Supplementary: Federated Learning for Clinical Structured Data: A Benchmark Comparison of Engineering and Statistical Approaches

A Details of FL Algorithms

A.1 GLORE

The Grid Binary Logistic Regression (GLORE)[1] calculates the traditional LR model in a distributed and privacy-preserving way via Newton-Raphson iteration[2]. For a federation of K-site with n_k records in each site k ($1 \leq k \leq K$), GLORE calculates the log-likelihood function based on $\sum_{k=1}^{K} n_k$ records is $l(\beta) = \sum_{i=1}^{\sum_{k=1}^{K} n_k} [y_i \log \pi(x_i, \beta) + (1 - y_i) \log(1 - \pi(x_i, \beta))]$, where $x_i = (1, x_{i,1}, ..., x_{i,m})$ for $i = 1, ..., \sum_{k=1}^{K} n_k$.

Algorithm 1: GLORE

Input: $K, \beta^{(0)}, \epsilon$

Client k uses local data to compute $W_k(\bar{X}_k, \beta^{(0)})$ and $\Pi_k(\bar{X}_k, \beta)$;

Each client k sends intermediary results θ_k back to the server, and then server aggregates them;

Server computes $\beta^{(1)}$ and sends back to clients;

while $\|\beta^{(t)} - \beta^{(t-1)}\| \geq \epsilon$ **do**

  Client k uses local data to compute $W_k(\bar{X}_k, \beta^{(t)})$ and $\Pi_k(\bar{X}_k, \beta)$ at t-th iteration;

  Each client k sends intermediary results θ_k back to the server, and then server aggregates them;

  Server computes $\beta^{(t+1)}$ and sends back to clients

end

Specifically, we have

$$
\begin{aligned}
\beta^{(t+1)} &= \beta^{(t)} - \left[\frac{\partial^2 l(\beta^{(t)})}{\partial \beta^{(t)} \partial \beta^{(t)T}}\right]^{-1} \frac{\partial l(\beta^{(t)})}{\partial \beta^{(t)}} \\
&= \beta^{(t)} + \left[\bar{X}^T W(\bar{X}, \beta^{(t)}) \bar{X}\right]^{-1} \bar{X}^T \left[\bar{Y} - \Pi(\bar{X}, \beta^{(t)})\right] \\
&= \beta^{(t)} + \left[\sum_{k=1}^{K} \bar{X}_k^T W_k(\bar{X}_k, \beta^{(t)}) \bar{X}_k\right]^{-1} \left\{\sum_{k=1}^{K} \bar{X}_k^T \left[\bar{Y}_k - \Pi_k(\bar{X}_k, \beta)\right]\right\},
\end{aligned}
$$

$$
W_k(\bar{X}_k, \beta) = \begin{bmatrix} \pi(x_{\sum_{j=1}^{k-1} n_j + 1}, \beta)(1 - \pi(x_{\sum_{j=1}^{k-1} n_j + 1}, \beta)) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \pi(x_{\sum_{j=1}^{k} n_j}, \beta)(1 - \pi(x_{\sum_{j=1}^{k} n_j}, \beta)) \end{bmatrix},
$$

$$
\Pi_k(\bar{X}_k, \beta) = \begin{bmatrix} \pi(x_{\sum_{j=1}^{k-1} n_j + 1}, \beta) \\ \vdots \\ \pi(x_{\sum_{j=1}^{k} n_j}, \beta) \end{bmatrix},
$$

$$
\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_K \end{bmatrix}, \bar{X}_k = \begin{bmatrix} x_{\sum_{j=1}^{k-1} n_j + 1} \\ \vdots \\ x_{\sum_{j=1}^{k} n_j} \end{bmatrix}, \bar{Y} = \begin{bmatrix} \bar{Y}_1 \\ \vdots \\ \bar{Y}_K \end{bmatrix}, \bar{Y}_k = \begin{bmatrix} y_{\sum_{j=1}^{k-1} n_j + 1} \\ \vdots \\ y_{\sum_{j=1}^{k} n_j} \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix}
$$

eTable 1: Notation list of parameters for FL algorithms **A.2-5**

Symbol	Definition	FedAvg	FedAvgM	q-FedAvg	FedProx
K	Total number of clients	√	√	√	√
k	Clients index	√	√	√	√
B	Local minibatch size	√	√	√	√
E	Local epochs	√	√	√	√
η	Learning rate	√	√	√	√
C	The fraction of clients that perform computation on each round	√	√		
w	Model parameters	√	√	√	√
$\mathcal{P}_k$	The set of indexes of data points on client k	√	√		
n_k	The number of the elements in $\mathcal{P}_k$	√	√		
T	Total communication rounds	√	√	√	√
p_k	Probability of device k being selected			√	√

A.2 FedAvg

FedAvg[3] trains and updates models through interaction between a server and multiple clients. Three key parameters (C, E and B) are used to control the amount of computations. On each round, the server selects a C-fraction of clients and the gradient of the loss is calculated on the client using local data. The global batch size is decided by C ($C = 1$ means non-stochastic gradient descent).

Algorithm 2: FedAvg

```
Server executes :
initialize w_0, T;
for each round t = 1, 2, ..., T do
    m ← max(C · K, 1);
    S_t ←(random set of m clients);
    for each client k ∈ S_t in parallel do
        w^k_{t+1} ←ClientUpdate(k, w_t);
    end
    m_t ← Σ_{k∈S_t} n_k;
    w_{t+1} ← Σ_{k∈S_t} (n_k/m_t) w^k_{t+1};
end
ClientUpdate(k, w): //Run on client k
B ←(split P_k into batches of size B);
for each local epoch i from 1 to E do
    for batch b ∈ B do
        w ← w − η ∇ ℓ(w; b);
    end
    return w to server;
end
```

A.3 FedAvgM

Based on FedAvg, FedAvgM[4, 5] introduces momentum v and momentum parameter β. In the original algorithm of FedAvg, the weights are updated by $w \leftarrow w - \Delta w$. FedAvgM instead updates the model by calculating

$$v \leftarrow \beta v + \Delta w$$
$$w \leftarrow w - v$$

By introducing momentum, SGD can speed up convergence, improve the stability of the optimisation and dampen the oscillations of parameter updates[4, 5].

Algorithm 3: FedAvgM

Server executes :
Initialize w_0, T, β, $v_t(v_0 = 0)$.
for *each round* $t = 1, 2, ..., T$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$(random set of m clients)
  for *each client* $k \in S_t$ **in parallel do**
    $w_{t+1}^k \leftarrow$ClientUpdate(k, w_t)
  end
  $m_t \leftarrow \sum_{k\in S_t} n_k$
  $\alpha \leftarrow \sum_{k\in S_t} \frac{n_k}{m_t}(w_{t+1}^k - w_t)$
  $v_{t+1} \leftarrow \beta v_t + (1-\beta)\alpha$
  $w_{t+1} \leftarrow w_t + v_{t+1}$
end
ClientUpdate(k, w)**:** //Run on client k
$\mathcal{B} \leftarrow$(split $\mathcal{P}_k$ into batches of size B)
for *each local epoch i from 1 to E* **do**
  for *batch* $b \in \mathcal{B}$ **do**
    $w \leftarrow w - \eta \bigtriangledown \ell(w; b)$
  end
  return w to server
end

A.4 q-FedAvg

Based on FedAvg, q-FedAvg[6] uses a more complicated dynamic weight determined by the Lipschitz constant (L) of the gradient[7].The parameter q can be tuned based on the desired level of fairness (with larger q inducing more fairness). The q-FedAvg is the same to FedAvg when $q = 0$.

Algorithm 4: q-FedAvg

Input: $m, E, B, T, q, \frac{1}{L}, \eta, w^0, p_k, k = 1, ..., K$
for $t = 0, ..., T-1$ **do**
  Sever selects a subset S_t of m devices at random (each device k is chosen with prob. p_k)
  Server sends w^t to all selected devices
  Each selected device k updates w^t for E epochs of SGD on F_k with step-size η to obtain $\bar{w}_k^{t+1}$
  Each selected device k computes:
  $\Delta w_k^t = L(w_k^t - \bar{w}_k^{t+1})$
  $\Delta_k^t = F_k^q(w^t)\Delta w_k^t$
  $h_k^t = qF_k^{q-1}(w^t)\|\Delta w_k^t\|^2 + LF_k^q(w^t)$
  Each selected device k sends Δ_k^t and h_k^t back to the server
  Server updates w^{t+1} as:
  $w^{t+1} = w^t - \frac{\sum_{k\in S_t}\Delta_k^t}{\sum_{k\in S_t} h_k^t}$
end

A.5 FedProx

Based on FedAvg, FedProx[8] improves its stability to data and system heterogeneity[8] by adding a proximal term to the objective function:

$$\min_w h_k(w; w^t) = F_k(w) + \frac{\mu}{2}\|w - w^t\|^2$$

This modification allows the model to reduce the impact of non-IID while tolerating system heterogeneity. The speed of convergence is related to the penalty constant μ in the proximal term. The γ_k^t-inexactness for client k at iteration t is defined as follows:

Definition 1 (γ_k^t-inexact solution)**.** *For a function $h_k(w; w_t) = F_k(w) + \frac{\mu}{2}\|w - w_t\|^2$, and $\gamma \in [0, 1]$, we say w^* is a γ_k^t-inexact solution of $\min_w h_k(w; w_t)$ if $\|\nabla h_k(w^*; w_t)\| \leq \gamma_k^t \|\nabla h_k(w_t; w_t)\|$, where $\nabla h_k(w; w_t) = \nabla F_k(w) + \mu(w - w_t)$. Note that a smaller γ_k^t corresponds to higher accuracy.*

By adjusting the value of γ, which varies from device to device, FedProx solves the local function imprecisely.

Algorithm 5: FedProx

```
Input: m, E, B, T, μ, γ, w^0, N, p_k, k = 1, ..., K
for t = 0, ..., T − 1 do
    Server selects a subset S_t of m devices at random (each
      device k is chosen with probability p_k);
    Server sends w^t to all chosen devices;
    Each chosen device k ∈ S_t finds a w_k^{t+1} which is a
      γ_k^t-inexact minimizer of: w_k^{t+1} ≈ arg min_w
      h_k(w; w^t) = F_k(w) + (μ/2)‖w − w^t‖^2;
    Each device k ∈ S_t sends w_k^{t+1} back to the server;
    Server aggregates the w's as w^{t+1} = (1/m) Σ_{k∈S_t} w_k^{t+1};
end
```

B Point estimates result of real data

eTable 2: Coefficients of logistic regression estimated by all methods using real data

Settings	Variables	Central	Site1	Site2	Site3	GLORE	FedAvg	FedAvgM	q-FedAvg	Fedprox
A	Age	0.5283	0.2876	0.3012	0.7496	0.5283	0.7507	0.7513	0.1613	0.5989
	Gender	0.8067	0.6085	1.5819	0.5202	0.8067	0.7164	0.7165	0.1033	0.7581
	Pulse	0.5257	0.7261	0.4703	0.5366	0.5257	0.5505	0.5495	0.1085	0.5147
	Respiration	0.1435	0.0651	0.2727	0.1171	0.1435	0.0843	0.0833	0.0227	0.1504
	Oxygen saturation	-0.1068	-0.1255	-0.0769	-0.1179	-0.1068	-0.1142	-0.1155	-0.0206	-0.1054
	Diastolic blood pressure	-0.3352	-1.8172	-1.0422	-0.0341	-0.3352	-0.3420	-0.3340	-0.0318	-0.0713
	Systolic blood pressure	-0.3478	-0.0339	-0.3429	-0.3908	-0.3478	-0.3872	-0.3884	-0.0860	-0.3959
	Myocardial infarction	-0.1664	0.3301	-0.7309	-0.0981	-0.1664	-0.0464	-0.0457	0.0110	-0.1376
	Congestive heart failure	0.6635	0.1732	0.3659	0.6303	0.6635	0.1954	0.1977	0.0373	0.6322
	Stroke	0.7243	1.3362	0.9881	0.5464	0.7243	0.1635	0.1662	0.0466	0.6720
	Dementia	1.1288	2.3570	1.7407	0.4880	1.1288	0.4802	0.4786	0.0561	0.8934
	Chronic pulmonary disease	-0.9380	-2.0481	-1.5959	-0.4554	-0.9380	0.1151	0.1135	-0.0326	-0.7420
	Peptic ulcer disease	-0.2813	0.8205	0.4857	-0.5416	-0.2813	-0.2742	-0.2697	-0.0037	-0.3127
	Kidney disease	0.0755	0.0813	-0.5726	0.3997	0.0755	0.2600	0.2571	0.0210	0.1513
B	Age	0.9209	1.1543	0.9620	0.8259	0.9209	0.8244	0.8245	0.1714	0.9065
	Gender	0.0912	0.1913	0.1166	0.0411	0.0912	0.0809	0.0809	-0.0086	0.0746
	Pulse	0.4868	0.6070	0.4929	0.4451	0.4868	0.4711	0.4711	0.0822	0.4846
	Respiration	0.2373	0.2058	0.2347	0.2603	0.2373	0.2064	0.2064	0.0391	0.2402
	Oxygen saturation	-0.1431	-0.2123	-0.1272	-0.1387	-0.1431	-0.0972	-0.0972	-0.0211	-0.1446
	Diastolic blood pressure	-0.1039	-0.0907	-0.1116	-0.1057	-0.1039	-0.2126	-0.2126	-0.0316	-0.1034
	Systolic blood pressure	-0.4786	-0.4584	-0.4149	-0.5406	-0.4786	-0.4857	-0.4857	-0.0793	-0.4853
	Myocardial infarction	0.9667	0.8295	0.7697	1.1644	0.9667	0.9455	0.9455	0.1039	1.0100
	Congestive heart failure	-0.1275	-0.4793	-0.0589	-0.0823	-0.1275	0.1697	0.1697	0.0208	-0.1047
	Stroke	0.8478	1.1449	0.8576	0.7589	0.8478	0.7597	0.7597	0.0774	0.8204
	Dementia	0.1418	0.4379	0.2844	-0.1462	0.1418	0.1291	0.1291	0.0150	0.0575
	Chronic pulmonary disease	-0.2402	-0.1530	-0.1379	-0.3417	-0.2402	0.0287	0.0287	-0.0130	-0.2641
	Peptic ulcer disease	1.0726	1.1972	0.9500	1.1183	1.0726	0.7154	0.7155	0.0304	1.0718
	Kidney disease	0.6570	0.6312	0.7366	0.6011	0.6570	0.5572	0.5572	0.0952	0.6466

Settings	Variables	Central	Site1	Site2	Site3	GLORE	FedAvg	FedAvgM	q-FedAvg	Fedprox
C	Age	0.5545	0.1995	0.5851	0.6355	0.5545	0.7214	0.7216	0.1559	0.5777
	Gender	0.8873	1.7086	0.9276	0.5553	0.8873	0.7166	0.7135	0.1116	0.8340
	Pulse	0.4663	0.5443	0.4393	0.4011	0.4663	0.5391	0.5386	0.1007	0.4549
	Respiration	0.1168	-0.4351	0.0993	0.2713	0.1168	0.0707	0.0713	0.0043	0.1206
	Oxygen saturation	-0.1476	-0.0986	-0.3258	-0.1395	-0.1476	-0.1322	-0.1328	-0.0279	-0.2011
	Diastolic blood pressure	-0.0260	-1.4738	0.3583	-0.0314	-0.0260	-0.3408	-0.3431	-0.0198	-0.0101
	Systolic blood pressure	-0.3766	0.1494	-0.5192	-0.4801	-0.3766	-0.3860	-0.3853	-0.0785	-0.4228
	Myocardial infarction	0.4372	0.8660	0.4413	0.3175	0.4372	-0.0299	-0.0256	0.0249	0.4067
	Congestive heart failure	-0.0771	-0.5931	-0.1349	-0.0266	-0.0771	0.2232	0.2221	0.0221	-0.0801
	Stroke	0.7810	0.3269	0.3885	1.3584	0.7810	0.1652	0.1740	0.0235	0.7955
	Dementia	0.5048	1.0787	0.7017	-0.1556	0.5048	0.5288	0.5233	0.0245	0.4036
	Chronic pulmonary disease	-0.1863	-0.1089	0.0534	-0.6521	-0.1863	0.1142	0.1081	-0.0021	-0.1743
	Peptic ulcer disease	0.1567	1.5816	0.1799	0.1388	0.1567	-0.2389	-0.2338	-0.0002	0.1003
	Kidney disease	-0.0445	0.1908	-0.2254	0.1576	-0.0445	0.2486	0.2485	0.0303	-0.0632
D	Age	0.9468	1.3055	0.8270	0.9776	0.9468	0.8245	0.8245	0.1718	0.9302
	Gender	0.0153	0.1425	0.0363	-0.0583	0.0153	0.0836	0.0836	-0.0091	0.0021
	Pulse	0.5332	0.5950	0.5218	0.5280	0.5332	0.4703	0.4703	0.0825	0.5325
	Respiration	0.1834	0.1090	0.2027	0.1891	0.1834	0.2049	0.2049	0.0349	0.1903
	Oxygen saturation	-0.1140	-0.1219	-0.1207	-0.1066	-0.1140	-0.0991	-0.0991	-0.0192	-0.1111
	Diastolic blood pressure	-0.1626	-0.1941	-0.1905	-0.1225	-0.1626	-0.2128	-0.2128	-0.0345	-0.1612
	Systolic blood pressure	-0.4887	-0.4602	-0.4805	-0.5079	-0.4887	-0.4811	-0.4811	-0.0795	-0.4853
	Myocardial infarction	1.0115	1.2010	0.9680	0.9926	1.0115	0.9412	0.9412	0.1161	0.9938
	Congestive heart failure	0.0422	-0.2224	0.2062	-0.0486	0.0422	0.1746	0.1746	0.0273	0.0767
	Stroke	0.7064	1.0565	0.6096	0.6621	0.7064	0.7593	0.7593	0.0740	0.6680
	Dementia	0.1184	0.3397	0.1140	0.0364	0.1184	0.1412	0.1412	0.0126	0.0984
	Chronic pulmonary disease	-0.1408	-0.3209	-0.2333	0.0254	-0.1408	0.0329	0.0329	-0.0116	-0.1303
	Peptic ulcer disease	0.9466	0.4710	0.7289	1.2841	0.9466	0.7179	0.7179	0.0255	0.9676
	Kidney disease	0.5497	0.7460	0.5559	0.4738	0.5497	0.5594	0.5594	0.0953	0.5349

Settings	Variables	Central	MIMIC	SGH	GLORE	FedAvg	FedAvgM	q-FedAvg	Fedprox
E	Age	0.7934	0.6575	0.8188	0.7934	0.8072	0.8055	0.1761	0.8116
	Gender	0.1110	0.7977	0.0232	0.1110	0.1397	0.1390	0.0393	0.0576
	Pulse	0.4769	0.4926	0.4732	0.4769	0.4721	0.4729	0.0946	0.4787
	Respiration	0.2131	0.1423	0.2244	0.2131	0.1946	0.1956	0.0181	0.2150
	Oxygen saturation	-0.0842	-0.1396	-0.0719	-0.0842	-0.0979	-0.0990	-0.0227	-0.0824
	Diastolic blood pressure	-0.2497	-0.0416	-0.2576	-0.2497	-0.2301	-0.2292	-0.0388	-0.2483
	Systolic blood pressure	-0.4644	-0.3884	-0.4872	-0.4644	-0.4687	-0.4685	-0.0876	-0.4751
	Myocardial infarction	0.9181	0.0322	1.0086	0.9181	0.8596	0.8602	0.0588	0.9889
	Congestive heart failure	0.2298	0.4716	0.2170	0.2298	0.1785	0.1789	0.0253	0.2158
	Stroke	0.6734	0.4740	0.7222	0.6734	0.7076	0.7073	0.0381	0.7127
	Dementia	0.2518	0.5591	0.2248	0.2518	0.1765	0.1774	0.0143	0.2290
	Chronic pulmonary disease	0.0456	-0.0481	0.0543	0.0456	0.0430	0.0431	0.0036	0.0510
	Peptic ulcer disease	0.6963	-0.4572	0.9108	0.6963	0.6300	0.6287	0.0030	0.8750
	Kidney disease	0.5109	0.3482	0.5404	0.5109	0.5302	0.5295	0.0665	0.5329

*Hyperparameter: FedAvg($\eta = 0.1$), FedAvgM($\eta = 0.1$), q-FedAvg($\eta = 0.1$), FedProx($\eta = 0.01, \mu = 0$)

C Point estimate results of GLORE

eTable 3: Coverage and average confidence intervals estimated by GLORE (setting I)

Small sample size									
Settings	**Estimated CIs**	$\beta_1 = -2$	$\beta_2 = 1$	$\beta_3 = 0.8$	$\beta_4 = 0.4$	$\beta_5 = 0.2$	$\beta_6 = 0.1$	$\beta_7 = 0$	$\beta_8 = 0$
Homogeneous	Coverage	0.970	0.930	0.960	0.930	0.960	0.940	0.960	0.960
	Lower mean	-2.134	0.914	0.719	0.323	0.119	0.021	-0.084	-0.071
	Upper mean	-1.892	1.095	0.891	0.483	0.275	0.176	0.071	0.084
Shift Mean(0.1)	Coverage	0.970	0.970	0.990	0.950	0.930	0.980	0.930	0.950
	Lower mean	-2.129	0.915	0.720	0.320	0.116	0.023	-0.080	-0.075
	Upper mean	-1.889	1.094	0.890	0.478	0.271	0.176	0.073	0.079
Shift Mean(0.2)	Coverage	0.960	0.940	0.980	0.930	0.940	0.950	0.930	0.950
	Lower mean	-2.128	0.914	0.721	0.320	0.118	0.024	-0.079	-0.074
	Upper mean	-1.889	1.092	0.890	0.477	0.271	0.176	0.073	0.078
Shift Mean(0.3)	Coverage	0.940	0.920	0.980	0.940	0.950	0.970	0.900	0.950
	Lower mean	-2.127	0.916	0.723	0.323	0.118	0.024	-0.078	-0.073
	Upper mean	-1.889	1.093	0.891	0.479	0.270	0.175	0.073	0.078
Shift Mean(0.4)	Coverage	0.960	0.940	0.980	0.930	0.970	0.960	0.970	0.950
	Lower mean	-2.126	0.916	0.722	0.325	0.118	0.024	-0.078	-0.073
	Upper mean	-1.888	1.092	0.889	0.480	0.269	0.174	0.072	0.077
Large sample size									
Settings	**Estimated CIs**	$\beta_1 = -2$	$\beta_2 = 1$	$\beta_3 = 0.8$	$\beta_4 = 0.4$	$\beta_5 = 0.2$	$\beta_6 = 0.1$	$\beta_7 = 0$	$\beta_8 = 0$
Homogeneous	Coverage	0.940	0.960	0.950	0.970	0.940	0.950	0.960	0.930
	Lower mean	-2.069	0.944	0.749	0.353	0.151	0.055	-0.044	-0.046
	Upper mean	-1.930	1.048	0.848	0.445	0.241	0.144	0.045	0.043
Shift Mean(0.1)	Coverage	0.950	0.940	0.920	0.950	0.970	0.880	0.960	0.970
	Lower mean	-2.073	0.951	0.752	0.352	0.154	0.056	-0.045	-0.045
	Upper mean	-1.935	1.054	0.850	0.443	0.243	0.144	0.043	0.043
Shift Mean(0.2)	Coverage	0.940	0.950	0.910	0.940	0.960	0.880	0.960	0.980
	Lower mean	-2.073	0.951	0.751	0.353	0.154	0.057	-0.044	-0.045
	Upper mean	-1.935	1.053	0.849	0.443	0.242	0.144	0.043	0.042
Shift Mean(0.3)	Coverage	0.940	0.940	0.890	0.950	0.940	0.890	0.970	0.990
	Lower mean	-2.072	0.950	0.752	0.354	0.154	0.058	-0.044	-0.044
	Upper mean	-1.935	1.052	0.848	0.444	0.241	0.145	0.043	0.043
Shift Mean(0.4)	Coverage	0.930	0.950	0.920	0.950	0.940	0.900	0.970	0.970
	Lower mean	-2.071	0.950	0.753	0.354	0.154	0.058	-0.043	-0.044
	Upper mean	-1.934	1.051	0.849	0.443	0.241	0.144	0.043	0.042

eTable 4: Coverage and average confidence intervals estimated by GLORE (setting II)

Small sample size									
Settings	**Estimated CIs**	$\beta_1 = -2$	$\beta_2 = 1$	$\beta_3 = 0.8$	$\beta_4 = 0.4$	$\beta_5 = 0.2$	$\beta_6 = 0.1$	$\beta_7 = 0$	$\beta_8 = 0$
Homogeneous	Coverage	0.970	0.930	0.960	0.930	0.960	0.940	0.960	0.960
	Lower mean	-2.134	0.914	0.719	0.323	0.119	0.021	-0.084	-0.071
	Upper mean	-1.892	1.095	0.891	0.483	0.275	0.176	0.071	0.084
Shift mean(0.1) SD(0.1)	Coverage	0.950	0.960	0.970	0.970	0.960	0.980	0.950	0.920
	Lower mean	-2.131	0.917	0.721	0.322	0.118	0.024	-0.079	-0.071
	Upper mean	-1.891	1.094	0.890	0.478	0.270	0.176	0.072	0.080
Shift mean(0.1) SD(0.2)	Coverage	0.960	0.930	0.970	0.960	0.950	0.980	0.950	0.920
	Lower mean	-2.127	0.917	0.722	0.322	0.117	0.025	-0.079	-0.069
	Upper mean	-1.888	1.093	0.889	0.476	0.268	0.174	0.071	0.080
Shift mean(0.1) SD(0.3)	Coverage	0.940	0.940	0.980	0.960	0.950	0.990	0.960	0.940
	Lower mean	-2.129	0.918	0.722	0.324	0.118	0.026	-0.078	-0.069
	Upper mean	-1.890	1.094	0.888	0.476	0.266	0.174	0.070	0.079
Shift mean(0.1) SD(0.4)	Coverage	0.940	0.940	0.980	0.970	0.970	0.980	0.970	0.950
	Lower mean	-2.130	0.921	0.724	0.323	0.118	0.026	-0.075	-0.069
	Upper mean	-1.890	1.095	0.889	0.475	0.266	0.173	0.071	0.078
Large sample size									
Settings	**Estimated CIs**	$\beta_1 = -2$	$\beta_2 = 1$	$\beta_3 = 0.8$	$\beta_4 = 0.4$	$\beta_5 = 0.2$	$\beta_6 = 0.1$	$\beta_7 = 0$	$\beta_8 = 0$
Homogeneous	Coverage	0.940	0.960	0.950	0.970	0.940	0.950	0.960	0.930
	Lower mean	-2.069	0.944	0.749	0.353	0.151	0.055	-0.044	-0.046
	Upper mean	-1.930	1.048	0.848	0.445	0.241	0.144	0.045	0.043
Shift mean(0.1) SD(0.1)	Coverage	0.960	0.970	0.920	0.940	0.960	0.890	0.980	0.980
	Lower mean	-2.071	0.950	0.753	0.353	0.156	0.056	-0.044	-0.045
	Upper mean	-1.933	1.052	0.850	0.443	0.243	0.143	0.043	0.043
Shift mean(0.1) SD(0.2)	Coverage	0.930	0.960	0.930	0.950	0.940	0.900	0.970	0.970
	Lower mean	-2.070	0.949	0.752	0.354	0.155	0.057	-0.044	-0.044
	Upper mean	-1.932	1.051	0.848	0.442	0.242	0.143	0.042	0.042
Shift mean(0.1) SD(0.3)	Coverage	0.940	0.940	0.940	0.950	0.970	0.900	0.960	0.980
	Lower mean	-2.070	0.949	0.754	0.354	0.155	0.057	-0.042	-0.043
	Upper mean	-1.932	1.050	0.849	0.442	0.241	0.143	0.043	0.042
Shift mean(0.1) SD(0.4)	Coverage	0.970	0.950	0.930	0.950	0.970	0.910	0.950	0.980
	Lower mean	-2.070	0.949	0.753	0.355	0.155	0.058	-0.042	-0.043
	Upper mean	-1.932	1.049	0.848	0.442	0.240	0.142	0.042	0.041

eTable 5: Coverage and average confidence intervals estimated by GLORE (setting III)

Small sample size									
Settings	**Estimated CIs**	$\beta_1 = -2$	$\beta_2 = 1$	$\beta_3 = 0.8$	$\beta_4 = 0.4$	$\beta_5 = 0.2$	$\beta_6 = 0.1$	$\beta_7 = 0$	$\beta_8 = 0$
Homogeneous	Coverage	0.970	0.930	0.960	0.930	0.960	0.940	0.960	0.960
	Lower mean	-2.134	0.914	0.719	0.323	0.119	0.021	-0.084	-0.071
	Upper mean	-1.892	1.095	0.891	0.483	0.275	0.176	0.071	0.084
Shift effect(0.1)	Site1 Coverage	0.000	0.120	0.230	0.740	0.940	0.940	0.960	0.940
	Site2 Coverage	0.800	0.900	0.910	0.910	0.960	0.950	0.960	0.940
	Site3 Coverage	0.560	0.800	0.850	0.880	0.960	0.970	0.960	0.940
	Lower mean	-2.210	0.951	0.747	0.336	0.125	0.024	-0.085	-0.072
	Upper mean	-1.960	1.136	0.922	0.498	0.283	0.181	0.072	0.085
Shift effect(0.2)	Site1 Coverage	0.000	0.000	0.000	0.220	0.870	0.900	0.950	0.950
	Site2 Coverage	0.420	0.660	0.800	0.860	0.940	0.950	0.950	0.950
	Site3 Coverage	0.030	0.270	0.420	0.790	0.930	0.910	0.950	0.950
	Lower mean	-2.270	0.977	0.768	0.348	0.130	0.026	-0.084	-0.073
	Upper mean	-2.014	1.164	0.945	0.512	0.289	0.184	0.074	0.085
Large sample size									
Settings	**Estimated CIs**	$\beta_1 = -2$	$\beta_2 = 1$	$\beta_3 = 0.8$	$\beta_4 = 0.4$	$\beta_5 = 0.2$	$\beta_6 = 0.1$	$\beta_7 = 0$	$\beta_8 = 0$
Homogeneous	Coverage	0.940	0.960	0.950	0.970	0.940	0.950	0.960	0.930
	Lower mean	-2.069	0.944	0.749	0.353	0.151	0.055	-0.044	-0.046
	Upper mean	-1.930	1.048	0.848	0.445	0.241	0.144	0.045	0.043
Shift effect(0.1)	Site1 Coverage	0.000	0.000	0.020	0.360	0.840	0.880	0.930	0.930
	Site2 Coverage	0.450	0.760	0.750	0.900	0.920	0.950	0.930	0.930
	Site3 Coverage	0.070	0.280	0.440	0.830	0.920	0.930	0.930	0.930
	Lower mean	-2.148	0.981	0.779	0.367	0.159	0.059	-0.044	-0.049
	Upper mean	-2.005	1.087	0.880	0.460	0.249	0.149	0.046	0.042
Shift effect(0.2)	Site1 Coverage	0.000	0.000	0.000	0.000	0.440	0.760	0.930	0.950
	Site2 Coverage	0.040	0.390	0.480	0.880	0.900	0.930	0.930	0.950
	Site3 Coverage	0.000	0.010	0.010	0.310	0.760	0.910	0.930	0.950
	Lower mean	-2.204	1.008	0.801	0.378	0.165	0.061	-0.046	-0.049
	Upper mean	-2.057	1.116	0.903	0.472	0.257	0.152	0.045	0.042

D Communication Cost

eTable 6: Communication cost of simulation data

Settings		GLORE		FedAvg		FedAvgM		q-FedAvg		FedProx	
		Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
	Homogeneous	5.59	5.48	10	10	10	10	30	30	10	20
I	Shift mean (0.1)	5.35	5.12	10	10	10	10	30	30	20	20
	Shift mean (0.2)	5.38	5.14	10	10	10	10	30	30	20	20
	Shift mean (0.3)	5.40	5.17	10	10	10	10	30	30	20	20
	Shift mean (0.4)	5.36	5.17	10	10	10	10	30	30	20	20
II	Shift mean (0.1) SD (0.1)	5.58	5.50	10	10	10	10	30	30	20	20
	Shift mean (0.1) SD (0.2)	5.78	5.87	10	10	10	10	30	30	20	20
	Shift mean (0.1) SD (0.3)	5.93	5.99	10	10	10	10	30	30	20	20
	Shift mean (0.1) SD (0.4)	5.98	6.00	10	10	10	10	30	30	20	20
III	Shift effect (0.1)	5.95	5.98	10	10	10	10	30	30	20	20
	Shift effect (0.2)	6.00	6.00	10	10	10	10	30	30	20	20

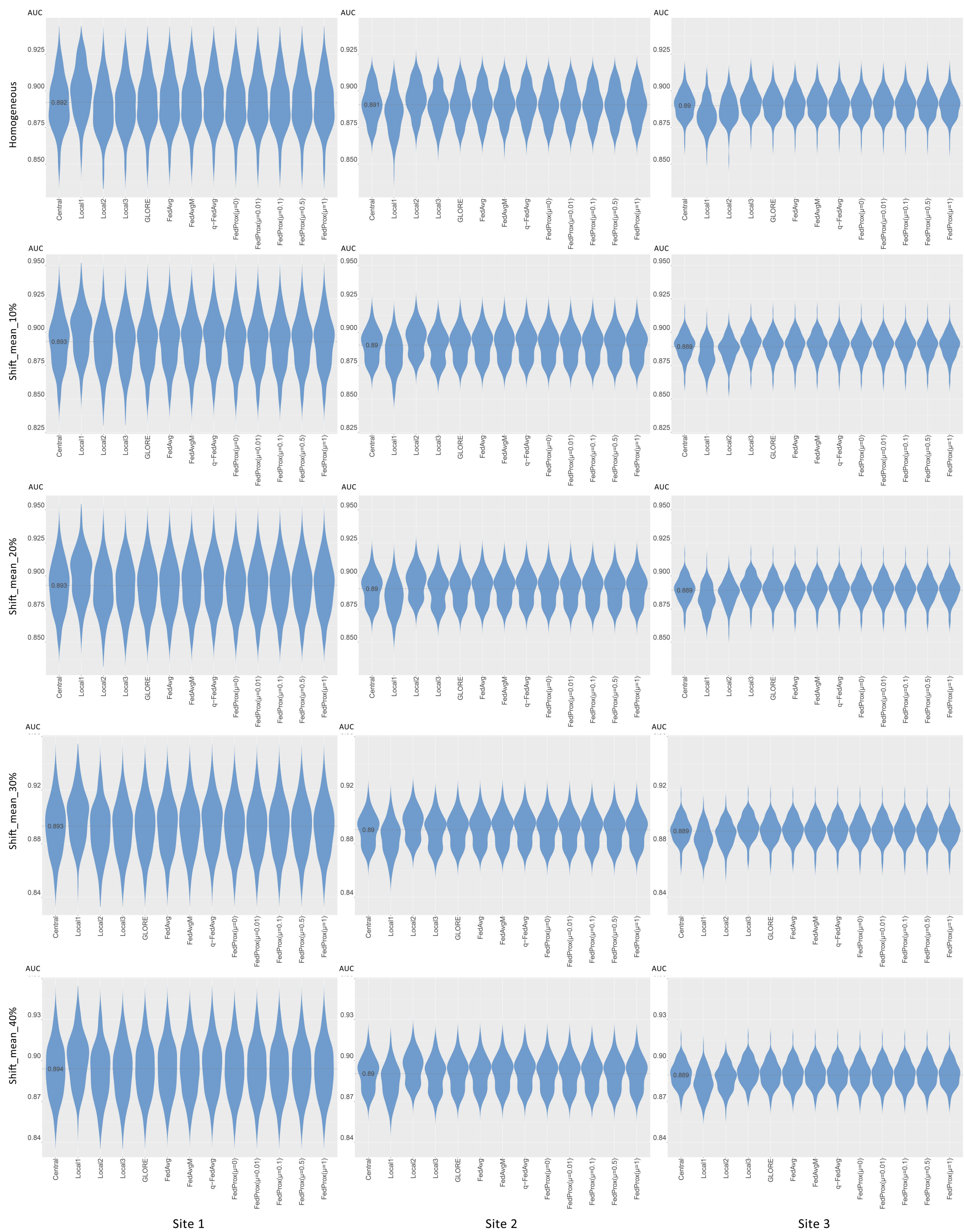
eTable 7: Communication cost of real data

Settings	GLORE	FedAvg	FedAvgM	q-FedAvg	Fedprox
A	7	10	10	30	100
B	7	10	10	30	50
C	7	10	10	30	100
D	7	10	10	30	50
E	7	10	10	30	50

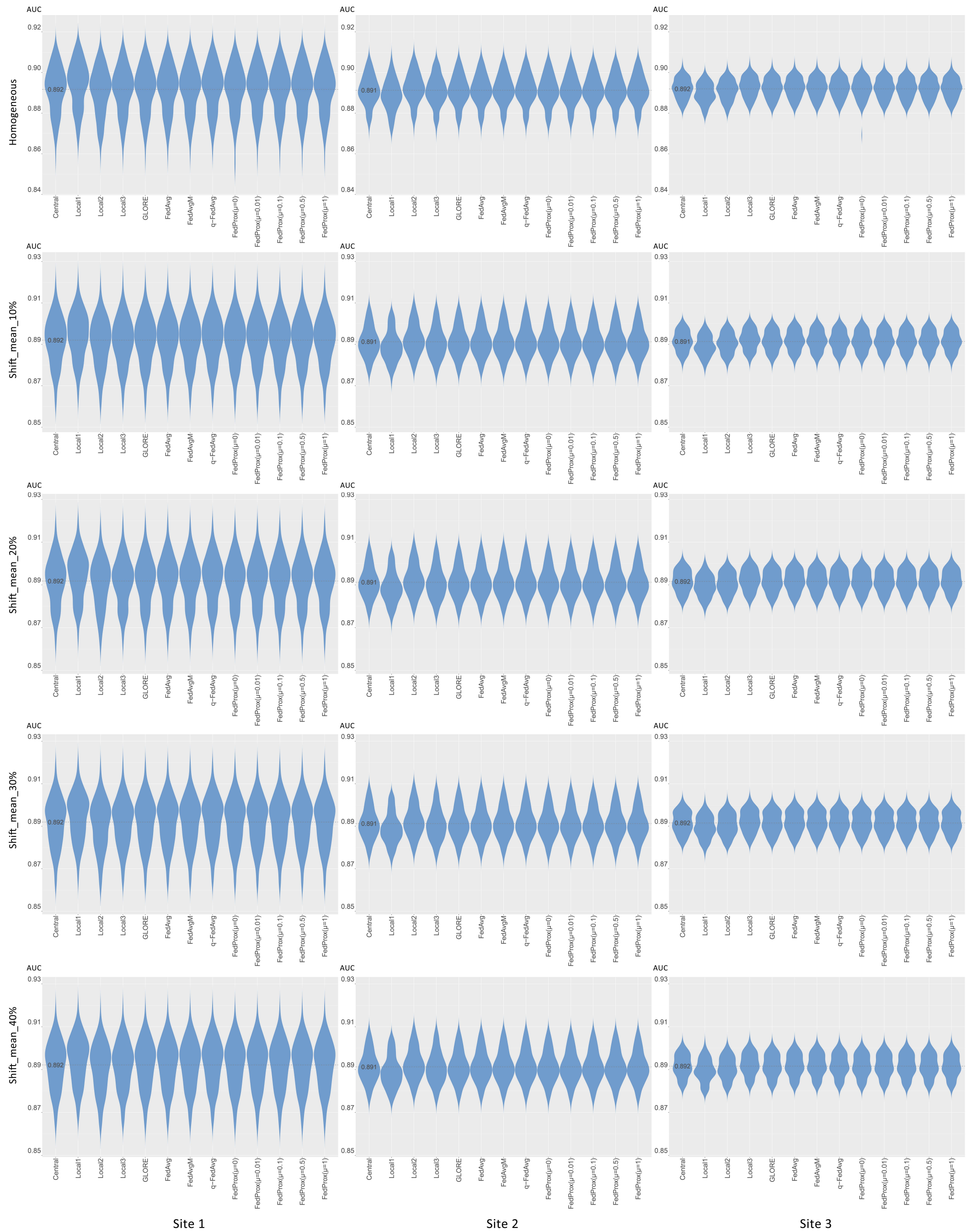
E Evaluations of simulation study

E.1 Prediction tasks

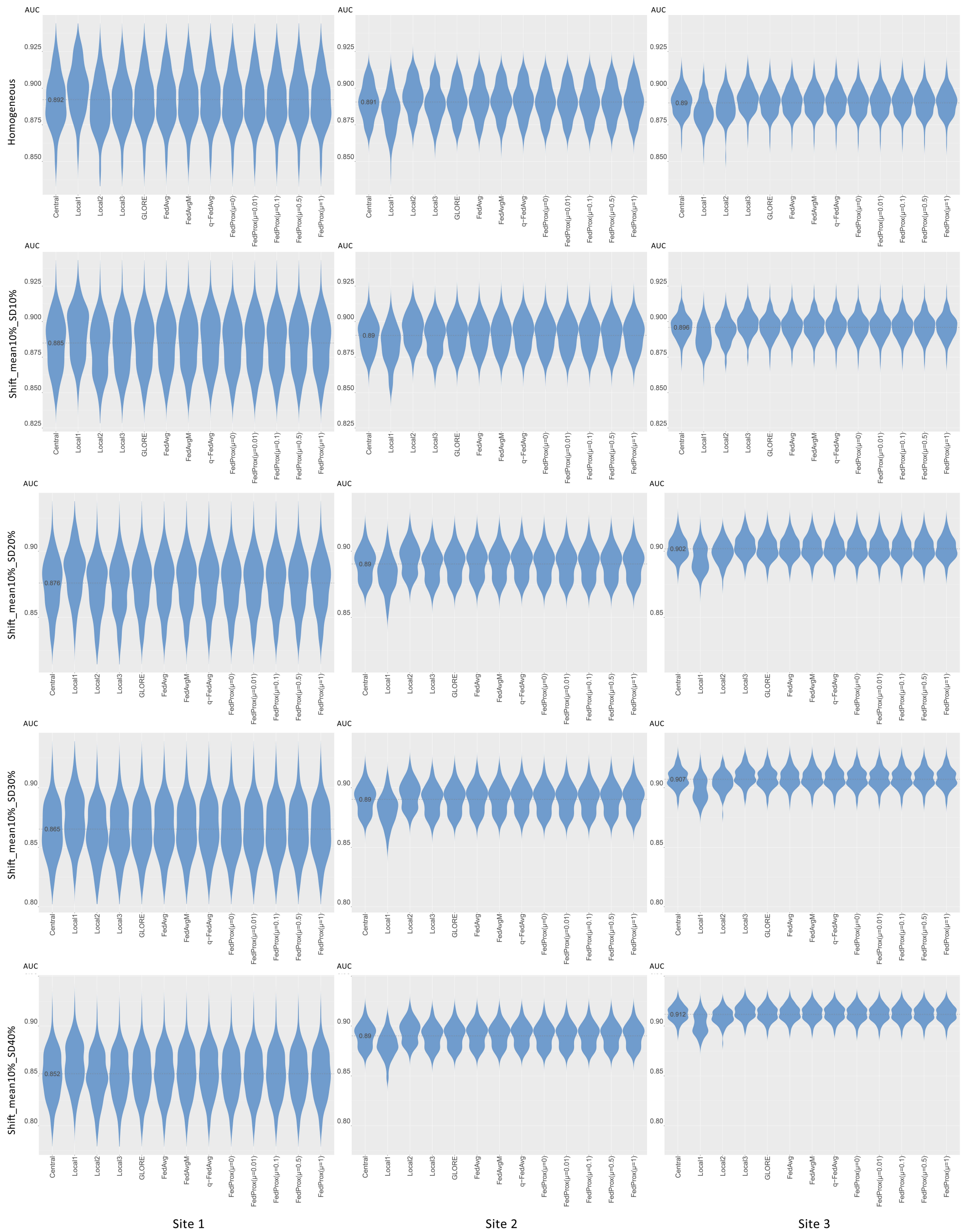
eFigure 1: FL Model comparisons by prediction performance under shifting of covariate mean with relatively small sample size



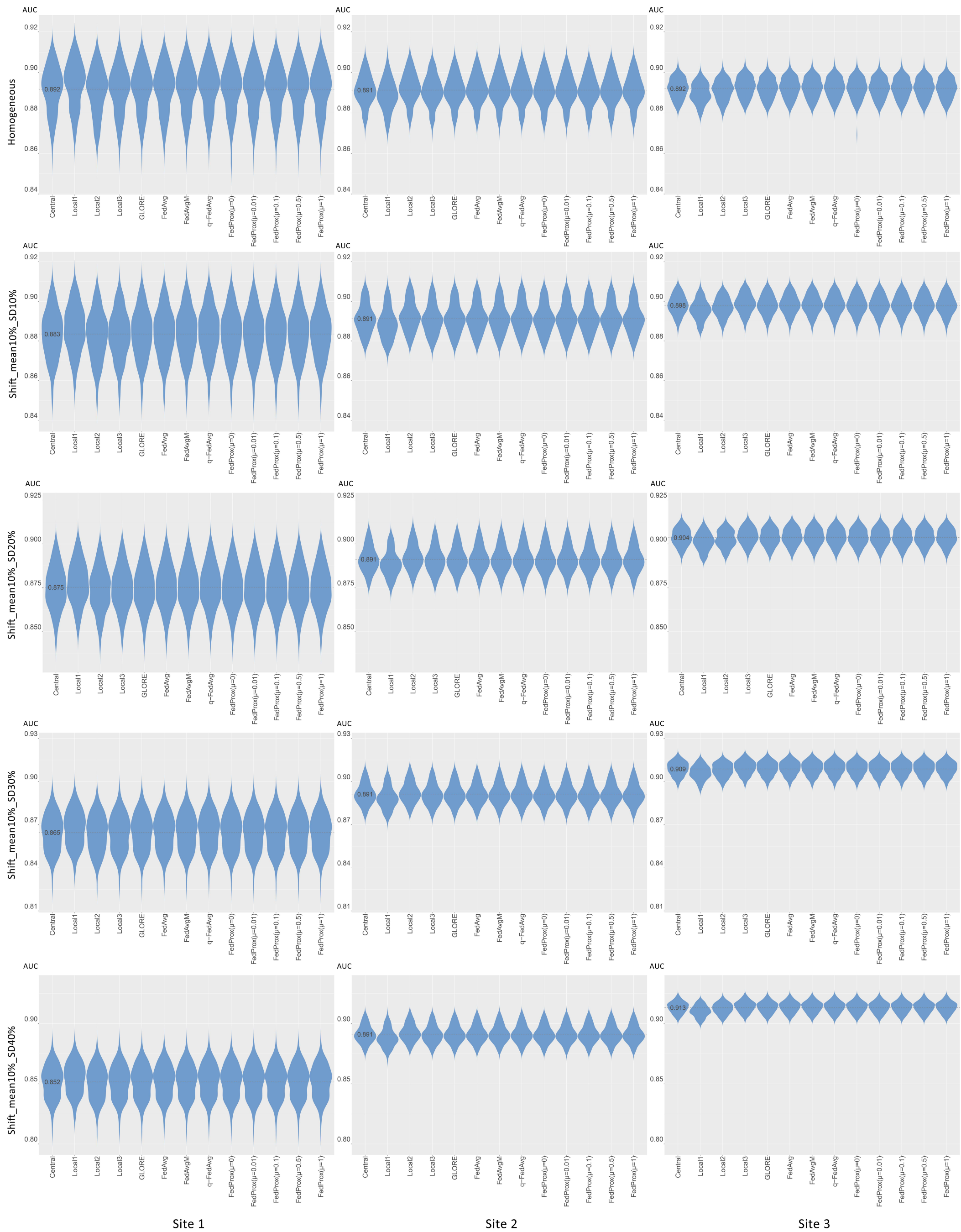
eFigure 2: FL Model comparisons by prediction performance under shifting of covariate mean with relatively large sample size



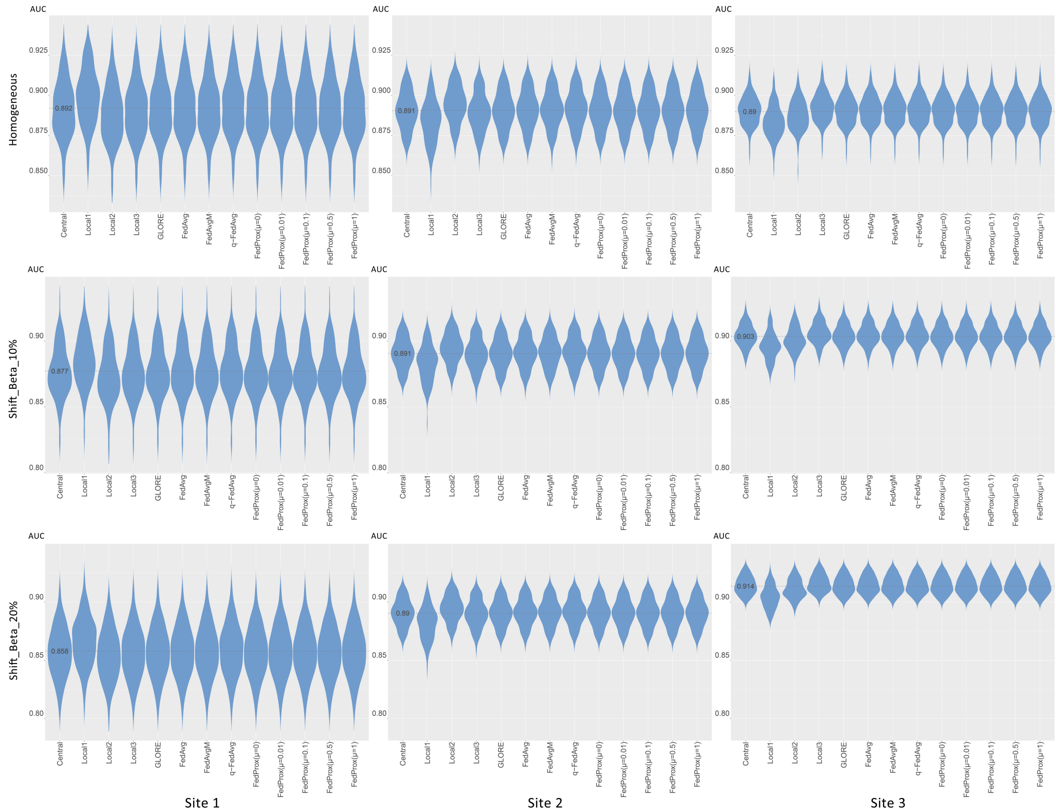
eFigure 3: FL Model comparisons by prediction performance under shifting of covariate standard deviation (SD) with relatively small sample size



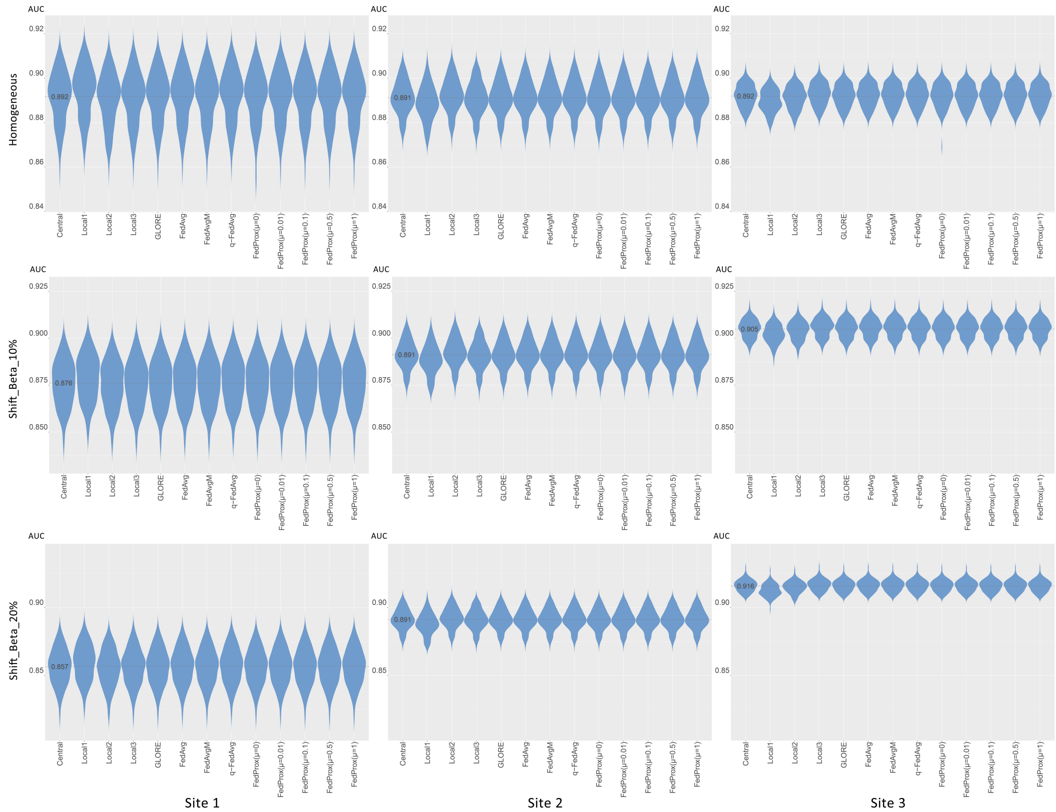
eFigure 4: FL Model comparisons by prediction performance under shifting of covariate SD with relatively large sample size



eFigure 5: FL Model comparisons by prediction performance under shifting of effect size with relatively small sample size

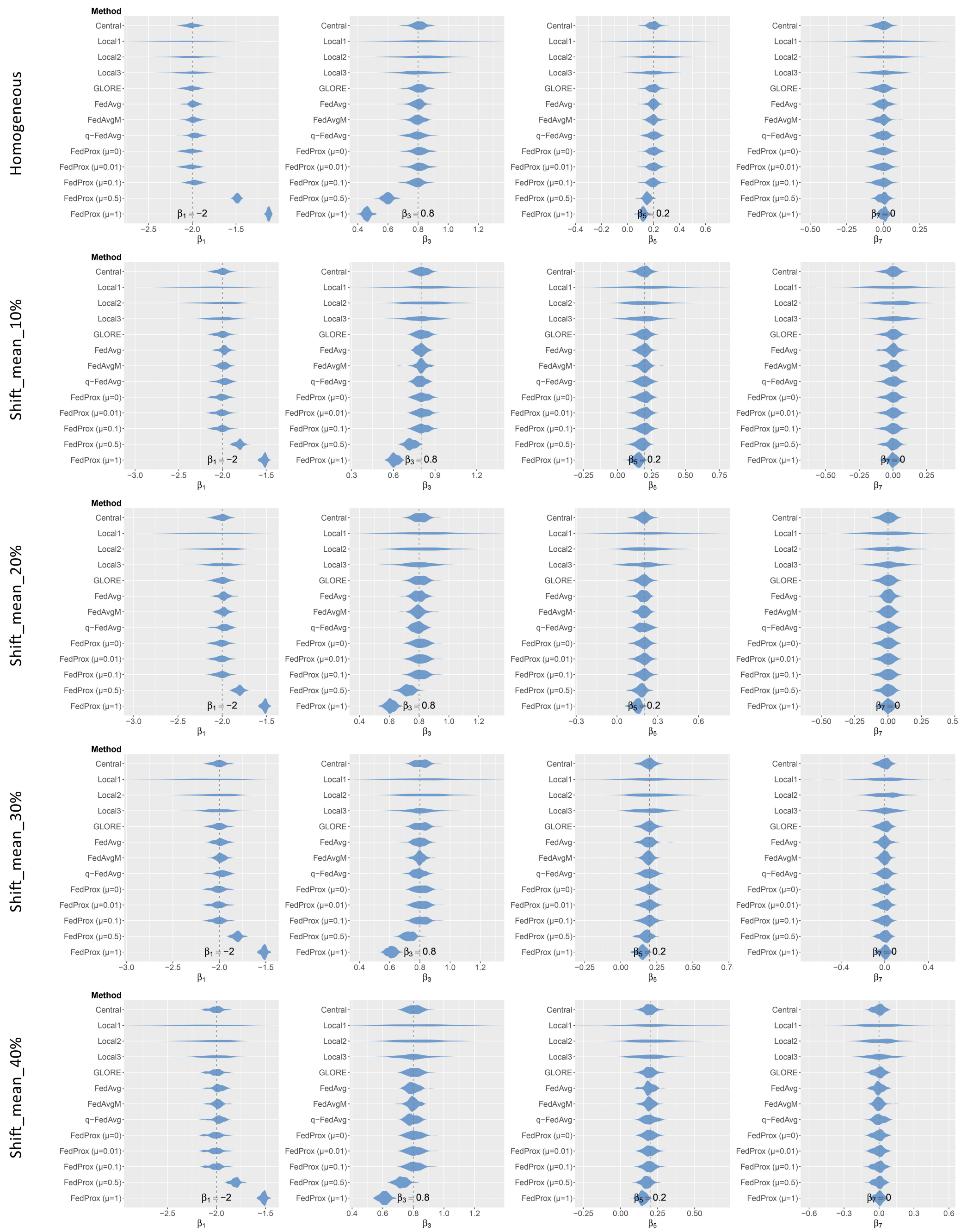


eFigure 6: FL Model comparisons by prediction performance under shifting of effect size with relatively large sample size

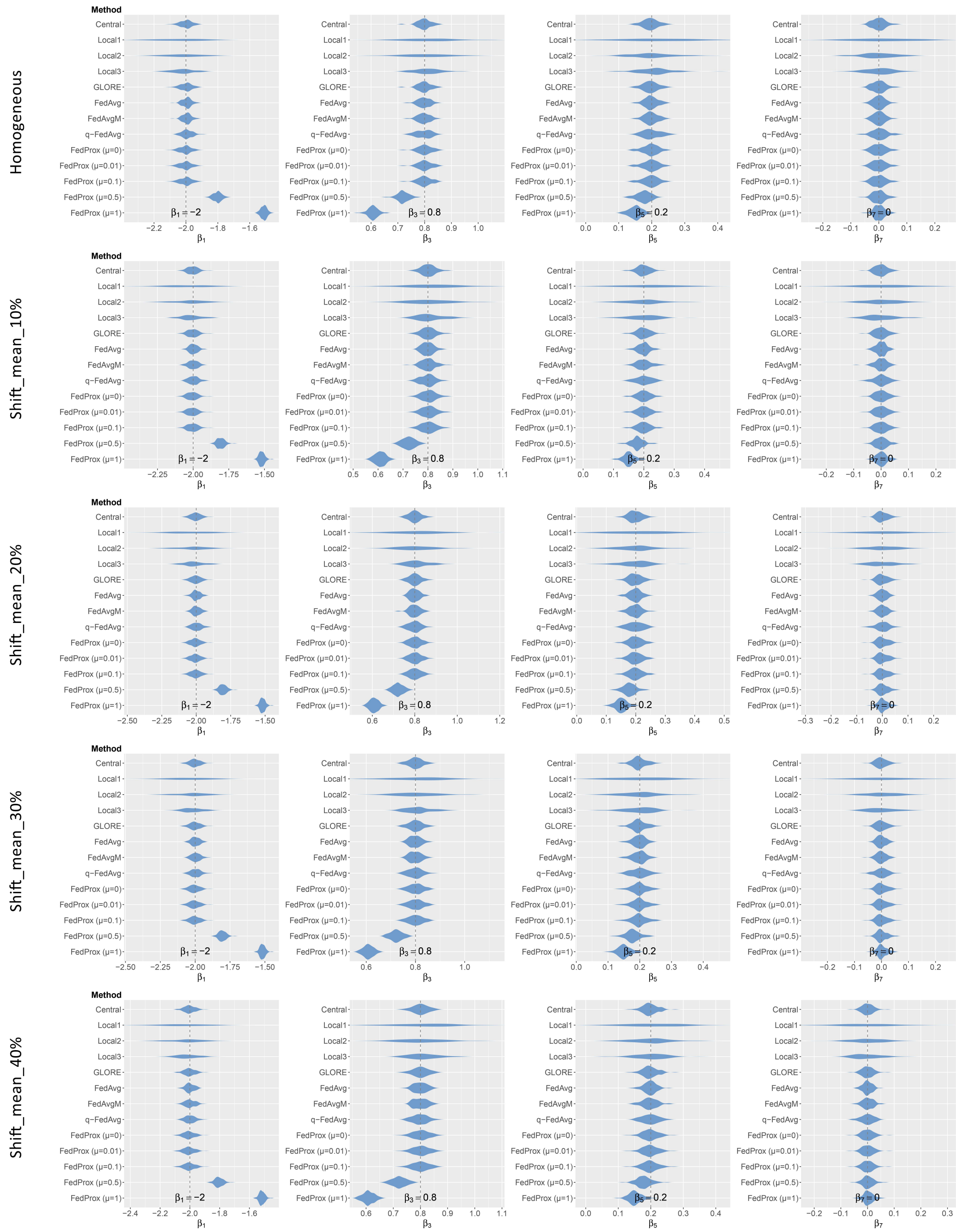


E.2 Point estimates

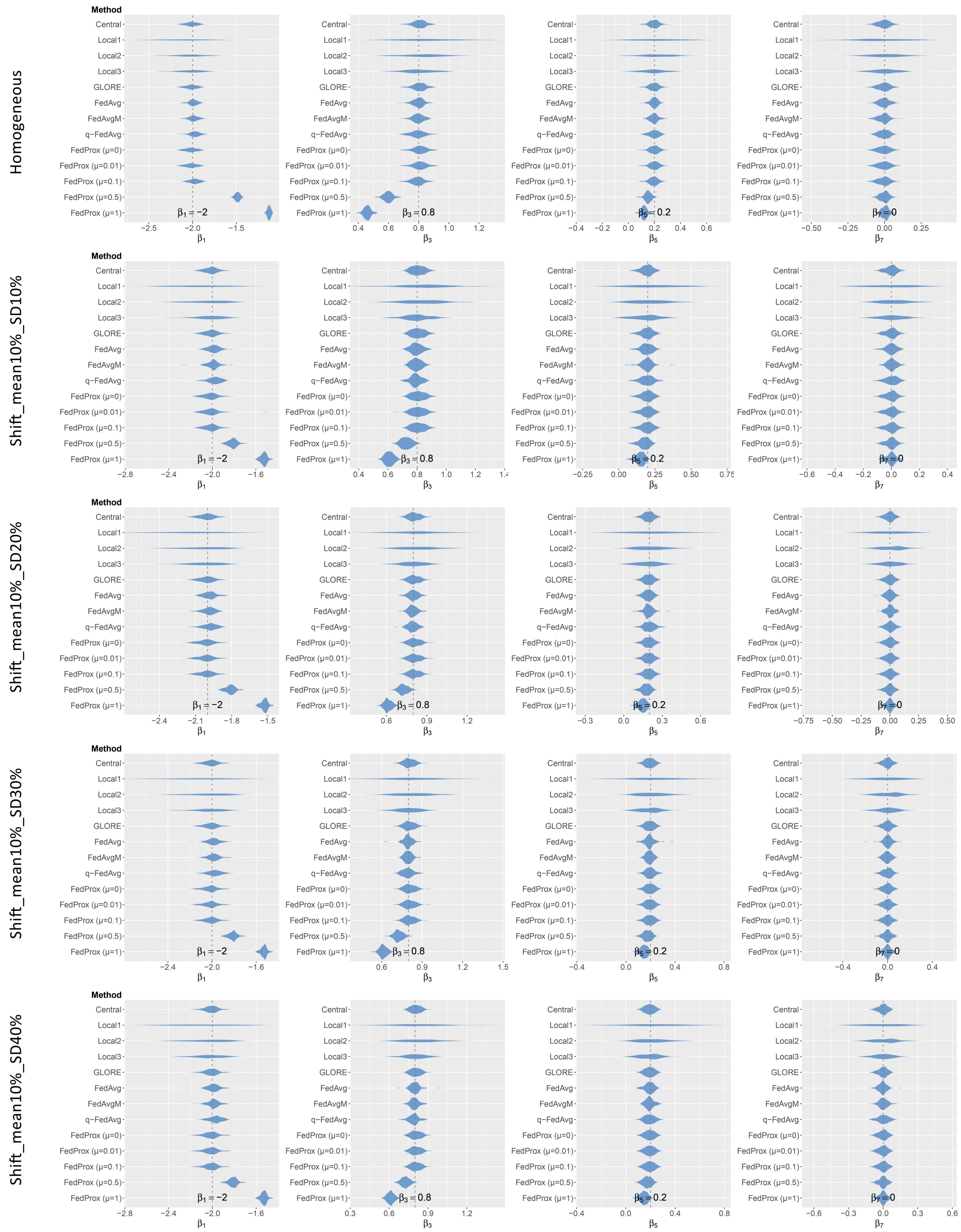
eFigure 7: FL Model comparisons by estimated coefficients under shifting of mean with relatively small sample size



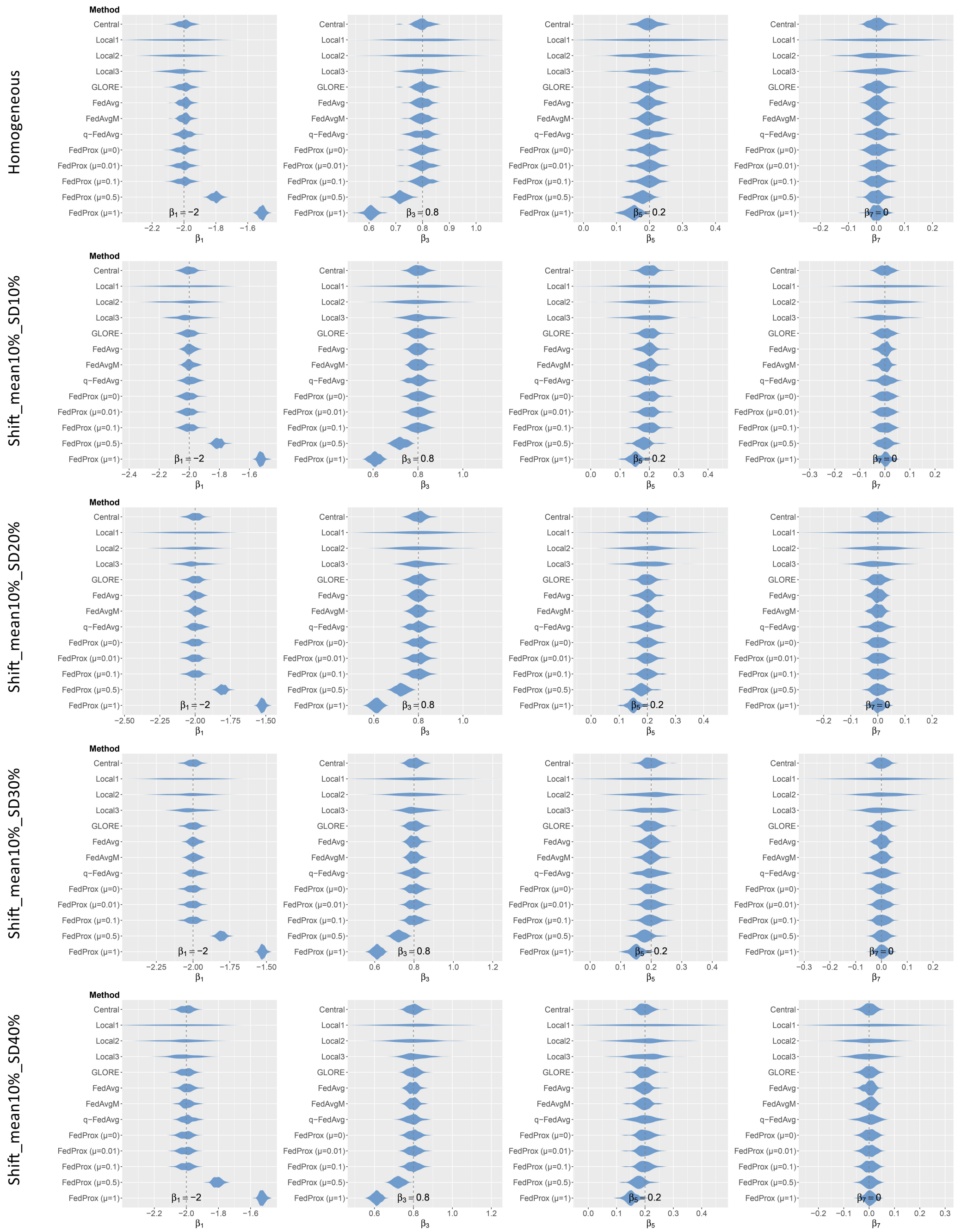
eFigure 8: FL Model comparisons by estimated coefficients under shifting of mean with relatively large sample size



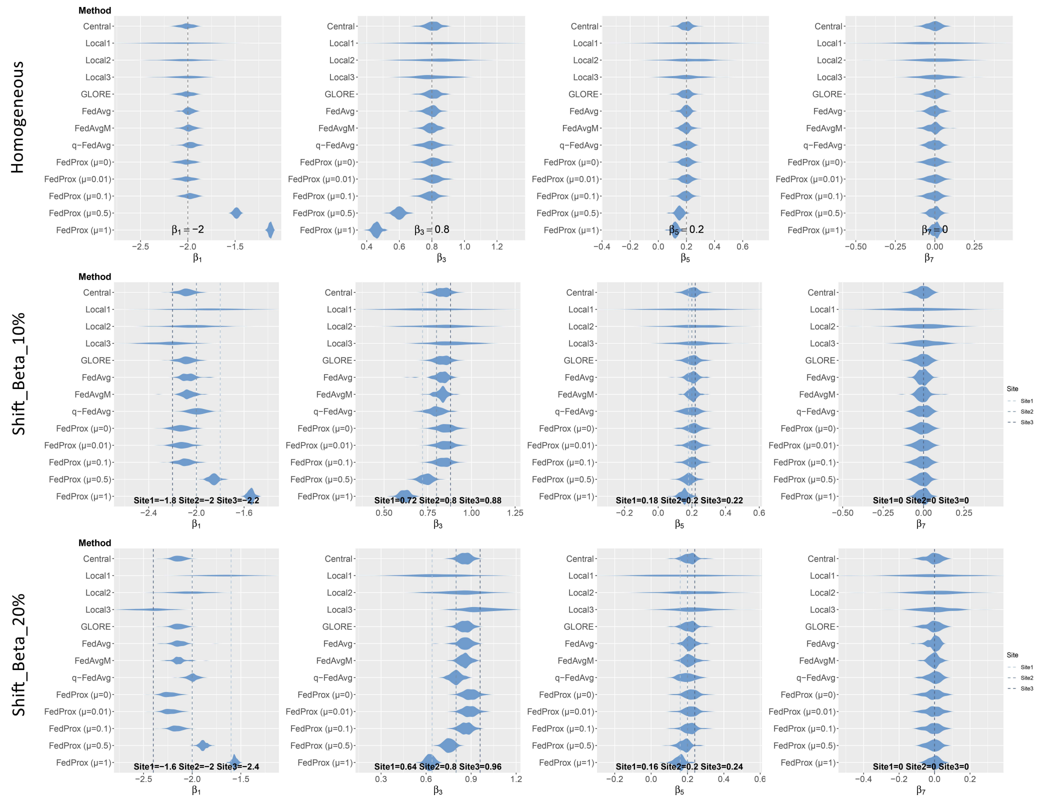
eFigure 9: FL Model comparisons by estimated coefficients under shifting of standard deviation (SD) with relatively small sample size



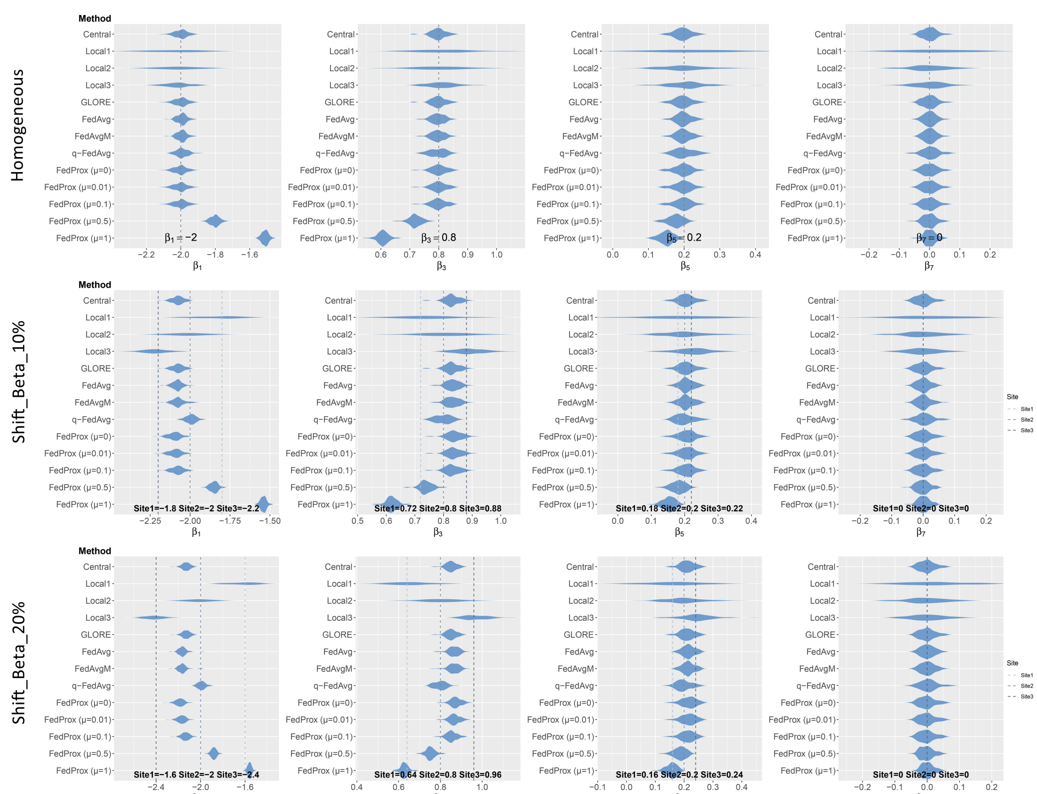
eFigure 10: FL Model comparisons by estimated coefficients under shifting of standard deviation (SD) with relatively large sample size



eFigure 11: FL Model comparisons by estimated coefficients under shifting of effect size with relatively small sample size



eFigure 12: FL Model comparisons by estimated coefficients under shifting of effect size with relatively small large size



References

[1] Y. Wu, X. Jiang, J. Kim, and L. Ohno-Machado, "Grid binary logistic regression (glore): building shared models without sharing data," *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 758–764, 2012.

[2] T. Minka, "A comparison of numerical optimizers for logistic regression," March 2003. [Online]. Available: https://www.microsoft.com/en-us/research/publication/comparison-numerical-optimizers-logistic-regression/

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Singh and J. Zhu, Eds., vol. 54. PMLR, 20–22 Apr 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[4] T.-M. H. Hsu, H. Qi, and M. Brown, "Measuring the effects of non-identical data distribution for federated visual classification," *arXiv preprint arXiv:1909.06335*, 2019.

[5] T. Salazar, M. Fernandes, H. Araujo, and P. H. Abreu, "Fair-fate: Fair federated learning with momentum," *arXiv preprint arXiv:2209.13678*, 2022.

[6] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," *arXiv preprint arXiv:1905.10497*, 2019.

[7] J. Bernstein, J. Zhao, K. Azizzadenesheli, and A. Anandkumar, "signsgd with majority vote is communication efficient and fault tolerant," *arXiv preprint arXiv:1810.05291*, 2018.

[8] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.