Online Appendix to
Monotone Ecological Inference

A Supplemental Figures/Tables

Tier	Uniquely Matched	Unmatched	Proportion Matched
Tier 1	1,028,172	764,810	.569
Tier 2	61,906	712,439	.603
Tier 3	129,298	582,697	.675
Tier 4	37,409	545,694	.695
Tier 5	35,231	510,427	.715
Tier 6	89,605	420,430	.764

Table S1: Summary of matched patient counts for the auxiliary individual-level dataset. Tier 1 contains the highest confidence matches and Tier 6 contains the lowest confidence matches (for details, see Appendix B). For individuals with multiple matches, one match is chosen at random.

	Tier 1	**Tier 1-2**	**Tier 1-3**	**Tier 1-4**	**Tier 1-5**	**Tier 1-6**
$E[Cov(Y, X_N \mid X)]$	-0.00463 (8e-05)	-0.00452 (8e-05)	-0.00451 (7e-05)	-0.00447 (7e-05)	-0.00455 (7e-05)	-0.00466 (7e-05)
$E[Cov(Y, X \mid N)]$	-0.00691 (0.00022)	-0.00703 (0.00021)	-0.0066 (0.0002)	-0.00652 (0.0002)	-0.00623 (0.0002)	-0.00592 (0.0002)

Table S2: Conditional covariance (standard error) estimated from the auxiliary individual-level dataset. Y indicates whether an individual is vaccinated, X indicates whether they are Republican, and X_N is the proportion of Republicans in their county. Individuals all have equal weight and are matched in a tiered approach, where Tier 1 contains the highest confidence matches and Tier 6 contains the lowest confidence matches. Individuals with multiple matches are discarded.

	Tier 1	**Tier 1-2**	**Tier 1-3**	**Tier 1-4**	**Tier 1-5**	**Tier 1-6**

$E[Cov(Y, X_N\|X)]$	-0.00463 (8e-05)	-0.00451 (8e-05)	-0.0045 (7e-05)	-0.00446 (7e-05)	-0.00459 (7e-05)	-0.00481 (7e-05)
$E[Cov(Y, X\|N)]$	-0.00698 (0.00022)	-0.00706 (0.00021)	-0.00638 (0.0002)	-0.00633 (0.0002)	-0.00594 (0.00019)	-0.00519 (0.00019)

Table S3: Conditional covariance (standard error) estimated from the auxiliary individual-level dataset. Y indicates whether an individual is vaccinated, X indicates whether they are Republican, and X_N is the proportion of Republicans in their county. Individuals all have equal weight and are matched in a tiered approach, where Tier 1 contains the highest confidence matches and Tier 6 contains the lowest confidence matches. For individuals with multiple matches, one match is chosen at random.

	Tier 1	**Tier 1-2**	**Tier 1-3**	**Tier 1-4**	**Tier 1-5**	**Tier 1-6**
$E[Cov(Y, X_N\|X)]$	-0.00361 (0.00042)	-0.00363 (0.00041)	-0.00353 (0.00038)	-0.00354 (0.00037)	-0.0036 (0.00037)	-0.00366 (0.00036)
$E[Cov(Y, X\|N)]$	-0.00847 (0.001)	-0.00876 (0.00096)	-0.00833 (0.00089)	-0.00712 (0.0009)	-0.00691 (0.00088)	-0.00656 (0.00087)

Table S4: Conditional covariance (standard error) estimated from the auxiliary individual-level dataset. Y indicates whether an individual is vaccinated, X indicates whether they are Republican, and X_N is the proportion of Republicans in their county. Individuals are re-weighted (see Appendix B) and are matched in a tiered approach, where Tier 1 contains the highest confidence matches and Tier 6 contains the lowest confidence matches. Individuals with multiple matches are discarded.

B Auxiliary Matched Analysis: Additional Detail

Data

Individual COVID-19 Vaccinations

We used data on individual COVID-19 vaccination status from the American Family Cohort (AFC). The AFC is derived from the American Board of Family Medicine PRIME Registry [Balraj et al., 2023]. PRIME is used by practices to track electronic health records data for Medicare and Medicaid reporting as well as measurement of social determinants of health [Phillips, 2017]. PRIME contains approximately 8 million patient records across 47 states [Balraj et al., 2023]. From this dataset, we obtained COVID-19 vaccination status following Hao et al., 2023 [Hao et al., 2023]. We considered all patients over 5 years old who did not die during 2021. Patients who had no history of receiving immunizations or visiting practices operating from January 1, 2019 to December 31, 2021 were excluded. We considered patients to be vaccinated if they received two COVID-19 vaccinations, or a single dose of the Janssen vaccine, before December 18, 2021 (n = 339,484). All other patients (n = 1,463,303) were considered unvaccinated.

Individual Voter Information

We used national voter records from the L2 database [L2, 2024]. This database is collected and assembled by the L2 organization and sold for use in academic, political, or business settings. In addition to minimizing data gathering steps, the database was constructed using a number of standardization and cleaning preprocessing steps that ensure greater comparability across states. We obtained voter name, state, date of birth, and registered political party for 214,125,844 voters nationally.

County-Level Voter Information

We used election returns at the county level from the 2020 Presidential Election from the [MIT Election Data and Science Lab, 2020] to provide the b_i in the individual-level analysis to check the signs of the covariance terms.

Matching

We used a tiered matching strategy to pair patients from the AFC dataset with voters from the L2 dataset. This strategy was informed by [Roos Jr et al., 1986], who implement a tiered matching system for records with few identifiers. We implement a similar system, where "Tier 1" matches represent high confidence, and "Tier 6" matches represent low confidence (see Table S5, S1).

We used the characteristics of the datasets to determine which factors to match on in each tier. In the L2 data, we found that recorded birth dates were non-uniformly distributed, with strong biases towards birth dates on the first of each month and on January 1. This suggests that the L2 data may record births as occurring on the first of the month or on January first when respondents only note their birth month or year. Thus, in some tiers, we allow for matches without identical birth dates as long as the birth month and year are the same.

We require an exact match on first and last name in all tiers. We considered matching on first initial and last name, but we found that these matching schemes resulted in clearly erroneous matches. We also considered matching on phonetic last name. However, since our datasets come from official voter and patient records, it is unlikely that last names are misspelled. Thus, a phonetic encoding may have led to incorrect matches.

In our final matching scheme, tier 1 requires matches at the highest level of specificity with exact matches across all available factors. We are most confident that tier 1 matches are correct, since they match for all six factors. Tier 6 only requires exact matches on three factors, so we are less confident that these matches are correct. The tiers are defined as follows:

- Tier 1: Unique match on exact first name, last name, state, and birth date (day/month/year)
- Tier 2: Unique match on exact first name, last name, state, and birth month/year
- Tier 3: Unique match on exact first name, last name, state, and birth year
- Tier 4: Unique match on exact first name, last name, and birth date (day/month/year)

- Tier 5: Unique match on exact first name, last name, and birth month/year
- Tier 6: Unique match on exact first name, last name, and birth year

We tested two different ways of considering ties, where multiple vaccination records or voters were matched to one voter or vaccination record in a given tier. In the first approach, we did not allow any ties ("zero-tie"), deleting ties from matched data at each tier. In the second approach, following [Roos Jr et al., 1986], we kept one record at random from each set of ties ("random-tie"). We present analyses of the zero-tie strategy in the main text. Analyses for the random-tie strategy are presented in the supplement.

Tier	**Uniquely Matched**	**Unmatched**	**Proportion Matched**
Tier 1	1,017,300	764,810	0.563
Tier 2	57,517	702,882	0.595
Tier 3	101,999	573,883	0.651
Tier 4	39,042	563,416	0.673
Tier 5	26,207	511,227	0.687
Tier 6	39,464	414,351	0.709

Table S5: Summary of matched patient counts for the auxiliary individual-level dataset. Tier 1 contains the highest confidence matches and Tier 6 contains the lowest confidence matches. Individuals with multiple matches are discarded.

When estimating covariances and confidence intervals, we tested two different weighting schemes. In the first scheme, each record was given equal

weight ("unit weight"). In the second scheme, we re-weighted records such that the sum of records for a given county in a given political party matched the number of voters of that party in that county, as calculated from our county-level voter information ("re-weighted"). Because many counties had few records in our matched dataset, we restricted the re-weighted data to only weight patients coming from counties with over 20 records – all other patients were given weight 0. We present analyses for the unit weight strategy in the main text. Analyses for the re-weighted strategy are presented in the supplement.

When patient counties recorded in the AFC and L2 data did not match, we chose to use the AFC county. For the zero-tie and random-tie approaches, 4,346 and 5,235 records respectively were removed due to missing AFC county. Additionally, in Alaska, the county-level dataset recorded Alaskan districts rather than counties. We removed all people with AFC counties in Alaska from the merged dataset. We also restricted to only Democrat and Republican voters. This left us with a total of 987,663 (437,722 Democrat, 549,941 Republican) and 1,061,517 (476,496 Democrat, 585,021 Republican) individuals matched in the zero-tie and random-tie approaches, respectively.

C Estimating Conditional Covariances

In this section, we describe how we estimate $\theta = \mathbb{E}[\mathrm{Cov}(A, B|C)]$, where C takes values in $\mathcal{C}$, and $\mathrm{Var}(\hat{\theta})$. We assume that the distribution of C is known, as is the distribution of $A|C$.

We know:

$$\theta = \sum_{c \in \mathcal{C}} \Pr(C = c) \operatorname{Cov}(A, B|C = c)$$

which we estimate by

$$\begin{aligned}
\hat{\theta} &= \sum_{c \in \mathcal{C}} \Pr(C = c) \sum_{i:C_i = c} \frac{1}{n_c - 1}(A_i - \bar{A})(B_i - \bar{B}) \\
&= \sum_{c \in \mathcal{C}} \frac{\Pr(C = c)}{n_c - 1} \sum_{i:C_i = c} (A_i - \bar{A})(B_i - \bar{B}) \\
&= \sum_{c \in \mathcal{C}} \Pr(C = c) \left(\frac{\sum_{i:C_i = c} (A_i - \bar{A})^2}{n_c - 1} \right) \left(\frac{\sum_{i:C_i = c} (A_i - \bar{A})(B_i - \bar{B})}{\sum_{i:C_i = c} (A_i - \bar{A})^2} \right) \\
&= \sum_{c \in \mathcal{C}} \Pr(C = c) \operatorname{Var}(A|C = c) \hat{\beta}_c
\end{aligned}$$

where $\hat{\beta}_c$ is the coefficient from the linear regression of A on B when $C = c$, n_c is the number of entries where $C = c$, and $\mathrm{Var}(A|C = c)$ is the variance of A for entries where $C = c$.

We now estimate $\mathrm{Var}(\hat{\theta})$.

$$\begin{aligned}
\mathrm{Var}(\hat{\theta}) &= \mathrm{Var}(\sum_{c \in \mathcal{C}} \Pr(C = c) \operatorname{Var}(A|C = c) \hat{\beta}_c) \\
&= \sum_{c \in \mathcal{C}} \mathrm{Var}(\Pr(C = c) \operatorname{Var}(A|C = c) \hat{\beta}_c) \\
&= \sum_{c \in \mathcal{C}} \Pr(C = c)^2 \operatorname{Var}(A|C = c)^2 \operatorname{Var}(\hat{\beta}_c) \\
&= \sum_{c \in \mathcal{C}} \Pr(C = c)^2 \operatorname{Var}(A|C = c)^2 \mathrm{SE}(\hat{\beta}_c)^2
\end{aligned}$$

Where $\text{SE}(\beta_c)^2$ is estimated standard error from the linear regression of A on B when $C = c$.

In our setting, we seek to estimate $\mathbb{E}\left[\text{Cov}(X_N, Y|X)\right]$ and $\mathbb{E}\left[\text{Cov}(X, Y|X_N)\right]$. We estimate $\mathbb{E}\left[\text{Cov}(X_N, Y|X)\right]$ as follows:

We calculate $Pr(X = 1)$ as the weighted sum of entries where $X = 1$ over the total weight of the dataset. We then calculate $Var(X_N|X = i)$. For each value of X, we then perform weighted least squares regression $(Y \sim X_N|X = i)$. We calculate the coefficient on X_N of this regression, $\hat{\beta}_i$. For our estimate for the standard error of $\mathbb{E}\left[\text{Cov}(X_N, Y|X)\right]$, we calculate the standard error of $\hat{\beta}_i$ and follow the formulas above.

For $\mathbb{E}\left[\text{Cov}(X, Y|X_N)\right]$, we first note a useful proposition:

Proposition 14. *We can write that:*

$$\mathbb{E}[\text{Cov}(X, Y|N)] = \mathbb{E}[\text{Cov}(X, Y|X_N)]$$

Using this proposition, we estimate $\mathbb{E}\left[\text{Cov}(X, Y|N)\right] = \mathbb{E}\left[\text{Cov}(X, Y|X_N)\right]$ as follows:

We round X_N to the nearest .05, creating 19 groups $(d_1, ...d_{19})$ of X_N values (in this case, X_N = county proportion Republican). We estimate $Pr(X_N \in d_i)$ as the weighted sum of entries with $X_N \in d_i$ over the total weight of the dataset. We also calculate the variance of X, $\text{Var}(X|X_N \in d_i)$. We then perform weighted least squares regression $(Y \sim X|X_N \in d_i)$ and calculate the coefficient on X of this regression, $\widehat{\beta_{d_i}}$. For our estimate for the standard error of $\mathbb{E}\left[\text{Cov}(X, Y|X_N)\right]$, we calculate the standard error of $\widehat{\beta_{d_i}}$, $\text{SE}(\widehat{\beta_{d_i}})$ and follow the formulas above.

Proof of ***Proposition 14****.* By the law of total covariance, the left side is:

$$\mathbb{E}[\text{Cov}(X, Y|N)] = \text{Cov}(X, Y) - \text{Cov}(\mathbb{E}[X|N], \mathbb{E}[Y|N])$$

and similarly the right side is :

$$\mathbb{E}[\text{Cov}(X, Y|X_N)] = \text{Cov}(X, Y) - \text{Cov}(\mathbb{E}[X|X_N], \mathbb{E}[Y|X_N])$$

So it is enough that:

$$\mathbb{E}[\mathbb{E}[X|N]\mathbb{E}[Y|N]] - \mathbb{E}[\mathbb{E}[X|N]]\mathbb{E}[\mathbb{E}[Y|N]] = \mathbb{E}[\mathbb{E}[X|X_N]\mathbb{E}[Y|X_N]] - \mathbb{E}[\mathbb{E}[X|X_N]]\mathbb{E}[\mathbb{E}[Y|X_N]].$$

But notice that the second term on both sides is $\mathbb{E}[X]\mathbb{E}[Y]$ by the law of iterated expectations, while $\mathbb{E}[X|X_N] = \mathbb{E}[X|N] = X_N$, by definition of X_N. Thus, we simply require that:

$$\mathbb{E}[X_N \cdot Y_N] = \mathbb{E}[X_N \cdot \mathbb{E}[Y|X_N]].$$

But

$$\mathbb{E}[X_N \cdot Y_N] = \mathbb{E}[\mathbb{E}[X_N \cdot Y_N|X_N]] =$$

by the Law of Iterated Expectations, and X_N is constant given X_N, so

$$\mathbb{E}[X_N \cdot Y_N] = \mathbb{E}[X_N \cdot \mathbb{E}[Y_N|X_N]].$$

Now since Y_N is $\mathbb{E}[Y|N]$, we can write that:

$$\mathbb{E}[X_N \cdot \mathbb{E}[Y_N|X_N]] = \mathbb{E}[X_N \cdot \mathbb{E}[\mathbb{E}[Y|N]|X_N]].$$

But note that:

$$\mathbb{E}[\mathbb{E}[Y|N]|X_N] = \mathbb{E}[Y|X_N]$$

by the (general version of the) Law of Iterated Expectations. Hence,

$$\mathbb{E}[X_N \cdot Y_N] = \mathbb{E}[X_N \cdot \mathbb{E}[Y|X_N]]$$

as desired.

□

D Relaxing Assumption 1

Now we turn to relaxing Assumption 1. In Assumption 1, we are assuming that any two neighborhoods that have the same group prevalence will also share the same expected outcome by group; relaxing this means allowing that neighborhoods with the same group prevalence have different expected outcomes. In this case, the analogues to Propositions 11 -13 apply in an average sense; i.e. we can obtain upper and lower bounds on the average disparity and average values of Y_n^j over the set of neighborhoods with the same group prevalence.

For each of our random variables, we use the notation $\tilde{\cdot}_n$ to indicate that we are averaging over the set of neighborhoods $n' \in \mathcal{N}$ such that $X_{n'} = X_n$. For example:

$$\begin{aligned}\tilde{Y}_n^1 :&= \mathbb{E}[Y_n^1 | n \in \mathcal{N}^{-1}(X_n)] \\ &= \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n^1\end{aligned}$$

where $\mathcal{N}^{-1}(x_n)$ is the set of $n \in \mathcal{N}$ such that $X_n = x_n$ and $w_n = p_n / \sum_{n \in \mathcal{N}^{-1}(x_n)} p_n$. We define $\tilde{Y}_n^0$ similarly, and $\tilde{D}_n$ as $\tilde{Y}_n^1 - \tilde{Y}_n^0$. We can also define the method of bounds similarly — $\tilde{Y}_{MOB,n}^{1+} = \sum_{n \in \mathcal{N}^{1+}(x_n)} w_n Y_{MOB,n}^{1+}$ and so on, and immediately have that:

$$\begin{aligned}\tilde{Y}_n^1 &\in [\tilde{Y}_{MOB,n}^{1-}, \tilde{Y}_{MOB,n}^{1+}], \\ \tilde{Y}_n^0 &\in [\tilde{Y}_{MOB,n}^{0-}, \tilde{Y}_{MOB,n}^{0+}],\end{aligned}$$

because

$$Y_n^1 \leq Y_{MOB,n}^{1+} \ \forall n \in \mathcal{N}^{-1}(x_n) \implies \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n^1 \leq \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_{MOB,n}^{1+} = \tilde{Y}_{MOB,n}^{1+}$$

and so on; similarly, we can write that:

$$\begin{aligned}\tilde{D}_n &\in [\tilde{D}_{MOB,n}^-, \tilde{D}_{MOB,n}^+] \\ &= [\tilde{Y}_n^{1-} - \tilde{Y}_n^{0+}, \tilde{Y}_n^{1+} - \tilde{Y}_n^{0-}].\end{aligned}$$

These define the method of bounds. We now turn to the equivalents of Propositions 11-13. As before, define $\mu(x_n) := \mathbb{E}[Y | X_n = x_n]$. Define also

$$\tilde{\delta}_{B,n} := \mathrm{Var}[X | N = n] \left(\tilde{Y}_n^1 - \tilde{Y}_n^0 \right)$$

and

$$\tilde{\delta}_{W,n} := \sum_{n \in \mathcal{N}^{-1}(X_n)} w_n \tilde{\delta}_{W,n}.$$

Note first:

Lemma 3. *Define $\tilde{\delta}_{W,n}$, $\tilde{D}_n$ as above. Then:*

$$\tilde{D}_n = \mu'(X_n) - \tilde{\delta}_{W,n}$$

Proof. Note that:

$$\mu(x_n) = \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n;$$

this can be re-written as:

$$\begin{aligned}\mu(x_n) &= \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n \cdot (X_n Y_n^1 + (1 - X_n) Y_n^0) \\ &= X_n \cdot \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n^1 + (1 - X_n) \cdot \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n^0\end{aligned}$$

since X_n is constant for all $n \in \mathcal{N}^{-1}(x_n)$. Then taking the derivative gives:

$$\begin{aligned}\mu'(X_n) &= X_n \cdot \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n^{1'} + (1 - X_n) \cdot \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n Y_n^{0'} + \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n (Y_n^1 - Y_n^0) \\ &= \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n \left(X_n \mu_1'(x_n) + (1 - X_n)\mu_0'(x_n)\right) + \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n (Y_n^1 - Y_n^0) \\ &= \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n \delta_{W,n} + \sum_{n \in \mathcal{N}^{-1}(x_n)} w_n (Y_n^1 - Y_n^0) \\ &= \tilde{\delta}_{W,n} + \tilde{D}_n.\end{aligned}$$

□

We can now state the generalized versions of Propositions 11-13:

Proposition 15 (General version of Proposition 11)**.** *(i) If $\tilde{\delta}_{B,n} \geq 0$ for some set of neighborhoods $\mathcal{N}^{-1}(x_n)$, then:*

$$\begin{aligned}\tilde{D}_n &\in [0, \tilde{D}^+_{MOB,n}] \\ \tilde{Y}_n^1 &\in [\tilde{Y}_n, \tilde{Y}^{1+}_{MOB,n}] \\ \tilde{Y}_n^0 &\in [\tilde{Y}^{0-}_{MOB,n}, \tilde{Y}_n]\end{aligned}$$

(ii) If $\tilde{\delta}_{B,n} \leq 0$ for some set of neighborhoods $\mathcal{N}^{-1}(x_n)$, then:

$$\begin{aligned}\tilde{D}_n &\in [\tilde{D}^-_{MOB,n}, 0], \\ \tilde{Y}_n^1 &\in [\tilde{Y}^{1-}_{MOB,n}, \tilde{Y}_n] \\ \text{and } \tilde{Y}_n^0 &\in [\tilde{Y}_n, \tilde{Y}^{0+}_{MOB,n}].\end{aligned}$$

(iii) The bounds in (i) and (ii) are sharp in the absence of additional information.

Proposition 16 (General version of Proposition 12)**.** *(i)Suppose that* $\tilde{\delta}_{W,n} \geq 0$ *for some neighborhoods* $\mathcal{N}^{-1}(x_n)$*. Then:*

$$\begin{aligned}
\tilde{Y}_n &\in \left[\tilde{D}^{-}_{MOB,n}, \min\left\{\mu'(X_n), \tilde{D}^{+}_{MOB,n}\right\}\right], \\
Y^1_n &\in \left[\tilde{Y}^{1-}_{MOB,n}, \min\left\{\tilde{Y}^{1+}_{MOB,n}, \tilde{Y}_n + (1-X_n)\mu'(X_n)\right\}\right], \\
\text{and } \tilde{Y}^0_n &\in \left[\max\left\{\tilde{Y}^{0-}_{MOB,n}, \tilde{Y}_n - X_n \cdot \mu'(X_n)\right\}\right]
\end{aligned}$$

(ii) Suppose that $\tilde{\delta}_{W,n} \leq 0$ *for some set of neighborhoods* $\mathcal{N}^{-1}(x_n)$*. Then:*

$$\begin{aligned}
\tilde{Y}_n &\in \left[\max\left\{\mu'(x_n), \tilde{D}^{-}_{MOB,n}\right\}, \tilde{D}^{+}_{MOB,n}\right] \\
\tilde{Y}^1_n &\in \left[\max\left\{\tilde{Y}^{1-}_{MOB,n}, \tilde{Y}_n + (1-X_n)\cdot\mu'(X_n)\right\}, \tilde{Y}^{1+}_{MOB,n}\right] \\
Y^0_n &\in \left[\tilde{Y}^{0-}_{MOB,n}, \min\left\{\tilde{Y}^{0+}_{MOB,n}, \tilde{Y}_n - X_n\cdot\mu'(X_n)\right\}\right]
\end{aligned}$$

(iii) The bounds in (i) and (ii) are sharp in the absence of further information.

Proposition 17 (General Version of Proposition 13)**.** *Suppose that* $\tilde{\delta}_{B,n} \cdot \tilde{\delta}_{W,n} \geq 0$*. Then either:*

(i) $\mu'(X_n) \geq 0$*, and:*

$$\begin{aligned}
\tilde{D}_n &\in [0, \min\{\mu'(X_n), \tilde{D}^{+}_{MOB,n}] \\
Y^1_n &\in \left[\tilde{Y}_n, \min\{\tilde{Y}_n + (1-X_n)\cdot\mu'(X_n), \tilde{Y}^{1+}_{MOB,n}\right] \\
\tilde{Y}^0_n &\in \left[\tilde{Y}_n, \min\{\tilde{Y}_n - X_n\mu'(X_n), \tilde{Y}^{0+}_{MOB,n}\right];
\end{aligned}$$

or (ii) $\mu'(X_n) \leq 0$*, and*

$$\begin{aligned}
\tilde{D}_n &\in \left[\max\{\mu'(X_n), \tilde{D}^{-}_{MOB,n}, 0\right], \\
\tilde{Y}^1_n &\in [\max\{\tilde{Y}^{1-}_{MOB,n}, \tilde{Y}_n + (1-X_n)\mu'(X_n)\}, \tilde{Y}_n] \\
\text{and } Y^0_n &\in \left[\tilde{Y}_n, \min\{\tilde{Y}_n - X_n\cdot\mu'(X_n), \tilde{Y}^{0+}_{MOB,n}\}\right]
\end{aligned}$$

(iii) The bounds in (i) and (ii) are sharp in the absence of further information.

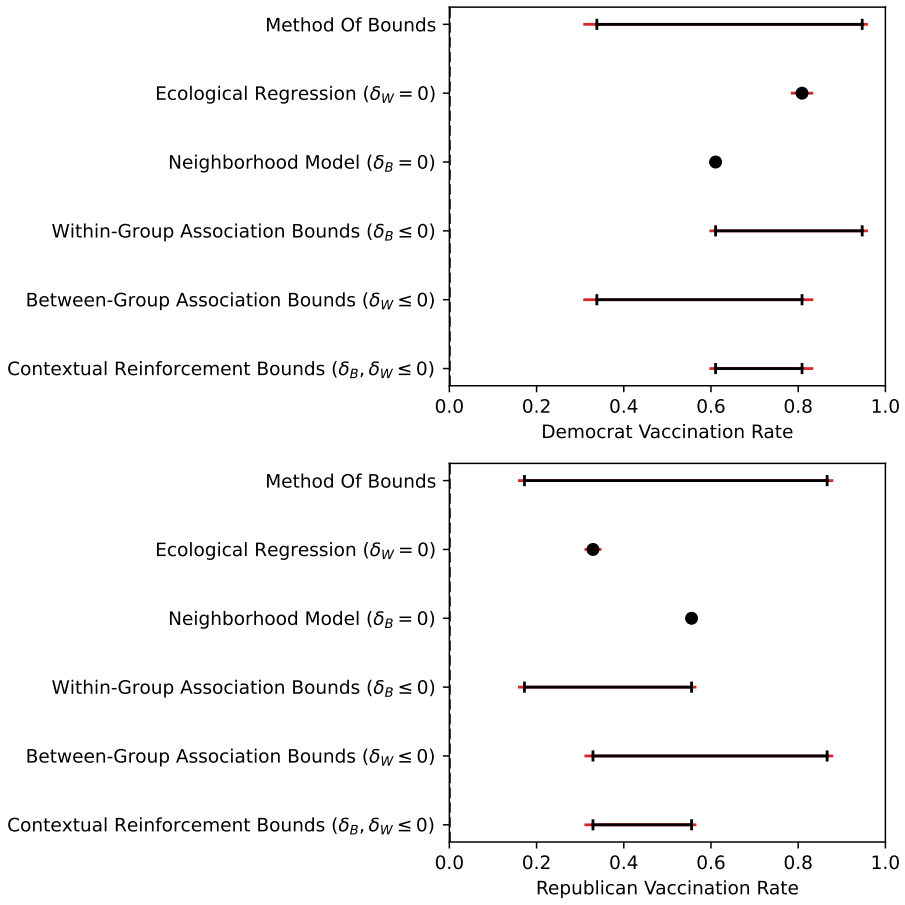
Proofs of these propositions follow similarly to those of the analogous special cases having replaced each quantity with its averaged counterpart.

E Proof of Theorems 2 and 3

Proof Sketch for Theorem 2. The key ideas are similar to those of Theorem 1, so we highlight only the key differences. Note that if $\delta_W \geq 0$ and $\delta_B \geq 0$, $Y^1 \leq Y^1_{ER}$ and $Y^1 \geq Y^1_{NM}$ by the previously derived formulas, so $Y^1_{NM} \leq Y^1 \leq Y^1_{ER}$; the opposite ordering holds when $\delta_W \leq 0$ and $\delta_B \leq 0$. The method of bounds must also hold, so Y^1_{ER} must be compared to Y^{1+}_{MOB} and the lesser of the two is an upper bound. (Note that by contrast, we need not compare Y^1_{NM} to Y^{1-}_{MOB} – the neighborhood model is always feasible by definition, so $Y^1_{NM} \geq Y^{1-}_{MOB}$). A similar statement can be made for Y^0_{ER}. Combining these gives the CR cells. For the case when $\delta_B \leq 0$ and $\delta_W \geq 0$, $Y^1 \leq Y^1_{N<}$ and $Y^1 \leq Y^1_{ER}$, as well as $Y^1 \leq Y^{1+}_{MOB}$ as always, so Y^1 must be less than the minimum of these three quantities; similarly, mutatis mutandis, with Y^1 and the maximum of Y^1_{ER}, Y^1_{NM}, and Y^{1-}_{MOB} in the opposite case. Sharpness results follow similarly as before: we use the feasibility of the boundary points of the range along with a mixture argument (as in the other proofs, using Lemma 2) to show that any point within the bounds is feasible. □

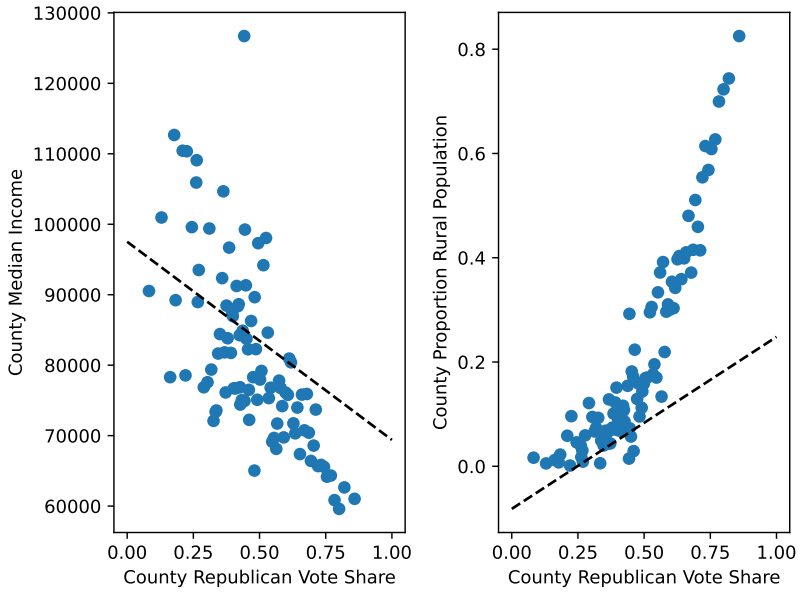
Proof Sketch for Theorem 3. This proof is similar to that of Theorem 3, with the concordant sign to ordering mapping reversed due to the sign difference in the biases laid out in Propositions 9 and 9. In particular, note that $\delta_B \geq 0 \implies Y^0 = Y^0_{NM} - \delta_B/(1 - \mathbb{E}[X]) \leq Y^0_{NM}$, while $\delta_W \geq 0 \implies Y^0 = Y^0_{ER} + \delta_W \mathbb{E}[X]/\operatorname{Var}(X_N) \geq Y^0_{ER}$. The rest of the proof follows similarly. □

Figure S1: Identification of Partisan Vaccination Rate



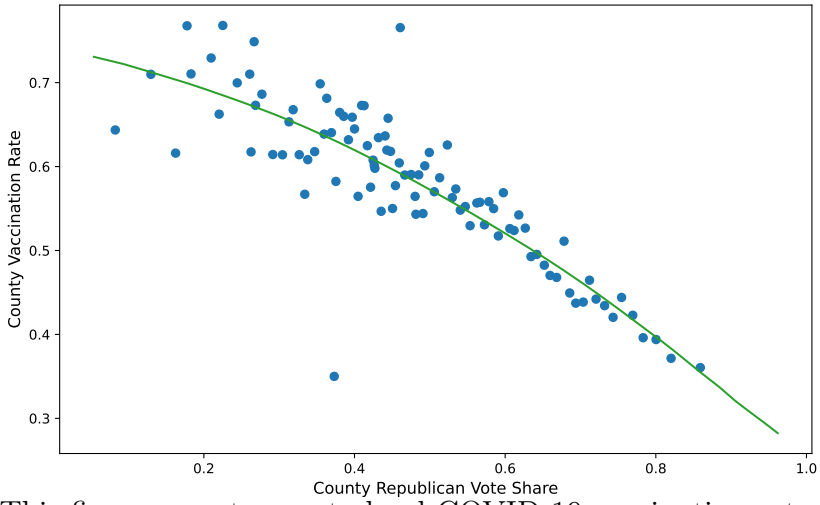
Notes: Mean vaccination rate for Democrats and Republicans (components of the partisan vaccination gap). Red bars are 95% confidence intervals following [Imbens and Manski, 2004]; the confidence intervals are based on standard errors from a county-level bootstrap with 1000 bootstrap replicates.

Figure S2: Median Income and Proportion Rural Population by Republican Vote Share



Notes: The figure reports county-level median income and proportion of the population living in rural areas. Counties are grouped into 100 equal-population bins (blue). Population-weighted linear regressions are also shown (black dotted lines). Household income data and proportion rural population were sourced from the 2023 5-year American Community Survey and the 2020 Decennial Census, respectively. Republican vote share was sourced from the MIT Election Data and Science Lab.

Figure S3: Local Linear Approximation to County Vaccine Data



Notes: This figure reports county-level COVID-19 vaccination rates by the share of voters in the county who voted for the Republican candidate in the 2020 presidential election. Counties are grouped into 100 equal-population bins (blue). The local linear approximation for county-level data is shown in green. This approximation uses an Epanechnikov kernel and 10-fold cross-validation to choose the bandwidth parameter.