Appendix: Improving precision through design and analysis in experiments with noncompliance

A-1 Placebo-Controlled Designs

Placebo-controlled trials provide an alternative approach for improving precision in experiments with noncompliance. Rather than a true control, in which there is no intervention, researchers administer a "placebo" encouragement, allowing compliance status to be measured in the placebo group. The researcher must assume that (1) the encouragement and placebo induce the same compliance and attrition patterns for individuals and (2) the placebo does not affect the outcome of interest. Researchers estimate the CACE using a difference-in-means among compliers in each group, leading to significant efficiency gains relative to instrumental variables (Nickerson, 2005) when compliance is less than 50% (Gerber and Green, 2012). For example, in the political rally example above, a placebo flyer could encourage attendance to a non-political rally, which allows the researcher to measure compliance (attendance) but is unlikely to affect voter turnout. The placebo-controlled design can be combined with a full control to improve precision and evaluate robustness to the necessary assumptions (Gerber, Green, et al., 2010). While placebo-controlled experiments effectively address noncompliance in the design stage, they are not always feasible and the required assumptions may prove too strong.

A-2 Extensions

A-2.1 Blocked Difference-in-Means with Larger Blocks

In the main text, we propose a blocked difference-in-means estimator with blocks of size 2, within which we can simply randomly assign treatment within the block. To construct our estimator, we drop all the blocks for which the unit assigned to treatment does not comply. However, when blocks are larger than size 2, then there is a chance that some of the units assigned to treatment comply, while others do not.

For blocks of larger than size 2, we propose a weighted blocked difference-in-means estimator, in which the weight of each block is determined by the proportion of complying units in the treatment group. More specifically:

$$\hat{\tau}_B = \sum_{b=1}^{B} \frac{n_C^{(b)}}{n_C} \left(\frac{1}{n_C^{(b)}} \sum_{i=1}^{n_b} Y_{ib} T_{ib} C_{ib} - \frac{1}{n_b} \sum_{i=1}^{n_b} Y_{ib}(1 - T_{ib}) \right),$$

where n_b is the total number of units in block b, n_C is the total number of observed compliers in the experiment, and $n_C^{(b)}$ is the total number of compliers observed in block b. This formulation is equivalent to running a weighted regression, where the weights are computed as the number of observed compliers in the encouragement group divided by the number of units assigned to encouragement. Blocks with zero units who comply are assigned a weight of zero in the regression.

A-2.2 Sensitivity Analysis

In the following subsection, we provide a two parameter sensitivity analysis for violations of block principal ignorability. We illustrate an example of the sensitivity analysis on the Get-Out-the-Vote data.

The bias of the blocked difference-in-means estimator under a violation of the block principal ignorability assumption is:

$$\text{Bias}(\hat{\tau}_B) = \underbrace{P(C_i = 0|Z_i = 0, B_i \in \mathcal{B}_c)}_{(a)} \cdot \Bigg(\underbrace{\mathbb{E}\big(\mathbb{E}(Y_i(0)|C_i = 1, B_i \in \mathcal{B}_c)\big) - \mathbb{E}\big(\mathbb{E}(Y_i(0)|C_i = 0, B_i \in \mathcal{B}_c)\big)}_{(b)} \Bigg). \tag{1}$$

where B_i represents the block that unit i belongs to, and $\mathcal{B}_c$ represents the set of blocks for which the unit assigned to treatment complies. The left term in Equation (1), denoted (a), is the probability that a unit assigned to control in a block where the unit assigned to treatment complies is a non-complier. If this probability is equal to zero, then all of the control units paired with a unit assigned to encouragement who complies are also compliers, and our block difference-in-means estimator of the CACE is unbiased, as:

$$\mathbb{E}(Y_i(0)|Z_i = 0, B_i \in \mathcal{B}_c) = \mathbb{E}(Y_i(0)|Z_i = 0, C_i = 1) = \mathbb{E}(Y_i(0)|C_i = 1),$$

where the last equality follows from the randomization in treatment assignment. The right term in Equation (1), denoted (b), represents the degree to which the block principal ignorability assumption is violated. More specifically, when block principal ignorability holds, then $\mathbb{E}(Y_i(0)|C_i = 1, B_i \in \mathcal{B}_c) = \mathbb{E}(Y_i(0)|C_i = 0, B_i \in \mathcal{B}_c)$, and the bias would subsequently be zero.

This bias decomposition lends itself naturally to a two parameter sensitivity analysis. The first parameter, from Equation (1) (a), is a probability (i.e., $P(C_i = 0|Z_i = 0, B_i \in \mathcal{B}_c)$) and is naturally bound by $[0, 1]$. In practice, if we assume that the worst-case scenario from blocking is that the blocks are uninformative, then $B_i \perp\!\!\!\perp C_i$, and:

$$P(C_i = 0|Z_i = 0, B_i \in \mathcal{B}_c) = P(C_i = 0|Z_i = 0) = 1 - P(C_i = 1)$$

This implies that if the covariates used for blocking are completely unrelated to compliance, this parameter should fall somewhere between $(0, 1-p)$. This is unlikely if the observed covariates $\mathbf{X}$ are associated with compliance among the encouragement group, which can be empirically evaluated. As the proportion of compliers in the experiment increases, assuming outcomes are finite, the bound on the potential bias will be tighter.

The second parameter is the average imbalance that occurs across blocks due to violations of the block principal ignorability assumption. This term can be bound by the range of the potential outcomes under control in the blocks for which we observe a complier:

$$\begin{aligned} &\mathbb{E}\big(\mathbb{E}(Y_i(0)|C_i = 1, B_i \in \mathcal{B}_c)\big) - \mathbb{E}\big(\mathbb{E}(Y_i(0)|C_i = 0, B_i \in \mathcal{B}_c)\big) \\ &\qquad\qquad \leq \max_i \{Y_i(0)|B_i \in \mathcal{B}_c\} - \min_i \{Y_i(0)|B_i \in \mathcal{B}_c\} \end{aligned}$$

The upper bound represents the range of control outcome values, across the complying blocks. The upper bound is reached if all the control outcomes that are compliers in the complying blocks are equal to the maximum value of $Y_i(0)$, while all the control outcomes that are non-compliers in the complying blocks are equal to the minimum value of $Y_i(0)$.

Researchers may consistently estimate the bound by using the observed sample quantities of

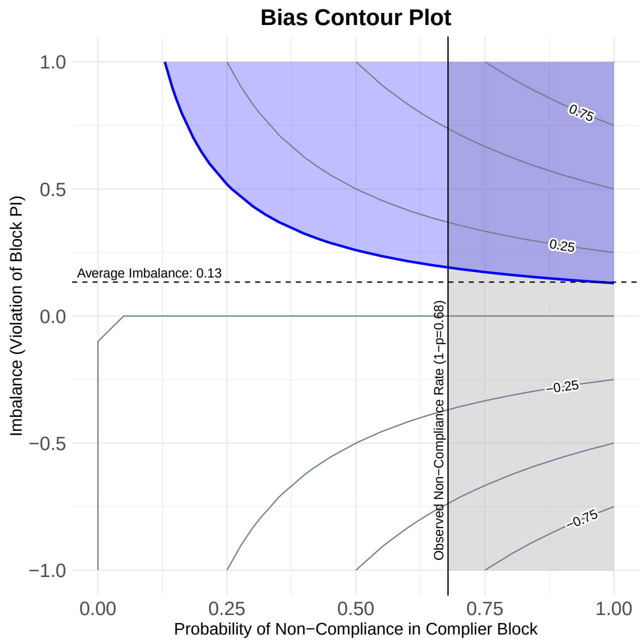

Figure A-2.1: The blue line and shaded blue region denote the killer confounding region, in which a bias of that magnitude would render the point estimate to be zero, or change signs. The shaded gray region denotes the where blocking would have to be worse than random. The dotted line denotes the difference between the average treatment outcome across compliers and non-compliers (i.e., average imbalance from compliance).

the range of the control units across the complying blocks to estimate the upper bound (Ferguson, 2017).

Thus, the bias as a whole can be upper bounded by the following:

$$\widehat{\text{Bias}}(\hat{\tau}_B) \leq (1-p) \cdot \left(\max_i \{Y_i | Z_i = 0, B_i \in \mathcal{B}_c\} - \min_i \{Y_i | Z_i = 0, B_i \in \mathcal{B}_c\} \right). \tag{2}$$

In practice, we expect the upper bound to be relatively conservative. Researchers may be able to tighten the bounds using substantive knowledge.

A-2.2.1 Illustration of Sensitivity Analysis: GOTV

To illustrate the sensitivity analysis, we apply the proposed framework to the Get-Out-the-Vote data, using the matched data for St. Paul. Because the outcome of interest is whether or not an individual voted, the total imbalance (i.e., term (b)) is bounded by $[-1, 1]$. We plot both of these parameters and the resulting bias using a contour plot in Figure A-2.1.

We plot the killer confounder region in Figure A-2.1 to denote the amount of bias that would result in a directional change in the point estimate. We consider the "worst-case" scenario that the blocks are formed at random. The overall observed compliance rate in the experiment is 32%. This implies that the first parameter should be upper bounded by $1 - p = 0.68$, denoted with the gray region. At this upper bound, in order for the bias to fall within the killer confounder region, the imbalance term across the control units would have to exceed 0.19 (represented by where the black line intersects the blue curve).

To try and understand if this level of imbalance is plausible or not, we note the following. If the blocks were truly random, then the imbalance term can be re-written as the average difference across the control units between the compliers and non-compliers. While we cannot measure this quantity, we can estimate the average difference across the encouraged units between the compliers and non-compliers (i.e., $\mathbb{E}(Y_i(1)|Z_i = 1, C_i = 1) - \mathbb{E}(Y_i(1)|Z_i = 1, C_i = 0)$). We find this quantity to be around 0.13, which is represented by the dashed line. Thus, in order for the bias to fall within the killer confounder region, the average imbalance across compliers and non-compliers for the control units would have to be 46% stronger than the average effect of complying across the units assigned to treatment. While this is possible, we argue that this is not likely. This indicates that, even in the worst-case scenario of random blocking, bias significant enough to change the substantive sign of the point estimate is unlikely. While the block principal ignorability assumption may not hold *exactly*, the blocks are likely not at-random either, thereby indicating robustness to the estimated CACE.

A-3 Simulation Studies

In this section, we provide details about the data-generating process, graphically displayed in Figure 1. We consider a scenario where the compliance rate is 10%, on average. The simulation is set up such that $X \in R^{n\times 4}$: $X \sim N(0, \Sigma)$, where:

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.25 & 0 \\ 0.5 & 1 & 0.25 & 0 \\ 0.5 & 0.25 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We define the potential outcomes and compliance status of each unit as:

$$Y = 2T + 15X_2 - 10X_2 \cdot T + X_3 + 2X_3 \cdot T + \varepsilon$$

$$C \sim \text{Binomial}(\Phi(0.25X_1 - 1.28)),$$

where $\varepsilon \sim N(0, 0.35)$ and T is an indicator denoting treatment receipt. Therefore, the covariates X_2 and X_3 affect the outcome Y, while X_1 determines compliance (see Figure 1). However, X_1 is correlated with both X_2 and X_3. As such, blocking on X_1 also accounts for some variation in the compliance, while blocking on either X_2 or X_3 can help account for variation in the outcome of the units. We note that this is a super-population-based simulation, however the performance of the estimators should be similar to a finite-sample set-up.

Table A-3.2 presents bias and MSE results for all of the simulation scenarios, including the placebo-controlled designs which are not included in Table 1 included in the main manuscript.

We also evaluate the performance of the variance estimators. Table A-3.1 shows that the 95% asymptotic confidence intervals have at least nominal coverage in Scenarios 1 and 2, where all of the estimators, except for the as-treated estimator (excluded), are unbiased. Note that our Block DiM estimators are conservative, as they rely on a Neyman-style estimator. However, in Table A-3.1, we see that the standard errors, despite being conservative, are significantly smaller than those for IV or the placebo-controlled design.

95% Interval Coverage Rates and Standard Error

	IV		IV (Block)		Block DiM		PSW		PSW (Block)		Placebo		Placebo (Block)	
	$\widehat{SE}$	Cov.	$\widehat{SE}$	Cov.	$\widehat{SE}$	Cov.	$\widehat{SE}$	Cov.	$\widehat{SE}$	Cov.	$\widehat{SE}$	Cov.	$\widehat{SE}$	Cov.
Scenario 1: Block on Compliance-Related Variables (X_1)														
1,000	9.67	0.95	8.39	0.95	3.33	0.98	3.61	0.98	3.43	0.99	4.10	0.95	3.95	0.96
2,000	6.78	0.95	5.89	0.95	2.36	0.97	2.55	0.99	2.43	0.99	2.89	0.95	2.79	0.96
5,000	4.26	0.95	3.71	0.96	1.49	0.97	1.61	0.99	1.53	0.99	1.83	0.96	1.76	0.96
10,000	3.00	0.95	2.61	0.96	1.05	0.97	1.14	0.99	1.08	0.99	1.29	0.96	1.25	0.96
Scenario 2: Block on Compliance and Outcome-Related Variables (X_1 & X_2)														
1,000	9.68	0.95	1.71	0.99	1.42	1.00	3.61	1.00	3.43	1.00	4.10	0.95	3.81	0.96
2,000	6.79	0.95	1.17	0.99	1.00	1.00	2.55	1.00	2.43	1.00	2.90	0.96	2.69	0.96
5,000	4.26	0.95	0.72	1.00	0.63	1.00	1.61	1.00	1.53	1.00	1.83	0.96	1.70	0.96
10,000	3.00	0.96	0.51	0.99	0.44	1.00	1.14	1.00	1.08	1.00	1.29	0.96	1.20	0.96

Table A-3.1: Summary of the coverage rates and average estimated standard errors. We see that the estimators all have at least nominal coverage.

Simulation Results

	MSE								Bias							
Sample Size	As Treated	IV	IV (Block)	Block DiM	PSW	PSW (Block)	Placebo Controlled	Placebo Controlled (Block)	As Treated	IV	IV (Block)	Block DIM	PSW	PSW (Block)	Placebo Controlled	Placebo Controlled (Block)
Scenario 1: Block on Compliance-Related Variables																
1,000	21.68	94.33	67.99	8.62	8.08	6.17	16.18	14.72	3.53	-0.13	0.10	0.00	-0.02	-0.01	-0.02	-0.02
2,000	16.75	44.37	33.68	4.20	3.81	3.08	8.06	7.29	3.51	-0.10	0.04	-0.00	-0.02	0.00	-0.03	0.01
5,000	14.25	18.25	13.39	1.69	1.52	1.20	3.16	2.87	3.54	0.03	0.04	-0.02	0.00	-0.02	0.03	-0.02
10,000	13.39	8.60	6.73	0.86	0.74	0.61	1.52	1.39	3.54	0.03	0.01	0.01	0.01	-0.00	-0.01	-0.02
Scenario 2: Block on Compliance and Outcome-Related Variables																
1,000	21.83	95.01	1.60	0.88	3.23	0.96	16.70	13.58	3.56	-0.02	0.02	-0.00	-0.00	-0.02	-0.03	-0.03
2,000	17.26	45.38	0.71	0.42	1.52	0.47	7.73	6.73	3.60	0.13	0.03	0.01	0.04	-0.00	0.08	0.02
5,000	14.33	16.88	0.25	0.16	0.58	0.18	3.07	2.61	3.55	0.12	-0.01	-0.01	0.02	-0.01	0.02	-0.03
10,000	13.46	8.66	0.12	0.08	0.29	0.09	1.48	1.34	3.55	-0.02	0.00	-0.00	0.01	-0.01	0.03	-0.01
Scenario 3: Block on Outcome-Related Variables																
1,000	21.01	93.42	1.27	0.83	2.79	0.95	15.31	13.33	3.53	-0.13	-0.00	0.15	0.17	0.14	0.05	-0.04
2,000	16.73	45.74	0.63	0.44	1.40	0.49	7.71	6.84	3.52	0.07	0.01	0.16	0.17	0.16	0.01	0.02
5,000	14.17	17.86	0.25	0.19	0.57	0.21	3.07	2.70	3.53	-0.03	-0.00	0.16	0.15	0.15	-0.03	-0.00
10,000	13.26	8.92	0.12	0.11	0.31	0.12	1.57	1.32	3.52	0.02	0.00	0.16	0.16	0.16	0.01	0.01

Table A-3.2: Summary of simulation results. The compliance rate is set at 10%. The simulation is run across varying sample sizes and varying blocking variables, for 5,000 total iterations for each sample size and blocking scenario. The experimental design is denoted in parentheses, with "CR" denoting complete randomization and "Block" denoting block-randomized.

A-3.1 Violating the Exclusion Restriction

To examine what happens under a violation of the exclusion restriction, we modify the equation of our potential outcomes in our simulation to include a direct effect from the the assignment variable Z:

$$Y = 2T + 15X_2 - 10X_2 \cdot T + X_3 + 2X_3 \cdot T + \gamma Z + \varepsilon,$$

where γ represents the degree to which the exclusion restriction is violated. We compare the performance of the different estimators when this occurs, varying γ in increments of 0.1 from 0 to 1.

Generally, when the exclusion restriction is violated, IV estimation incurs a large degree of bias. As expected, we see that the blocked difference-in-means and PSW estimators do not exhibit bias, as neither of these rely on the exclusion restriction assumption. The violation does not affect the efficiency much in this context, so increases in the MSE of the IV estimators are primarily driven by bias. Figure A-3.2 visualizes the results for the estimators' performance under exclusion restriction violations.

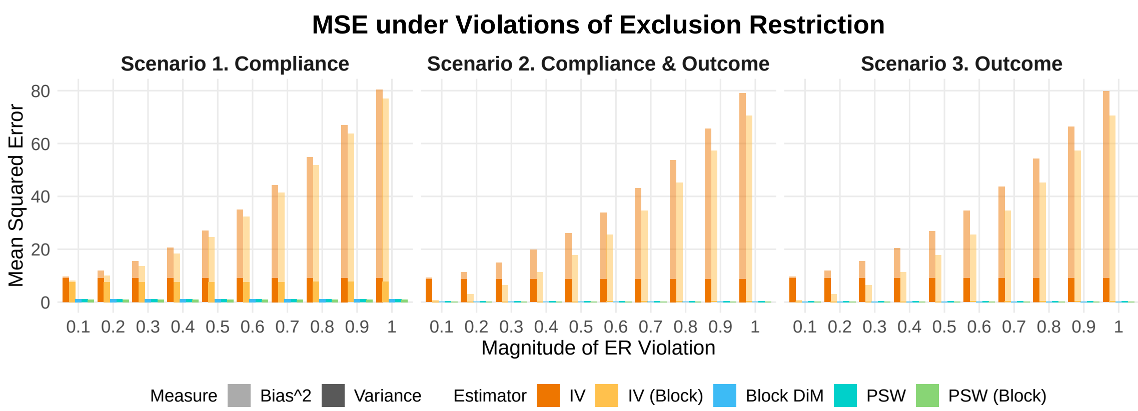

Figure A-3.2: Estimator Performance under Exclusion Restriction Violation.

A-3.2 Violating Principal Ignorability

In our original simulations we show that since X_1 and X_2 are correlated, blocking on just X_2 (i.e., covariates related to outcome) significantly reduced bias, despite not fully meeting principal ignorability. Because the performance of the principal ignorability approaches largely depends on the choice of covariates used for blocking or analysis, a natural question is how 'good' do those covariates have to be? To investigate this question, which involves varying degrees of violations of principal ignorability, we examine the behavior of the estimators when incorporating variables correlated with those sufficient for (block) principal ignorability. In order to do so, we construct a set of proxy variables that are correlated with X_1 and X_2:

$$X_i' = \rho X_i + \sqrt{1-\rho^2} \cdot \varepsilon,$$

where $i \in \{1, 2\}$, $\varepsilon \sim N(0, 1)$, and ρ is the level of correlation. For the simulation, we vary ρ to be between $[0, 1]$, at increments of 0.1, where when $\rho = 0$, we are effectively controlling for noise, thus significantly violating PI, and when $\rho = 1$, we are blocking or weighting on the same variables in

City	Blocking Set: Restricted			
	IV (Block)	PSW	PSW (Block)	Blocked DiM
Bridgeport	8.8%	46.0%	46.7%	37.7%
Columbus	16.6%	62.6%	63.1%	61.3%
Detroit	20.0%	54.6%	58.5%	53.7%
Minneapolis	24.5%	64.2%	66.7%	66.0%
Raleigh	24.4%	41.3%	48.7%	46.7%
St. Paul	16.7%	53.0%	56.7%	52.1%

Table A-4.3: The percentage reduction in estimated standard error relative to the IV estimator under complete randomization.

Scenario 1, which meets PI. Results are presented in Section 4.1 of the main manuscript.

A-4 GOTV Application: Additional Analyses

In this section, we provide additional details related to our empirical analysis based on Green, Gerber, and Nickerson (2003)'s Get-Out-the-Vote experiment (GOTV).

A-4.1 Variable Selection

In the empirical evaluation in the main manuscript, we used all available covariates in our analysis. In this section, we demonstrate a method for covariate selection when constructing blocks using a variable importance measure. We begin by fitting two different Random Forest models (Breiman, 2001) for compliance and for outcome. The compliance model is fit only using the units assigned to encouragement, and the outcome model is fit only across the control outcomes. Both models used all available pretreatment covariates in Figure A-4.3. We estimate the variable importance measures for each covariate in the model to determine which should be included in the blocking set using a permutation-based importance measure (Janitza, Tutz, and Boulesteix, 2016). We keep the covariates that result in over a 1% increase in prediction error from permuting.

Using this restricted set of variables to perform blocking, we still see substantial gains to precision similar to using the full set. Table A-4.3 presents the gains in precision relative to IV under complete randomization. Table A-4.5 presents the numerical results. As with the full set, in sites where we can better predict compliance, we see more substantial gains in precision. The point estimates are substantively similar to the original analysis as well as the results using the full set of covariates, bolstering the claim that even this restricted set of covariates does a good job of meeting, or nearly meeting, principal ignorability.

A-4.2 Predicting Compliance Scores

One way to bolster the claim of principal ignorability is to evaluate how well compliance can be predicted in the encouragement group. To demonstrate this, we calculate the accuracy of the compliance score model, first estimated using the full set of covariates, and then estimated using the restricted set of covariates described above. Overall, using a Random Forest model, we are able to predict the compliance status of individuals with a 70-90% accuracy rate, depending on experimental site, using a threshold of 0.5 for our predictions. We also calculate the area under the ROC curve (AUC) for a threshold-invariant summary of performance. We find that for all cases,

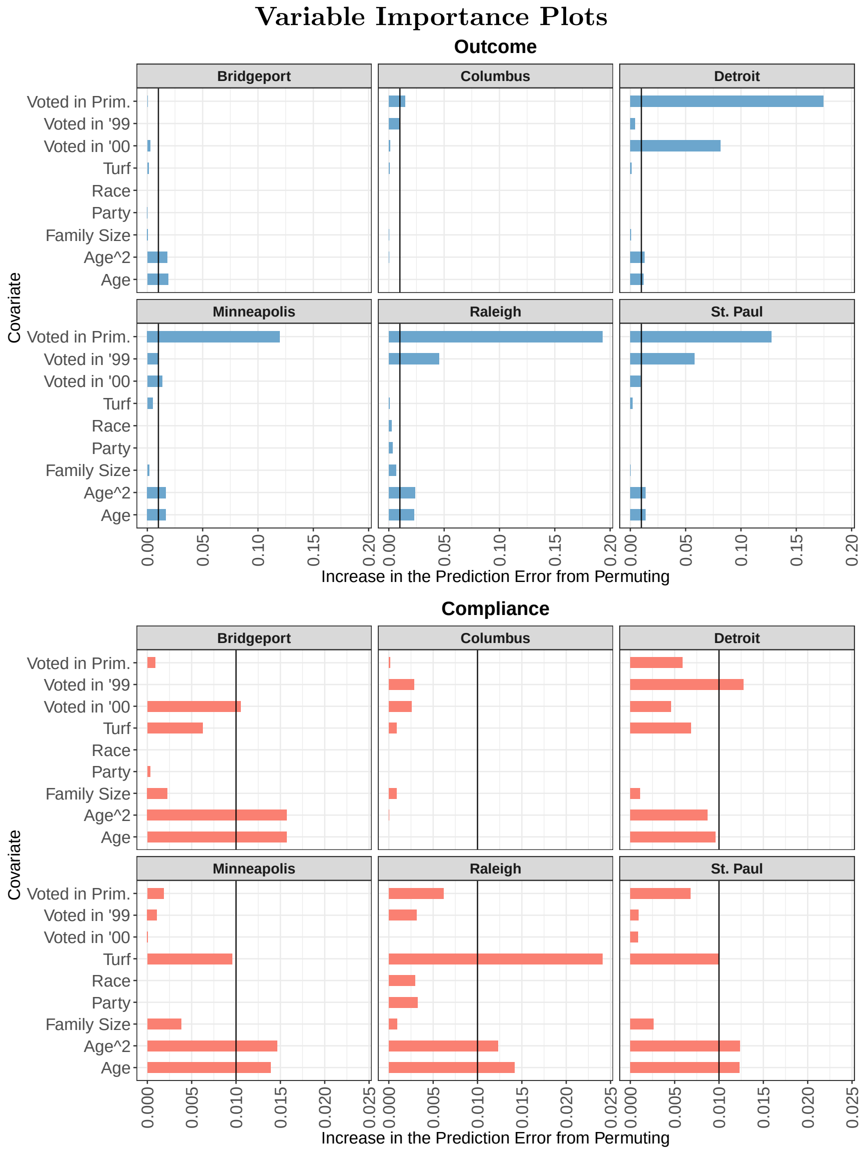

Figure A-4.3: Variable importance plots for a Random Forest model fit to estimate control outcomes, and a Random Forest model fit to estimate compliance. We denote the 1% cutoff used in covariate selection with a black line.

Accuracy of Compliance Score Models

	Blocking Set			
	Full		Restricted	
City	Accuracy	AUC	Accuracy	AUC
Bridgeport	81%	0.97	72%	0.70
Columbus	86%	0.66	86%	0.50
Detroit	72%	0.94	71%	0.53
Minneapolis	83%	0.95	81%	0.60
Raleigh	95%	0.99	61%	0.62
St. Paul	73%	0.96	68%	0.65

Table A-4.4: Accurate of compliance model using prediction accuracy and AUC. Predictions use a 0.5 threshold, which can optimized in practice to maximize fit.

GOTV Point Estimates ($\times$ 100)

Cities	Blocking Variables	IV	IV (Block)	PSW	PSW (Block)	Block DiM
Blocking Set: Full						
Bridgeport	Voted ('00), Age,	14.11	13.31	15.17	14.94	10.89
	Age^2	(5.43)	(5.17)	(2.93)	(2.91)	(3.51)
Raleigh	Primary, Voted ('99),	-0.26	3.94	4.20	6.10	2.62
	Age^2, Age, Turf	(3.44)	(2.41)	(2.03)	(1.66)	(1.68)
Minneapolis	Primary, Voted ('99),	10.15	11.65	8.85	9.14	6.39
	Age^2, Age, Turf	(8.74)	(6.46)	(3.13)	(2.92)	(3.07)
Detroit	Primary, Voted ('99),	8.29	8.95	4.82	5.51	3.82
	Voted ('00), Age	(4.60)	(3.26)	(2.17)	(1.92)	(1.91)
Columbus	Primary, Voted ('99)	6.02	9.64	5.07	5.66	5.42
		(8.34)	(6.81)	(3.08)	(3.00)	(3.24)
St. Paul	Primary, Voted ('99),	14.37	14.37	13.68	13.68	12.96
	Age, Age^2	(6.45)	(5.10)	(3.03)	(2.76)	(3.02)
Blocking Set: Restricted						
Bridgeport	Voted ('00), Age	14.52	14.52	15.28	15.28	12.90
		(5.42)	(4.94)	(2.93)	(2.89)	(3.38)
Raleigh	Primary, Turf	-4.46	4.99	2.30	6.47	6.56
		(3.48)	(2.63)	(2.05)	(1.79)	(1.86)
Minneapolis	Primary, Age^2	10.15	12.03	6.07	6.44	6.39
		(8.74)	(6.60)	(3.13)	(2.91)	(2.97)
Detroit	Primary, Voted ('99)	8.42	8.95	1.85	2.30	2.37
		(4.60)	(3.68)	(2.09)	(1.91)	(2.13)
Columbus	Primary, Voted ('99)	10.24	10.84	5.86	5.76	8.43
		(8.21)	(6.84)	(3.07)	(3.03)	(3.17)
St. Paul	Primary, Age	14.37	14.37	10.39	10.39	14.65
		(6.45)	(5.37)	(3.04)	(2.79)	(3.09)

Table A-4.5: Point estimates and standard errors, in parentheses, for the CACE for IV, PSW, and blocked difference-in-means estimators, under both complete randomization and blocking.

the AUC is at least 0.5. In cases we are using the full set of blocking variables, the AUC for most of the sites is high (i.e., 0.94-0.99), indicating that we are able to build a highly predictive compliance model. For this reason, we focus on the full set of covariates in our main application.

A-4.3 Numerical Results

Figure 4 in the main manuscript presents the results of our re-analysis of the personal canvassing effects from Green, Gerber, and Nickerson (2003). Table A-4.5 presents the numeric results, as well as the results for IV and PSW under complete randomization. We note that point estimates are very similar across estimators. This robustness provides credibility to the PI assumption. In Raleigh, the PSW and block-DIM estimates are similar to the blocked IV estimate, but all are substantively larger than the IV estimate under complete randomization. The difference is driven primarily by the block-randomization, indicating it is more likely related to imbalances in the original data than sensitivity to the identifying assumptions.

A-5 Proofs and Derivations

A-5.1 Proof of Consistency of Block Difference-in-Means under Block Principal Ignorability

Under block principal ignorability, given n blocks, with two units in each block, the following is a consistent estimator for CACE:

$$\hat{\tau}_B = \frac{1}{\sum_{b=1}^{B} C_b} \sum_{b=1}^{B} C_b \hat{\tau}_b, \tag{3}$$

where $\hat{\tau}_b$ is the difference-in-means estimated within the b-th block, and C_b denotes whether or not there exists an observed complier in the b-th block. As such, $\sum_{b=1}^{B} C_b = np$ is effectively the number of blocks that contain a complier, which is equivalent to the number of treated units that comply.

Proof: Define $\mathcal{B}_c$ the set of blocks where the unit assigned to treatment is revealed to be a complier (i.e., $\mathcal{B}_c = \{B_i | C_i = 1, Z_i = 1\}$).

$$\begin{aligned}
&\mathbb{E}(Y_i(1) - Y_i(0) \mid C_i = 1) \\
&\quad = \mathbb{E}(Y_i(1) \mid C_i = 1, Z_i = 1) - \mathbb{E}(Y_i(0) \mid C_i = 1, Z_i = 1) \\
&\quad = \mathbb{E}(Y_i(1) \mid C_i = 1, Z_i = 1) - \mathbb{E}\big(\mathbb{E}(Y_i(0) \mid C_i = 1, Z_i = 1, B_i \in \mathcal{B}_c)\big) \\
&\quad = \mathbb{E}(Y_i(1) \mid C_i = 1, Z_i = 1) - \mathbb{E}\big(\mathbb{E}(Y_i(0) \mid Z_i = 1, B_i \in \mathcal{B}_c)\big) \\
&\quad = \mathbb{E}(Y_i(1) \mid C_i = 1, Z_i = 1) - \mathbb{E}\big(\mathbb{E}(Y_i(0) \mid Z_i = 0, B_i \in \mathcal{B}_c)\big)
\end{aligned} \tag{4}$$

The first line follows from the definition of the complier average causal effect; the second line from ignorability of treatment assignment due to randomization; the third from the law of iterated expectation; fourth from block principal ignorability; and the fifth from ignorability of treatment assignment due to randomization. The last line is made up of observable quantities. We can apply the law of iterated expectations once again to re-write Equation 4 as:

$$\begin{aligned}
&\mathbb{E}(Y_i(1) \mid C_i = 1, Z_i = 1) - \mathbb{E}\big(\mathbb{E}(Y_i(0) \mid Z_i = 0, , B_i \in \mathcal{B}_c)\big) \\
&\qquad = \mathbb{E}\big(\mathbb{E}(Y_i(1) \mid Z_i = 1, B_i \in \mathcal{B}_c)\big) - \mathbb{E}\big(\mathbb{E}(Y_i(0) \mid Z_i = 0, B_i \in \mathcal{B}_c)\big) \\
&\qquad = \mathbb{E}\big(\mathbb{E}(Y_i(1) \mid Z_i = 1, B_i \in \mathcal{B}_c) - \mathbb{E}(Y_i(0) \mid Z_i = 0, B_i \in \mathcal{B}_c)\big)
\end{aligned} \tag{5}$$

The sample analog of Equation 5 will simply be the difference-in-means across each block in $\mathcal{B}_c$, which is what we have defined as $\hat{\tau}_B$:

$$\hat{\tau}_B := \frac{1}{\sum_{b=1}^{B} C_b} \sum_{b=1}^{B} C_b \hat{\tau}_b,$$

where $\hat{\tau}_b$ is the difference-in-means estimated within the b-th block, and C_b denotes whether or not there exists an observed complier in the b-th block.

To show that $\hat{\tau}_B$ is a consistent estimator for CACE, we can take the expectation of both the numerator and the denominator. We begin with the numerator:

$$\begin{aligned}
\mathbb{E}\left(\sum_{b=1}^{B} C_b\hat{\tau}_b\right) &= \sum_{b=1}^{B}\mathbb{E}(C_b\hat{\tau}_b) \\
&= B\cdot\mathbb{E}(C_b\hat{\tau}_b) \\
&= B\cdot\mathbb{E}(C_b\hat{\tau}_b \mid C_b = 1)P(C_b = 1) \\
&= B\cdot\mathbb{E}(\hat{\tau}_b \mid C_b = 1)P(C_i = 1 \mid T_i = 1) \\
&= B\cdot\mathbb{E}(\hat{\tau}_b \mid C_b = 1)P(C_i = 1)
\end{aligned}$$

Now taking the expectation of the denominator:

$$\begin{aligned}
\mathbb{E}\left(\sum_{b=1}^{B} C_b\right) &= \sum_{b=1}^{B}\mathbb{E}(C_b) \\
&= B\cdot\mathbb{E}(C_b) \\
&= B\cdot P(C_i = 1 \mid T_i = 1) \\
&= B\cdot P(C_i = 1)
\end{aligned}$$

As such, applying Weak Law of Large Numbers and the Slutsky's Theorem:

$$\begin{aligned}
\hat{\tau}_B \xrightarrow{p} &\ \frac{\mathbb{E}\left(\sum_{b=1}^{B} C_b\hat{\tau}_b\right)}{\mathbb{E}\left(\sum_{b=1}^{B} C_b\right)} \\
= &\ \frac{B\cdot\mathbb{E}(\hat{\tau}_b \mid C_b = 1)P(C_i = 1)}{B\cdot P(C_i = 1)} \\
= &\ \mathbb{E}(\hat{\tau}_b \mid C_b = 1)
\end{aligned}$$

Because $\mathbb{E}(\hat{\tau}_b \mid C_b = 1)$ is equivalent to the quantity on Equation 5:

$$\equiv \mathbb{E}(Y_i(1) - Y_i(0) \mid C_i = 1)$$

$\square$

A-5.2 Variance of Block Difference-in-Means

Let there be n total blocks (i.e., each block comprises of 2 units). C_b is an indicator for whether or not block b is considered a "complier block". More specifically, $C_b = 1$ if the unit assigned to treatment in the b-th block is a complier. Then, the approximate variance of the blocked difference-in-means estimator, $\hat{\tau}_B$, is:

$$\text{var}(\hat{\tau}_B) \approx \frac{1}{(np)^2}\sum_{b=1}^{B}\text{var}_b(\hat{\tau}_b)P(C_b = 1) + \mathbb{E}_b(\hat{\tau}_b)^2P(C_b = 1)(1 - P(C_b = 1), \tag{6}$$

where the subscript b denotes that the quantity is computed over block b. This can be estimated using linear regression with cluster-robust standard errors. We provide a derivation and show that the cluster-robust standard error estimator is conservative in Technical Appendix Section TA-2.1.

A-5.3 Comparison of Variances

In the simulation studies below, we empirically evaluate the performance of the instrumental variables, principal score weighting, and block difference-in-means estimators, with and without blocking. We show that blocking, particularly on variables related to the outcome, can greatly improve the precision of each of these estimators, regardless of the required identifying assumption. In Technical Appendix Section TA-1 we formalize the efficiency gains from blocking for the instrumental variables and principal score weighting estimators.

In our simulations we also show the the precision gains of our proposed block difference-in-means estimator over instrumental variables with blocking, as well as the principal score weighted estimator, which we formalize in Technical Appendix Section TA-1.

References

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Ferguson, Thomas (2017). *A course in large sample theory*. Routledge.

Gerber, Alan and Donald Green (2012). *Field experiments: Design, analysis, and interpretation.* WW Norton.

Gerber, Alan, Donald Green, et al. (2010). "Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments". In: *Political Analysis* 18.3, pp. 297–315. DOI: 10.1093/pan/mpq008.

Green, Donald, Alan Gerber, and David Nickerson (2003). "Getting out the vote in local elections: Results from six door-to-door canvassing experiments". In: *The Journal of Politics* 65.4, pp. 1083–1096.

Janitza, Silke, Gerhard Tutz, and Anne-Laure Boulesteix (2016). "Random forest for ordinal responses: prediction and variable selection". In: *Computational Statistics & Data Analysis* 96, pp. 57–73.

Nickerson, David (2005). "Scalable Protocols Offer Efficient Design for Field Experiments". In: *Political Analysis* 13.3, pp. 233–252.