

Communication-efficient federated logistic lasso for OMOP patient-level prediction: preliminary evaluation across five databases

Egill A. Fridgeirsson¹, Martijn Schuemie², Jenna M. Reps^{1,2}

¹Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands

²Johnson & Johnson, Titusville, NJ, USA

Background

Federated patient-level prediction could support multi-database model development in the Observational Medical Outcomes Partnership (OMOP) Common Data Model when patient-level data cannot be pooled. Communication-efficient one-shot and low-communication methods are attractive for this setting, but many such methods become difficult to use with the very high-dimensional feature spaces typical of patient-level prediction. Recent OHDSI work suggests that a constrained, reusable set of phenotype predictors plus age and sex can retain much of the performance of models trained with thousands of candidate predictors across several tasks and databases[1]. This creates a practical setting in which federated L1-regularized logistic regression methods may be feasible. However, the tradeoff between communication-efficient surrogate methods and iterative federated optimization remains unclear, especially when database populations and covariate distributions differ. We conducted a federated method evaluation using five distinct OMOP data sources, implemented in a simulated federation in which each database was treated as a separate site.

Methods

The five OMOP databases were Merative MarketScan® Multi-State Medicaid Database (MDCD), Merative MarketScan® Medicare Supplemental Database (MDCR), Merative MarketScan® Commercial Claims and Encounters (CCAE), Japan Medical Data Center (JMDC), and Optum® de-identified Electronic Health Record dataset (Optum® EHR). Leave-one-database-out external validation trained models on four databases and evaluated discrimination on the fifth. Tasks were deep vein thrombosis (DVT) and major bleed in tetraplegia, and bleed after warfarin. Predictors were age, sex, and constrained phenotypes from [1]. Time-at-risk was days 1-365 for the tetraplegia tasks and days 1-30 for bleed after warfarin. People with prior outcomes were excluded. Across databases, tasks included 125,481-1,150,425 cohort entries and 4,413-17,373 outcomes (1.5%-5.0% overall outcome rates, Table 1).

Table 1.

Task	Time-at-risk	Total N	Total outcomes	Site N range	Site outcome count range	Site outcome rate range
DVT in Tetraplegia	1 – 365 days	143,065	4,413	3,172-73,131	65-1,828	2.0%-5.3%
Major bleed in tetraplegia	1 – 365 days	125,481	6,335	2,628-68,075	97-2,526	3.7%-10.3%
Bleed after warfarin	1-30 days	1,150,425	17,373	5,823-661,435	35-9,316	0.6%-2.9%

The centralized Cyclops pooled lasso was treated as an oracle benchmark rather than a deployable federated method [2]. Its penalty was selected using internal 10-fold cross-validation on the combined four-database training data. For local coefficient averaging, each training database fitted a lasso model with internal 10-fold cross-validation, then coefficients were averaged with sample-size weights. DualAvg was an iterative federated optimizer for the pooled L1-regularized logistic objective [3]. Its penalty was selected using inner leave-one-training-site-out cross-validation. ODAL1 and ODAL2 were few-exchange surrogates for unpenalized logistic regression using aggregate gradient summaries, with ODAL2 additionally using Hessian summaries [4]. ADAP1 was a penalized first-order surrogate with lead-site five-fold cross-validation [5]. Surrogate methods used the largest training database as the lead site. ODAL1 and ODAL2 used feature filtering matched to pooled lasso. All tuning used training data only. Federation was simulated on one machine.

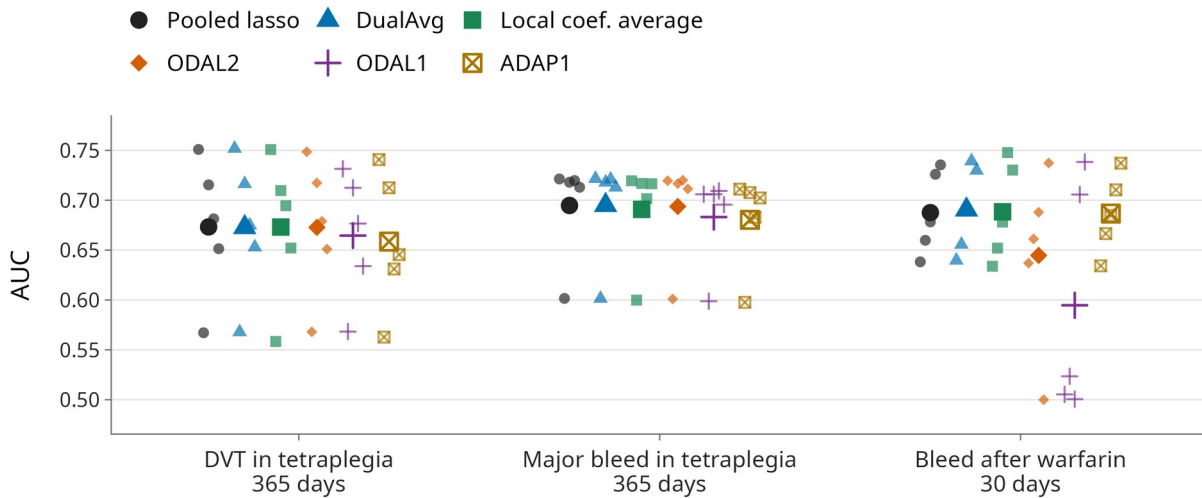
Results

Across 15 task-database evaluations, mean AUC was 0.685 for pooled lasso, 0.686 for DualAvg, and 0.684 for sample-size-weighted averaging of local lasso coefficients. These averages gave equal weight to each task-database evaluation. Both DualAvg and coefficient averaging had observed mean discrimination close to the centralized benchmark (Figure 1).

ODAL2 had an overall mean AUC of 0.670. Its task-level means were close to pooled lasso for DVT in tetraplegia (0.673 versus 0.673) and major bleed in tetraplegia (0.694 versus 0.695), but lower for bleed after warfarin (0.645 versus 0.688). The warfarin evaluation with JMDC held out produced extreme coefficients and AUC 0.500. All observed fold AUCs were retained. ODAL1 and ADAP1 had overall mean AUCs of 0.647 and 0.675, respectively.

For phenotype and sex coefficients, excluding age and intercept, median Pearson correlation with pooled lasso was 0.994 for DualAvg, 0.929 for local coefficient averaging, and 0.969 for ODAL2. Corresponding median root mean squared errors were 0.021, 0.060, and 0.046. Each metric was calculated within a paired task-database evaluation and summarized across all 15 evaluations. ODAL2 fits an unpenalized surrogate, whereas pooled lasso shrinks coefficients, so exact coefficient agreement is not expected.

A Held-out-database AUCs and task means



B Paired AUC difference from pooled lasso

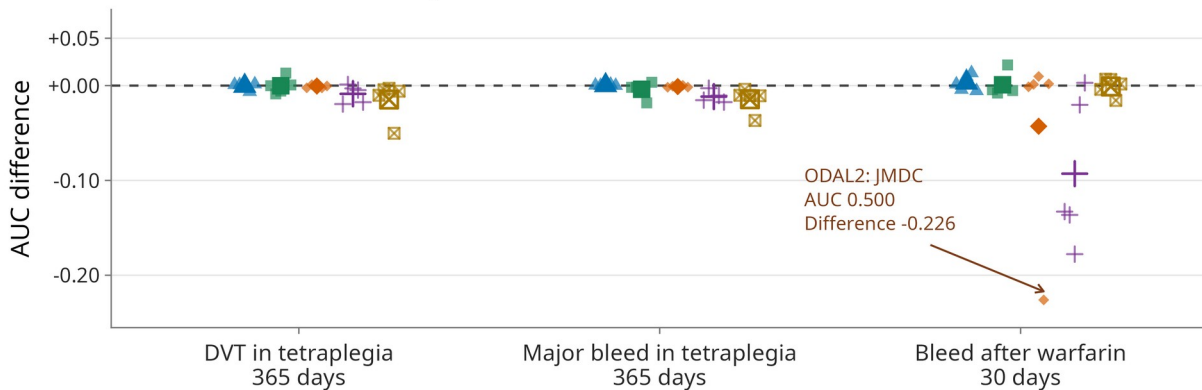


Figure 1. External-validation discrimination. Panel A shows AUCs and panel B shows differences from pooled lasso in the same held-out database. Small markers show the five held-out databases and larger markers show equally weighted task means. All observed fold AUCs are included.

Conclusions

Across three prediction tasks and five databases, sample-size-weighted coefficient averaging and DualAvg achieved mean AUCs close to pooled lasso. ODAL2 was also competitive on the two one-year tasks but less consistent for bleeding after warfarin. These results highlight the value of simple coefficient averaging as a baseline for federated prediction.

References

1. Reps JM, Wong J, Fridgeirsson EA, Kim C, John LH, Williams RD, et al. Finding a constrained number of predictor phenotypes for multiple outcome prediction. *BMJ Health Care Inform.* 2025 Jan 16;32(1):e101227. doi:10.1136/bmjhci-2024-101227 PubMed PMID: 39824520; PubMed Central PMCID: PMC11751890.

2. Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive parallelization of serial inference algorithms for a complex generalized linear model. *ACM Transactions on Modeling and Computer Simulation*. 2013 Jan 1;23(1). doi:10.1145/2414416.2414791
3. Yuan H, Zaheer M, Reddi S. Federated Composite Optimization. In: *Proceedings of the 38th International Conference on Machine Learning* [Internet]. PMLR; 2021 [cited 2026 May 18]. p. 12253–66. Available from: <https://proceedings.mlr.press/v139/yuan21d.html>
4. Duan R, Boland MR, Liu Z, Liu Y, Chang HH, Xu H, et al. Learning from electronic health records across multiple sites: A communication-efficient and privacy-preserving distributed algorithm. *J Am Med Inform Assoc*. 2020 Mar 1;27(3):376–85. doi:10.1093/jamia/ocz199
5. Liu X, Duan R, Luo C, Ogdie A, Moore JH, Kranzler HR, et al. Multisite learning of high-dimensional heterogeneous data with applications to opioid use disorder study of 15,000 patients across 5 clinical sites. *Sci Rep*. 2022 Jun 30;12(1):11073. doi:10.1038/s41598-022-14029-9