北京大学中国经济研究中心
China Center for Economic Research
讨论稿系列
Working Paper Series

E2026018 2026-11-19

When Listwise Deletion Works: Causal Diagrams and Complete-Case Regression

Wenxin Lin Shiyao Liu

ABSTRACT

Listwise deletion (LWD) appears more common than imputation: at least 50% of the observational causal studies we review in leading journals state or imply it, while 7% describe an imputation model. We develop a framework to assess when complete-case regression consistently estimates the population average treatment effect (ATE). With a constant unit-level treatment effect, two checks suffice: whether covariates and the missingness indicator satisfy the adjustment criterion in an augmented causal diagram (DAG), and whether the estimator pools within-covariate comparisons appropriately, possibly using inverse propensity weighting. A collider–mediator illustration highlights the first check: the same missing-at-random (MAR) mechanism can require different adjustment choices. With heterogeneous effects, conditional effects must be transportable to the population and be reweighted to its population distribution. These formal results provide applied researchers with practical checks for assessing complete-case regression, recover existing guidance as special cases, and are illustrated with applications across fields.

When Listwise Deletion Works: Causal Diagrams and Complete-Case Regression[1]

Wenxin Lin[2] Shiyao Liu[3]

First Draft: March 25, 2026.

This Draft: November 19, 2026.

ABSTRACT

Listwise deletion (LWD) appears more common than imputation: at least 50% of the observational causal studies we review in leading journals state or imply it, while 7% describe an imputation model. We develop a framework to assess when complete-case regression consistently estimates the population average treatment effect (ATE). With a constant unit-level treatment effect, two checks suffice: whether covariates and the missingness indicator satisfy the adjustment criterion in an augmented causal diagram (DAG), and whether the estimator pools within-covariate comparisons appropriately, possibly using inverse propensity weighting. A collider–mediator illustration highlights the first check: the same missing-at-random (MAR) mechanism can require different adjustment choices. With heterogeneous effects, conditional effects must be transportable to the population and be reweighted to its population distribution. These formal results provide applied researchers with practical checks for assessing complete-case regression, recover existing guidance as special cases, and are illustrated with applications across fields.

Word count: 9,960

[1]Authors are listed in alphabetical order. We thank Jongyoon Baik, Adam Berinsky, Frederick Boehmke, Naoki Egami, Macartan Humphreys, Kosuke Imai, Hongyi Jiang, In Song Kim, Dean Knox, Licheng Liu, Naijia Liu, Adeline Lo, Jacob Montgomery, Xun Pang, Yiqing Xu, Teppei Yamamoto, Soichiro Yamauchi, Leo Yang Yang, Junjian Yi, Wei Zhong, Meixi Zhuang, and Cai Zuo, as well as participants at the Poster Session of the 2026 Annual Meeting of the Society for Political Methodology, the Second Asia-Pacific Politics and Public Administration Conference, the 2026 Visions in Methodology conference at the Hertie School, the 3rd Forum for Women in Political Science (China) in Nanjing, and seminar participants of the Political Economy workshop at Peking University and at the University of Macau, for helpful comments. We used Anthropic's Claude (Opus 5 and 5.5 and Sonnet 5 and 5.5), OpenAI's Codex (GPT-5.6, GPT-6, GPT-6.1, and GPT-6.2), and Alibaba's Qwen 3.8. All three tools were used to assist with language editing, LaTeX formatting, drafting, coding, debugging, pipeline management, and reproducibility checks. We also used Claude Opus 5.5 to conduct adversarial reviews. Claude Opus 5.5 and Claude Sonnet 5 coded articles for the review of missing-data practices (see Appendix A for details). All AI-assisted passages and code were reviewed, revised, and approved by the authors, who take full responsibility for the manuscript's scientific content.

[2]National School of Development, Peking University.

[3]China Center for Economic Research, Institute of South-South Cooperation and Development, National School of Development, Peking University; Governance/Lab, Massachusetts Institute of Technology. Corresponding author: `shiyaoliu at nsd.pku.edu.cn`.

1 Introduction

Missing data are common in social science research. Through a review of published observational causal studies in the *American Journal of Political Science* (AJPS) and *American Political Science Review* (APSR) between 2019 and 2024, Huang and Liu (2026) found that more than 50% of the studies mentioned missing values. Our review of a reconstructed sample of theirs finds that 80 of 158 articles (50.6%) state or imply listwise deletion in the main text, compared with 42 (26.6%) reporting any fill-in method and 11 (7.0%) describing an imputation model. Further, listwise deletion could also arise from software defaults: in R, `lm()`, `glm()`, and `ivreg()` drop incomplete observations under default `na.action = na.omit`, and mediation analyses may inherit these sample restrictions from their component models[4].

Multiple imputation uses partially observed records and can reduce bias and improve precision (King et al. 2001; Honaker and King 2010; Lall 2016), but under nonignorable missingness it can perform worse than listwise deletion (Pepinsky 2018). Listwise deletion avoids an imputation model and can yield consistent coefficient estimates under some nonrandom missingness (Arel-Bundock and Pelc 2018), yet it discards observed information and can introduce selection bias (Heckman 1979; King et al. 2001; Huang and Liu 2026).

In this paper, we study when regression after listwise deletion can be justified. Among the 80 articles that state or imply listwise deletion, 45 (56%) report at least one regression estimated after dropping cases with missing values[5]. Thus, we consider the following causal-regression workflow for listwise deletion: a researcher chooses an adjustment set intended to identify the population average treatment effect (ATE) and a regression specification intended to estimate it as if all required values were observed, then applies listwise deletion and reports the treatment coefficient from the complete-case regression as an estimate of that ATE. We develop a formal framework for this decision by making the assumed causes of missingness explicit and deriving conditions under which the complete-case regression consistently estimates the researcher's target.

[4] `mediate()` refits mediator and outcome models on their overlapping observations when `dropobs = TRUE` (Tingley et al. 2014).

[5] The regression count is a lower bound. Appendix A details the sample construction and coding rules.

Taking the oracle regression's identification of a variance-weighted average treatment effect as given, we formalize sufficient conditions under which regression after deletion consistently estimates the population ATE. Under a constant unit-level treatment effect, researchers must check two boxes: whether treatment comparisons remain valid after selection into the complete-case sample, and whether the fitted regression pools those comparisons appropriately. The first asks researchers to add the missing indicator and its causes to the DAG used to justify their original adjustment set, then reassess adjustment and treatment overlap. The second asks whether the outcome specification remains correct among complete cases, or whether treatment-propensity weighting can instead recover their average effect. With heterogeneous effects, a third box connects complete-case comparisons to the population target: when conditional effects are transportable to the population, researchers need to average them over the population covariate distribution rather than the complete-case covariate distribution. These checks provide a formal framework for assessing the causal and modeling assumptions behind complete-case regression, which require substantive judgment beyond statistical tests: as Subsection 4.2 illustrates, the same observed data can imply different conclusions about the validity of listwise deletion under different causal structures.

We also show why each box requires a separate check. In one simulation, selection leaves within-covariate treatment comparisons valid but changes the outcome specification, causing complete-case regression to pool those comparisons incorrectly. In the collider–mediator illustration, the same missing mechanism and MAR classification call for different adjustment choices depending on whether a covariate is a mediator or a confounder. Finally, the reweighting applications show how valid complete-case effects can differ from the population ATE when treatment effects vary and the covariate distributions differ.

Our approach builds on graphical analyses of selection bias and missing data. Bareinboim and Pearl (2012), Bareinboim, Tian and Pearl (2014), and Bareinboim and Pearl (2016) study when statistical and causal quantities can be recovered from data after selection. Mohan and Pearl (2021) develop the corresponding missing-data perspective. Daniel et al. (2012) and Saadati and Tian (2020) establish graphical conditions for causal analysis using complete records.

Ross, Breskin and Westreich (2020) show how effect modification and differences in covariate distributions affect marginal complete-case estimates. Hughes et al. (2019), most closely related, use causal diagrams to illustrate when complete-case regression yields unbiased estimates, taking the validity of the full-data analysis as given, but do not provide formal proofs. Building on their perspective, we derive and prove sufficient conditions that distinguish valid treatment comparisons among complete cases, which graphical criteria address, from the regression's pooling of those comparisons, which they alone do not guarantee. Together, these results bring graphical criteria for listwise deletion into the regression workflow used by applied political scientists, who, to our knowledge, do not currently use causal diagrams to justify deletion. Our framework thus gives applied researchers a practical set of checks for assessing whether their complete-case regression consistently estimates the ATE of interest.

Our framework also reconciles the differing recommendations of Pepinsky (2018) and Arel-Bundock and Pelc (2018) by specifying the causal and modeling assumptions under which they apply, complements guidance on choosing imputation models (Marbach 2022) by clarifying when deletion is justified in the first place, and generalizes the identification conditions underlying Huang and Liu (2026): we clarify when their imputation step is needed to recover the population covariate distribution and when complete-case effects can instead be averaged using an observed population distribution.

In Section 2, we use a simulation to build intuition for the two sufficient conditions: with a constant unit-level treatment effect, we contrast three missing-not-at-random mechanisms under which complete-case regression respectively recovers the population ATE, requires a pooling remedy, and fails because within-covariate comparisons are invalid.

Section 3 then derives formal sufficient conditions under which complete-case regression consistently estimates the population ATE: when within-strata treatment comparisons among complete cases identify causal effects, when regression pools them appropriately, and how the resulting effect connects to the population ATE under constant or heterogeneous effects.

Section 4 revisits conventional guidance, recovering familiar recommendations as special cases

and showing how a covariate's role as mediator or confounder changes the validity of listwise deletion under the same missingness pattern and MAR classification.

Section 5 develops reweighting for heterogeneous effects, generalizing the identification argument underlying Huang and Liu's (2026) two-stage projection estimator, and then illustrates it in a reanalysis of Bodea and Hicks (2015). Appendix F shows the potential of extension into generalized linear models.

Section 6 brings all three checks together in a reanalysis of the canvassing experiment (Kalla and Broockman 2023), illustrating the application of the three box roadmap we propose.

2 Motivating Simulation: Identification and Weighting

We begin with a simulation to illustrate conditions under which complete-case regression consistently estimates the population ATE, the causal effect that the oracle adjustment set identifies. The oracle sample contains $N_{\text{or}} = 20{,}000$ observations generated by the following process, with a constant unit-level treatment effect $\tau = 1$:

$$Y_i = D_i + X_i + U_i + \xi_i, \qquad \xi_i \sim N(0, 0.5^2).$$

The pretreatment confounder X_i is uniform on $\{-1, 0, 1\}$, while the unobserved cause of the outcome, U_i, is also uniform on $\{-1, 1\}$. X_i, U_i, and ξ_i are mutually independent. Treatment assignment depends only on X_i: $D_i \mid X_i = x \sim \text{Bernoulli}((1 + x^2)/4 + 0.30x)$.

In the observed data, Y_i and D_i are available for all 20,000 observations, but X_i is missing for some. Let $M_i = 0$ when X_i is observed and $M_i = 1$ otherwise. We consider three mechanisms governing whether X_i is observed:

$$\text{A:} \quad P(M_i = 0 \mid X_i = x, U_i = u) = 0.95 - 0.30x^2,$$

$$\text{B:} \quad P(M_i = 0 \mid X_i = x, U_i = u) = (0.95 - 0.30x^2)(1 + 0.50ux^2),$$

$$\text{C:} \quad P(M_i = 0 \mid X_i = x, U_i = u, D_i = d) = (0.95 - 0.30x^2)[1 + 0.50u(2d - 1)x^2].$$

All three mechanisms are not missing at random (NMAR): whether X_i is observed depends on its unobserved value. Under B, observation further depends on U_i, while under C, it additionally depends on the interaction between U_i and D_i. Averaging over U_i, each mechanism gives $P(M_i = 0 \mid X_i = x) = 0.95 - 0.30x^2$. Because X_i is uniform on $\{-1, 0, 1\}$, the missingness rate is $P(M_i = 1) = 0.25$ in all three cases.

An analyst could use listwise deletion to fit the following linear regression among complete cases:

$$Y_i = \alpha_{\text{cc}} + \beta_{\text{cc}} D_i + \gamma_{\text{cc}} X_i + \varepsilon_{\text{sim},i}, \qquad M_i = 0.$$

One rationale for this choice is the advice[6] of Arel-Bundock and Pelc (2018, p. 243):

> Under [the pattern of missingness driven by values of the independent variables], LWD can be used without fear that it will introduce bias in regression estimates.

Alternatively, an analyst might use multiple imputation to fit the same regression, guided by the general expectation expressed by King et al. (2001, p. 51):

> Thus, multiple imputation will normally be better than, and almost always not worse than, listwise deletion.

[6] An analyst might read Arel-Bundock and Pelc (2018) as endorsing listwise deletion here, because the oracle error $U_i + \xi_i$ is independent of X_i and D_i and missingness depends only on X_i. But their advice also requires the complete-case indicator to be independent of the regression error—exactly what mechanisms B and C violate, since M_i is associated with $U_i + \xi_i$ given X_i and D_i. Their advice therefore excludes these cases.

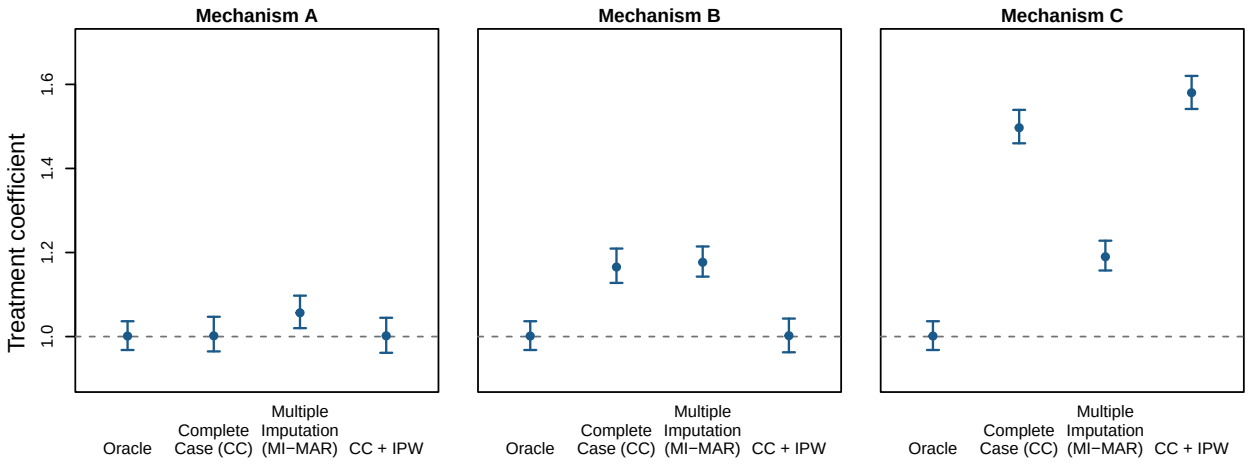


Figure 1: Estimated Treatment Coefficients under Different Missing Mechanisms

Note: Points show mean treatment-coefficient estimates across 500 replications. Bars show the 2.5th and 97.5th percentiles. The dashed line marks the true effect, $\tau = 1$. MI-MAR uses a missing-at-random imputation model, while CC + IPW uses inverse treatment-propensity weights. See Appendix Table B1 for numerical results.

We examine how both choices perform under each mechanism, benchmarking against the oracle estimator and adding a complete-case estimator that uses inverse treatment-propensity weights (CC + IPW) (Choi, Dekkers and le Cessie 2019). Figure 1 summarizes the treatment-coefficient estimates across 500 replications:

Under mechanism A, the unweighted complete-case estimates are centered on $\tau = 1$, consistent with Arel-Bundock and Pelc (2018)'s result that listwise deletion need not bias regression estimates when missingness depends on regressor values. The multiple-imputation estimates instead average 1.057: being NMAR in mechanism A, the consistency guarantee for correctly specified imputation under MAR does not apply. Under mechanism B, both the unweighted complete-case and multiple-imputation estimates are centered above $\tau = 1$. For complete-case regression, M_i depends on both X_i and U_i and is therefore associated with the oracle regression error $U_i + \xi_i$, violating the indicator–error independence condition of Arel-Bundock and Pelc (2018). The CC + IPW estimates remain centered near one. Under mechanism C, none of the three feasible estimators produces estimates centered on $\tau = 1$.

Sources of Error These patterns reflect two sources of error: treatment comparisons within X_i strata may fail to identify τ, or the pooled regression may fail to preserve valid within-stratum comparisons in its treatment coefficient. Table 1 reports the mean estimated intercepts and treatment coefficients from separate complete-case regressions of Y_i on D_i in each X_i stratum.

Table 1: Complete-Case Regression Coefficients within Each X_i

	Mechanism A ($M=0$)			Mechanism B ($M=0$)			Mechanism C ($M=0$)		
Coefficient	$X=-1$	$X=0$	$X=1$	$X=-1$	$X=0$	$X=1$	$X=-1$	$X=0$	$X=1$
Intercept	−1.000	−0.000	1.000	−0.499	−0.000	1.500	−1.501	−0.000	0.501
	(0.018)	(0.016)	(0.038)	(0.016)	(0.016)	(0.033)	(0.016)	(0.016)	(0.035)
D	0.999	1.004	1.002	1.001	1.004	1.001	2.002	1.004	2.001
	(0.041)	(0.032)	(0.043)	(0.036)	(0.032)	(0.037)	(0.036)	(0.032)	(0.040)

Note: Entries are mean coefficients across 500 replications. Parentheses below give their standard deviations.

Weighting To examine pooling, we first allow the intercept and treatment coefficient to vary across values of X_i:

$$Y_i = \alpha_{X_i} + \beta_{X_i} D_i + \varepsilon_{\text{str},i}, \qquad M_i = 0, \qquad E[\varepsilon_{\text{str},i} \mid D_i, X_i, M_i = 0] = 0.$$

Among complete cases with $X_i = x$, untreated units have mean outcome α_x, and treated units have mean outcome $\alpha_x + \beta_x$. Thus, β_x is their observed difference in means. Whether this difference identifies the causal effect τ depends on the missing mechanism. The residual $\varepsilon_{\text{str},i}$ captures individual deviations from these group means.

Under mechanisms A and B, $\beta_{-1} = \beta_0 = \beta_1 = \tau = 1$, as reflected in Table 1. The baseline means α_x differ: $(\alpha_{-1}, \alpha_0, \alpha_1) = (-1, 0, 1)$ under A and $(-0.5, 0, 1.5)$ under B. A linear X_i term captures α_x under A but not under B. Thus, the pooled complete-case regression on D_i and linear X_i consistently estimates τ under A but not under B, due to the misspecification of the control variables X_i under B.

Under mechanism B, selection into complete cases depends on U_i: at $X_i = \pm 1$, units with $U_i = 1$ are more likely to be complete than units with $U_i = -1$. The oracle regression error therefore

has a nonzero conditional mean among complete cases, $\mathbb{E}[U_i + \xi_i \mid X_i, D_i, M_i = 0] = X_i^2/2$. Since $Y_i(0) = X_i + U_i + \xi_i$, the mean untreated outcome becomes $\mathbb{E}[Y_i(0) \mid X_i, M_i = 0] = X_i + X_i^2/2$. The added $X_i^2/2$ cannot be represented by the intercept and linear X_i term in the fitted regression. Thus, correct oracle specification, linear X_i, does not ensure correct specification among complete cases.

Alternatively, IPW can estimate causal effects without specifying how X_i enters the outcome regression (Rosenbaum and Rubin 1983; Hirano, Imbens and Ridder 2003). Under mechanism B, the IPW estimator in Figure 1 weights each complete case by the inverse probability of its observed treatment given X_i and $M_i = 0$. Its estimates are centered near the complete-case ATE, which equals the constant population ATE[7].

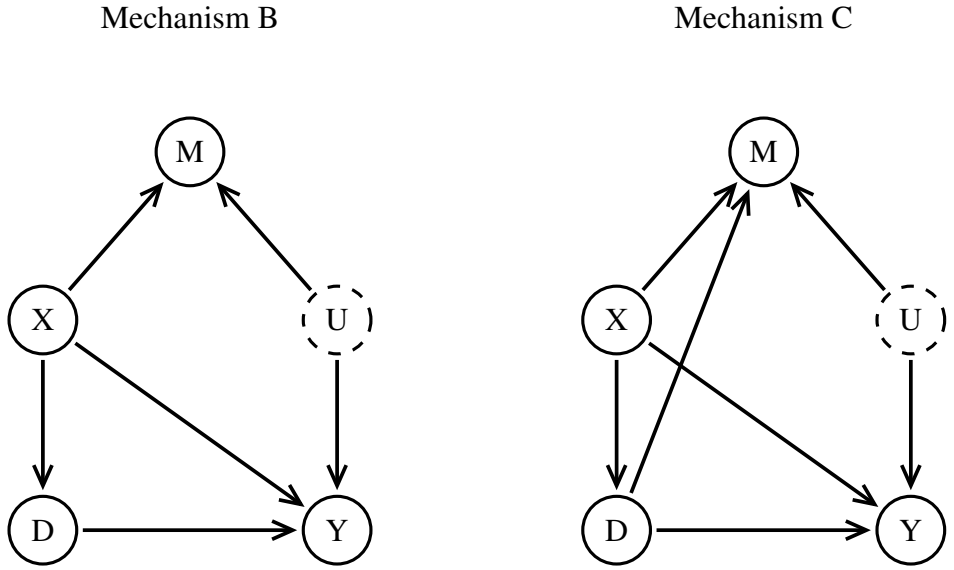


Figure 2: Causal Diagrams for Missing Mechanisms B and C

Back-door adjustment criterion We next compare mechanisms B and C. In Figure 1, the IPW complete-case estimates are centered near $\tau = 1$ under B but not C, although treatment probabilities are correctly specified within each X_i stratum under both mechanisms. The difference is that $\beta_x = \tau = 1$ in every X_i stratum under B, whereas β_{-1} and β_1 do not identify τ under C.

[7]Appendix B.1 explains how IPW and saturated regression weight the within-X comparisons differently under mechanisms B and C.

To assess why the within-stratum comparisons differ, we use Directed Acyclic Graphs (DAGs) from Pearl (1995), including the missingness indicator M_i as a node, to trace the paths that conditioning on complete-case status opens. Figure 2 presents the DAGs for mechanisms B and C.

Under B, conditioning on the collider M_i opens the noncausal path $D_i \leftarrow X_i \rightarrow M_i \leftarrow U_i \rightarrow Y_i$, but conditioning on X_i blocks it. Thus, $\{X_i, M_i\}$ satisfies the adjustment criterion, yielding $Y_i(d) \perp D_i \mid X_i, M_i$ for $d \in \{0,1\}$—the within-complete-case exchangeability condition emphasized by Choi, Dekkers and le Cessie (2020). The within-X_i treatment contrast among complete cases therefore identifies $\tau = 1$.

Under C, conditioning on $M_i = 0$ opens the noncausal path $D_i \rightarrow M_i \leftarrow U_i \rightarrow Y_i$, which conditioning on X_i does not block. Thus, $\{X_i, M_i\}$ fails the adjustment criterion's path-blocking requirement. In this simulation, the within-X_i treatment contrasts at $X_i = -1$ and $X_i = 1$ do not identify τ.

To sum up, all three mechanisms are NMAR and, by construction, share the same probability of missingness conditional on X_i, yet they differ in whether complete-case regression consistently estimates the population ATE, for two reasons. Under A, the within-X_i comparisons identify τ and the pooled linear regression recovers it. Under B, the comparisons still identify τ, and entering X_i categorically or using treatment IPW restores consistent estimation. Under C, the comparisons fail to identify τ in two strata, and neither approach recovers it. These differences motivate the formal results in Section 3, which establish sufficient causal and modeling conditions for complete case regressions to consistently estimate the population ATE.

3 Identification of Population ATE under Complete-Case Regression

In this section, we first prove that two conditions are sufficient for complete-case regression to consistently estimate the complete-case average treatment effect: (a) treatment comparisons within covariate strata identify the stratum-specific causal effects, and (b) the fitted estimator aggregates

those comparisons correctly. Under a constant unit-level treatment effect, the complete-case ATE equals the population ATE. With heterogeneous effects, recovering the population ATE requires a third condition: the conditional effects identified among complete cases must also transport to the population, and complete cases must cover the population covariate support. These effects can then be averaged over the population covariate distribution.

Our proof develops in four parts. First, we define the oracle regression benchmark and distinguish missing patterns from missing mechanisms. Second, we use a DAG with a missing indicator to assess within-covariate treatment comparisons among complete cases. Third, we examine when outcome specification or treatment-propensity weighting allows regression to aggregate those comparisons correctly. Finally, we connect the complete-case ATE to the population ATE. Section 5 develops reweighting under heterogeneous effects.

3.1 Oracle Regression and Missingness

We first distinguish the missing mechanism from the missing pattern. A missing pattern describes which variables are missing for a unit. A missing mechanism specifies how the probability of that pattern depends on the full-data values (Little and Rubin 2019). Missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR) classify missing mechanisms[8]. While frequently used, these classifications usually do not specify which variables are missing or, more importantly for our framework, what causes the missingness.

Missing Pattern and Missing Mechanism In our motivating simulation, mechanisms A, B, and C are all NMAR and share the same missing pattern: X_i is missing for some units. They differ, however, in how observation of X_i depends on X_i, U_i, and D_i. These differences determine whether complete-case regression consistently estimates τ and, if not, whether saturation or treatment IPW can help. The NMAR label alone therefore does not distinguish the assumptions that matter for

[8]Under MCAR, missingness probabilities are independent of data values. Under MAR, missingness may depend on variables whose values are observed, but not on the values that are missing under that pattern. NMAR permits dependence on those missing values.

these estimators[9].

Formally, for a unit i, let $X_i = (X_i^1, \ldots, X_i^p)$ collect the pretreatment covariates in the researcher's chosen adjustment set[10], and let $V_i = (Y_i, D_i, X_i^1, \ldots, X_i^p)$ additionally contain the treatment indicator D_i and the outcome Y_i.

Definition 1 (Missing pattern and omnibus missing indicator)**.** *For unit i, let M_i^Y, M_i^D, and $M_i^{X^j}$, for $j = 1, \ldots, p$, equal one when the corresponding value of Y_i, D_i, or X_i^j is missing, and zero otherwise. The vector*

$$\mathbf{M}_i = (M_i^Y, M_i^D, M_i^{X^1}, \ldots, M_i^{X^p})$$

is unit i's missing pattern. Complete-case regression uses only units for which every component of V_i is observed. To represent this restriction, we summarize the missing pattern with an omnibus indicator. Define

$$M_i = \mathbf{1}\left\{ M_i^Y + M_i^D + \sum_{v=1}^{p} M_i^{X^v} > 0 \right\}.$$

Thus $M_i = 0$ if and only if Y_i, D_i, and every component of X_i are observed[11].

If a researcher changes the adjustment set, the missing indicator and complete-case sample may also change, so the researcher must redefine the indicator and reassess the conditions for the revised specification. The target population ATE remains fixed across these specifications because it is defined over the population from which the oracle sample is drawn, as described below.

Appendix D.2 provides the conventional probabilistic definition of a missing mechanism (Little and Rubin 2019; Seaman et al. 2013). In the next subsection, we give the mechanism a causal formulation, building on graphical models of missing data (Mohan and Pearl 2021).

Our analysis from now on focuses on the missing mechanism: what causes a unit to enter the complete-case sample and how that selection affects treatment effect identification and estimation.

[9]The importance of specifying the missing mechanism precisely has also been raised by Seaman et al. (2013), who distinguish different formal definitions of MAR.

[10]Throughout, we take the chosen adjustment set to include any pretreatment treatment-effect moderators used in the analysis.

[11]The omnibus missing indicator corresponds to Wooldridge's (2002, p. 553) selection indicator with reversed coding: $M_i = 1 - s_i$. Both indicators are defined for all units in the underlying population, including those excluded from estimation. They can therefore represent listwise deletion due to missing values in the outcome, regressors, or both.

We represent selection into this sample using the omnibus missing indicator M_i, regardless of which variable makes a unit incomplete.

Oracle Samples To assess how selection into complete cases affects the regression workflow, we first define its benchmark before listwise deletion: the oracle sample, in which all required values are observed and for which the researcher chooses an adjustment set and regression specification. We distinguish two versions of these sampled units to connect the regression benchmark to the causal effect of interest.

The oracle potential-outcome sample contains both potential outcomes and lets us define the causal effect. The oracle sample contains each unit's realized outcome, treatment, and required covariates, including values missing in the observed data. We call the chosen regression applied to the oracle sample the oracle regression and ask whether fitting the same specification to complete cases recovers the population ATE that this adjustment set identifies.

We maintain Assumption A1 (SUTVA) to link potential outcomes to each unit's realized outcome, and denote the population distribution from which the oracle potential-outcome sample is drawn by P_{po}. The assumption and formal definitions of both oracle samples are in Appendix D.

Oracle Regression The target causal estimand is the population average treatment effect:

$$\tau \equiv \mathbb{E}_{P_{\text{po}}}[Y_i(1) - Y_i(0)].$$

We take the researcher's chosen oracle regression as our benchmark:

$$Y_i = \beta_{\text{or}} D_i + b(X_i)^\top \gamma_{\text{or}} + \varepsilon_{\text{or},i}. \tag{1}$$

Here $b(X_i)$ contains an intercept and the chosen covariate terms. β_{or} and γ_{or} are the population treatment and covariate coefficients, and $\varepsilon_{\text{or},i}$ is the residual. Their sample estimates are $\widehat{\beta}_{\text{or}}$ and $\widehat{\gamma}_{\text{or}}$.

We assume that the researcher's causal identification assumptions make X_i a valid adjustment

set, so that the population ATE would be identified if all required values were observed. Even with a valid adjustment set, however, the oracle regression coefficient β_{or} may not equal the population ATE. It equals τ under a constant treatment effect with correctly specified covariate terms, or under a constant treatment propensity, as in a randomized experiment. With heterogeneous effects, it is instead a variance-weighted average of conditional effects (Angrist and Pischke 2009; Aronow and Samii 2016). We therefore target τ directly and do not require $\beta_{\text{or}} = \tau$. Under standard OLS regularity conditions, the sample coefficient $\widehat{\beta}_{\text{or}}$ converges in probability to β_{or}. Although we focus on the population ATE, some results also apply to other causal estimands targeted by the oracle regression, as discussed after Proposition 3.

After listwise deletion, the researcher fits the same specification to complete cases:

$$Y_i = \beta_{\text{cc}} D_i + b(X_i)^\top \gamma_{\text{cc}} + \varepsilon_{\text{cc},i}, \qquad M_i = 0. \tag{2}$$

Here β_{cc} and γ_{cc} are the population regression coefficients among complete cases, and $\varepsilon_{\text{cc},i}$ is the corresponding residual. Under standard OLS regularity conditions, the sample treatment coefficient $\widehat{\beta}_{\text{cc}}$ converges in probability to β_{cc}.

To summarize the problem, our question is whether complete-case regression consistently estimates the population ATE identified by the oracle regression's adjustment set. When $\beta_{\text{or}} = \tau$, this requires

$$\beta_{\text{cc}} = \beta_{\text{or}}.$$

When β_{or} differs from τ, equality between the two regression coefficients would no longer be sufficient to recover the population ATE. In either case, we take the oracle adjustment set's validity as given and focus on the additional causal and modeling conditions needed for regression after listwise deletion to recover τ.

3.2 DAG for Within-X Complete-Case Comparisons

We first use a causal DAG to establish when treatment comparisons within covariate strata among complete cases identify their conditional treatment effects. The next subsection then examines when regression pools these comparisons correctly.

Consider complete cases with the same pre-treatment covariate values $X_i = x$, and let $\delta_{\text{cc}}(x)$ denote their average treatment effect. When both treatment arms have positive probability in this stratum, the population coefficient from regressing Y_i on an intercept and D_i equals the treated-minus-control difference in mean outcomes. We show when this difference identifies $\delta_{\text{cc}}(x)$.

We augment the causal DAG for the oracle regression variables Y_i, D_i, X_i with the omnibus missing indicator M_i and its causes. This recasts *selection on being observed*, the restriction that defines complete-case regression, as *selection on the observable indicator* M_i, specifically $M_i = 0$. As Figure 2 illustrates, this conditioning can open noncausal paths, where $\{X_i, M_i\}$ satisfies the adjustment criterion under mechanism B but not mechanism C.

Adjustment criterion To assess whether $\{X_i, M_i\}$ is a valid adjustment set, we use the criterion of Shpitser, VanderWeele and Robins (2010, Definition 5), a generalization of the back-door criterion of Pearl (1995) (see also Cinelli, Forney and Pearl 2024), stated here for a single treatment D and outcome Y.

Definition 2 (Causal missingness mechanism for listwise deletion)**.** *Let G be a causal DAG containing the oracle variables Y_i, D_i, X_i, the omnibus missing indicator M_i, and any other relevant observed or latent causes. The causal missingness mechanism for listwise deletion consists of the causal relationships leading to M_i and the process determining it, while G represents its causal structure.*

Definition 3 (Adjustment criterion for a single treatment)**.** *Let G be a causal DAG with treatment D, outcome Y, and a set $\mathscr{Z}$ of variables disjoint from $\{D, Y\}$. The set $\mathscr{Z}$ satisfies the adjustment criterion if:*

1. *$\mathscr{Z}$ contains no node other than D on a directed path from D to Y, nor any descendant of such a node*[12].

2. *Conditioning on $\mathscr{Z}$ blocks every noncausal path from D to Y*[13].

A path is blocked given $\mathscr{Z}$ if it contains a noncollider in $\mathscr{Z}$, or a collider with neither itself nor any descendant in $\mathscr{Z}$ (Shpitser, VanderWeele and Robins 2010, Section 2). Applying the criterion to $\mathscr{Z} = \{X_i, M_i\}$ gives the following result.

Proposition 1 (Adjustment criterion, conditional ignorability, and within-X complete-case comparison)**.** *Write $\delta_{\text{cc}}(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, M_i = 0]$. If $\{X_i, M_i\}$ satisfies the adjustment criterion relative to (D_i, Y_i) in G, then:*

(i) in every causal model inducing G, for each $d \in \{0,1\}$,

$$Y_i(d) \perp D_i \mid X_i, M_i;$$

(ii) for any complete-case stratum $X_i = x$ with treatment overlap, the population coefficient on D_i from regressing Y_i on an intercept and D_i equals

$$\mathbb{E}[Y_i \mid D_i = 1, X_i = x, M_i = 0] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x, M_i = 0] = \delta_{\text{cc}}(x).$$

With a constant unit-level treatment effect $Y_i(1) - Y_i(0) = \tau$ for every unit, $\delta_{\text{cc}}(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, M_i = 0] = \tau$.

Proof in Appendix D.

Proposition 1 shows that $\delta_{\text{cc}}(x)$ is identified when the covariates and missing indicator satisfy the adjustment criterion and both treatment arms are represented within each complete-case covariate stratum, which provides the foundation for aggregating the conditional effects later.

[12] A node counts as its own descendant. Unlike the back-door criterion, the adjustment criterion can permit a descendant of D on a branch that does not lead to Y, provided all noncausal paths are blocked. See Appendix D for the general forms.

[13] A noncausal path is one that is not directed entirely from D to Y (Shpitser, VanderWeele and Robins 2010, Section 2)

Although the criterion is formal, checking it adds little to the diagram work that already supports the oracle regression. A researcher who uses a DAG to choose its adjustment set can add the missing indicator M_i and its causes, including relevant unobserved ones, to that same diagram and ask whether the original set still works among complete cases.

3.3 Pooling Within-X Comparisons into Regression

As mechanism B in Section 2 illustrates, valid within-covariate treatment comparisons do not ensure that the pooled regression consistently estimates the complete-case ATE. The following two propositions give sufficient conditions under which unweighted complete-case regression does so.

Proposition 2 (Correct specification and the complete-case ATE)**.** *Suppose $\{X_i, M_i\}$ satisfies the adjustment criterion and complete-case treatment overlap holds (Assumption A2). Consider the complete-case regression*

$$Y_i = g(X_i)^\top \alpha_{\text{cc}} + D_i h(X_i)^\top \beta_{\text{cc}} + \varepsilon_{\text{cc},i}, \qquad M_i = 0,$$

where $g(X_i)$ contains an intercept and the chosen baseline covariate terms, and $h(X_i)$ contains an intercept and the chosen terms interacted with treatment. Suppose this regression is correctly specified in conditional mean:

$$\mathbb{E}[\varepsilon_{\text{cc},i} \mid D_i, X_i, M_i = 0] = 0.$$

Then, under SUTVA (Assumption A1),

$$\delta_{\text{cc}}(x) = h(x)^\top \beta_{\text{cc}},$$

and averaging over the complete-case covariate distribution gives the complete-case ATE, τ_{cc}:

$$\tau_{\text{cc}} \equiv \mathbb{E}[h(X_i)^\top \beta_{\text{cc}} \mid M_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0].$$

Let $\widehat{\beta}_{\text{cc}}$ be the complete-case OLS estimate and n_{cc} the number of complete cases. Under standard regularity conditions (Assumptions A3 and A4 in Appendix D),

$$\widehat{\tau}_{\text{cc}} \equiv \frac{1}{n_{\text{cc}}} \sum_{i:M_i=0} h(X_i)^\top \widehat{\beta}_{\text{cc}} \xrightarrow{p} \tau_{\text{cc}}.$$

Proof in Appendix D.3.

When $h(X_i) = 1$, there are no treatment–covariate interactions, β_{cc} is a scalar, and the conditional treatment effect is the same for every X_i, so the coefficient on D_i is that common effect. Equation (2) is this no-interaction case, and Proposition 2 generalizes it by allowing β_{cc} to be a vector. When interactions are included, $h(X_i)$ varies with the covariates, the coefficient on D_i alone describes only the reference group, and the complete-case ATE must be recovered by averaging the fitted conditional effects over the covariate distribution observed among complete cases.

Proposition 2 identifies two requirements for consistently estimating the complete-case ATE, τ_{cc}: valid within-covariate comparisons and their correct aggregation. The adjustment criterion, together with treatment overlap, makes those comparisons causally valid. Correct specification of the complete-case outcome regression makes the aggregation valid[14]: it allows the fitted conditional treatment effects to be averaged over complete cases, and with the stated regularity conditions that average is consistent for τ_{cc}.

However, correct specification after listwise deletion may be harder to establish than researchers expect. A regression correctly specified for the oracle sample may be misspecified after deletion because selection can change how the mean outcome depends on the covariates (Heckman 1979). Section 2's mechanism B illustrates this: a linear term for X_i is sufficient in the oracle regression, but the complete-case regression must capture a separate baseline mean for each X_i stratum[15]. Thus, assessing whether the specification remains adequate after deletion may require additional

[14]Proposition 2 generalizes the specification and selection argument of Arel-Bundock and Pelc (2018): their indicator–error independence condition, together with correct oracle specification, implies the complete-case zero conditional mean condition required here. Our graphical condition additionally establishes the causal interpretation as the complete-case ATE, which their coefficient-preservation result does not address. See Appendix D.4 for a discussion.

[15]A more flexible outcome model may help accommodate changes in the conditional mean. Developing diagnostics for the complete-case specification is a direction for future work.

assumptions about unobserved outcome determinants and their role in missingness, even when the oracle regression is correctly specified.

Treatment-propensity weighting, not IPW for missingness[16], offers another way to pool valid within-covariate comparisons without requiring a correctly specified outcome equation (Rosenbaum and Rubin 1983; Hirano, Imbens and Ridder 2003). Appendix Proposition D.1 states the standard IPW result for complete cases (Hirano, Imbens and Ridder 2003; Choi, Dekkers and le Cessie 2019). Here, we consider its use in a complete-case regression.

Proposition 3 (Graphical criterion for complete-case IPW regression)**.** *Suppose* $\{X_i, M_i\}$ *satisfies the adjustment criterion and complete-case treatment overlap holds (Assumption A2).*

Among complete cases, consider the IPW-weighted working regression

$$Y_i = \alpha + \beta D_i + X_i^\top \gamma + \varepsilon_{\text{ipw},i}, \qquad \omega_i = \frac{D_i}{e_0(X_i)} + \frac{1 - D_i}{1 - e_0(X_i)},$$

where $e_0(x) = P(D_i = 1 \mid X_i = x, M_i = 0)$*. Let* β^* *denote the population counterpart of the IPW-weighted regression coefficient on* D_i *among complete cases. Then, under SUTVA and standard regularity conditions (Assumptions A1, A4, and A6 in Appendix D),*

$$\beta^* = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0].$$

If the propensity score is estimated consistently, in the sense that the estimated inverse weights converge uniformly (Assumption A5), the sample coefficient $\widehat{\beta}$*, computed using* $\widehat{e}_0$*, converges in probability to* β^**.*

Proposition 3 follows from Proposition 1 and Proposition D.2, which applies Gabriel et al. (2024) to complete cases.

Among complete cases, the weights make each treatment arm reproduce the complete-case covariate distribution and set the weighted treatment probability to one-half within every supported

[16]IPW for missingness weights complete cases by the inverse probability of being a complete case to address selection due to missingness (Seaman and White 2013; Hughes et al. 2019). Here, we weight by the inverse treatment probability among complete cases to pool their within-covariate treatment comparisons.

stratum $X_i = x$ (Hernán and Robins 2020, Section 2.4). The weighted regression therefore pools the valid within-X comparisons without the correctly specified complete-case outcome mean that Proposition 2 requires, relying instead on the estimated propensity weights.

In practice, weighting adds one input, a propensity model for treatment among complete cases. The weighted comparisons are most reliable where the data have treatment overlap, and the within-X comparisons still require the adjustment criterion to hold in the augmented DAG. As the motivating simulation shows, weighting addresses the pooling problem under mechanism B but cannot repair the invalid comparisons under mechanism C.

Corollary 1 (Constant complete-case propensity: unweighted OLS equals IPW)**.** *Suppose the conditions of Proposition 3 hold, and additionally there exists $\bar{e}_0 \in (0,1)$ such that $e_0(x) = \bar{e}_0$ for every x in the complete-case support. Then, for any chosen $g(X_i)$ containing an intercept, the unweighted additive complete-case regression*

$$Y_i = g(X_i)^\top \alpha_{\text{cc}} + \beta_{\text{cc}} D_i + \varepsilon_{\text{cc},i}, \qquad M_i = 0,$$

has $\beta_{\text{cc}} = \tau_{\text{cc}}$ regardless of whether $g(X_i)^\top \alpha_{\text{cc}}$ correctly specifies the complete-case baseline mean, and, under Assumption A3, $\widehat{\beta}_{\text{cc}} \xrightarrow{p} \tau_{\text{cc}}$: unweighted OLS coincides with IPW whenever every complete-case unit shares the same treatment propensity.

Proof in Appendix D.6.

A constant complete-case propensity makes D_i exactly mean-independent of any function of X_i among complete cases, so adding, omitting, or misspecifying baseline controls does not affect the consistency of the treatment coefficient (Lin 2013). Unlike Proposition 2, this route does not require correctly specifying the shape of treatment-effect heterogeneity: the additive specification $h(X_i) = 1$ recovers the complete-case ATE regardless of how $\delta_{\text{cc}}(x)$ varies with x.

Finally, we show in Appendix D.3 that Propositions 1–3 still hold when treatment itself affects missingness, for example when treatment induces attrition.

Connection to the Oracle Target Both regression approaches consistently estimate the complete-case ATE under their respective assumptions[17]. To recover the population ATE, we also need to establish that the complete-case ATE equals the population ATE. The next subsection gives conditions for this connection under constant and heterogeneous treatment effects.

3.4 When Complete-Case Regression Recovers the Population ATE

Under a constant unit-level treatment effect, the connection is immediate: $\tau_{\text{cc}} = \tau$. Either regression approach therefore consistently estimates the population ATE under its respective assumptions.

Corollary 2 (Constant unit-level treatment effect)**.** *Suppose there is a constant treatment effect for every unit,* $Y_i(1) - Y_i(0) = \tau$*, and* $P(M_i = 0) > 0$*. Then*

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0)] = \tau.$$

Under the conditions of Proposition 2, $\tau_{\text{cc}} = \tau$*, and* $\widehat{\tau}_{\text{cc}}$ *converges in probability to* τ*.*

Under the conditions of Proposition 3, $\beta^* = \tau$*. Under Assumption A5 (estimated inverse-weight consistency),* $\widehat{\beta}$ *also converges in probability to* τ*.*

Under the conditions of Corollary 1, $\beta_{\text{cc}} = \tau$ *as well, and* $\widehat{\beta}_{\text{cc}}$ *converges in probability to* τ*.*

With heterogeneous effects, however, τ_{cc} may not equal the population ATE, τ. Recovering the latter requires connecting the effects among complete cases to those in the population.

To make that connection, we assume *mean exchangeability of selection and treatment effect heterogeneity*: at every value of the covariates used for transport, the average treatment effect is the same among complete cases and in the population, though effects may still vary across those covariates. The condition is more plausible when the observed covariates capture characteristics that influence both completeness and responses to treatment, and it may fail when an unobserved

[17]The researcher may instead target the average treatment effect on the treated (ATT). Proposition 2 extends to the complete-case ATT by averaging fitted effects over treated complete cases. Treatment IPW for the ATT requires different weights.

characteristic does (Egami and Hartman 2023; Chen and Huang 2026). This condition also relaxes the outcome-response ignorability assumption of Huang and Liu (2026)[18].

Corollary 3 (Heterogeneous treatment effects)**.** *Suppose the conditions of Proposition 1 hold with adjustment covariates X_i. Let W_i be a subvector of X_i, possibly X_i itself, and let F_{po}^W denote its population distribution. Suppose $\mathbb{E}[|Y_i(d)|] < \infty$ for $d \in \{0,1\}$. Further, suppose* mean exchangeability of selection and treatment effect heterogeneity given *W_i holds: for almost every w with respect to F_{po}^W,*

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid W_i = w, M_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) \mid W_i = w].$$

Suppose also that $P(M_i = 0 \mid W_i = w) > 0$ for almost every such w. Define

$$\tau_{\text{cc}}(w) = \mathbb{E}[Y_i(1) - Y_i(0) \mid W_i = w, M_i = 0].$$

Then

$$\tau = \int \tau_{\text{cc}}(w)\, dF_{\text{po}}^W(w).$$

Under Assumptions A1 (SUTVA) and A2 (complete-case treatment overlap),

$$\tau_{\text{cc}}(w) = \int \delta_{\text{cc}}(x)\, dF_{\text{cc}}(x \mid W_i = w),$$

where $F_{\text{cc}}(x \mid W_i = w)$ is the distribution of X_i among complete cases with $W_i = w$, and

$$\delta_{\text{cc}}(x) = \mathbb{E}[Y_i \mid D_i = 1, X_i = x, M_i = 0] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x, M_i = 0].$$

Proof in Appendix D.8.

Applying Corollary 3 asks researchers for one new judgment. The first requirement, valid within-X_i comparisons, is the adjustment-criterion check already made for Proposition 1, and

[18]For full-covariate transport, given valid within-X_i treatment comparisons in both the population and complete cases, outcome-response ignorability (Yang, Wang and Ding 2019; Huang and Liu 2026) is sufficient for mean exchangeability but not necessary. For example, selection on an unobserved characteristic that raises both potential outcomes equally changes their conditional means while preserving the conditional average treatment effect.

positive complete-case probability across the support of W_i is an overlap question the data can answer. The new judgment is mean exchangeability given W_i, which data alone cannot establish. With all three in hand, researchers recover τ by averaging $\tau_{\text{cc}}(w)$ over F_{po}^{W}, as Section 5 discusses. Further, the covariates used for adjustment X_i can be different from the covariate used for transport W_i. If transport covariates contain missingness, imputation as in Huang and Liu (2026) may recover their target distribution under additional assumptions.

Practical Advice To sum up, empirical researchers can assess whether their complete-case regression recovers the population ATE by checking three boxes, as in Table 2. Box 1 asks whether within-covariate treatment comparisons remain valid after listwise deletion. Researchers who already use a DAG to justify their adjustment set need to add the missing indicator and its causes, then reassess the comparisons. Box 2 asks whether the regression appropriately aggregates those comparisons, either through a correctly specified complete-case outcome model or treatment-propensity weighting. If the target is the average treatment effect among all units with complete data (τ_{cc}), Boxes 1 and 2 suffice for the regression-based estimator $\widehat{\tau}_{\text{cc}}$[19] to consistently estimate it. Box 3 asks whether the complete-case effects connect to the population target. Under a constant unit-level treatment effect, the complete-case ATE equals the population ATE. With heterogeneous effects, researchers also need to assess the transport and support conditions and average the conditional effects over the population covariate distribution (Section 5). The next section shows how this framework incorporates and extends conventional guidance on complete-case regression.

[19]With treatment IPW, $\widehat{\tau}_{\text{cc}}$ is the treatment coefficient from the weighted regression. With a correctly specified outcome regression, it is the average fitted treatment effect among complete cases, $\widehat{\tau}_{\text{cc}} = n_{\text{cc}}^{-1} \sum_{i:M_i=0} h(X_i)^{\top} \widehat{\beta}_{\text{cc}}$. In the simplest case, $h(X_i) = 1$, this reduces to the treatment coefficient $\widehat{\beta}_{\text{cc}}$.

Table 2: From Oracle Regression to Complete-Case Estimation: Conditions and Checks

Step	What researchers should do	Result
Starting point: the oracle regression	Choose a regression specification whose adjustment set identifies the causal target if all required values were observed. Apply that specification among complete cases[1].	Subsection 3.1
Box 1: Valid comparisons	Augment the original DAG with the omnibus missing indicator and its causes. Check the adjustment criterion and treatment overlap among complete cases. These checks establish valid within-stratum treatment comparisons.	Proposition 1
Box 2: OLS pooling (correct specification)	Correctly specify the complete-case conditional outcome mean. For each complete case, calculate the difference between its fitted outcomes under treatment and control, then average these differences over complete cases to estimate the complete-case ATE.	Proposition 2
Alternative Box 2: IPW pooling	Estimate treatment propensity scores among complete cases and construct consistent inverse weights. IPW regression consistently estimates the complete-case ATE.	Proposition 3
Alternative Box 2: Constant propensity	If the complete-case treatment propensity is constant (for example, a randomized experiment with constant assignment probability and response independent of treatment within covariate strata), regress the outcome on treatment and any chosen baseline covariates. This regression consistently estimates the complete-case ATE.	Corollary 1
Boxes 1 and 2 only	If the target is the ATE among all units with complete data (τ_{cc}), Boxes 1 and 2 suffice for consistent estimation.	
Box 3 under a constant unit-level treatment effect	A constant unit-level treatment effect makes the complete-case effect equal the population ATE.	Corollary 2
Box 3 under heterogeneous effects	Assess mean exchangeability of selection and treatment effect heterogeneity and complete-case support. Section 5 develops how to average conditional effects over the population covariate distribution.	Corollary 3

Notes: Boxes 1 and 2 apply under both constant and heterogeneous effects. Regularity conditions are presumed throughout.

[1] The framework may extend to other oracle estimands, such as the LATE identified by an instrumental-variable regression. Such an extension requires assumptions ensuring instrument validity after selection and connecting the complete-case LATE to the oracle LATE.

4 Revisiting Conventional Guidance on Complete-Case Regression

In this Section, we first show how our framework recovers conventional guidance for missingness driven by regression covariates or the outcome. We follow Arel-Bundock and Pelc (2018) in assuming correct oracle specification and their indicator–error independence condition, which together satisfy the specification requirement of Proposition 2. Our framework thus recovers their recommendations while allowing weaker specification and independence restrictions. We then use the collider–mediator dilemma to show why the justification for complete-case regression depends on the causal graph and the effect of interest.

4.1 Recovering Familiar Recommendations

To begin, we maintain treatment overlap, standard regularity conditions, and a constant conditional ATE (Appendix D.4), so that valid complete-case treatment comparisons recover the population ATE through the correctly specified regression. The remaining question is whether the adjustment criterion holds after adding the missing indicator and its causes to the oracle DAG. The same graphical assessment still applies under heterogeneous effects (Subsection 3.4 and Section 5), and when researchers use the IPW-regression route of Proposition 3.

Table 3 reviews recommendations on complete-case analysis (CCA), also known as listwise deletion, and multiple imputation, adapted from Pepinsky (2018). The NMAR(X) row, for example, records that complete-case regression remains consistent when missingness depends on the missing regressor's value (Arel-Bundock and Pelc 2018). We use our augmented DAG to justify the whole CCA column and recover these recommendations.

Under MCAR, the missing indicator M_i is independent of the oracle outcome, treatment, and covariates. We represent it as an isolated node in the DAG, so adding M_i preserves identification by the original adjustment set.

Table 3: Conventional Guidance on Complete-Case Analysis and Multiple Imputation

Pattern	Mechanism	CCA	Imputation
X/Y	MCAR	✓	✓
X/Y	MAR(X)	✓	✓
X	MAR(X,Y)	×	✓
X	NMAR(X)	✓	×
X/Y	NMAR(X,Y)	×	×

Notes: Adapted from Pepinsky (2018). X includes all regressors, including treatment D. Pattern indicates which variables contain missing values; parentheses indicate the drivers of missingness. A check indicates consistency under the stated model assumptions; a cross indicates no such guarantee. For CCA, the checked rows rely on the specification and indicator–error independence conditions of Arel-Bundock and Pelc (2018).

Missingness driven by treatment and covariates Figure 3 shows a causal structure in which missingness depends only on treatment D_i and the confounder vector X_i. The same structure represents three missingness patterns, each with a different variable missing and the others observed, and with different MAR/NMAR classifications:

1. **Missing outcome:** Y_i contains missing values, but its missingness depends only on the observed D_i, X_i. This is MAR and corresponds to the second row of Table 3.

2. **Missing treatment:** D_i contains missing values, and its own value affects missingness. This is NMAR and corresponds to the fourth row, where the table's X includes treatment.

3. **Missing covariates:** Some covariates in X_i contain missing values, and missingness depends on those covariates' own values. This is also NMAR and corresponds to the fourth row.

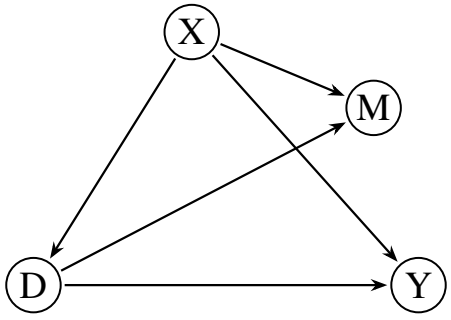


Figure 3: Selection Driven by Treatment and Covariates

In the original DAG without M_i, adjusting for X_i blocks the noncausal path $D_i \leftarrow X_i \rightarrow Y_i$, identifying the population ATE from the oracle sample. Restricting to complete cases additionally conditions on the collider $D_i \rightarrow M_i \leftarrow X_i$, opening the path $D_i \rightarrow M_i \leftarrow X_i \rightarrow Y_i$, which remains blocked by adjustment for X_i. Moreover, M_i is on a separate branch from the causal path $D_i \rightarrow Y_i$. Thus, $\{X_i, M_i\}$ satisfies the adjustment criterion. Under the maintained assumptions, complete-case regression thus consistently estimates the population ATE in all three cases, recovering the recommendation that missingness driven by regressors need not invalidate regression estimates, whether the mechanism is MAR or NMAR (Arel-Bundock and Pelc 2018).

Missingness driven by the outcome Figure 4 retains the same treatment–outcome relationship and confounder, but missingness now instead depends only on Y_i. This DAG likewise represents three cases, each with a different variable missing and the others observed, and with different MAR/NMAR classifications:

1. **Missing outcome:** Y_i contains missing values, and missingness depends on its own value. This is NMAR and illustrates the outcome-driven selection covered by the fifth row of Table 3.

2. **Missing treatment:** D_i contains missing values, but missingness depends only on the observed Y_i. This is MAR and corresponds to the third row.

3. **Missing covariates:** X_i contains missing values, but missingness depends only on the observed Y_i. This is also MAR and corresponds to the third row.

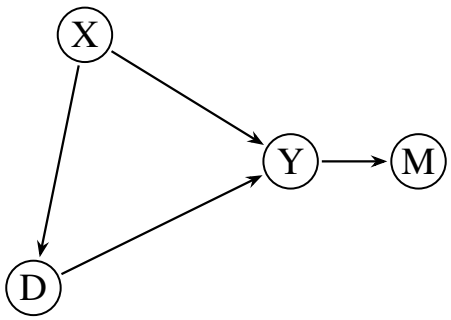


Figure 4: Selection Driven by the Outcome

Restricting to complete cases conditions on M_i, a descendant of Y_i on the causal path $D_i \to Y_i$. Thus, $\{X_i, M_i\}$ violates the first condition of the adjustment criterion, and complete-case regression is no longer guaranteed to consistently estimate the population ATE. This recovers the warning that outcome-driven missingness can invalidate complete-case regression (Heckman 1979), even under MAR (Arel-Bundock and Pelc 2018)[20].

Together, the two DAGs show that whether complete-case regression is justified depends on the missingness mechanism in the augmented DAG, rather than the missingness pattern or its MAR/NMAR classification.

4.2 The Collider–Mediator Dilemma

The preceding examples show how our framework recovers existing recommendations for complete-case regression. We now use it to examine a setting for which those recommendations provide no clear guidance: a variable that explains missingness and also mediates the treatment effect.

Rivlin-Angert, Yakter and Sheffer (2025) study the relationship between the Big Five personality traits and political attitudes using a survey of Jewish Israelis. Their original dataset contains 925 respondents. They use listwise deletion for all regressions, leaving 816 complete cases for specifications with sociodemographic controls[21]. Our illustration draws on their analysis of conscientiousness and support for a two-state solution, specifically the mediating role of authoritarian tendency.

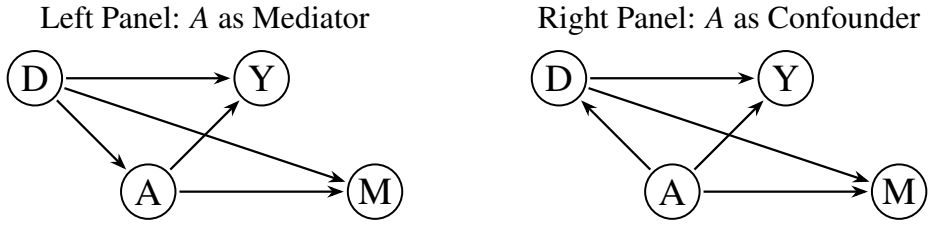
The authors do not discuss the missingness mechanism underlying their data, so we retain the 803 respondents with complete observations on all variables used in this illustration and impose additional missingness. We assume that respondents with below-average conscientiousness and above-average authoritarian tendencies have a 50% probability of having their outcome value unreported[22]. The missingness mechanism is therefore known in our synthetic data. In practice,

[20] A notable exception is a logistic outcome model: if selection depends only on the outcome, the treatment log odds ratio is preserved (Prentice and Pyke 1979), but risk differences, and thus the ATE, are not.

[21] The OLS regressions include all five personality traits, with additional specifications controlling for gender, age group, education, and religious self-identification.

[22] Research on social desirability shows that respondents may conceal politically sensitive preferences, while research

researchers must enter that mechanism themselves into their augmented DAG. The left panel of Figure 5 augments the paper's original mediation structure with the missing indicator and its assumed causes[23].



A: authoritarian tendency
D: conscientiousness
Y: support for a two-state solution
M: missing indicator for Y, with $M = 1$ when the response is missing

Figure 5: Authoritarian Tendency as a Mediator or Confounder

Why the mediator creates a dilemma While the original study distinguishes average direct and causal mediation effects, our illustration focuses on the total effect of conscientiousness on support for a two-state solution[24]. Under the assumed causal structure and correct regression specification, regressing Y_i on D_i, with the baseline controls but without the mediator A_i, consistently estimates this total effect.

Analyzing the dataset after listwise deletion conditions on $M_i = 0$, opening the noncausal path $D_i \rightarrow M_i \leftarrow A_i \rightarrow Y_i$, on which M_i is a collider. Moreover, M_i is a descendant of the mediator A_i on the causal path $D_i \rightarrow A_i \rightarrow Y_i$, so conditioning on it violates the first condition of the adjustment criterion. Thus, in contrast to the case of the oracle sample, a regression of Y_i on D_i can fail to consistently estimate the total effect.

The guidance that listwise deletion preserves the oracle regression coefficients when missingness is driven by included regressors (Arel-Bundock and Pelc 2018) may tempt researchers to add A_i to the regression. Indeed, adjusting for A_i blocks the opened noncausal path, but it also blocks the

on survey effort links conscientiousness to careless responding (Carmines and Nassar 2021; Zamarro et al. 2018).

[23]The other personality traits and sociodemographic controls are omitted from the diagram.

[24]The authors describe their analysis as "causal mediation analysis," while acknowledging its suggestive nature (Rivlin-Angert, Yakter and Sheffer 2025, p. 404), so we use it here as an illustration.

mediated causal path. Thus, adding the mediator to the complete-case regression does not restore identification of the total effect. The treatment coefficient, instead, identifies the controlled direct effect, which equals the average direct effect when the specification has no treatment–mediator interaction, under sequential ignorability (Imai, Keele and Tingley 2010).

Illustration with Synthetic Data To demonstrate this, we draw 10,000 bootstrap samples of 803 respondents each, sampling with replacement from the complete-case dataset. In each sample, we impose the specified missing mechanism using the sample means of conscientiousness and authoritarian tendency as the thresholds. Specifically,

$$\Pr(M_i = 1 \mid D_i, A_i) = \begin{cases} 0.5, & D_i < \overline{D} \text{ and } A_i > \overline{A}, \\ 0, & \text{otherwise.} \end{cases}$$

On average, 11.5 percent of outcomes are missing. In each bootstrap sample, we estimate the following models, all of which include the authors' sociodemographic controls.

- Oracle estimator: regress Y_i on D_i, without the mediator A_i, on the oracle dataset before imposing missingness.
- Oracle with A_i (Oracle + A): regress Y_i on D_i and A_i on the oracle dataset.
- Complete-case estimator (CCA): regress Y_i on D_i, without A_i, on the complete-case dataset where $M_i = 0$.
- Complete-case estimator with A_i (CCA + A): regress Y_i on D_i and A_i after listwise deletion.
- Multiple-imputation estimator (MI): impute missing Y_i using D_i, A_i, and the baseline controls, then regress Y_i on D_i, without A_i, on the imputed datasets and pool the estimates (Little and Rubin 2019).

Figure 6 summarizes the empirical distributions of these estimators. Neither CCA nor CCA + A

is centered around the oracle total-effect benchmark of -0.560[25]. Instead, CCA + A is concentrated near the oracle controlled direct-effect benchmark of -0.305. These results are consistent with the predictions from the augmented DAG.

We also observe that the multiple-imputation estimates are centered near the oracle total-effect benchmark of -0.560. This is consistent with the imposed MAR mechanism: missingness depends only on observed conscientiousness and authoritarian tendency. Multiple imputation enables researchers to include both variables and the baseline controls in the imputation model while keeping the mediator out of the total-effect regression[26].

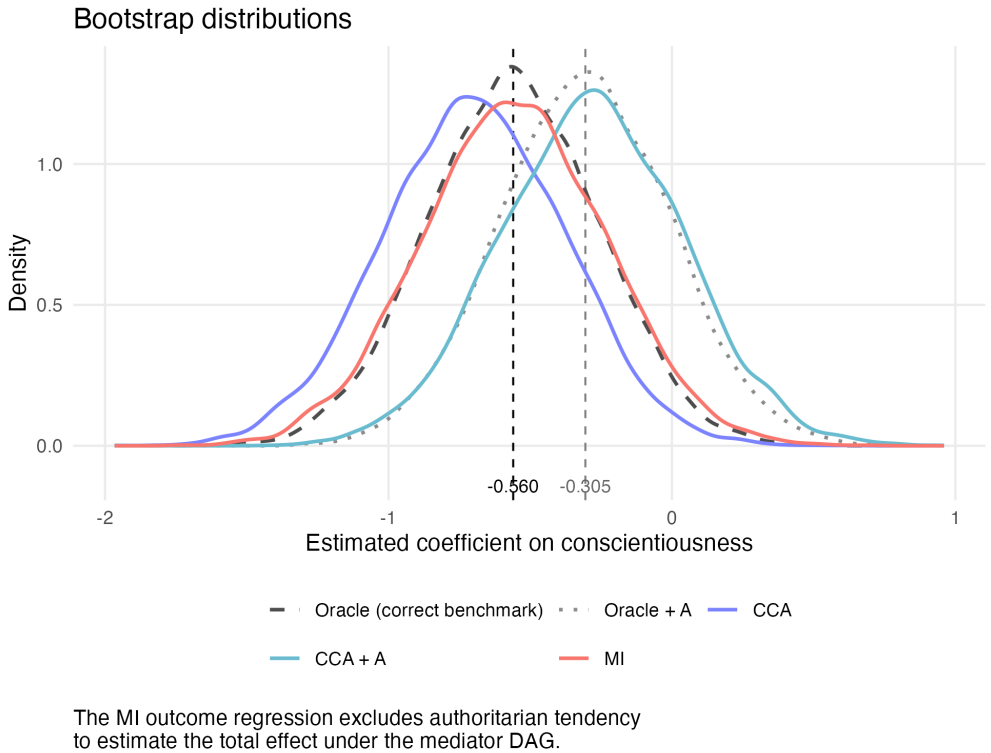


Figure 6: Bootstrap Illustration with Authoritarianism as a Mediator

[25]Our oracle regressions yield total- and controlled-direct-effect estimates of -0.560 and -0.305, compared with -0.564 and -0.310 in Rivlin-Angert, Yakter and Sheffer (2025, Table 4). The small differences arise because they report averages over 1,000 quasi-Bayesian draws, while we report OLS point estimates.

[26]G-computation could instead recover the total effect under suitable assumptions by combining a complete-case outcome model with the mediator distribution conditional on treatment and baseline controls, estimated from the original sample.

When authoritarian tendency is a confounder We now consider an alternative substantive theory in which authoritarian tendency is a confounder of the relationship between conscientiousness and support for a two-state solution. Political science research sometimes treats authoritarianism as a relatively enduring predisposition that shapes political attitudes (Stenner 2005; Engelhardt, Feldman and Hetherington 2023). The right panel of Figure 5 represents this alternative causal structure, retaining the same missing mechanism.

The missingness mechanism is MAR in both left and right DAGs, but the causal role of A_i differs, as shown by the reversed arrow between D_i and A_i. When A_i is a confounder, conditioning on M_i opens the noncausal path $D_i \rightarrow M_i \leftarrow A_i \rightarrow Y_i$. Adjusting for A_i blocks both this path and the original back-door path $D_i \leftarrow A_i \rightarrow Y_i$. Under the maintained identification and regression-specification assumptions, complete-case regression adjusting for A_i therefore consistently estimates the total effect. Consistent with the augmented DAG, Figure 7 shows that CCA + A is centered near the total-effect benchmark of -0.305[27].

In this illustration, we show that the same dataset, missingness pattern, and MAR classification may call for different strategies under different causal structures. Adjusting for authoritarian tendency identifies the total effect when it is a confounder, but blocks part of that effect when it is a mediator. Different causal structures can generate observationally equivalent data, so researchers must specify their causal assumptions to assess whether complete-case analysis remains justified after conditioning on the missing indicator. Appendix E further supplies a separate simulation of the dilemma, including missing-outcome and missing-treatment cases.

[27]The oracle regression adjusting for authoritarian tendency yields the same coefficient, -0.305, as in the mediation illustration, but it identifies the total effect under the confounder DAG. CCA + A averages -0.285 because, in the finite sample used for the bootstrap, respondents subject to missingness happen to lie slightly above the fitted line – deleting some flattens the slope. This offset would vanish as the original sample grows if the population conditional mean of the outcome is linear in authoritarian tendency.

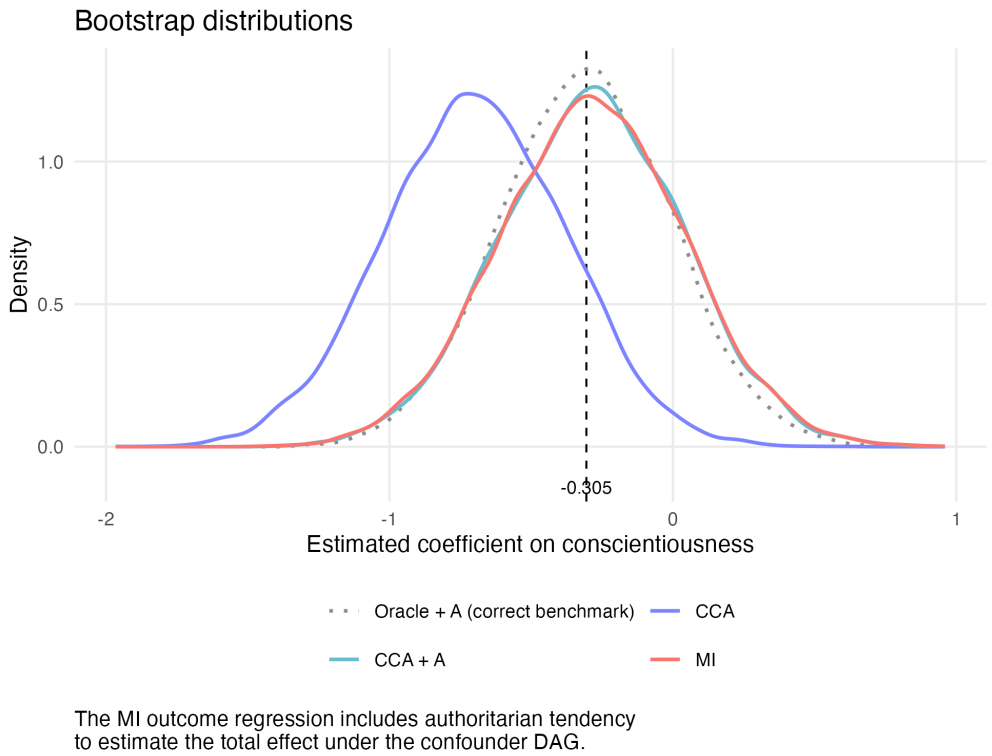


Figure 7: Bootstrap Illustration with Authoritarianism as a Confounder

5 Reweighting Complete-Case Comparisons: Heterogeneous Effects

In this section, we extend the framework to heterogeneous treatment effects by reweighting complete-case comparisons to recover population averages. Appendix F further applies the same approach to generalized linear models.

5.1 Reweighting for Heterogeneous Treatment Effects

Let W_i denote the covariates used for transport in Corollary 3. Let $\delta_{\text{cc}}(w)$ and $\delta_{\text{po}}(w)$ denote the conditional average treatment effects among complete cases and in the population, respectively. Let F_{cc} and F_{po} denote the distributions of W_i among complete cases and in the population. Then

$$\tau_{\text{cc}} = \int \delta_{\text{cc}}(w)\, dF_{\text{cc}}(w), \qquad \tau = \int \delta_{\text{po}}(w)\, dF_{\text{po}}(w).$$

Under mean exchangeability of selection and treatment effect heterogeneity, $\delta_{\text{cc}}(w) = \delta_{\text{po}}(w)$ for almost every w with respect to F_{po}. With complete-case support, we therefore have

$$\tau_{\text{cc}} - \tau = \int \delta_{\text{cc}}(w)\, dF_{\text{cc}}(w) - \int \delta_{\text{cc}}(w)\, dF_{\text{po}}(w).$$

Under these conditions, the difference arises from averaging the same conditional treatment effects over different covariate distributions. This decomposition also situates the two-stage projection estimator of Huang and Liu (2026). In the full-covariate case $W_i = X_i$, their outcome-response ignorability and treatment-identification conditions imply $\delta_{\text{cc}}(w) = \delta_{\text{po}}(w)$. As some confounders are missing, they estimate outcome models among complete cases, impute missing covariates such that they can project the fitted models onto the completed covariate data to average predicted conditional effects over the full sample.

Our framework generalizes this identification argument by specifying graphical conditions under which complete-case comparisons identify $\delta_{\text{cc}}(x)$ (Proposition 1). It also generalizes the transport step in two ways: mean exchangeability allows selection to change conditional outcome means as long as the conditional treatment effect is preserved, and the transport covariates W_i may be a subvector of the adjustment covariates X_i. Rather than first pooling these comparisons into the complete-case ATE, τ_{cc}, through the regression approaches in Propositions 2 and 3, Corollary 3 provides a direct route to the population ATE: given identified within-X_i effects, average $\delta_{\text{cc}}(x)$ over the complete-case distribution of X_i within each $W_i = w$ group, then average the resulting effects over the population distribution of W_i, assuming mean exchangeability and complete-case support given W_i.

5.2 Reanalysis of Bodea and Hicks (2015)

Bodea and Hicks (2015) study the conditional effects of central bank independence (CBI) using data on 78 countries from 1973 to 2008. They find that CBI lowers inflation in democracies but not in nondemocracies. Their country fixed-effects regressions interact the CBI index with six

measures of democracy and democratic institutions: Polity (Marshall and Gurr 2020), Freedom House (Freedom House 2021), executive constraints (Marshall and Gurr 2020), checks and balances (Scartascini, Cruz and Keefer 2021), political constraints (Henisz 2000; 2012), and press freedom (Freedom House 2017).

We reanalyze their study using the authors' updated and expanded CBI data, which cover 143 countries and the European Central Bank from 1970 to 2020 (Bodea and Hicks 2023), along with updated data from the original sources of the moderators. For 1973–2008, the expanded data contain 132 countries with CBI and Polity observed[28], of which 115 (2,198 country-years) have every control observed. The outcome is signed log inflation, and effects are reported per 0.1 increase in the CBI index. Without a moderator, we reproduce the authors' original CBI coefficient of -0.465 (SE 0.143) per unit of CBI. The expanded data yield -0.546 (Figure 8, leftmost panel).

In the original data, 2,061 country-years in 79 countries have CBI and Polity observed during the analysis period. After listwise deletion, 1,769 country-years in 78 countries enter the authors' baseline regression. Our main comparison concerns heterogeneous effects in the expanded data, retaining the economic controls and country fixed effects. To account for possible nonlinearity in the moderating effects (Hainmueller, Mummolo and Xu 2019), we divide each measure into three categories and fit a country fixed-effects regression within each. Holding these category-specific estimates fixed, we compare their averages using complete-case and target-population shares (Table G1, Panel B). Although CBI is continuous, the population-averaging result in Corollary 3 can be extended to category-specific CBI slopes, taking their causal interpretation and consistent estimation as given[29].

Bodea and Hicks (2015) do not specify a missingness mechanism for their main regression variables. For this illustration, we assume the causal structure in Figure 3 and treatment overlap among complete cases, so that the within-covariate comparisons satisfy Box 1. We also assume that

[28] Some updated-CBI countries and the ECB are not in the expanded data (see the notes to Table G1).

[29] The arguments in Propositions 1–2 and Corollary 3 can be adapted to continuous treatments. A sufficient DGP is $Y_{it}(d) = a_{i,w} + \mu_w(H_{it}) + s_w d + u_{it}$, where w is a pretreatment moderator category and H_{it} contains pretreatment controls and outcome history. Under valid adjustment, treatment support, consistency, no interference, and a correctly specified complete-case conditional mean, s_w is the category-specific causal slope. If these slopes transport to the target categories, averaging $0.1 s_w$ using target shares gives the target average effect per 0.1 increase in CBI.

the complete-case outcome regression is correctly specified, satisfying Box 2. With these conditions in place, we focus on Box 3 and show that even valid complete-case comparisons may not recover the population ATE when treatment effects vary and listwise deletion changes the distribution of moderators.

Following Corollary 3, with the moderator as W_i, we assume mean exchangeability of selection and treatment effect heterogeneity[30]: the average effect within each moderator category is the same among complete cases and in the target population, in addition to complete-case support. We then hold the estimated category-specific effects fixed and average them using their population shares to recover the population average.

Whereas Bodea and Hicks (2015) focus on the conditional effects themselves, we consider a case in which researchers are concerned with the population ATE. Suppose the target population consists of country-years in the countries covered by the expanded CBI data. As Figure 8 shows, for three of the six moderators, the complete-case average is negative and statistically significant, whereas the average reweighted to the target population's moderator distribution is not statistically distinguishable from zero. For four of the six moderators, the difference between the complete-case and reweighted averages is statistically significant at the 5% level using unadjusted p-values.

This is because listwise deletion changes the moderator distribution: democracies and countries with stronger democratic institutions are overrepresented among complete cases[31], so the complete-case average places more weight on effects in these categories than the population average. Interpreting the complete-case average as a population average can therefore be misleading. The same pattern also holds when the target population is all country-years covered by Polity5, a larger population than the country-years in countries covered by the expanded CBI data. Table G1 reports the numerical results.

[30]Here, we assume that each category's fixed-effects slope equals its complete-case average effect, which in turn equals its target-population average. For example, under Polity, only 11 of the 38 complete-case autocracies have within-country CBI variation. The slope identified from these 11 countries must therefore represent the average effect across all complete-case autocracies and, under mean exchangeability, all target-population autocracies.

[31]As Panel C of Table G1 shows, in the expanded data, democracies account for 67.2% of complete-case country-years, while the corresponding share among country-years in the target population is 48.4%.

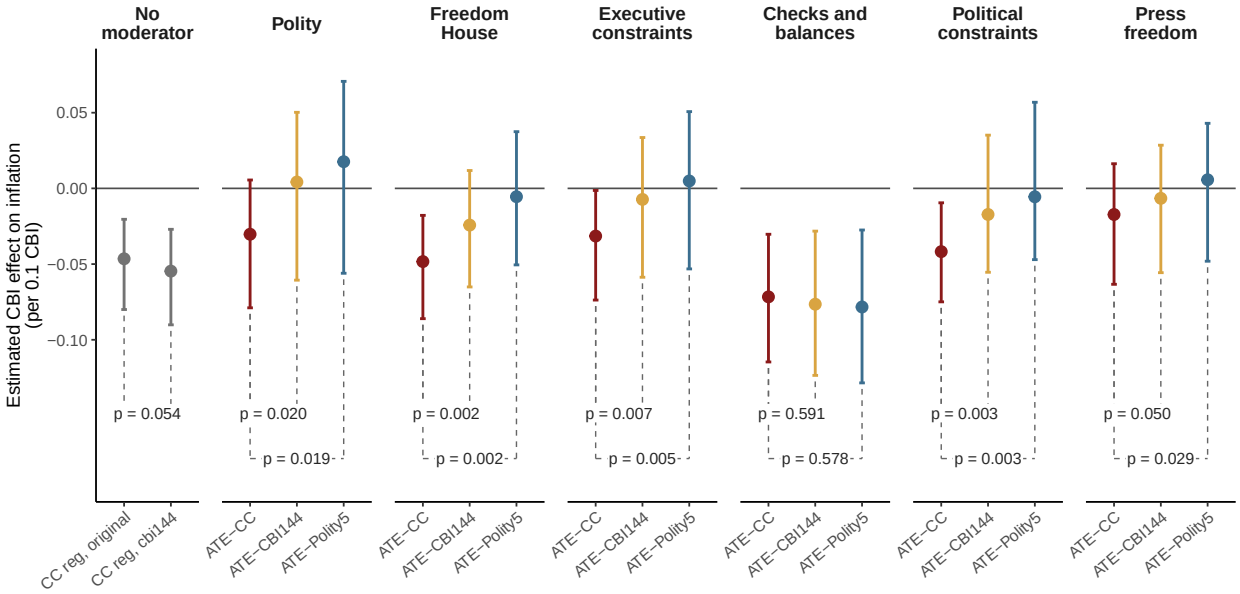


Figure 8: Effect of Central Bank Independence on Inflation, Complete Cases and Reweighted Averages

Notes: Points are average effects of a 0.1 increase in the CBI index on signed log inflation. Intervals are 95% country-cluster bootstrap intervals. The first panel reports complete-case (CC) regression coefficients for the original data and the expanded data (cbi144) without a moderator. In the other panels, ATE-CC uses complete-case category shares, ATE-CBI144 uses shares among countries covered by the updated CBI data (where moderator observed), and ATE-Polity5 uses shares among countries covered by Polity5 (where moderator observed). Dashed brackets give two-sided bootstrap p-values for differences between linked estimates. Holm-adjusted p values reported in Table G1. Across the six moderators, the lowest proportion of countries with within-country CBI variation in any category ranges from 25% to 38%.

6 Checking All Three Boxes: Complete-Case Analysis in a Canvassing Experiment

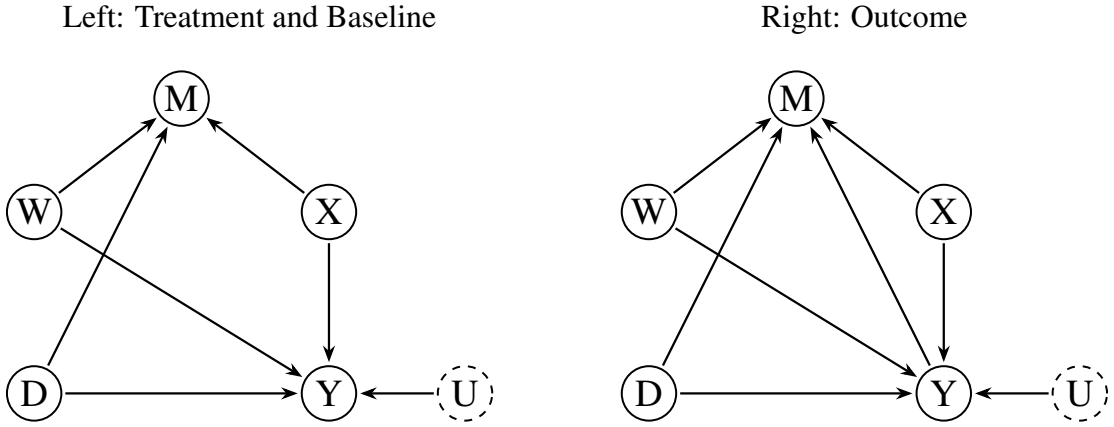
The preceding two sections illustrate the use of DAGs and reweighting to assess complete-case regression. We now reanalyze Kalla and Broockman (2023) to show how researchers can justify listwise deletion through all three checks: the validity of treatment comparisons using a DAG, the pooling of those comparisons through treatment weighting, and reweighting to recover the population ATE.

Their Study 1 examines whether canvassing conversations reduce prejudice toward undocumented immigrants. We compare their full intervention, which used personal stories and perspective-taking, with a placebo conversation on an unrelated topic, and use their first-follow-up outcome index, measured approximately one week after canvassing. Among the 1,900 participants in these two arms, 728 (38.3 percent) have no observed first-follow-up outcome, leaving 1,172 complete cases. Treatment assignment, baseline education, and the regression controls are fully observed. The missing pattern is purely on the follow-up outcomes: 45.0 percent of participants without a college degree and 30.9 percent of those with one lack the outcome.

Box 1: Do Treatment Comparisons Remain Valid after Listwise Deletion? Let D denote assignment to the full intervention rather than the placebo, W baseline college education, X the remaining baseline controls, and Y the first-follow-up outcome. Together, (W,X) form the pretreatment covariate vector, the X in the general framework. Let $M = 1$ indicate that Y is missing. We consider two possible missing mechanisms, represented by two DAGs in Figure 9.

In the left panel of Figure 9, missingness has three causes: treatment D, baseline college education W, and the remaining baseline controls X, which include prior immigration attitudes. College-educated participants may be more likely to answer an emailed follow-up survey, and the canvassing itself may change willingness to respond[32].

[32]The authors find no evidence of differential attrition: follow-up response is 61.28% in the placebo arm and 62.09% in the treatment arm. We further test whether response depends on treatment conditional on education and the other



D: assignment to the full intervention rather than the placebo
X: baseline outcome items and other baseline controls, excluding education
U: unobserved causes of the follow-up outcome
W: baseline college education
Y: first-follow-up outcome
M: missing indicator for Y

Figure 9: Two Possible Missing Mechanisms in the Canvassing Experiment

Alternatively, one may believe that follow-up response further depends on post-treatment attitudes. For example, among participants with the same baseline characteristics, those who would express less favorable attitudes toward undocumented immigrants might be less willing to respond. This adds the arrow $Y \to M$, shown in the right panel of Figure 9.

In the left panel, conditioning on the collider M opens the noncausal paths $D \to M \leftarrow W \to Y$ and $D \to M \leftarrow X \to Y$. Adjusting for W and X blocks both paths. Moreover, M is not a descendant of a node on the causal path, so the adjustment criterion in Proposition 1 is satisfied. In the right panel, M is a descendant of Y on the causal path $D \to Y$. Conditioning on M therefore violates the first requirement of the adjustment criterion, so complete-case comparisons are no longer justified by Proposition 1.

Thus, under the left DAG, Box 1 is checked. Under the right DAG, there is no guarantee for valid treatment comparisons under listwise deletion, and researchers need to consider alternative strategies. For the rest of this illustration, we focus on the left DAG.

baseline controls. In a linear probability model with treatment–covariate interactions, a household-clustered Wald test jointly rejects the treatment and interaction terms ($\chi^2(36) = 61.28$, $p = 0.0054$). This evidence favors allowing $D \to M$.

Box 2: Does the Regression Appropriately Pool the Treatment Comparisons? We next assess whether the regression appropriately pools treatment comparisons within baseline-covariate strata into the complete-case ATE. The authors randomly assigned households within blocks, with equal probabilities of assignment to each arm[33]. Under the left panel of the DAG, treatment and baseline characteristics jointly determine whether the follow-up outcome is missing, so the complete-case treatment propensity may vary across baseline-covariate strata. We therefore use treatment IPW under Proposition 3 to pool the comparisons, estimating treatment probabilities among complete cases with a probit model including education and other baseline controls.

The IPW estimate is 0.0932 (Table H1). Under Boxes 1 and 2, it consistently estimates the complete-case ATE, the average effect among participants who responded to the first follow-up. We also estimate effects separately by education, fitting a treatment propensity model and treatment-IPW regression in each group. The estimates are 0.0497 for participants without a college degree and 0.1107 for those with one. We use these estimates for reweighting below.

Box 3: Does the Complete-Case Regression Recover the Population ATE? Having checked Boxes 1 and 2, we ask whether a complete-case regression that omits the education moderator recovers the population ATE. When effects vary by education, the complete-case ATE averages conditional effects over the education distribution among complete cases, whereas the population ATE averages population conditional effects over the population distribution (Corollary 3).

Suppose mean exchangeability of selection and treatment effect heterogeneity holds at the level of education, so that within each education group the average treatment effect is the same among complete cases and in the population. If complete cases are represented in every population education group, we can estimate the population[34] ATE by averaging the complete-case education-specific effects using the population shares.

In the data, college-educated participants account for 53.2% of complete cases, but 47.5% of

[33]For this application, we assume i.i.d. sampling of households rather than individuals, in which consistency still holds under regularity conditions.

[34]Population refers to the popuation represented by the 1,900 participants under original authors' recruitment and sampling procedure

the target sample. Averaging the education-specific effects using the target shares yields 0.0787, compared[35] with 0.0925 from the pooled complete-case regression. Their difference is -0.0138, with a 95% paired household-bootstrap interval of $[-0.0488, 0.0174]$.

Thus, reweighting to the population share does not produce a statistically significant change in the estimated effect. This is partly due to the fact that the treatment effect does not statistically differ by education ($p = 0.281$, two-sided household bootstrap). If treatment effects do differ, however, the statistically insignificant difference between the complete-case and reweighted estimates in this finite sample does not establish that the complete-case estimator converges to the population ATE. The difference between the education distributions of complete cases and the population also would not disappear as the sample size grows, so does the resulting difference in averaging weights.

7 Concluding Remarks

This article asks when complete-case regression consistently estimates the population average treatment effect. It does so when complete-case treatment comparisons identify the relevant conditional effects (under the adjustment criterion and treatment overlap) and the regression estimator pools those comparisons appropriately (through correct specification or IPW). With heterogeneous effects, recovering the population ATE also requires averaging the conditional effects over the population covariate distribution. Most importantly, MAR or NMAR labels alone do not establish these conditions.

Our framework brings existing guidance on complete-case regression into an augmented causal diagram. Recommendations for missingness driven by covariates or the outcome, as summarized by Pepinsky (2018) and incorporating the result of Arel-Bundock and Pelc (2018), follow as special cases. The collider–mediator illustration shows why MAR labels alone cannot determine whether listwise deletion is justified: with the same data, missing pattern, and MAR classification, adjusting for authoritarian tendency recovers the total effect when it is a confounder, but not when it is a

[35] This comparison captures both treatment IPW and reweighting by education. Using the same education-specific estimates, changing from complete-case to target shares gives a difference of -0.0035.

mediator. Because these causal structures can be observationally equivalent, researchers must state the assumptions they use to justify listwise deletion, just as they do when choosing an adjustment set for the oracle regression.

For empirical researchers, Table 2 summarizes the checks: start with the oracle regression's adjustment set, which would identify the population ATE, assess treatment comparisons in its augmented DAG, verify that the estimator pools them appropriately, and, with heterogeneous effects, average conditional effects over the target population's covariate distribution. Our Bodea–Hicks reanalysis illustrates that a negative, statistically significant complete-case average can become indistinguishable from zero after reweighting. In the canvassing reanalysis, the complete-case and reweighted estimates are not statistically distinguishable, but this finite-sample result does not establish consistency: any population gap caused by averaging over the wrong distribution persists as the sample grows. Our framework gives researchers a formal basis for assessing listwise deletion and its consequences. By making the assumed causes of missingness explicit, researchers can explain when deletion is appropriate for their causal question.

When the conditions cannot be established, researchers have two broad options. They may turn to methods that address the missing part of the data, such as imputation or IPW for missingness, and assess the assumptions those methods require. Alternatively, they may retain listwise deletion and use bounds or sensitivity analysis to assess what can be learned when the conditions for complete-case regression are uncertain.

Future work could develop diagnostics for the outcome specification and extend the framework to other causal estimands, such as the ATT and the complier average treatment effect, and to estimation and inference with continuous treatment-effect modifiers. Software could also help researchers construct augmented causal diagrams and assess whether they satisfy the graphical conditions for valid complete-case comparisons. Finally, we note that even when complete-case regression is consistent, the number of complete observations can become very small as the number of variables with missing values grows (Wang and Aronow 2023). In such cases, consistency provides limited assurance for the available sample, and imputation may be needed to leverage

information from incomplete observations.

References

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arel-Bundock, Vincent and Krzysztof J. Pelc. 2018. "When Can Multiple Imputation Improve Regression Estimates?" *Political Analysis* 26(2):240–245.

Aronow, P. M. and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60(1):250–267.

Bareinboim, Elias, Jin Tian and Judea Pearl. 2014. Recovering from Selection Bias in Causal and Statistical Inference. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI Press pp. 2410–2416.
URL: `https://ftp.cs.ucla.edu/pub/stat_ser/r425.pdf`

Bareinboim, Elias and Judea Pearl. 2012. Controlling Selection Bias in Causal Inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. Vol. 22 of *Proceedings of Machine Learning Research* pp. 100–108.
URL: `https://proceedings.mlr.press/v22/bareinboim12.html`

Bareinboim, Elias and Judea Pearl. 2016. "Causal Inference and the Data-Fusion Problem." *Proceedings of the National Academy of Sciences* 113(27):7345–7352.

Bodea, Cristina and Raymond Hicks. 2015. "Price Stability and Central Bank Independence: Discipline, Credibility, and Democratic Institutions." *International Organization* 69(1):35–61.

Bodea, Cristina and Raymond Hicks. 2023. "Central Bank Independence Data." `https://github.com/arpie71/CBIdata`. Accessed November 6, 2026.

Carmines, Edward and Rita Nassar. 2021. "How Social Desirability Bias Affects Immigration Attitudes in a Hyperpolarized Political Environment." *Social Science Quarterly* 102(4):1803–1811.

Chen, Zhongren and Melody Huang. 2026. "Generalizing Causal Effects with Noncompliance: Application to Deep Canvassing Experiments." *Political Analysis* pp. 1–18.
URL: https://doi.org/10.1017/pan.2026.10063

Choi, Jungyeon, Olaf M. Dekkers and Saskia le Cessie. 2019. "A Comparison of Different Methods to Handle Missing Data in the Context of Propensity Score Analysis." *European Journal of Epidemiology* 34(1):23–36.

Choi, Jungyeon, Olaf M. Dekkers and Saskia le Cessie. 2020. "Authors' Reply: A Comparison of Different Methods to Handle Missing Data in the Context of Propensity Score Analysis." *European Journal of Epidemiology* 35(1):89–91.

Cinelli, Carlos, Andrew Forney and Judea Pearl. 2024. "A Crash Course in Good and Bad Controls." *Sociological Methods & Research* 53(3):1071–1104.

Daniel, Rhian M., Michael G. Kenward, Simon N. Cousens and Bianca L. De Stavola. 2012. "Using Causal Diagrams to Guide Analysis in Missing Data Problems." *Statistical Methods in Medical Research* 21(3):243–256.

Egami, Naoki and Erin Hartman. 2023. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review* 117(3):1070–1088.

Engelhardt, Andrew M., Stanley Feldman and Marc J. Hetherington. 2023. "Advancing the Measurement of Authoritarianism." *Political Behavior* 45(2):537–560.

Freedom House. 2017. "Freedom of the Press: Data, Editions 1980–2017." Historical data workbook, https://freedomhouse.org/sites/default/files/2020-02/FOTP1980-FOTP2017_Public-Data.xlsx. Accessed November 2026.

Freedom House. 2021. "Freedom in the World: Country and Territory Ratings and Statuses, 1973–2021." Historical data workbook, https://freedomhouse.org/sites/default/files/2021-02/Country_and_Territory_Ratings_and_Statuses_FIW1973-2021.xlsx.

Gabriel, Erin E., Michael C. Sachs, Torben Martinussen, Ingeborg Waernbaum, Els Goetghebeur, Stijn Vansteelandt and Arvid Sjölander. 2024. "Inverse Probability of Treatment Weighting with Generalized Linear Outcome Models for Doubly Robust Estimation." *Statistics in Medicine* 43(3):534–547.

Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27(2):163–192.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153–161.

Henisz, Witold J. 2000. "The Institutional Environment for Economic Growth." *Economics and Politics* 12(1):1–31.

Henisz, Witold J. 2012. "The Political Constraint Index (POLCON) Dataset: 2012 Release." The Wharton School, University of Pennsylvania, https://mgmt.wharton.upenn.edu/faculty/heniszpolcon/polcondataset/.

Hernán, Miguel A. and James M. Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
URL: https://miguelhernan.org/whatifbook

Hirano, Keisuke, Guido W. Imbens and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71(4):1161–1189.

Honaker, James and Gary King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54(2):561–581.

Huang, Melody and Naijia Liu. 2026. "Missing Data in Confounders." Working paper.
URL: `https://naijialiu.github.io/CV/confounders.pdf`

Hughes, Rachael A., Jon Heron, Jonathan A. C. Sterne and Kate Tilling. 2019. "Accounting for Missing Data in Statistical Analyses: Multiple Imputation Is Not Always the Answer." *International Journal of Epidemiology* 48(4):1294–1304.

Imai, Kosuke, Luke Keele and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15(4):309–334.

Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4):765–789.

Kalla, Joshua L. and David E. Broockman. 2023. "Which Narrative Strategies Durably Reduce Prejudice? Evidence from Field and Survey Experiments Supporting the Efficacy of Perspective-Getting." *American Journal of Political Science* 67(1):185–204.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.

Lall, Ranjit. 2016. "How Multiple Imputation Makes a Difference." *Political Analysis* 24(4):414–433.

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7(1):295–318.

Little, Roderick J. A. and Donald B. Rubin. 2019. *Statistical Analysis with Missing Data*. 3 ed. John Wiley & Sons.

Marbach, Moritz. 2022. "Choosing Imputation Models." *Political Analysis* 30(4):597–605.

Marshall, Monty G. and Ted Robert Gurr. 2020. "Polity5: Political Regime Characteristics and Transitions, 1800–2018." Center for Systemic Peace, https://www.systemicpeace.org/inscrdata.html.

Mohan, Karthika and Judea Pearl. 2021. "Graphical Models for Processing Missing Data." *Journal of the American Statistical Association* 116(534):1023–1037.

Pearl, Judea. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82(4):669–688.

Pepinsky, Thomas B. 2018. "A Note on Listwise Deletion versus Multiple Imputation." *Political Analysis* 26(4):480–488.

Prentice, R. L. and R. Pyke. 1979. "Logistic Disease Incidence Models and Case-Control Studies." *Biometrika* 66(3):403–411.

Rivlin-Angert, Naama, Alon Yakter and Lior Sheffer. 2025. "Do Personality Traits Predict Voter Attitudes When Politics Is Structured around Conflict? Lessons from Israel." *Public Opinion Quarterly* 89(2):389–414.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.

Ross, Rachael K., Alexander Breskin and Daniel Westreich. 2020. "When Is a Complete-Case Approach to Missing Data Valid? The Importance of Effect-Measure Modification." *American Journal of Epidemiology* 189(12):1583–1589.

Saadati, Mojdeh and Jin Tian. 2020. Adjustment Criteria for Recovering Causal Effects from Missing Data. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019*. Vol. 11906 of *Lecture Notes in Artificial Intelligence* Springer pp. 561–577.

Scartascini, Carlos, Cesi Cruz and Philip Keefer. 2021. "The Database of Political Institutions (DPI): 2020." IDB Open Data. https://doi.org/10.60966/gn3s-at31.

Seaman, Shaun, John Galati, Dan Jackson and John Carlin. 2013. "What Is Meant by "Missing at Random"?" *Statistical Science* 28(2):257–268.

Seaman, Shaun R. and Ian R. White. 2013. "Review of Inverse Probability Weighting for Dealing with Missing Data." *Statistical Methods in Medical Research* 22(3):278–295.

Shpitser, Ilya, Tyler J. VanderWeele and James M. Robins. 2010. On the Validity of Covariate Adjustment for Estimating Causal Effects. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. AUAI Press pp. 527–536.

Stenner, Karen. 2005. *The Authoritarian Dynamic*. Cambridge University Press.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele and Kosuke Imai. 2014. "mediation: R Package for Causal Mediation Analysis." *Journal of Statistical Software* 59(5):1–38.
URL: `https://www.jstatsoft.org/article/view/v059i05`

Wang, J. Sophia and P. M. Aronow. 2023. "Listwise Deletion in High Dimensions." *Political Analysis* 31(1):149–155.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yang, Shu, Linbo Wang and Peng Ding. 2019. "Causal Inference with Confounders Missing Not at Random." *Biometrika* 106(4):875–888.

Zamarro, Gema, Albert Cheng, M. Danish Shakeel and Collin Hitt. 2018. "Comparing and Validating Measures of Non-Cognitive Traits: Performance Task Measures and Self-Reports from a Nationally Representative Internet Panel." *Journal of Behavioral and Experimental Economics* 72:51–60.

When Listwise Deletion Works:

Causal Diagrams and Complete-Case Regression

Supporting Information

Table of Contents

A Review of Missing Data in *APSR* and *AJPS*, 2019–2024 **1**

B Numerical Results for the Motivating Simulation **2**
B.1 Saturated Regression and IPW under Mechanisms B and C 2

C MAR Companion to the Motivating Simulation **3**

D Technical Definitions, Assumptions, and Proofs **5**
D.1 Adjustment criterion 5
D.2 SUTVA, Oracle Samples, Missing Mechanism, and Proof of Proposition 1 6
D.3 Proof of Proposition 2 7
D.4 Oracle specification and the indicator–error independence condition of Arel-Bundock and Pelc (2018) 9
D.5 Proof of Proposition D.1 12
D.6 Proof of Corollary 1 14
D.7 Proof of Corollary 2 15
D.8 Proof of Corollary 3 15
D.9 Complete-case IPW regression 16

E Simulation of the Collider–Mediator Dilemma **20**

F Reweighting for Generalized Linear Models **21**

G Estimates for the Reanalysis of Bodea and Hicks (2015) **23**

H Treatment-IPW Comparison for the Canvassing Experiment **24**

A Review of Missing Data in *APSR* and *AJPS*, 2019–2024

To approximate Huang and Liu (2026), we screened 2019–2024 Crossref records for causal terms, such as *causal*, *effect of*, and *confound*, and excluded records mentioning experimental designs. We chose this rule partly to approximate their sample sizes before full-text eligibility review. Full-text eligibility review retained 100 *APSR* and 58 *AJPS* observational causal studies.

Claude Opus 5.5 coded ten pilot articles, and Claude Sonnet 5 coded the remaining articles' main texts and available appendices. We reviewed the AI-generated coding using the same standards we apply to work produced by a human research assistant and found the results reliable. We asked the agent to code deletion as stated when named and implied when missing values were given as the reason for dropping observations, while not to count unexplained sample-size differences. Table A1 summarizes the results.

Table A1: Missing data in eligible observational causal studies

	APSR ($n = 100$)	*AJPS* ($n = 58$)
Mention of missing values		
Main text	65 (65%)	42 (72%)
Main text, including restricted-availability analyses	88 (88%)	52 (90%)
Main text or appendix	95 (95%)	55 (95%)
Huang and Liu (2026), main text	70/130 (54%)	41/90 (46%)
Case deletion		
Stated or implied, main text	47 (47%)	33 (57%)
Stated, main text	3 (3%)	1 (2%)
Stated or implied, main text or appendix	68 (68%)	39 (67%)
Baseline column, complete or available cases	50 (50%)	29 (50%)
Regression on a deleted sample, of deletion articles	27/47 (57%)	18/33 (55%)
Imputation		
Imputation model, main text	7 (7%)	4 (7%)
Imputation model, main text or appendix	10 (10%)	8 (14%)
Any fill-in, main text	27 (27%)	15 (26%)
Any fill-in, appendix only	15 (15%)	7 (12%)
Huang and Liu (2026), imputation model	fewer than 7	fewer than 5

Note: A paper can fall in several rows. Restricted-availability analyses are those limited to the periods or units for which a variable exists. The baseline-column row comes from the separate baseline-specification codebook. The regression row uses the articles with deletion stated or implied in the main text as its denominator and is a lower bound.

B Numerical Results for the Motivating Simulation

Table B1: Estimated Treatment Coefficients under Different Missing Mechanisms

Mechanism	Regression	Mean	SD	2.5th pct.	97.5th pct.
A	Oracle	1.001	0.018	0.968	1.037
A	Complete case	1.002	0.021	0.965	1.047
A	MI (MAR model)	1.057	0.020	1.020	1.097
A	Complete case with IPW	1.002	0.022	0.961	1.045
B	Oracle	1.001	0.018	0.968	1.037
B	Complete case	1.165	0.020	1.128	1.209
B	Complete case, saturated X	1.002	0.021	0.964	1.044
B	MI (MAR model)	1.177	0.019	1.143	1.214
B	Complete case with IPW	1.002	0.021	0.963	1.043
C	Oracle	1.001	0.018	0.968	1.037
C	Complete case	1.497	0.021	1.460	1.539
C	Complete case, saturated X	1.541	0.022	1.503	1.584
C	MI (MAR model)	1.190	0.018	1.157	1.228
C	Complete case with IPW	1.580	0.021	1.542	1.620

Note: MI imputes missing X_i from observed D_i and Y_i using a multinomial MAR working model with 20 imputations. The treatment coefficient is pooled across completed datasets. SD and percentiles summarize variation across the 500 replications.

B.1 Saturated Regression and IPW under Mechanisms B and C

Recall the within-X_i complete-case model from the main text, $Y_i = \alpha_{X_i} + \beta_{X_i} D_i + \varepsilon_{\text{str},i}$. Here β_x is the observed treated-minus-untreated contrast at $X_i = x$. For both mechanisms, the *Complete case, saturated X* row in Table B1 estimates, among $M_i = 0$ cases, the unweighted regression

$$Y_i = \alpha_{X_i}^{\text{sat}} + \beta_{\text{sat}} D_i + r_i^{\text{sat}}, \qquad M_i = 0.$$

The fitted intercept $\alpha_{X_i}^{\text{sat}}$ varies across the three values of X_i, while β_{sat} is common to them. r_i^{sat} is the regression residual. The simulation implements these intercepts with X_i, $\mathbf{1}\{X_i = -1\}$, and $\mathbf{1}\{X_i = 1\}$. In our DGP, one of these terms is dropped because of perfect collinearity.

The *Complete case with IPW* row estimates the weighted regression

$$Y_i = \alpha_{\text{IPW}} + \beta_{\text{IPW}} D_i + \gamma_{\text{IPW}} X_i + r_i^{\text{IPW}}, \qquad M_i = 0,$$

with $\omega_i = 1/\hat{e}_0(X_i)$ for treated units and $\omega_i = 1/[1-\hat{e}_0(X_i)]$ for untreated units. $\hat{e}_0(x)$ is the treated proportion among complete cases with $X_i = x$, truncated to $[0.01, 0.99]$ before weighting. r_i^{IPW} is the regression residual.

The two mean treatment coefficients in Table B1 agree under B: 1.002 for saturated OLS and 1.002 for IPW. Under C, they differ: 1.541 and 1.580, respectively.

To explain the difference, let $e_0(x) = P(D_i = 1 \mid X_i = x, M_i = 0)$. In the population, the two coefficients pool the β_x contrasts as

$$\begin{aligned} \text{IPW:} \quad & \sum_x P(X = x \mid M = 0)\,\beta_x, \\ \text{Saturated-}X\text{ OLS:} \quad & \frac{\sum_x P(X = x \mid M = 0) e_0(x)[1 - e_0(x)]\,\beta_x}{\sum_x P(X = x \mid M = 0) e_0(x)[1 - e_0(x)]}. \end{aligned}$$

Thus, saturated-X OLS weights each stratum by the within-stratum variance of treatment as well as its complete-case frequency (Angrist and Pischke 2009). Under B, $\beta_x = 1$ for every x, so the different weights produce the same population coefficient. Under C, $(\beta_{-1}, \beta_0, \beta_1) = (2, 1, 2)$, so the weights produce different coefficients.

C MAR Companion to the Motivating Simulation

We retain the outcome process, treatment assignment, and regression from Section 2. The oracle sample has 20,000 observations, $Y_i = D_i + X_i + U_i + \xi_i$, and a constant unit-level treatment effect $\tau = 1$. We add an always-observed pretreatment noisy proxy $\widetilde{X}_i = X_i + U_i + \nu_i$, where $\nu_i \sim N(0, 0.25^2)$. Only X_i can be missing.

Let $M_i = 0$ when X_i is observed. We consider three mechanisms:

$$\begin{aligned} \text{A:} \quad & P(M_i = 0 \mid D_i, \widetilde{X}_i) = 0.875 - 0.30 D_i, \\ \text{B:} \quad & P(M_i = 0 \mid D_i, \widetilde{X}_i) = \kappa \mathbf{1}\{\widetilde{X}_i \geq -1.5\}, \\ \text{C:} \quad & P(M_i = 0 \mid D_i, \widetilde{X}_i) = \begin{cases} p_0, & D_i = 0, \\ \kappa \mathbf{1}\{\widetilde{X}_i \geq -1.5\}, & D_i = 1. \end{cases} \end{aligned}$$

The constants $\kappa \simeq 0.9000$ and $p_0 \simeq 0.6946$ give a 25% missingness rate under each mechanism. All three are MAR because observation depends only on the fully observed D_i and $\widetilde{X}_i$. As in Section 2,

we report complete-case regression and treatment IPW; we also report linear and correctly specified multiple imputation (MI)[36].

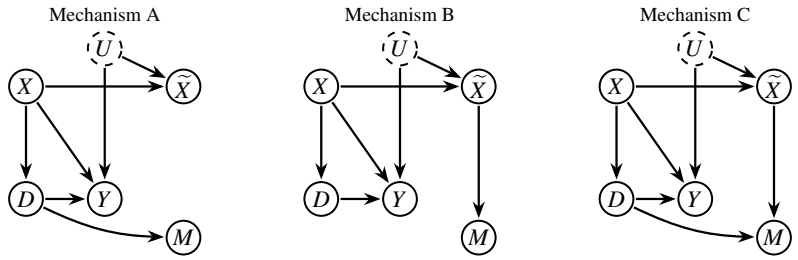


Figure C1: Causal Diagrams for the MAR Missing Mechanisms

Note: The dashed circle marks unobserved U_i. Only X_i can be missing; $M_i = 0$ denotes a complete case.

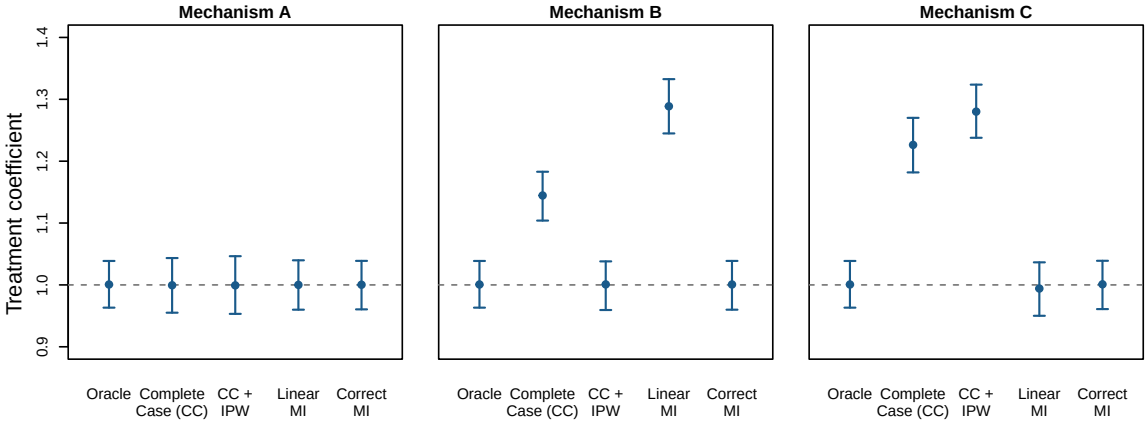


Figure C2: Estimated Treatment Coefficients under MAR Missing Mechanisms

Note: Points show mean treatment-coefficient estimates across 500 replications. Bars show the 2.5th and 97.5th percentiles. The dashed line marks $\tau = 1$. CC + IPW uses inverse treatment-propensity weights.

Figure C2 shows the same progression as the NMAR simulation. Moving from A to B illustrates the need to pool valid within-X_i comparisons appropriately; moving from B to C shows how the adjustment criterion can fail. Correctly specified MI remains centered near the target under all three MAR mechanisms, although linear MI misses the target under B.

[36]Linear MI uses a multinomial-logit model for X_i with linear predictors D_i, Y_i, and $\widetilde{X}_i$. Correctly specified MI adds $\log\cosh\{(Y_i - D_i - x)/0.5^2 + (\widetilde{X}_i - x)/0.25^2\}$ for $x \in \{-1, 0, 1\}$. Both use 20 imputations.

D Technical Definitions, Assumptions, and Proofs

D.1 Adjustment criterion

Adjustment criterion in the original paper Definition 5 of Shpitser, VanderWeele and Robins (2010) defines the adjustment criterion as follows, using **D** for their treatment set X to distinguish it from our covariates X_i. Let G be a causal DAG, and let **D**, **Y**, and $\mathscr{Z}$ be pairwise disjoint sets of treatment, outcome, and observed adjustment variables. Let $G_{\overline{\mathbf{D}}}$ denote the graph obtained by deleting all arrows into nodes in **D**. A directed path from **D** to **Y** is *proper* if its only treatment node is its starting node. The set $\mathscr{Z}$ satisfies the adjustment criterion relative to $(\mathbf{D}, \mathbf{Y})$ in G if:

1. In $G_{\overline{\mathbf{D}}}$, no member of $\mathscr{Z}$ may descend from a node $v \notin \mathbf{D}$ that lies on a proper causal path from **D** to **Y** in G.

2. $\mathscr{Z}$ blocks every noncausal path from **D** to **Y** in G.

A node counts as its own descendant. A noncausal path is one that is not directed from its treatment endpoint to its outcome endpoint.

Condition 1 with a single treatment Our analysis has one treatment D and one outcome Y, so every directed path from D to Y is proper. For any node $v \neq D$ on such a path, no directed path from v to a descendant can pass through D, because that would create a cycle. Deleting arrows into D therefore leaves the descendants of v unchanged:

$$\mathrm{De}_{G_{\bar{D}}}(v) = \mathrm{De}_G(v).$$

Condition 1 thus excludes nodes other than D on directed paths from D to Y, together with their descendants in the original DAG.

Condition 2 with a single treatment Condition 2 is unchanged: $\mathscr{Z}$ must block every noncausal path from D to Y in the original graph G, as stated in Section 3.

D.2 SUTVA, Oracle Samples, Missing Mechanism, and Proof of Proposition 1

Assumption A1 (Stable unit treatment value assumption (SUTVA))**.** *Each treatment level has a single outcome-relevant version. Let* $\mathbf{d} = (d_1, \ldots, d_n)$ *be a treatment assignment vector. For every unit* i *and every* $\mathbf{d}$*, the unit's potential outcome depends only on its own assignment:* $Y_i(\mathbf{d}) = Y_i(d_i)$*. For the realized assignment,* $Y_i = Y_i(D_i)$*.*

Definition 4 (Oracle potential-outcome sample and oracle sample)**.** *Let* $Y_i(d)$ *be unit* i*'s potential outcome under treatment* $d \in \{0,1\}$*. The oracle potential-outcome sample is*

$$\mathscr{S}_{\text{po}} = \{(Y_i(0), Y_i(1), D_i, X_i, \mathbf{M}_i) : i = 1, \ldots, n\}, \qquad (Y_i(0), Y_i(1), D_i, X_i, \mathbf{M}_i) \overset{\text{i.i.d.}}{\sim} P_{\text{po}},$$

where P_{po} *is the joint population distribution of the respective unit-level vector.*

Under SUTVA, let $Y_i = Y_i(D_i)$ *denote unit* i*'s realized outcome. The oracle sample is*

$$\mathscr{S}_{\text{or}} = \{(Y_i, D_i, X_i, \mathbf{M}_i) : i = 1, \ldots, n\}, \qquad (Y_i, D_i, X_i, \mathbf{M}_i) \overset{\text{i.i.d.}}{\sim} P_{\text{or}},$$

where P_{or} *is the joint distribution induced by* P_{po} *and* $Y_i = Y_i(D_i)$*. The sample contains every unit's realized values, including those missing in the observed data, and retains each unit's missing pattern.*

Missing mechanism

Definition 5 (Missing mechanism)**.** *The missing mechanism specifies the conditional probability of each missing pattern given the full realized data:*

$$\pi(\mathbf{m} \mid y, d, x) = P(\mathbf{M}_i = \mathbf{m} \mid Y_i = y, D_i = d, X_i = x).$$

Assumption A2 (Complete-case treatment overlap)**.** *Let* $e_0(x) = P(D_i = 1 \mid X_i = x, M_i = 0)$*. For every* x *in the complete-case support,* $0 < e_0(x) < 1$*.*

Proof of Proposition 1(i). Apply Shpitser, VanderWeele and Robins (2010, Theorem 4) with their treatment X equal to D_i and their adjustment set Z equal to $\{X_i, M_i\}$. Their intervention-defined counterfactual Y_x is our potential outcome $Y_i(d)$ when $x = d$. □

Proof of Proposition 1(ii). Fix a supported complete-case stratum $X_i = x, M_i = 0$ satisfying treatment overlap. The first equality follows from consistency, $Y_i = Y_i(D_i)$, which means that a unit assigned d has realized outcome $Y_i(d)$, while the second follows from Proposition 1(i):

$$\begin{aligned}\mathbb{E}[Y_i \mid D_i = d, X_i = x, M_i = 0] &= \mathbb{E}[Y_i(d) \mid D_i = d, X_i = x, M_i = 0] \\ &= \mathbb{E}[Y_i(d) \mid X_i = x, M_i = 0].\end{aligned}$$

Subtracting the expressions for $d = 1$ and $d = 0$ gives

$$\begin{aligned}&\mathbb{E}[Y_i \mid D_i = 1, X_i = x, M_i = 0] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x, M_i = 0] \\ &\quad = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, M_i = 0] = \delta_{\mathrm{cc}}(x).\end{aligned}$$

Because D_i is binary, the population coefficient in a regression on an intercept and D_i within this stratum equals this difference in means. Under a constant unit-level treatment effect, $\delta_{\mathrm{cc}}(x) = \tau$. □

D.3 Proof of Proposition 2

Assumption A3 (Complete-case OLS regularity)**.** *Let $Z_i = (g(X_i)^\top, D_i h(X_i)^\top)^\top$, where g and h are fixed finite-dimensional vectors of chosen terms, each including an intercept. For the additive specification, take $g(X_i) = b(X_i)$ and $h(X_i) = 1$. Suppose $P(M_i = 0) > 0$,*

$$\mathbb{E}[Y_i^2 + \|Z_i\|^2 + \|h(X_i)\|^2 \mid M_i = 0] < \infty,$$

and $G_0 = \mathbb{E}[Z_i Z_i^\top \mid M_i = 0]$ is positive definite.

Proof of Proposition 2. Correct specification of the complete-case conditional mean gives

$$\mathbb{E}[Y_i \mid D_i, X_i, M_i = 0] = g(X_i)^\top \alpha_{\mathrm{cc}} + D_i h(X_i)^\top \beta_{\mathrm{cc}}.$$

Under Assumption A2 (complete-case treatment overlap), both treatment arms have positive probability at each supported $X_i = x$. Subtracting their conditional means cancels the baseline term $g(x)^\top \alpha_{\mathrm{cc}}$, giving

$$\mathbb{E}[Y_i \mid D_i = 1, X_i = x, M_i = 0] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x, M_i = 0] = h(x)^\top \beta_{\mathrm{cc}}.$$

By SUTVA (Assumption A1), a unit assigned d has realized outcome $Y_i(d)$. The adjustment

criterion then makes the comparison causal, as established in Proposition 1(ii). Hence

$$\delta_{\mathrm{cc}}(x) = h(x)^\top \beta_{\mathrm{cc}}.$$

Averaging over $X_i \mid M_i = 0$ and applying the law of iterated expectations gives

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0] = \mathbb{E}[h(X_i)^\top \beta_{\mathrm{cc}} \mid M_i = 0] = \tau_{\mathrm{cc}}.$$

The integrability needed for this averaging follows from Assumption A4.

For consistency, let $\vartheta_{\mathrm{cc}} = (\alpha_{\mathrm{cc}}^\top, \beta_{\mathrm{cc}}^\top)^\top$. Multiplying the zero conditional residual mean by Z_i and averaging gives the complete-case population OLS normal equations:

$$\mathbb{E}[Z_i(Y_i - Z_i^\top \vartheta_{\mathrm{cc}}) \mid M_i = 0] = 0.$$

Their solution is unique under Assumption A3. Define

$$\widehat{G}_0 = \frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} Z_i Z_i^\top, \qquad \widehat{q}_0 = \frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} Z_i Y_i.$$

The i.i.d. sampling specified in Subsection 3.1, positive complete-case probability, and finite moments in Assumption A3 imply, by the law of large numbers applied to selected sums and the complete-case fraction,

$$\widehat{G}_0 \xrightarrow{p} G_0, \qquad \widehat{q}_0 \xrightarrow{p} G_0 \vartheta_{\mathrm{cc}}.$$

Positive definiteness of G_0 and the continuous mapping theorem give

$$\widehat{\vartheta}_{\mathrm{cc}} = \widehat{G}_0^{-1} \widehat{q}_0 \xrightarrow{p} \vartheta_{\mathrm{cc}}.$$

In particular, $\widehat{\beta}_{\mathrm{cc}} \xrightarrow{p} \beta_{\mathrm{cc}}$. The finite moments of $h(X_i)$ also give

$$\frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} h(X_i) \xrightarrow{p} \mathbb{E}[h(X_i) \mid M_i = 0].$$

Taking the inner product of this sample average with $\widehat{\beta}_{\mathrm{cc}}$ therefore yields

$$\widehat{\tau}_{\mathrm{cc}} = \left\{ \frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} h(X_i) \right\}^\top \widehat{\beta}_{\mathrm{cc}} \xrightarrow{p} \mathbb{E}[h(X_i) \mid M_i = 0]^\top \beta_{\mathrm{cc}} = \tau_{\mathrm{cc}}.$$

□

Treatment-induced attrition The complete-case ATE, $\tau_{\text{cc}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0]$, averages effects among units observed under the realized assignment. If treatment affects attrition, this group may differ under another assignment. In this case, Propositions 1–3 still hold. Under Corollary 3's transport and support conditions, averaging complete-case conditional effects over the population covariate distribution recovers the population ATE.

D.4 Oracle specification and the indicator–error independence condition of Arel-Bundock and Pelc (2018)

Their regression in our notation Arel-Bundock and Pelc (2018, p. 241) start from the full-data model $Y = \mathbf{X}\beta + \varepsilon$ and let $Q_i = 1$ if the ith observation is fully observed. Their complete-case model and estimator are

$$QY = Q\mathbf{X}\beta_c + Q\varepsilon, \qquad Q = \text{diag}(Q_1, \ldots, Q_n),$$

$$\widehat{\beta}_c = (\mathbf{X}^\top Q\mathbf{X})^{-1}\mathbf{X}^\top QY = \beta + (\mathbf{X}^\top Q\mathbf{X})^{-1}\mathbf{X}^\top Q\varepsilon.$$

Under the usual assumptions of the classical linear model, $\widehat{\beta}_c$ is unbiased if Q is independent of ε. In our notation, $Q_i = 1 - M_i$, the ith row of $\mathbf{X}$ is $(D_i, b(X_i)^\top)$, and $\beta = (\beta_{\text{or}}, \gamma_{\text{or}}^\top)^\top$, with residual $\varepsilon_i = \varepsilon_{\text{or},i}$. Their full-data model is therefore the oracle regression of Subsection 3.1, and their complete-case model is the same regression fitted to units with $M_i = 0$. Because the classical linear model conditions on the regressors, their independence of Q and ε is conditional on $\mathbf{X}$, which in our notation is $M_i \perp \varepsilon_{\text{or},i} \mid D_i, X_i$.

Constant conditional ATE in their model Treatment enters their model through the single coefficient β_{or}, so $\mathbb{E}[Y_i \mid D_i = 1, X_i = x] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x] = \beta_{\text{or}}$ for every x. When X_i is a valid adjustment set for the oracle sample, this contrast is the conditional ATE, $\delta_{\text{po}}(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x]$. Their model therefore assumes a constant conditional ATE, $\delta_{\text{po}}(x) = \beta_{\text{or}}$ for all x, while

allowing unit-level effects to vary within covariate strata through $\varepsilon_{\text{or},i}$. When effects instead vary across x, they advise modeling that heterogeneity (Arel-Bundock and Pelc 2018, p. 242), the role of our Section 5.

For the additive special case $g(X_i) = b(X_i)$ and $h(X_i) = 1$, the indicator–error independence condition of Arel-Bundock and Pelc (2018), written in our notation as $M_i \perp \varepsilon_{\text{or},i} \mid D_i, X_i$, together with correct oracle specification, implies the zero conditional mean requirement in Proposition 2. Proposition 2 relaxes these specification and independence restrictions. Its additional graphical condition establishes the causal interpretation as the complete-case ATE, which is not addressed by their coefficient-preservation result.

Sufficient conditions Suppose the oracle conditional mean is correctly specified:

$$Y_i = \beta_{\text{or}} D_i + b(X_i)^\top \gamma_{\text{or}} + \varepsilon_{\text{or},i}, \qquad \mathbb{E}[\varepsilon_{\text{or},i} \mid D_i, X_i] = 0.$$

Suppose also that

$$M_i \perp \varepsilon_{\text{or},i} \mid D_i, X_i,$$

and Assumption A3 (complete-case OLS regularity) holds. Then the same regression among complete cases satisfies

$$\mathbb{E}[\varepsilon_{\text{cc},i} \mid D_i, X_i, M_i = 0] = 0.$$

If $\{X_i, M_i\}$ also satisfies the adjustment criterion and Assumptions A1 (SUTVA), A2 (complete-case treatment overlap), and A4 (complete-case population conditions) hold, then

$$\tau_{\text{cc}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0], \qquad \widehat{\tau}_{\text{cc}} \xrightarrow{p} \tau_{\text{cc}}.$$

Proof. Conditional independence and the correctly specified oracle conditional mean imply, on the complete-case support,

$$\mathbb{E}[\varepsilon_{\text{or},i} \mid D_i, X_i, M_i = 0] = \mathbb{E}[\varepsilon_{\text{or},i} \mid D_i, X_i] = 0.$$

Consequently,

$$\mathbb{E}[Y_i \mid D_i, X_i, M_i = 0] = \beta_{\text{or}} D_i + b(X_i)^\top \gamma_{\text{or}}.$$

These coefficients satisfy the complete-case population OLS normal equations. Assumption A3 ensures their uniqueness, giving $\beta_{\text{cc}} = \beta_{\text{or}}$ and $\gamma_{\text{cc}} = \gamma_{\text{or}}$. Thus the complete-case residual $\varepsilon_{\text{cc},i}$ equals $\varepsilon_{\text{or},i}$, and

$$\mathbb{E}[\varepsilon_{\text{cc},i} \mid D_i, X_i, M_i = 0] = 0.$$

Under the adjustment criterion and Assumptions A1, A2, and A4, Proposition 2 therefore gives

$$\tau_{\text{cc}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0],$$

and consistency of $\widehat{\tau}_{\text{cc}}$ under the maintained i.i.d. sampling in Subsection 3.1. □

These assumptions are stronger than the complete-case conditional mean condition in Proposition 2: the proof needs only preservation of the zero conditional residual mean, whereas conditional independence preserves the entire conditional outcome distribution given the regressors.

Violation of conditional independence in Mechanism B Mechanism B violates this independence condition. Its oracle residual is $U_i + \xi_i$, with $\mathbb{E}[U_i + \xi_i \mid D_i, X_i] = 0$. Because selection depends on U_i, however,

$$\mathbb{E}[U_i + \xi_i \mid D_i, X_i, M_i = 0] = X_i^2/2.$$

Thus the oracle residual is not conditionally independent of M_i given D_i, X_i, and the same intercept-and-linear-X_i specification no longer captures the complete-case mean. Nevertheless, the graphical adjustment condition still gives valid within-X_i treatment comparisons. Saturated outcome adjustment or treatment IPW can recover their common effect, as shown in the motivating simulation.

MAR does not generally imply the indicator–error independence condition: it permits selection on observed outcome values (Seaman et al. 2013). If only X_i can be missing while Y_i, D_i are observed, $P(M_i = 0 \mid Y_i, D_i, X_i) = P(M_i = 0 \mid Y_i, D_i)$ is MAR, yet selection through Y_i may depend on the oracle residual. If only Y_i can be missing and D_i, X_i are fully observed, MAR gives the

indicator–error independence condition. With a correctly specified oracle conditional mean, it then preserves the zero conditional residual mean among complete cases.

Why transport is not needed despite treatment heterogeneity The indicator–error independence condition of Arel-Bundock and Pelc (2018), together with a zero conditional mean for the oracle residual, gives $\mathbb{E}[Y_i \mid D_i, X_i, M_i = 0] = \mathbb{E}[Y_i \mid D_i, X_i]$: selection leaves the conditional mean of the outcome unchanged. The treated-minus-untreated contrast at each x is therefore the same among complete cases as in the oracle sample. When both contrasts are causal (Proposition 1), this says $\delta_{\text{cc}}(x) = \delta_{\text{po}}(x)$. Under a constant conditional ATE, both equal τ, so the complete-case ATE equals the population ATE without a separate transport assumption, and our comparison with Arel-Bundock and Pelc (2018) rests on the same restriction on effects as theirs.

D.5 Proof of Proposition D.1

Assumption A4 (Complete-case population conditions)**.** $P(M_i = 0) > 0$ *and* $\mathbb{E}[|Y_i(d)| \mid M_i = 0] < \infty$ *for* $d \in \{0, 1\}$.

Assumption A5 (Estimated inverse-weight consistency)**.** *With probability tending to one,* $0 < \widehat{e}_0(x) < 1$ *throughout the complete-case support, and the estimated inverse weights converge uniformly over that support:*

$$\sup_x \left| \frac{1}{\widehat{e}_0(x)} - \frac{1}{e_0(x)} \right| = o_p(1), \qquad \sup_x \left| \frac{1}{1 - \widehat{e}_0(x)} - \frac{1}{1 - e_0(x)} \right| = o_p(1).$$

Proposition D.1 (Complete-case treatment IPW)**.** *Let* $e_0(x) = P(D_i = 1 \mid X_i = x, M_i = 0)$. *Suppose Assumption A1 (SUTVA), Assumption A2 (complete-case treatment overlap), and Assumption A4 (complete-case population conditions) hold, and* $Y_i(d) \perp D_i \mid X_i, M_i = 0$ *for* $d \in \{0, 1\}$. *Then*

$$\mathbb{E}\left[\frac{D_i Y_i}{e_0(X_i)} - \frac{(1 - D_i) Y_i}{1 - e_0(X_i)} \,\middle|\, M_i = 0 \right] = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0].$$

Under Assumption A5, the sample analogue

$$\widehat{\tau}_{\mathrm{IPW}} = \frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} \left\{ \frac{D_i Y_i}{\widehat{e}_0(X_i)} - \frac{(1-D_i)Y_i}{1-\widehat{e}_0(X_i)} \right\}, \qquad n_{\mathrm{cc}} = \sum_{i=1}^{n} \mathbf{1}\{M_i = 0\},$$

converges in probability to $\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0]$.

Here and below, $o_p(1)$ denotes convergence to zero in probability, and $O_p(1)$ denotes boundedness in probability.

Proof of Proposition D.1 For treated units, Assumption A1 (SUTVA) gives $D_i Y_i = D_i Y_i(1)$. The condition $Y_i(1) \perp D_i \mid X_i, M_i = 0$ in Proposition D.1 and the definition of $e_0(x)$ then imply, for each supported x,

$$\begin{aligned} \mathbb{E}\left[\frac{D_i Y_i}{e_0(X_i)} \,\middle|\, X_i = x, M_i = 0 \right] &= \frac{\mathbb{E}[D_i \mid X_i = x, M_i = 0]}{e_0(x)} \mathbb{E}[Y_i(1) \mid X_i = x, M_i = 0] \\ &= \mathbb{E}[Y_i(1) \mid X_i = x, M_i = 0]. \end{aligned}$$

The same argument for untreated units gives

$$\mathbb{E}\left[\frac{(1-D_i)Y_i}{1-e_0(X_i)} \,\middle|\, X_i = x, M_i = 0 \right] = \mathbb{E}[Y_i(0) \mid X_i = x, M_i = 0].$$

Subtracting and averaging over $X_i \mid M_i = 0$ proves the stated identity.

Specifically, define the sample analogue using the true complete-case propensity score as

$$\widetilde{\tau}_{\mathrm{IPW}} = \frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} \left\{ \frac{D_i Y_i}{e_0(X_i)} - \frac{(1-D_i)Y_i}{1-e_0(X_i)} \right\}.$$

Let $\psi_i = D_i Y_i / e_0(X_i) - (1-D_i)Y_i / [1-e_0(X_i)]$ when $M_i = 0$, and set $\psi_i = 0$ otherwise. The weighted terms have finite absolute means: Assumption A1, the conditional independence in Proposition D.1, and Assumption A4 give

$$\begin{aligned} \mathbb{E}\left[\frac{D_i |Y_i|}{e_0(X_i)} \,\middle|\, M_i = 0 \right] &= \mathbb{E}[|Y_i(1)| \mid M_i = 0] < \infty, \\ \mathbb{E}\left[\frac{(1-D_i)|Y_i|}{1-e_0(X_i)} \,\middle|\, M_i = 0 \right] &= \mathbb{E}[|Y_i(0)| \mid M_i = 0] < \infty. \end{aligned}$$

Because $(Y_i(0), Y_i(1), D_i, X_i, \mathbf{M}_i) \overset{\text{i.i.d.}}{\sim} P_{\mathrm{po}}$, as specified in the oracle potential-outcome sample def-

inition, the complete cases are draws from the conditional distribution given $M_i = 0$. Under Assumption A4, the law of large numbers gives

$$\widetilde{\tau}_{\mathrm{IPW}} = \frac{n^{-1}\sum_{i=1}^{n} \mathbf{1}\{M_i = 0\}\psi_i}{n^{-1}\sum_{i=1}^{n} \mathbf{1}\{M_i = 0\}} \xrightarrow{p} \mathbb{E}[\psi_i \mid M_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0].$$

Moreover,

$$|\widehat{\tau}_{\mathrm{IPW}} - \widetilde{\tau}_{\mathrm{IPW}}| \le \max\left\{ \sup_x \left| \frac{1}{\widehat{e}_0(x)} - \frac{1}{e_0(x)} \right|, \sup_x \left| \frac{1}{1-\widehat{e}_0(x)} - \frac{1}{1-e_0(x)} \right| \right\} \frac{1}{n_{\mathrm{cc}}} \sum_{i:M_i=0} |Y_i|.$$

Assumption A5 makes the maximum of the two uniform inverse-weight differences $o_p(1)$. Under the i.i.d. sampling and finite-moment conditions in Assumption A4, the complete-case sample mean $n_{\mathrm{cc}}^{-1}\sum_{i:M_i=0}|Y_i|$ is $O_p(1)$. Thus $\widehat{\tau}_{\mathrm{IPW}} - \widetilde{\tau}_{\mathrm{IPW}} = o_p(1)$, and $\widehat{\tau}_{\mathrm{IPW}}$ converges to the same population contrast.

D.6 Proof of Corollary 1

Under the conditions of Proposition 3, Proposition 1(i) gives $Y_i(d) \perp D_i \mid X_i, M_i$ for $d \in \{0,1\}$, so the IPW identity in Proposition D.1 identifies the complete-case ATE. With $e_0(x) = \bar{e}_0$ constant, treatment is independent of X_i among complete cases, and $P(D_i = 1 \mid M_i = 0) = \mathbb{E}[e_0(X_i) \mid M_i = 0] = \bar{e}_0$. Thus, the distribution of X_i is the same in the treatment and control groups, and the IPW identity reduces to the difference between their average outcomes. For the treated term,

$$\begin{aligned}\mathbb{E}\left[\frac{D_i Y_i}{\bar{e}_0} \middle| M_i = 0 \right] &= \frac{P(D_i = 1 \mid M_i = 0)}{\bar{e}_0} \mathbb{E}[Y_i \mid D_i = 1, M_i = 0] \\ &= \mathbb{E}[Y_i \mid D_i = 1, M_i = 0].\end{aligned}$$

Similarly, the untreated term equals $\mathbb{E}[Y_i \mid D_i = 0, M_i = 0]$. Hence,

$$\mathbb{E}[Y_i \mid D_i = 1, M_i = 0] - \mathbb{E}[Y_i \mid D_i = 0, M_i = 0] = \tau_{\mathrm{cc}}.$$

By the Frisch–Waugh–Lovell theorem, β_{cc} is the coefficient from regressing Y_i on the residual from projecting D_i on $g(X_i)$ among complete cases. Because $\mathbb{E}[D_i \mid X_i, M_i = 0] = \bar{e}_0$ and $g(X_i)$ includes an intercept, that residual is $D_i - \bar{e}_0$. Let $\mu_d = \mathbb{E}[Y_i \mid D_i = d, M_i = 0]$ for $d \in \{0,1\}$. Since

D_i is binary,

$$\begin{aligned}\mathrm{Cov}(D_i, Y_i \mid M_i = 0) &= \bar{e}_0\mu_1 - \bar{e}_0\{\bar{e}_0\mu_1 + (1-\bar{e}_0)\mu_0\} \\ &= \bar{e}_0(1-\bar{e}_0)(\mu_1 - \mu_0), \\ \mathrm{Var}(D_i \mid M_i = 0) &= \bar{e}_0(1-\bar{e}_0).\end{aligned}$$

Subtracting $\bar{e}_0$ from D_i changes neither its covariance with Y_i nor its variance. Thus,

$$\begin{aligned}\beta_{\mathrm{cc}} &= \frac{\mathrm{Cov}(D_i - \bar{e}_0, Y_i \mid M_i = 0)}{\mathrm{Var}(D_i - \bar{e}_0 \mid M_i = 0)} \\ &= \mathbb{E}[Y_i \mid D_i = 1, M_i = 0] - \mathbb{E}[Y_i \mid D_i = 0, M_i = 0] = \tau_{\mathrm{cc}}.\end{aligned}$$

Consistency follows as in Proposition 2. ■

D.7 Proof of Corollary 2

Under a constant unit-level treatment effect, $\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = m] = \tau$. Thus, by the law of total expectation,

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \int \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = m]\, dF_M(m) = \int \tau\, dF_M(m) = \tau.$$

In particular, $\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0] = \tau$. Substituting this equality into Proposition 2 gives $\tau_{\mathrm{cc}} = \tau$ and consistency of $\widehat{\tau}_{\mathrm{cc}}$ under that proposition's conditions. Proposition 3 gives $\beta^* = \tau$ under its own conditions and consistency of $\widehat{\beta}$ under Assumption A5 (estimated inverse-weight consistency). Corollary 1 gives $\beta_{\mathrm{cc}} = \tau_{\mathrm{cc}} = \tau$ and consistency of $\widehat{\beta}_{\mathrm{cc}}$ under its own conditions.

D.8 Proof of Corollary 3

The complete-case support condition in Corollary 3, $P(M_i = 0 \mid W_i = w) > 0$ for F^W_{po}-almost every w, ensures that $\tau_{\mathrm{cc}}(w)$ is defined throughout the population support of W_i up to a set of probability zero. By mean exchangeability given W_i and the law of iterated expectations,

$$\begin{aligned}\int \tau_{\mathrm{cc}}(w)\, dF^W_{\mathrm{po}}(w) &= \int \mathbb{E}[Y_i(1) - Y_i(0) \mid W_i = w]\, dF^W_{\mathrm{po}}(w) \\ &= \mathbb{E}[Y_i(1) - Y_i(0)] = \tau.\end{aligned}$$

For the observed-outcome representation, Assumption A1 (SUTVA), Proposition 1(i) (conditional ignorability given the full adjustment vector), and Assumption A2 (complete-case treatment overlap)

give, for $d \in \{0,1\}$,

$$\begin{aligned}\mathbb{E}[Y_i \mid D_i = d, X_i = x, M_i = 0] &= \mathbb{E}[Y_i(d) \mid D_i = d, X_i = x, M_i = 0] \\ &= \mathbb{E}[Y_i(d) \mid X_i = x, M_i = 0].\end{aligned}$$

Subtracting the expression for $d = 0$ from that for $d = 1$ gives $\delta_{\mathrm{cc}}(x)$. Since W_i is a subvector of X_i, iterated expectations within complete cases give

$$\begin{aligned}\int \delta_{\mathrm{cc}}(x)\, dF_{\mathrm{cc}}(x \mid W_i = w) &= \mathbb{E}[Y_i(1) - Y_i(0) \mid W_i = w, M_i = 0] \\ &= \tau_{\mathrm{cc}}(w).\end{aligned}$$

This inner averaging uses only the covariate support among complete cases within each $W_i = w$ group. The transport and support assumptions at the W_i level then justify the outer population average. When $W_i = X_i$, the inner distribution is concentrated at $x = w$, recovering the full-covariate result.

D.9 Complete-case IPW regression

The following result applies Gabriel et al. (2024, Section 3 and Supporting Information S1) to complete cases.

Assumption A6 (Complete-case regression moments)**.** $\mathbb{E}[Y_i(d)^2 \mid M_i = 0] < \infty$ *for* $d \in \{0,1\}$, $\mathbb{E}[\|X_i\|^2 \mid M_i = 0] < \infty$*, and* $\mathbb{E}[(1, X_i^\top)^\top (1, X_i^\top) \mid M_i = 0]$ *is positive definite.*

Proposition D.2 (Complete-case IPW regression)**.** *Let* $\omega_i = D_i / e_0(X_i) + (1 - D_i)/\{1 - e_0(X_i)\}$*, and let* $\widehat{\omega}_i$ *be its analogue with* $\widehat{e}_0$ *in place of* e_0*. Let* $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ *minimize*

$$\sum_{i:M_i=0} \widehat{\omega}_i \left(Y_i - \alpha - \beta D_i - X_i^\top \gamma\right)^2.$$

Suppose the conditions of Proposition D.1 and Assumption A6 (complete-case regression moments) hold. Then the population weighted least-squares coefficient on D_i *among complete cases equals* $\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0]$*, whether or not* $\mathbb{E}[Y_i \mid D_i, X_i, M_i = 0]$ *is linear in* D_i *and* X_i*. Under Assumption A5,* $\widehat{\beta}$ *converges in probability to* $\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0]$.

Proof of Proposition D.2 Let $Z_i = (1, D_i, X_i^\top)^\top$ be the regressor vector and $\theta = (\alpha, \beta, \gamma^\top)^\top$ be the coefficient vector.

Step 1: The population coefficients exist and are unique. Let $\phi(X_i)$ be any function with $\mathbb{E}[|\phi(X_i)| \mid M_i = 0] < \infty$. Because $\mathbb{E}[D_i \mid X_i, M_i = 0] = e_0(X_i)$, treatment overlap gives

$$\mathbb{E}[\omega_i D_i \phi(X_i) \mid X_i, M_i = 0] = \frac{\phi(X_i)}{e_0(X_i)} \mathbb{E}[D_i \mid X_i, M_i = 0] = \phi(X_i).$$

The same calculation with $1 - D_i$ gives $\mathbb{E}[\omega_i (1 - D_i)\phi(X_i) \mid X_i, M_i = 0] = \phi(X_i)$. Taking expectations over $X_i \mid M_i = 0$ yields

$$\mathbb{E}[\omega_i D_i \phi(X_i) \mid M_i = 0] = \mathbb{E}[\omega_i (1 - D_i)\phi(X_i) \mid M_i = 0] = \mathbb{E}[\phi(X_i) \mid M_i = 0]. \tag{D.1}$$

For a treated complete case, ω_i is the inverse treatment probability $1/e_0(X_i)$, while for an untreated complete case, it is $1/\{1 - e_0(X_i)\}$. Equation (D.1) shows that weighting either treatment arm this way reproduces the covariate moments of the complete-case population.

Hence $Q = \mathbb{E}[\omega_i Z_i Z_i^\top \mid M_i = 0]$ is finite under Assumption A6.

For any $a = (a_0, a_1, a_2^\top)^\top$, (D.1) gives

$$\begin{aligned} a^\top Q a &= \mathbb{E}\Big[\omega_i (a^\top Z_i)^2 \mid M_i = 0\Big] \\ &= \mathbb{E}\left[\frac{D_i}{e_0(X_i)}(a_0 + a_1 + a_2^\top X_i)^2 + \frac{1 - D_i}{1 - e_0(X_i)}(a_0 + a_2^\top X_i)^2 \,\middle|\, M_i = 0\right] \\ &= \mathbb{E}\Big[(a_0 + a_1 + a_2^\top X_i)^2 \mid M_i = 0\Big] + \mathbb{E}\Big[(a_0 + a_2^\top X_i)^2 \mid M_i = 0\Big]. \end{aligned}$$

The first equality follows from $Q = \mathbb{E}[\omega_i Z_i Z_i^\top \mid M_i = 0]$ and $a^\top Z_i Z_i^\top a = (a^\top Z_i)^2$. The second substitutes the definitions of Z_i and ω_i: when $D_i = 1$, $a^\top Z_i = a_0 + a_1 + a_2^\top X_i$ and $\omega_i = 1/e_0(X_i)$; when $D_i = 0$, $a^\top Z_i = a_0 + a_2^\top X_i$ and $\omega_i = 1/\{1 - e_0(X_i)\}$.

The expression for $a^\top Q a$ is the conditional expectation of two squared quantities. Each square is nonnegative. If their sum has expectation zero, both must be zero almost surely among complete cases. Thus

$$a_0 + a_1 + a_2^\top X_i = 0 \quad \text{and} \quad a_0 + a_2^\top X_i = 0 \qquad \text{almost surely given } M_i = 0.$$

Subtracting the second equation from the first gives $a_1 = 0$.

By the positive-definiteness condition in Assumption A6, $a_0 + a_2^\top X_i = 0$ almost surely implies $a_0 = 0$ and $a_2 = 0$. Hence $a = 0$, so Q is positive definite.

Let $s_i = (1 + D_i^2 + X_i^\top X_i)^{1/2}$, the length of the regressor vector Z_i. Similarly, we have

$$\begin{aligned}\mathbb{E}[\omega_i s_i |Y_i| \mid M_i = 0] &= \sum_{d\in\{0,1\}} \mathbb{E}\Big[(1 + d^2 + X_i^\top X_i)^{1/2} |Y_i(d)| \,\Big|\, M_i = 0\Big] \\ &\le \sum_{d\in\{0,1\}} \Big\{\mathbb{E}[1 + d^2 + X_i^\top X_i \mid M_i = 0]\Big\}^{1/2} \big\{\mathbb{E}[Y_i(d)^2 \mid M_i = 0]\big\}^{1/2} < \infty,\end{aligned}$$

For the equality, SUTVA gives $Y_i = Y_i(d)$ when $D_i = d$, and then $s_i = (1 + d^2 + X_i^\top X_i)^{1/2}$. Conditional independence of $Y_i(d)$ and D_i given $X_i, M_i = 0$ lets the treatment probability $e_0(X_i)$ for $d = 1$, or $1 - e_0(X_i)$ for $d = 0$, cancel its inverse in ω_i. The inequality applies the Cauchy–Schwarz inequality to $(1 + d^2 + X_i^\top X_i)^{1/2}$ and $|Y_i(d)|$ for each d. The final bound is finite by Assumption A6. Because each component of Z_i satisfies $|Z_{ij}| \le s_i$, this bound makes every component of $\mathbb{E}[\omega_i Z_i Y_i \mid M_i = 0]$ finite. Thus $q = \mathbb{E}[\omega_i Z_i Y_i \mid M_i = 0]$ is finite.

The population weighted least-squares coefficient vector minimizes the expected weighted squared error

$$\mathscr{L}(\theta) = \mathbb{E}[\omega_i (Y_i - Z_i^\top \theta)^2 \mid M_i = 0].$$

Expanding the square and using the definitions of Q and q gives

$$\begin{aligned}\mathscr{L}(\theta) &= \mathbb{E}[\omega_i Y_i^2 \mid M_i = 0] - 2\theta^\top \mathbb{E}[\omega_i Z_i Y_i \mid M_i = 0] + \theta^\top \mathbb{E}[\omega_i Z_i Z_i^\top \mid M_i = 0]\theta \\ &= \mathbb{E}[\omega_i Y_i^2 \mid M_i = 0] - 2\theta^\top q + \theta^\top Q \theta.\end{aligned}$$

The first term is finite because the same weighting argument gives $\mathbb{E}[\omega_i Y_i^2 \mid M_i = 0] = \sum_{d\in\{0,1\}} \mathbb{E}[Y_i(d)^2 \mid M_i = 0] < \infty$. Since Q and q are also finite, $\mathscr{L}(\theta)$ is finite for every θ.

The first-order condition gives $Q\theta = q$. Because Q is positive definite, the objective is strictly convex and its unique minimizer is $\theta^* = (\alpha^*, \beta^*, \gamma^{*\top})^\top = Q^{-1} q$. Substituting the definitions of Q and q into $Q\theta^* = q$ gives

$$\mathbb{E}[\omega_i Z_i (Y_i - Z_i^\top \theta^*) \mid M_i = 0] = 0. \tag{D.2}$$

Step 2: The coefficient on D_i is the complete-case ATE.

Let $r_i = Y_i - \alpha^* - \beta^* D_i - X_i^\top \gamma^*$, where $\theta^* = (\alpha^*, \beta^*, \gamma^{*\top})^\top = Q^{-1}q$ is the population weighted least-squares coefficient vector.

Since $Z_i = (1, D_i, X_i^\top)^\top$, equation (D.2) can be written as

$$\mathbb{E}\left[\omega_i \begin{pmatrix} r_i \\ D_i r_i \\ X_i r_i \end{pmatrix} \middle| M_i = 0\right] = 0.$$

Its first two components give $\mathbb{E}[\omega_i r_i \mid M_i = 0] = 0$ and $\mathbb{E}[\omega_i D_i r_i \mid M_i = 0] = 0$. Hence $\mathbb{E}[\omega_i (1 - D_i) r_i \mid M_i = 0] = 0$.

As in the proof of Proposition D.1, applying IPW to the regression residuals, together with SUTVA and $Y_i(d) \perp D_i \mid X_i, M_i = 0$, gives, for each supported x,

$$\begin{aligned} \mathbb{E}[\omega_i D_i r_i \mid X_i = x, M_i = 0] &= \mathbb{E}[Y_i(1) \mid X_i = x, M_i = 0] - \alpha^* - \beta^* - x^\top \gamma^*, \\ \mathbb{E}[\omega_i (1 - D_i) r_i \mid X_i = x, M_i = 0] &= \mathbb{E}[Y_i(0) \mid X_i = x, M_i = 0] - \alpha^* - x^\top \gamma^*. \end{aligned}$$

For the first equality, untreated units contribute zero because $D_i = 0$. Among treated units at $X_i = x$, the residual is $Y_i(1) - \alpha^* - \beta^* - x^\top \gamma^*$ and the weight is $1/e_0(x)$. Since the treated fraction is $e_0(x)$, the fraction and its inverse cancel. Conditional ignorability then replaces the treated mean of $Y_i(1)$ with its mean among all complete cases at x. Thus, at fixed x, the weighted treated residual mean equals the mean treated potential outcome minus the treated fitted value $\alpha^* + \beta^* + x^\top \gamma^*$. The same reasoning applies to the untreated term.

Averaging over $X_i \mid M_i = 0$ and setting both expressions to zero gives

$$\begin{aligned} \mathbb{E}[Y_i(1) \mid M_i = 0] &= \alpha^* + \beta^* + \mathbb{E}[X_i \mid M_i = 0]^\top \gamma^*, \\ \mathbb{E}[Y_i(0) \mid M_i = 0] &= \alpha^* + \mathbb{E}[X_i \mid M_i = 0]^\top \gamma^*. \end{aligned}$$

Subtracting yields $\beta^* = \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0]$.

We note here that the argument does not assume a linear model for $\mathbb{E}[Y_i \mid D_i, X_i, M_i = 0]$. To identify β^*, it uses only the intercept and D_i components of (D.2). Further, the terms involving γ^* cancel when the treated and untreated equations are subtracted.

Step 3: Consistency. Let $\widetilde{Q}$ and $\widetilde{q}$ be the complete-case sample averages of $\omega_i Z_i Z_i^\top$ and $\omega_i Z_i Y_i$, and let $\widehat{Q}$ and $\widehat{q}$ be the same averages with $\widehat{\omega}_i$ in place of ω_i. Whenever $\widehat{Q}$ is invertible, $\widehat{\theta} = \widehat{Q}^{-1}\widehat{q}$. By Step 1 these weighted terms have finite means. As in the proof of Proposition D.1, the law of large numbers gives $\widetilde{Q} \xrightarrow{p} Q$ and $\widetilde{q} \xrightarrow{p} q$. Let

$$r_n = \max\left\{ \sup_x \left| \frac{1}{\widehat{e}_0(x)} - \frac{1}{e_0(x)} \right|, \sup_x \left| \frac{1}{1-\widehat{e}_0(x)} - \frac{1}{1-e_0(x)} \right| \right\},$$

which is $o_p(1)$ by Assumption A5. Since $|\widehat{\omega}_i - \omega_i| \le r_n$,

$$\|\widehat{Q} - \widetilde{Q}\| \le r_n \frac{1}{n_{\text{cc}}} \sum_{i:M_i=0} s_i^2, \qquad \|\widehat{q} - \widetilde{q}\| \le r_n \frac{1}{n_{\text{cc}}} \sum_{i:M_i=0} s_i |Y_i|.$$

Both averages are $O_p(1)$ by Assumption A6 and the Cauchy–Schwarz inequality, so both differences are $o_p(1)$. Hence $\widehat{Q} \xrightarrow{p} Q$ and $\widehat{q} \xrightarrow{p} q$. Because Q is invertible, $\widehat{Q}$ is invertible with probability tending to one. The continuous mapping theorem then gives $\widehat{\theta} \xrightarrow{p} \theta^*$, and in particular $\widehat{\beta} \xrightarrow{p} \mathbb{E}[Y_i(1) - Y_i(0) \mid M_i = 0]$.

E Simulation of the Collider–Mediator Dilemma

We assume the following data generation process, where the total causal effect of D on Y is $0.8 \times 1.5 + 0.8 = 2$, while the controlled direct effect (Imai et al. 2011) of D on Y is 0.8. The outcome Y is missing according to the latent index $L_i = 1.5D_i + 2A_i + \eta_i$. The threshold c in the missingness determination equation is selected such that the expected probability of missingness is 30%, i.e. $\mathbb{E}[M_i] = \mathbb{P}(L_i \ge c) = 0.3$.

$$\begin{aligned}
D_i &\sim \mathit{Bernoulli}(0.5) \\
A_i &= 1 + 0.8D_i + u_i, \qquad u_i \sim N(0,1) \\
Y_i &= 1 + 0.8D_i + 1.5A_i + \zeta_i, \qquad \zeta_i \sim N(0,1) \\
M_i &= I\{1.5D_i + 2A_i + \eta_i \ge c\}, \qquad \eta_i \sim N(0,1)
\end{aligned}$$

We draw 10,000 samples of 5,000 observations. As in the main-text illustration, we estimate the oracle regression, complete-case regression with and without A_i, and multiple imputation. Here, the

synthetic models use the variables in the DGP above, without additional baseline controls. For MI, we impute Y using (D,A), or D using (Y,A), and then regress Y on D in the imputed datasets (Little and Rubin 2019).

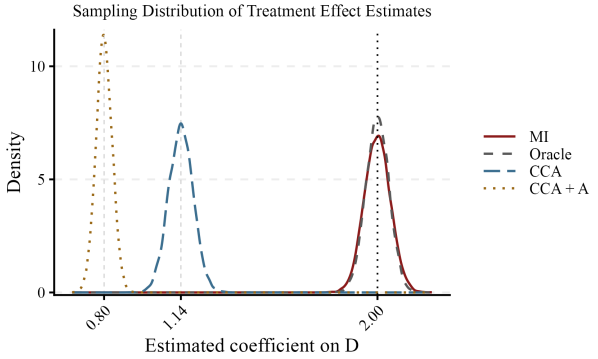


Figure E1: Sampling Distribution of the Estimated Total Effect of D on Y: Missing Y

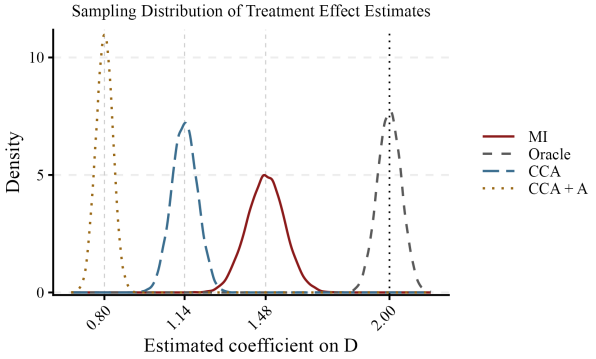


Figure E2: Sampling Distribution of the Estimated Total Effect of D on Y: Missing D

Simulation Results When Y is missing (Figure E1), neither complete-case regression estimates the total effect consistently: without A, selection opens a noncausal path; controlling for A blocks that path but also blocks the mediated effect. CCA+A instead estimates the controlled direct effect, while MI recovers the total effect under the specified MAR mechanism. When D is missing under the same index (Figure E2), neither complete-case regression nor MI recovers the total effect; missingness in D depends on its own value.

F Reweighting for Generalized Linear Models

The distinction between conditional comparisons and their aggregation also applies to generalized linear models. Suppose the oracle conditional mean is

$$m_{\text{po}}(d,x) := \mathbb{E}[Y_i \mid D_i = d, X_i = x] = \ell^{-1}\left(\alpha + \beta_D d + x^\top \gamma\right).$$

For continuous D_i, the marginal effect is

$$a_{\text{po}}(d,x) := \frac{\partial m_{\text{po}}(d,x)}{\partial d} = \beta_D \left(\ell^{-1}\right)'\left(\alpha + \beta_D d + x^\top \gamma\right).$$

Although β_D is constant, the outcome-scale marginal effect generally varies with d and x. The oracle response function has a causal interpretation if X_i satisfies the adjustment criterion and SUTVA and treatment overlap hold.

Let $m_{\text{cc}}(d,x) := \mathbb{E}[Y_i \mid D_i = d, X_i = x, M_i = 0]$ and $a_{\text{cc}}(d,x) := \partial m_{\text{cc}}(d,x)/\partial d$. With H_{cc} and H_{po} denoting the complete-case and population distributions of (D_i, X_i), the corresponding average marginal effects are

$$\theta_{\text{cc}} = \int a_{\text{cc}}(d,x)\, dH_{\text{cc}}(d,x), \qquad \theta_{\text{po}} = \int a_{\text{po}}(d,x)\, dH_{\text{po}}(d,x).$$

As in Subsection 5.1, recovering the population average requires both transport of the conditional effects and averaging over the population distribution. A sufficient condition for transporting the marginal-effect function is

$$\mathbb{E}[Y_i \mid D_i = d, X_i = x, M_i = 0] = \mathbb{E}[Y_i \mid D_i = d, X_i = x],$$

together with complete-case support; equality of the response functions implies equality of their derivatives where these exist. We can then fit a correctly specified GLM among complete cases and average its fitted marginal effects over H_{po}. If regressors are missing, their target distribution must be recovered, as discussed after Corollary 3. Consistent estimation of the GLM coefficients alone does not ensure consistent estimation of the population AME.

For binary treatment, use the probability contrast $\delta_{\text{cc}}(x) = m_{\text{cc}}(1,x) - m_{\text{cc}}(0,x)$ in place of a derivative. Mean exchangeability of selection and treatment effect heterogeneity can transport this contrast without requiring equality of both response levels. Its population average then follows the same averaging argument as in Subsection 5.1.

G Estimates for the Reanalysis of Bodea and Hicks (2015)

Table G1: Reanalysis of Bodea and Hicks (2015): Supporting Estimates

Panel A: CBI effect on inflation without a moderator

Version	Countries in dataset	CC countries	CC observations	Coefficient per 1.0 CBI (SE)	Difference, expanded minus original
Original data	79	78	1,769	−0.465 (0.143)	−0.081 [−0.182, 0.002] $p = 0.054$
Expanded data (cbi144)	132	115	2,198	−0.546 (0.154)	

Panel B: Average effect by moderator, expanded data, per 0.1 CBI

			Updated-CBI countries			All countries			
Moderator	Observations (countries)	ATE-CC	Weighted	Difference	p-value Raw Holm	Weighted	Difference	p-value Raw Holm	Thinnest support
Polity	2,198 (115)	−0.030 [−0.079, 0.006]	0.004 [−0.061, 0.050]	+0.034 [0.007, 0.060]	0.020 0.060	0.018 [−0.056, 0.071]	+0.048 [0.010, 0.082]	0.019 0.057	24.6%
Freedom House	2,231 (116)	−0.048 [−0.086, −0.018]	−0.024 [−0.065, 0.012]	+0.024 [0.009, 0.045]	0.002 0.012	−0.005 [−0.051, 0.037]	+0.043 [0.018, 0.074]	0.002 0.012	33.3%
Executive constraints	2,174 (115)	−0.031 [−0.074, −0.001]	−0.007 [−0.059, 0.034]	+0.024 [0.006, 0.046]	0.007 0.028	0.005 [−0.053, 0.051]	+0.036 [0.010, 0.065]	0.005 0.020	37.0%
Checks and balances	2,073 (116)	−0.072 [−0.115, −0.030]	−0.077 [−0.124, −0.028]	−0.005 [−0.018, 0.011]	0.591 0.591	−0.078 [−0.128, −0.027]	−0.007 [−0.025, 0.014]	0.578 0.578	38.2%
Political constraints	2,273 (116)	−0.042 [−0.075, −0.010]	−0.017 [−0.055, 0.035]	+0.025 [0.009, 0.053]	0.003 0.015	−0.006 [−0.047, 0.057]	+0.036 [0.014, 0.076]	0.003 0.015	32.8%
Press freedom	1,933 (116)	−0.017 [−0.063, 0.016]	−0.007 [−0.056, 0.029]	+0.011 [0.000, 0.022]	0.050 0.100	0.006 [−0.048, 0.043]	+0.023 [0.003, 0.040]	0.029 0.058	29.0%

Panel C: Polity categories, expanded data

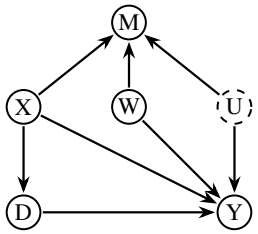
Category	CC observations	Share, CC	Share, cbi144	Share, worldwide	Slope per 0.1 CBI	Countries (with CBI variation)
Autocracy (−10 to −6)	282	12.8%	29.7%	36.1%	+0.140	38 (11)
Anocracy (−5 to 5)	439	20.0%	21.9%	23.7%	−0.033	57 (14)
Democracy (6 to 10)	1,477	67.2%	48.4%	40.2%	−0.062	83 (53)

Notes: CC means complete-case. The outcome is the signed log of inflation. All models include country fixed effects and the controls of Bodea and Hicks (2015): lagged log inflation, lagged money growth, lagged log GDP, lagged GDP growth, lagged trade openness, a de jure exchange-rate regime indicator, lagged fiscal balance, election indicators, and lagged world inflation. Panel A also includes Polity and reports coefficients and their difference per unit of CBI. The original-data estimate reproduces their Table 1, Model 5. The expanded data add country-years from countries covered by the updated CBI data, with controls rebuilt from public sources. The updated-CBI country target is labeled cbi144; the updated CBI data cover 143 countries and the European Central Bank. The expanded data exclude updated-CBI countries that lack Polity or CBI scores in 1973–2008, and the ECB has no country code. The ATE-CBI144 and ATE-Polity5 target shares use only country-years in 1973–2008 for which the moderator is observed, so ATE-CBI144 does not cover every updated-CBI country. CC have every control observed. Panel B fits a separate regression for each moderator category and averages the category slopes using the CC shares, the shares among the updated-CBI countries

(cbi144), or the shares across all countries Polity5 covers (worldwide), 1973–2008. Differences are the reweighted minus the CC average. Intervals are 95% percentile intervals from 2,000 country-cluster bootstrap draws, all draws succeeded. Target shares are held fixed in the bootstrap; CC shares are recomputed in each draw. Panel B reports raw and Holm-adjusted two-sided p-values (raw/adjusted), with adjustment across the six moderators separately for each target; Panel A reports its unadjusted p-value. Thinnest support is the smallest share, across a moderator's categories, of countries whose CBI changes within the category. The moderator categories are defined as follows: Polity, -10 to -6, -5 to 5, and 6 to 10; Freedom House average rating, at most 2.5, above 2.5 through 5, and above 5; executive constraints, 1–3, 4–6, and 7; checks and balances, 1–2, 3, and 4 or more; political constraints, at most 0.2, above 0.2 through 0.4, and above 0.4; and press freedom, not free, partly free, and free.

H Treatment-IPW Comparison for the Canvassing Experiment

In another hypothetical example, had the researcher used a randomization scheme in which treatment probabilities differed by participants' place of residence, assignment would depend on a component of X, creating the arrow $X \to D$. Suppose further that follow-up response depended on both X and an unobserved determinant of the outcome, U, as in Figure H1. Adjusting for X would still preserve valid within-X treatment comparisons, satisfying Box 1. However, selection could make the baseline outcome mean nonlinear in X, as in mechanism B of the motivating simulation. A regression controlling only linearly for X could then pool those comparisons incorrectly, so Box 2 would require an additional specification or weighting check.



Note: In this hypothetical design, W is baseline college education, X includes residence and the other baseline controls, and dashed U is unobserved. Treatment probabilities vary with X. Conditioning on $M = 0$ opens a noncausal path, which adjustment for X and W blocks. With treatment overlap, the adjustment criterion is satisfied among complete cases.

Figure H1: A Hypothetical Conditionally Randomized Design with Selection on an Unobserved Outcome Determinant

Table H1: Complete-Case and Probit-IPW Estimates in the Two-Arm Canvassing Sample

Group	CC observations	CC regression	Probit-IPW	Difference (IPW minus CC)
All complete cases	1,172	0.0925 (0.0226) [0.0483, 0.1357]	0.0932 (0.0231) [0.0474, 0.1375]	0.0007 (0.0062) [-0.0112, 0.0134]

Notes: CC means complete-case. The sample includes only the full-intervention and placebo arms of Study 1 in Kalla and Broockman (2023). The outcome is the first-follow-up omnibus index of immigrant attitudes and policy preferences. Estimates are index units; bootstrap standard errors are in parentheses and 95% percentile intervals are in brackets. The CC regression includes the authors' baseline controls except college education. A probit of treatment assignment on education and these controls estimates the complete-case treatment propensity $\widehat{e}_0(W,X)$. IPW estimates come from a regression of the outcome on an intercept and treatment, weighted by $D/\widehat{e}_0(W,X) + (1-D)/[1-\widehat{e}_0(W,X)]$. No weights are trimmed. In 2,000 household-bootstrap draws, households are resampled from the target sample, and both regressions and the probit model are refitted. All draws succeeded. Differences use the same paired draws. These weights address treatment assignment among complete cases, rather than response selection or population education composition. The use of estimated propensities is an illustrative choice; the actual design assigns equal probabilities within matched household blocks.

Table H2: Complete-Case Estimation and Reweighting by Education

	Pooled CC	No college	College
Complete-case treatment effects			
Treatment effect	0.0925	0.0497	0.1107
Bootstrap SE	(0.0226)	(0.0410)	(0.0372)
Complete-case observations	1,172	548	624
Education shares and averages			
Complete-case share	—	46.8%	53.2%
Target-sample share	—	52.5%	47.5%
Average using complete-case shares		0.0822	
Average using target-sample shares		0.0787	
Difference: reweighted minus pooled CC		−0.0138	
95% bootstrap interval for difference		[−0.0488, 0.0174]	

Notes: CC means complete-case. The sample comprises 1,900 participants in the full-intervention and placebo arms who confirmed their identity at the door and were eligible for follow-up surveys, of whom 1,172 have the first-follow-up outcome. Education-specific effects come from separate treatment-IPW regressions within each education group. Within each group, a separate complete-case probit models treatment assignment using the remaining baseline controls. The pooled CC regression omits education and includes the same baseline controls. The difference compares the target-share average of education-specific effects with the pooled CC regression. The complete-case-share average and the pooled CC estimate can differ in a finite sample because they use different specifications. Standard errors and percentile intervals use 2,000 paired household-bootstrap draws from the target sample, refitting the propensity model and regressions and recomputing the education shares. All draws succeeded. No weights are trimmed. Calculations use unrounded values.