Federated Simulated Clinical Trials Technical Details

Wanjun Gu

Federated Simulated Clinical Trials for GLP-1

1. Notation and Setup

We consider J institutions (sites), indexed by $j = 1, \dots, J$.

At institution j, we observe data for n_j patients:

$$\mathcal{D}_j = \{(X_{ij}, T_{ij}, Y_{ij})\}_{i=1}^{n_j}$$

where:

- $X_{ij} \in \mathbb{R}^p$: baseline covariates (comorbidities, labs, demographics, medications, etc.).
- $T_{ij} \in \{0, 1\}$: treatment indicator (e.g., GLP-1 use vs no GLP-1).
- Y_{ij}: outcome of interest (can be continuous, binary, time-to-event, etc.).

We work under the **potential outcomes** framework:

- $Y_{ij}(1)$: potential outcome if patient i at site j is treated with GLP-1.
- $Y_{ij}(0)$: potential outcome if not treated.

The primary estimand could be, for example, the population ATE (average treatment effect):

$$\text{ATE} = \mathbb{E}[Y(1) - Y(0)]$$

possibly within strata (e.g., ancestry, co-medication).

2. Federated Learning Layer

We assume each site fits a parameterized model locally, **without sharing raw data**.

Let $f(\cdot;\theta)$ denote a generic predictive or causal model (e.g., logistic regression, Cox model, boosted trees, or a transformer-based foundation model). The model class can differ across components (outcome model, propensity model), but we treat θ abstractly as a parameter vector with uncertainty.

2.1 Local Training

At site j, we obtain an estimate $\hat{\theta}_j$ by minimizing a site-specific empirical loss:

$$\hat{\theta}_j = \arg\min_{\theta} L_j(\theta) = \arg\min_{\theta} \frac{1}{n_j} \sum_{i=1}^{n_j} \ell(f(X_{ij};\theta), Z_{ij})$$

where $\ell(\cdot,\cdot)$ is an appropriate loss (e.g., negative log-likelihood, squared error) and Z_{ij} is the relevant target (e.g., Y_{ij} for outcome models, T_{ij} for propensity models).

Each site also computes a measure of parameter uncertainty, e.g. an approximate covariance:

$$\widehat{\Sigma}_j \approx \mathrm{Var}(\hat{\theta}_j)$$

which could come from:

- The inverse observed Fisher information (for GLMs);
- The Hessian of the loss (for general ML models);
- Posterior covariance (for Bayesian models).

Only $\hat{\theta}_j$ and uncertainty summaries (e.g. $\widehat{\Sigma}_j$) are shared, not the individual-level data.

2.2 Federated Aggregation

We define site-level quality or reliability weights ω_j, reflecting:

- Data volume (n_j);
- Data completeness/structure (e.g., EHR coverage);
- Signal quality (e.g., missingness, measurement noise).

A natural choice is:

$$\omega_j \propto q_j \, n_j, \quad \text{with} \quad \sum_{j=1}^{J} \omega_j = 1$$

where $q_j \in (0, 1]$ encodes a data-quality score (e.g., q_j closer to 1 for UCSF/UC-wide EHR, lower for sparse community sites).

A simple federated estimator is:

$$\hat{\theta}_{\text{fed}} = \sum_{j=1}^{J} \omega_j \, \hat{\theta}_j$$

For models with known variance estimates, we can use precision-weighted meta-analysis:

$$\hat{\theta}_{\text{fed}} = \left(\sum_{j=1}^{J} \widehat{\Sigma}_j^{-1} \right)^{-1} \left(\sum_{j=1}^{J} \widehat{\Sigma}_j^{-1} \hat{\theta}_j \right)$$

This formulation is agnostic to the specific model class: linear, tree-based, or deep/foundation models — all that matters is we can represent them with parameters θ and uncertainty Σ.

3. Simulated Clinical Trial Layer

3.1 Propensity Scores and Covariate Balance

At each site j, we estimate the propensity score:

$$e_j(X_{ij}) = \Pr(T_{ij} = 1 \mid X_{ij})$$

using a local model (logistic regression, gradient boosting, neural net, etc.) with parameters ϕ_j:

$$\hat{\phi}_j = \arg\min_{\phi} \frac{1}{n_j} \sum_{i=1}^{n_j} \ell_{\text{CE}}(g(X_{ij}; \phi), T_{ij})$$

where $g(\cdot; \phi)$ outputs a probability.

We can then either:

- **Federate the propensity model** (aggregate $\hat{\phi}_j$ into $\hat{\phi}_{\text{fed}}$ as above), or
- **Keep site-specific propensity scores** and combine at the estimand level.

For inverse probability of treatment weighting (IPTW), we define subject-level weights:

$$w_{ij}^{\mathrm{IPTW}} = \begin{cases} \frac{1}{\hat{e}_j(X_{ij})}, & T_{ij} = 1 \\ \frac{1}{1-\hat{e}_j(X_{ij})}, & T_{ij} = 0 \end{cases}$$

Alternative, more stable weights (e.g., overlap weights) are:

$$w_{ij}^{\mathrm{overlap}} = \begin{cases} 1 - \hat{e}_j(X_{ij}), & T_{ij} = 1 \\ \hat{e}_j(X_{ij}), & T_{ij} = 0 \end{cases}$$

These weights induce a “pseudo-population” that mimics a randomized trial by balancing covariates between treated and untreated groups.

3.2 Propensity Score Matching

For matching, at each site j, we construct pairs (or sets) of treated and untreated patients with similar propensity scores:

$$\left|\hat{e}_j(X_{ij}) - \hat{e}_j(X_{i'j})\right| \leq \delta$$

for some caliper δ, using nearest-neighbor or optimal matching algorithms.

Matched sets define a simulated randomized cohort within each site. Effect estimates (e.g., difference in means, hazard ratios) are computed locally, then combined federatively.

3.3 Estimating Treatment Effects

Within each site, a weighted or matched estimator of the ATE could be:

$$\hat{\tau}_j = \frac{\sum_i w_{ij} T_{ij} Y_{ij}}{\sum_i w_{ij} T_{ij}} - \frac{\sum_i w_{ij}(1 - T_{ij}) Y_{ij}}{\sum_i w_{ij}(1 - T_{ij})}$$

The site-level $\hat{\tau}_j$ and its variance $\widehat{\mathrm{Var}}(\hat{\tau}_j)$ are then aggregated:

$$\hat{\tau}_{\mathrm{fed}} = \frac{\sum_{j=1}^{J} \omega_j \hat{\tau}_j}{\sum_{j=1}^{J} \omega_j} \quad \text{or} \quad \hat{\tau}_{\mathrm{fed}} = \frac{\sum_{j=1}^{J} \hat{\tau}_j / \widehat{\mathrm{Var}}(\hat{\tau}_j)}{\sum_{j=1}^{J} 1/\widehat{\mathrm{Var}}(\hat{\tau}_j)}$$

depending on whether we use quality weights or inverse-variance weights.

4. Heterogeneous Institutions and Data Quality

We explicitly model heterogeneity across institutions via:

1. **Data-quality scores** q_j (as above),
2. **Random effects / hierarchical structure**, and
3. **Different missingness patterns and feature sets**.

4.1 Hierarchical / Random Effects Model

We can write a hierarchical outcome model, for example a logistic model for a binary endpoint:

$$\text{logit}\ \Pr(Y_{ij} = 1 \mid T_{ij}, X_{ij}, j) = \alpha_j + \beta T_{ij} + X_{ij}^{\top}\gamma$$

with:

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$$

and possibly:

$$\beta_j \sim \mathcal{N}(\beta, \sigma_\beta^2)$$

if we allow site-specific treatment effects.

This can be estimated in a federated manner by each site providing sufficient statistics or gradients/Hessians, rather than raw data.

4.2 Handling Sparse Sites

For sites with sparse data:

- We regularize site-specific parameters toward the global mean (shrinkage).
- Their contribution to the federated estimator is down-weighted via q_j and/or larger estimated variances.

This allows us to still leverage GLP-1 signal from smaller or less organized institutions without letting them dominate the estimates.

5. Stratification by Ancestry, Co-Medications, and Interactions

Let the covariate vector be structured as:

$$X_{ij} = (A_{ij}, D_{ij}, C_{ij})$$

where:

- A_{ij}: ancestry-related variables (e.g., self-reported European/African/East Asian, or genetic PCs).
- D_{ij}: co-medication indicators (e.g., concomitant drugs with potential interactions).
- C_{ij}: other clinical covariates (age, sex, BMI, comorbidities, labs, etc.).

5.1 Effect Heterogeneity via Interactions

We explicitly model interactions between GLP-1 treatment and key covariates. For example, in a logistic regression:

$$\text{logit } \Pr(Y_{ij} = 1 \mid T_{ij}, X_{ij}) = \alpha + \beta T_{ij} + X_{ij}^{\top}\gamma + T_{ij}A_{ij}^{\top}\delta_A + T_{ij}D_{ij}^{\top}\delta_D + T_{ij}C_{ij}^{\top}\delta_C$$

Here, δ_A, δ_D, and δ_C capture **differential treatment effects** across ancestry, drug co-exposures, and other clinical characteristics.

We can also define subgroup-specific ATEs:

- By ancestry group a:
$$\text{ATE}(a) = \mathbb{E}[Y(1) - Y(0) \mid A = a]$$
- By drug-interaction profile d:
$$\text{ATE}(d) = \mathbb{E}[Y(1) - Y(0) \mid D = d]$$

These can be estimated via stratified IPTW, matching within strata, or directly via the interaction model above.

5.2 Stratified Propensity and Outcome Models

To improve balance within key strata (e.g., ancestry categories), propensity models may include flexible interactions:

$$\operatorname{logit} e_j(X_{ij}) = \eta_{0j} + A_{ij}^\top \eta_A + D_{ij}^\top \eta_D + C_{ij}^\top \eta_C + \text{(selected interactions)}$$

Outcome models are similarly enriched to capture non-linear and high-dimensional interactions (e.g., using tree-based or neural architectures). The federated approach only requires that each site can provide:

- Local parameter estimates (or gradients),
- Uncertainty estimates,
- Or summary effect estimates for each stratum.

6. Extension to Complex Models (Transformers / Foundation Models)

For high-dimensional EHR representations (textual notes, longitudinal trajectories), we can introduce a representation model:

$$H_{ij} = \Phi(\mathcal{E}_{ij}; \psi)$$

where:

- $\mathcal{E}_{ij}$ is the raw EHR sequence (notes, visits, codes).
- Φ is a transformer or foundation model with parameters ψ.
- H_{ij} is a low-dimensional embedding used in place of or alongside X_{ij}.

Federated training proceeds by:

1. Each site computing local gradients $\nabla_\psi L_j(\psi)$ on its own data.
2. A secure aggregation protocol combining gradients:

$$\nabla_\psi L_{\text{fed}}(\psi) = \sum_{j=1}^{J} \omega_j \, \nabla_\psi L_j(\psi)$$

3. Updating ψ at a central server (without ever seeing raw data).

The embeddings H_{ij} are then used in the causal framework above (propensity modeling, outcome modeling, stratification).