Evaluating Federated Learning's Application in Epidemiology: Identifying Risk Factors for Treatment-Resistant Depression

Yaki Saciuk, MPH[1]*; Echo Wang, DrPH[2]*; Hongtao Zhang, PhD[2]; Lawrence Gould, PhD[2]; Moshe Hoshen, PhD[1]; Alan Apter, MD[1]; Tal Patalon, MD, MBA[3]; Mehmet Burcu, PhD, MS, FISPE[2]; Sam Joo, PhD[2]; Julia Dibello, PhD[2]#; Sivan Gazit, MD, MA[1]#

[1]Kahn Sagol Maccabi (KSM) Research & Innovation Center, Maccabi Healthcare Services, Tel Aviv, Israel; [2]Merck & Co., Inc., Rahway, NJ, USA; [3]Maccabi Healthcare Services, Tel Aviv, Israel. *Co-first authors. #Co-corresponding authors.

Background and objectives

Treatment-resistant depression (TRD) is commonly defined as nonremission after at least 2 adequate antidepressant treatments.[1] Risk factors include severe symptoms, depression chronicity, psychiatric/medical comorbidities, and psychosocial factors.[2] However, inconsistent findings between studies limit generalizability.[3] The federated learning (FL) approach enables analysis across data sources while preserving privacy. Despite its increasing adoption in machine learning tasks, FL's application in statistical inferences remains limited. Federated data networks provide a major opportunity to generate real-world evidence (RWE) from privacy-restricted markets and data sources, and understanding the quality of federated analysis is critical to its adoption in practice. Therefore, we implemented federated algorithms for statistical modeling (termed *federated analysis*, FA) to identify TRD risk factors and compare the federated results to those obtained on centralized data.

Methods

Study population and TRD definition

The study includes active members who have been continuously enrolled in Maccabi Healthcare Services (MHS) healthcare plan, with at least 2 years of continuous membership prior to the index date. TRD was defined as initiating third-line treatment within 2 years of first-line therapy.

Federated analysis (FA) setup (Figure 1)

Splitting the data – creating an emulated multicenter study

The MHS cohort should be divided into 2 distinct and non-overlapping subcohorts, site (a) and site (b), ideally yielding 2 sets of data that are non-independent and non-identically distributed (non-IID). The criterion for splitting is members' place of residence – "living in peripheral Israel" or "living in central Israel." Peripheral regions generally correspond to a lower average socioeconomic stratum but still have people from a full range of socioeconomic statuses (SESs).

Implementing federated analysis methodology in a statistical inference study (Figure 2)

Our federated framework was implemented on the Rhino Health Federated Computing Platform. Federated descriptive statistics were computed separately at each site and aggregated to produce summary baseline characteristics. We implemented federated versions of generalized linear models (GLMs) (logistic regression) and Cox proportional hazard (PH) models based on established methods in the literature. The GLM framework follows the approach described by Cellamare et al.,[4] while the Cox PH estimation adopts the distributed procedure introduced by Lu et al.[5] Variable selection was stabilized through 10-fold cross-validation: In each iteration, backward stepwise elimination The Akaike information criterion (AIC-based) was applied to 9 folds, and predictors retained in ≥50% of iterations were refitted in the final model. In the federated setting, we evaluated 2 feature selection strategies. The first, a site-level union approach (naïve approach), defined the federated feature set as the union of predictors selected independently at each site. The second was a fully federated stepwise backward elimination procedure that identified features jointly across sites without requiring disclosure of local selections.

Metrics for evaluation

- **Model performance metrics for centralized and federated models:** C statistic, sensitivity, specificity, positive predictive value (PPV), Hosmer-Lemeshow statistic, and Brier score
- **Evaluate the similarity between centralized and federated coefficients**
  - Mean difference between the models' coefficients
  - Decision agreement – determining if the same conclusion is drawn from the two estimates based on statistical significance and the estimate direction
  - Estimate agreement – evaluating whether the aggregated point estimate falls within the confidence interval of the unified one
  - Overlap between the confidence intervals (CI)
- **Model fairness and potential biases:** Algorithmic bias, or model fairness, is defined as the differential results or performance of predictive models that may lead to differential allocation or outcomes between subgroups.[6] In this study, we will use selected fairness measures (ie, false-positive rate difference) to evaluate if the federated model reduces algorithm bias compared to the centralized model between certain demographic groups, such as biological sex, social sector, and SES

Figure 1. Federated analysis setup and workflow

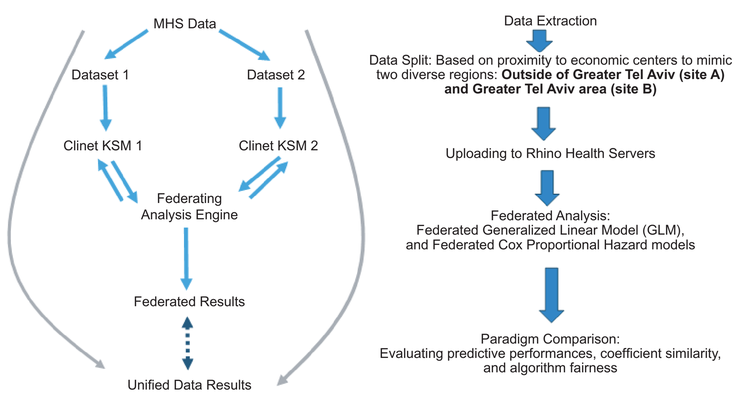
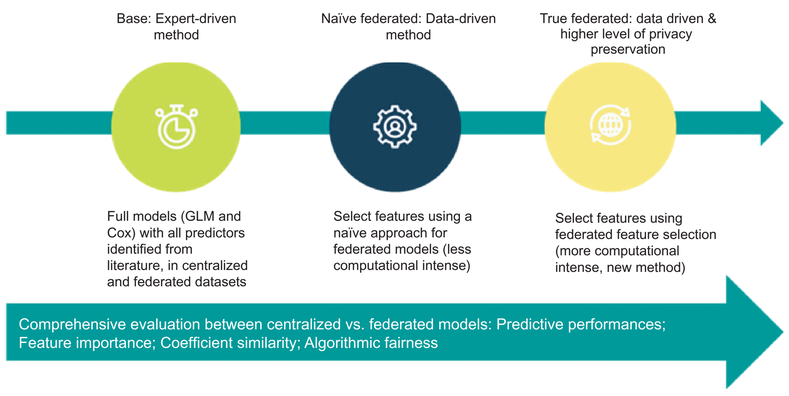


Figure 2. Federated analysis scenarios evaluated that mimic typical real-world analyses



Results

Table 1. Characteristics of patients with MDD during their first episode during the research period

	Outside Greater Tel Aviv area (Site A)	Greater Tel Aviv area (Site B)	Total population (centralized)	Total population (FA)
	(N=4915)	(N=2357)	(N=7272)	(N=7272)
Age at first MDD episode				
Mean (SD)	49.4 (17.3)	48.1 (17.4)	48.9 (17.3)	48.9 (17.3)
Median [Min, Max]	49.7 [18.0, 99.0]	47.4 [18.0, 97.0]	49.1 [18.0, 99.0]	49.0 [17.7, 98.9]
Sex (n, %)				
F	3032 (61.7%)	1436 (60.9%)	4468 (61.4%)	4468 (61.4%)
M	1883 (38.3%)	921 (39.1%)	2804 (38.6%)	2804 (38.6%)
Socioeconomic status (n, %)				
Middle	2950 (60.0%)	1174 (49.8%)	4124 (56.7%)	4124 (56.7%)
Low	492 (10.0%)	74 (3.1%)	566 (7.8%)	566 (7.8%)
High	1473 (30.0%)	1109 (47.1%)	2582 (35.5%)	2582 (35.5%)
Social sector (n, %)				
General	4439 (90.3%)	2222 (94.3%)	6661 (91.6%)	6661 (91.6%)
Arab	208 (4.2%)	20 (0.8%)	228 (3.1%)	228 (3.1%)
Orthodox	268 (5.5%)	115 (4.9%)	383 (5.3%)	383 (5.3%)
Cerebrovascular accident (CVA) (n, %)[a]				
No	4855 (98.8%)	2328 (98.8%)	7183 (98.8%)	7183 (98.8%)
Yes	60 (1.2%)	29 (1.2%)	89 (1.2%)	89 (1.2%)
Myocardial infarction (MI) (n, %)				
No	4714 (95.9%)	2274 (96.5%)	6988 (96.1%)	6988 (96.1%)
Yes	201 (4.1%)	83 (3.5%)	284 (3.9%)	284 (3.9%)
Atrial fibrillation (AF) (n, %)				
No	4813 (97.9%)	2297 (97.5%)	7110 (97.8%)	7110 (97.8%)
Yes	102 (2.1%)	60 (2.5%)	162 (2.2%)	162 (2.2%)
Congestive heart failure (CHF)				
No	4870 (99.1%)	2333 (99.0%)	7203 (99.1%)	7203 (99.1%)
Yes	45 (0.9%)	24 (1.0%)	69 (0.9%)	69 (0.9%)
Ischemic heart disease (no-MI) (n, %)				
No	4757 (96.8%)	2279 (96.7%)	7036 (96.8%)	7036 (96.8%)
Yes	158 (3.2%)	78 (3.3%)	236 (3.2%)	236 (3.2%)
Peripheral vascular disease (PVD) (n, %)				
No	4845 (98.6%)	2319 (98.4%)	7164 (98.5%)	7164 (98.5%)
Yes	70 (1.4%)	38 (1.6%)	108 (1.5%)	108 (1.5%)
Hypertension (n, %)[a]				
No	3505 (71.3%)	1778 (75.4%)	5283 (72.6%)	5283 (72.6%)
Yes	1410 (28.7%)	579 (24.6%)	1989 (27.4%)	1989 (27.4%)
Diabetes mellitus (n, %)[a]				
No	4144 (84.3%)	2013 (85.4%)	6157 (84.7%)	6157 (84.7%)
Yes	771 (15.7%)	344 (14.6%)	1115 (15.3%)	1115 (15.3%)
Immunosuppression (n, %)[a]				
No	4688 (95.4%)	2253 (95.6%)	6941 (95.4%)	6941 (95.4%)
Yes	227 (4.6%)	104 (4.4%)	331 (4.6%)	331 (4.6%)
Chronic kidney disease (CKD) (n, %)[a]				
No	4904 (99.8%)	2346 (99.5%)	7250 (99.7%)	7250 (99.7%)
Yes	11 (0.2%)	11 (0.5%)	22 (0.3%)	22 (0.3%)
Human immunodeficiency virus (HIV) (n, %)[a]				
No	4905 (99.8%)	2344 (99.4%)	7249 (99.7%)	7249 (99.7%)
Yes	10 (0.2%)	13 (0.6%)	23 (0.3%)	23 (0.3%)
Inflammatory bowel disease (IBD) (n, %)[a]				
No	4828 (98.2%)	2322 (98.5%)	7150 (98.3%)	7150 (98.3%)
Yes	87 (1.8%)	35 (1.5%)	122 (1.7%)	122 (1.7%)
Cancer (n, %)[a]				
No	4439 (90.3%)	2125 (90.2%)	6564 (90.3%)	6564 (90.3%)
Yes	476 (9.7%)	232 (9.8%)	708 (9.7%)	708 (9.7%)
Autism (n, %)[b]				
No	4904 (99.8%)	2353 (99.8%)	7257 (99.8%)	7257 (99.8%)
Yes	11 (0.2%)	4 (0.2%)	15 (0.2%)	15 (0.2%)
Dementia (n, %)[a]				
No	4745 (96.5%)	2310 (98.0%)	7055 (97.0%)	7055 (97.0%)
Yes	170 (3.5%)	47 (2.0%)	217 (3.0%)	217 (3.0%)
Hypothyroidism (n, %)[c]				
No	4685 (95.3%)	2271 (96.4%)	6956 (95.7%)	6956 (95.7%)
Yes	230 (4.7%)	86 (3.6%)	316 (4.3%)	316 (4.3%)
Hyperthyroidism (n, %)[c]				
No	4875 (99.2%)	2337 (99.2%)	7212 (99.2%)	7212 (99.2%)
Yes	40 (0.8%)	20 (0.8%)	60 (0.8%)	60 (0.8%)
Thyroiditis (n, %)[c]				
No	4852 (99.4%)	2312 (99.4%)	7164 (99.4%)	7164 (99.4%)
Yes	27 (0.6%)	13 (0.6%)	40 (0.6%)	40 (0.6%)
Polypharmacy – therapeutic groups (n)[d]				
Mean (SD)	5.6 (4.9)	5.1 (4.7)	5.5 (4.9)	5.5 (4.8)
Median [Min, Max]	4 [0, 33]	4 [0, 30]	4 [0, 33]	4 [0, 32]
Corticosteroids in the last 3 months (n, %)				
No	4732 (96.3%)	2266 (96.1%)	6998 (96.2%)	6998 (96.2%)
Yes	183 (3.7%)	91 (3.9%)	274 (3.8%)	274 (3.8%)
Charlson Comorbidity Index (CCI) (n, %)				
0	2017 (41.0%)	1064 (45.1%)	3081 (42.4%)	3081 (42.4%)
1	937 (19.1%)	424 (18.0%)	1361 (18.7%)	1361 (18.7%)
2	601 (12.2%)	258 (10.9%)	859 (11.8%)	859 (11.8%)
3	418 (8.5%)	158 (6.7%)	576 (7.9%)	576 (7.9%)
≥4	942 (19.2%)	453 (19.2%)	1395 (19.2%)	1395 (19.2%)
Number of previous depressive disorder episodes (n, %)[b]				
No previous episodes	2196 (44.7%)	1011 (42.9%)	3207 (44.1%)	3207 (44.1%)
Single previous episode	2200 (44.8%)	1095 (46.5%)	3295 (45.3%)	3295 (45.3%)
More than 1 previous episode	519 (10.6%)	251 (10.6%)	770 (10.6%)	770 (10.6%)
Anxiety disorder (n, %)[b]				
No	2531 (51.5%)	1259 (53.4%)	3790 (52.1%)	3790 (52.1%)
Yes	2384 (48.5%)	1098 (46.6%)	3482 (47.9%)	3482 (47.9%)
Adjustment disorder (n, %)[b]				
No	3203 (65.2%)	1575 (66.8%)	4778 (65.7%)	4778 (65.7%)
Yes	1712 (34.8%)	782 (33.2%)	2494 (34.3%)	2494 (34.3%)
Personality disorders (n, %)[b]				
No	4530 (92.2%)	2101 (89.1%)	6631 (91.2%)	6631 (91.2%)
Yes	385 (7.8%)	256 (10.9%)	641 (8.8%)	641 (8.8%)
History of suicide or self-harm (n, %)[b]				
No	4867 (99.0%)	2335 (99.1%)	7202 (99.0%)	7202 (99.0%)
Yes	48 (1.0%)	22 (0.9%)	70 (1.0%)	70 (1.0%)
Postpartum depression (PPD) (n, %)[b]				
No	4894 (99.6%)	2343 (99.4%)	7237 (99.5%)	7237 (99.5%)
Yes	21 (0.4%)	14 (0.6%)	35 (0.5%)	35 (0.5%)
Sleep disorders (n, %)[b]				
No	4902 (99.7%)	2353 (99.8%)	7255 (99.8%)	7255 (99.8%)
Yes	13 (0.3%)	4 (0.2%)	17 (0.2%)	17 (0.2%)
PTSD (n, %)[b]				
No	4450 (90.5%)	2196 (93.2%)	6646 (91.4%)	6646 (91.4%)
Yes	465 (9.5%)	161 (6.8%)	626 (8.6%)	626 (8.6%)
Fibromyalgia (n, %)[b]				
No	4558 (92.7%)	2238 (95.0%)	6796 (93.5%)	6796 (93.5%)
Yes	357 (7.3%)	119 (5.0%)	476 (6.5%)	476 (6.5%)
Hyperkinesis (n, %)[b]				
No	4495 (91.5%)	2123 (90.1%)	6618 (91.0%)	6618 (91.0%)
Yes	420 (8.5%)	234 (9.9%)	654 (9.0%)	654 (9.0%)
Body mass index (BMI) (n, %)				
Normal	1510 (30.7%)	806 (34.2%)	2316 (31.8%)	2316 (31.8%)
Overweight	1322 (26.9%)	571 (24.2%)	1893 (26.0%)	1893 (26.0%)
Obesity	1104 (22.5%)	444 (18.8%)	1548 (21.3%)	1548 (21.3%)
Underweight	163 (3.3%)	85 (3.6%)	248 (3.4%)	248 (3.4%)
Unknown	816 (16.6%)	451 (19.1%)	1267 (17.4%)	1267 (17.4%)
Smoking status (n, %)				
Never	2218 (45.1%)	1027 (43.6%)	3245 (44.6%)	3245 (44.6%)
Current/Past	1366 (27.8%)	644 (27.3%)	2010 (27.6%)	2010 (27.6%)
Unknown	1331 (27.1%)	686 (29.1%)	2017 (27.7%)	2017 (27.7%)
Psychiatric hospitalization at baseline (n, %)				
No	4767 (97.0%)	2302 (97.7%)	7069 (97.2%)	7069 (97.2%)
Yes	148 (3.0%)	55 (2.3%)	203 (2.8%)	203 (2.8%)
Substance abuse (n, %)[b]				
No	4820 (98.1%)	2307 (97.9%)	7127 (98.0%)	7127 (98.0%)
Yes	95 (1.9%)	50 (2.1%)	145 (2.0%)	145 (2.0%)
Alcohol abuse (n, %)[b]				
No	4844 (98.6%)	2328 (98.8%)	7172 (98.6%)	7172 (98.6%)
Yes	71 (1.4%)	29 (1.2%)	100 (1.4%)	100 (1.4%)
Nursing home residents (n, %)				
No	4826 (98.2%)	2319 (98.4%)	7145 (98.3%)	7145 (98.3%)
Yes	89 (1.8%)	38 (1.6%)	127 (1.7%)	127 (1.7%)
Visits to family medicine/GP in last 2 years				
0 visits	166 (3.4%)	50 (2.1%)	216 (3.0%)	216 (3.0%)
Up to 20 visits	2466 (50.2%)	1305 (55.4%)	3771 (51.9%)	3771 (51.9%)
More than 20 visits	2283 (46.4%)	1002 (42.5%)	3285 (45.2%)	3285 (45.2%)
Visits to psychiatrist in last 2 years				
0 visits	2641 (53.7%)	1139 (48.3%)	3780 (52.0%)	3780 (52.0%)
>0 visits	2274 (46.3%)	1218 (51.7%)	3492 (48.0%)	3492 (48.0%)
Visits to psychologist in last 2 years				
0 visits	4407 (89.7%)	2100 (89.1%)	6507 (89.5%)	6507 (89.5%)
>0 visits	508 (10.3%)	257 (10.9%)	765 (10.5%)	765 (10.5%)
Calendar year (n, %)				
2015	572 (11.6%)	271 (11.5%)	843 (11.6%)	843 (11.6%)
2016	587 (11.9%)	324 (13.7%)	911 (12.5%)	911 (12.5%)
2017	647 (13.2%)	292 (12.4%)	939 (12.9%)	939 (12.9%)
2018	725 (14.8%)	343 (14.6%)	1068 (14.7%)	1068 (14.7%)
2019	773 (15.7%)	376 (16.0%)	1149 (15.8%)	1149 (15.8%)
2020	734 (14.9%)	353 (15.0%)	1087 (14.9%)	1087 (14.9%)
2021	877 (17.8%)	398 (16.9%)	1275 (17.5%)	1275 (17.5%)
Time since the last antidepressant treatment				
More than 2 years	1318 (26.8%)	574 (24.4%)	1892 (26.0%)	1892 (26.0%)
Between 1 and 2 years	171 (3.5%)	91 (3.9%)	262 (3.6%)	262 (3.6%)
Up to 1 year	3426 (69.7%)	1692 (71.8%)	5118 (70.4%)	5118 (70.4%)
Total previous antidepressant treatment days[e]				
Mean (SD)	1179 (1760)	1320 (1920)	1224.4 (1814.6)	1224.4 (1814.3)
Median [Min, Max]	344 [0, 9405]	412 [0, 9376]	360 [0, 9400]	359.5 [0, 9400.4]
Previous use of antidepressants				
SSRI	3542 (72.1%)	1761 (74.7%)	5303 (72.9%)	5303 (72.9%)
SNRI	1795 (36.5%)	860 (36.5%)	2655 (36.5%)	2655 (36.5%)
NASSA	1011 (20.6%)	521 (22.1%)	1532 (21.1%)	1532 (21.1%)
NDRI	564 (11.5%)	342 (14.5%)	906 (12.5%)	906 (12.5%)
Vortioxetine	244 (5.0%)	171 (7.3%)	415 (5.7%)	415 (5.7%)
TCA	1088 (22.1%)	539 (22.9%)	1627 (22.4%)	1627 (22.4%)
SARI	283 (5.8%)	163 (6.9%)	446 (6.1%)	446 (6.1%)
NaRI	150 (3.1%)	98 (4.2%)	248 (3.4%)	248 (3.4%)
Add-on medications	958 (19.5%)	542 (23.0%)	1500 (20.6%)	1500 (20.6%)
Treatment lag (in days) from diagnosis to start of therapy				
Mean (SD)	10.1 (21.8)	9.72 (20.8)	10.0 (21.5)	10.0 (21.5)
Median [Min, Max]	1 [0, 180]	1 [0, 171]	1 [0, 180]	2 [0, 180]

[a]Based on the Maccabi Healthcare Services (MHS) registry.
[b]Based on diagnoses recorded up to 3 years prior to the onset of the first MDD episode.
[c]Based on diagnoses recorded up 1 year prior to the onset of the first MDD episode.
[d]In the 6 months prior to MDD onset.
[e]Cumulative antidepressant exposure was calculated per patient by summing the number of prescribed days for each individual medication.
Concurrent prescriptions (ie, medications dispensed on the same date) contributed additively to the total, such that overlapping periods across different substances were not collapsed.
[f]SSRI, selective serotonin reuptake inhibitors; SNRI, serotonin and norepinephrine reuptake inhibitors; NASSA, noradrenergic and specific serotonergic antidepressants; NDRI, norepinephrine and dopamine reuptake inhibitors; TCA, tricyclic antidepressants; SARI, serotonin antagonist and reuptake inhibitors; NaRI, selective noradrenergic reuptake inhibitors; MDD, major depressive disorder; PTSD, Post-traumatic stress disorder.
Add-on medications include olanzapine, queiapine, risperdone, aripiprazole, and lithium.

Table 2. Comparison of risk factor coefficients for treatment-resistant depression (TRD) using logistic regression: centralized vs federated analysis (with features selected using federated backward elimination approach)

Federated and centralized models both selected 24 predictors in independent feature selection processes, with 23 of them overlapping. Federated logistic model showed high concordance in coefficient estimates with centralized model, demonstrated by absolute mean difference, decision agreement, estimate agreement, and CI overlap. Similar high concordance was seen in the naïve federated approach (FA) and base scenario (data not shown).

	Metrics for assessing the level of agreement between aggregated federated analysis and centralized analysis							
	Outside Greater Tel Aviv area (Site A)	Greater Tel Aviv area (Site B)	Total population (centralized)	Federated analysis	Abs. mean difference	Decision agreement[a]	Estimate agreement[b]	Overlap[c]
Atrial fibrillation	0.43 (0.24, 0.74)*	0.58 (0.29, 1.08)	0.5 (0.32, 0.75)*	0.52 (0.34, 0.79)*	0.02	Yes	Yes	95.7%
Anxiety disorder	1.26 (1.1, 1.44)*	1.36 (1.13, 1.65)*	1.28 (1.14, 1.43)*	1.27 (1.14, 1.42)*	0.01	Yes	Yes	100.0%
Personality disorder	1.6 (1.27, 2.02)*		1.35 (1.12, 1.63)*	1.35 (1.12, 1.63)*	0.00	Yes	Yes	100.0%
Polypharmacy: log of no. of therapeutic groups	1.1 (1.04, 1.17)*	1.11 (1.02, 1.21)*	1.1 (1.04, 1.16)*	1.1 (1.04, 1.16)*	0.00	Yes	Yes	100.0%
Time since the last treatment (Reference more than 2 years)								
Between 1 and 2 years	0.92 (0.58, 1.43)	1.32 (0.73, 2.33)	1.12 (0.78, 1.58)	1.11 (0.78, 1.58)	0.01	Yes	Yes	100.0%
Up to 1 year	1.53 (1.19, 1.97)*	1.49 (1.03, 2.16)*	1.55 (1.26, 1.92)*	1.56 (1.26, 1.93)*	0.01	Yes	Yes	100.0%
Total previous antidepressant treatment days (log days)	0.96 (0.92, 1)	0.94 (0.89, 1)	0.95 (0.92, 0.98)*	0.95 (0.92, 0.98)*	0.00	Yes		100.0%
Previous use of antidepressants: NASSA	1.15 (0.97, 1.37)		1.15 (1, 1.33)*	1.15 (1, 1.32)*	0.00	Yes	Yes	100.0%
Previous use of antidepressants: SNRI	1.29 (1.11, 1.52)*	1.28 (1.02, 1.6)*	1.28 (1.12, 1.45)*	1.27 (1.12, 1.45)*	0.01	Yes	Yes	100.0%
Previous use of antidepressants: Vortioxetine		1.6 (1.13, 2.27)*	1.27 (1.01, 1.58)*	1.28 (1.03, 1.59)*	0.01	Yes	Yes	96.6%
Previous use of antidepressants: NaRI	1.34 (0.94, 1.92)		1.34 (1.01, 1.77)*	1.33 (1, 1.76)*	0.01	Yes	Yes	98.7%
Previous use of antidepressants: NDRI	1.3 (1.06, 1.59)*		1.18 (1, 1.39)*	1.19 (1.01, 1.4)*	0.01	Yes	Yes	97.5%
Previous use of antidepressants: Add-on medications	1.31 (1.1, 1.56)	1.78 (1.41, 2.24)*	1.46 (1.27, 1.68)*	1.46 (1.27, 1.68)*	0.00	Yes	Yes	100.0%
Previous use of antidepressants: SARI		1.64 (1.15, 2.34)*	1.19 (0.95, 1.47)	1.18 (0.95, 1.47)	0.01	Yes	Yes	100.0%
Previous use of antidepressants: TCA	1.34 (1.14, 1.58)*	1.46 (1.16, 1.84)*	1.35 (1.18, 1.55)*	1.35 (1.18, 1.54)*	0.00	Yes	Yes	100.0%
Treatment lag (in 30-day units) between diagnosis and therapy initiation	0.76 (0.68, 0.85)*	0.81 (0.69, 0.94)*	0.78 (0.71, 0.85)*	0.77 (0.71, 0.85)*	0.01	Yes	Yes	100.0%
Smoking status (Reference nonsmokers)								
Current or past	1.11 (0.95, 1.3)		1.11 (0.97, 1.26)	1.1 (0.96, 1.25)	0.01	Yes	Yes	96.6%
Unknown	0.88 (0.74, 1.03)		0.89 (0.78, 1.02)	0.88 (0.77, 1.01)	0.01	Yes	Yes	95.8%
Nursing home residents	1.45 (0.91, 2.29)		1.4 (0.95, 2.05)	1.42 (0.97, 2.09)	0.02	Yes	Yes	98.2%
Psychiatric hospitalization at baseline	1.48 (1.03, 2.12)*		1.4 (1.03, 1.9)*	1.41 (1.03, 1.91)*	0.01	Yes	Yes	100.0%
Visits to family medicine/GP in last 2 years (Reference 0 visits)								
Up to 20 visits			0.79 (0.57, 1.13)	0.8 (0.57, 1.12)	0.01	Yes	Yes	100.0%
More than 20 visits			0.89 (0.63, 1.27)	0.9 (0.63, 1.28)	0.01	Yes	Yes	100.0%
Dementia			0.75 (0.54, 1.03)	0.77 (0.55, 1.09)	0.02	Yes	Yes	98.2%
Substance abuse	0.67 (0.4, 1.08)		0.73 (0.49, 1.08)	0.73 (0.49, 1.08)	0.00	Yes	Yes	100.0%
Visits to psychiatrist in last 2 years		0.8 (0.65, 0.99)*	0.91 (0.8, 1.03)	0.9 (0.8, 1.02)	0.01	Yes	Yes	100.0%
Adjustment disorder	1.11 (0.96, 1.27)		1.09 (0.97, 1.23)	1.09 (0.97, 1.23)	0.00	Yes	Yes	100.0%
Socioeconomic status (Reference middle SES)								
Low	0.75 (0.59, 0.94)*		0.79 (0.64, 0.98)*		—	—	—	—
High	0.93 (0.8, 1.08)		0.97 (0.86, 1.09)		—	—	—	—
Age at first MDD episode (Reference 18, 25)								
(25, 35)		0.79 (0.53, 1.17)		0.83 (0.65, 1.04)	—	—	—	—
(35, 45)		0.65 (0.43, 0.98)*		0.92 (0.73, 1.16)	—	—	—	—
(45, 55)		1.09 (0.74, 1.61)		1.09 (0.87, 1.37)	—	—	—	—
(55, 65)		0.98 (0.65, 1.49)		0.93 (0.73, 1.17)	—	—	—	—
(65, 75)		0.9 (0.58, 1.41)		0.87 (0.66, 1.13)	—	—	—	—
≥75		1.1 (0.67, 1.82)		0.9 (0.66, 1.22)	—	—	—	—
Sex (male)	1.12 (0.97, 1.29)				—	—	—	—
Cerebrovascular accident	0.51 (0.25, 0.98)*				—	—	—	—
Corticosteroids in the last 3 m.	1.32 (0.95, 1.83)	0.71 (0.42, 1.15)			—	—	—	—
PTSD		1.4 (0.98, 2)			—	—	—	—
Calendar year of MDD onset		0.96 (0.91, 1.01)			—	—	—	—
Thyroiditis		2.41 (0.71, 8.1)			—	—	—	—

[a]Decision agreement assesses whether 2 effect estimates lead to the same conclusion. If the original estimate is significant, decision agreement occurs when the federated estimate is also significant and on the same side of the null. If the original estimate is nonsignificant, decision agreement occurs when the federated estimate is also nonsignificant.
[b]Estimate agreement checks whether the federated estimate aligns with the original study's confidence interval. Agreement is achieved if the federated point estimate falls within the original study's confidence interval.
[c]Confidence interval overlap measures the extent to which the confidence intervals of the original study estimate and the federated data estimate overlap, expressed as an average percentage.

Table 3. Performance of TRD risk factor logistic regression models after feature selection

Although the model performed differently in the two sites, the two federated approaches showed highly similar predictive performances compared to the centralized model across different metrics.

	Outside Greater Tel Aviv area (Site A) (N=4915)	Greater Tel Aviv area (Site B) (N=2357)	Centralized approach (N=7272)	True FA approach (N=7272)	Naïve FA approach (N=7272)
Hosmer-Lemeshow Test (Chi-square, statistic, *P* value)	12.05, 0.15	13.37, 0.1	10.26, 0.25	10.2, 0.25	10.7, 0.22
C statistic (%)	65.1%	65.90%	64.6%	64.6%	64.7%
Brier score	0.18	0.19	0.18	0.18	0.18
Sensitivity[a]	57%	66%	57%	59%	58%
Specificity[a]	66%	59%	64%	63%	63%
PPV[a]	36%	39%	36%	36%	36%

[a]Estimated at the threshold of 26%, which maximizes Youden's Index (J statistic) for the full model applied to the total population.
PPV, Positive Predictive Value.

Table 4. Algorithmic bias (measured as difference in false-negative rate [FNR] between subgroups) in logistic regression model for different sites and scenarios

Overall, Arab and low-SES groups are more likely to be underdiagnosed (positive and greater FNR difference in the table), but it differs by each site. Naïve federated approach seems to retain the algorithmic biases. True federated approach reduced algorithmic bias, while it could also change the algorithmic bias dramatically, including direction, indicating variability at subgroup level.

		False negative rate difference (compared to Reference group)				
	Reference group	A: Outside Greater Tel Aviv area	B: Greater Tel Aviv area	Centralized approach	True FA approach	Naïve FA approach
Orthodox Jews	Geneneral population	4.5%	2.0%	2.4%	–1.7%	3.9%
Arabs	Geneneral population	10.3%	7.7%	11.2%	2.0%	10.8%
High SES	Middle SES	7.6%	1.9%	3.8%	1.0%	3.6%
Low SES	Middle SES	20.3%	8.8%	13.3%	–0.9%	16.4%

SES, socioeconomic status.

Conclusions

- This is one of the first studies that applied federated analysis to statistical inferences and conducted comprehensive evaluations against centralized data
- Federated GLM and Cox modeling achieved high concordance with centralized analysis, while caution should be applied when analyzing subgroups
- Presented a repeatable end-to-end process for federated analysis, regardless of platform
- Improved confidence in conducting RWE studies on federated data networks, encouraging applications to leverage the existing networks

References

1. Sharman Moser S, et al. *BMC Psychiatry.* 2022;22(1):541.
2. Lundberg J, et al. *JAMA Psychiatry.* 2023;80(2):167-175.
3. Gronemann FH, et al. *J Affect Disord.* 2020;261:221-229.
4. Cellamare M, et al. *Algorithms.* 2022;15(7):243.
5. Lu CL, et al. *J Am Med Inform Assoc.* 2015;22(6):1212-1219.
6. Wang HE, et al. *J Biomed Inform.* 2024;156:104683.

Presented at AMIA 2025 Annual Symposium; Atlanta, GA; November 15-19, 2025.