Supplemental Materials for "Priming bias versus post-treatment bias in experimental designs"*

Matthew Blackwell† Jacob R. Brown‡ Sophie Hill§
Kosuke Imai¶ Teppei Yamamoto‖

November 26, 2024

Contents

*To be published online.

†Associate Professor, Department of Government, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge MA 02138. Email: mblackwell@gov.harvard.edu, URL: https://mattblackwell.org

‡Assistant Professor, Department of Political Science, Boston University. 232 Bay State Road, Boston, MA 02215. Email: jbrown13@bu.edu, URL: https://jacobrbrown.com

§PhD Student, Department of Government, Harvard University. 1737 Cambridge Street, Cambridge MA 02138. Email: sophie_hill@g.harvard.edu, URL: sophie-e-hill.com

¶Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge MA 02138. Email: imai@harvard.edu URL: https://imai.fas.harvard.edu

‖Professor, Faculty of Political Science and Economics, Waseda University, Tokyo, Japan. Email: tyamam@waseda.jp, URL: http://web.mit.edu/teppei/www

A Proofs

A.1 Pre-test bounds

Proof of Proposition 4. Note that we can write $\tau_{pre}(m) = P_{10m} - P_{00m}$. From expression 7, we can write the true CATE under priming monotonicity as

$$\tau(m) = P_{10m} - P_{00m} - \{\mathbb{P}(Y_i^*(1) = 0, Y_i(1) = 1 \mid M_i^* = m) - \mathbb{P}(Y_i^*(0) = 0, Y_i(0) = 1 \mid M_i^* = m)\}.$$

We can bound the unknown probabilities as

$$0 \leq \mathbb{P}(Y_i^*(t) = 0, Y_i(t) = 1 \mid M_i^* = m) \leq P_{t0m}.$$

Thus, a sharp upper bound on $\tau(m)$ would because

$$P_{10m} - P_{00m} + P_{00m} = P_{10m},$$

and a sharp lower bound would be

$$P_{10m} - P_{00m} - P_{10m} = -P_{00m},$$

which establishes the bounds for $\tau(m)$. For the upper $\delta = \tau(1) - \tau(0)$, we simply use the upper bound for $\tau(1)$ and the lower bound for $\tau(0)$ and vice versa for the lower bound for δ. □

Proof of Proposition 2. Under priming mononoticity, we can write the true CATE as

$$\tau(m) = \tau_{pre}(m) - \{\mathbb{P}(Y_i^*(1) = 0, Y_i(1) = 1 \mid M_i^* = m) - \mathbb{P}(Y_i^*(0) = 0, Y_i(0) = 1 \mid M_i^* = m)\}.$$

The restriction on the amount of priming bias implies that

$$0 \leq \mathbb{P}(Y_i^*(t) = 0, Y_i(t) = 1 \mid M_i^* = m) \leq \theta,$$

for all t. Thus, clearly we have

$$\tau(m) \in \left[\tau_{pre}(m) - \theta, \tau_{pre}(m) + m\right].$$

Taking the maximum and minimum of $\tau(1)$ and $\tau(0)$, respectively, establishes the upper bound for δ. Reversing this gives the lower bound.

□

A.2 Post-test randomization bounds

Proof of Proposition 3. Under Assumption 3, the information about the parameter of interest comes from $P_t = \mathbb{P}(Y_i = 1 \mid T_i = t, Z_i = 1)$ alone. This is because the distribution of the post-test moderators provides no information about the pre-test moderator. Recall that

$$P_t = \pi_{t1}Q_* + \pi_{t0}(1 - Q_*), \tag{1}$$

where $\pi_{tm} = \mathbb{P}[Y_i(t,1) = 1 \mid M_i^* = m]$ and $Q_* = \mathbb{P}[M_i^* = 1]$.

Below, we show how to derive the upper bound for δ. The derivation of the lower bound is similar. Conditional on Q_*, we can define the following linear program:

$$\begin{aligned} &\max \pi_{11} - \pi_{01} - \pi_{10} + \pi_{00} \\ &\text{subject to } \pi_{t1}Q_* + \pi_{t0}(1 - Q_*) = P_t \quad \text{for } t = 0, 1, \\ &0 \leq \pi_{tm} \leq 1 \quad \forall (t, d) \in \{0, 1\}^2 \end{aligned}$$

We can convert this to an augmented form by adding slack variables,

$$\begin{aligned} &\max \pi_{11} - \pi_{01} - \pi_{10} + \pi_{00} \\ &\text{subject to } \pi_{t1}Q_* + \pi_{t0}(1 - Q_*) = P_t \quad \text{for } t = 0, 1 \\ &\pi_{tm} + s_{tm} = 1 \quad \forall (t, m) \in \{0, 1\}^2 \\ &\{\pi_{11}, \pi_{01}, \pi_{10}, \pi_{00}, s_{11}, s_{01}, s_{10}, s_{00}\} \geq 0. \end{aligned}$$

The feasibility of various basic solutions here will depend on the relationship between the observed probabilities and Q_*. In Table SM.1, we show basic feasible solutions for the four different conditions relating P_1 and P_0 to Q_*. Under each condition, it is straightforward to determine that

Table SM.1: Optimal solutions to the linear program under different conditions

Condition	π_{11}	π_{10}	π_{01}	π_{00}	s_{11}	s_{01}	s_{10}	s_{00}	Value
$P_1 > Q_*, P_0 > 1-Q_*$	$\frac{P_1}{Q_*}$	0	0	$\frac{P_0}{1-Q_*}$	$1-\frac{P_1}{Q_*}$	1	1	$1-\frac{P_0}{1-Q_*}$	$\frac{P_1}{Q_*}+\frac{P_0}{1-Q_*}$
$P_1 < Q_*, P_0 > 1-Q_*$	1	$\frac{P_1-Q_*}{1-Q_*}$	0	$\frac{P_0}{1-Q_*}$	0	1	1	$1-\frac{P_0}{1-Q_*}$	$\frac{1-P_1+P_0}{1-Q_*}$
$P_1 > Q_*, P_0 < 1-Q_*$	$\frac{P_1}{Q_*}$	0	$\frac{P_0-1+Q_*}{Q_*}$	1	$1-\frac{P_1}{Q_*}$	1	1	0	$\frac{P_1-P_0+1}{Q_*}$
$P_1 < Q_*, P_0 < 1-Q_*$	1	$\frac{P_1-Q_*}{1-Q_*}$	$\frac{P_0-1+Q_*}{Q_*}$	1	0	1	1	0	$\frac{1-P_1}{1-Q_*}+\frac{1-P_0}{Q_*}$

the basic feasible solution is also optimal since there is no entering variable that can increase the value of the quantity of interest. Thus, we know that

$$\delta \leq \min\left\{\frac{1+P_1-P_0}{Q_*},\ \frac{1-P_1+P_0}{1-Q_*},\ \frac{1-P_0}{Q_*}+\frac{1-P_1}{1-Q_*},\ \frac{P_1}{Q_*}+\frac{P_0}{1-Q_*}\right\}$$

A similar derivation shows that

$$\delta \geq \max\left(-\frac{1-P_1+P_0}{Q_*},\ -\frac{1+P_1-P_0}{1-Q_*},\ -\frac{P_0}{Q_*}-\frac{P_1}{1-Q_*},\ -\frac{1-P_1}{Q_*}-\frac{1-P_0}{1-Q_*}\right).$$

Taking a look at these bounds, suppose that $P_1 \geq 1-P_0$, then the upper bound as a function of Q_* is:

$$U(Q_*) = \begin{cases} \frac{P_0+(1-P_1)}{1-Q_*} & \text{if } Q_* \leq 1-P_0 \\ \frac{1-P_0}{Q_*}+\frac{1-P_1}{1-Q_*} & \text{if } Q_* \in [1-P_0, P_1] \\ \frac{P_1+(1-P_0)}{Q_*} & \text{if } Q_* \geq P_1 \end{cases}$$

Notice that the numerators of all these functions are positive, so the first bounding function $(P_0+1-P_1)/(1-Q_*)$ is montonically increasing over its range and the third, $(P_1+(1-P_0))/Q_*$ is monotonically decreasing over its range. Finally, inspection of the middle bounding function shows that it is convex over its range. This implies that the this function must have its maximum at one of the bound points, $1-P_0$ or P_1. Taking the maximum of these two values, and comparing them to the maximum two values from the situation when $P_1 \leq 1-P_0$ gives the expression of the bounds in the result. A similar approach applies to the lower bounds as well.

Sharpness of these bounds is implied by the linear nature of the optimization function and the convexity of the feasible set. If these bounds were not sharp, this would imply that that there are bounds sharper than these that contain all values of δ consistent with the data and maintained assumptions. But this is clearly contradicted by the fact that the solutions in Table SM.1 are feasible and would fall outside these supposedly sharper bounds. □

A.3 Bounds under additional assumptions

To derive bounds under additional assumptions, we first derive bounds conditional on the strata probabilities.

Lemma SM.1. *The bounds on δ for given values of $\boldsymbol{\rho}$ are $\delta \in [\delta_L(\boldsymbol{\rho}), \delta_U(\boldsymbol{\rho})]$, where*

$$\delta_U(\boldsymbol{\rho}) = P_{111}\left(\frac{Q_{11}}{Q_*}\right) - P_{011}\left(\frac{Q_{01}}{Q_*}\right) - P_{110}\left(\frac{1-Q_{11}}{1-Q_*}\right) + P_{010}\left(\frac{1-Q_{01}}{1-Q_*}\right)$$
$$+ \frac{1}{Q_*(1-Q_*)}\min\left\{\begin{matrix} P_{110}(1-Q_{11}) \\ \rho_{011}+\rho_{001} \\ Q_* - P_{111}Q_{11} \end{matrix}\right\} + \frac{1}{Q_*(1-Q_*)}\min\left\{\begin{matrix} P_{011}Q_{01} \\ \rho_{110}+\rho_{010} \\ 1-Q_*-P_{010}(1-Q_{01}) \end{matrix}\right\},$$

and,

$$\delta_L(\boldsymbol{\rho}) = P_{111}\left(\frac{Q_{11}}{Q_*}\right) - P_{011}\left(\frac{Q_{01}}{Q_*}\right) - P_{110}\left(\frac{1-Q_{11}}{1-Q_*}\right) + P_{010}\left(\frac{1-Q_{01}}{1-Q_*}\right)$$
$$+ \frac{1}{Q_*(1-Q_*)}\max\left\{\begin{matrix} -P_{111}Q_{11} \\ -\rho_{110}-\rho_{100} \\ -1+Q_*+P_{110}(1-Q_{11}) \end{matrix}\right\} + \frac{1}{Q_*(1-Q_*)}\max\left\{\begin{matrix} -P_{011}Q_{01} \\ -\rho_{001}-\rho_{101} \\ -Q_*+P_{011}Q_{01} \end{matrix}\right\}.$$

Proof. Conditional on ρ_s and Q_*, deriving the bounds on δ is a standard linear programming problem. We now describe the process for deriving these bounds at a general level. Without any assumptions, we are interested in maximizing or minimizing the objective function,

$$\begin{aligned}\delta =& P_{111}\left(\frac{Q_{11}}{Q_*}\right) - P_{011}\left(\frac{Q_{01}}{Q_*}\right) - P_{110}\left(\frac{1-Q_{11}}{1-Q_*}\right) + P_{010}\left(\frac{1-Q_{01}}{1-Q_*}\right) \\ &+ \mu_{011}(1,1)\frac{\rho_{011}}{Q_*(1-Q_*)} + \mu_{001}(1,1)\frac{\rho_{001}}{Q_*(1-Q_*)} \\ &- \mu_{110}(1,1)\frac{\rho_{110}}{Q_*(1-Q_*)} - \mu_{100}(1,1)\frac{\rho_{100}}{Q_*(1-Q_*)} \\ &+ \mu_{110}(0,1)\frac{\rho_{110}}{Q_*(1-Q_*)} + \mu_{010}(0,1)\frac{\rho_{010}}{Q_*(1-Q_*)} \\ &- \mu_{101}(0,1)\frac{\rho_{101}}{Q_*(1-Q_*)} - \mu_{001}(0,1)\frac{\rho_{001}}{Q_*(1-Q_*)},\end{aligned}$$

subject to the constraints

$$\begin{aligned} P_{111}Q_{11} &= \mu_{111}(1,1)\rho_{111} + \mu_{101}(1,1)\rho_{101} + \mu_{110}(1,1)\rho_{110} + \mu_{100}(1,1)\rho_{100} \\ P_{011}Q_{01} &= \mu_{111}(0,1)\rho_{111} + \mu_{011}(0,1)\rho_{011} + \mu_{110}(0,1)\rho_{110} + \mu_{010}(0,1)\rho_{010} \\ P_{110}(1-Q_{11}) &= \mu_{011}(1,1)\rho_{011} + \mu_{110}(1,1)\rho_{110} + \mu_{010}(1,1)\rho_{010} + \mu_{000}(1,1)\rho_{000} \\ P_{010}(1-Q_{01}) &= \mu_{101}(0,1)\rho_{101} + \mu_{001}(0,1)\rho_{001} + \mu_{100}(0,1)\rho_{100} + \mu_{000}(0,1)\rho_{000} \\ 0 \le \mu_s(t) &\le 1, \quad \forall s, t. \end{aligned}$$

For this step, we do not need to specify constraints on ρ_s because we consider them fixed (and Q_* is a linear function of ρ_s). The simplex tableau method yields the given bounds. □

Proof of Proposition 4. Recall the constraints on the strata probabilities:

$$Q_{11} = \rho_{111} + \rho_{101} + \rho_{110} + \rho_{100}$$

$$Q_{01} = \rho_{111} + \rho_{011} + \rho_{110} + \rho_{010}$$

$$Q_* = \rho_{111} + \rho_{011} + \rho_{101} + \rho_{001},$$

Under monotonicity, we only have strata $S_i \in \{111, 110, 010, 100, 000\}$, so we have $Q_* = \mathbb{P}[M_i^* = 1] = \rho_{111}$ and $\rho_{110} + \rho_{010} = Q_{01} - Q_*$ and $\rho_{110} + \rho_{100} = Q_{11} - Q_*$. Plugging these values into the bounds from Lemma SM.1, we obtain

$$\delta_U(Q_*) = P_{111}\left(\frac{Q_{11}}{Q_*}\right) - P_{011}\left(\frac{Q_{01}}{Q_*}\right) - P_{110}\left(\frac{1-Q_{11}}{1-Q_*}\right) + P_{010}\left(\frac{1-Q_{01}}{1-Q_*}\right)$$
$$+ \frac{1}{Q_*(1-Q_*)} \min \left\{ \begin{matrix} P_{110}(1-Q_{11}) \\ 0 \\ Q_* - P_{111}Q_{11} \end{matrix} \right\} + \frac{1}{Q_*(1-Q_*)} \min \left\{ \begin{matrix} P_{011}Q_{01} \\ Q_{01} - Q_* \\ 1 - Q_* - P_{010}(1-Q_{01}) \end{matrix} \right\},$$

and,

$$\delta_L(Q_*) = P_{111}\left(\frac{Q_{11}}{Q_*}\right) - P_{011}\left(\frac{Q_{01}}{Q_*}\right) - P_{110}\left(\frac{1-Q_{11}}{1-Q_*}\right) + P_{010}\left(\frac{1-Q_{01}}{1-Q_*}\right)$$
$$+ \frac{1}{Q_*(1-Q_*)} \max \left\{ \begin{matrix} -P_{111}Q_{11} \\ -Q_{11} - Q_* \\ -1 + Q_* + P_{110}(1-Q_{11}) \end{matrix} \right\} + \frac{1}{Q_*(1-Q_*)} \max \left\{ \begin{matrix} -P_{011}Q_{01} \\ 0 \\ -Q_* + P_{011}Q_{01} \end{matrix} \right\}.$$

We further simplify the upper bound expression by noting that $P_{110}(1-Q_{11}) \geq 0$ and $Q_{01} - Q_* \leq 1 - Q_* - P_{010}(1-Q_{01})$. The lower bound simplifies because $Q_{11} - Q_* \leq 1 - Q_* - P_{110}(1-Q_{11})$ and $P_{011}Q_{01} \geq 0$. Removing these extraenous conditions gives the result in the text.

□

Proof of Proposition 5. Under the maintained assumptions, $Q_* = Q_{01}$, which we plug into the expression of Proposition 4. Then, the result is immediate upon noting that $P_{011}Q_{01} - Q_{01} \leq 0$, $P_{011}Q_{01} \geq 0$ and rearranging terms. □

For calculating the bounds under the sensitivity constraints, we can take the bounds from Lemma SM.1 and solve a corresponding linear programming problem to optimize them with respect to the principal strata probabilities. For example, depending the observed data, the upper bound will depend on

$\rho_{011}+\rho_{001}$, $\rho_{110}+\rho_{010}$, or $\rho_{011}+\rho_{001}+\rho_{110}+\rho_{010}$. To find the upper bound across values of $\boldsymbol{\rho}$, we apply the linear programming machinery to finding the upper bound for each of these quantities subject to the constraints that

$$\begin{aligned} Q_{11} &= \rho_{111}+\rho_{101}+\rho_{110}+\rho_{100} \\ Q_{01} &= \rho_{111}+\rho_{011}+\rho_{110}+\rho_{010} \\ Q_{*} &= \rho_{111}+\rho_{011}+\rho_{101}+\rho_{001}, \end{aligned}$$

where $0 \leq \rho_s \leq 1$ for all s and $\sum_{s\in\mathcal{S}} \rho_s = 1$. Note that for the sensitivity analysis, we may impose additional constraints on ρ_s in this step. As an example, for the objection function of $\rho_{011}+\rho_{001}$, we have the upper bound

$$\min\left\{1-Q_{11}, Q_{*}, 1-Q_{01}+Q_{*}, \frac{1}{2}(1-Q_{11}+Q_{*}), 1+Q_{01}-Q_{11}\right\}.$$

Plugging these bounds into the upper bound $\delta_U(\boldsymbol{\rho})$ will yield an upper bound purely as a function of observed parameters and Q_* and γ (the sensitivity parameter). Under some of our assumptions, inspection of the resulting functions reveals that the maximum of these functions can only occur at a handful of critical values of Q_* which can be evaluated and compared quickly. Otherwise, we use a standard optimization routine to find the value of Q_* that maximizes the upper bound or minimizes the lower bound.

B Estimation and Inference Details

The discussion of the bounds in the main text focused on population level bounds—that is, we identified the bounds in terms of population quantities such as P_{tzm}. Estimation and inference for the bounds with a sample poses some important difficulties. The most obvious way to estimate these bounds is to plug in sample version of the population quantities and solve the above linear programming problem to obtain bounds. Unfortunately, the asymptotic distribution of estimators based on this plug-in approach do not have the standard asymptotics due to the lack of differentiability of the bounds as a function of the data. Andrews and Han (2009) show that naive bootstrap methods are not valid for these types of estimators due to these issues.

Below we mostly focus on the random placement design, where estimation and inference is the most complicated. In the pre-test design, the bounds are simple functions of the parameters of interest, so we can use standard asymptotic variance estimators, combined with the approach of Imbens and Manski (2004) to obtain confidence intervals for the parameters of interest. For the post-test bounds, we obtain standard errors for the bounds based on the nonparametric bootstrap and then use these in the approach of Imbens and Manski (2004). In simulations, we found this to have similar performance to the more complicated *union bound* approach of Ye et al. (2023).

B.1 Estimation and inference for the random placement design

For the random placement design, we do not generally have closed-form expressions for the bounds, and so we can reformulate the estimation and inference problem based on moment conditions that feed into sample versions of a population criterion function (Chernozhukov, Hong and Tamer, 2007). In the general analysis of the random placement design, we define the parameters of our model as

$$\psi_{y_1 y_0, m_1}(t, m_0) = \Pr[Y_i^*(t) = y_1, Y_i(t) = y_0, M_i(t) = m_1 \mid M_i^* = m_0],$$

so that the parameter of interest can be written as

$$\delta = \sum_{y_0 m_1} \psi_{1 y_0 m_1}(1,1) - \psi_{1 y_0 m_1}(0,1) - \psi_{1 y_0 m_1}(1,0) + \psi_{1 y_0 m_1}(0,0),$$

with constraints

$$0 \leq \psi_{y_1 y_0, m_1}(t, m_0) \leq 1$$
$$\sum_{y_1=0}^{1} \sum_{y_0=0}^{1} \sum_{m_1=0}^{1} \psi_{y_1 y_0, m_1}(t, m_0) = 1.$$

Let $W_i = (Y_i, M_i, T_i, Z_i)$ be the observed data vector with possible realized value $w \equiv (w_y, w_m, w_t, w_z)$. Abusing notation, we let $w_1 = (w_y, w_m, w_t, 1)$ and $w_0 = (w_y, w_m, w_t, 0)$. The randomization and consistency assumptions imply the following moment conditions:

$$\begin{aligned}\mathbb{E}[g_{w_1}(W_i,\psi) \mid Z_i = 1] &= \mathbb{P}(Y_i = w_y, M_i = w_m, T_i = w_t \mid Z_i = 1) \\ &\quad - Q_* \left(\sum_{y_0} \psi_{w_y y_0 w_m}(w_t, 1) \right) - (1 - Q_*) \left(\sum_{y_0} \psi_{w_y y_0 w_m}(w_t, 0) \right) \\ \mathbb{E}[g_{w_0}(W_i,\psi) \mid Z_i = 0] &= \mathbb{P}(Y_i = w_y, T_i = w_t \mid M_i = w_m, Z_i = 0) - \left(\sum_{y_1} \sum_{m_1} \psi_{y_1 w_y m_1}(w_t, w_m) \right).\end{aligned}$$

There are $d_1 = 8$ for the post-test data and $d_0 = 8$ restrictions for the pre-test data.

Define $r_d(\psi)$ encode the deterministic restrictions on the ψ values and let $\Psi_d^\dagger = \{\psi \,:\, r_d(\psi) \geq 0\}$ be the values of the underlying parameters that satisfy these restrictions. These restrictions include assumptions like monotonicity that would cause certain ψ values to be set to zero or the sensitivity analysis specifications that limit the size of a group of ψ values. Under these maintained assumptions, we can characterize the identified set as

$$\Psi^\star = \{\psi \in \Psi_d^\dagger \,:\, \mathbb{E}[g_{w_1}(W_i,\psi)] = 0, E[g_{w_0}(W_i,\psi)] = 0 \ \forall w_1, w_0 \in \{0,1\}^3\}.$$

One way to define a distance from the identified set is with a population criterion function. Let the moment conditions be indexed by j such that $j = 1, \ldots, 8$ correspond to $\{g_{w_1}\}$ and $j = 9, \ldots, 16$ correspond to $\{g_{w_0}\}$. Then, the population criterion (loss) function is

$$L(\psi) = \sum_{j=1}^{16} |\mathbb{E}[g_j(W_i,\psi)]| \tag{2}$$

Following Torgovitsky (2019), we use absolute value loss here to ensure that we can leverage linear programming techniques for computational convenience. We can obtain an empirical version of the criterion,

$$L_n(\psi) = \sum_{j=1}^{16} \sqrt{n} |\overline{g}_j(W_i,\psi)|, \tag{3}$$

where, for instance,

$$\begin{aligned}\overline{g}_{w_1}(W_i,\psi) &= \mathbb{P}_n(Y_i = w_y, M_i = w_m, T_i = w_t \mid Z_i = 1) \\ &\quad - \widehat{Q}_* \left(\sum_{y_0} \psi_{w_y y_0 w_m}(w_t, 1) \right) - (1 - \widehat{Q}_*) \left(\sum_{y_0} \psi_{w_y y_0 w_m}(w_t, 0) \right),\end{aligned}$$

and $\mathbb{P}_n$ is the in-sample distribution.

We could proceed with estimating the bounds for a given set of assumptions by searching over the parameter space such that $L_n = 0$, but this is often a fragile approach. In particular, it may be the case that restrictions hold in the population but fail to hold in empirical samples due to sampling variability so that the minimum value of L_n is strictly greater than 0. As an alternative approach, we can first find the minimum value of L_n in the sample and then find extreme values of the parameter under parameter values that are close to that minimum.

We first define the sample minimum of the criterion function under the maintained deterministic restrictions,

$$\overline{L}_n = \inf_{\psi \in \Psi_d^\dagger} L_n(\psi). \tag{4}$$

We can then estimate the upper and lower bounds by finding the minimum and maximum values of δ that come close to this value:

$$\widehat{\delta}_L = \min_{\psi \in \Psi_d^\dagger} \delta(\psi) \text{ s.t. } L_n(\psi) \leq \overline{L}_n(1 + \epsilon_n),$$
$$\widehat{\delta}_U = \max_{\psi \in \Psi_d^\dagger} \delta(\psi) \text{ s.t. } L_n(\psi) \leq \overline{L}_n(1 + \epsilon_n).$$

The tuning parameter ϵ_n controls how close we require the criterion function of the bounds to be to the overall sample minimizer. This approach requires $\epsilon_n \to 0$ as $n \to \infty$. We take $\epsilon_n = 0.25$ in our implementation, which has shown good performance in simulations.

For the confidence intervals, we use the nonparametric bootstrap to obtain standard error estimates of each of the bounds and then apply the (Imbens and Manski, 2004) approach. In simulations, we found this approach to be slightly conservative and other competing methods, such as Chernozhukov, Newey and Santos (2023), to undercover the true parameter slightly, at least in simulations similar to our data example. Thus, we use the nonparametric bootstrap approach for our confidence intervals.

C Parametric Bayesian Approach to Incorporate Covariates

The nonparametric bounds above are sharp in the sense that they leverage all information about the outcome, moderator, treatment, and question order. Researchers, however, often have additional data in the form of covariates that may help reduce the uncertainty of their estimates. Here, we consider a Bayesian parametric model of the principal strata approach to the pre-test, post-test, and random placement designs, building on the work of Mealli and Pacini (2013) (see also Imbens and Rubin, 1997; Hirano et al., 2000). Unlike the nonparametric bounds approach, a Bayesian model allows us to incorporate prior information about the data-generating process in a smooth and flexible manner.[1]

C.1 The Model

Our approach focuses on a data augmentation strategy that models the joint distribution of the outcomes and the principal strata, the latter of which are not directly observable. We allow the distribution of the potential outcomes and principal strata conditional on those strata to further depend on covariates via a binomial and multinomial logistic model, respectively:

$$\mathbb{P}(Y_i = 1 \mid T_i = t, Z_i = z, S_i = s, \boldsymbol{X}_i) = \mu_{is}(t, z) = \text{logit}^{-1}(\alpha_{tz|s} + \boldsymbol{X}_i'\boldsymbol{\beta}),$$
$$\mathbb{P}(S_i = s \mid \boldsymbol{X}_i) = \rho_{is} = \frac{\exp(\boldsymbol{X}_i'\boldsymbol{\psi}_s)}{\sum_{j \in \mathcal{S}} \exp(\boldsymbol{X}_i'\boldsymbol{\psi}_j)},$$

where $\boldsymbol{X}_i$ denotes observed pre-treatment covariates that might be predictive of unit i's outcome and principal strata. Note that the strata probabilities do not depend on T_i and Z_i due to randomization. We gather the parameters as $\boldsymbol{\alpha} = \{\alpha_{tz|s}\}$ and $\boldsymbol{\psi} = \{\boldsymbol{\psi}_s\}$. We can easily incorporate assumptions like monotonicity and stable moderators by simply restricting the space of possible principal strata $\mathcal{S}$.

Our goal is to make inferences about the posterior distribution of these parameters and the

[1] Levis et al. (2023) proposes a way to incorporate covariates on nonparametric bounds when the quantity of interest can be written as an average of covariate-specific quantities. Unfortunately, we cannot write the interaction in this way because it is the difference between two different CATEs that condition on different subsets of the data. One could use their approach on each of the individual CATEs and combine those bounds for the interaction, but the resulting bounds would not be sharp.

ultimate quantities of interest, $\tau(d)$ and δ. There are two ways to represent these quantities under this parametric model, resulting in two different types of posterior distributions. The first is based on *population* inference and derives expressions for $\tau(d)$ and δ purely in terms of the parameters of the model. The second is based on *in-sample* inference and derives expressions for $\tau(d)$ and δ in terms of potential outcomes in a particular sample.

For the population inference approach, we first note that due to consistency and randomization, we have $\mu_{is}(t,z) = \mathbb{P}(Y_i(t,z) = 1 \mid S_i = s, \boldsymbol{X}_i)$. Thus, we can write the values of the quantities of interest for a given unit as,

$$\tau_i(m) = \mathbb{E}\left[Y_i^*(1) - Y_i^*(0) \mid M_i^* = m, X_i\right] \;=\; \sum_{s \in \mathcal{S}_d^*} (\mu_{is}(1,1) - \mu_{is}(0,1))\rho_{is},$$

and $\delta_i = \tau_i(1) - \tau_i(0)$, where we omit the implied dependence on $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\psi})$ and remember that $\mathcal{S}_d^*$ is the set of strata levels such that $M_i^* = m$, for $m \in \{0,1\}$. Using the empirical distribution of covariates, the average of these conditional mean differences and interactions will equal the overall quantities of interest, i.e., $\tau(m) = \sum_{i=1}^n \tau_i(m)/n$ and $\delta = \sum_{i=1}^n \delta_i/n$.

The in-sample versions of the quantities of interest are more straightforward, since they are simply the conditional mean differences and interaction among the units in the sample,

$$\tau_s(m) = \frac{\sum_{i=1}^n \mathbb{I}(M_i^* = m)\left\{Y_i^*(1) - Y_i^*(0)\right\}}{\sum_{i=1}^n \mathbb{I}(M_i^* = m)},$$

and $\delta_s = \tau_s(1) - \tau_s(0)$. Obviously, across repeated samples, we can relate these to the population quantities as $\mathbb{E}[\tau_s(m)] = \tau(m)$ and $\mathbb{E}[\delta_s] = \delta$.

We develop an efficient Markov Chain Monte Carlo (MCMC) algorithm to take draws from the posterior and then calculate these quantities of interest. Our Gibbs sampler can also be simplified and used for inference in the absence of pre-treatment covariates, which can be viewed as a Bayesian alternative to uncertainty estimation for the partially identified parameters discussed in Section 5. We provided details of these algorithms in Supplemental Materials D.

This Bayesian approach has the advantage of easily incorporating covariates, but it does require us to select prior distribution for the model parameters, some of which are unidentified in the frequentist sense. Thus, the identification of these parameters will depend on the prior. To investigate this, we take draws of the prior predictive distribution under different prior structures, which we show in Figure SM.1. All the priors we considered are symmetric, but uniform priors on the model parameters lead to somewhat informative priors on the ultimate quantity of interest. Thus, we rely on more dispersed priors for the simulations and the application. We discuss the choice of prior distribution more fully in Supplemental Materials D.

We conduct two simulation studies to demonstrate the gains in efficiency from the monotonicity and stability assumptions and the incorporation of covariates in the Bayesian approach. The first simulation varies the assumptions of the data-generating process and compares the posterior variance of these distributions across combinations of our two assumptions. The second simulation varies the predictive power of the covariates on the outcome and the strata in the data-generating process and compares the variance of the posterior distributions from Gibbs run on each simulated data set with and without incorporating covariates. We present these results in Supplemental Materials D.1.

D MCMC Algorithm

In this section we describe our MCMC algorithm for the Bayesian model of Section C. Our goal is to sample from the joint distribution of the parameters and the principal strata indicator,

$$\mathbb{P}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{S} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{Z}, \boldsymbol{M}) \propto \prod_{i=1}^{n} \left(\sum_{s \in \mathcal{S}_i} \left[\mathbb{P}(Y_i \mid T_i, Z_i, S_i = s, \boldsymbol{X}_i) \mathbb{P}(S_i = s \mid \boldsymbol{X}_i) \right]^{\mathbb{I}(S_i = s)} \right) \mathbb{P}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\psi}),$$

where $\mathcal{S}_i = \mathcal{S}(T_i, Z_i, M_i)$ are the set of principal strata to which unit i could possibly belong. When the set of observed pre-treatment covariates ($\boldsymbol{X}_i$) is empty, the parameter space reduces to that of a standard finite mixture model, and sampling from the joint posterior is straightforward. With $\boldsymbol{X}_i$, Bayesian inference for the model is more complicated. Traditionally, Bayesian inference for logistic regression models has been challenging due to a lack of a simple Gibbs sampling algorithm. Recently, however, Polson, Scott and Windle (2013) introduced a simple data-augmentation strategy based on the Pólya-Gamma (PG) distribution, obviating the need for approximate methods or precise tuning of a Metropolis-Hastings algorithm. We use this approach for both the binary and multinomial logistic regression models for the outcome and principal strata, respectively. This allows a simple Gibbs structure where the full conditional posterior distributions of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\boldsymbol{\psi}$ are Normal conditional on specific draws from the PG distribution.

Conditional on the other parameters, then the full conditional posterior of the principal strata follows a similar form to Hirano et al. (2000),

$$\mathbb{P}(S_i = s \mid Y_i, \mathbf{X}_i, T_i, Z_i, M_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\psi}) = \frac{\mu_{is}(T_i, Z_i)^{Y_i}(1 - \mu_{is}(T_i, Z_i))^{1-Y_i}\rho_{is}}{\sum_{k \in \mathcal{S}_i} \mu_{ik}(T_i, Z_i)^{Y_i}(1 - \mu_{ik}(T_i, Z_i))^{1-Y_i}\rho_{ik}},$$

where we suppress the dependence of μ_{is} and ρ_{is} on the model parameters. Repeatedly drawing from these full conditional posterior distributions should provide a sample from the above joint posterior and allow for posterior inference in the usual manner. In each iteration, $r \in \{1, \dots, R\}$, of the algorithm, we have draws

$$\left(\left\{ \widehat{S}_i^{(r)} \right\}_{i=1}^{n}, \widehat{\boldsymbol{\psi}}^{(r)}, \widehat{\boldsymbol{\alpha}}^{(r)}, \widehat{\boldsymbol{\beta}}^{(r)} \right).$$

We can use these draws to generate draws of the population and in-sample versions of the quantity of interest. Given that $\widehat{S}_i^{(r)}$ is the imputed principal strata imputed for unit i in the rth draw from the posterior, we let

$$\widehat{\mu}_i^{(r)}(t, z) = \widehat{\mu}_{i, \widehat{S}_i^{(r)}}(t, z, \widehat{\boldsymbol{\alpha}}^{(r)}, \widehat{\boldsymbol{\beta}}^{(r)})$$

be the mean of the potential outcomes conditional on that imputed principal strata. Furthermore, let $\widehat{\rho}_{is}^{(r)}$ be the rth draw of the predicted probabilities of each principal strata for each unit. Then, we can calculate the population quantity as

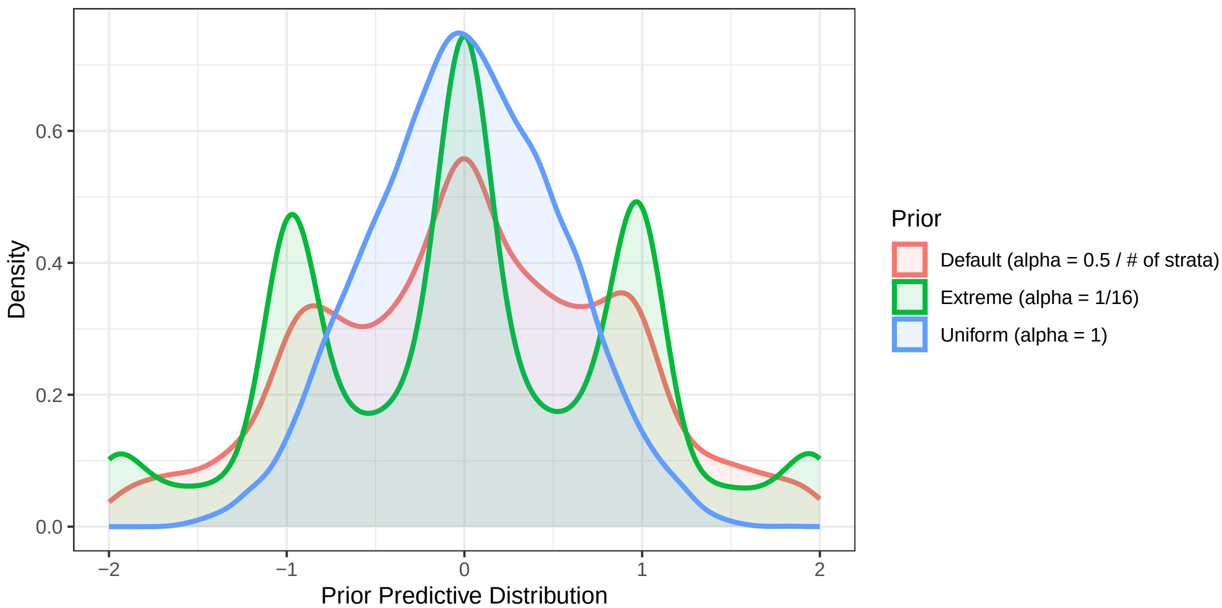
$$\widehat{\delta}_p^{(r)} = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{s \in \mathcal{S}_1^*} (\widehat{\mu}_{is}^{(r)}(1,1) - \widehat{\mu}_{is}^{(r)}(0,1))\widehat{\rho}_{is}^{(r)} \right) - \left(\sum_{s \in \mathcal{S}_0^*} (\widehat{\mu}_{is}^{(r)}(1,1) - \widehat{\mu}_{is}^{(r)}(0,1))\widehat{\rho}_{is}^{(r)} \right),$$

For the in-sample quantity, we can then draw *imputed* values of the missing potential outcomes themselves $\widehat{Y}_i^{*,(r)}(1) \sim \text{Bin}(\widehat{\mu}_i^{(r)}(1,1))$ and $\widehat{Y}_i^{*,(r)}(0) \sim \text{Bin}(\widehat{\mu}_i^{(r)}(0,1))$. We can combine this with the imputed value of M_i^*, which mechanically derives from $\widehat{S}_i^r$, to get the rth draw from the posterior of δ_s,

$$\widehat{\delta}_s^{(r)} = \frac{\sum_{i=1}^n \widehat{M}_i^{*,(r)} \left\{ \widehat{Y}_i^{*,(r)}(1) - \widehat{Y}_i^{*,(r)}(0) \right\}}{\sum_{i=1}^n \widehat{M}_i^{*,(r)}} - \frac{\sum_{i=1}^n \left(1 - \widehat{M}_i^{*,(r)}\right) \left\{ \widehat{Y}_i^{*,(r)}(1) - \widehat{Y}_i^{*,(r)}(0) \right\}}{\sum_{i=1}^n \left(1 - \widehat{M}_i^{*,(r)}\right)}.$$

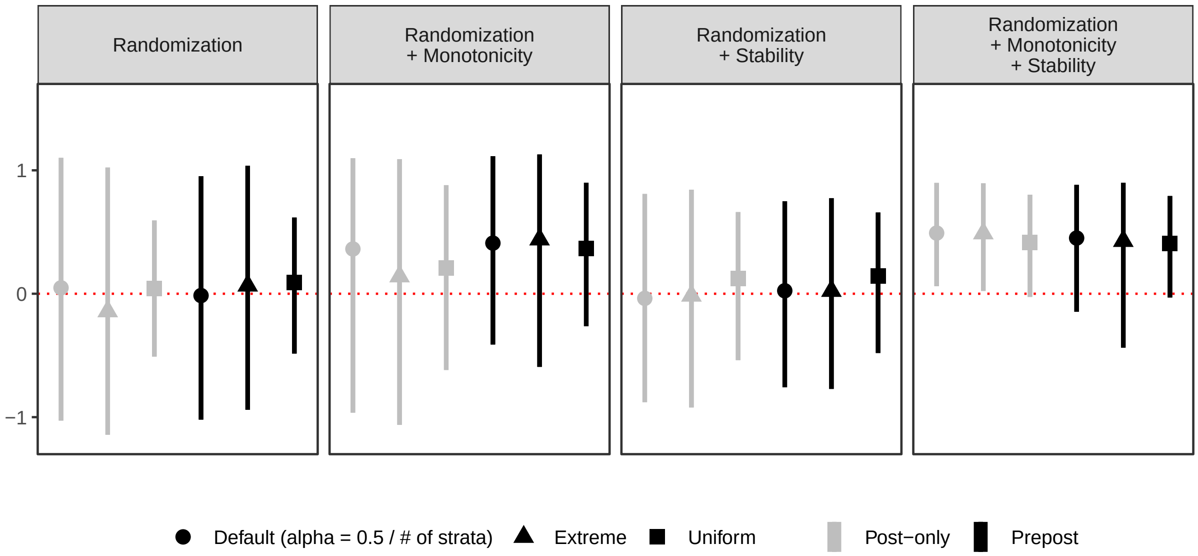
Broadly speaking, we would not expect very large differences between these two targets, except for slightly less posterior variance for the in-sample version.

Figure SM.1: Prior predictive distribution of the parameter under three different prior distributions: (red) the default priors that scales a Jeffreys prior by the number of principal strata; (blue) a uniform prior on all parameters; and (green) a more extreme prior that has $\alpha = 1/16$.



As discussed in the main text, the priors need careful attention because they drive the identification of the parameters that are unidentified by the likelihood. One additional complication comes from how the ultimate quantity of interest is a function of the parameters so we cannot directly place, for example, a uniform prior on δ. Figure SM.1 shows the prior predictive distribution for interaction with three different priors when monotonicity and stable moderator under control are assumed and there are no covariates. The uniform prior on all parameters results in a prior on δ that has more density in the center of identified range than we might expect. This result is similar to how sums of uniform random variables are not themselves uniform. We can counteract this issue

Figure SM.2: Comparing Bayesian estimates for δ for default, extreme, and uniform priors



Notes: Figure shows posterior means and 95% credible intervals for δ under different sets of assumptions, applied to either the post-only data (grey) or the combined pre-post data (black). Estimates are shown with default, extreme, and uniform priors, (denoted by circles, triangles, and squares, respectively). We follow the original authors in using age, gender, education, and closeness to one's ethnic group as covariates. The naïve OLS estimates are included for comparison.

by reducing the Dirichlet and Beta hyperparameters below 1 to put more density at extreme values of the parameters compared to the center. Dropping these parameters down to 1/16 (in green) leads to more mass on strata means closer to 0 or 1 and strata probabilities closer to 0 and 1. In terms of the interaction, this leads to more mass at the values -2, -1, 0, 1, and 2. Our default prior (red) is one that scales the hyperparameters by the inverse of the number of strata to achieve something closer to a uniform distribution.

Additionally, we re-ran the Gibbs empirical analysis of the Horowitz and Klaus (2020) study, adjusting the priors to the extreme values or uniform values in the previous simulation. The results are displayed in Figure SM.2, demonstrating the general consistency of the point estimates across starting priors, although there is some fluctuations in the variance of the results.

D.1 Simulation Evidence for the Bayesian Approach

Simulation Study I. In the first simulation study, we generate simulated data with $n = 1000$ constructed using a data generating process that matches the Bayesian posterior, pre-specifying coefficient values for the outcome and principal strata models, randomly drawing values of Z, T, and three covariates X_1, X_2, and X_3, and generating values of Y and M from the models. Tables SM.2 and SM.3 in the Supplemental Materials display the β coefficients for the outcome and ψ coefficients for the for the true data generating process (DGP). The DGP assumes that monotonicity and stable moderator under control both hold so that there are three feasible strata ($\mathcal{S} = \{000, 100, 111\}$). Thus, in this setting it would be most appropriate to incorporate both assumptions into the MCMC algorithm for sampling from the posterior distribution. Since these assumptions narrow the non-parametric bounds, we expect the assumptions to reduce variance of the posterior distribution of δ.

To test this, we perform a Monte Carlo simulation with 1,000 iterations. For each iteration, we calculate the posterior distribution of δ with the same data across four different versions of the MCMC algorithm: enforcing just the monotonicity assumption, enforcing just the stable moderator under control assumption, enforcing neither assumption, and enforcing both assumptions. Each run of our MCMC algorithm consists of 4 chains with 2,000 iterations each, 200 burn-in (or warm-up) iterations, and a thinning parameter of 2. Both in-sample and population δ values are calculated at each iteration and the variance of the posterior is calculated from a sample of 1,000 draws from the posterior. This is done for each of the 1,000 simulated datasets, and for each dataset we compute the percent reduction in variance compared to the MCMC algorithm with no assumptions when using the algorithm with the monotonicity assumption, the stable moderator under control assumption, or both assumptions.

Figure SM.3 presents boxplots for the distribution of reductions in variance for each combination of assumptions. Both the monotonicity and stable moderator assumptions on their own reduce the variance compared to no assumptions, while making both assumptions reduces the variance even further. The monotonicity assumption showed a median posterior variance reduction of 33.0% for

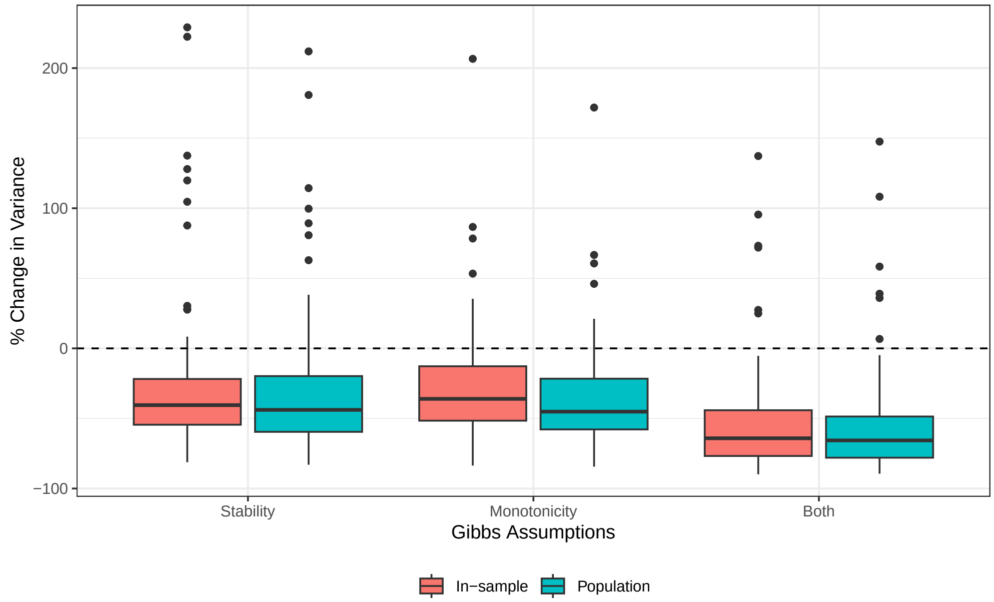


Figure SM.3: Variance reduction from different combinations of assumptions. Boxplots present distribution of % variance reduction of δ from the MCMC algorithm with the labeled assumptions compared to same algorithm with no assumptions, across 1,000 draws of simulated data. For each simulation iteration, 4 chains were run for each combination of assumptions. MCMC parameters: 2,000 iterations, 200 burn-in, 2 thinning parameter, simulated data $n = 1,000$.

the in-sample δ and 39.9% for the population δ. The stable moderator under control assumption on reduced the posterior variance by a median reduction of 38.9% (in-sample δ) and 41.9% (population δ). The MCMC algorithm with both assumptions exhibited a posterior variance reduction of 64.2% (in-sample δ) and 65.2% (population δ).

Simulation Study II. In the second simulation study, we drew a series of simulated datasets under different conditions where the covariates had a weak, medium, or strong correspondence with the outcome and principal strata in the data generating processes. Thus, there were six total conditions: Weak, Medium, and Strong influence in the outcome DGP; and Weak, Medium, and Strong influence in the principal strata DGP. The values of the coefficients for these conditions are β and ψ values of 0, 0.25, and 0.5, respectively. When varying the influence of covariates in the outcome DGP, the influence of covariates on the strata was held constant, and the influence in the outcome model was similarly held constant when varying influence in the strata DGP. Fixed values

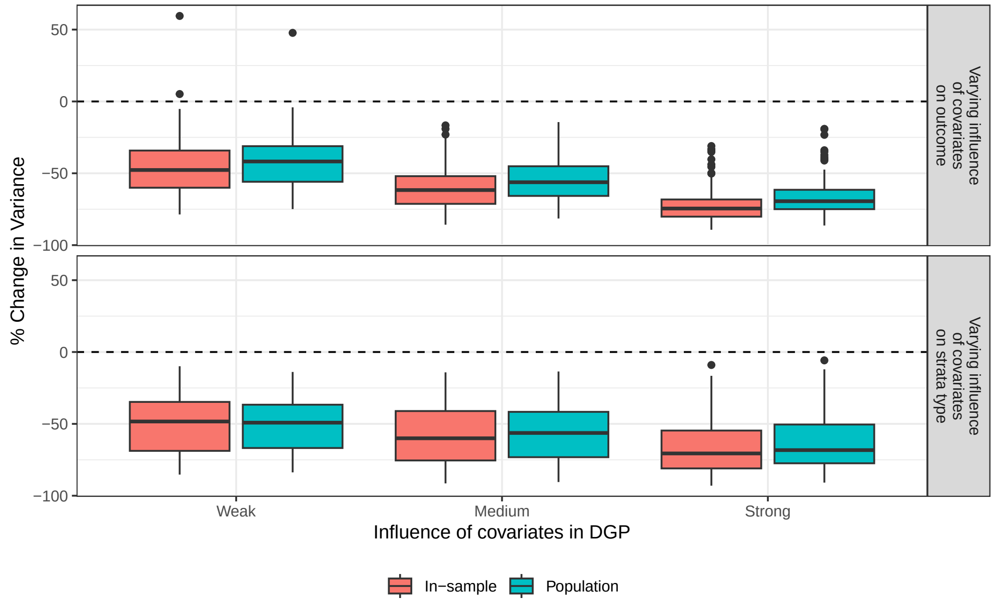


Figure SM.4: Variance reduction from incorporation of covariates. Boxplots present distribution of % variance reduction of δ from Gibbs with covariates compared to MCMC without covariates, across 1,000 draws of simulated data. For each draw 4 MCMC chains were run for each combination of assumptions. MCMC parameters: 2,000 iterations, 200 burn-in, 2 thinning parameter, simulated data $n = 1,000$.

of the β's and ψ's are shown in the Supplemental Materials in Tables SM.4 and SM.5.

For each condition, we drew 1,000 simulated datasets and ran the MCMC algorithm twice: one time incorporating covariates and one time omitting them. Each MCMC run consisted of the same iterations, burn-in, and thinning parameters as in the previous simulation study. We again calculate in-sample and population δ values for each iteration of the Gibbs and calculate the variance of the posterior distribution and the % variance reduction comparing the Gibbs with covariates to that without. Figure SM.4 presents boxplots for the distribution in variance reduction. When we vary the influence of covariates on the outcome, we see a clear variance reduction in all conditions, and we observe a larger reduction as the influence of covariates on the outcome in the DGP increases. When testing the impact of incorporating covariates across different levels of influence in the DGP on the strata, the pattern is less pronounced, with overall reduction increases in all conditions but slightly lower reductions in the Medium than Weak condition. The Strong condition still has the largest variance reduction overall, however, so in general the efficiency gains from incorporating covariates are increasing as the influence of covariates on strata in the data increases.

E Additional Simulation Details

Table SM.2: β Values for DGP in Bayesian Assumptions Simulation

Variable	β
(Intercept)	-2.00
X1	1.00
X2	0.15
X3 (Medium)	0.24
X3 (Large)	0.28
T	0.83
Z	-0.01
T:Z	0.11
S111	0.41
S100	0.62
T:S111	0.01
T:S100	0.23
Z:S111	0.20
Z:S100	-0.02
T:Z:S111	-0.90
T:Z:S100	0.09

Table SM.3: ψ Values for DGP in Bayesian Assumptions Simulation

	S111	s100	s000
(Intercept)	-2.06	-1.00	0.00
X1	2.00	1.50	0.00
X2	0.50	0.17	0.00
X3 (Medium)	1.35	-0.28	0.00
X3 (Large)	1.75	-1.01	0.00

E.1 Comparing the sharp bounds and Bayesian approach without covariates

Figure SM.5 displays the non-parametric bounds (with 95% confidence intervals) and Bayesian estimates (posterior means with 95% credible intervals) for δ under different sets of assumptions

Table SM.4: Fixed β Values in Covariate Simulation

Variable	β
(Intercept)	-1.00
X1	1.00
X2	0.50
X3 (Medium)	0.50
X3 (Large)	0.28
T	0.83
Z	-0.01
T:Z	0.11
S111	0.41
S100	0.62
T:S111	2.00
T:S100	-0.13
Z:S111	0.50
Z:S100	0.10
T:Z:S111	0.05
T:Z:S100	0.01

Table SM.5: Fixed ψ values in Bayesian Covariate Simulation

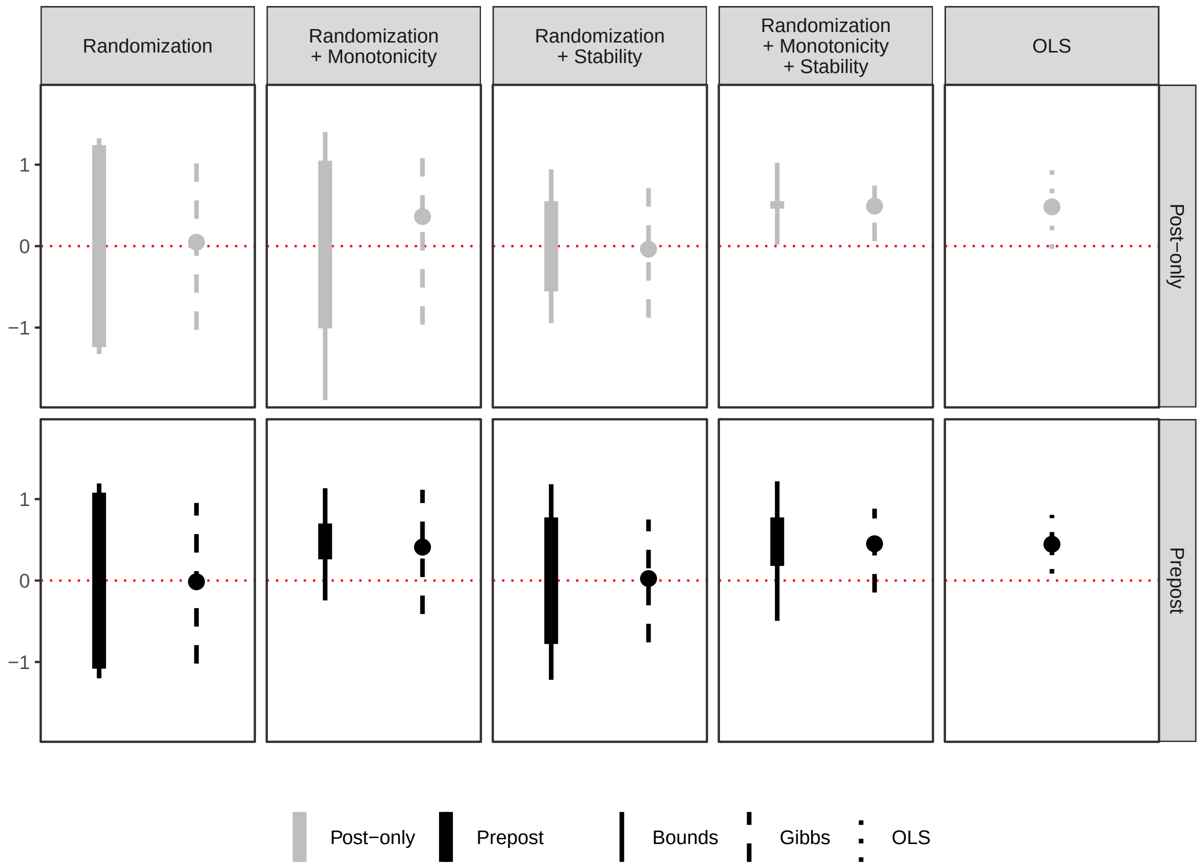
	S111	S100	S000
(Intercept)	-2.06	-1.00	0.00
X1	2.00	1.50	0.00
X2	0.50	0.17	0.00
X3 (Medium)	1.35	-0.28	0.00
X3 (Large)	1.75	-1.01	0.00

applied to the post-only data (in grey) and to the combined pre-post data (in black). We also include the naïve OLS estimate with 95% confidence interval for comparison in the final panel.

E.2 Incorporating covariates into the Bayesian approach

Thus far, our Bayesian estimates have omitted covariates to aid comparison with the nonparametric bounds. However, as discussed above, a key attraction of the Bayesian approach is the ease with which we can incorporate additional information. Figure SM.6 presents posterior means with 95% credible intervals, both with and without covariates, under different assumptions. We follow the original authors in using age, gender, education, and closeness to one's ethnic group as covariates.

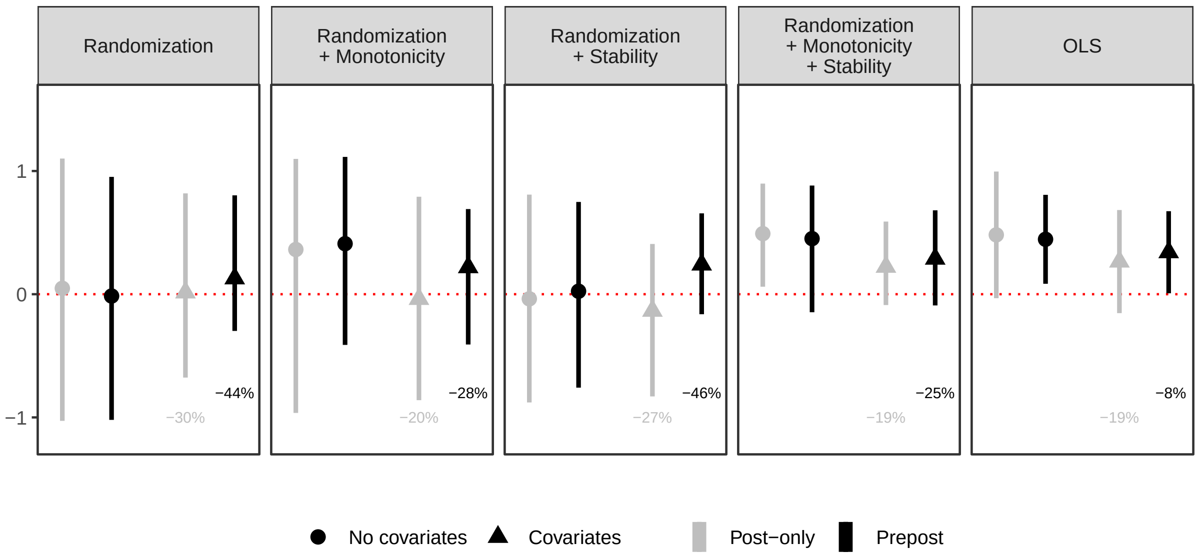
Figure SM.5: Comparing non-parametric bounds and Bayesian estimates for δ under different assumptions



Notes: The figure shows nonparametric bounds and Bayesian estimates of the quantity of interest under different sets of assumptions applied to either the post-only data (grey) or the combined pre-post data (black). The thick bars denote the width of the bounds, and thinner lines denote the 95% confidence intervals around the bounds. Across the first four panels, the thin lines with dots denote the Bayesian posterior mean and 95% credible interval. This estimate included no covariates to facilitate comparison with the nonparametric bounds. For the final panel ("OLS"), the thin lines with dots denote the OLS estimate and 95% confidence interval.

Including covariates significantly tightens the credible intervals, especially when fewer assumptions are imposed. For example, when only randomization is assumed, the width of the 95% credible interval shrinks by more than 40% when including covariates. While including covariates does not alter our substantive conclusions in this case, it does show that incorporating additional information can lead to large gains in precision. Since researchers often include a wide range of control variables

Figure SM.6: Comparing Bayesian estimates for δ with and without covariates



Notes: Figure shows posterior means and 95% credible intervals for δ under different sets of assumptions applied to either the post-only data (grey) or the combined pre-post data (black). Estimates are shown with and without the inclusion of covariates (denoted by triangles and circles, respectively), and the numbers indicate the reduction in the width of the credible intervals due to the inclusion of covariates for the post-only data (in grey) and the combined pre-post data (in black). We follow the original authors in using age, gender, education, and closeness to one's ethnic group as covariates. The naïve OLS estimates are included for comparison.

in the design of a survey experiment, flexibly leveraging this information is a key advantage of the Bayesian approach.

References

Andrews, Donald W. K. and Sukjin Han. 2009. "Invalidity of the bootstrap and the m out of n bootstrap for confidence interval endpoints defined by moment inequalities." *The Econometrics Journal* 12:S172–S199.

Chernozhukov, Victor, Han Hong and Elie Tamer. 2007. "Estimation and Confidence Regions for Parameter Sets in Econometric Models1." *Econometrica* 75(5):1243–1284.

Chernozhukov, Victor, Whitney K. Newey and Andres Santos. 2023. "Constrained Conditional Moment Restriction Models." *Econometrica* 91(2):709–736.

Hirano, Keisuke, Guido W. Imbens, Donald B. Rubin and Xiao-Hua Zhou. 2000. "Assessing the effect of an influenza vaccine in an encouragement design." *Biostatistics* 1(1):69–88.

Horowitz, Jeremy and Kathleen Klaus. 2020. "Can politicians exploit ethnic grievances? An experimental study of land appeals in Kenya." *Political Behavior* 42(1):35–58.

Imbens, Guido W. and Charles F Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72(6):1845–1857.

Imbens, Guido W. and Donald B. Rubin. 1997. "Bayesian inference for causal effects in randomized experiments with noncompliance." *The Annals of Statistics* 25(1):305–327.

Levis, Alexander W., Matteo Bonvini, Zhenghao Zeng, Luke Keele and Edward H. Kennedy. 2023. "Covariate-assisted bounds on causal effects with instrumental variables.".

Mealli, Fabrizia and Barbara Pacini. 2013. "Using Secondary Outcomes to Sharpen Inference in Randomized Experiments With Noncompliance." *Journal of the American Statistical Association* 108(503):1120–1131.

Polson, Nicholas G., James G. Scott and Jesse Windle. 2013. "Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables." *Journal of the American Statistical Association* 108(504):1339–1349.

Torgovitsky, Alexander. 2019. “Nonparametric Inference on State Dependence in Unemployment.” *Econometrica* 87(5):1475–1505.

Ye, Ting, Luke Keele, Raiden Hasegawa and Dylan S. Small. 2023. “A Negative Correlation Strategy for Bracketing in Difference-in-Differences.” *Journal of the American Statistical Association* pp. 1–13.