Workload-Preserving Differentially Private Synthetic Data for Causal Inference via Maximum-Entropy Calibration

Amir Asiaee[1] **Kaveh Aryan**[2]

[1]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[2]Department of Informatics, King's College London, London, UK

Abstract

Workload-based differentially private (DP) synthetic data methods privately measure aggregate queries and post-process the noisy answers into synthetic records. Generic workloads can achieve strong distributional fidelity, but causal estimands such as the average treatment effect (ATE) depend on treatment-arm balance and outcome moments that generic marginals need not preserve. We propose *causal workloads*: DP query sets designed around the orthogonal moments used by doubly robust causal estimators. The released workload can be used directly by stable moment-map estimators or reconstructed by maximum-entropy calibration into reusable synthetic data; our theory decomposes ATE error into sampling, privacy, workload-approximation, Monte Carlo, and calibration terms. We also introduce CAUSAL-AIM, an adaptive workload selector, and a noise-aware multiple-imputation (NA+MI) procedure for confidence intervals from DP synthetic data. Because the workload is released once, the same DP synthetic table can support ATE, ATT, and subgroup analyses without additional privacy spending. Empirically, causal workloads are most useful at strict privacy budgets and for calibrated uncertainty, while generic workloads often retain an advantage for point RMSE as privacy relaxes. The broader lesson is a tradeoff: distributional fidelity can help point accuracy, but valid causal inference requires preserving causal moments and propagating DP noise rather than treating synthetic rows as real.

1 INTRODUCTION

Differential privacy is increasingly used for public release of sensitive tabular datasets in domains such as health, education, and social science [Dwork and Roth, 2014]. A common deployment pattern is to release DP synthetic microdata so that downstream analysts can reuse existing statistical and machine learning workflows [Liang et al., 2026]. State-of-the-art DP synthesis methods often follow a select–measure–reconstruct pipeline: select a set of low-dimensional queries, measure them privately, and reconstruct a distribution (often maximum entropy / graphical model) from the noisy answers [McKenna et al., 2021, 2022]. However, *causal inference* introduces additional structure and fragility. Even if synthetic data match many marginals, causal estimands can be biased by subtle distortions in overlap, confounding structure, or conditional outcome models. Recent work emphasizes that DP noise inflates the variance of causal estimators and invalidates naive confidence intervals unless it is explicitly modeled [Farzam and Sapiro, 2024, Schröder et al., 2025a,b, Ohnishi and Awan, 2024].

This paper asks a concrete question:

> *Which DP queries must a synthetic data mechanism preserve to enable valid causal inference, and what are the fundamental privacy–causal tradeoffs?*

We answer by designing *causal workloads* and giving explicit bounds for stable workload-based estimators and for the calibrated synthetic-data route.

Key idea: causal workloads as orthogonal moments. Modern causal estimators, including doubly robust (DR) estimators [Robins et al., 1994] and double/debiased machine learning [Chernozhukov et al., 2018], rely on score functions whose expectation identifies the target estimand. We propose to choose a workload that directly measures these orthogonal moments, or a basis approximation thereof, under DP. The released workload can then be used in two ways: the direct q-route evaluates a stable estimator as a function of the released moments, while the synthetic-data route reconstructs a maximum-entropy distribution matched to those moments and samples reusable synthetic records. Our

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:133–160.

theory controls the moment-level error in both routes, with additional calibration and Monte Carlo terms for synthetic-data analysis.

Why synthetic data at all? If the goal is ATE only, one could release a DP ATE estimate directly. Synthetic data is valuable when many analysts ask many questions, or when we want to preserve a whole *class* of causal estimands (ATE, average treatment effect on the treated (ATT) $\mathbb{E}[Y(1) - Y(0) \mid T{=}1]$, subgroup effects, balancing diagnostics, etc.) without repeated access to the private data. Causal workloads formalize this "class of analyses" viewpoint; Section 7 provides direct evidence that a single release supports ATE, ATT, and subgroup analyses with no additional privacy spending, a reuse property structurally unavailable to direct DP ATE estimators.

Our contributions.

1. We define *causal workloads*: DP query sets built from orthogonal-score moments, with two downstream routes: direct moment plug-ins and maximum-entropy synthetic-data reconstruction.
2. We prove finite-sample error bounds that connect ATE error to released-workload error, and decompose synthetic-data error into sampling, privacy, workload-approximation, Monte Carlo, and calibration terms. All proofs are collected in Appendix E.
3. We introduce CAUSAL-AIM, an adaptive causal workload selector, and NA+MI, a noise-aware multiple-imputation procedure for confidence intervals from DP synthetic data.
4. Empirically, Causal + NA+MI is the only private method with near-nominal coverage on all four benchmarks ($99.8{-}100\%$ at $\varepsilon \leq 1$), while naive synthetic-data analyses remain at or below 35.2%; the same release also supports ATE, ATT, and subgroup analyses without extra privacy spending.

2 RELATED WORK

A brief DP-synthesis primer. A mechanism is (ε, δ)-differentially private if replacing one record changes its output distribution by at most a factor e^{ε}, up to slack δ [Dwork and Roth, 2014]. Workload-based DP synthetic-data methods first choose a set of aggregate queries (the *workload*), release noisy answers to those queries, and then reconstruct a distribution whose query answers match the noisy measurements. The final synthetic records are sampled from this reconstructed distribution; reconstruction and sampling are post-processing, so they spend no additional privacy budget. This select–measure–reconstruct template underlies MWEM (Multiplicative Weights Exponential Mechanism; Hardt et al., 2012), which iteratively fixes poorly matched queries; Private-PGM, which fits a graphical model to noisy low-order marginals [McKenna et al., 2019]; MST (Maximum-Spanning-Tree marginal selection with Private-PGM reconstruction; McKenna et al., 2021), which won the NIST synthetic-data challenge; and AIM (Adaptive and Iterative Mechanism; McKenna et al., 2022), which adaptively selects marginals against a target workload.

Workload-based DP synthetic data. A large class of methods measures a set of marginals / low-dimensional queries and reconstructs a distribution by imposing structure such as a graphical-model factorization or an optimization objective over candidate distributions, including Private-PGM [McKenna et al., 2019] and its NIST-MST/MST instantiations [McKenna et al., 2021], the adaptive AIM mechanism [McKenna et al., 2022], and recent optimal-transport based approaches [Donhauser et al., 2024]. Benchmarks suggest these methods can outperform GAN-style synthesis on tabular data for many metrics [Tao et al., 2021, Bowen and Liu, 2020]. These methods are designed for general-purpose fidelity; none explicitly targets causal estimands.

Inference from DP synthetic data. Naively treating DP synthetic data as real can produce invalid inference, e.g. inflated type-I error [Perez et al., 2024]. Noise-aware procedures combine Bayesian modeling of the DP noise with multiple imputation to recover calibrated uncertainty [Räisä et al., 2023]. For record-level synthetic microdata—tables whose rows represent individual units rather than only aggregate summaries—classical multiple-imputation inference dates back to Reiter [2003] and Raghunathan et al. [2003]. Our noise-aware multiple imputation (NA+MI) procedure builds on this line but specializes it to causal estimands by choosing workloads informed by the orthogonal score structure.

Causal inference under DP. Recent papers propose DP causal estimators with point and interval estimation guarantees. PrivATE [Schröder et al., 2025a] develops doubly robust DP confidence intervals for ATE using output perturbation. Ohnishi and Awan [2024] propose DP covariate balancing with finite-sample guarantees. DP-CATE [Schröder et al., 2025b] extends orthogonal learning to heterogeneous treatment effects under DP. Farzam and Sapiro [2024] study causal inference under DP broadly and analyze how DP mechanisms can inflate ATE variance and distort individual-level effects. Earlier, Niu et al. [2022] proposed DP estimation of heterogeneous causal effects, and concurrent work by Lebeda et al. [2025] develops a model-agnostic DP causal framework. Our work focuses instead on *synthetic data* as the release object: rather than releasing a single DP estimate, we release a DP synthetic dataset that supports a whole class of causal analyses.

Double/debiased machine learning. Our causal workload design draws on the orthogonal moment framework

of Chernozhukov et al. [2018], which shows that Neyman-orthogonal scores enable root-n inference for treatment effects when nuisance functions are estimated at slower rates. We use this structure to choose workloads that preserve the moments needed for orthogonal scores.

Causal prior for synthetic data. A related line encodes causal or analysis-specific prior knowledge *into the synthetic generation process itself*: reward-guided generation steers a generator toward a target downstream utility such as regression-coefficient recovery [Jackson et al., 2026], and post-hoc constraint wrappers inject partial causal structure, such as trusted or forbidden edges and monotonicity, into arbitrary tabular generators [Asiaee et al., 2026]. Our objective is different: rather than encoding a prior during generation, we design the DP measurements so that the released dataset preserves a target causal estimand, and we account for the DP noise in downstream inference.

3 PROBLEM SETUP

We observe i.i.d. data $D = \{(X_i, T_i, Y_i)\}_{i=1}^n$, where $T \in \{0, 1\}$ is treatment. The target is the average treatment effect (ATE) $\tau := \mathbb{E}[Y(1) - Y(0)]$.

Assumption 1 (Unconfoundedness and positivity)**.** $(Y(1), Y(0)) \perp T \mid X$ and the propensity score [Rosenbaum and Rubin, 1983] $e(x) := \Pr(T = 1 \mid X = x)$ satisfies $e(x) \in [\eta, 1 - \eta]$ almost surely for some $\eta > 0$.

Under Assumption 1, $\tau = \mathbb{E}[m_1(X) - m_0(X)]$, where $m_t(x) := \mathbb{E}[Y \mid T = t, X = x]$.

4 CAUSAL WORKLOADS AND MAXIMUM-ENTROPY CALIBRATION

Figure 1 gives an overview of the full release-and-inference procedure.

4.1 A WORKLOAD THAT TARGETS ORTHOGONAL MOMENTS

Let $\phi : \mathcal{X} \to \mathbb{R}^p$ be a feature map (e.g. spline basis, random Fourier features, tree-based leaves, or indicators for a discretization of X). Define the *moment vectors*

$$q_t^{(0)}(D) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{T_i = t\}\, \phi(X_i) \in \mathbb{R}^p, \tag{1}$$

$$q_t^{(1)}(D) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{T_i = t\}\, Y_i\, \phi(X_i) \in \mathbb{R}^p. \tag{2}$$

Here, $q_t^{(0)}$ records treatment-arm feature masses, and $q_t^{(1)}$ records outcome-weighted feature moments. As a simple bin-indicator example, each coordinate of $q_t^{(1)}$ is the joint arm-bin mass times the mean outcome in that bin, with the mass scaled by $1/n$. These four blocks are the private measurements used in the experimental pipeline. Second-order Gram moments are useful for some direct regression plug-ins, but they are not part of the default release; Appendix A gives the feature-encoding example and the extension.

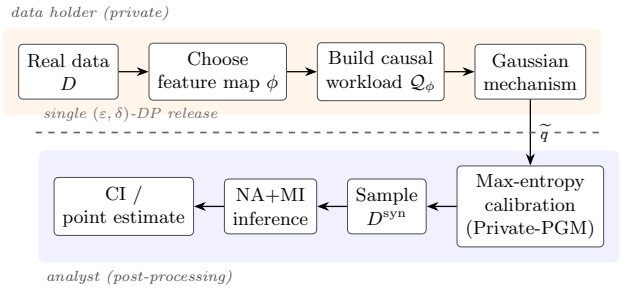

Figure 1: End-to-end pipeline. The data holder chooses the feature map ϕ, builds the causal workload, and releases Gaussian-mechanism-noised moments once. Downstream analysis is post-processing: one may use the noisy moments directly in a stable moment-map estimator, or reconstruct a maximum-entropy distribution (via Private-PGM), sample synthetic data, and run NA+MI inference.

Definition 1 (Causal workload)**.** Fix a feature map ϕ. The *causal workload* $\mathcal{Q}_\phi$ used in our main procedure is the ordered collection of query functions whose answers are $\{q_0^{(0)}, q_1^{(0)}, q_0^{(1)}, q_1^{(1)}\}$. Its stacked answer is

$$q(D) := \left[q_0^{(0)}(D), q_1^{(0)}(D), q_0^{(1)}(D), q_1^{(1)}(D)\right] \in \mathbb{R}^{4p}.$$

Here "workload" means the collection of DP queries, while $q(\cdot)$ denotes the realized answer vector.

This base $4p$ workload is what we release in the experiments. It is *one principled choice*, not a claimed optimum: it targets the first-order treatment and outcome moments used by orthogonal-score inference on nuisances projected onto the ϕ basis (Proposition 2, Appendix B), without asserting minimax optimality, uniqueness, or a lower bound.

4.2 PRIVATE MEASUREMENT OF THE WORKLOAD

Let $W = (X, T, Y)$ and let $m = 4p$ be the dimension of the causal workload $q(D)$ defined above. For $a \in [m]$, let $h_a(W_i)$ denote record i's contribution to coordinate a, so $q_a(D) = n^{-1} \sum_i h_a(W_i)$; for any distribution P over W, $q_a(P) = \mathbb{E}_P h_a(W)$. We use $q(D)$ for the exact, non-released empirical answer on the confidential data; $q^\star = q(P^\star)$ for the population answer under the true data-generating law $P^\star$; and $\widetilde{q}$ for the DP noisy release. The fixed causal mechanism measures all m coordinates once with

the Gaussian mechanism:

$$\widetilde{q} = q(D) + Z, \qquad Z \sim \mathcal{N}(0, \sigma^2 I_m).$$

Assume Y is clipped to $[-B, B]$ and $\|\phi(X)\|_2 \le \phi_{\max}$. Let $\Delta_2 := \sup_{D \sim D'} \|q(D) - q(D')\|_2$ be the replacement-adjacency ℓ_2 sensitivity. $\Delta_2 \le \frac{2\phi_{\max}\sqrt{1+B^2}}{n}$. The standard Gaussian mechanism [Dwork and Roth, 2014] uses $\sigma = \frac{\Delta_2\sqrt{2\log(1.25/\delta)}}{\varepsilon}$. For an adaptive mechanism, the same notation is applied to the measured coordinate set $S \subset [m]$: $\widetilde{q}_S$ denotes the stored noisy answers on S.

4.3 TWO ROUTES FROM THE RELEASED WORKLOAD FOR CAUSAL ESTIMATION

The DP object produced by measurement is a noisy aggregate vector, not yet a synthetic dataset. It can be used in two ways. In the *direct moment route* (or q-route), an estimator that can be written as a stable function $\widehat{\tau}_{\text{est}}(q)$ is evaluated directly at $\widetilde{q}$ (or at the measured coordinates $\widetilde{q}_S$); Section 5 analyzes this route because it exposes how moment error becomes causal-estimation error. In the *synthetic-data route*, we reconstruct a distribution P^{syn} whose workload answers match the noisy measurements, sample synthetic records from that distribution, and then run standard causal estimators such as DR / augmented inverse-propensity weighting (AIPW) and NA+MI. The experiments use this synthetic-data route; the theory controls the part of this route that is expressible through the reconstructed workload answer $q(P^{\text{syn}})$, plus finite synthetic-sampling error. The exponential-family form below is precisely the reconstruction step: it turns noisy moments into a distribution, and sampling from that distribution produces the synthetic data. We present each route in two versions: a fixed mechanism that measures all queries at once ($S = [m]$), and an adaptive mechanism, CAUSAL-AIM, that builds S over rounds.

4.4 MAXIMUM-ENTROPY RECONSTRUCTION FOR SYNTHETIC DATA

Let $\mathcal{P}$ be a class of distributions over $W = (X, T, Y)$ (discrete support or parametric). Given a measured coordinate set $S \subset [m]$ (with $S = [m]$ for the fixed full workload), we define the calibrated synthetic distribution as an information projection:

$$P^{\text{syn}} = \arg\min_{P \in \mathcal{P}} \mathrm{KL}(P \,\|\, P_0) \text{ s.t. } \forall a \in S,\ q_a(P) = \widetilde{q}_a, \quad (3)$$

where P_0 is a reference distribution, such as a uniform distribution on the discretized support, an independent product of noisy one-way marginals, or a public population prior. In the discrete case, the solution is an exponential family:

$$p^{\text{syn}}(w) \propto p_0(w) \exp\Big(\sum_{a \in S} \xi_a h_a(w)\Big)$$

for Lagrange multipliers ξ. Appendix C gives the derivation and the relaxed form used when noisy constraints are infeasible. This is the same mathematical object used in Private-PGM [McKenna et al., 2019, 2021]. Thus reconstruction means fitting the maximum-entropy distribution p_ξ whose measured workload answers match $\widetilde{q}_S$ up to the DP noise scale; synthetic-data generation is the subsequent sampling of records from p_ξ. The fixed synthesis procedure is given in Appendix C.

4.5 ADAPTIVE QUERY SELECTION: CAUSAL-AIM

AIM selects queries adaptively to reduce worst-case error in a generic workload [McKenna et al., 2022]. The fixed mechanism measures the whole causal workload once, so $S = [m]$. CAUSAL-AIM is the adaptive variant: it builds S over rounds by selecting the next feature group to measure using a private estimate of causal utility. Here a candidate ϕ_r can be a single feature or a group such as all bins of one covariate; let $I(r) \subset [m]$ denote the corresponding coordinates of the released workload. At iteration k, CAUSAL-AIM (i) starts from the current measured set S_{k-1} and its reconstructed model P_{k-1}, (ii) evaluates a private upper bound on ATE error using an orthogonal score residual $\psi(W; \theta, \zeta)$, where ψ is the ATE influence function and ζ collects nuisance functions, and (iii) selects the next feature group ϕ_{r_k}, or its corresponding moment coordinates $I(r_k)$, before refitting to obtain P_k; Appendix D gives details. Selected groups are removed from the candidate set: once $I(r)$ has been measured, its noisy answer is reused in later refits rather than measured again. After any round, the stored noisy vector $\widetilde{q}_{S_k}$ can be used directly in a moment-map estimator or passed through maximum-entropy reconstruction to generate synthetic data; Algorithm 1 displays the synthetic-data route used in our experiments. This yields a causal-utility-driven workload that is usually much smaller than "all marginals up to order 2".

5 THEORY: CAUSAL ERROR BOUNDS AND TRADEOFFS

This section tracks how error in the released workload affects ATE estimation along the two routes. We first analyze direct moment plug-ins, where the estimator is a function of the released workload, and then add the reconstruction-calibration and finite-synthetic-sample terms needed for the synthetic-data route. Thus the theoretical object is an estimator with an explicit workload argument, $\widehat{\tau}_{\text{est}}(q)$; this is narrower than an arbitrary downstream learner run on synthetic rows. We state the main decomposition for the ATE; the same weighted-estimand argument gives ATT and subgroup effects when the corresponding weights are represented by the workload (Appendix E.1).

Algorithm 1 CAUSAL-AIM (high-level)

Require: Data D, privacy budget (ε, δ), feature groups $\mathcal{R}$, iterations K, reference distribution P_0

1: Initialize selected coordinates $S_0 \leftarrow \emptyset$, available set $A_0 \leftarrow \mathcal{R}$, and current reconstruction $P^{(0)} \leftarrow P_0$
2: **for** $k = 1, \ldots, \min(K, |\mathcal{R}|)$ **do**
3: Compute an internal utility $u_k(r)$ for each unmeasured group $r \in A_{k-1}$, estimating the ATE-error reduction from adding coordinates $I(r)$
4: Select r_k using the exponential mechanism applied to $\{u_k(r) : r \in A_{k-1}\}$, then update $S_k \leftarrow S_{k-1} \cup I(r_k)$ and $A_k \leftarrow A_{k-1} \setminus \{r_k\}$
5: Measure only the new coordinates $I(r_k)$ with the Gaussian mechanism and store their noisy answers
6: Refit P_k by solving (3) on all stored noisy coordinates S_k
7: **end for**
8: Output stored noisy moments $\widetilde{q}_{S_K}$; for synthetic release, output P_K and sample $D^{\mathrm{syn}} \sim P_K$ (post-processing)

5.1 BOUNDING ATE ERROR BY MOMENT ERROR

The two routes in Section 4 share the same first question: how accurately does DP release the workload moments? For the direct moment route, the second question is how sensitive an ATE estimator is to perturbing those moments. For the second question, call a direct moment-map estimator $\widehat{\tau}_{\mathrm{est}}(q)$ L_{est}-*stable* on a workload domain if

$$|\widehat{\tau}_{\mathrm{est}}(q) - \widehat{\tau}_{\mathrm{est}}(q')| \le L_{\mathrm{est}} \left\|q - q'\right\|_\infty \quad \forall \text{admissible } q, q'.$$

Here q may be the population answer $q^\star = q(P^\star)$, the empirical answer $q(D)$, the DP release $\widetilde{q}$, or the reconstructed answer $q(P^{\mathrm{syn}})$. The formal stability examples below analyze the direct route; the later decomposition composes the same stability logic with the calibration and Monte Carlo terms that appear when the released workload is routed through maximum-entropy reconstruction and synthetic sampling. This separates the generic moment-level guarantee from software-specific choices about the reconstruction solver and downstream nuisance fitting.

Example 1: projected ridge plug-in. Some direct regression plug-ins require the arm-specific Gram moment $G_t(D) := n^{-1} \sum_i \mathbb{1}\{T_i = t\} \phi(X_i)\phi(X_i)^\top$; write $q_t^{(2)}(D) := \mathrm{vec}\{G_t(D)\}$. For a workload-answer vector containing the needed blocks, write $G_t(q)$ for the Gram block, $r_t(q) = q_t^{(1)}$, and $\bar{\phi}(q) = q_0^{(0)} + q_1^{(0)}$. Here "projection" means the ridge least-squares approximation of the arm-specific outcome regression $m_t(x)$ in the span of ϕ. The arm-t coefficient is $\widehat{\beta}_{t,\lambda}(q) = \{G_t(q) + \lambda I\}^{-1} r_t(q)$, with corresponding projected conditional average treatment effect (CATE) and ATE:

$$\widehat{\tau}_\lambda(x; q) := \phi(x)^\top \{\widehat{\beta}_{1,\lambda}(q) - \widehat{\beta}_{0,\lambda}(q)\}, \tag{4}$$

$$\widehat{\tau}_\lambda(q) := \bar{\phi}(q)^\top \{\widehat{\beta}_{1,\lambda}(q) - \widehat{\beta}_{0,\lambda}(q)\}. \tag{5}$$

Theorem 1 (Projected ridge moment stability)**.** *Assume $Y \in [-B, B]$ and $\|\phi(X)\|_2 \le \phi_{\max}$. Fix $\lambda > 0$ and use $\widehat{\tau}_\lambda(q)$ as defined in* (5)*. Suppose that for each $q_\circ \in \{q, q'\}$ and $t \in \{0, 1\}$, $G_t(q_\circ) \succeq \kappa I$ for some $\kappa \ge 0$ and $\left\|\widehat{\beta}_{t,\lambda}(q_\circ)\right\|_2 \le R_\beta$. Then for two admissible workload-answer vectors q, q' of the same type (same coordinate ordering and same rule for G_t),*

$$|\widehat{\tau}_\lambda(q) - \widehat{\tau}_\lambda(q')| \;\le\; L_\phi \left\|q - q'\right\|_\infty, \; L_\phi \le C_\phi \left(1 + \frac{1}{\kappa + \lambda}\right).$$

The constant C_ϕ depends on $(\phi_{\max}, R_\beta, p)$ but not on n, ε, or the particular DP noise draw; Appendix E.2 gives one explicit choice.

Remark 1 (Ridge and rank). The useful quantity is $\kappa + \lambda$: ridge compensates for small eigenvalues of the arm-specific Gram matrices. If $\lambda = 0$, the same statement requires $\kappa > 0$; if $\kappa = 0$, one must take $\lambda > 0$ to make the inverse stable.

Example 2: clipped IPW/AIPW moment maps. Inverse-propensity weighting (IPW) estimates $\mathbb{E}[Y(1)]$ and $\mathbb{E}[Y(0)]$ by reweighting observed treated and control outcomes by inverse treatment probabilities. With partition-cell features, those reweighted sums can be written directly as functions of the released moments. For partition-cell features, let $p_{tj}(q) = q_{tj}^{(0)}$, $r_{tj}(q) = q_{tj}^{(1)}$, $s_j(q) = p_{0j}(q) + p_{1j}(q)$, and define the clipped cell propensity $\widehat{e}_j(q) = \mathrm{clip}\{p_{1j}(q)/s_j(q), \eta, 1 - \eta\}$. The direct clipped IPW moment map is

$$\widehat{\tau}_{\mathrm{IPW}}(q) = \sum_j \left\{ \frac{r_{1j}(q)}{\widehat{e}_j(q)} - \frac{r_{0j}(q)}{1 - \widehat{e}_j(q)} \right\}.$$

AIPW adds the usual outcome-regression augmentation to the same cell-level IPW score; Appendix E.3 gives the full expression.

Proposition 1 (Clipped IPW/AIPW moment stability)**.** *For partition-cell features, bounded outcomes, lower-bounded cell masses, and propensities clipped to $[\eta, 1 - \eta]$, the direct moment-map IPW and AIPW estimators are L_{IPW}- and L_{AIPW}-stable. The constants scale polynomially in $1/\eta$ and the inverse minimum cell mass; Appendix E.3 gives explicit bounds. This is standard clipped-score stability [Rosenbaum and Rubin, 1983, Robins et al., 1994, Chernozhukov et al., 2018]; the paper-specific step is plugging the resulting L_{est} into the DP moment-release decomposition.*

For either example, setting $q = q(D)$ and $q' = \widetilde{q}$ converts the deterministic stability bound into a privacy-noise bound. For ridge, the following corollary also records the conditioning requirement for the noisy Gram blocks.

Corollary 1 (Noise-calibrated ridge)**.** *Suppose the workload coordinates entering $\widehat{\tau}_\lambda(q)$ have independent Gaussian noise with coordinate scale σ, and set $\delta_m = c\sigma\sqrt{\log(m/\gamma)}$. On the event $\|\widetilde{q} - q(D)\|_\infty \le \delta_m$ and $\max_t \left\|\widetilde{G}_t - G_t\right\|_{\mathrm{op}} \le \lambda/2$, the noisy ridge matrices remain well conditioned and*

$$|\widehat{\tau}_\lambda(\widetilde{q}) - \widehat{\tau}_\lambda(q(D))| \lesssim C_\phi \left(1 + \frac{1}{\kappa + \lambda}\right) \delta_m.$$

For full joint-cell bases with diagonal Gram matrices, the operator-norm condition follows from $\lambda \gtrsim \delta_m$; for explicitly released dense Gram blocks, standard Gaussian-matrix bounds give the analogous sufficient choice $\lambda \gtrsim \sigma\sqrt{p + \log(1/\gamma)}$.

Here and below, SNR means the coordinatewise signal-to-noise ratio $|\widetilde{q}_a|/\sigma_a$. Appendix E.2 gives the conditioning intuition and explains how ridge complements the SNR calibration filter.

Theorem 1 and Proposition 1 are two examples of the same moment-stability template: ridge covers direct regression plug-ins that may require Gram moments, while IPW/AIPW covers cell-based causal scores computable from the base treatment and outcome moments.

5.2 ACCURACY OF THE DP MOMENT RELEASE

The previous subsection is deterministic once a moment-error bound is available. The next result supplies that bound for the Gaussian mechanism: it converts the sensitivity of the released workload into the coordinate error used by the stability examples and the decomposition below.

Theorem 2 (DP moment accuracy)**.** *Let $q(D) = n^{-1}\sum_i h(W_i) \in \mathbb{R}^m$ be the released workload answer, and suppose $\|h(W)\|_2 \le H_q$ for every record. Measuring $q(D)$ with the Gaussian mechanism under (ε, δ)-DP yields, with probability at least $1 - \gamma$,*

$$\|\widetilde{q} - q(D)\|_\infty = O\left(\frac{H_q\sqrt{\log(m/\gamma)\log(1/\delta)}}{n\varepsilon}\right).$$

For the base causal workload, $H_q = \phi_{\max}\sqrt{1 + B^2}$. If optional Gram blocks are appended for a direct regression plug-in, one may instead take $H_q = \{\phi^2_{\max}(1 + B^2) + \phi^4_{\max}\}^{1/2}$.

Remark on norms. Theorems 1 and 2 both use $\|\phi(X)\|_2 \le \phi_{\max}$; indicator features additionally satisfy $\|\phi(X)\|_\infty \le 1$, and for p-dimensional indicator features both hold simultaneously with $\phi_{\max} \le \sqrt{p}$.

5.3 OVERALL ATE ERROR DECOMPOSITION

This subsection does not introduce a third route; it composes the direct-route moment-stability bound with the two extra errors introduced when the released workload is converted into synthetic microdata. At the population-reconstruction level, the controlled estimand is $\widehat{\tau}_{\mathrm{est}}\{q(P^{\mathrm{syn}})\}$. When this quantity is estimated from n_{syn} sampled synthetic records, write the result as $\widehat{\tau}^{\mathrm{syn}}$ and treat finite synthetic sample size as an additional sampling term. The experiments instantiate this release–reconstruct–analyze path with DR/AIPW on synthetic microdata; the theorem formalizes the workload-matching error budget that this pipeline is designed to control.

Because the iterative max-entropy solver satisfies the noisy constraints only approximately, the bound carries an explicit calibration term. For a measured or retained coordinate set S (with $S = [m]$ for the fixed full workload when no thresholding is used), let Π_S denote coordinate projection: $\Pi_S v = (v_a : a \in S)$. In implementation, Π_S is the Boolean mask over retained moment coordinates. Define

$$\mathrm{CalGap}(S, \sigma) := L_{\mathrm{est}} \left\|\Pi_S\{\widetilde{q} - q(P^{\mathrm{syn}})\}\right\|_2,$$

where L_{est} is the appropriate estimator Lipschitz constant (L_ϕ for the ridge plug-in of Theorem 1, or the IPW/AIPW constants in Proposition 1 and Appendix E.3) and $q(P^{\mathrm{syn}})$ the workload moments of the solver output.

Theorem 3 (ATE error decomposition)**.** *Under Assumption 1, the boundedness conditions above, and an estimator Lipschitz condition $|\widehat{\tau}_{\mathrm{est}}(q) - \widehat{\tau}_{\mathrm{est}}(q')| \le L_{\mathrm{est}}\|q - q'\|_\infty$ for the retained workload coordinates, let $\widehat{\tau}^{\mathrm{syn}}$ be the empirical estimate of $\widehat{\tau}_{\mathrm{est}}\{q(P^{\mathrm{syn}})\}$ formed from n_{syn} i.i.d. draws from the solver output P^{syn}. This condition holds for the ridge plug-in by Theorem 1 and for partition-feature clipped IPW/AIPW by Proposition 1. Then*

$$|\widehat{\tau}^{\mathrm{syn}} - \tau| \le \underbrace{O_p\big(L_{\mathrm{est}} n^{-1/2}\big)}_{\textit{sampling}} + \underbrace{O_p\left(\frac{L_{\mathrm{est}} H_q\sqrt{\log m\ \log(\frac{1}{\delta})}}{n\varepsilon}\right)}_{\textit{privacy}} + \underbrace{\mathrm{Approx}(\phi; S)}_{\textit{workload}} + \underbrace{O_p\big(n_{\mathrm{syn}}^{-1/2}\big)}_{\textit{Monte Carlo}} + \underbrace{\mathrm{CalGap}(S, \sigma)}_{\textit{calibration}},$$

where $\mathrm{Approx}(\phi; S)$ is the approximation error of representing $m_t(x)$ and $e(x)$ using the retained workload coordinates. When P^{syn} satisfies the constraints exactly, the fifth term vanishes and the bound reduces to the four-term decomposition.

Remark 2 (SNR thresholding controls the calibration gap). Retaining only high-signal coordinates $S_\tau = \{a : |\widetilde{q}_a|/\sigma_a \ge \tau_{\mathrm{SNR}}\}$ and running the solver to noise-level tolerance $|\widetilde{q}_a - q_a(P^{\mathrm{syn}})| \le c_{\mathrm{cal}}\sigma_a$ on S_τ yields $\mathrm{CalGap}(S_\tau, \sigma) \le L_{\mathrm{est}} c_{\mathrm{cal}}\bar{\sigma}\sqrt{m_{\mathrm{kept}}}$, where $m_{\mathrm{kept}} = |S_\tau|$ and $\bar{\sigma}$ bounds the per-coordinate noise scale (Corollary 3, Appendix E.5). Discarded moments are not free: their effect moves into $\mathrm{Approx}(\phi; S_\tau)$, so thresholding trades a smaller, controllable calibration gap against possible approximation bias. Empirically, which side of the tradeoff binds is data-dependent: in our workload-dimension ablation on IHDP

(Figure 9), the variance side dominates—without thresholding, richer workloads reduce RMSE throughout the tested range, and thresholding lowers bias but costs more variance than it saves—while wider workloads or stricter budgets push toward the bias-dominated regime the threshold guards against. Because CalGap is computable from the release (the solver's residual is recorded per run), the tradeoff can be monitored rather than assumed.

Corollary 2 (Design rule for causal utility)**.** *To keep DP distortion below sampling noise, it suffices (up to logs) that*

$$\frac{H_q\sqrt{\log m}}{n\varepsilon} \ll \frac{1}{\sqrt{n}} \iff \varepsilon \gg H_q\sqrt{\frac{\log m}{n}}.$$

This recovers a natural "ε vs. n" threshold analogous to parametric DP inference [Smith, 2011], with the outcome bound B controlling the difficulty.

6 UNCERTAINTY QUANTIFICATION FROM DP SYNTHETIC DATA

Recent work shows that synthetic-data inference must account for both sampling variability and DP noise [Räisä et al., 2023, Perez et al., 2024]. We propose a causal version of NA+MI:

1. Treat the measured DP moments $\widetilde{q}_S$ as noisy observations of latent true moments q_S, with $S = [m]$ for the fixed full workload and adaptive S for CAUSAL-AIM.
2. Sample M posterior draws $q_S^{(\ell)} \sim \mathcal{N}(\widetilde{q}_S, \Sigma_S)$, $\ell = 1, \ldots, M$: the asymptotic Gaussian posterior for q_S given $\widetilde{q}_S$ under a flat prior, with covariance given by the known Gaussian-mechanism noise on the measured coordinates.
3. For each draw, construct $P^{\mathrm{syn},(\ell)}$ via (3) on the measured coordinates and sample a synthetic dataset $D^{\mathrm{syn},(\ell)}$.
4. Compute $\widehat{\tau}^{(\ell)}$ on each synthetic dataset and combine using Rubin's MI rules [Rubin, 1987].

Appendix F gives a bias-aware variant that inflates the MI variance by an estimated workload-approximation term, with a coverage guarantee; Appendix G states the full procedure as pseudocode. The result is a confidence interval that widens as ε decreases (stronger privacy), similar in spirit to Räisä et al. [2023].

7 EXPERIMENTS

We evaluate three questions: whether causal workloads improve ATE accuracy over generic workloads, whether NA+MI yields calibrated confidence intervals, and when adaptive CAUSAL-AIM improves over a fixed workload.

7.1 SETUP

Benchmarks. We use five benchmark settings where the target ATE is known, so both point error and coverage can be evaluated directly.

- **IHDP** [Hill, 2011]: $n = 747$ children from the Infant Health and Development Program benchmark, with 25 child and family covariates. Treatment indicates the program intervention, the observed outcome is a continuous cognitive score, and the standard semi-simulated release supplies potential-outcome means for the ground-truth ATE.
- **Twins** [Louizos et al., 2017]: $n = 11{,}400$ same-sex twin-pair records with 30 birth and maternal covariates. The benchmark constructs a confounded treatment indicator by sampling heavier-versus-lighter twin status from a covariate-dependent propensity; the outcome is one-year mortality, with both potential mortality outcomes observed within each pair.
- **ACIC 2016** [Dorie et al., 2019]: $n = 4{,}802$ units from the competition covariate matrix, with anonymized real covariates and three categorical fields one-hot encoded into 82 numeric columns. We use DGP 7, which simulates confounded treatment assignment and continuous potential outcomes with heterogeneous treatment effects; the ATE is computed from the exported potential outcomes.
- **LaLonde/NSW** [LaLonde, 1986]: the National Supported Work experimental sample with $n = 445$ subjects (185 treated and 260 controls). Covariates include age, education, race, marital status, and pre-program earnings; the outcome is 1978 earnings, and the experimental difference-in-means is the reference effect.
- **ACS**: California 2018 American Community Survey person microdata from `folktables` [Ding et al., 2021]. We use 20 demographic covariates and simulate confounded treatment and continuous potential outcomes at $n \in \{1000, 5000, 20000\}$, so the ATE is known by construction.

Baselines. All private methods release synthetic data and then estimate the ATE on synthetic records using doubly robust analysis.

- **Non-private DR**: an oracle doubly robust estimator with cross-fitting on the confidential data.
- **MST + naive DR**: generic MST-style synthesis [McKenna et al., 2021], followed by standard DR inference that ignores DP noise.
- **AIM + naive DR**: generic AIM synthesis [McKenna et al., 2022], followed by the same naive DR analysis.
- **Causal workload + naive DR**: our fixed causal workload mechanism, analyzed without the NA+MI uncertainty correction.

- **Causal workload + NA+MI**: the fixed causal workload mechanism with noise-aware multiple imputation.
- **CAUSAL-AIM + NA+MI**: adaptive causal workload selection with the same NA+MI analysis.

Defaults and metrics. Default runs use $L = 5$ quantile bins, $M = 20$ MI draws, $n_{\text{syn}} = 10n$, clipped standardized outcomes, DR analysis on synthetic data, and no SNR thresholding or calibration ridge in the main pipeline; we keep $n_{\text{syn}} = 10n$ for comparability across methods, although the ablation shows $n_{\text{syn}} = n$ already suffices. We report ATE RMSE and empirical 95% coverage over $\varepsilon \in \{0.5, 1, 2, 5\}$ with $\delta = 1/n^2$, and run ablations over workload dimension, MI draws, synthetic sample size, overlap strength, and adaptive rounds K. Appendix H gives feature-construction details, metric definitions, and ablation grids; additional figures appear in Appendix J.

7.2 RESULTS

Table 1 summarizes all methods at ε=1, Figures 2 and 3 show RMSE and coverage across privacy budgets, and Figure 5 shows the adaptive fixed-workload comparison on ACIC.

Generic workloads win on RMSE at loose budgets, but lose on coverage. MST + naive DR achieves the lowest private RMSE on all four benchmarks at $\varepsilon \geq 2$ and on three of four at $\varepsilon = 1$. At strict budgets, causal workload + NA+MI becomes more competitive, matching or beating MST on two of four benchmarks at $\varepsilon = 0.5$ and on IHDP at $\varepsilon = 1$. However, MST's point accuracy comes with invalid uncertainty quantification: at $\varepsilon \leq 1$, all naive methods have coverage at most 35.2% (Figure 3).

NA+MI restores calibrated inference. Causal workload + NA+MI is the only private method with near-nominal coverage at $\varepsilon \leq 1$ on all four benchmarks (99.8−100%; Table 1). The price is wider intervals: NA+MI intervals are 34–300× wider than naive intervals at strict budgets (Figure 4) because they account for DP measurement noise. This is the intended behavior when privacy noise dominates the sampling error; the narrow naive intervals are overconfident rather than genuinely informative.

Adaptive workloads are operating-point choices. On ACIC at $\varepsilon \geq 1$, CAUSAL-AIM substantially reduces RMSE relative to the fixed causal workload, especially at small K (Figure 5). At $\varepsilon = 0.5$, additional adaptive rounds can hurt because each round consumes privacy budget that may exceed the gain from better feature targeting. On IHDP, no tested K recovers the fixed workload's calibration (Appendix K), so fixed causal workload + NA+MI remains the conservative default when coverage is the priority.

Robustness and reuse. On the ACS semi-synthetic study, NA+MI again provides the coverage advantage: at ($n = 1000, \varepsilon = 0.5$), Causal + NA+MI has coverage 1.00 while MST + naive DR has coverage 0.07, and NA+MI also wins RMSE in that cell (0.88 vs. 1.17). Ablations show that $M = 20$ MI draws is conservative, synthetic sample sizes above $n_{\text{syn}} = n$ add little, and SNR thresholding is mainly useful as a safeguard when moments approach the noise floor (Figures 9, 10, 11, and 12). Finally, one DP synthetic release supports multiple estimands: ATE, ATT, and a subgroup effect all achieve nominal coverage in the reuse experiment (Appendix K).

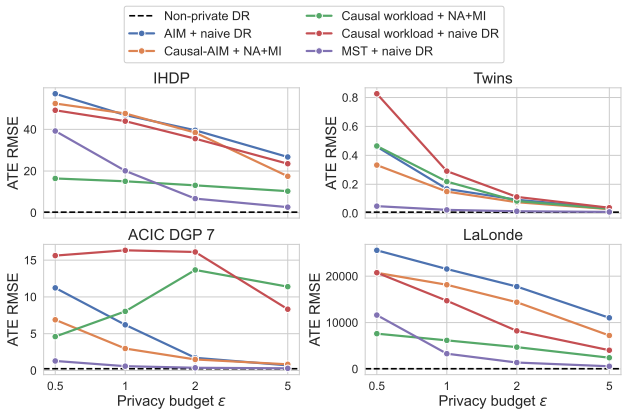

Figure 2: ATE RMSE versus privacy budget (ε) on four benchmarks ($n_{\text{rep}} = 500$, $\delta = 1/n^2$). MST + naive DR leads on RMSE at $\varepsilon \geq 2$; at $\varepsilon = 0.5$, Causal + NA+MI matches or beats it on half the benchmarks, and on IHDP at $\varepsilon = 1$. The value of causal workload methods lies chiefly in enabling calibrated coverage (Figure 3).

Fidelity is not causal utility. Generic workloads dominate marginal-fidelity metrics such as average marginal total-variation distance (TVD) in nearly every configuration (Figure 13), yet those metrics do not rank methods by ATE error or coverage. This is the empirical counterpart of the workload argument: preserving generic low-dimensional marginals is not enough when the downstream target depends on treatment-arm outcome and balance moments.

Practical takeaway. Use Causal + NA+MI when valid uncertainty quantification is the primary goal, especially at strict privacy budgets. Use CAUSAL-AIM when point RMSE is prioritized and its operating point is favorable for the dataset. Use generic workloads when the release is intended for broad marginal fidelity or exploratory analysis and calibrated causal intervals are not required.

8 CONCLUSION

This paper argues that DP synthetic data for causal inference should preserve the moments used by the causal estimand, not only generic distributional fidelity. The proposed causal workload releases treatment-arm feature masses and outcome-feature moments, reconstructs a maximum-entropy

Table 1: ATE RMSE and 95% CI Coverage at $\varepsilon = 1.0$, $\delta = 1/n^2$. Each cell is RMSE / Coverage; among private methods, bold marks the best value and underlining marks the second-best value in each dataset column.

Method	IHDP	Twins	ACIC	LaLonde
Non-private DR	0.259/0.940	0.007/1.000	0.236/1.000	47.593/1.000
MST + naive	20.100/0.014	**0.023**/0.352	**0.583**/0.236	**3290.027**/0.118
AIM + naive	46.863/0.002	0.168/0.064	6.206/0.060	21546.320/0.008
Causal + naive	43.942/0.024	0.291/0.078	16.338/0.016	14707.613/0.048
Causal + NA+MI	**15.074/0.998**	0.219/**1.000**	8.034/**1.000**	6163.055/**1.000**
Causal-AIM + NA+MI	47.633/0.346	0.150/0.956	2.981/0.718	18147.736/0.678

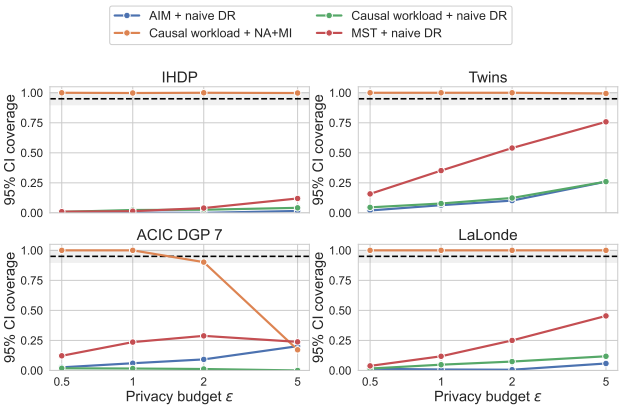

Figure 3: Empirical 95% CI coverage across privacy budgets and four datasets ($n_{\text{rep}} = 500$). The dashed line marks nominal 0.95. All naive methods are at or below 35.2% at $\varepsilon \leq 1$. Causal workload + NA+MI achieves $99.8-100\%$ coverage at $\varepsilon \leq 1$ on all four benchmarks; Coverage can decline at high ε due to approximation bias (Appendix L).

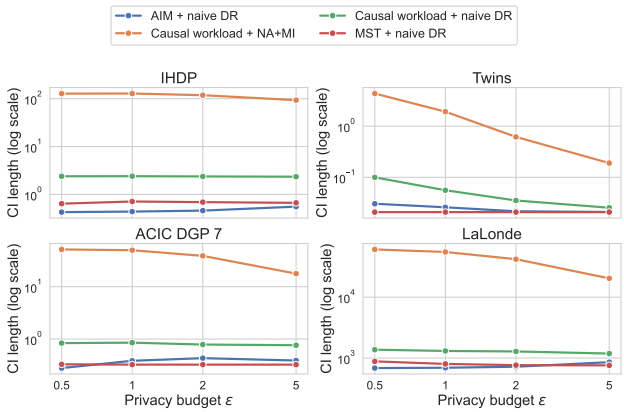

Figure 4: Average 95% CI length across methods, datasets, and privacy budgets ($n_{\text{rep}} = 500$). Noise-aware MI intervals are 34–300× wider than naive intervals at $\varepsilon \leq 1$ (depending on dataset and comparator) because they account for both DP noise and sampling variability. Naive intervals are narrow but far below nominal coverage (Figure 3), so their apparent precision is spurious.

synthetic distribution, and supports noise-aware inference through NA+MI. The theory tracks how DP moment error, workload approximation, calibration mismatch, and synthetic Monte Carlo error enter ATE estimation; the experiments show the corresponding tradeoff: generic workloads often lead on RMSE at loose budgets, but their naive intervals undercover severely, while causal workload + NA+MI provides calibrated inference at strict privacy budgets. The distinction between the direct q-route and the synthetic-data route clarifies what the theory controls and what the release enables: one measured workload can be used directly by stable moment maps or converted into reusable synthetic records for ATE, ATT, subgroup effects, and model checks under the same privacy cost. The main limitations are the need to choose a feature map ϕ, approximation bias at high ε, wider intervals when DP noise dominates, and our focus on the AIM/Private-PGM reconstruction family; Appendix L discusses these limitations, the coverage–privacy diagnostic, and future directions such as continuous feature maps, richer subgroup/CATE workloads, and end-to-end DP-aware feature selection.

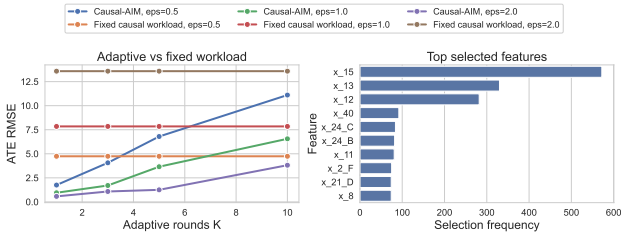

Figure 5: ATE RMSE of CAUSAL-AIM (adaptive) versus a fixed causal workload on ACIC DGP 7 ($n_{\text{rep}} = 100$). The left panel overlays all privacy budgets: horizontal lines mark the fixed workload and curves show CAUSAL-AIM as a function of adaptive rounds K; the right panel shows the most frequently selected features. At $\varepsilon \geq 1$, CAUSAL-AIM reduces RMSE by $78-96\%$ over the fixed workload at $K \leq 3$ ($53-91\%$ at K=5). At $\varepsilon = 0.5$, additional rounds degrade performance because each round consumes budget that exceeds the benefit of better feature targeting; few rounds ($K \leq 3$) are recommended at low ε.

Acknowledgements

AA was partially supported by NIH grant R01MH139379 and by Patient-Centered Outcomes Research Institute (PCORI) award ME-2023C1-32148.

Reproducibility. All experiments, tables, and figures in this paper can be reproduced with the code, data-processing scripts, and notebooks available at `https://github.com/AsiaeeLab/causal-aim`.

References

Amir Asiaee, Zhuohui J. Liang, and Chao Yan. CausalWrap: Model-agnostic causal constraint wrappers for tabular synthetic data, 2026.

P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130, 2016. doi: 10.1111/rssb.12158.

Claire McKay Bowen and Fang Liu. Comparative study of differentially private data synthesis methods. *Statistical Science*, 35(2):280–307, 2020.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring adult: New datasets for fair machine learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 6478–6490, 2021.

Konstantin Donhauser, Javier Abad, Neha Hulkund, and Fanny Yang. Privacy-preserving data release leveraging optimal transport and particle gradient descent. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024.

Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott, and Dan Cervone. Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science*, 34(1):43–68, 2019.

Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*, volume 9 of *Foundations and Trends in Theoretical Computer Science*. Now Publishers, 2014.

Amirhossein Farzam and Guillermo Sapiro. Causal inference under differential privacy: Challenges and mitigation strategies. In *Causal Representation Learning Workshop at NeurIPS*, 2024. CRL@NeurIPS 2024.

Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pages 3713–3722. PMLR, 2021.

Peter Grünwald and Thijs van Ommen. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Analysis*, 12(4): 1069–1103, 2017. doi: 10.1214/17-BA1085.

Moritz Hardt, Katrina Ligett, and Frank McSherry. A simple and practical algorithm for differentially private data release. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 25, 2012.

Jennifer L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

Nicholas J. Jackson, Natalia Espinosa-Dice, Chao Yan, and Bradley A. Malin. Reward-guided generation improves the scientific utility of synthetic biomedical data. *medRxiv*, 2026. doi: 10.64898/2026.03.11.26348077. Preprint; PMCID: PMC13015641.

Haim Kaplan, Shachar Schnapp, and Uri Stemmer. Differentially private approximate quantiles. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162 of *Proceedings of Machine Learning Research*, pages 10751–10761. PMLR, 2022.

Robert J. LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76(4):604–620, 1986.

Christian Janos Lebeda, Mathieu Even, Aurélien Bellet, and Julie Josse. Model agnostic differentially private causal inference. *arXiv preprint arXiv:2505.19589*, 2025.

Jing Lei. Differentially private M-estimators. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 24, pages 361–369, 2011.

Zhuohui J. Liang, Zhuohang Li, Nicholas J. Jackson, Kevin H. Guo, Yanink Caro-Vega, Ronaldo I. Moreira, Fabio Paredes, Jordany Bernadin, Diana Varela, Carina Cesar, Alessandro Blasimme, Jessica M. Perkins, Amir Asiaee, Stephany N. Duda, Bradley A. Malin, Bryan E. Shepherd, and Chao Yan. Generating synthetic multinational longitudinal cohorts for clinically grounded HIV research. *Nature Communications*, 2026. doi: 10.1038/s41467-026-74492-0.

Christos Louizos, Uri Shalit, Joris M. Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

Ryan McKenna, Daniel Sheldon, and Gerome Miklau. Graphical-model based estimation and inference for differential privacy. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pages 4187–4196. PMLR, 2019.

Ryan McKenna, Gerome Miklau, and Daniel Sheldon. Winning the NIST contest: A scalable and general approach to differentially private synthetic data. *arXiv preprint arXiv:2108.04978*, 2021.

Ryan McKenna, Brett Mullins, Daniel Sheldon, and Gerome Miklau. AIM: An adaptive and iterative mechanism for differentially private synthetic data. *Proceedings of the VLDB Endowment*, 15(11):2599–2612, 2022. doi: 10.14778/3551793.3551817.

Jeffrey W. Miller and David B. Dunson. Robust Bayesian inference via coarsening. *Journal of the American Statistical Association*, 114(527):1113–1125, 2019. doi: 10.1080/01621459.2018.1469995.

Fengshi Niu, Harsha Nori, Brian Quistorff, Rich Caruana, Donald Ngwe, and Aadharsh Kannan. Differentially private estimation of heterogeneous causal effects. In *Proceedings of the First Conference on Causal Learning and Reasoning (CLeaR)*, volume 177 of *Proceedings of Machine Learning Research*, pages 618–633. PMLR, 2022.

Yuki Ohnishi and Jordan Awan. Differentially private covariate balancing causal inference. *arXiv preprint arXiv:2410.14789*, 2024.

Ileana Montoya Perez, Parisa Movahedi, Valtteri Nieminen, Antti Airola, and Tapio Pahikkala. Does differentially private synthetic data lead to synthetic discoveries? *Methods of Information in Medicine*, 63:35–51, 2024. doi: 10.1055/a-2385-1355.

Trivellore E. Raghunathan, Jerome P. Reiter, and Donald B. Rubin. Multiple imputation for statistical disclosure limitation. *Journal of Official Statistics*, 19(1):1–16, 2003.

Ossi Räisä, Joonas Jälkö, Samuel Kaski, and Antti Honkela. Noise-aware statistical inference with differentially private synthetic data. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 206 of *Proceedings of Machine Learning Research*, pages 3620–3643. PMLR, 2023.

Jerome P. Reiter. Inference for partially synthetic, public use microdata sets. *Survey Methodology*, 29(2):181–188, 2003.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York, 1987.

Maresa Schröder, Justin Hartenstein, and Stefan Feuerriegel. PrivATE: Differentially private confidence intervals for average treatment effects. *arXiv preprint arXiv:2505.21641*, 2025a.

Maresa Schröder, Valentyn Melnychuk, and Stefan Feuerriegel. Differentially private learners for heterogeneous treatment effects. In *International Conference on Learning Representations (ICLR)*, 2025b.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pages 813–822, 2011.

Yuchao Tao, Ryan McKenna, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. Benchmarking differentially private synthetic data generation algorithms. *arXiv preprint arXiv:2112.09238*, 2021.

James Watson and Chris Holmes. Approximate models and robust decisions. *Statistical Science*, 31(4):465–489, 2016. doi: 10.1214/16-STS592.

Workload-Preserving Differentially Private Synthetic Data for Causal Inference via Maximum-Entropy Calibration (Supplementary Material)

Amir Asiaee[1] **Kaveh Aryan**[2]

[1]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[2]Department of Informatics, King's College London, London, UK

A FEATURE ENCODINGS AND GRAM MOMENTS

This appendix clarifies the feature-encoding point used in Section 5. The experiments use *concatenated one-hot* features: each covariate is discretized separately, one-hot encoded separately, and then the groups are stacked. For example, with age in $\{\text{young}, \text{old}\}$ and sex in $\{\text{female}, \text{male}\}$, the experimental-style feature vector is

$$\phi_{\text{cat}}(X) = \big(1,\ 1\{\text{young}\},\ 1\{\text{old}\},\ 1\{\text{female}\},\ 1\{\text{male}\}\big).$$

The base workload then releases, for each arm, noisy versions of the age-bin masses, sex-bin masses, and the corresponding outcome-weighted moments.

A *full joint-cell* encoding instead uses indicators for the Cartesian-product cells, e.g.

$$\phi_{\text{joint}}(X) = \big(1\{\text{young}, \text{female}\},\ 1\{\text{young}, \text{male}\},\ 1\{\text{old}, \text{female}\},\ 1\{\text{old}, \text{male}\}\big).$$

Exactly one coordinate of $\phi_{\text{joint}}(X)$ is active for each record. Therefore

$$\phi_{\text{joint}}(X)\phi_{\text{joint}}(X)^\top \text{ is diagonal, and } G_t(D) = \text{diag}\{q_t^{(0)}(D)\}.$$

For concatenated one-hot features, off-diagonal entries are real co-occurrence moments; for instance,

$$\frac{1}{n}\sum_i 1\{T_i = t\}1\{\text{old}_i\}1\{\text{male}_i\}$$

is not determined by the separate old and male counts. Thus a direct ridge plug-in under concatenated or overlapping features would need to append $q_t^{(2)} = \text{vec}(G_t)$ as optional queries. Our experiments do not use this augmented Gram release; they release the base $4p$ workload, reconstruct synthetic data, and then run DR/AIPW on the synthetic records.

B FIRST-ORDER SUFFICIENCY OF THE ORTHOGONAL-SCORE WORKLOAD

This section clarifies why the causal workload of Section 4 is *one principled choice* rather than a claimed optimum. Let $\zeta = (m_0, m_1, e)$ denote the nuisances (η remains the positivity constant), $\psi(W; \theta, \zeta)$ a Neyman-orthogonal score for a smooth target θ, and $\mathcal{F}_\phi$ the linear nuisance class spanned by ϕ.

Proposition 2 (First-order sufficiency)**.** *Restrict the nuisance functions to the feature class spanned by ϕ. For smooth causal targets identified by a Neyman-orthogonal score, the workload $\mathcal{Q}_\phi = \{q_0^{(0)}, q_1^{(0)}, q_0^{(1)}, q_1^{(1)}\}$ identifies the arm-specific feature masses, outcome-feature moments, and treatment-balance moments used by projected outcome, propensity, IPW, and AIPW moment maps. Direct ridge plug-ins may additionally require Gram moments, as discussed in Appendix A; these are optional extensions rather than part of the default experimental release.*

Proof. The outcome-feature moments $q_t^{(1)} = \mathbb{E}[1\{T = t\}Y\phi(X)]$ encode arm-specific outcome structure in the chosen feature class, while $q_1^{(0)} = \mathbb{E}[T\phi(X)]$ and $q_0^{(0)} = \mathbb{E}[(1 - T)\phi(X)]$ encode treatment-feature balance. Once projected nuisances or direct cell scores are identified from these moments, orthogonality makes remaining nuisance errors second order, so a synthetic distribution matching the causal workload supports inference for the projected target subject to the explicit privacy, approximation, Monte Carlo, and calibration terms. □

Necessity intuition and scope. If two local alternatives in $\mathcal{F}_\phi$ agree on a release but differ in an arm-specific outcome-feature or treatment-feature moment, some orthogonal-score target has different first-order pathwise derivatives under the two alternatives, so a release that cannot distinguish those directions cannot support uniformly regular root-n inference over the projected class. This is a first-order identification argument only: it is *not* a minimax lower bound, does *not* prove uniqueness of the workload, and does *not* show that the base query count is optimal among all DP mechanisms; those questions are outside the present scope (Appendix L).

C MAXIMUM-ENTROPY RECONSTRUCTION DETAILS

This section provides details of the reconstruction step used in Algorithm 2. Assume first that $W = (X, T, Y)$ has finite support $\mathcal{W}$, as in a discretized synthetic-data model. Let $h_a : \mathcal{W} \to \mathbb{R}$, $a = 1, \ldots, m$, be the ordered workload query functions, with $m = 4p$ for the base causal workload and larger m if optional moments are appended. For a measured coordinate set $S \subset [m]$, write $h_S(w) = (h_a(w) : a \in S)^\top$ and $\widetilde{q}_S = (\widetilde{q}_a : a \in S)^\top$. Let $p_0(w) > 0$ be a base distribution. The role of p_0 is to say what distribution we prefer among all distributions that match the released moments: it may be uniform over the discretized support, an independent product distribution, a product distribution fit to already measured one-way marginals, or a public prior from an external population.

If the noisy constraints are feasible exactly, reconstruction is the information projection

$$\min_{p \in \Delta(\mathcal{W})} \sum_{w \in \mathcal{W}} p(w) \log \frac{p(w)}{p_0(w)} \quad \text{s.t.} \quad \sum_{w \in \mathcal{W}} p(w) h_a(w) = \widetilde{q}_a,\ a \in S.$$

The Lagrangian is

$$\mathcal{L}(p, \xi, \alpha) = \sum_w p(w) \log \frac{p(w)}{p_0(w)} - \xi^\top \left(\sum_w p(w) h_S(w) - \widetilde{q}_S \right) + \alpha \left(\sum_w p(w) - 1 \right).$$

Differentiating with respect to $p(w)$ and setting the derivative to zero gives

$$\log \frac{p(w)}{p_0(w)} + 1 - \xi^\top h_S(w) + \alpha = 0,$$

and therefore

$$p_\xi(w) = \frac{p_0(w) \exp\{\xi^\top h_S(w)\}}{Z(\xi)}, \qquad Z(\xi) = \sum_{u \in \mathcal{W}} p_0(u) \exp\{\xi^\top h_S(u)\}.$$

The dual objective can be written as

$$\max_{\xi \in \mathbb{R}^{|S|}} \left\{ \xi^\top \widetilde{q}_S - \log Z(\xi) \right\},$$

whose gradient is $\widetilde{q}_S - \mathbb{E}_{p_\xi}[h_S(W)]$ and whose Hessian is minus the covariance of $h_S(W)$ under p_ξ. Thus optimizing the dual searches for multipliers ξ whose exponential-family distribution matches the released noisy moments.

Because $\widetilde{q}_S$ contains DP noise, exact feasibility is not guaranteed. For example, noise can make a released cell mass slightly negative, or make several marginal constraints mutually inconsistent. Implementations therefore solve a relaxed weighted calibration problem such as

$$\min_{P \in \mathcal{P}} \frac{1}{2} \left\| \Sigma_S^{-1/2} \Pi_S \{ q(P) - \widetilde{q} \} \right\|_2^2 + \rho \, \mathrm{KL}(P \| P_0),$$

where Σ_S is the known DP noise covariance on the measured coordinates, so high-noise coordinates are automatically downweighted. Equivalently, one can solve the exact dual only on retained high-signal coordinates and stop when residuals are at the noise scale, e.g. $|q_a(P) - \widetilde{q}_a| \le c_{\mathrm{cal}} \sigma_a$. This is the sense in which Algorithm 2 uses Private-PGM-style reconstruction: it fits a structured exponential-family / graphical-model distribution to noisy measured moments rather than trying to estimate the full joint table without structure, then samples synthetic records from the fitted distribution.

Algorithm 2 Fixed Causal-Workload Synthesis Procedure

Require: Confidential data D, query functions $\{h_a\}_{a=1}^m$, privacy budget (ε, δ), base distribution P_0, synthetic size n_{syn}

1: Compute exact empirical answers $q_a(D) = n^{-1} \sum_i h_a(W_i)$ for $a = 1, \ldots, m$
2: Compute the sensitivity bound Δ_2 and Gaussian scale $\sigma = \Delta_2 \sqrt{2 \log(1.25/\delta)}/\varepsilon$
3: Release $\widetilde{q} = q(D) + Z$, where $Z \sim \mathcal{N}(0, \sigma^2 I_m)$
4: Optionally retain high-signal coordinates $S_\tau = \{a : |\widetilde{q}_a|/\sigma_a \geq \tau_{\text{SNR}}\}$; otherwise set $S_\tau = [m]$
5: Fit P^{syn} by solving the relaxed maximum-entropy calibration problem on $\Pi_{S_\tau} \widetilde{q}$
6: Draw n_{syn} records D^{syn} from P^{syn}

Ensure: Synthetic dataset D^{syn} and recorded noise scale σ

D CAUSAL-AIM SELECTION DETAILS

At each round k, CAUSAL-AIM computes an internal utility for each unmeasured candidate group r, estimating the reduction in ATE error if its coordinate block $I(r) \subset [m]$ were added to the measured workload. This is an adaptive greedy selection rule with private randomized selection: candidates are scored against the current reconstructed model, one group is selected through the exponential mechanism, its coordinates are measured once, and the model is refit before the next round. Previously measured groups are removed from the candidate set and are not re-measured; later rounds reuse their stored noisy answers.

The score is motivated by the orthogonal-score residual

$$\Delta_r = \left\| \mathbb{E}_{P_{k-1}}[\psi(W; \hat{\theta}_{k-1}, \hat{\zeta}_{k-1}) \cdot \phi_r(X)] \right\|_2^2,$$

where ψ is the influence function for the ATE and $(\hat{\theta}_{k-1}, \hat{\zeta}_{k-1})$ are the current nuisance estimates under P_{k-1}. Here $W = (X, T, Y)$; for the ATE, with nuisances $\zeta = (m_0, m_1, e)$ and target τ, a standard orthogonal score is

$$\psi(W; \tau, \zeta) = m_1(X) - m_0(X) + \frac{T\{Y - m_1(X)\}}{e(X)} - \frac{(1-T)\{Y - m_0(X)\}}{1 - e(X)} - \tau.$$

In the implementation, this is operationalized by a moment-space proxy. Let $q^{(k)}$ be the current noisy moment vector assembled from the coordinates measured so far and from the public base distribution on unmeasured coordinates. A typical utility is

$$u_k(r) = \widehat{L}_k \left\| \Pi_{I(r)} \{q(D) - q^{(k-1)}\} \right\|_2,$$

where $\widehat{L}_k$ is an empirical Lipschitz/overlap factor and $\Pi_{I(r)}$ projects onto the coordinates for group r. Large utility means that the current reconstruction leaves a causally weighted moment block poorly matched. The utilities are used only inside the selection mechanism and are not released. If Δ_u bounds the sensitivity of u_k, the exponential mechanism selects

$$\Pr(r_k = r) \propto \exp\left\{ \frac{\varepsilon_{\text{score},k}\, u_k(r)}{2\Delta_u} \right\}, \qquad r \in A_{k-1}.$$

Thus larger utility means larger selection probability; if one instead defines a loss, the utility is the negative loss.

E PROOFS

E.1 WEIGHTED ATES: ATT AND SUBGROUP EFFECTS

The ATE decomposition extends to weighted effects whenever the weighting rule is represented by the retained workload coordinates. For a nonnegative weight function ω, define

$$\tau_\omega := \frac{\mathbb{E}[\omega(X)\{m_1(X) - m_0(X)\}]}{\mathbb{E}[\omega(X)]}.$$

This includes subgroup effects with $\omega(X) = \mathbb{1}\{X \in G\}$, and ATT-type effects with $\omega(X) = e(X)$, provided the subgroup or propensity weights are represented by the workload.

Algorithm 3 CAUSAL-AIM Selection and Measurement

Require: Data D, groups $\mathcal{R}$, coordinate blocks $\{I(r)\}_{r\in\mathcal{R}}$, budget (ε,δ), rounds K, public base moment vector $q^{(0)} = q(P_0)$

Split the budget into scoring and measurement portions, and split each portion across rounds by sequential composition
Initialize measured coordinates $S_0 \leftarrow \emptyset$, available groups $A_0 \leftarrow \mathcal{R}$, and current moment proxy $q^{(0)}$
for $k = 1, \ldots, \min(K, |\mathcal{R}|)$ **do**
    For each $r \in A_{k-1}$, compute an internal utility $u_k(r)$ for adding coordinate block $I(r)$
    Select r_k from A_{k-1} with the exponential mechanism, with probability proportional to $\exp\{\varepsilon_{\mathrm{score},k} u_k(r)/(2\Delta_u)\}$
    Release noisy answers $\widetilde{q}_{I(r_k)} = q_{I(r_k)}(D) + Z_k$ with the Gaussian mechanism
    Update $S_k \leftarrow S_{k-1} \cup I(r_k)$ and $A_k \leftarrow A_{k-1} \setminus \{r_k\}$
    Update $q^{(k)}$ by replacing coordinates $I(r_k)$ with $\widetilde{q}_{I(r_k)}$ and retaining all earlier noisy measurements
    Refit P_k by maximum-entropy calibration to the stored noisy answers on S_k
end for
Output stored noisy moments $\widetilde{q}_{S_K}$; for synthetic release, output P_K and sample synthetic records from P_K

Proposition 3 (Weighted-effect moment stability)**.** *Let $\widehat{\tau}_\omega(q) = N_\omega(q)/D_\omega(q)$ be a direct moment-map estimator using only retained workload coordinates. Assume that, on the admissible workload domain, $D_\omega(q) \geq \rho_\omega > 0$, $|N_\omega(q)| \leq M_\omega D_\omega(q)$, and*

$$|N_\omega(q) - N_\omega(q')| \leq L_{N,\omega} \|q - q'\|_\infty, \qquad |D_\omega(q) - D_\omega(q')| \leq L_{D,\omega} \|q - q'\|_\infty.$$

Then $\widehat{\tau}_\omega$ is Lipschitz in the workload moments:

$$|\widehat{\tau}_\omega(q) - \widehat{\tau}_\omega(q')| \leq L_\omega \|q - q'\|_\infty, \qquad L_\omega = \frac{L_{N,\omega} + M_\omega L_{D,\omega}}{\rho_\omega}.$$

Consequently, the proof of Theorem 3 applies to τ_ω after replacing L_{est} by L_ω and replacing $\mathrm{Approx}(\phi; S)$ by the weighted approximation error $\mathrm{Approx}_\omega(\phi; S)$.

Proof. Let $\delta_q = \|q - q'\|_\infty$. By adding and subtracting $N_\omega(q')/D_\omega(q)$,

$$\left|\frac{N_\omega(q)}{D_\omega(q)} - \frac{N_\omega(q')}{D_\omega(q')}\right| \leq \frac{|N_\omega(q) - N_\omega(q')|}{D_\omega(q)} + |N_\omega(q')| \left|\frac{1}{D_\omega(q)} - \frac{1}{D_\omega(q')}\right|.$$

The first term is at most $L_{N,\omega}\delta_q/\rho_\omega$. For the second term, use $|N_\omega(q')| \leq M_\omega D_\omega(q')$ and $D_\omega(q) \geq \rho_\omega$ to obtain

$$|N_\omega(q')| \frac{|D_\omega(q) - D_\omega(q')|}{D_\omega(q) D_\omega(q')} \leq \frac{M_\omega L_{D,\omega}}{\rho_\omega} \delta_q.$$

Combining the two displays gives the Lipschitz constant. The decomposition proof of Theorem 3 then applies verbatim with target τ_ω and represented target $\tau_{\omega,\phi,S}$. □

For a subgroup effect, $\omega(X) = \mathbb{1}\{X \in G\}$ is represented whenever G is a union of workload cells; the denominator condition is a minimum subgroup-mass condition. For ATT in a partition basis, $\omega(X) = e(X)$ is represented by the cell propensity $e_j = p_{1j}/(p_{0j} + p_{1j})$, computed from $q^{(0)}$ and clipped as in Appendix E.3; the same cell-mass and clipping assumptions give the required Lipschitz constants. This proposition does not claim a full pointwise CATE guarantee; pointwise CATE requires additional approximation assumptions at the target covariate value.

E.2 PROOF OF THEOREM 1: ATE LIPSCHITZNESS

Throughout, $G_t = n^{-1}\sum_i \mathbb{1}\{T_i = t\}\phi(X_i)\phi(X_i)^\top$ (equivalently $q_t^{(2)} = \mathrm{vec}(G_t)$ when released explicitly) and $r_t = q_t^{(1)}$ are the arm-specific Gram and outcome moments, $\bar{\phi} := q_0^{(0)} + q_1^{(0)}$ is the empirical feature mean, and $u = (G_0, G_1, r_0, r_1, \bar{\phi})$ collects the moment inputs. The ATE plug-in is $\widehat{\tau}_\lambda(u) = \bar{\phi}^\top(\widehat{\beta}_{1,\lambda} - \widehat{\beta}_{0,\lambda})$ with $\widehat{\beta}_{t,\lambda} = (G_t + \lambda I)^{-1} r_t$.

Lemma 1 (Ridge perturbation bound, full form)**.** *Under the assumptions of Theorem 1, consider two inputs u, u' with $G'_t + \lambda I \succeq (\kappa + \lambda) I/2$ and $\left\|\widehat{\beta}_{t,\lambda}(u)\right\|_2 \vee \left\|\widehat{\beta}_{t,\lambda}(u')\right\|_2 \le R_\beta$ for $t = 0, 1$. Then*

$$|\widehat{\tau}_\lambda(u) - \widehat{\tau}_\lambda(u')| \le 2R_\beta \left\|\bar{\phi} - \bar{\phi}'\right\|_2 + \frac{2\phi_{\max}}{\kappa + \lambda} \sum_{t=0}^{1} \left\{ \|r_t - r'_t\|_2 + R_\beta \|G_t - G'_t\|_{\mathrm{op}} \right\}. \tag{6}$$

Proof. Let $A_t = G_t + \lambda I$ and $A'_t = G'_t + \lambda I$, so $\left\|A_t^{-1}\right\|_{\mathrm{op}} \le 1/(\kappa + \lambda)$ and $\left\|(A'_t)^{-1}\right\|_{\mathrm{op}} \le 2/(\kappa + \lambda)$. The identity $A_t^{-1} r_t - (A'_t)^{-1} r'_t = A_t^{-1}(r_t - r'_t) + A_t^{-1}(G'_t - G_t)(A'_t)^{-1} r'_t$ gives

$$\left\|\widehat{\beta}_{t,\lambda}(u) - \widehat{\beta}_{t,\lambda}(u')\right\|_2 \le \frac{\|r_t - r'_t\|_2 + R_\beta \|G_t - G'_t\|_{\mathrm{op}}}{\kappa + \lambda},$$

with constants at most doubled when only the lower bound on A'_t is used. Splitting the plug-in difference,

$$\begin{aligned} &|\bar{\phi}^\top(\widehat{\beta}_{1,\lambda}(u) - \widehat{\beta}_{0,\lambda}(u)) - (\bar{\phi}')^\top(\widehat{\beta}_{1,\lambda}(u') - \widehat{\beta}_{0,\lambda}(u'))| \\ &\le \left\|\bar{\phi} - \bar{\phi}'\right\|_2 \left\|\widehat{\beta}_{1,\lambda}(u) - \widehat{\beta}_{0,\lambda}(u)\right\|_2 \\ &\quad + \left\|\bar{\phi}'\right\|_2 \sum_{t=0}^{1} \left\|\widehat{\beta}_{t,\lambda}(u) - \widehat{\beta}_{t,\lambda}(u')\right\|_2, \end{aligned}$$

where $\left\|\widehat{\beta}_{1,\lambda}(u) - \widehat{\beta}_{0,\lambda}(u)\right\|_2 \le 2R_\beta$ by the radius assumption and $\left\|\bar{\phi}'\right\|_2 \le \phi_{\max}$ since $\bar{\phi}'$ averages feature vectors bounded by $\phi_{\max}$. Substituting the coefficient bound proves (6). □

Proof of Theorem 1. Let $\delta_q = \|q - q'\|_\infty$ denote the maximum coordinatewise perturbation across the moment blocks entering $u = (G_0, G_1, r_0, r_1, \bar{\phi})$. Then $\left\|\bar{\phi} - \bar{\phi}'\right\|_2 \le \sqrt{p}\,\delta_q$, $\|r_t - r'_t\|_2 \le \sqrt{p}\,\delta_q$, and $\|G_t - G'_t\|_{\mathrm{op}} \le p\,\delta_q$. Substituting these inequalities into Lemma 1 gives

$$|\widehat{\tau}_\lambda(q) - \widehat{\tau}_\lambda(q')| \le \left\{ 2R_\beta \sqrt{p} + \frac{4\phi_{\max}(\sqrt{p} + R_\beta p)}{\kappa + \lambda} \right\} \delta_q.$$

Thus one valid main-text choice is

$$C_\phi = 2R_\beta \sqrt{p} + 4\phi_{\max}(\sqrt{p} + R_\beta p),$$

which yields $L_\phi \le C_\phi\{1 + (\kappa + \lambda)^{-1}\}$. For propensity-clipped IPW and orthogonal-score estimators, clipping $\widehat{e}(x)$ to $[\eta, 1 - \eta]$ makes $a \mapsto 1/a$ and $a \mapsto 1/(1 - a)$ Lipschitz with constants of order η^{-2} on the clipped interval, and the same decomposition applies with additional polynomial factors in $1/\eta$. The unregularized case follows by setting $\lambda = 0$ when $\kappa > 0$. □

Remark 3 (Why the coefficient radius is explicit). The $O(1/(\kappa+\lambda))$ dependence for Gram perturbations uses $\left\|\widehat{\beta}_{t,\lambda}\right\|_2 \le R_\beta$, enforceable by coefficient clipping or a bounded parameter set (both post-processing). Without a radius, the same identity yields a valid bound with the Gram term scaling as $O(\|r_t\|_2 / (\kappa + \lambda)^2)$. We state the radius explicitly so the theorem does not hide a regularity condition inside the constant.

Proof of Corollary 1. By Gaussian tails and a union bound, with probability at least $1 - \gamma$, the released workload vector obeys $\|\widetilde{q} - q(D)\|_\infty \le \delta_m = c_0 \sigma \sqrt{\log(m/\gamma)}$. In a full joint-cell basis the Gram perturbation is diagonal, so $\max_t \left\|\widetilde{G}_t - G_t\right\|_{\mathrm{op}} \le \delta_m$. For explicitly released dense Gram blocks, the same conclusion follows under the stated operator-norm event; a standard Gaussian-matrix bound gives this event when λ is at least a constant multiple of $\sigma\sqrt{p + \log(1/\gamma)}$.

$$\lambda_{\min}(\widetilde{G}_t + \lambda I) \ge \kappa + \lambda - \left\|\widetilde{G}_t - G_t\right\|_{\mathrm{op}} \ge \kappa + \lambda/2$$

whenever $\max_t \left\|\widetilde{G}_t - G_t\right\|_{\mathrm{op}} \le \lambda/2$. Theorem 1 then gives the displayed main-text rate. □

Conditioning intuition. With full joint-cell indicators, an arm-specific Gram matrix loses rank exactly when some cell contains few or no records from that treatment arm. With concatenated bin indicators or other overlapping bases, the same issue appears through missing co-occurrences or near-collinear Gram blocks. DP noise of scale σ further perturbs each released coordinate, so at small ε and large p the effective κ can be zero and the unregularized moment-to-ATE map is not Lipschitz. Corollary 1 restores stability by setting λ at the privacy-noise scale. This plays a different role from SNR thresholding: SNR thresholding stops the max-entropy solver from chasing moments that are indistinguishable from privacy noise, while ridge keeps the nuisance inverse stable when the retained moments still leave an arm's Gram matrix near-singular.

E.3 STABILITY OF CLIPPED IPW AND AIPW SCORES

This appendix makes Proposition 1 explicit for partition-feature workloads. The argument is standard perturbation calculus for clipped inverse-propensity scores; it is included only to show how the same moment-error bound propagates through IPW/AIPW post-processing.

Lemma 2 (Clipped IPW/AIPW perturbation)**.** *Let $\phi_j(x) = \mathbb{1}\{x \in A_j\}$, $j = 1, \dots, p$, be partition indicators, $|Y| \le B$, and define $p_{tj}(q) = q_{tj}^{(0)}$, $s_j(q) = p_{0j}(q) + p_{1j}(q)$, and $r_{tj}(q) = q_{tj}^{(1)}$. For two admissible moment vectors q, q', set $\delta_q = \|q - q'\|_\infty$ and $\Delta_e = \max_j |\widehat{e}_j(q) - \widehat{e}_j(q')|$, where $\widehat{e}_j(\cdot) \in [\eta, 1-\eta]$. Then the clipped IPW functional*

$$\widehat{\tau}_{\mathrm{IPW}}(q) = \sum_{j=1}^{p} \left\{ \frac{r_{1j}(q)}{\widehat{e}_j(q)} - \frac{r_{0j}(q)}{1 - \widehat{e}_j(q)} \right\}$$

satisfies

$$|\widehat{\tau}_{\mathrm{IPW}}(q) - \widehat{\tau}_{\mathrm{IPW}}(q')| \le \frac{2p}{\eta}\delta_q + \frac{B}{\eta^2}\Delta_e.$$

If $\widehat{e}_j(q) = \mathrm{clip}\{p_{1j}(q)/s_j(q), \eta, 1-\eta\}$ and $\min_{j, q_\circ \in \{q, q'\}} s_j(q_\circ) \ge \rho$, then $\Delta_e \le 3\delta_q/\rho$.

Now suppose $|\widehat{m}_{tj}(q)| \le B_m$ and set $\Delta_m = \max_{t,j} |\widehat{m}_{tj}(q) - \widehat{m}_{tj}(q')|$. The partition AIPW functional

$$\begin{aligned}\widehat{\tau}_{\mathrm{AIPW}}(q) = &\sum_{j=1}^{p} s_j(q)\{\widehat{m}_{1j}(q) - \widehat{m}_{0j}(q)\} \\ &+ \sum_{j=1}^{p} \frac{r_{1j}(q) - p_{1j}(q)\widehat{m}_{1j}(q)}{\widehat{e}_j(q)} - \sum_{j=1}^{p} \frac{r_{0j}(q) - p_{0j}(q)\widehat{m}_{0j}(q)}{1 - \widehat{e}_j(q)}\end{aligned}$$

satisfies

$$|\widehat{\tau}_{\mathrm{AIPW}}(q) - \widehat{\tau}_{\mathrm{AIPW}}(q')| \le \left(4B_m p + \frac{2p(1+B_m)}{\eta}\right)\delta_q + 2\left(1 + \frac{1}{\eta}\right)\Delta_m + \frac{B + B_m}{\eta^2}\Delta_e.$$

Consequently, if the fitted nuisance maps obey $\Delta_e \le L_e\delta_q$ and $\Delta_m \le L_m\delta_q$, clipped IPW and AIPW are Lipschitz in the released moment vector with the displayed estimator-specific constants.

Proof. For $u, v \in [\eta, 1-\eta]$, $|1/u - 1/v| \le |u - v|/\eta^2$ and $|1/(1-u) - 1/(1-v)| \le |u-v|/\eta^2$. Apply these inequalities after adding and subtracting terms with either the moments held fixed or the nuisance functions held fixed. The partition basis gives $\sum_j |r_{1j}| + \sum_j |r_{0j}| \le B$, $\sum_j p_{1j} + \sum_j p_{0j} \le 1$, and $\sum_j |s_j(q) - s_j(q')| \le 2p\delta_q$. These bounds yield the IPW and AIPW displays. For the cell propensity ratio, clipping is nonexpansive and the map p_{1j}/s_j has perturbation at most $3\delta_q/\rho$ when $s_j \ge \rho$. □

E.4 PROOF OF THEOREM 2: DP MOMENT ACCURACY

Write $q(D) = n^{-1}\sum_i h(W_i)$ with $\|h(W_i)\|_2 \le H_q$. Under replacement adjacency, changing one record changes $q(D)$ by at most

$$\Delta_2 \le \frac{2H_q}{n}.$$

Applying the Gaussian mechanism to $q(D)$ adds $Z \sim \mathcal{N}(0, \sigma^2 I_m)$ with $\sigma = \Delta_2 \sqrt{2\log(1.25/\delta)}/\varepsilon$. By a standard Gaussian tail bound and a union bound over the m coordinates, with probability at least $1-\gamma$,

$$\|Z\|_\infty \le \sigma\sqrt{2\log(m/\gamma)}.$$

Substituting the expression for σ yields the stated rate. For the base workload, $\|h(W)\|_2^2 = (1+Y^2)\,\|\phi(X)\|_2^2 \le \phi_{\max}^2(1+B^2)$. For an optional Gram-augmented workload, one additional active Gram block contributes $\left\|\phi(X)\phi(X)^\top\right\|_F^2 = \|\phi(X)\|_2^4 \le \phi_{\max}^4$, giving $H_q^2 \le \phi_{\max}^2(1+B^2)+\phi_{\max}^4$. In the experiments, we use $B=5$ (outcomes clipped to $[-5,5]$). □

E.5 PROOF OF THEOREM 3: ATE ERROR DECOMPOSITION

Let P^{syn} be the actual solver output. Write $\widehat{\tau}(q)$ for the population moment-map estimator and $\widehat{\tau}^{\mathrm{syn}}$ for its empirical version computed from n_{syn} synthetic draws. Decompose the error via the triangle inequality:

$$\begin{aligned}|\widehat{\tau}^{\mathrm{syn}} - \tau| \le\ & \underbrace{|\tau - \tau_{\phi,S}|}_{\text{(iii) approximation}} + \underbrace{|\tau_{\phi,S} - \widehat{\tau}(q^\star)|}_{\text{= 0 by definition}} + \underbrace{|\widehat{\tau}(q^\star) - \widehat{\tau}(q(D))|}_{\text{(i) sampling}} \\ & + \underbrace{|\widehat{\tau}(q(D)) - \widehat{\tau}(\widetilde{q})|}_{\text{(ii) privacy}} + \underbrace{|\widehat{\tau}(\widetilde{q}) - \widehat{\tau}(q(P^{\mathrm{syn}}))|}_{\text{(v) calibration}} + \underbrace{|\widehat{\tau}(q(P^{\mathrm{syn}})) - \widehat{\tau}^{\mathrm{syn}}|}_{\text{(iv) Monte Carlo}},\end{aligned} \tag{7}$$

where $\tau_{\phi,S}$ is the target represented by the retained workload coordinates.

Term (i): By the CLT, $\|q(D) - q^\star\|_\infty = O_p(n^{-1/2})$ (concentration of bounded averages). By the estimator Lipschitz condition, $|\widehat{\tau}(q^\star) - \widehat{\tau}(q(D))| \le L_{\mathrm{est}} O_p(n^{-1/2})$.

Term (ii): By Theorem 2, $\|\widetilde{q} - q(D)\|_\infty = O_p(H_q\sqrt{\log m \cdot \log(1/\delta)}/(n\varepsilon))$. Applying the same Lipschitz condition gives the privacy term.

Term (iii): This is $\mathrm{Approx}(\phi; S) := |\tau - \tau_{\phi,S}|$, the bias from representing outcome and propensity models in a finite retained basis. This term is zero when m_t and e are linear in ϕ and nonzero otherwise; its magnitude depends on the richness of ϕ.

Term (iv): The synthetic data estimator $\widehat{\tau}^{\mathrm{syn}}$ is computed from n_{syn} i.i.d. draws from P^{syn}. By the CLT for the synthetic population, this contributes $O_p(n_{\mathrm{syn}}^{-1/2})$.

Term (v), calibration: If exact matching is infeasible because the measurements are noisy, define the relaxation residual directly as $\Delta_{\mathrm{cal}}(S) := \Pi_S\{\widetilde{q} - q(P^{\mathrm{syn}})\}$. The estimator Lipschitz condition bounds the calibration term by

$$|\widehat{\tau}(\widetilde{q}) - \widehat{\tau}(q(P^{\mathrm{syn}}))| \le L_{\mathrm{est}}\,\|\Delta_{\mathrm{cal}}(S)\|_2 = \mathrm{CalGap}(S, \sigma).$$

Combining all five terms yields the stated bound. □

Corollary 3 (SNR thresholding controls the calibration gap)**.** *Let $S_\tau = \{a : |\widetilde{q}_a|/\sigma_a \ge \tau_{\mathrm{SNR}}\}$ be the retained set, where σ_a is the Gaussian-mechanism standard deviation of coordinate a. Suppose the max-entropy solver is run only on S_τ and stopped at noise-level tolerance*

$$|\widetilde{q}_a - q_a(P^{\mathrm{syn}})| \le c_{\mathrm{cal}}\,\sigma_a \qquad \textit{for all } a \in S_\tau, \tag{8}$$

for a fixed numerical constant c_{cal}. If $\sigma_a \le \bar{\sigma}$ on S_τ, then $\mathrm{CalGap}(S_\tau, \sigma) \le L_{\mathrm{est}} c_{\mathrm{cal}} \bar{\sigma}\sqrt{m_{\mathrm{kept}}}$ with $m_{\mathrm{kept}} = |S_\tau|$; in the homoskedastic case $\bar{\sigma} = \sigma$, so $\mathrm{CalGap}(S_\tau, \sigma) = O(L_{\mathrm{est}}\sigma\sqrt{m_{\mathrm{kept}}})$.

Proof. By definition of CalGap and the tolerance (8),

$$\mathrm{CalGap}(S_\tau, \sigma) = L_{\mathrm{est}}\Big(\sum_{a\in S_\tau} |\widetilde{q}_a - q_a(P^{\mathrm{syn}})|^2\Big)^{1/2} \le L_{\mathrm{est}} c_{\mathrm{cal}}\Big(\sum_{a\in S_\tau}\sigma_a^2\Big)^{1/2},$$

and the claim follows from $\sigma_a \le \bar{\sigma}$ and $|S_\tau| = m_{\mathrm{kept}}$. The tolerance condition is essential: thresholding alone cannot bound arbitrary optimization error; the guarantee is for the SNR-thresholded, noise-tolerant calibration rule. Moments excluded by S_τ are not counted in $\mathrm{CalGap}(S_\tau, \sigma)$ because the solver no longer matches them; their effect moves into $\mathrm{Approx}(\phi; S_\tau)$, which is the bias–variance tradeoff measured in the workload-dimension ablation. □

F BIAS-AWARE NA+MI INTERVAL AND THE SNR DEPLOYMENT DIAGNOSTIC

This section gives the coverage-corrected interval motivated in Appendix L; it turns the coverage–privacy paradox into a measurable diagnostic and a conservative correction. Throughout, $\widehat{\tau}_{\mathrm{NA+MI}}$ is the NA+MI point estimate, $\sigma^2_{\mathrm{NA+MI}}$ its Rubin-rules variance (which accounts for DP moment noise), and $\widehat{\mathrm{Approx}}(\phi)$ an estimate of the workload approximation bias defined below.

Proposition 4 (Bias-aware NA+MI interval)**.** *Suppose the estimator admits the expansion* $\widehat{\tau}_{\mathrm{NA+MI}} - \tau = b_\phi + \sigma_{\mathrm{NA+MI}} Z + o_p(\sigma_{\mathrm{NA+MI}} + |b_\phi|)$ *with* $Z \Rightarrow \mathcal{N}(0,1)$*, where* b_ϕ *is the leading workload approximation bias, and that the diagnostic is conservative:* $\Pr\{|b_\phi| \le \widehat{\mathrm{Approx}}(\phi) + o_p(\sigma_{\mathrm{NA+MI}})\} \to 1$*. Define* $\widetilde{\sigma}^2 = \sigma^2_{\mathrm{NA+MI}} + \widehat{\mathrm{Approx}}(\phi)^2$ *and* $\mathrm{CI}_{\mathrm{corr}}(\alpha) = \widehat{\tau}_{\mathrm{NA+MI}} \pm z_{1-\alpha/2}\,\widetilde{\sigma}$*. Then for confidence levels with* $z_{1-\alpha/2} \ge \sqrt{3}$ *(including standard 95% intervals),* $\mathrm{CI}_{\mathrm{corr}}(\alpha)$ *has asymptotic coverage at least* $1-\alpha$*. When* $b_\phi = 0$ *it reduces to the usual NA+MI interval; when* $|b_\phi| \gg \sigma_{\mathrm{NA+MI}}$ *it becomes conservative rather than collapsing around a biased center.*

Proof. Condition on the event $|b_\phi| \le A + o_p(\sigma_{\mathrm{NA+MI}})$ with $A = \widehat{\mathrm{Approx}}(\phi)$. Ignoring the lower-order term, coverage is bounded below by $\Pr\{|\sigma_{\mathrm{NA+MI}} Z + b_\phi| \le z_{1-\alpha/2}\sqrt{\sigma^2_{\mathrm{NA+MI}} + A^2}\}$, which for fixed A is minimized over $|b_\phi| \le A$ at $|b_\phi| = A$, where it equals

$$f(a) = \Phi\big(zs(a) - a\big) + \Phi\big(zs(a) + a\big) - 1,$$

with $z = z_{1-\alpha/2}$, $a = A/\sigma_{\mathrm{NA+MI}}$, and $s(a) = \sqrt{1+a^2}$. Differentiating,

$$f'(a) = \phi\big(zs(a) - a\big)\Big(\tfrac{za}{s(a)} - 1\Big) + \phi\big(zs(a) + a\big)\Big(\tfrac{za}{s(a)} + 1\Big),$$

where ϕ denotes the standard normal density. If $za/s(a) \ge 1$ then $f'(a) \ge 0$ directly. Otherwise $f'(a) \ge 0$ is equivalent to $\exp\{-2zas(a)\} \ge \frac{1-za/s(a)}{1+za/s(a)}$, i.e., writing $y = za/s(a)$, to $\operatorname{arctanh}(y) \ge y/(1-y^2/z^2)$. For $z^2 \ge 3$ this holds termwise: $\operatorname{arctanh}(y) = \sum_{k\ge 0} y^{2k+1}/(2k+1)$, $y/(1-y^2/z^2) = \sum_{k\ge 0} y^{2k+1}/z^{2k}$, and $z^{2k} \ge 3^k \ge 2k+1$ for $k \ge 1$. Hence f is nondecreasing on $a \ge 0$ whenever $z \ge \sqrt{3}$, so $f(a) \ge f(0) = 2\Phi(z) - 1 = 1 - \alpha$. □

Estimating $\widehat{\mathrm{Approx}}(\phi)$**.** Let $\widehat{\psi}_\phi(W)$ be the estimated orthogonal ATE score with nuisances fitted in the ϕ basis, centered at $\widehat{\tau}_{\mathrm{NA+MI}}$, and let $b(X) \in \mathbb{R}^r$ be a fixed diagnostic dictionary richer than ϕ, scaled so that $\mathbb{E}[b(X)b(X)^\top] \approx I$ under the synthetic distribution. With the residual score moment $\widehat{R}_\phi = \big\| n_{\mathrm{syn}}^{-1} \sum_i \widehat{\psi}_\phi(W_i^{\mathrm{syn}})\, b(X_i^{\mathrm{syn}}) \big\|_2$, a practical conservative choice is $\widehat{\mathrm{Approx}}(\phi) = \widehat{L}_{\mathrm{diag}} \widehat{R}_\phi$, where $\widehat{L}_{\mathrm{diag}}$ is the same empirical Lipschitz/overlap constant used to score candidate features in CAUSAL-AIM. Large residual moments indicate that the current workload leaves systematic orthogonal-score structure unexplainedin the regime at high $\varepsilon$where DP noise shrinks but approximation bias remains. Both quantities are computable from the synthetic release and a public pilot or held-out covariate sample, so the ratio $\widehat{\mathrm{Approx}}(\phi)/\sigma_{\mathrm{NA+MI}}$ can be tracked at deployment.

Practical rule. Keep moment j iff $\mathrm{SNR}_j := |\widetilde{q}_j|/\sigma_j \ge \tau_{\mathrm{SNR}}$ (default $\tau_{\mathrm{SNR}} = 3$) as the calibration filter, and report the corrected interval $\widehat{\tau}_{\mathrm{NA+MI}} \pm z_{1-\alpha/2}(\sigma^2_{\mathrm{NA+MI}} + \widehat{\mathrm{Approx}}(\phi)^2)^{1/2}$. If $\widehat{\mathrm{Approx}}(\phi) \ll \sigma_{\mathrm{NA+MI}}$, the correction is negligible and DP noise dominates; if $\widehat{\mathrm{Approx}}(\phi) \gtrsim \sigma_{\mathrm{NA+MI}}$, the analysis is approximation-dominated, which is the diagnostic signature of the coverage–privacy paradox at high ε.

Connection to robust Bayes. The correction treats workload approximation as local model misspecification: NA+MI accounts for posterior uncertainty induced by DP noise, while the $\widehat{\mathrm{Approx}}(\phi)^2$ term enlarges the interval to cover discrepancy between the working workload model and the data-generating law. This correction is the additive, workload-aware analogue of robust-Bayes and misspecification-aware posterior calibration [Bissiri et al., 2016, Watson and Holmes, 2016, Grünwald and van Ommen, 2017, Miller and Dunson, 2019], which inflate posterior spread to remain valid under model misspecification.

Algorithm 4 Noise-Aware Multiple Imputation (NA+MI)

Require: Measured DP moments $\widetilde{q}_S$ on coordinates S, mechanism noise covariance Σ_S, number of draws M, synthetic size n_{syn}, level α, optional SNR threshold τ_{SNR}

Compute the retained set $S_\tau = \{j \in S : |\widetilde{q}_j|/\sigma_j \geq \tau_{\mathrm{SNR}}\}$ once from $\widetilde{q}_S$ (take $S_\tau = S$ if thresholding is off)

for $\ell = 1, \ldots, M$ **do**

  Draw $q_S^{(\ell)} \sim \mathcal{N}(\widetilde{q}_S, \Sigma_S)$ ▷ flat-prior Gaussian posterior on measured coordinates

  $P^{\mathrm{syn},(\ell)} \leftarrow$ max-entropy calibration of (3) on $\Pi_{S_\tau} q_S^{(\ell)}$ (Algorithm 2)

  $D^{\mathrm{syn},(\ell)} \leftarrow n_{\mathrm{syn}}$ ancestral-sampling draws from $P^{\mathrm{syn},(\ell)}$

  $(\widehat{\tau}^{(\ell)}, \widehat{v}^{(\ell)}) \leftarrow$ doubly robust ATE estimate and its variance estimate on $D^{\mathrm{syn},(\ell)}$

end for

$\bar{\tau} \leftarrow M^{-1}\sum_\ell \widehat{\tau}^{(\ell)}$; $W_M \leftarrow M^{-1}\sum_\ell \widehat{v}^{(\ell)}$; $B_M \leftarrow (M-1)^{-1}\sum_\ell (\widehat{\tau}^{(\ell)} - \bar{\tau})^2$

$T_M \leftarrow W_M + (1 + 1/M)B_M$; $\nu_M \leftarrow (M-1)\big(1 + \frac{W_M}{(1+1/M)B_M}\big)^2$ ▷ Rubin's rules [Rubin, 1987]

Ensure: Point estimate $\bar{\tau}$ and interval $\bar{\tau} \pm t_{\nu_M, 1-\alpha/2}\sqrt{T_M}$

G NA+MI: DETAILED PSEUDOCODE

Algorithm 4 gives an implementation-level version of the noise-aware multiple-imputation procedure of Section 6.

The between-imputation component B_M is what widens the interval as ε decreases: smaller ε means larger DP noise, more dispersed posterior draws $q_S^{(\ell)}$, and hence more variable $\widehat{\tau}^{(\ell)}$. The bias-aware variant of Appendix F replaces T_M by $T_M + \widehat{\mathrm{Approx}}(\phi)^2$.

H EXPERIMENTAL PROTOCOL AND SUPPORTING RESULTS

H.1 FEATURE CONSTRUCTION AND DEFAULTS

For each dataset, continuous covariates are discretized into $L = 5$ quantile bins and the resulting indicator variables define ϕ. When quantile edges coincide, bins collapse, so the effective dimension is data-dependent; IHDP yields $p = 48$ at $L = 5$. The small-L default is consistent with practice in DP quantile estimation, where a handful of privately estimated quantiles per variable is reliable at moderate budgets [Gillenwater et al., 2021, Kaplan et al., 2022, Lei, 2011].

Unless stated otherwise, experiments use $n_{\mathrm{syn}} = 10n$ synthetic records per release, $M = 20$ MI draws, no max-entropy calibration ridge ($\alpha_{\mathrm{cal}} = 0$), and no SNR thresholding in the main pipeline. Outcomes are standardized by fixed public constants and clipped at $B = 5$; constants and clip fractions are reported in Appendix I. The non-private oracle uses unstandardized, unclipped outcomes. Estimated propensities in the DR analysis are clipped to $[0.02, 0.98]$ (a conservative implementation bound, distinct from the positivity constant η of Assumption 1 and from the DGP overlap parameter swept in the overlap ablation); continuous-outcome DR nuisance fits use a fixed ridge linear model, and binary outcomes use logistic regression. Main benchmark experiments use $n_{\mathrm{rep}} = 500$ replications, adaptive and ACS studies use 100, and ablations use 200. Every run logs its full configuration, code version, and data source.

H.2 METRICS

All metrics are computed over n_{rep} independent replications. Let $\widehat{\tau}^{(r)}$ be the ATE estimate, $\mathrm{CI}^{(r)}$ its 95% confidence interval, and τ the true ATE.

- **ATE bias**: $\mathrm{Bias} = \big|n_{\mathrm{rep}}^{-1}\sum_r \widehat{\tau}^{(r)} - \tau\big|$.
- **ATE RMSE**: $\mathrm{RMSE} = \{n_{\mathrm{rep}}^{-1}\sum_r (\widehat{\tau}^{(r)} - \tau)^2\}^{1/2}$.
- **CI coverage**: $\mathrm{Cov} = n_{\mathrm{rep}}^{-1}\sum_r \mathbb{1}\{\tau \in \mathrm{CI}^{(r)}\}$, targeting the nominal 95% level.
- **CI length**: $\mathrm{Len} = n_{\mathrm{rep}}^{-1}\sum_r |\mathrm{CI}^{(r)}|$.
- **Marginal fidelity**: $\mathrm{TVD} = d^{-1}\sum_{j=1}^{d} \mathrm{TV}(\hat{P}_j, \hat{P}_j^{\mathrm{syn}})$, the average total-variation distance across one-dimensional marginals.

H.3 EXPERIMENT GRID AND ABLATIONS

- **Causal versus generic workloads**: compare ATE RMSE and bias for causal workload synthesis, MST, and AIM over $\varepsilon \in \{0.5, 1, 2, 5\}$ with $\delta = 1/n^2$.
- **Coverage calibration**: compare naive intervals on generic and causal synthetic data against NA+MI intervals under the same privacy budgets.
- **Adaptive selection**: compare CAUSAL-AIM against the fixed causal workload at the same total privacy budget, sweeping $K \in \{1, 3, 5, 10\}$ adaptive rounds.
- **Workload dimension**: vary quantile-bin granularity $L \in \{3, 5, 10, 20\}$, with and without SNR thresholding at $\tau_{\mathrm{SNR}} = 3$.
- **MI draws**: vary $M \in \{5, 10, 20, 50, 100\}$.
- **Synthetic sample size**: vary $n_{\mathrm{syn}}/n \in \{1, 2, 5, 10, 50\}$.
- **Overlap**: vary the positivity constant η in the ACIC DGP and measure the effect on RMSE and coverage.

I ACS SEMI-SYNTHETIC DGP DETAILS

Outcome standardization constants. For DP measurement, each dataset's outcome is standardized as $(Y - m_{\mathrm{pub}})/s_{\mathrm{pub}}$ using the fixed public constants below (treated as public benchmark metadata, consistent with the public clip bound B), then clipped at $B = 5$; all reported estimates and intervals are converted back to original units. The non-private oracle uses unstandardized, unclipped outcomes.

Dataset	m_{pub}	s_{pub}	clip fraction
IHDP	12.12	15.43	0.6%
Twins	0	1	0%
ACIC	−0.317	5.706	0.1%
LaLonde/NSW	5,301	6,624	0.4%
ACS	0	1	1.4%

We use 20 demographic covariates from the 2018 California ACS via `folktables` [Ding et al., 2021]: age, education, marital status, relationship, disability, employment status, citizenship, migration, Hispanic origin, ancestry, nativity, deafness, blindness, cognitive difficulty, sex, race, PUMA, state, class of worker, and place of birth; structurally missing categorical codes are encoded as their own category.

Treatment assignment. Let X_n denote column-standardized covariates. We draw $\beta \sim \mathcal{N}(0, I_d)$, normalize to unit ℓ_2 norm, and set the propensity score $e(x) = \mathrm{expit}\big(\beta^\top x + 0.8\sin(x_1) - 0.4\,x_2 x_3\big)$, rescaled to $[0.1, 0.9]$ to enforce overlap. Treatment is drawn as $T_i \sim \mathrm{Bernoulli}(e(X_i))$.

Outcome model. We draw $g \sim \mathcal{N}(0, I_d)$ (normalized) and define the baseline outcome $\mu_0(x) = g^\top x + 0.6\cos(x_3) + 0.3(x_1^2 - x_2)$. Individual treatment effects are $\tau(x) = 0.8 + 0.5\tanh(x_1) + 0.25\sin(x_4 - x_5)$. Observed outcomes are $Y_i = \mu_0(X_i) + \tau(X_i)\,T_i + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, 1.1^2)$. The ground-truth ATE is the mean of the sampled potential-outcome differences, $\tau = n^{-1}\sum_i \{Y_i(1) - Y_i(0)\}$.

J SUPPLEMENTARY EXPERIMENTAL FIGURES

All figures below supplement the main-text results in Section 7; they are ordered as benchmark diagnostics, ACS validation, design ablations, and the fidelity-versus-causal-utility diagnostic. Bias patterns for the main benchmark comparison appear in Figure 6.

K SUPPLEMENTARY EMPIRICAL ANALYSES

This appendix reports supplementary experiments; all use the defaults of Section 7 at $n_{\mathrm{rep}} = 200$ (scalability: $n_{\mathrm{rep}} = 100$) unless noted.

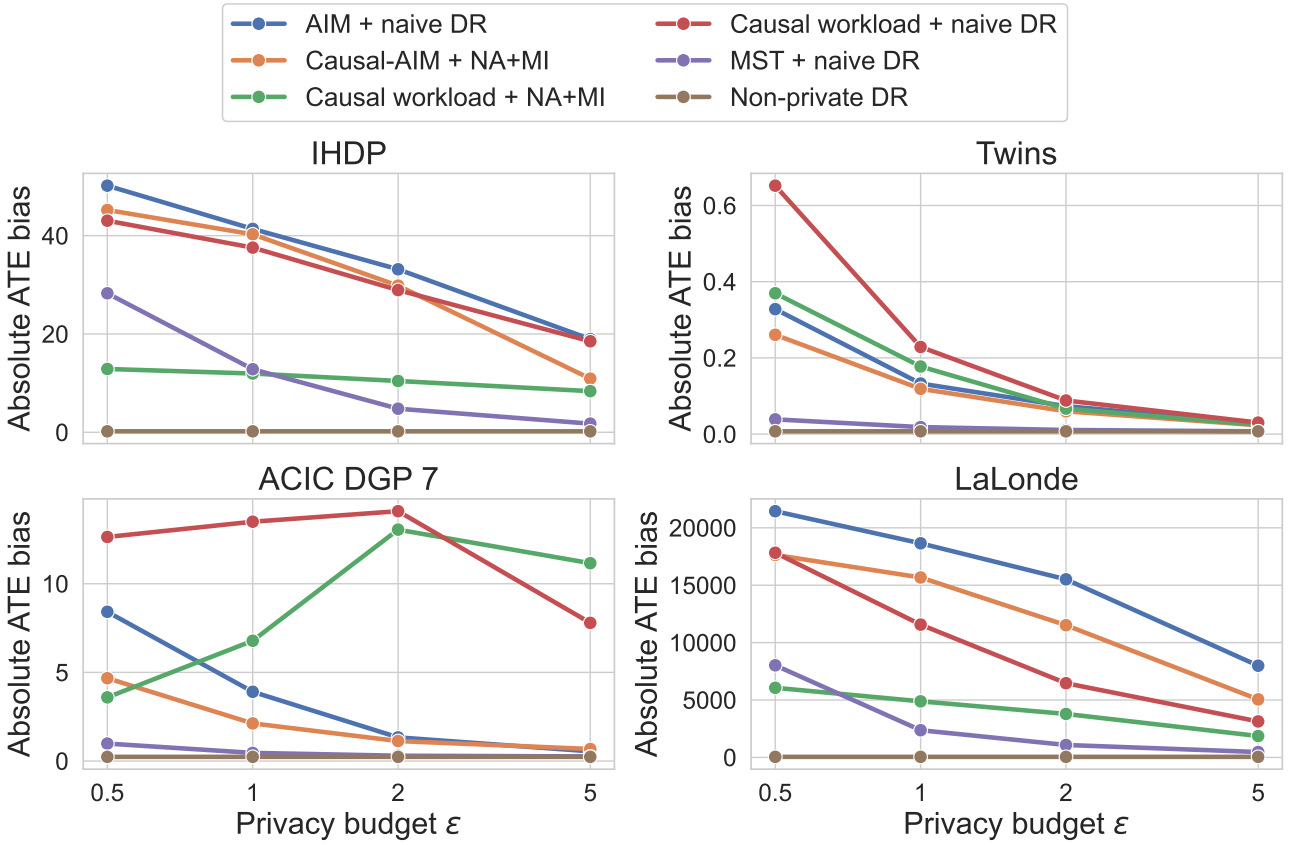

Figure 6: Benchmark bias comparison: Mean absolute ATE error ($n_{\text{rep}}^{-1} \sum_r |\hat{\tau}^{(r)} - \tau|$) across privacy budgets and four benchmark datasets ($n_{\text{rep}} = 500$). Bias patterns by dataset and privacy level mirror the RMSE rankings in Figure 2. At low ε (≤ 1), all methods show substantial bias relative to the non-private oracle; MST + naive DR has the lowest bias on most datasets, with Causal + NA+MI lower on IHDP (and on LaLonde at $\varepsilon = 0.5$). As ε grows, all methods converge toward the oracle.

K.1 MULTI-ESTIMAND REUSE FROM ONE SYNTHETIC RELEASE

From a *single* DP synthetic release, we estimate the ATE, ATT, and a subgroup effect with NA+MI intervals and no additional privacy spending. All eighteen dataset–budget–estimand cells attain nominal coverage (Table 2). Direct DP ATE estimators cannot offer this reuse: each answered query spends additional privacy budget.

Table 2: Multi-estimand reuse from a single DP release: 95% CI coverage (RMSE, original outcome units) of NA+MI intervals for ATE, ATT, and a subgroup effect, all computed from one synthetic dataset per replication ($n_{\text{rep}} = 200$).

Dataset	ε	ATE	ATT	Subgroup
IHDP	0.5	1.00 (15.98)	1.00 (17.48)	1.00 (15.44)
IHDP	1	1.00 (15.63)	1.00 (18.31)	1.00 (14.96)
ACIC	0.5	1.00 (4.36)	1.00 (6.79)	1.00 (4.45)
ACIC	1	1.00 (7.57)	1.00 (9.42)	1.00 (7.59)
LaLonde/NSW	0.5	1.00 (7,613)	1.00 (8,042)	1.00 (7,840)
LaLonde/NSW	1	1.00 (6,688)	1.00 (6,945)	1.00 (6,849)

K.2 DIRECT DP ATE PROXY COMPARISON

The comparison (Table 3) is mixed: the proxies win on single-estimand RMSE on ACIC, while Causal + NA+MI wins on IHDP and attains full coverage everywhere; and the synthetic-data release supports multiple estimands at shared privacy

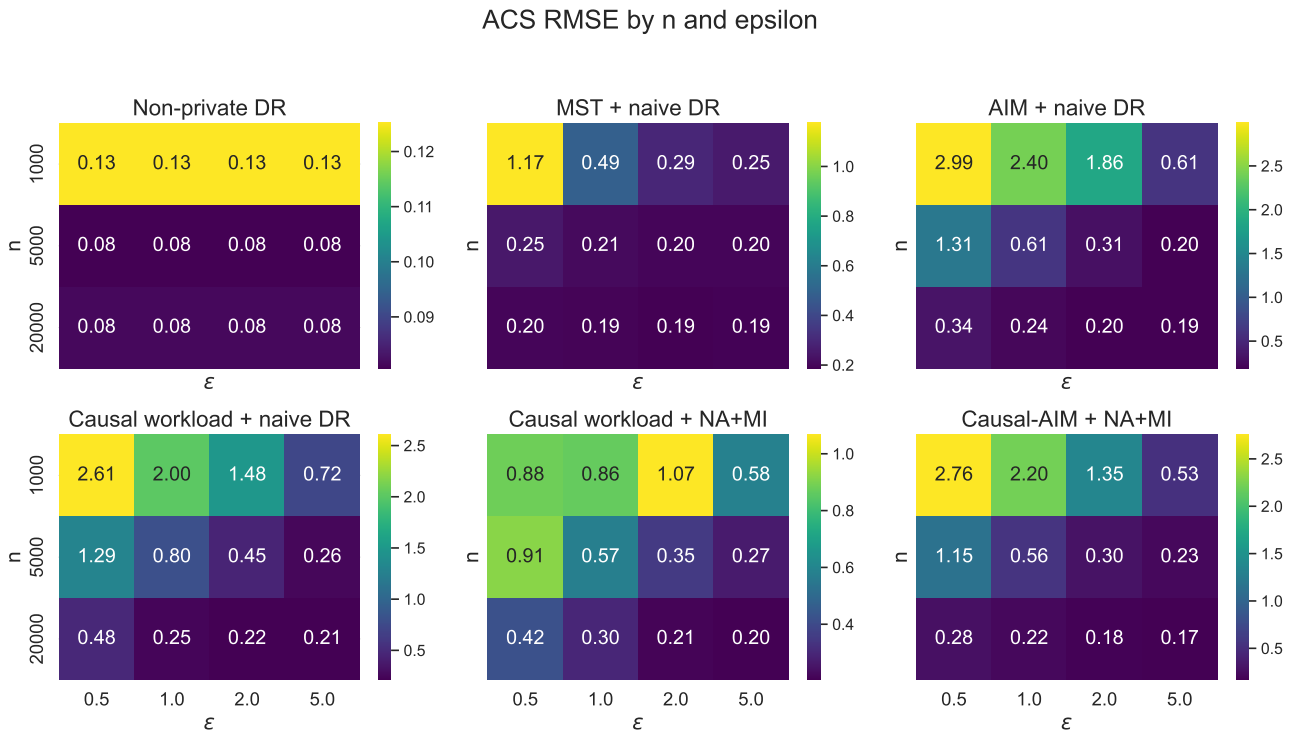

Figure 7: ACS study: ATE RMSE heatmaps on the ACS semi-synthetic study ($n_{\text{rep}} = 100$). Rows index sample size $n \in \{1000, 5000, 20000\}$ and columns index $\varepsilon \in \{0.5, 1, 2, 5\}$. Each panel shows one method. At small n and low ε, all private methods have elevated RMSE; at large n and ε, all converge toward the non-private oracle level. NA+MI is competitive on RMSE at strict budgets (winning at n=1000, ε=0.5) and is the only method with valid coverage (Figure 8).

Table 3: RMSE and coverage against output-perturbation *proxies* of direct DP ATE estimators. The proxies apply DP only at the output rather than to the propensity fit, a weaker DP regime than the published methods [Schröder et al., 2025a, Ohnishi and Awan, 2024]; these results should therefore be read as a favorable contextual approximation of the direct-estimator family. Cells are RMSE / coverage, $n_{\text{rep}} = 200$.

Dataset	ε	PrivATE-style	OA-style	Causal+NA+MI
IHDP	0.5	41.8 / 0.97	39.8 / 0.96	16.4 / 1.00
IHDP	1	18.6 / 0.98	22.0 / 0.96	14.8 / 1.00
ACIC	0.5	2.96 / 0.94	2.74 / 0.97	4.27 / 1.00
ACIC	1	1.47 / 0.93	1.34 / 0.96	8.30 / 1.00

cost (Table 2).

K.3 HYBRID CAUSAL AND GENERIC WORKLOADS

A 50/50 split of the privacy budget between causal moments and generic one-way marginals ($n_{\text{rep}} = 200$; RMSE / coverage):

Dataset	ε	Causal+NA+MI	Hybrid+NA+MI	MST + naive DR
IHDP	0.5	17.6 / 1.00	19.2 / 1.00	36.2 / 0.03
IHDP	1	14.8 / 0.99	18.7 / 1.00	18.5 / 0.01
ACIC	0.5	4.48 / 1.00	5.79 / 1.00	1.20 / 0.14
ACIC	1	8.13 / 1.00	**5.17 / 1.00**	0.64 / 0.22

The hybrid Pareto-improves on the fixed causal workload on ACIC at $\varepsilon = 1$ (lower RMSE at equal coverage) but trails on IHDP; it is an alternative when both point accuracy and coverage matter, not a new default.

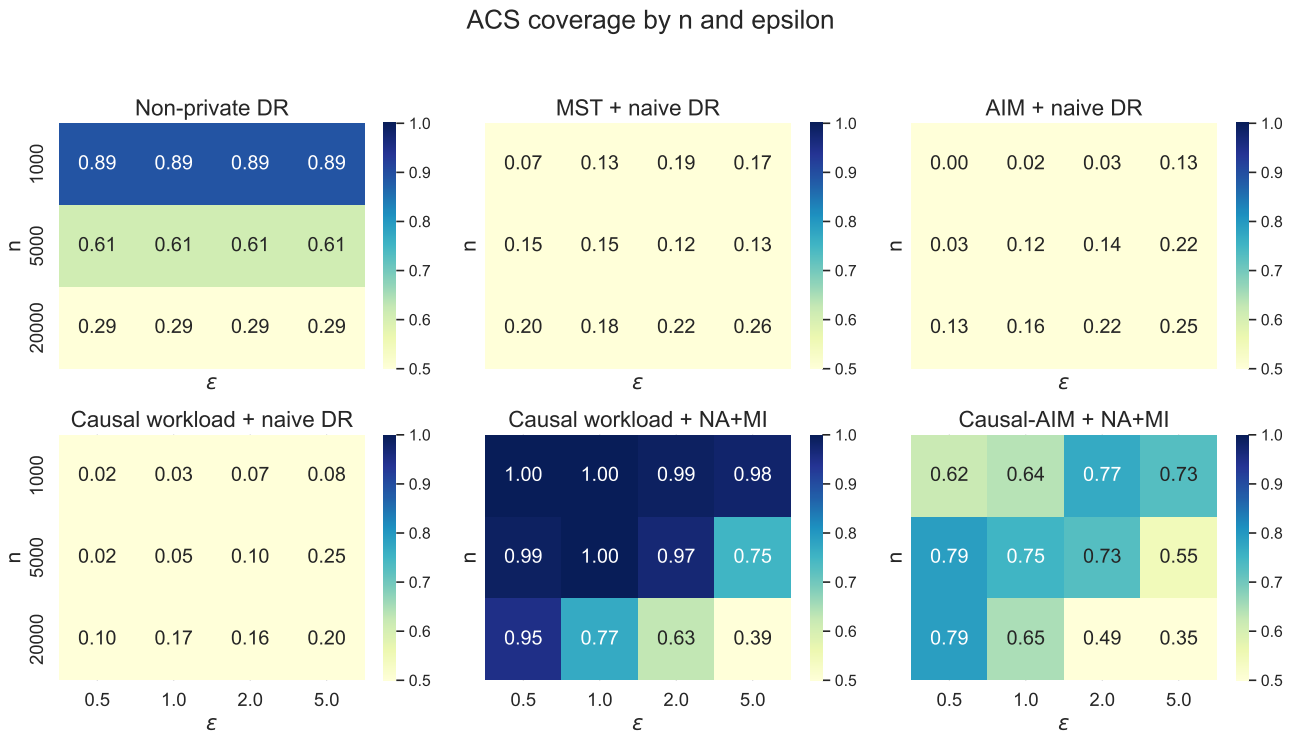

Figure 8: ACS study: Empirical 95% CI coverage on the same ACS grid as Figure 7 ($n_{\text{rep}} = 100$). Causal workload + NA+MI achieves coverage 1.00 at (n=1000, ε=0.5), while MST + naive DR achieves only 0.07. The coverage advantage of NA+MI is most pronounced at low ε and moderate n, where DP noise dominates the error.

K.4 ADAPTIVE SELECTION OPERATING POINT

On IHDP ($K \in \{1, 2, 3, 5, 10\}$, $n_{\text{rep}} = 200$; RMSE / coverage), no K recovers the fixed workload's calibration:

ε	K=1	K=2	K=3	K=5	K=10
0.5	42.2 / 0.02	50.7 / 0.11	50.8 / 0.20	50.3 / 0.31	49.1 / 0.60
1	29.1 / 0.04	37.2 / 0.11	42.1 / 0.16	43.1 / 0.38	42.5 / 0.61
2	14.6 / 0.17	22.7 / 0.14	31.5 / 0.11	34.4 / 0.44	39.8 / 0.64

Contrast with ACIC (Figure 5): whether CAUSAL-AIM helps is dataset-dependent, which is the operating-point message of Section 7.

K.5 SCALABILITY IN COVARIATE DIMENSION

On ACS at $n = 5000$, $\varepsilon = 1$, and $n_{\text{rep}} = 100$, increasing the number of covariates raises both RMSE and synthesis time, with mirror descent on the ϕ-basis as the dominant cost. The $d = 40$ and $d = 60$ configurations extend the 20 base ACS covariates with deterministic derived features (squares, pairwise interactions, and threshold indicators).

Covariates d	RMSE	Median synthesis time
20	0.56	2.9s
40	1.07	5.1s
60	1.38	8.5s

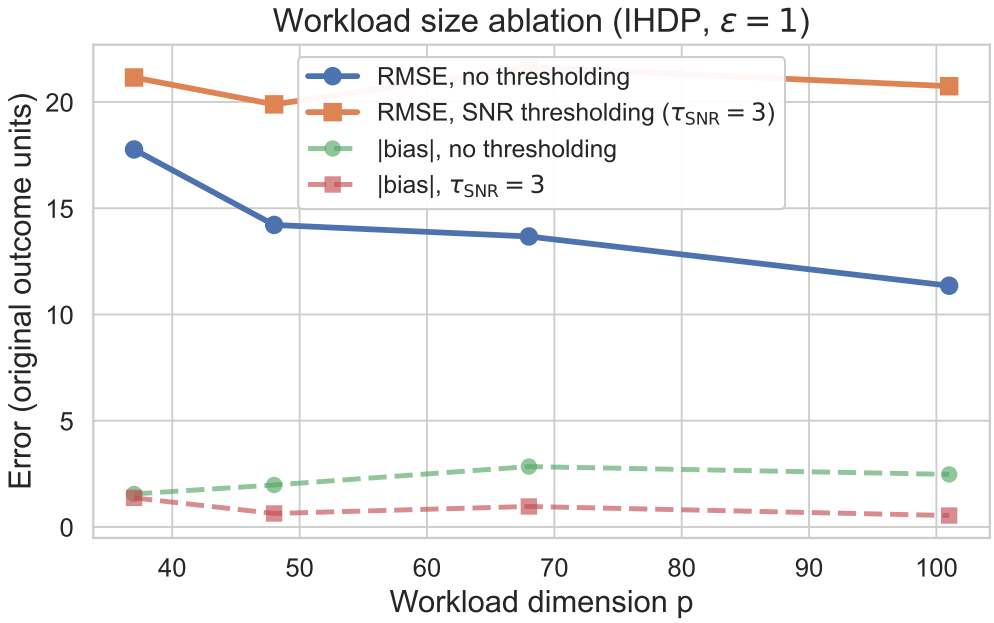

Figure 9: Workload dimension on IHDP at $\varepsilon = 1$ ($n_{\text{rep}} = 200$), with and without SNR thresholding ($\tau_{\text{SNR}} = 3$). Without thresholding, RMSE decreases with p in the tested range ($p \leq 101$); with $\tau_{\text{SNR}} = 3$, RMSE is roughly flat, with lower bias and more variance; the bias-dominated regime of Remark 2 is not reached on this dataset.

K.6 CALIBRATION-RIDGE SENSITIVITY

On IHDP at $\varepsilon = 1$ ($n_{\text{rep}} = 200$), RMSE / coverage across max-entropy calibration ridge values $\alpha_{\text{cal}} \in \{0.001, 0.01, 0.1, 1.0\}$: $14.9/1.00$, $14.2/1.00$, $15.3/1.00$, $18.6/1.00$. Performance is insensitive to this solver regularization over three orders of magnitude (mild degradation only at $\alpha_{\text{cal}} = 1$); this ablation is distinct from the theoretical ridge parameter λ in Corollary 1, and SNR thresholding remains the main calibration filter (Remark 2).

L ADDITIONAL DISCUSSION, LIMITATIONS, AND FUTURE WORK

Why generic fidelity is insufficient. Causal inference requires preserving conditional relationships such as $Y \mid T, X$, not only low-dimensional distributional marginals. A synthetic dataset that matches $P(X, T)$ and $P(X, Y)$ but distorts $P(Y \mid T, X)$ can have good marginal fidelity and poor ATE accuracy. The causal workload makes the target-specific requirement explicit by measuring the moments that identify the orthogonal score in the chosen feature basis.

Practical guidance. A practitioner should specify the target estimand, derive the corresponding orthogonal score and moment requirements, choose a feature map ϕ that approximates those moments, generate synthetic data with a fixed causal workload or CAUSAL-AIM, and report NA+MI uncertainty. On our benchmarks, richer feature maps helped throughout the tested range (L up to 20 bins on IHDP), so the practical default is the richest ϕ the budget supports, using SNR thresholding and the measured CalGap as safeguards when moments approach the noise floor.

Coverage–RMSE tradeoff. The experiments reveal a tradeoff: MST often leads on point RMSE at moderate-to-loose budgets, but its naive confidence intervals are far below nominal coverage at strict budgets. Causal workload + NA+MI is the conservative recommendation when calibrated uncertainty quantification is essential. CAUSAL-AIM + NA+MI is useful when lower RMSE is prioritized and the operating point is favorable, while hybrid causal + generic workloads can help when both point accuracy and coverage matter (Appendix K).

Coverage–privacy paradox. NA+MI coverage can decrease as ε increases because DP noise shrinks while workload approximation bias remains. In ACIC, Causal + NA+MI coverage drops from 1.00 at $\varepsilon \leq 1$ to 0.90 at $\varepsilon = 2$ and 0.17 at $\varepsilon = 5$, while the other three main benchmarks remain at $0.99{-}1.00$. The diagnostic ratio $\widehat{\mathrm{Approx}}(\phi)/\sigma_{\text{NA+MI}}$ identifies

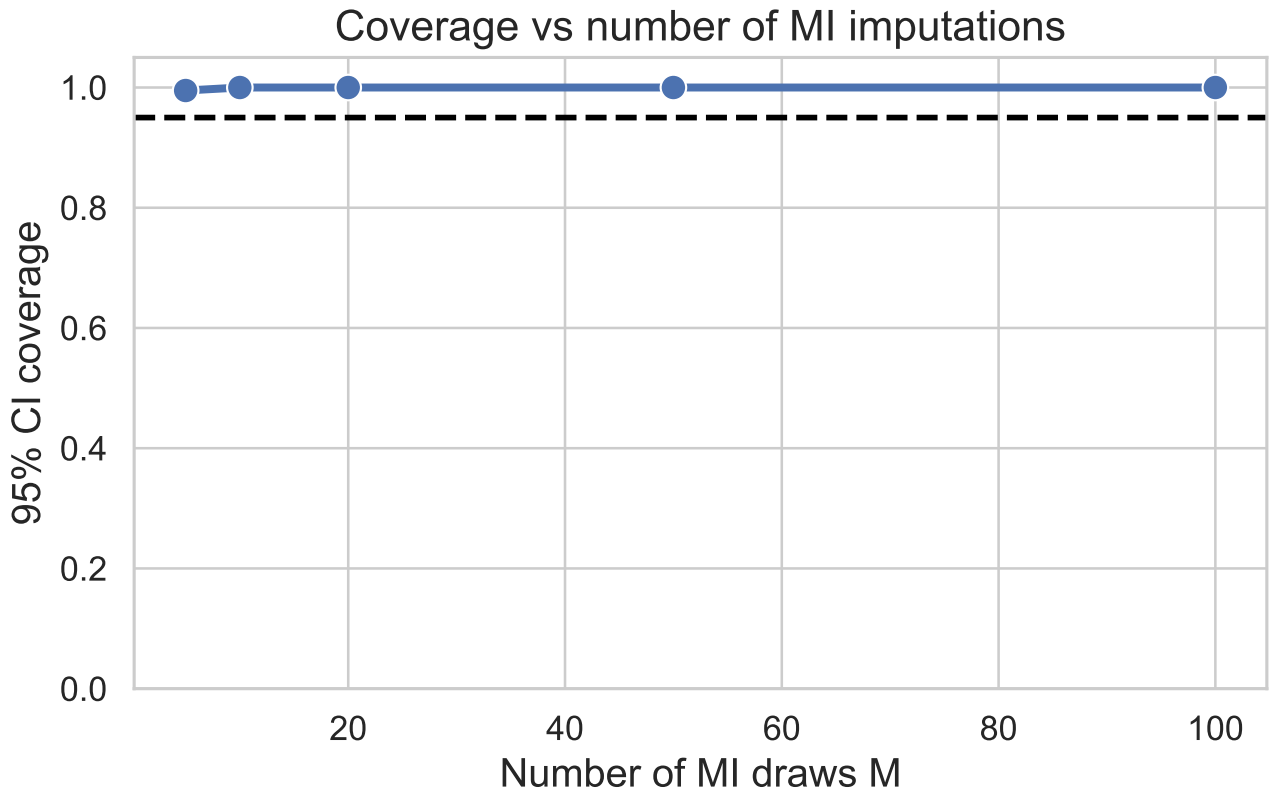

Figure 10: MI draws ablation: empirical 95% CI coverage versus number of imputation draws M on IHDP at $\varepsilon = 1$ ($n_{\text{rep}} = 200$). Coverage is at or above 0.995 for every M tested (0.995 at M=5, 1.000 for $M \geq 10$): the interval is near-nominal already at small M. We use M=20 as a conservative default for noise-aware MI.

this approximation-dominated regime, and Appendix F proves a conservative bias-aware interval that inflates variance by $\widehat{\text{Approx}}(\phi)^2$.

Limitations. The implementation instantiates the framework on the AIM/Private-PGM family, with MST as the generic comparator; the workload-design principle should transfer to other select–measure–reconstruct generators, but we have not evaluated modern generator families beyond this class. The feature map ϕ and discretization of continuous covariates introduce approximation error, especially when the true outcome surface is highly nonlinear. Theorem 3 gives a useful moment-level decomposition, but constants can be loose and the theorem does not fully analyze every downstream fitted-nuisance routine run on sampled synthetic rows. Finally, valid NA+MI intervals can be wide at strict privacy budgets, especially on small benchmarks such as LaLonde/NSW.

Future work. Natural extensions include structured graphical-model solvers and convex relaxations for better calibration; continuous covariates through private kernel mean embeddings or random Fourier features; subgroup-specific and CATE workloads; automated DP-aware selection of ϕ; and information-theoretic lower bounds linking overlap, workload dimension, and unavoidable privacy distortion.

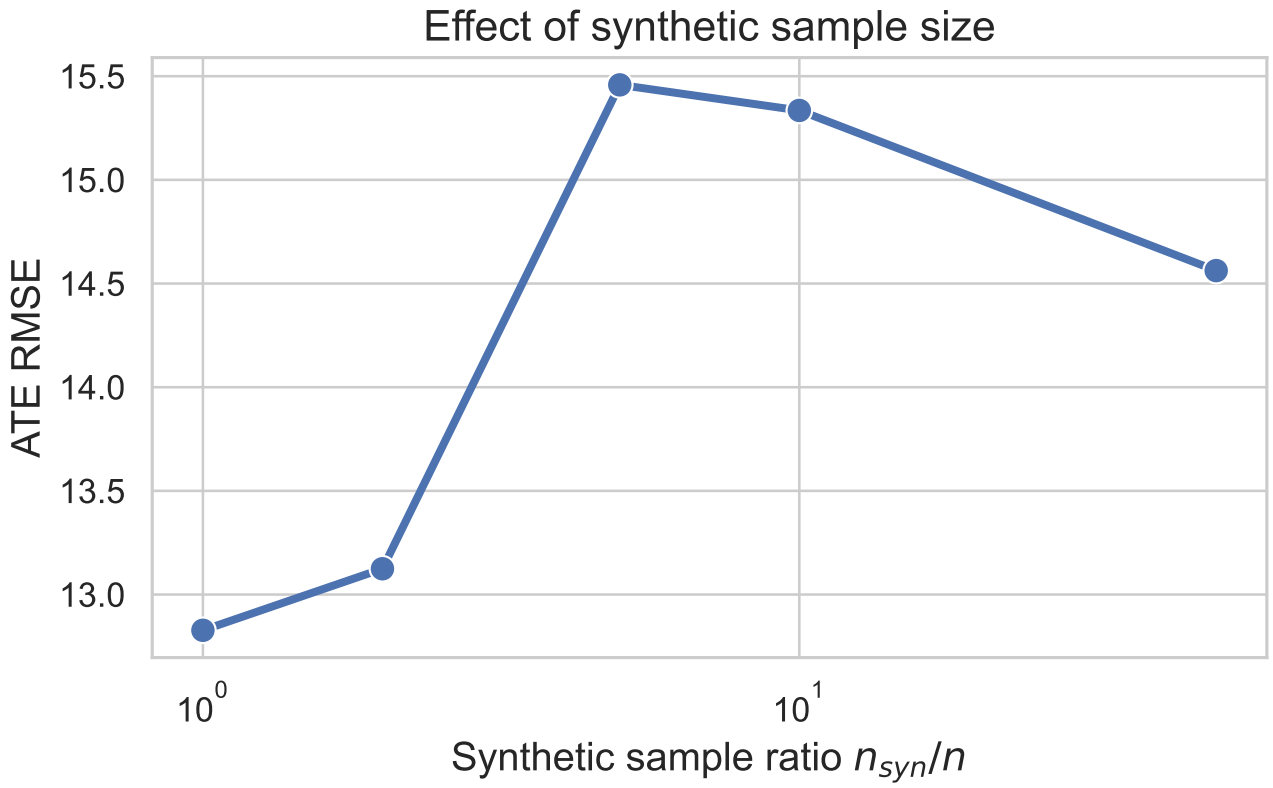

Figure 11: Synthetic sample size ablation: ATE RMSE versus the ratio n_{syn}/n on IHDP at $\varepsilon = 1$ ($n_{\text{rep}} = 200$). RMSE varies modestly (range 12.8–15.5) across ratios from 1 to 50, with no benefit from oversampling: the Monte Carlo term $O_p(n_{\text{syn}}^{-1/2})$ in Theorem 3 is dominated by DP noise already at $n_{\text{syn}} = n$. In practice, $n_{\text{syn}} = n$ suffices.

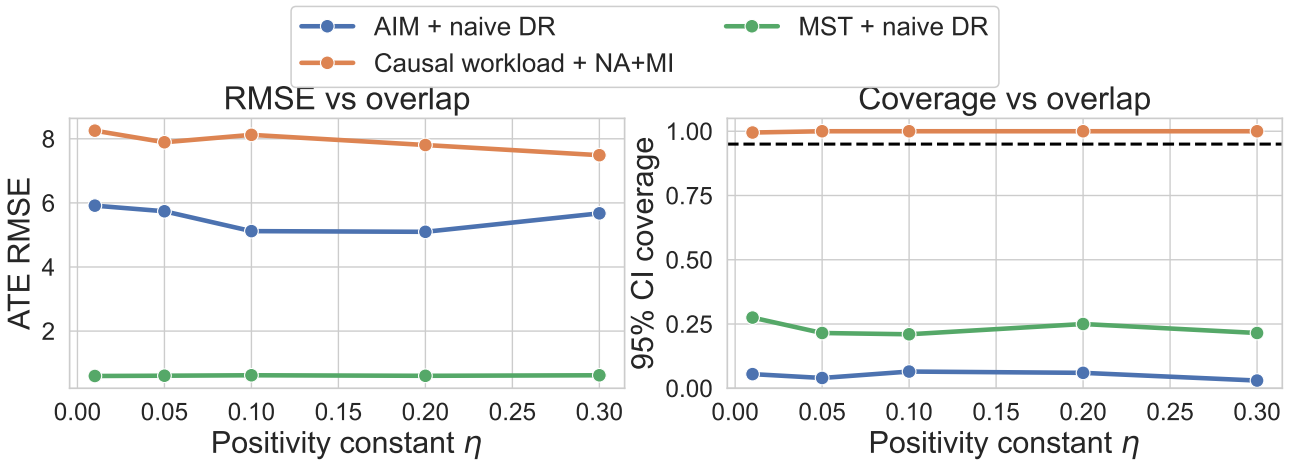

Figure 12: Overlap ablation on ACIC DGP 7 at $\varepsilon = 1$ ($n_{\text{rep}} = 200$). The positivity constant η varies from 0.3 (strong overlap) to 0.01 (near violation). As overlap weakens, MST + naive DR maintains low RMSE (0.60–0.62) but poor coverage (0.21–0.28), while Causal workload + NA+MI has higher RMSE (7.5–8.3) with near-nominal coverage (0.995–1.00) at every η. This confirms that the coverage–RMSE tradeoff persists across overlap regimes and that NA+MI remains the only method with usable confidence intervals.

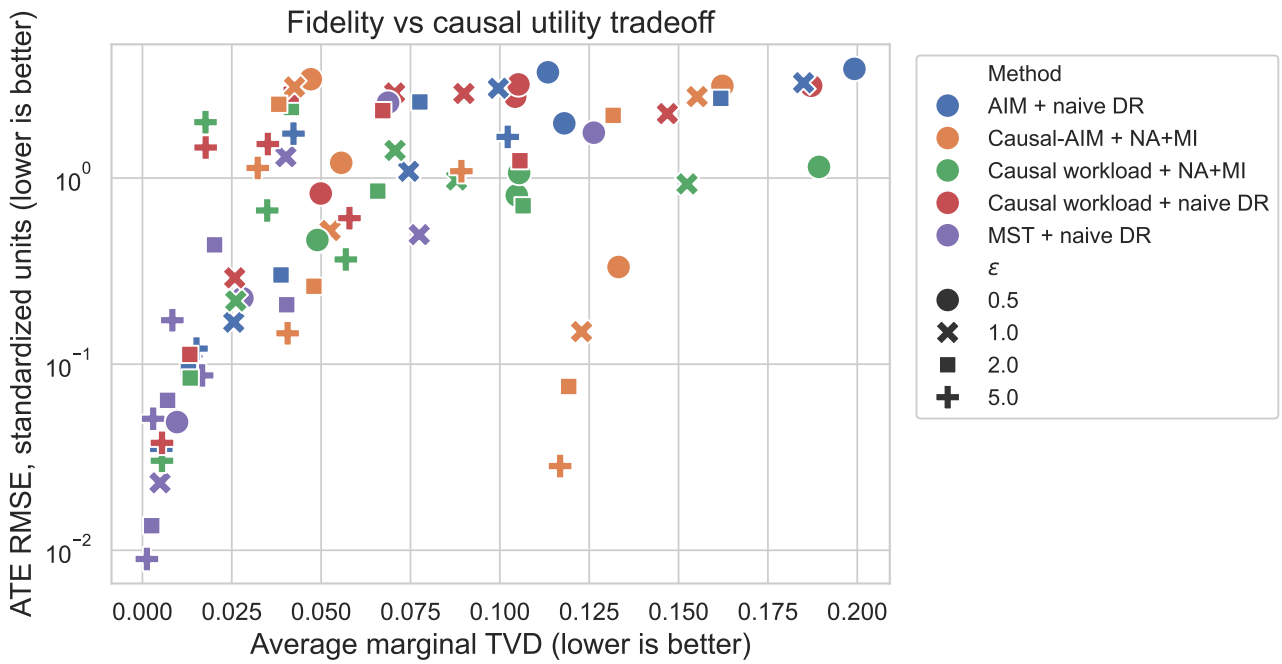

Figure 13: Marginal fidelity versus causal utility across all method–dataset–ε configurations ($n_{\text{rep}} = 500$). Each point represents one experimental cell; the x-axis is average per-marginal TVD and the y-axis is ATE RMSE in standardized outcome units (log scale), so datasets with different outcome scales are comparable. Generic workloads (MST) dominate on marginal fidelity (TVD) in nearly every configuration and on RMSE at $\varepsilon \geq 2$; at $\varepsilon = 0.5$, causal workloads with NA+MI match or beat MST on RMSE on half the benchmarks (and on IHDP at $\varepsilon = 1$) despite worse TVD. Distributional fidelity therefore does not order methods by causal utility; the case for causal workloads rests on noise-aware uncertainty quantification (Figure 3).