PrivATE: Differentially Private Average Treatment Effect Estimation for Observational Data

Quan Yuan[1,2], Xiaochen Li[3], Linkang Du[4], Min Chen[5], Mingyang Sun[6], Yunjun Gao[1], Shibo He[1], Jiming Chen[1,7], and Zhikun Zhang[1]

1 浙江大學 ZHEJIANG UNIVERSITY

2 UNIVERSITY of VIRGINIA

3 UNC GREENSBORO

4 西安交通大學 XI'AN JIAOTONG UNIVERSITY

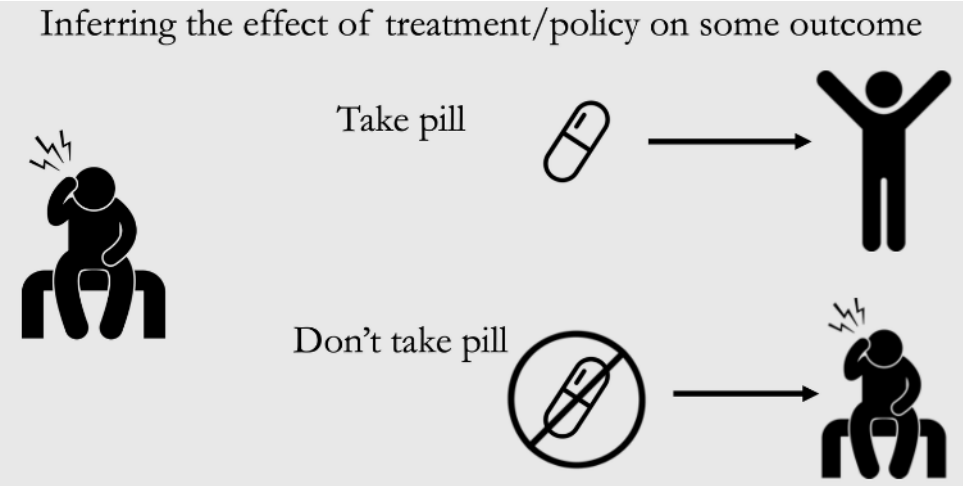
5 VU UNIVERSITY AMSTERDAM

6 PEKING UNIVERSITY

7 杭州電子科技大學 HANGZHOU DIANZI UNIVERSITY

Background

- Causal Inference
  - Inferring the effects of any treatment/policy/intervention/etc.



Medical Care

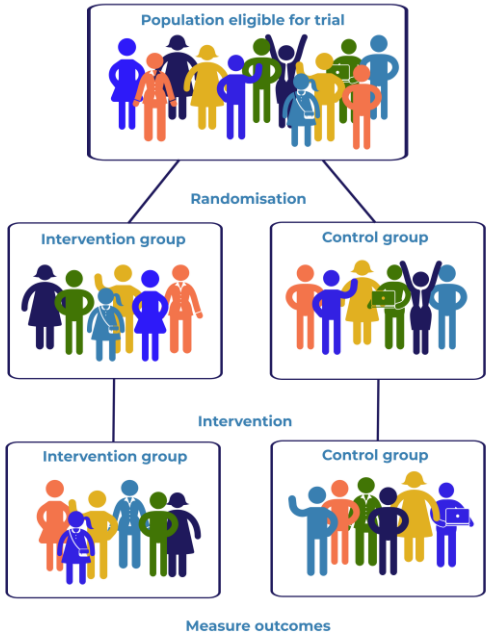
Take the policy A ⟶ High Income

Don't take the policy A ⟶ Low Income

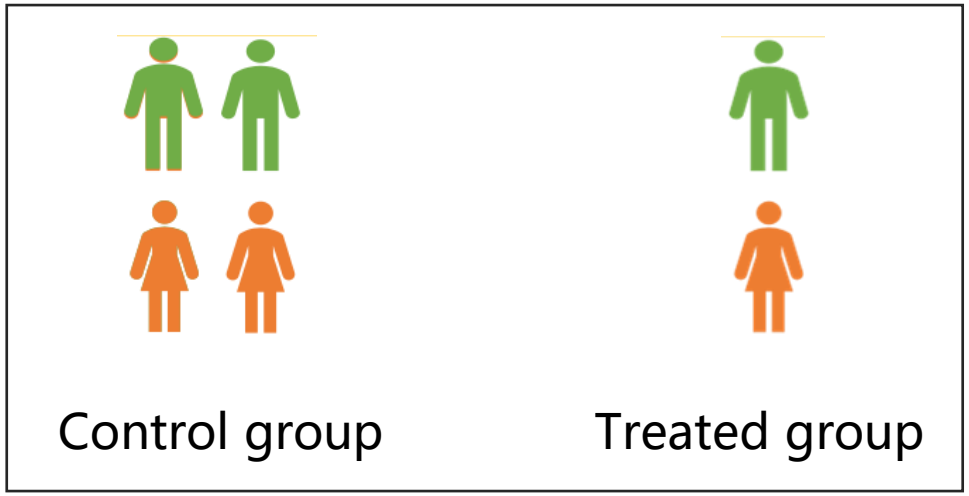
Economic Policy

Background

- Typical Settings
  - RCT (Randomized Controlled Trial)
  - Observational Study (not intervening in individual grouping)



RCT



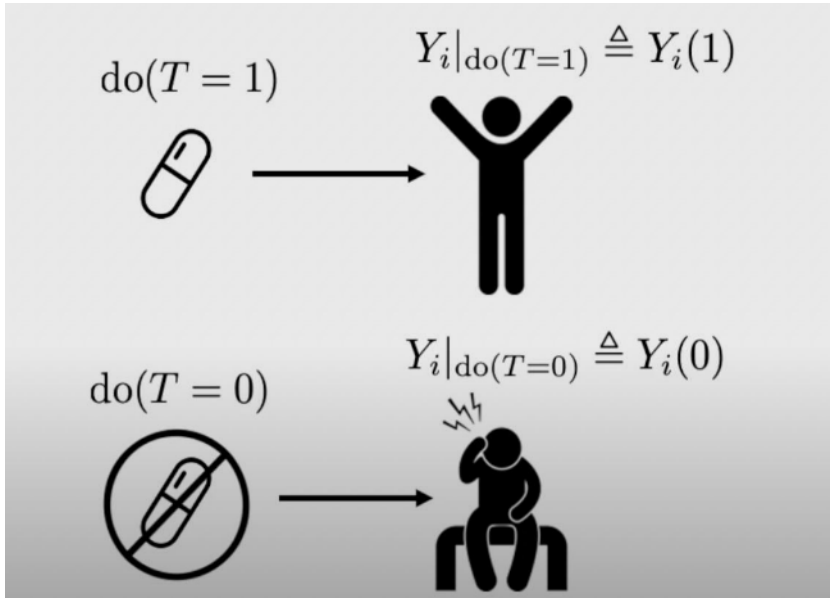
Observational Study

(more practical)

Background

Average Treatment Effect (ATE)

- The mean of individual treatment effects for all samples



ATE: $\mathbb{E}(Y(1) - Y(0))$

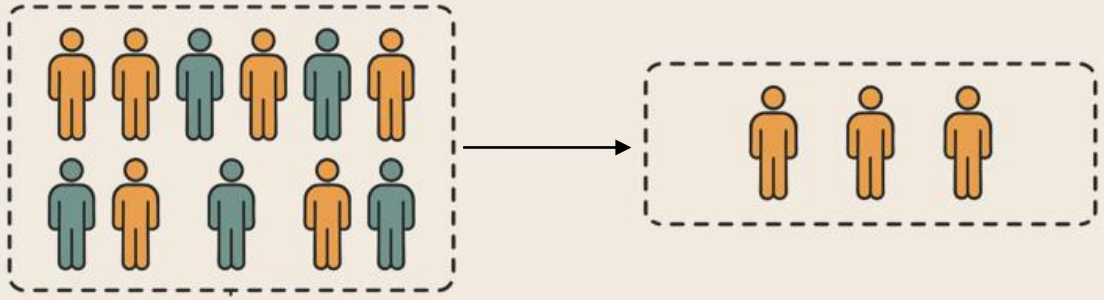
$\boldsymbol{T}$: observed treatment (whether to take the pill)
$\boldsymbol{X}$: observed covariates (such as age, gender…)
$\boldsymbol{Y}$: observed outcome (health status)
$\boldsymbol{i}$: a specific sample/individual
$\boldsymbol{Y_i(1)}$: potential outcome under treatment
$\boldsymbol{Y_i(0)}$: potential outcome under no treatment

Background

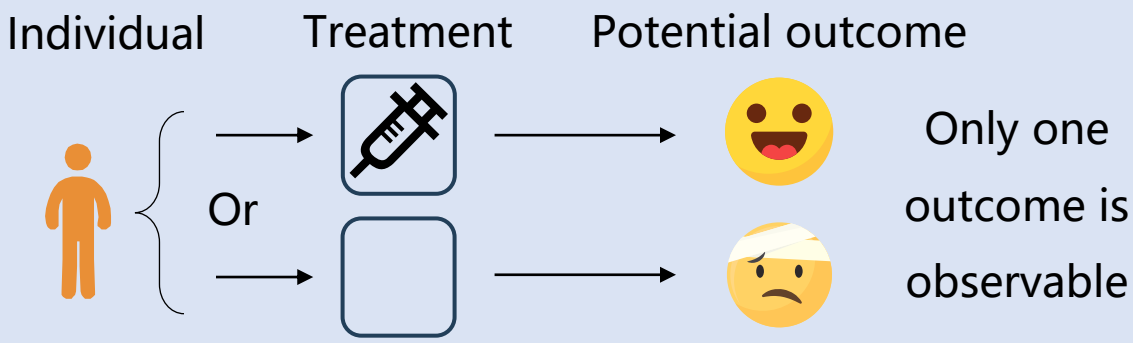
ATE Estimation

Problem

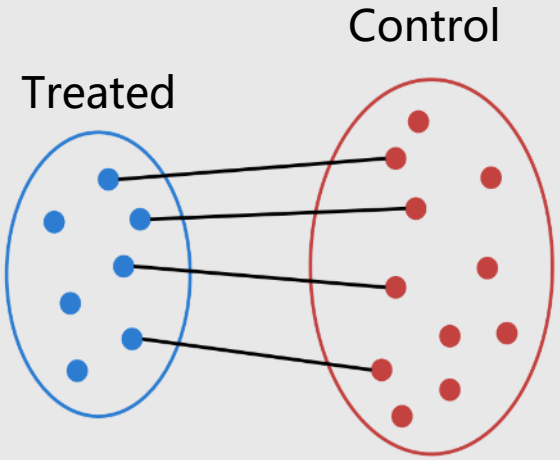
1. Selection Bias



2. Missing Information



Solution



Sample Matching

Background

- Privacy Leakage
  - The data used in causal inference often contain sensitive information

PII

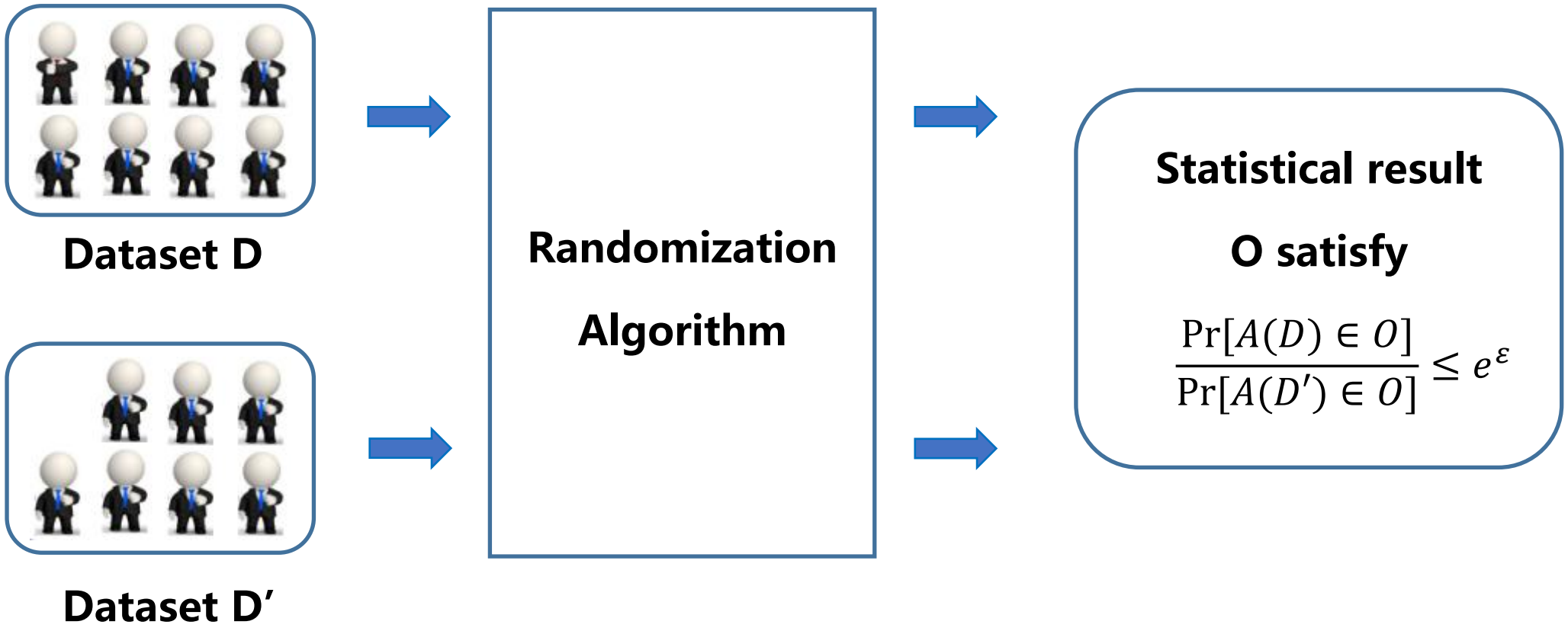
Financial

Health

Business

Problem Definition

- Differential Privacy (DP)



ATE Estimation

- Existing Solution

Limitation

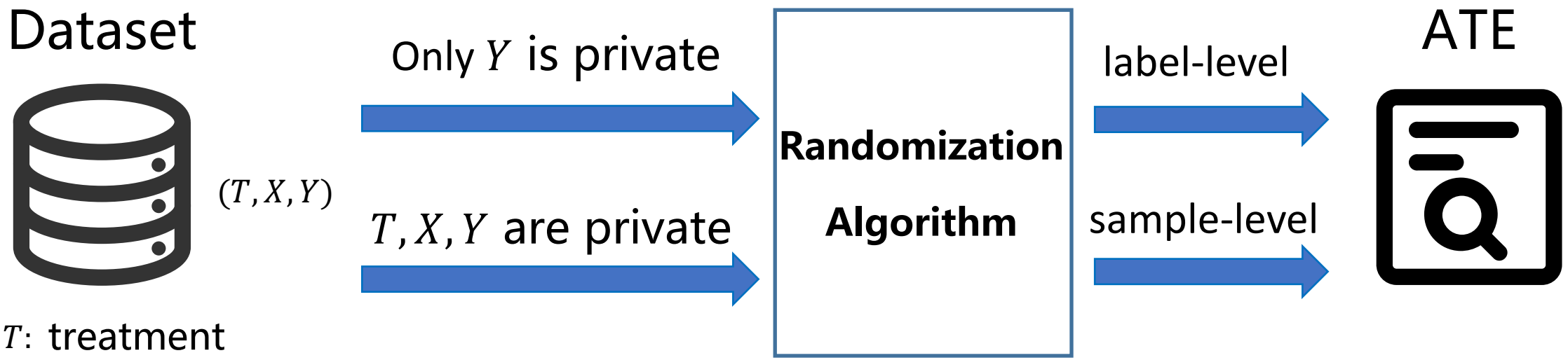
- **Specific assumption**
- **Partial protection**
- **Fixed configuration**

Our aim

- **Universal assumption**
- **Comprehensive protection**
- **Flexible configuration**

Problem Definition

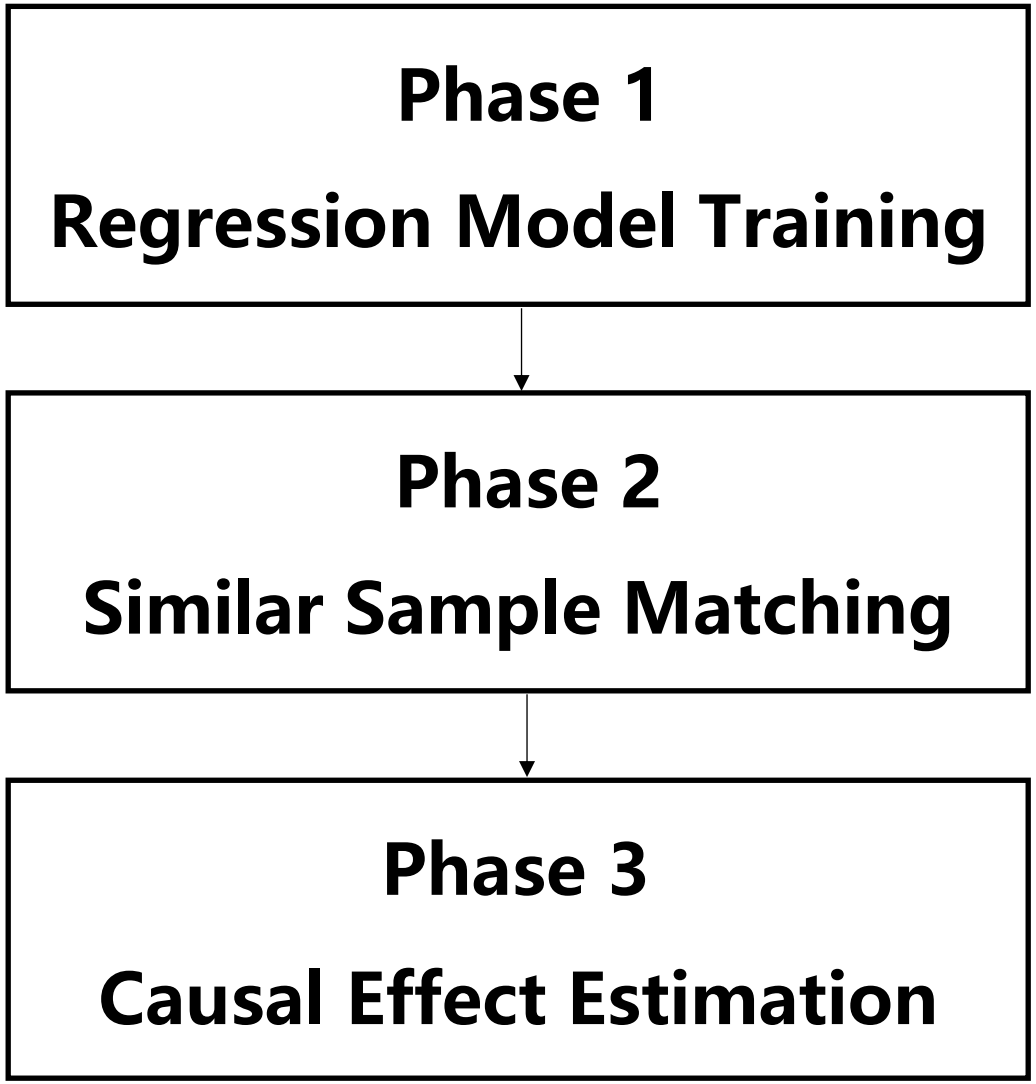
- Label-level and Sample-level



T: treatment
X: covariates
Y: outcome

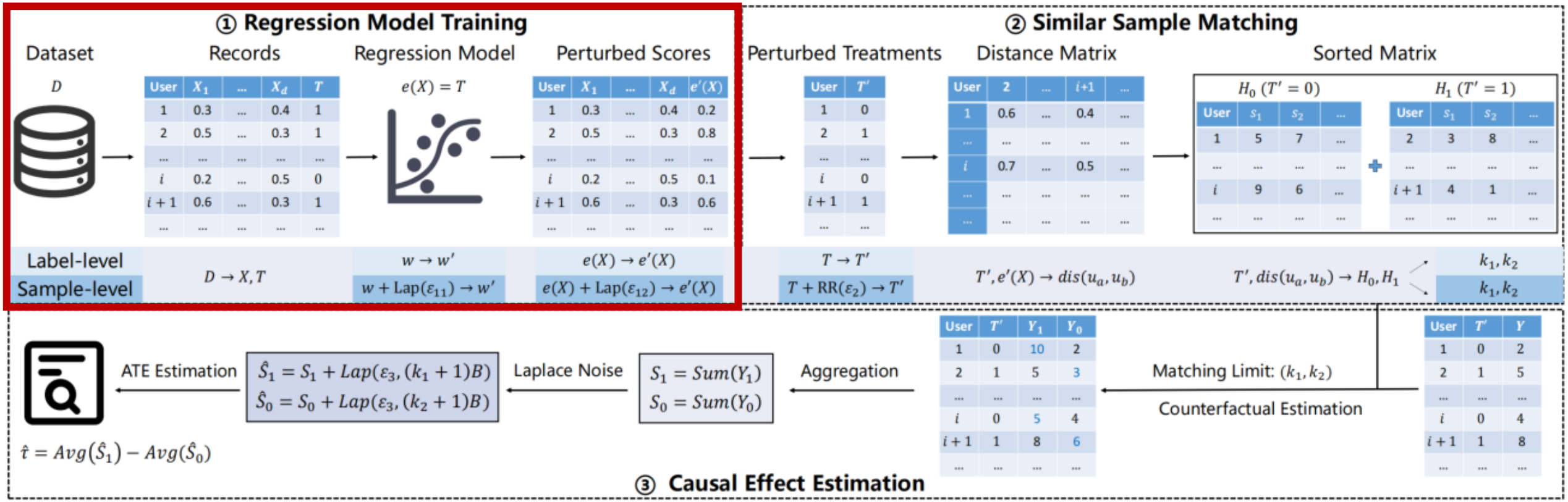
Our goal: Publish the ATE value under label-level/sample-level privacy while ensuring high utility

Workflow of PrivATE



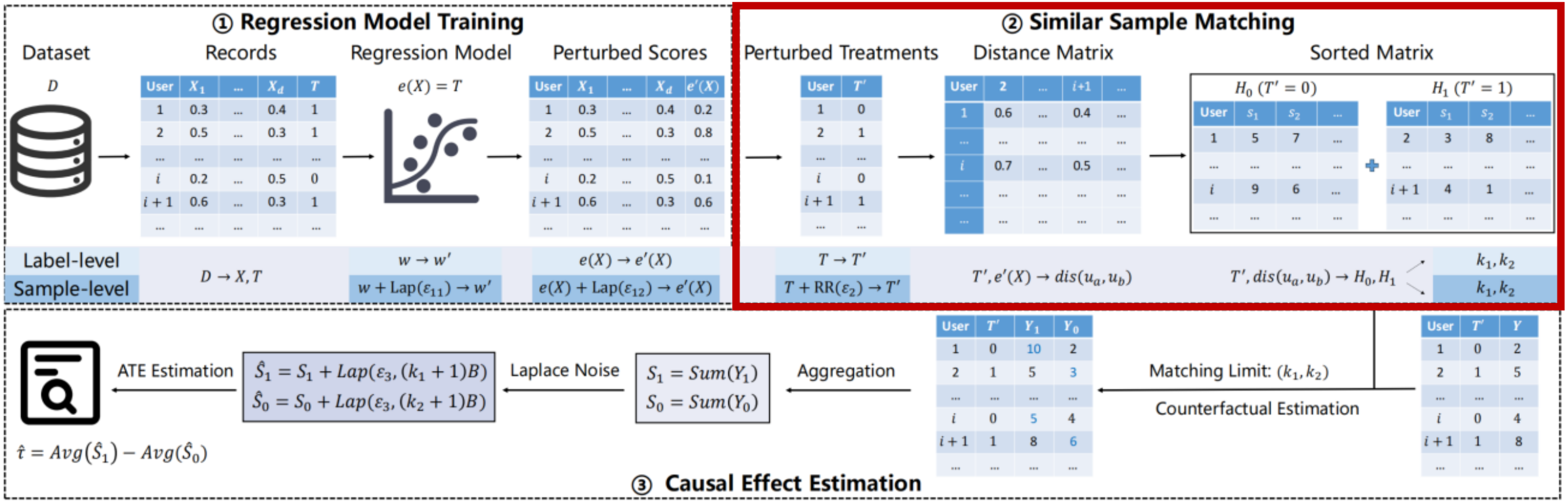
Workflow of PrivATE

- Phase 1: Regression Model Training
- Phase 2: Similar Sample Matching
- Phase 3: Causal Effect Estimation



Workflow of PrivATE

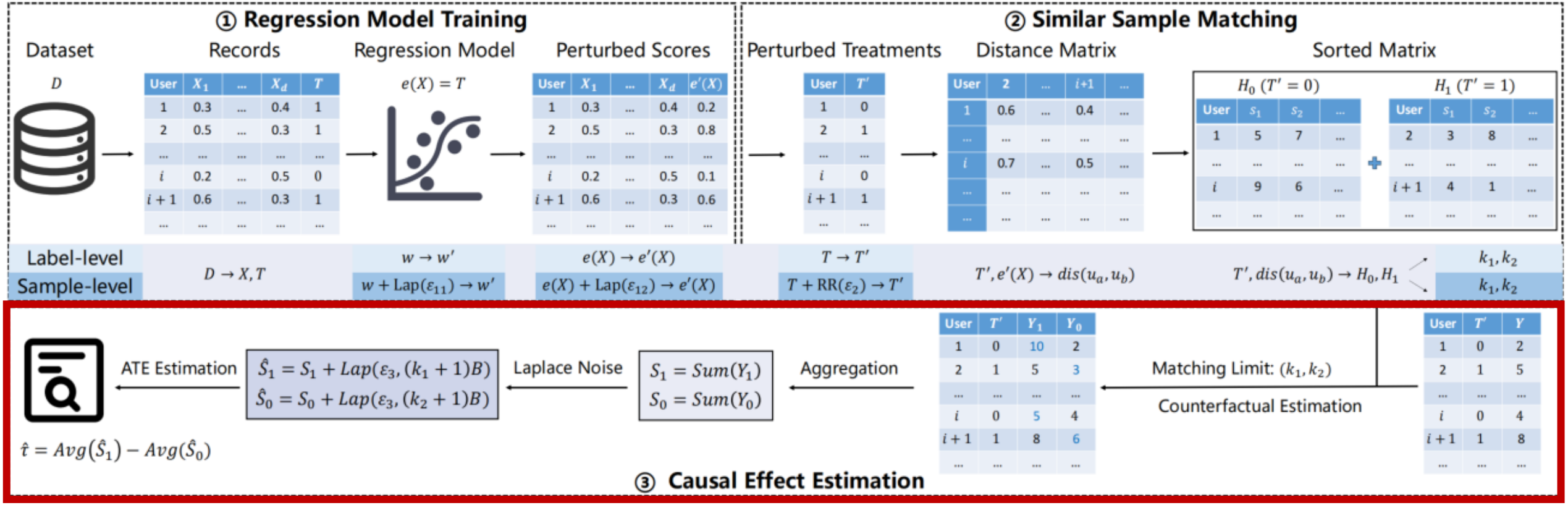
- Phase 1: Regression Model Training
- Phase 2: Similar Sample Matching
- Phase 3: Causal Effect Estimation



Workflow of PrivATE

- Phase 1: Regression Model Training
- Phase 2: Similar Sample Matching
- Phase 3: Causal Effect Estimation

$$\mathbb{E}[(\hat{S} - S)^2] = \mathrm{Var}[\hat{S}] + \mathrm{Bias}[\hat{S}]^2$$



Error Analysis

Potential outcome

$$\mathbb{E}[(\hat{S} - S)^2] = \mathrm{Var}[\hat{S}] + \mathrm{Bias}[\hat{S}]^2$$

$$\mathrm{Var}[\hat{S}] \approx \frac{2k^2B^2}{\varepsilon^2} \qquad \mathrm{Bias}[\hat{S}] \approx c \cdot B \cdot n_1 \cdot \frac{M_1}{k}$$

Minimize

Label-level

$$k^* = \sqrt{\frac{\varepsilon \cdot c \cdot n_1 \cdot M_1}{2}}$$
$$k_f = \min(\max(round(k^*), 1), M_1)$$

Sample-level

$$k^* = \sqrt{\frac{\varepsilon_3 \cdot h \cdot n_1' \cdot M_1'}{2}}$$
$$k_f = \max(round(k^*), 1)$$

Observed noisy value

$\boldsymbol{k}$: Matching limit

$\boldsymbol{B}$: The maximum variation range of outcome

$\boldsymbol{c}$: Error coefficient

$\boldsymbol{n_1}$: $\max(n_t, n_c)$, the number of samples in the group

$\boldsymbol{M_1}$: True average maximum number of matches

Experiment Setup

Dataset

Datasets	Treated	Control	Total	Type
IHDP [31]	139	608	747	Semi-real
Lalonde [32]	185	260	445	Real
ACIC [33]	858	3944	4802	Semi-real
Synth [18]	489	511	1000	Synthetic

Metric

- Relative error

Experiment Setup

Competitors

- DP ATE estimation
  - IPW-PP[1]
  - SmoothDPM[2]
  - DPCI[3]
- DP data synthesis
  - AIM[4]
  - PrivSyn[5]

[1] 2019 ArXiv Privacy-preserving Causal Inference via Inverse Probability Weighting

[2] 2024 SatML Differentially Private Multi-Site Treatment Effect Estimation

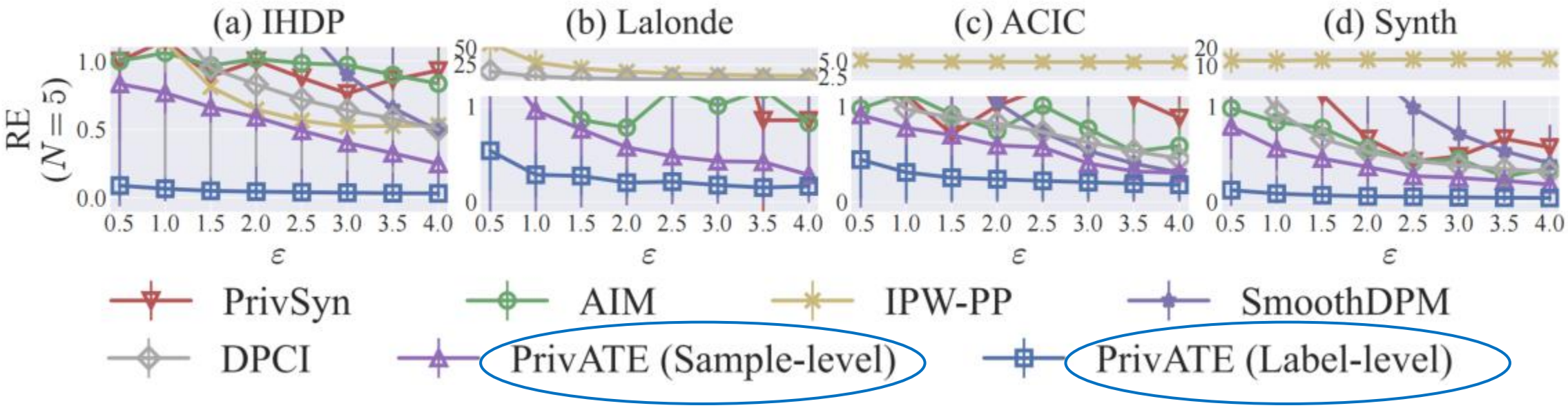
[3] 2025 ArXiv PrivATE: Differentially Private Confidence Intervals for Average Treatment Effects

[4] 2022 VLDB AIM: An Adaptive and Iterative Mechanism for Differentially Private Synthetic Data

[5] 2021 USENIX Security PrivSyn: Differentially Private Data Synthesis

Performance

End-to-end comparison

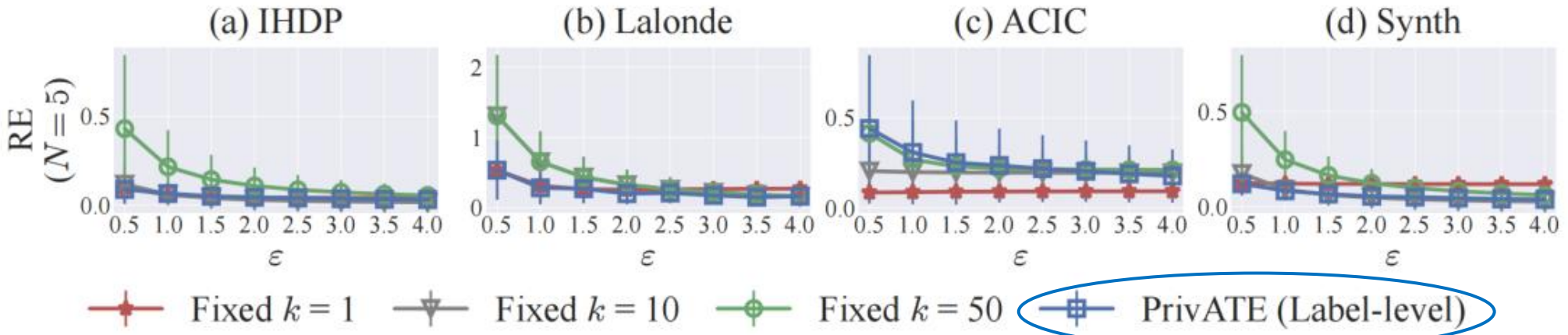


PrivATE outperforms other methods across various datasets and privacy budgets.

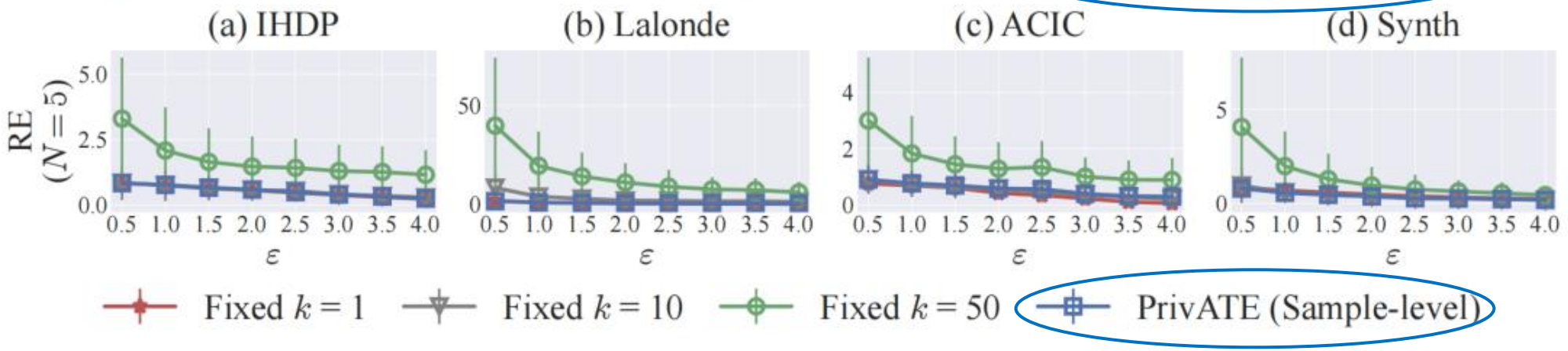
Performance

◻ Choice of matching limit

Label-level



Sample-level

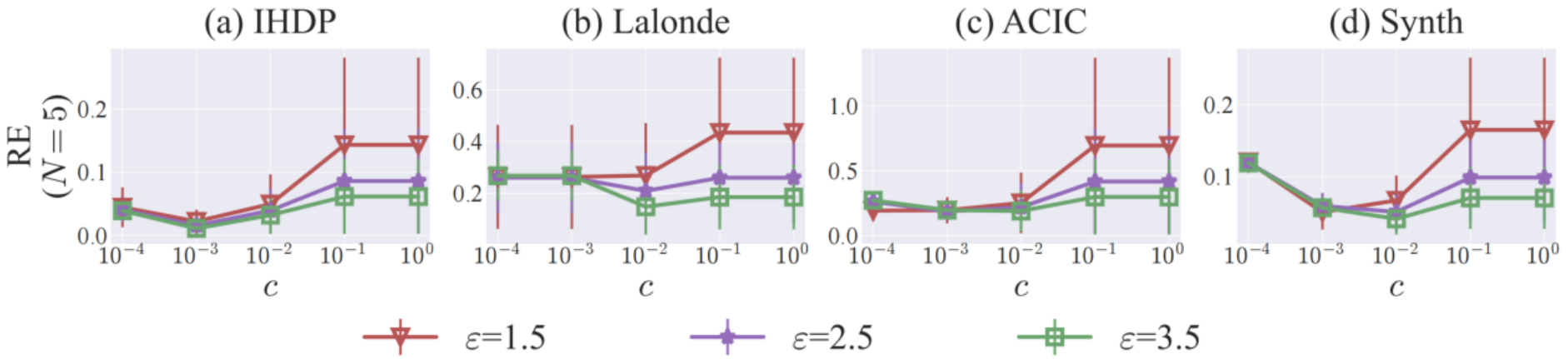


Fixed k cannot effectively handle various scenarios.

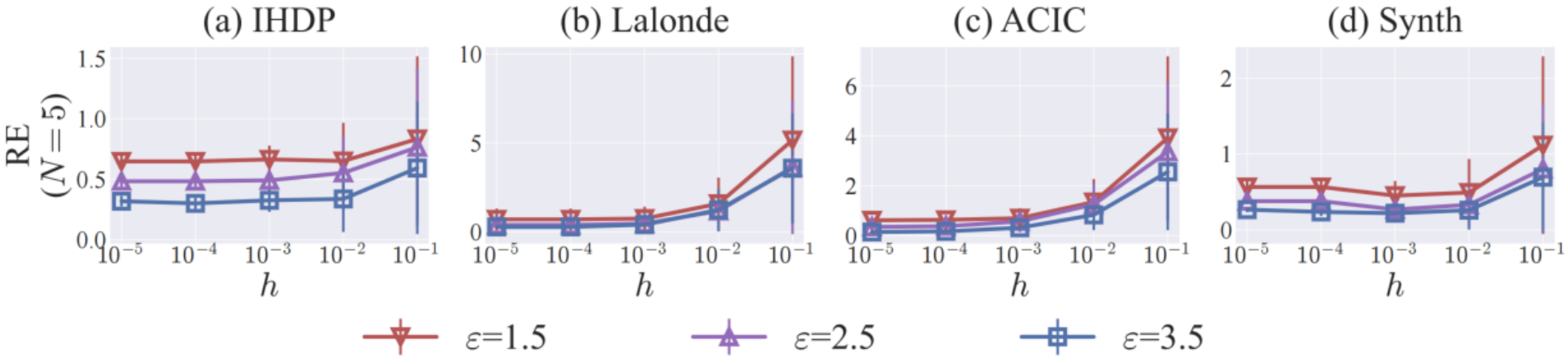
Performance

◻ Impact of error coefficient

Label-level



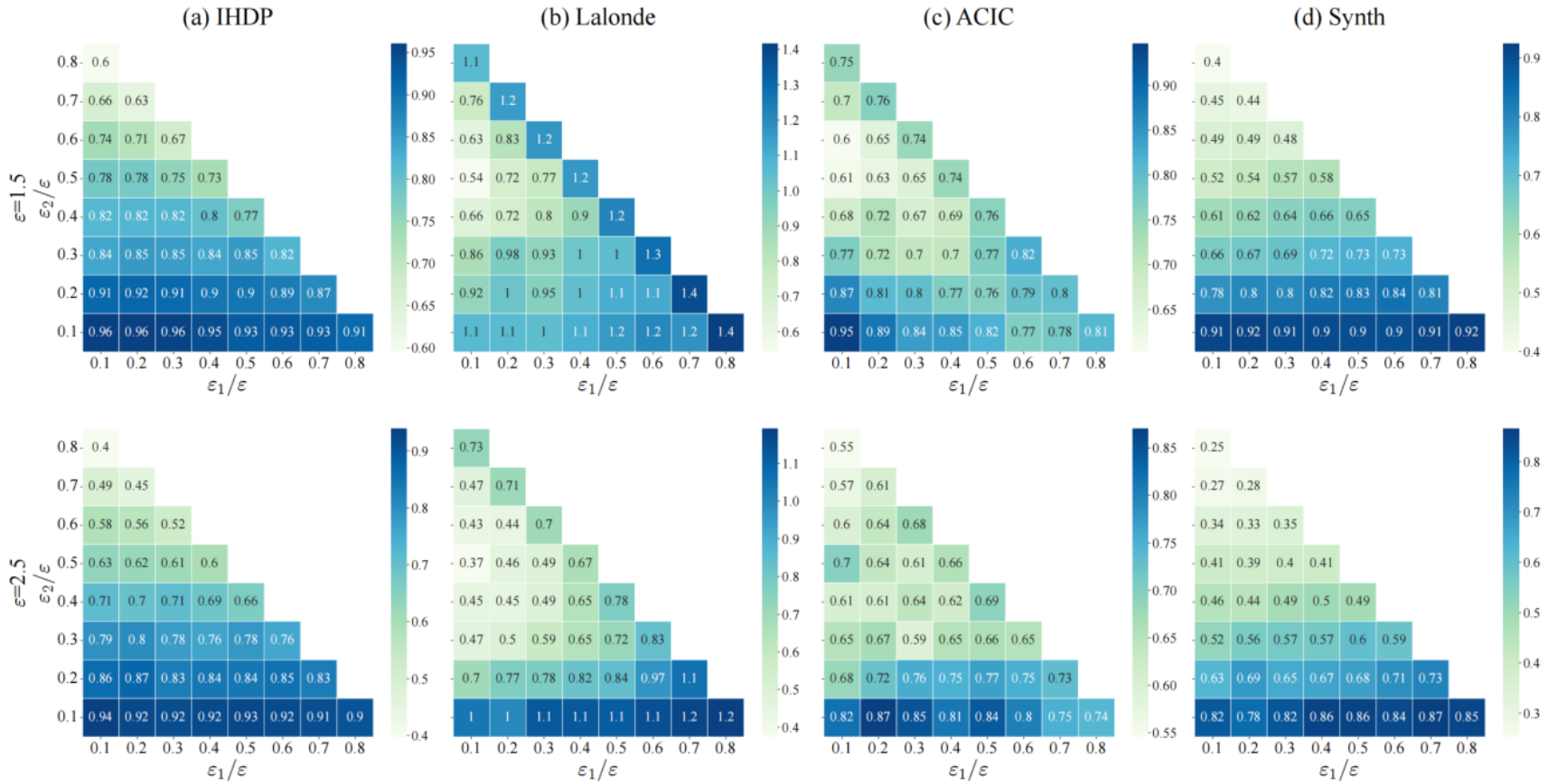
Sample-level



Too large or too small coefficients are difficult to achieve low RE.

Performance

◻ Impact of budget allocation



No optimal allocation for all datasets and privacy budgets.

Conclusion

- A more practical and effective privacy-preserving ATE estimation framework under DP
- Two levels of privacy protection (i.e., label-level and sample-level) to satisfy different tradeoffs between utility and privacy
- An adaptive matching limit determination mechanism to strike a balance between reducing global sensitivity and improving matching accuracy
- An extensive evaluation on multiple datasets to illustrate the superiority of PrivATE

PrivATE: Differentially Private Average Treatment Effect Estimation for Observational Data

Full paper

Thank you for your attention!