A GLS Estimator for Combining Data from Probability and Nonprobability Samples

Emily Berg
Iowa State University
Joint work with Chengpeng Zeng and Zhengyuan Zhu

August 4, 2024

Outline

- Introduction: Why combine probability and non-probability samples?
  - Review existing methods
  - Overview our approach
- GLS estimators for data integration
  - Complete response in the probability sample
  - Nonresponse in the probability sample
- Simulations to evaluate alternative estimates of the mean
  - Compare alternatives with no nonresponse in the probability sample
  - Nonresponse in the probability sample

Contrasting Probability and Nonprobability Samples

	Probability surveys	**Non-probability Surveys**

Strengths	• Known inclusion probability
	  ○ Unbiased estimates
	  ○ Design-consistency
	• Quality control in data collection
		• Inexpensive
		• Naturally collect variables of interest
		• Large sample size

Weaknesses	• Rising nonresponse rates • Increasing costs	• Unknown selection mechanism ○ Selection bias • Lack clear data collection protocols

- Baker et al. (2013)

Three Cases

- Notation
  - Y = primary variable of interest
  - X = auxiliary variable

	Prob. Samp.		Nonprob. Samp.	
	X	Y	X	Y
Case 1:	✓		✓	✓
Case 2:	✓	✓	✓	
Case 3:	✓	✓	✓	✓

- Beaumont (2020), Wu (2022), Valliant (2020), Savitsky et al. (2022)

Related Literature: Case 1 and 2

- Case 1 (common): Y only in nonprobability sample
  - Propensity score (Valliant & Dever 2011)
    - Weight by inverse of estimated probability of participation in nonprobability sample
  - Mass imputation (Kim et al. 2021)
    - Impute Y for every element of probability sample
  - Doubly-robust (Chen et al. 2020)
    - Combines strengths of mass imputation and propensity score estimators
- Case 2 (rare): Y only in probability sample
  - Medous et al. (2022)

Related Literature: Case 3

- Elliott & Valliant (2017)
  - Requires the probability of inclusion in prob sample for elements only in the nonprobability sample
  - Otherwise, use a model for the probability of inclusion in the probability sample
- Kim et al. (2021)
  - Requires knowledge of which probability sample elements are in the nonprobability sample
- Zhu et al. (2023)
  - Define a convex combination of estimators based on prob and nonprob samples
  - Ignore correlation between the two estimators
- Chen et al. (2023)
  - Preliminary test procedure that simplifies to the probability sample in the presence of nonignorable selection

Case 3: Motivating Applications

- Survey of attitudes toward adopting energy conservation strategies (Iowa State University)
  - Probability sample to be combined with volunteer internet panel data
- Health parameters such as body mass index
  - NHANES probability sample
  - Administrative data

Innovations of Our Case 3 Procedure

- Develop a GLS estimator to combine estimates from probability and nonprobability samples
  - Accounts for correlation between the two estimates
- Explore the possibility of allowing nonignorable selection into the nonprobability sample
- Allow for nonresponse in the probability sample

Framework

- Finite population: $(x_i, y_i) \overset{iid}{\sim} F(x, y)\ i = 1, \ldots, N$
  - **Parameter of interest**: $\mu_y = N^{-1} \sum_{i=1}^{N} y_i$
- Probability sample B $(S_B) \subset \{1, \ldots, N\}$
  - $I_i = I(i \in B)$
  - **Known**: $\pi_i = P(I_i = 1)$, $d_i^B = \pi_i^{-1}$
  - **Known**: $\pi_{ij} = P(I_i = 1, I_j = 1)$
  - x_i, y_i observed for $i \in B$
- Nonprobability sample A $(S_A) \subset \{1, \ldots, N\}$
  - $\delta_i = I(i \in A)$
  - **Unknown**: $p_i^A = P(\delta_i = 1)$
  - x_i, y_i observed for $i \in A$
- Conditions
  - $p_i^A > 0 : i = 1, \ldots, N$
  - $\delta_i \perp I_i \mid x_i, y_i$

Propensity Score Model

$$p_i^A(\boldsymbol{\alpha}) = \frac{\exp(\alpha_0 + \alpha_1 x_i + \alpha_2 y_i)}{1 + \exp(\alpha_0 + \alpha_1 x_i + \alpha_2 y_i)}$$

- Ignorable: $\alpha_2 = 0$
- Nonignorable: $\alpha_2 \neq 0$

Pseudo-Likelihood Estimator (Chen et al. 2020)

$$\hat{\boldsymbol{\alpha}} = argmax_{\boldsymbol{\alpha}} \ell^*(\boldsymbol{\alpha}), \quad U(\hat{\boldsymbol{\alpha}}) = \mathbf{0}$$

$$\ell^*(\boldsymbol{\alpha}) = \sum_{i \in S_A} (\alpha_0 + \alpha_1 x_i + \alpha_2 y_i) - \sum_{i \in S_B} d_i^B log(1 - p_i^A(\boldsymbol{\alpha}))$$

$$U(\boldsymbol{\alpha}) = \sum_{i \in S_A} (1, x_i, y_i)' - \sum_{i \in S_B} d_i^B p_i^A(\boldsymbol{\alpha})(1, x_i, y_i)'$$

Two Estimators of μ_y

Horvitz-Thompson

$$\hat{\theta}^B = \frac{1}{N} \sum_{i \in S_B} d_i^B y_i.$$

Propensity-Score

$$\hat{\theta}^A = \frac{1}{N} \sum_{i \in S_A} \frac{1}{p_i^A(\hat{\alpha})} y_i.$$

How can we combine the two estimators?

An Estimated Generalized Least Squares (EGLS) Estimator

$$\hat{\mu}_y = (\mathbf{1}'\hat{\boldsymbol{V}}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\boldsymbol{V}}^{-1}(\hat{\theta}^A, \hat{\theta}^B)'$$

$$\hat{V}(\hat{\mu}_y) = (\mathbf{1}'\hat{\boldsymbol{V}}^{-1}\mathbf{1})^{-1}$$

$$\hat{\boldsymbol{V}} = \begin{pmatrix} \hat{V}_A & \hat{C}_{AB} \\ \hat{C}_{AB} & \hat{V}_B \end{pmatrix}$$

- $\hat{V}_A$ is an estimate of the variance of $\hat{\theta}_A$, $\hat{V}_B$ is an estimate of the variance of $\hat{\theta}_B$, and $\hat{C}_{AB}$ is an estimate of the covariance between $\hat{\theta}_A$ and $\hat{\theta}_B$.
- We obtain the estimated variances and covariances from Taylor linearization

Taylor Linearization

Theorem 1: Let $n = min\{n_A, n_B\}$, where n_A is the expected sample size for S_A and n_B is the expected sample size for S_B. Let $\boldsymbol{z}_i = (1, x_i, y_i)'$. Assume the following regularity conditions:

- $lim_{n\to\infty} N^{-1} \sum_{i\in S_A} p_i^A(\boldsymbol{\alpha})^{-1}(1 - p_i^A(\boldsymbol{\alpha}))y_i\boldsymbol{z}_i' = \boldsymbol{\gamma}_{y,\infty}$, where $\boldsymbol{\gamma}_{y,\infty}$ is a fixed vector.
- $lim_{n\to\infty} N^{-1} \sum_{i\in S_B} d_i^B \boldsymbol{z}_i p_i^A(\boldsymbol{\alpha})(1 - p_i^A(\boldsymbol{\alpha}))\boldsymbol{z}_i' = \boldsymbol{M}_\infty$, where $\boldsymbol{M}_\infty$ is a positive definite matrix.
- $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = O_p(n^{-0.5})$.

Then,

$$\hat{\theta}^A - \theta = \frac{1}{N}\sum_{i\in S_A}\{\frac{1}{p_i^A(\boldsymbol{\alpha})}y_i - \theta - \boldsymbol{\gamma}_{y,\infty}\boldsymbol{M}_\infty^{-1}\boldsymbol{z}_i\}$$
$$+ N^{-1}\boldsymbol{\gamma}_{y,\infty}\boldsymbol{M}_\infty^{-1}\sum_{i\in S_B} d_i^B p_i^A(\boldsymbol{\alpha})\boldsymbol{z}_i + O_p(n^{-1}).$$

Obtaining $\hat{V}$

$$\hat{V}_A = \left(\frac{1}{N}\right)^2 \sum_{i \in S_A} \frac{1}{p_i^A(\hat{\boldsymbol{\alpha}})} \left(\frac{1}{p_i^A(\hat{\boldsymbol{\alpha}})} - 1\right) \hat{v}_i^2 + \left(\frac{1}{N}\right)^2 \hat{\boldsymbol{\gamma}}' \sum_{i \in S_B} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} \hat{\boldsymbol{\eta}}_i \hat{\boldsymbol{\eta}}_i' \hat{\boldsymbol{\gamma}},$$

where $\hat{v}_i = y_i - \hat{\boldsymbol{\gamma}}' \hat{\boldsymbol{\eta}}_i, \quad \hat{\boldsymbol{\eta}}_i = (1, x_i, y_i) p_i^A(\hat{\boldsymbol{\alpha}}),$

$$\hat{\boldsymbol{\gamma}} = \left[\sum_{i \in S_A} \frac{1 - p_i^A(\hat{\boldsymbol{\alpha}})}{p_i^A(\hat{\boldsymbol{\alpha}})} y_i (1, x_i, y_i) \right] \hat{M}^{-1},$$

and $\hat{M} = \sum_{i \in S_B} d_i^B (1, x_i, y_i)' p_i^A(\hat{\boldsymbol{\alpha}})(1 - p_i^A(\hat{\boldsymbol{\alpha}}))(1, x_i, y_i).$

$$\hat{V}_B = \left(\frac{1}{N}\right)^2 \sum_{i \in S_B} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} y_i^2, \quad \hat{C}_{AB} = \hat{\boldsymbol{\gamma}} \left(\frac{1}{N}\right)^2 \sum_{i \in S_B} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} \hat{\boldsymbol{\eta}}_i y_i.$$

A Hajek Version

Hajek

$$\hat{\theta}^B = \frac{1}{\sum_{i \in S_B} d_i^B} \sum_{i \in S_B} d_i^B y_i.$$

Propensity-Score

$$\hat{\theta}^A = \frac{1}{\sum_{i \in S_A} \frac{1}{p_i^A(\hat{\alpha})}} \sum_{i \in S_A} \frac{1}{p_i^A(\hat{\alpha})} y_i.$$

$$\hat{\mu}^{HJ} = (\mathbf{1}' \hat{\boldsymbol{V}}_{HJ}^{-1} \mathbf{1})^{-1} \mathbf{1}' \hat{\boldsymbol{V}}_{HJ}^{-1} (\hat{\theta}^A, \hat{\theta}^B)'$$

$$\hat{V}_{HJ}(\hat{\mu}^{HJ}) = (\mathbf{1}' \hat{\boldsymbol{V}}_{HJ}^{-1} \mathbf{1})^{-1}$$

$$\hat{\boldsymbol{V}}_{HJ} = \begin{pmatrix} \hat{V}_A & \hat{C}_{AB} \\ \hat{C}_{AB} & \hat{V}_B \end{pmatrix}$$

- Similar Taylor linearizations for Hajek estimator

Nonresponse for Sample B

- Nonresponse is common in probability samples
- Generalize to allow nonresponse in the probability sample

Nonresponse for Sample B: Set-up

- x_i: $i \in S_B$
- y_i: $R_i = 1$
- MAR condition: $y_i \perp R_i \mid x_i$

Propensity Score Model

$$P(R_i = 1 \mid x_i; \boldsymbol{\rho}) = p_i^B(\boldsymbol{\rho}) = \frac{\exp(\rho_0 + \rho_1 x_i)}{1 + \exp(\rho_0 + \rho_1 x_i)},$$

The estimator of $\boldsymbol{\rho} = (\rho_0, \rho_1)'$ satisfies $\boldsymbol{U}_B(\hat{\boldsymbol{\rho}}) = \mathbf{0}$, where

$$\boldsymbol{U}_B(\boldsymbol{\rho}) = \sum_{i \in S_B} d_i^B (R_i - p_i^B(\boldsymbol{\rho}))(1, x_i)'.$$

Modified P-L Estimator

- Extend the pseudo-likelihood estimator of $\boldsymbol{\alpha}$ to account for nonresponse in sample S_B.
- Define the estimator of $\boldsymbol{\alpha}$ to satisfy $\boldsymbol{U}_A(\hat{\boldsymbol{\alpha}}) = \mathbf{0}$, where

$$\boldsymbol{U}_A(\boldsymbol{\alpha}) = \sum_{i \in S_A} (1, x_i, y_i)' - \sum_{i \in S_B} R_i d_i^B \frac{1}{p_i^B(\hat{\boldsymbol{\rho}})} p_i^A(\hat{\boldsymbol{\alpha}})(1, x_i, y_i)' = \mathbf{0}.$$

Two Estimators Under Nonresponse

Estimator for the non-probability sample:

$$\hat{\theta}_*^A = \frac{1}{N} \sum_{i \in S_A} \frac{1}{p_i^A(\hat{\boldsymbol{\alpha}})} y_i.$$

Estimator based on the probability sample:

$$\hat{\theta}_*^B = \frac{1}{N} \sum_{i \in S_B} R_i d_i^B \frac{1}{p_i^B(\hat{\boldsymbol{\rho}})} y_i.$$

EGLS estimator:

$$\begin{aligned} \hat{\mu}^* &= (\mathbf{1}' \hat{\boldsymbol{V}}_*^{-1} \mathbf{1})^{-1} \mathbf{1}' \hat{\boldsymbol{V}}_*^{-1} (\hat{\theta}_*^A, \hat{\theta}_*^B)' \\ \hat{V}_*(\hat{\mu}^*) &= (\mathbf{1}' \hat{\boldsymbol{V}}_*^{-1} \mathbf{1})^{-1} \\ \hat{\boldsymbol{V}}_* &= V((\hat{\theta}_*^A, \hat{\theta}_*^B)) \end{aligned}$$

- $\hat{\boldsymbol{V}}_*$ from Taylor linearization

Taylor Linearization Under Nonresponse

Theorem 2: Under further regularity conditions,

$$\hat{\theta}_*^A = \frac{1}{N}\left[\sum_{i\in S_A} v_i^{*A} + \sum_{i\in S_B} d_i^B \eta_i^{*B}\right] + O_p(n^{-1}),$$

where $v_i^{*A} = y_i - \boldsymbol{\gamma}_{y,\infty}\boldsymbol{M}_{\alpha,\infty}^{-1}\boldsymbol{z}_i p_i^A$ and $\eta_i^{*B} = \boldsymbol{\gamma}_{y,\infty}\boldsymbol{M}_{\alpha,\infty}^{-1}\boldsymbol{\eta}_i^B$. Also,

$$\hat{\theta}_*^B = \frac{1}{N}\sum_{i\in S_B} d_i^B v_i^B + O_p(n^{-1}),$$

where $v_i^B = R_i \frac{1}{p_i^B} y_i - \boldsymbol{\gamma}_{b,\infty}^* \boldsymbol{M}_{bb,\infty}^{-1}\boldsymbol{x}_i(R_i - p_i^B)$.

Estimating Variances

$$\hat{V}_A = \left(\frac{1}{N}\right)^2 \left[\sum_{i \in S_A} \frac{1}{p_i^A(\hat{\boldsymbol{\alpha}})}\left(\frac{1}{p_i^A(\hat{\boldsymbol{\alpha}})} - 1\right)(\hat{v}_i^{*A})^2 + \sum_{i \in S_B}\left(\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}\right)(\hat{\eta}_i^{*B})^2\right],$$

where $\hat{v}_i^{*A} = y_i - \hat{\boldsymbol{\gamma}}_A'\hat{M}_\alpha^{-1}(1, x_i, y_i)' p_i^A(\hat{\boldsymbol{\alpha}})$, $\hat{\eta}_i^{*B} = \hat{\boldsymbol{\gamma}}_A'\hat{M}_\alpha^{-1}\hat{\boldsymbol{\eta}}_i^B$,

$$\hat{\boldsymbol{\gamma}}_A = \sum_{i \in S_A} \frac{1 - p_i^A(\hat{\boldsymbol{\alpha}})}{p_i^A(\hat{\boldsymbol{\alpha}})} y_i (1, x_i, y_i),$$

$$\hat{M}_\alpha = \sum_{i \in S_B} R_i d_i^B \frac{1}{p_i^B(\hat{\boldsymbol{\rho}})} p_i^A(\hat{\boldsymbol{\alpha}})(1 - p_i^A(\hat{\boldsymbol{\alpha}}))(1, x_i, y_i)'(1, x_i, y_i),$$

$$\hat{\boldsymbol{\eta}}_i^B = R_i \frac{1}{p_i^B(\hat{\boldsymbol{\rho}})} p_i^A(\hat{\boldsymbol{\alpha}})(1, x_i, y_i)' - \hat{\boldsymbol{\gamma}}_b'\hat{M}_{bb}^{-1}(1, x_i)'(R_i - p_i^B(\hat{\boldsymbol{\rho}})),$$

$$\hat{\boldsymbol{\gamma}}_b = \sum_{i \in S_B} R_i d_i^B \frac{1 - p_i^B(\hat{\boldsymbol{\rho}})}{p_i^B(\hat{\boldsymbol{\rho}})}(1, x_i, y_i)'(1, x_i),$$

$$\hat{M}_{bb} = \sum d_i^B (1, x_i)' p_i^B(\hat{\boldsymbol{\rho}})(1 - p_i^B(\hat{\boldsymbol{\rho}}))(1, x_i).$$

Estimating Variances

The estimated variance of $\hat{\theta}^B$ is of the form

$$\hat{V}_B = \left(\frac{1}{N}\right)^2 \sum_{i \in S_B} \sum_{j \in S_B} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j}\right) (\hat{v}_i^B)^2,$$

where

$$\hat{v}_i^B = R_i \frac{1}{\hat{p}_i^B} y_i - \hat{\gamma}_b^* \boldsymbol{M}_{bb}^{-1} \boldsymbol{x}_i (R_i - \hat{p}_i^B),$$

$\hat{p}_i^B = p_i^B(\hat{\boldsymbol{\rho}})$, and

$$\hat{\gamma}_b^* = N^{-1} \sum_{i \in S_B} d_i^B (\hat{p}_i^B)^{-1} (1 - \hat{p}_i^B) y_i (1, x_i).$$

Finally, the estimated covariance is of the form

$$\hat{C}_{AB} = \left(\frac{1}{N}\right)^2 \sum_{i \in S_B} \sum_{j \in S_B} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j}\right) (\hat{v}_i^B)(\hat{\eta}_i^{*B})$$

Simulation studies: Set-Up

Finite Population

Generate

$$y_i = 1 + x_i + x_i^2 + e_i, \;\; i = 1, 2, \ldots, N = 10000,$$

where $x_i \overset{iid}{\sim} Uniform(-2, 2)$, and $e_i \overset{iid}{\sim} N(0, 2)$.

S_A Sample

Select the non-probability sample S_A as a Poisson sample, where

$$logit(p_i^A) = \boldsymbol{z}_i'\boldsymbol{\alpha} = \alpha_0 + \alpha_x x_i + \alpha_y y_i.$$

S_B Sample

Select the probability sample S_B as a Poisson sample, where

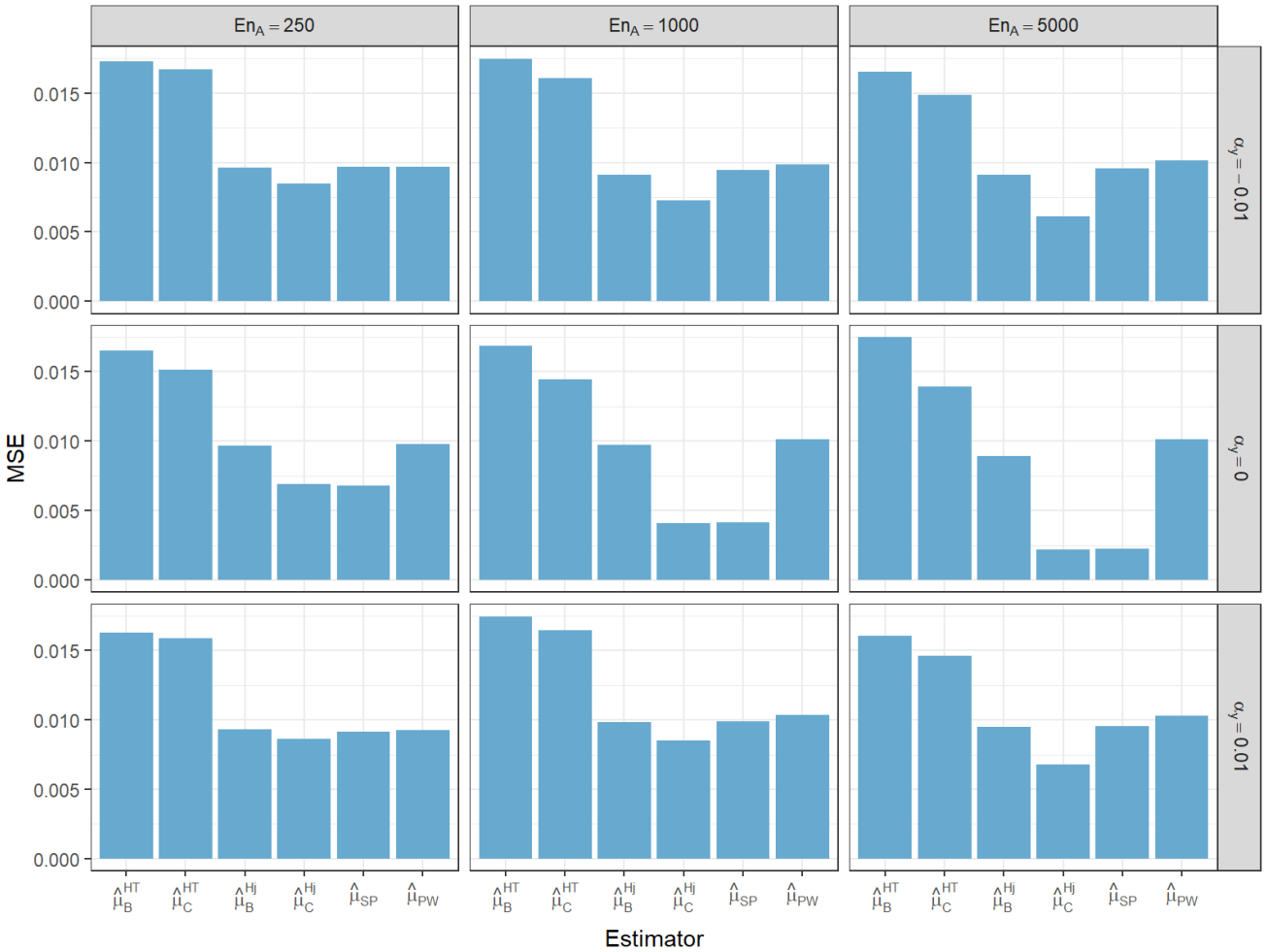
$$\pi_i = \frac{E(n_B)c_i}{\sum_{j=1}^{N} c_j}, \quad c_i = e^{|x_i|} + 1$$

Simulation studies: no nonresponse

- Evaluate the proposed estimators $\hat{\mu}_C$ of HT and Hajek type under conditions where the ignorability assumption is either met or violated.
- Alternative methods:
  - Proposed: $\hat{\mu}_C$ (HT and Hajek)
  - Prob. sample only: $\hat{\mu}_B$ (HT and Hajek)
  - Semi-parametric estimator: $\hat{\mu}_{SP}$ (Zhu et al. 2023)
  - Pseudo-weight estimator: $\hat{\mu}_{PW}$ (Elliott & Valliant 2017)
- Select the target non-probability sample size as $E(n_A) = 250, 1000, 5000$, fix probability sample size as $E(n_B) = 500$, and adjust $\alpha_y = 0, \pm 0.01$.
- MC sample size 2000

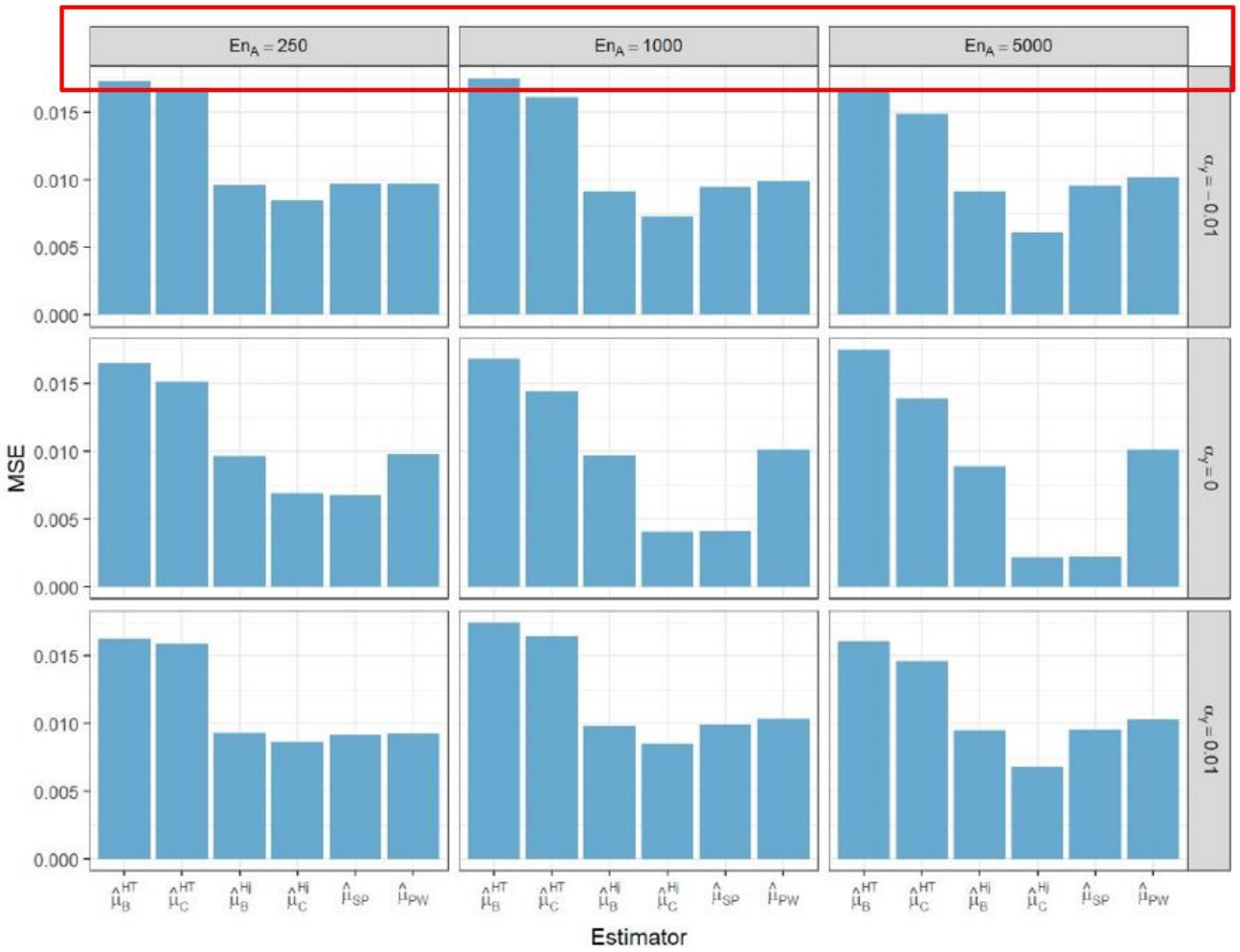
Simulation studies: no nonresponse

MC MSE of alternative estimates of μ_y



Simulation studies: no nonresponse

MC MSE of alternative estimates of μ_y

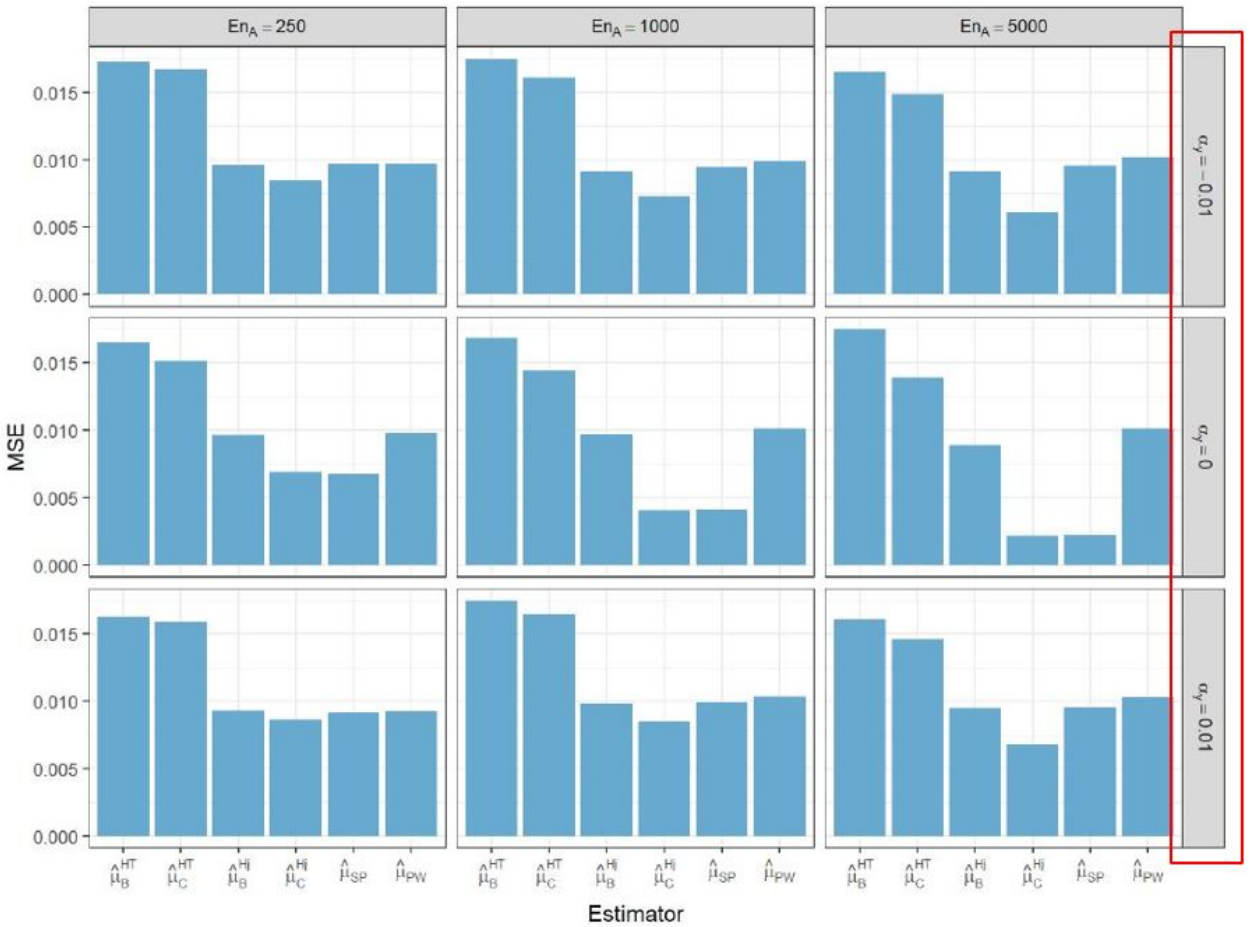


Simulation studies: no nonresponse

MC MSE of alternative estimates of μ_y

Simulation studies: no nonresponse

MC MSE of alternative estimates of μ_y



Simulation studies: nonresponse in the probability sample

- Incorporate ignorable missingness of y_i in S_B by applying $p_i^B(\rho) = logit^{-1}(\rho_0 + \rho_x x_i)$, for $\rho_x = 0, \pm 0.2$ and adjusting ρ_0 to achieve missing rates at 20% and 40%.
- Calculate the MSE, the MC variance and the MC mean of the estimated variance as

$$MSE = \frac{1}{M}\sum_{m=1}^{M}(\hat{\mu}^{(m)} - \mu)^2$$

$$V_{MC} = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\mu}^{(m)} - M^{-1}\sum_{m=1}^{M}\hat{\mu}^{(m)})^2,$$

$$\bar{V} = \frac{1}{M}\sum_{m=1}^{M}\hat{V}(\hat{\mu}^{(m)}).$$

Simulation studies: nonresponse

		$\hat{\theta}_A^*$			$\hat{\theta}_B^*$			$\hat{\mu}_C^*$		
ρ_x	%missing	MSE	V_{MC}	$\bar{V}$	MSE	V_{MC}	$\bar{V}$	MSE	V_{MC}	$\bar{V}$
0	0	1.679	1.680	1.661	1.692	1.693	1.662	1.571	1.570	1.544
0	20	1.997	1.993	1.858	1.998	1.996	1.857	1.851	1.837	1.684
0	40	2.326	2.326	2.184	2.263	2.264	2.159	2.014	2.006	1.894
-0.2	20	1.986	1.986	1.870	1.979	1.980	1.858	1.873	1.868	1.691
-0.2	40	2.258	2.255	2.244	2.212	2.211	2.215	2.039	2.026	1.937
0.2	20	1.821	1.821	1.844	1.830	1.831	1.850	1.674	1.667	1.677
0.2	40	2.221	2.222	2.147	2.215	2.215	2.142	2.012	2.010	1.872

Discussion

- GLS estimator to combine probability and nonprobability samples when Y observed in both
  - Straightforward to calculate
  - Allows for nonresponse in the probability sample
  - Simulations support the proposed procedures

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J. & Tourangeau, R. (2013), 'Summary report of the aapor task force on non-probability sampling', *Journal of survey statistics and methodology* **1**(2), 90–143.

Beaumont, J.-F. (2020), 'Are probability surveys bound to disappear for the production of official statistics?', *Survey Methodology* **46**(1), 1–29.

Chen, Y., Li, P. & Wu, C. (2020), 'Doubly robust inference with nonprobability survey samples', *Journal of the American Statistical Association* **115**(532), 2011–2021.

Chen, Y., Li, P. & Wu, C. (2023), 'Dealing with undercoverage for non-probability survey samples', *Survey Methodology* **49**(2).

Elliott, M. R. & Valliant, R. (2017), 'Inference for nonprobability samples'.

Kim, J. K., Park, S., Chen, Y. & Wu, C. (2021), 'Combining non-probability and probability survey samples through mass imputation', *Journal of the Royal Statistical Society Series A: Statistics in Society* **184**(3), 941–963.

Medous, E., Goga, C., Ruiz-Gazen, A., Beaumont, J.-F., Dessertaine, A. & Puech, P. (2022), 'QR prediction for statistical data integration'.

Savitsky, T. D., Williams, M. R., Gershunskaya, J., Beresovsky, V. & Johnson, N. G. (2022), 'Methods for combining probability and nonprobability samples under unknown overlaps', *arXiv preprint arXiv:2208.14541* .

Valliant, R. (2020), 'Comparing alternatives for estimation from nonprobability samples', *Journal of Survey Statistics and Methodology* **8**(2), 231–263.

Valliant, R. & Dever, J. A. (2011), 'Estimating propensity adjustments for volunteer web surveys', *Sociological Methods & Research* **40**(1), 105–137.

Wu, C. (2022), 'Statistical inference with non-probability survey samples', *Survey Methodology* **48**(2), 283–311.

Zhu, T., Gamble, L. J., Klapman, M., Xue, L. & Lesser, V. M. (2023), 'Using auxiliary information in probability survey data to improve pseudo-weighting in nonprobability samples: A copula model approach', *Journal of Survey Statistics and Methodology* p. smad032.

Thank You!