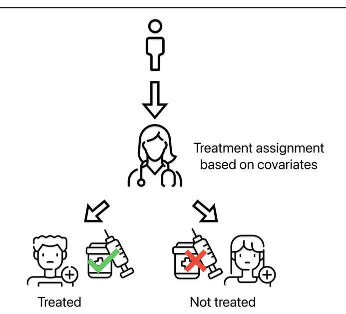
Federated Causal Inference on Multi-Site Observational Data via Propensity Score Aggregation

Rémi Khellaf [1] Aurélien Bellet [1] Julie Josse [1]

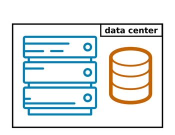
[1]INRIA, PreMeDICaL team

Motivation



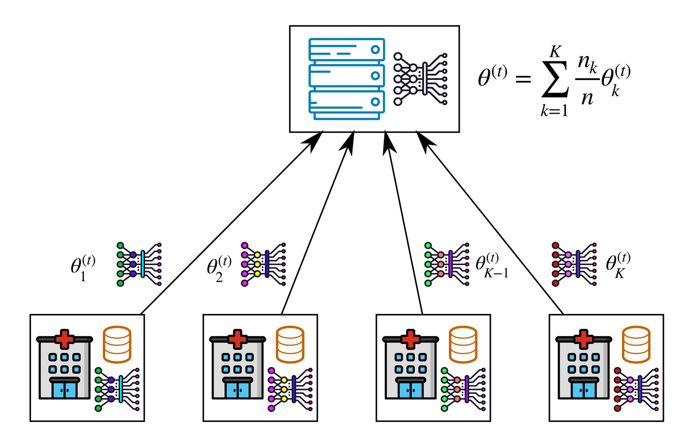
Estimating the **Average Treatment Effect** (ATE) from **observational data requires large sample sizes** due to positivity constraints, high dimensionality, small populations (rare diseases), costly data collection, and slow nonparametric convergence.

→ A **multicentric framework** is essential.



Data centralization is often infeasible: Privacy laws (HIPAA, GDPR, PIPL), no "consent and release", competitive concerns, logistics.

→ **Federated Learning (FL)** allows collaboration across sites without centralizing their data. We build **federated estimators** (via FedAvg) and compare them with pooled and local baselines.



Use-cases: Vaccine efficacy across hospitals, minimum wage policies across governments, credit scoring across banks.

Multicentric Causal Inference Modelling

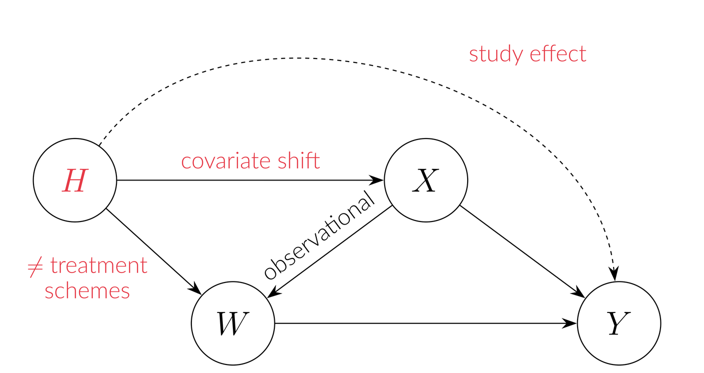
Decentralization challenge: handling heterogeneity across sites.

Goal: Estimate the ATE over the population of the K decentralized participating sites, defined as a mixture:

$$\tau := \mathbb{E}\big[\mathbb{E}[Y(1) - Y(0) \mid H]\big] = \sum_{k=1}^{K} \rho_k \tau_k$$

Data & Notations:

- $X \in \mathbb{R}^{n \times d}$ (covariates), $W \in \{0,1\}^n$ (treatment), $Y \in \mathbb{R}^n$
- $H \in \{1, \dots, K\}$ (site index)
- $Z_k = \{X_{k,i}, W_{k,i}, Y_{k,i}, H_{k,i} = k\}$ (local datasets)
- $\mu_w(x) = \mathbb{E}[Y \mid W = w, X = x]$
- $e_k(x) = \mathbb{P}(W = 1 \mid X = x, H = k)$
- $e(x) = \mathbb{P}(W = 1 \mid X = x)$ (global PS)
- τ_k, τ (local and global ATE)



Core Identification Assumptions:

A **Unconfoundedness:** $\{Y(1), Y(0)\} \perp\!\!\!\perp W \mid (X, H)$

B **Ignorability on sites:** $\{Y(1), Y(0)\} \perp\!\!\!\perp H \mid X$

C **Consistency:** $Y = WY(1) + (1 - W)Y(0)$

$A + B \implies \{Y(1), Y(0)\} \perp\!\!\!\perp H \mid (X, W)$ *(No site-effects)*

Overlap & Support Conditions:

D.1 **Global Positivity:** $\forall x \in \mathcal{X}, 0 < e(x) < 1$

D.2 **Local Positivity:** $\forall x \in \mathcal{X}, \forall k \in [K], 0 < e_k(x) < 1$ *(Much stronger!)*

Pooled Estimators (Centralized Oracle):

Learns nuisance functions on $\cup_{k=1}^{K} Z_k$.

IPW: $\tau(e, Z) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{W_i Y_i}{e(X_i)} - \frac{(1 - W_i) Y_i}{1 - e(X_i)} \right)$

AIPW: $\tau(\mu_1, \mu_0, e, Z) = \frac{1}{n} \sum_{i=1}^{n} \Big(\mu_1(X_i) - \mu_0(X_i) + \frac{W_i(Y_i - \mu_1(X_i))}{e(X_i)} - \frac{(1 - W_i)(Y_i - \mu_0(X_i))}{1 - e(X_i)} \Big)$

→ **Infeasible** due to data sharing constraints.

Meta-Analysis on Aggregated Data:

$$\hat{\tau}_{\text{meta}} = \sum_{k=1}^{K} \hat{\rho}_k \hat{\tau}_k \quad \text{with } \hat{\tau}_k = \begin{cases} \tau(\hat{e}_k, Z_k) & \text{(IPW)} \\ \tau(\hat{\mu}_{1k}, \hat{\mu}_{0k}, \hat{e}_k, Z_k) & \text{(AIPW)} \end{cases}$$

- **SW:** $\hat{\rho}_k = \frac{n_k}{n}$ • **IVW:** $\hat{\rho}_k = \mathbb{V}(\hat{\tau}_k)^{-1}$

→ **Problem 1:** Fails with small n_k or if local overlap (**D.2**) is violated at any single site.

→ **Problem 2:** Coarser grain of aggregation handles heterogeneity less precisely.

→ **Solution:** Leverage **Federated Learning** for individual-level analysis without centralizing.

Federated Flexible (A)IPW from Observational Studies[1]

Observational studies often provide large, representative samples ("real-world data"). However, in multicentric settings, sites rarely share a common treatment assignment $e_k \neq e_{k'}$. This hinders learning a single global PS. We note that

$$e(x) = \mathbb{P}(W_i = 1 \mid X_i = x) = \sum_{k=1}^{K} \mathbb{P}(H_i = k \cap W_i = 1 \mid X_i) = \sum_{k=1}^{K} \omega_k(X_i) e_k(X_i)$$

Primary Framework: Membership Weights (MW)

$$\omega_k(X_i) = \mathbb{P}(H_i = k \mid X_i)$$

1. **Learning:** Fit $\hat{\omega}_k(X_i)$ via FedAvg (Logistic Regressions, NNs) or Federated Random Forests[2].
2. **PS Aggregation:** Sites share locally estimated PS $\hat{e}_k$. Server computes:
$$\hat{e}_{\text{MW}}(x) = \sum_{k=1}^{K} \hat{\omega}_k(x) \hat{e}_k(x)$$
3. **Causal Estimate:** Sites compute (A)IPW locally: $\hat{\tau}_k^{\text{MW}} = \tau(\hat{\mu}_1, \hat{\mu}_0, \hat{e}_{\text{MW}}, Z_k)$.
4. **Global ATE:** Server averages results: $\hat{\tau}_{\text{MW}} = \sum_{k=1}^{K} \frac{n_k}{n} \hat{\tau}_k^{\text{MW}}$.

Ideal when: X strongly explains H (e.g., geography) and K is small.
Advantages: Highest flexibility for nuisance functions.
Costs: Multi-round communication (floats exchanged = # rounds × K× parameters).

Alternative Approach: Density Ratios (DW)

$$\omega_k(x) = \rho_k f_k(x) / f(x)$$

Where $f(x)$ is the global covariate mixture $\sum_{l=1}^{K} \rho_l f_l(x)$ and $\rho_k \approx \frac{n_k}{n}$.

Parametrization ($f_k(x; \theta_k)$):

1. **Moment Sharing:** Sites share basic parameters $\hat{\theta}_k$ (means, covariances) alongside local $\hat{e}_k$.
2. **PS Aggregation:** Estimate the global PS as:
$$\hat{e}_{\text{DW}}(x) = \frac{\sum_{k=1}^{K} \frac{n_k}{n} f(x; \hat{\theta}_k) \hat{e}_k(x)}{\sum_{l=1}^{K} \frac{n_l}{n} f(x; \hat{\theta}_l)}$$

(Steps 3 and 4 follow identically to the MW framework).

Ideal when: A priori knowledge on $X|H$ exists; low dimensionality.
Advantages: One-shot communication ($O(Kd^2)$ floats exchanged).
Costs: Highly parametric structure.

Theoretical Properties & Overlap Advantage

Key Property: Our Federated framework **only** requires the weaker assumption of global overlap (D.1), allowing inclusion of external control arms.

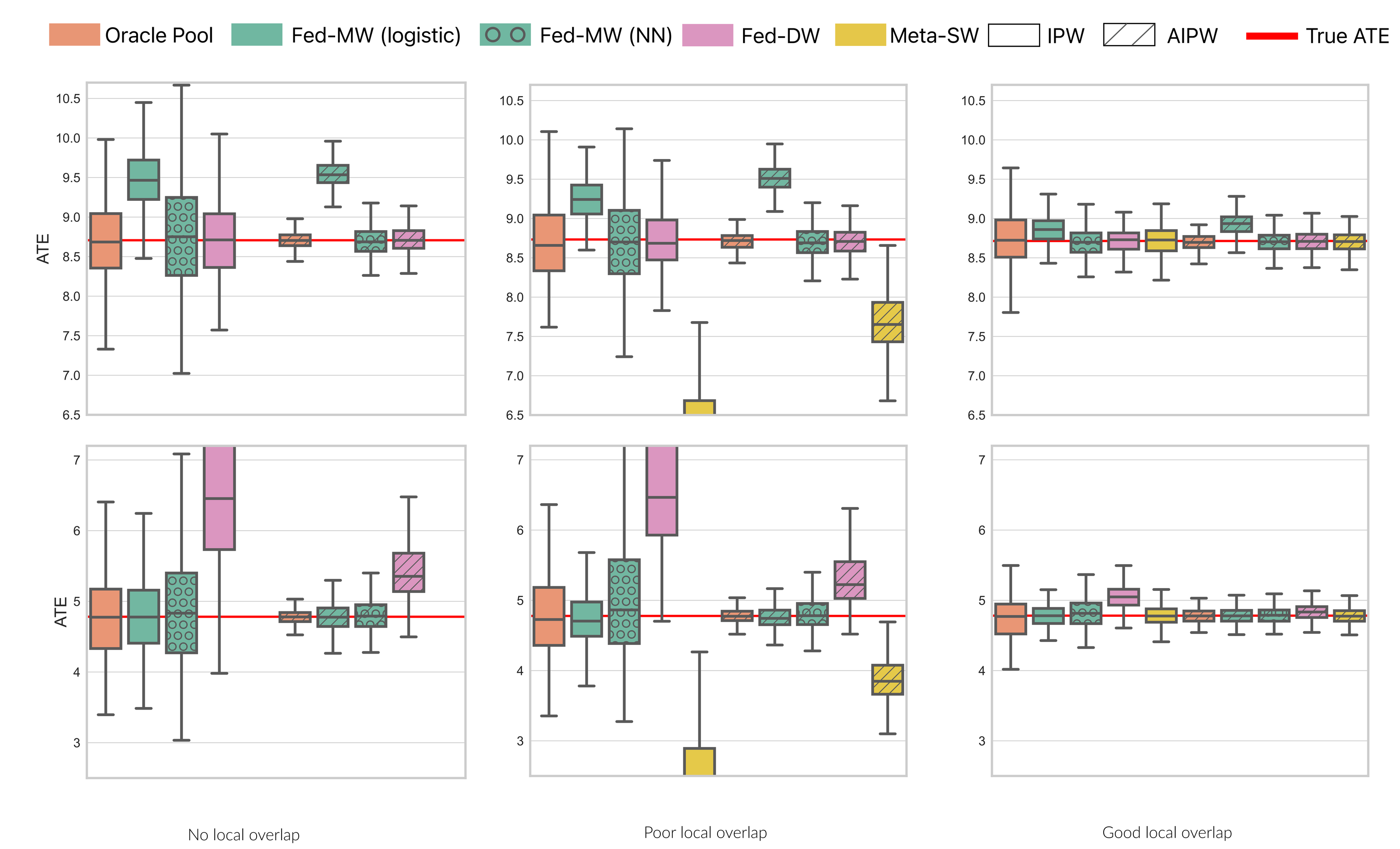
Heterogeneous PS ($e_k \neq e_{k'}$):

- **Failure of Meta-Methods:** Standard Meta-IPW and Meta-AIPW bias heavily if D.2 fails anywhere. MW and DW remain unbiased.
- **Robustness:** Safely handles shifts in treatment assignments, overlaps, and covariate distributions.
- **Double Robustness:** Federated AIPWs achieve **semi-parametric efficiency** $\mathbb{V}^{\infty}(\hat{\tau}_{\text{AIPW}}^{\text{fed}}) = \mathbb{V}^{\infty}(\hat{\tau}_{\text{AIPW}}^{\star})$ if error rates are $\max\{\|\hat{\mu}_w - \mu_w\|, \|\hat{e}_k - e_k\|, \|\hat{\omega}_k - \omega_k\|\} = o_P(n^{-1/4})$.

Homogeneous PS ($\forall k, e_k(x) = e(x)$):

- Local and global overlap equate (D.1 $\iff$ D.2).
- Meta and Federated estimators with oracle PS are equal.
- **Efficiency Gain:** Estimating $e(x)$ globally via FL yields better estimates than local learning.

Experiments



Top row: $X|H = k \sim \mathcal{N}(\mu_k, \Sigma_k)$. **Bottom row:** $H|X \sim$ Multinomial Logistic. μ_1, μ_0 are misspecified → double robustness

Simulation Insights & Empirical Behavior:

- **No Local Overlap:** Meta-analysis baseline fails completely due to division-by-zero errors. Federated estimators remain stable.
- **Poor Local Overlap:** Meta-analysis remains highly volatile and biased. Federated framework is consistent.
- **Good Local Overlap:** The efficiency gains over classical post-hoc meta-analysis become marginal.
- **MW vs DW Trade-off:** MW integrates flexible ML (NNs, Random Forests) without assumptions on the data generating process, at the cost of a higher variance.

References:

[1] This Paper: "Federated Causal Inference on Multi-Site Observational Data via Propensity Score Aggregation." *ICML 2026.*

[2] Federated Trees: "Principled Federated Random Forests for Heterogeneous Data." *2026 Preprint.*

[3] Federated RCTs: "Federated causal inference: Multi-study ATE estimation beyond meta-analysis." *AISTATS 2025.*

[4] Kairouz, Peter, and H. Brendan McMahan. "Advances and open problems in federated learning.", 2021

[5] Rosenbaum, Paul R., and Donald B. Rubin. "The central role of the propensity score in observational studies for causal effects.", 1983

[6] R. Xiong et al. "Federated causal inference in heterogeneous observational data.", 2023

remi.khellaf@inria.fr