariate in a CRT, where the intercluster correlation δ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.



Web Figure 11. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the $n_{\text{C-IPRW}}$, $n_{\text{C-known}}$, $n_{\text{C-approx}}$ and $n_{\text{C-standard}}$ formulas when weighting by a baseline continuous covariate in a CRT, where the intercluster correlation δ is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.

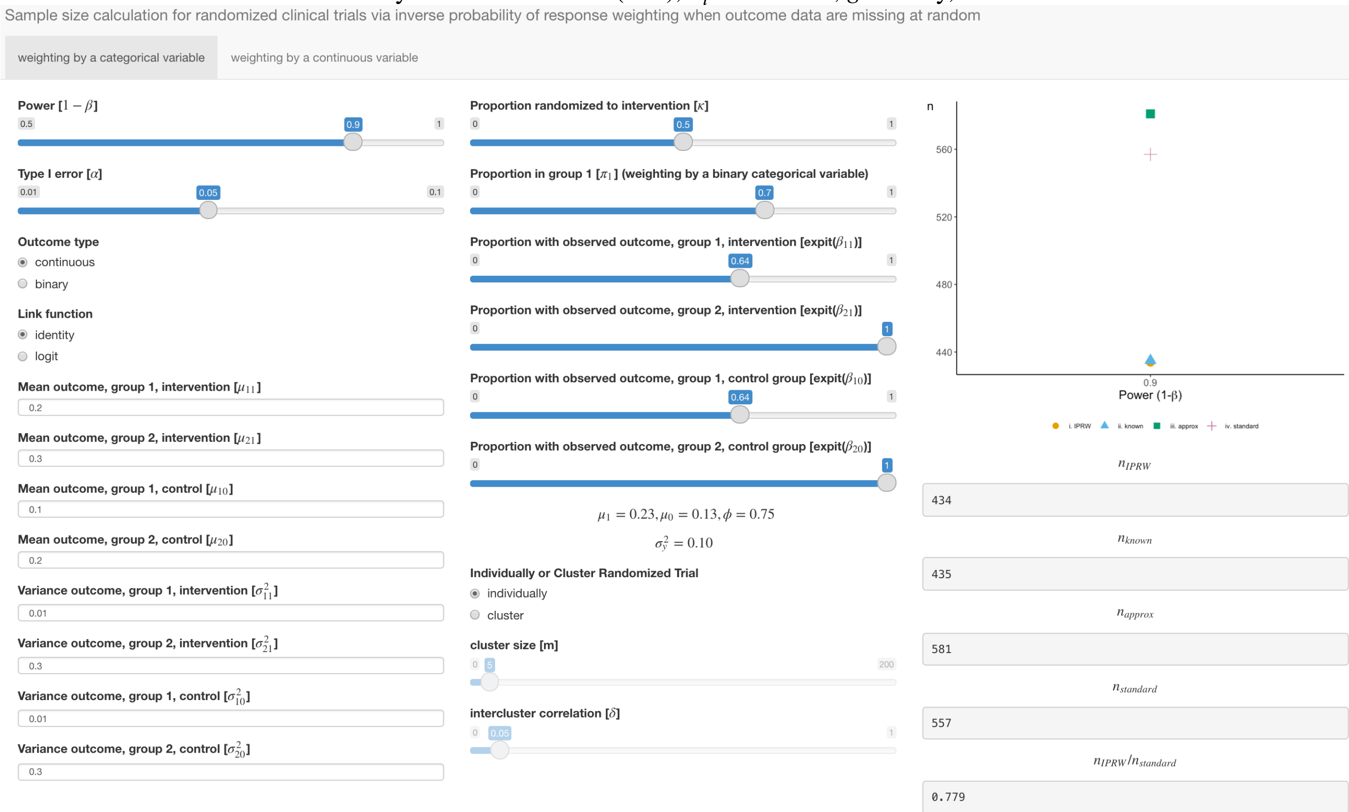


Web Figure 12. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the $n_{\text{C-IPRW}}$, $n_{\text{C-known}}$, $n_{\text{C-approx}}$ and $n_{\text{C-standard}}$ formulas when weighting by a baseline binary covariate in a CRT, where the cluster size m is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.

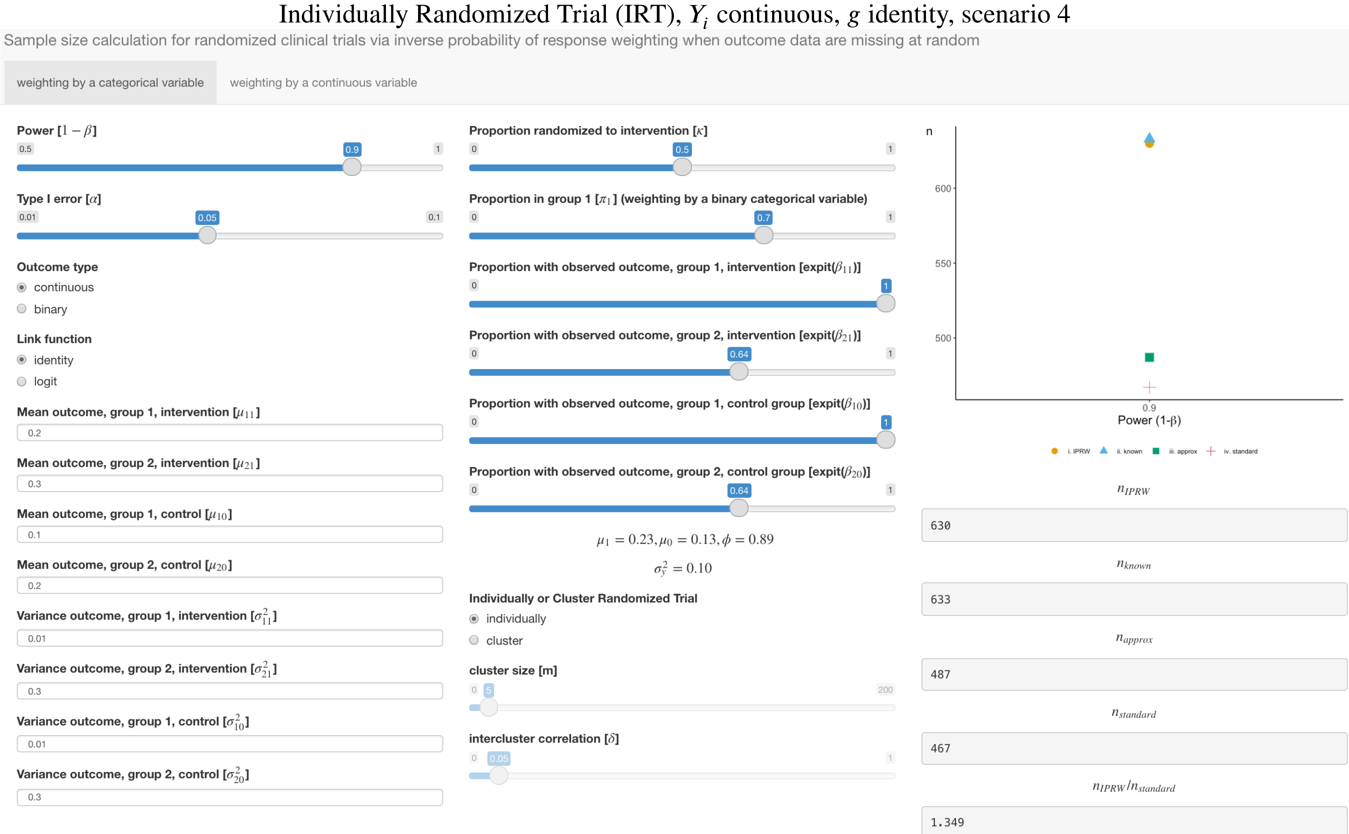


Web Figure 13. (a) Datasets generated with the sample size calculated to have 90% power at a two-sided 5% significance level using the $n_{\text{C-IPRW}}$, $n_{\text{C-known}}$, $n_{\text{C-approx}}$ and $n_{\text{C-standard}}$ formulas when weighting by a baseline continuous covariate in a CRT, where the cluster size m is varied on the x-axis. (b) Simulation results displaying the empirical power for each sample size with the IPRW estimator.

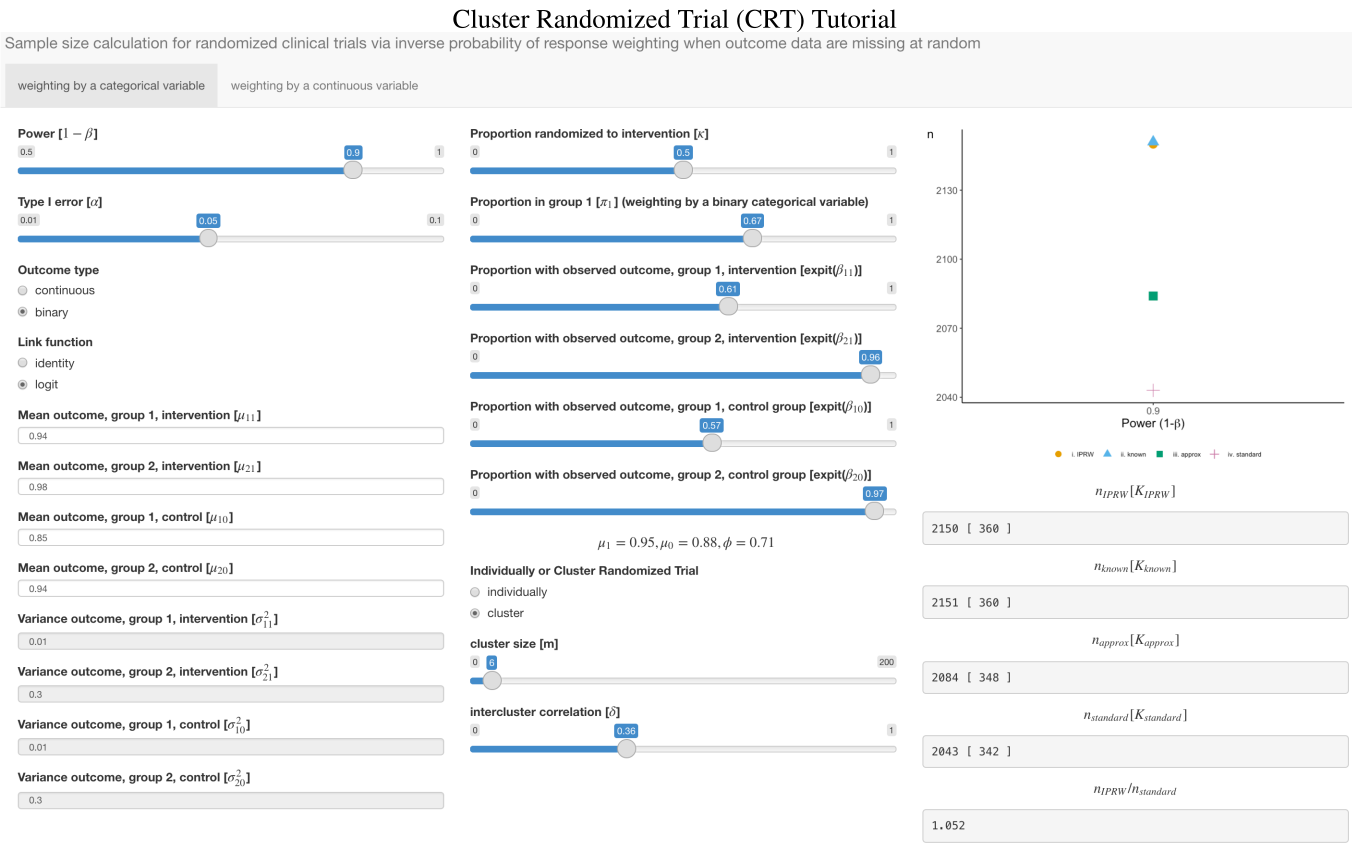
Individually Randomized Trial (IRT), Y_i continuous, g identity, scenario 3



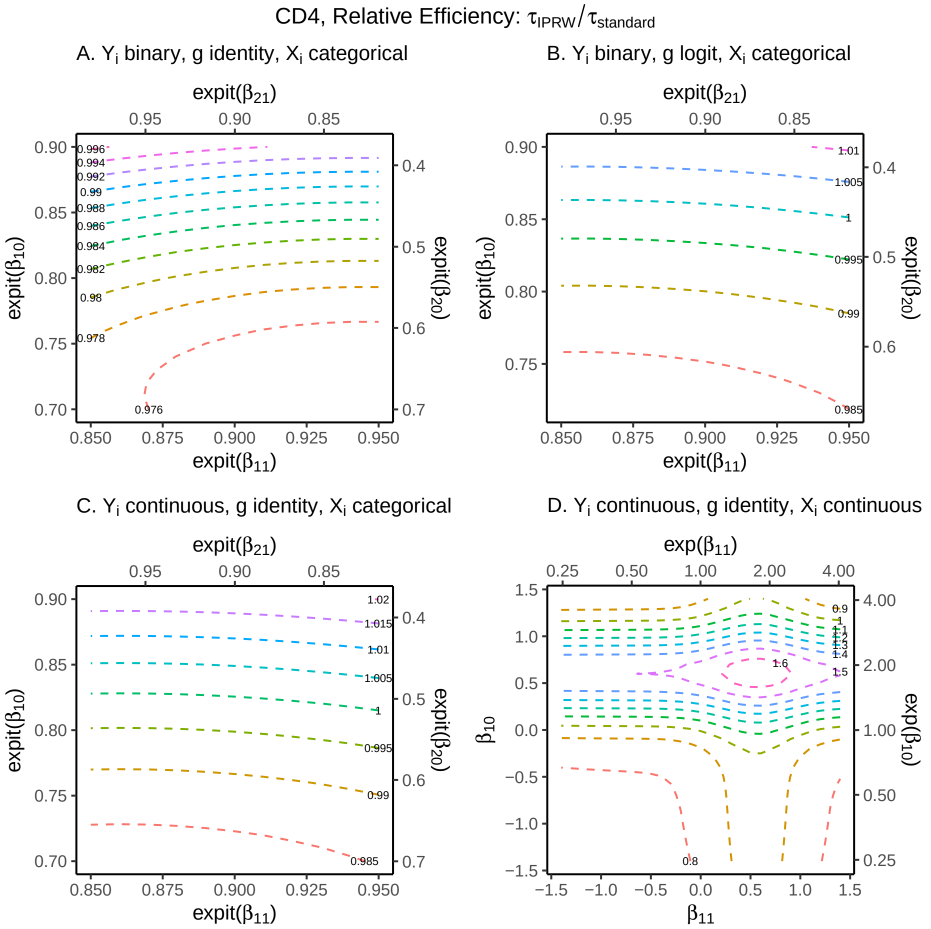
Individually Randomized Trial (IRT), Y_i continuous, g identity, scenario 4



Web Figure 14. R shiny app illustration to calculate the sample size by the n_{IPRW}, n_{known}, n_{approx} and n_{standard} formulas. Note, sample sizes were rounded up to the nearest even number to enable simulation of half the participants in each randomized intervention group.

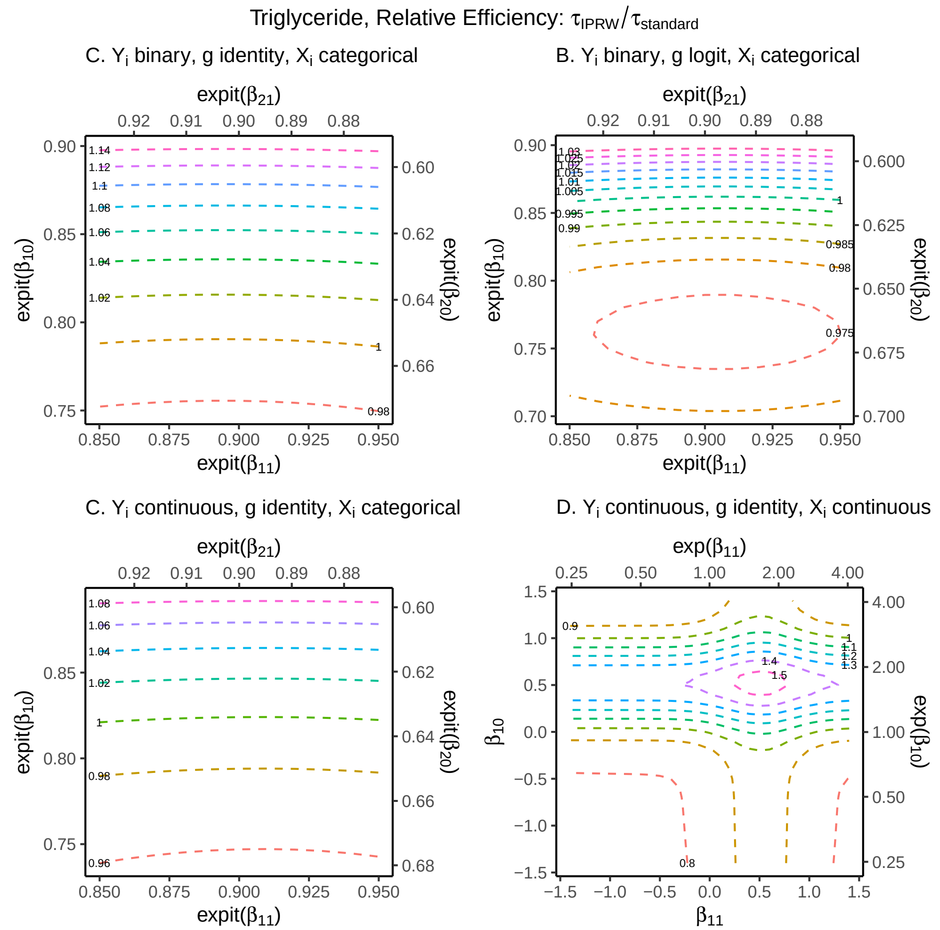


Web Figure 15. R shiny app CRT tutorial to calculate sample size by $n_{\text{IPRW}}[K_{\text{IPRW}}]$, $n_{\text{known}}[K_{\text{known}}]$, $n_{\text{approx}}[K_{\text{approx}}]$ and $n_{\text{standard}}[K_{\text{standard}}]$ for the example in Section 4.1.



Web Figure 16. Relative efficiency based on the inverse probability of response weighted (IPRW) versus standard approach for a CD4 count outcome. The contours display the ratio $\tau_{\text{IPRW}}/\tau_{\text{standard}}$ for different magnitudes of the association between the fully observed baseline covariate (X_i) and the probability of the outcome being observed $\{P(R_i|X_i, Z_i)\}$.

$$\tau_{\text{IPRW}} = \begin{cases} A. & \sum_{c=1}^{C}\left[\kappa^{-1}\pi_c\left\{\frac{\mu_{c1}(1-\mu_{c1})}{\text{expit}(\beta_{c1})} + (\mu_{c1}-\mu_1)^2\right\} + (1-\kappa)^{-1}\pi_c\left\{\frac{\mu_{c0}(1-\mu_{c0})}{\text{expit}(\beta_{c0})} + (\mu_{c0}-\mu_0)^2\right\}\right] \\ & = 2\times 0.38\left\{\frac{0.23}{\text{expit}(\beta_{11})} + 0.06\right\} + 2\times 0.38\left\{\frac{0.20}{\text{expit}(\beta_{10})} + 0.08\right\} \\ & +2\times 0.62\left\{\frac{0.20}{\text{expit}(\beta_{21})} + 0.02\right\} + 2\times 0.62\left\{\frac{0.19}{\text{expit}(\beta_{20})} + 0.03\right\} \\ B. & \sum_{c=1}^{C}\left[\kappa^{-1}\frac{\pi_c}{\{\mu_1(1-\mu_1)\}^2}\left\{\frac{\mu_{c1}(1-\mu_{c1})}{\text{expit}(\beta_{c1})} + (\mu_{c1}-\mu_1)^2\right\} + (1-\kappa)^{-1}\frac{\pi_c}{\{\mu_0(1-\mu_0)\}^2}\left\{\frac{\mu_{c0}(1-\mu_{c0})}{\text{expit}(\beta_{c0})} + (\mu_{c0}-\mu_0)^2\right\}\right] \\ & = \frac{2\times 0.38}{0.06}\left\{\frac{0.23}{\text{expit}(\beta_{11})} + 0.06\right\} + \frac{2\times 0.38}{0.06}\left\{\frac{0.20}{\text{expit}(\beta_{10})} + 0.08\right\} \\ & +\frac{2\times 0.62}{0.06}\left\{\frac{0.20}{\text{expit}(\beta_{21})} + 0.02\right\} + \frac{2\times 0.62}{0.06}\left\{\frac{0.19}{\text{expit}(\beta_{20})} + 0.03\right\} \\ C. & \sum_{c=1}^{C}\left[\kappa^{-1}\pi_c\left\{\frac{\sigma_{c1}^2}{\text{expit}(\beta_{c1})} + (\mu_{c1}-\mu_1)^2\right\} + (1-\kappa)^{-1}\pi_c\left\{\frac{\sigma_{c0}^2}{\text{expit}(\beta_{c0})} + (\mu_{c0}-\mu_0)^2\right\}\right] \\ & = 2\times 0.38\left\{\frac{0.24}{\text{expit}(\beta_{11})} + 0.05\right\} + 2\times 0.38\left\{\frac{0.19}{\text{expit}(\beta_{10})} + 0.06\right\} \\ & +2\times 0.62\left\{\frac{0.28}{\text{expit}(\beta_{21})} + 0.01\right\} + 2\times 0.62\left\{\frac{0.24}{\text{expit}(\beta_{20})} + 0.02\right\} \\ D. & \text{no closed-form, based on Gauss-Hermite quadrature with 100 quadrature points} \end{cases}$$



Web Figure 17. Relative efficiency based on the inverse probability of response weighted (IPRW) versus standard approach for a triglyceride outcome. The contours display the ratio $\tau_{\text{IPRW}}/\tau_{\text{standard}}$ for different magnitudes of the association between the fully observed baseline covariate (X_i) and the probability of the outcome being observed $\{P(R_i|X_i, Z_i)\}$.

$$
\tau_{\text{IPRW}} = \begin{cases}
A. & \sum_{c=1}^{C}\left[\kappa^{-1}\pi_c\left\{\frac{\mu_{c1}(1-\mu_{c1})}{\text{expit}(\beta_{c1})} + (\mu_{c1}-\mu_1)^2\right\} + (1-\kappa)^{-1}\pi_c\left\{\frac{\mu_{c0}(1-\mu_{c0})}{\text{expit}(\beta_{c0})} + (\mu_{c0}-\mu_0)^2\right\}\right] \\
 & = 2\times 0.65\left\{\frac{0.19}{\text{expit}(\beta_{11})} + 0.03\right\} + 2\times 0.65\left\{\frac{0.14}{\text{expit}(\beta_{10})} + 0.02\right\} \\
 & +2\times 0.35\left\{\frac{0.19}{\text{expit}(\beta_{21})} + 0.10\right\} + 2\times 0.35\left\{\frac{0.23}{\text{expit}(\beta_{20})} + 0.11\right\} \\
B. & \sum_{c=1}^{C}\left[\kappa^{-1}\frac{\pi_c}{\{\mu_1(1-\mu_1)\}^2}\left\{\frac{\mu_{c1}(1-\mu_{c1})}{\text{expit}(\beta_{c1})} + (\mu_{c1}-\mu_1)^2\right\} + (1-\kappa)^{-1}\frac{\pi_c}{\{\mu_0(1-\mu_0)\}^2}\left\{\frac{\mu_{c0}(1-\mu_{c0})}{\text{expit}(\beta_{c0})} + (\mu_{c0}-\mu_0)^2\right\}\right] \\
 & = \frac{2\times 0.65}{0.06}\left\{\frac{0.19}{\text{expit}(\beta_{11})} + 0.03\right\} + \frac{2\times 0.65}{0.05}\left\{\frac{0.14}{\text{expit}(\beta_{10})} + 0.02\right\} \\
 & +\frac{2\times 0.35}{0.06}\left\{\frac{0.19}{\text{expit}(\beta_{21})} + 0.10\right\} + \frac{2\times 0.35}{0.05}\left\{\frac{0.23}{\text{expit}(\beta_{20})} + 0.11\right\} \\
C. & \sum_{c=1}^{C}\left[\kappa^{-1}\pi_c\left\{\frac{\sigma^2_{c1}}{\text{expit}(\beta_{c1})} + (\mu_{c1}-\mu_1)^2\right\} + (1-\kappa)^{-1}\pi_c\left\{\frac{\sigma^2_{c0}}{\text{expit}(\beta_{c0})} + (\mu_{c0}-\mu_0)^2\right\}\right] \\
 & = 2\times 0.65\left\{\frac{0.33}{\text{expit}(\beta_{11})} + 0.09\right\} + 2\times 0.65\left\{\frac{0.28}{\text{expit}(\beta_{10})} + 0.04\right\} \\
 & +2\times 0.35\left\{\frac{0.32}{\text{expit}(\beta_{21})} + 0.31\right\} + 2\times 0.35\left\{\frac{0.39}{\text{expit}(\beta_{20})} + 0.23\right\} \\
D. & \text{no closed-form, based on Gauss-Hermite quadrature with 100 quadrature points}
\end{cases}
$$