Supplemental Material: Revisiting Performance Metrics for Prediction with Rare Outcomes

Statistical Methods in Medical Research
XX(X):1–??

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

Supplemental Material: Additional Simulation Study Details

We include additional details of the simulation study modeled after the isolated AVR cohort for 30-day mortality outcome. Our data-generation process for the other cohort simulations is consistent with this general description. Variables with observed rate of less than 1% in our data analyses were not included in the simulations. Further details are found in Section 4 of the manuscript and all simulation code is available on our companion GitHub page [blinded for review].

The nine continuous variables described in Section 4 of our manuscript represented: age, height, weight, cross clamp time, perfuse time, hemodialysis ejection fraction, creatine level, body surface area, and body size. The 36 binary covariates represented: sex, government insurance, HMO insurance, commercial insurance, Medicaid, state-specific government insurance, Medicare, self-insurance/no insurance, first surgery, hypertension, family history of coronary arterial disease, chronic lung disease, immunosuppressant, pulmonary valve disease, coronary valve disease, diabetes, previous cardiovascular intervention, endocarditis, treated endocarditis, previous myocardial infarction, hemodialysis ejection fraction done, congestive heart failure, aortic valve insufficiency, tricuspid valve insufficiency, mitral valve insufficiency, pulmonary valve insufficiency, other cardiac procedure, intraoperative blood products used, left main disease, adenosine disphosphate inhibitors, aspirins, beta blockers, inotropes, steroids, lipid lowering drugs, and anticoagulants.

Seven categorical variables (including the valves) with more than two categories were generated from multinomial distributions, as described in Section 4 of the manuscript. The covariates were race/ethnicity (caucasian, Black, hispanic, other), number of diseased veins (0, 1, 2, 3), New York Heart Association class (Class 1, Class 2, Class 3), ejection fraction (<30, 30, 40+), hemodialysis ejection fraction method (left ventricular, echo, other), and surgical urgency (elective, urgent, emergent/salvage), along with the 11 valve groups. Outcomes for simulation settings 1, 2, and 3 were generated from Bern(e_{y^1}), Bern(e_{y^2}), and Bern(e_{y^3}), where $e_{y^{(\cdot)}}$ is the proportion of events, respectively with equations shown below.

$$
\begin{aligned}
\text{logit}(e_{y^1}) = & - 3.75 + 0.45\ \text{Age} + 0.13\ \text{Sex} + 0.22\ \text{Caucasian} + 0.24\ \text{Black} \\
& - 0.29\ \text{Hispanic} + 0.06\ \text{Government Insurance} - 0.13\ \text{Government, medicaid} \\
& + 0.05\ \text{Government, medicare} + 0.10\ \text{Government, state specific} \\
& - 0.21\ \text{Commercial Insurance} - 0.19\ \text{HMO Insurance} \\
& + 0.22\ \text{None/Self Insurance} + 0.00\ \text{Family history of CAD} \\
& + 0.05\ \text{Hypertension} + 0.57\ \text{Chronic lung disease} \\
& + 0.15\ \text{Immunosuppressant} + 0.40\ \text{PV disease} \\
& + 0.38\ \text{Coronary valve disease} + 0.42\ \text{Diabetes} \\
& + 0.43\ \text{Endocarditis} - 0.11\ \text{Treated Endocarditis} \\
& + 0.03\ \text{Previous CV Intervention} - 0.01\ \text{Previous MI} \\
& + 0.60\ \text{Congestive heart failure} - 0.11\ \text{NYHA Class 1} \\
& - 0.38\ \text{NYHA Class 2} - 0.37\ \text{NYHA Class 3} - 1.12\ \text{HDEF done} \\
& + 0.01\ \text{HDEF} + 0.32\ \text{HDEF Method LV} + 0.30\ \text{HDEF Method Echo} \\
& - 0.21\ \text{HDEF Method Other} + 0.46\ \text{EF Category} > 40 \\
& + 0.42\ \text{EF Category } 30 - 0.08\ \#\ \text{diseased vein 1} \\
& - 0.05\ \#\ \text{diseased vein 2} - 0.42\ \#\ \text{diseased vein 3} \\
& - 0.11\ \text{AV insufficiency} - 0.12\ \text{MV insufficiency} + 0.16\ \text{TV insufficiency} \\
& - 0.26\ \text{PV insufficiency} + 0.10\ \text{First surgery} - 1.11\ \text{Elective Surgery} \\
& - 0.63\ \text{Urgent Surgery} + 0.26\ \text{ADP Inhibitors} + 0.18\ \text{Anticoagulants} \\
& - 0.23\ \text{Aspirin} + 0.04\ \text{Betablocker} + 0.01\ \text{Inotropes} - 0.62\ \text{Lipid lowering} \\
& + 0.70\ \text{Steroids} - 0.25\ \text{Other Cardiac Procedure} + 0.41\ \text{IBPR} \\
& + 0.22\ \text{Left main disease} - 2.13\ \text{Body surface area} - 0.01\ \text{Body size.} \\
& + 0.79\ \text{Height} + 1.49\ \text{Weight} \\
& + 0.13\ \text{Creatinine} + 0.66\ \text{Perfus time} - 0.42\ \text{Cross clamp time} \\
& - 1.50\ \text{Valve 2} - 1.02\ \text{Valve 4} - 1.51\text{Valve 5} \\
& + 0.00\ \text{Valve 6} - 1.01\ \text{Valve 7} - 0.87\ \text{Valve 9} \\
& + 1.25\ \text{Valve 10} - 1.26\ \text{Valve 11} - 0.93\ \text{Valve 12} \\
& - 1.02\ \text{Valve 13} - 0.84\ \text{Valve 14.}
\end{aligned}
$$

$$
\begin{aligned}
\text{logit}(e_{y^2}) = & - 2.75 + 2.13\ \text{Age} + 0.27\ \text{Sex} + 0.00\ \text{Caucasian} \\
& + 0.19\ \text{Black} + 0.68\ \text{Hispanic} \\
& + 0.48\ \text{Government insurance} - 0.50\ \text{Government, medicaid} \\
& + 0.13\ \text{Government, medicare} - 0.68\ \text{Government, state specific} \\
& + 0.00\ \text{Commercial Insurance} + 0.20\ \text{HMO Insurance} \\
& + 0.53\ \text{None/Self Insurance} + 0.20\ \text{Family history of CAD} \\
& + 0.12\ \text{Hypertension} + 0.00\ \text{Chronic lung disease} \\
& - 0.14\ \text{Immunosuppressant} + 0.35\ \text{PV disease} \\
& + 0.50\ \text{Coronary valve disease} + 0.65\ \text{Diabetes} \\
& + 0.76\ \text{Endocarditis} - 1.28\ \text{Treated Endocarditis} \\
& - 0.37\ \text{Previous CV Intervention} + 0.48\ \text{Previous MI} \\
& + 0.16\ \text{Congestive heart failure} - 0.20\ \text{NYHA Class 1} \\
& - 0.30\text{NYHA Class 2} - 0.34\ \text{NYHA Class 3} \\
& - 0.82\ \text{HDEF done} + 0.00\ \text{HDEF} - 0.13\ \text{HDEF Method LV} \\
& - 0.32\ \text{HDEF Method Echo} - 0.47\ \text{HDEF Method Other} \\
& + 0.68\ \text{EF Category} > 40 + 1.28\ \text{EF Category 30} \\
& + 0.19\ \#\ \text{diseased vein 1} - 0.46\ \#\ \text{diseased vein 2} \\
& - 1.55\ \#\ \text{diseased vein 3} + 0.00\ \text{AV insufficiency} \\
& + 0.32\ \text{MV insufficiency} + 0.00\ \text{TV insufficiency} \\
& - 0.24\ \text{PV insufficiency} + 0.00\ \text{First surgery} - 1.74\ \text{Elective Surgery} \\
& - 1.28\ \text{Urgent Surgery} + 0.53\ \text{ADP Inhibitors} - 0.28\ \text{Anticoagulants} \\
& - 0.26\ \text{Aspirin} + 0.32\ \text{Betablocker} - 0.96\ \text{Inotropes} - 0.58\ \text{Lipid lowering} \\
& + 0.85\ \text{Steroids} - 0.46\ \text{Other Cardiac Procedure} + 0.28\ \text{IBPR} \\
& + 1.12\ \text{Left main disease} - 0.11\ \text{Body surface area} + 0.13\ \text{Body size.} \\
& - 0.24\ \text{Height} + 0.86\ \text{Weight} + 0.13\ \text{Creatinine} \\
& + 1.02\ \text{Perfus time} - 0.57\ \text{Cross clamp time} - 1.48\ \text{Valve 2} \\
& - 1.67\ \text{Valve 4} - 1.81\text{Valve 5} + 0.00\ \text{Valve 6} - 2.34\ \text{Valve 7} \\
& - 1.73\ \text{Valve 9} + 0.75\ \text{Valve 10} - 1.74\ \text{Valve 11} - 1.64\ \text{Valve 12} \\
& - 2.53\ \text{Valve 13} - 1.57\ \text{Valve 14} - 0.39\ \text{Sex} \times \text{Diabetes} - 1.57\ \text{Age}^2 \\
& - 0.81\ \text{Weight} \times \text{Hypertension} + 0.45\ \text{Age} \times \text{Congestive heart failure.}
\end{aligned}
$$

$$
\begin{aligned}
\text{logit}(e_{y^3}) = & - 4.25 - 0.25 \text{ Previous CV Intervention} \\
& + 0.50 \text{ Previous MI} + 0.44 \text{ Congestive heart failure} \\
& - 0.11 \text{ NYHA Class 1} - 0.28 \text{NYHA Class 2} \\
& - 0.28 \text{ NYHA Class 3} - 1.15 \text{ HDEF done} \\
& + 0.11 \text{ HDEF} + 0.05 \text{ HDEF Method LV} \\
& - 0.14 \text{ HDEF Method Echo} - 0.05 \text{ HDEF Method Other} \\
& + 0.73 \text{ EF Category} > 40 + 1.38 \text{ EF Category 30} \\
& + 0.37 \text{ \# diseased vein 1} - 0.11 \text{ \# diseased vein 2} \\
& - 1.18 \text{ \# diseased vein 3} - 0.01 \text{ AV insufficiency} \\
& + 0.45 \text{ MV insufficiency} - 0.11 \text{ TV insufficiency} \\
& - 0.24 \text{ PV insufficiency} - 0.04 \text{ First surgery} \\
& - 1.63 \text{ Elective Surgery} - 1.18 \text{ Urgent Surgery} \\
& + 0.62 \text{ ADP Inhibitors} - 0.08 \text{ Anticoagulants} \\
& - 0.16 \text{ Aspirin} + 0.45 \text{ Betablocker} - 0.81 \text{ Inotropes} \\
& - 0.46 \text{ Lipid lowering} + 0.83 \text{ Steroids} \\
& - 0.45 \text{ Other Cardiac Procedure} + 0.40 \text{ IBPR} \\
& + 1.15 \text{ Left main disease} + 1.43 \text{ Body surface area} \\
& + 0.31 \text{ Body size} - 0.87 \text{ Height} - 0.87 \text{ Weight} \\
& + 0.13 \text{ Creatinine} + 1.01 \text{ Perfus time} - 0.60 \text{ Cross clamp time} \\
& - 0.00 \text{ Valve1} - 1.50 \text{ Valve 2} - 1.57 \text{ Valve 4} \\
& - 1.82 \text{ Valve 5} - 1.23 \text{ Valve 6} - 1.58 \text{ Valve 7} \\
& - 0.99 \text{ Valve 9} + 1.25 \text{ Valve 10} - 1.00 \text{ Valve 11} \\
& - 0.88 \text{ Valve 12} - 1.86 \text{Valve 13} + 0.25 \text{ Valve 14.}
\end{aligned}
$$

Supplemental Material: Additional Tables & Figures

Predictors	Isolated AVR	AVR or AVR & CABG	AVR or AVR & MVR	Any AVR
Age (mean, years)	68	70	68	71
Male (%)	58	63	58	62
Race (%)				
Caucasian	92	93	91	93
Black	2	2	2	2
Hispanic	3	2	3	2
Other	3	4	3	4
Latino (%)	3	3	3	3
Body surface area (mean, m^2)	2	2	2	2
Body Size (mean, cm/kg)	2	2	2	2
Height (mean, cm)	169	169	169	169
Weight (mean, kg)	84	83	84	83
Creatinine (mean, mg/dL)	1	1	1	1
Perfus Time (mean, min)	110	130	115	132
Cross Clamp Time (mean, min)	80	98	84	100
Government Insurance (%)	63	66	63	66
Commercial Insurance (%)	42	40	41	40
HMO Insurance (%)	18	16	17	16
None/Self Insurance (%)	2	2	2	2
Government payor (%)				
Military	1	1	1	1
State specific plan	5	4	5	4
Medicare	50	55	50	55
Medicaid	7	6	7	6
None	37	34	37	34
Medicare Fee-for-Service (%)	13	13	13	12
Hospital ID (%)				
A	15	17	15	14
B	10	10	10	10
C	7	7	7	7
D	16	14	16	14
E	2	2	2	2
F	2	3	2	3
G	6	6	6	6
H	13	13	13	13
I	3	3	3	3
J	7	7	7	7
K	3	3	3	3
L	5	5	5	5
M	4	4	4	4
N	7	6	7	6

Table 1. Baseline Covariates. *Features observed at baseline for each cohort. HMO: health maintenance organization*

Predictors (%)	Isolated AVR	AVR or AVR & CABG	AVR or AVR & MVR	Any AVR
ADP Inhibitors				
Yes	1	2	1	2
Contraindicated	2	1	2	1
Anticoagulants				
Yes	12	16	13	17
Contraindicated	2	1	2	1
Aspirins				
Yes	49	57	49	57
Contraindicated	1	1	1	1
Beta blockers				
Yes	48	56	48	55
Contraindicated	6	5	6	5
Inotropes				
Yes	1	1	1	1
Contraindicated	2	1	2	1
Steroids				
Yes	3	4	3	4
Contraindicated	2	1	2	1
Coumadin	1	1	1	2
Lipid Lowering	41	46	41	45
Intravenous Nitrates	1	2	1	2

Table 2. Medication-Related Baseline Covariates. *Medication used at baseline for each cohort.*

Predictors	Isolated AVR	AVR or AVR & CABG	AVR or AVR & MVR	Any AVR
Family History CAD (%)	17	20	17	20
Hypertension (%)	73	78	73	78
Chronic Lung Disease (%)	17	18	17	18
Immunosuppressant (%)	4	4	4	4
Pulmonary Valve Disease (%)	7	12	7	12
Coronary Valve Disease (%)	13	15	13	15
Diabetes (%)	26	31	26	30
Endocarditis (%)	6	4	7	4
Treated Endocarditis (%)	3	2	3	2
Previous CV Intervention (%)	23	25	23	25
Previous MI (%)	11	19	10	19
Previous MI (within 7 days, %)	11	19	10	19
CHF (%)	36	38	38	38
NYHA Class (%)				
1	5	4	5	4
2	20	19	20	19
3	26	29	27	29
4	49	48	48	48
Cardiogenic Shock (%)	1	1	1	1
Other Cardiac Procedure (%)	6	4	6	4
IBPR (%)	32	37	33	38
Left Main Disease (%)	2	9	2	9
HDEF Done (%)	97	97	97	97
HDEF Method (%)				
Left ventiricular	19	23	19	23
Echo	72	68	72	68
Other	3	4	3	4
EF Category (%)				
<30	7	7	7	8
30	5	7	5	6
40+	88	86	88	86
HDEF (mean)	55	54	55	54
# of Diseased Veins				
0	80	45	80	48
1	10	18	10	17
2	4	16	4	15
3	6	21	6	20
Aortic Valve Insufficiency (%)	69	67	70	67
Mitral Valve Insufficiency (%)	75	74	76	74
Tricuspid Valve Insufficiency (%)	67	65	67	65
Pulmonary Valve Insufficiency (%)	46	43	46	43
First surgery (%)	72	71	71	71
Surgical urgency (%)				
Elective	77	69	76	69
Urgent	22	30	23	30
Emergent or salvage	1	1	1	1

Table 3. Comorbidity-Related Baseline Covariates. *Comorbidities observed at baseline for each cohort. CAD: coronary arterial disease; CV: cardiovascular; MI: myocardial infarction; CHF: congestive heart failure; NYHA: New York Heart Association; IBPR: intraoperative blood products refused; HDEF: hemo data-ejection fraction; EF: Ejection fraction; 'Other Cardiac Procedure' refers to cardiac procedures other than coronary artery bypass grafting (CABG) or valve procedures.*

Prepared using sagej.cls

	Valve Type, % (n)			
	Isolated AVR	AVR or AVR & CABG	AVR or AVR & MVR	Any AVR
Mechanical				
Group 1	0.5 (34)	0.4 (49)	0.6 (42)	0.5 (57)
Group 2	1.0 (67)	0.8 (87)	1.2 (85)	0.9 (105)
Group 3	0.4 (27)	0.3 (32)	0.4 (27)	0.3 (32)
Group 4	11 (685)	8.0 (926)	11 (779)	8.6 (1020)
Group 5	1.7 (107)	1.3 (151)	1.8 (126)	1.4 (170)
Group 6	3.8 (248)	2.9 (332)	3.9 (269)	2.9 (353)
Bioprosthetic				
Group 7	5.6 (361)	5.7 (660)	5.5 (376)	5.7 (675)
Group 8	*	*	*	*
Group 9	4.6 (299)	4.4 (509)	4.7 (312)	4.4 (522)
Group 10	7.8 (505)	8.2 (940)	7.8 (531)	8.1 (966)
Group 11	2.3 (149)	3.0 (350)	2.3 (156)	3.0 (357)
Group 12	36 (2308)	40 (4660)	35 (2385)	40 (4737)
Group 13	5.9 (381)	5.9 (682)	5.9 (402)	5.8 (703)
Group 14	20 (1304)	18 (2127)	20 (1339)	18 (2162)

Table 4. Percentage of Types of Valves in Each Cohort. *These valves are grouped by manufacturer and generation specific subtypes . Cells with $<$ 10 events were suppressed and replaced with *.*

Cohort	n	30 Day (%)	1 Year (%)
AVR or AVR & CABG	11502	2.4	6.9
AVR or AVR & MVR	6824	1.8	5.7
Any AVR	11854	2.4	7.1

Table 5. 30-Day and 1-Year Mortality Rates for Three Cohorts.

	Valve Type (%)			
	30 Day		1 Year	
	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
Mechanical				
Group 1	1.7	0.5	1.1	0.5
Group 2	0.9	1.0	0.6	1.1
Group 3	0.9	0.4	0.3	0.4
Group 4	7.8	1.1	7.4	11
Group 5	0.9	1.6	0.9	1.7
Group 6	3.4	3.8	3.1	3.9
Bioprosthetic				
Group 7	5.2	5.5	7.1	5.5
Group 8	*	*	*	*
Group 9	4.3	4.6	4.3	4.6
Group 10	6.1	7.8	5.7	7.9
Group 11	2.6	2.3	2.0	2.3
Group 12	43	35	44	35
Group 13	1.7	5.9	4.6	5.9
Group 14	21	20	19	20

Table 6. Percentage of Valves by Mortality Outcome in Isolated AVR Cohort. *Cells with $<$ 10 events were suppressed and replaced with *.*

	Algorithm	Hyperparameters	Tuning method
1.	Logistic regression		
	i. without penalty		
	ii. with Firth's correction		
	iii. with lasso penalty	variable-wise sparsity	CV
	iv. with TS lasso penalty	variable-wise sparsity	CV
	v. with group lasso penalty	group-wise sparsity	CV
	vi. with SG lasso penalty,		
	a. group sparsity = 0.15	group-wise sparsity	CV
	b. group sparsity = 0.50	group-wise sparsity	CV
	c. group sparsity = 0.85	group-wise sparsity	CV
2.	Random forest		
	i. node size = 1	# predictors in tree	
		size of the tree	OOB
	ii. node size = 50	# predictors in tree	
		size of the tree	OOB
	iii. node size = 100	# predictors in tree	
		size of the tree	OOB
3.	Gradient boosted trees		
	i. step size shrinkage = 0.3		
	maximum tree depth = 6		
	ii. step size shrinkage = 0.7		
	maximum tree depth = 6		
	iii. step size shrinkage= 0.3		
	maximum tree depth = 15		
	iv. step size shrinkage = 0.7		
	maximum tree depth = 15		
4.	BART		
	i. number of trees = 50		
	base = 0.95		
	power = 2		
	k = 2		
	quantile of the prior = 0.9		
5.	Neural networks		
	i. # units in hidden layer = 1		
	ii. # unit in hidden layer = 3		
	iii. # units hidden layer = 4		
6.	SVM (radial kernel)	cost parameter	CV

Table 7. Hyperparameters and Related Tuning Methods for Algorithms in the Extended Ensemble. *SG is sparse group, TS is treatment-specific, SVM is support vector machine, CV is cross-validation, OOB is out of bag, and for BART, k determines the prior probability that $E(Y|X)$ is between (-3, 3).*

Simulation setting	Data generation	Predictors for fitting
1	$\text{logit}(e_{y^1}) = X_{\text{sim}}\beta$	X_{sim}
2	$\text{logit}(e_{y^2}) = X_{\text{sim}}\beta + X_{\text{sim},I}\beta_I$	X_{sim}
3	$\text{logit}(e_{y^3}) = X_{\text{sim},1}\beta_1 + X_{\text{sim},2}\beta_2$	$X_{\text{sim},1}$ and $X_{\text{sim},3}$

Table 8. Data Generation and Predictors Under Different Simulation Settings. X *denotes full set of covariates;* X_{sim} *is the union of* $X_{sim,1}$, $X_{sim,2}$, *and* $X_{sim,3}$. $X_{sim,I}$ *includes interactions between selected variables from* X_{sim}.

	Mortality Rate (%)					
	Setting 1		Setting 2		Setting 3	
Cohort	30 Day	1 Year	30 Day	1 Year	30 Day	1 Year
Isolated AVR	1.9	5.7	1.7	5.7	1.9	5.9
AVR or AVR & MVR	2.0	5.9	2.0	5.8	2.0	6.2
AVR or AVR & CABG	2.6	6.6	2.9	7.2	2.9	7.3
Any AVR	2.7	7.3	2.4	7.0	2.9	7.7

Table 9. Mortality Rates in Simulated Data.

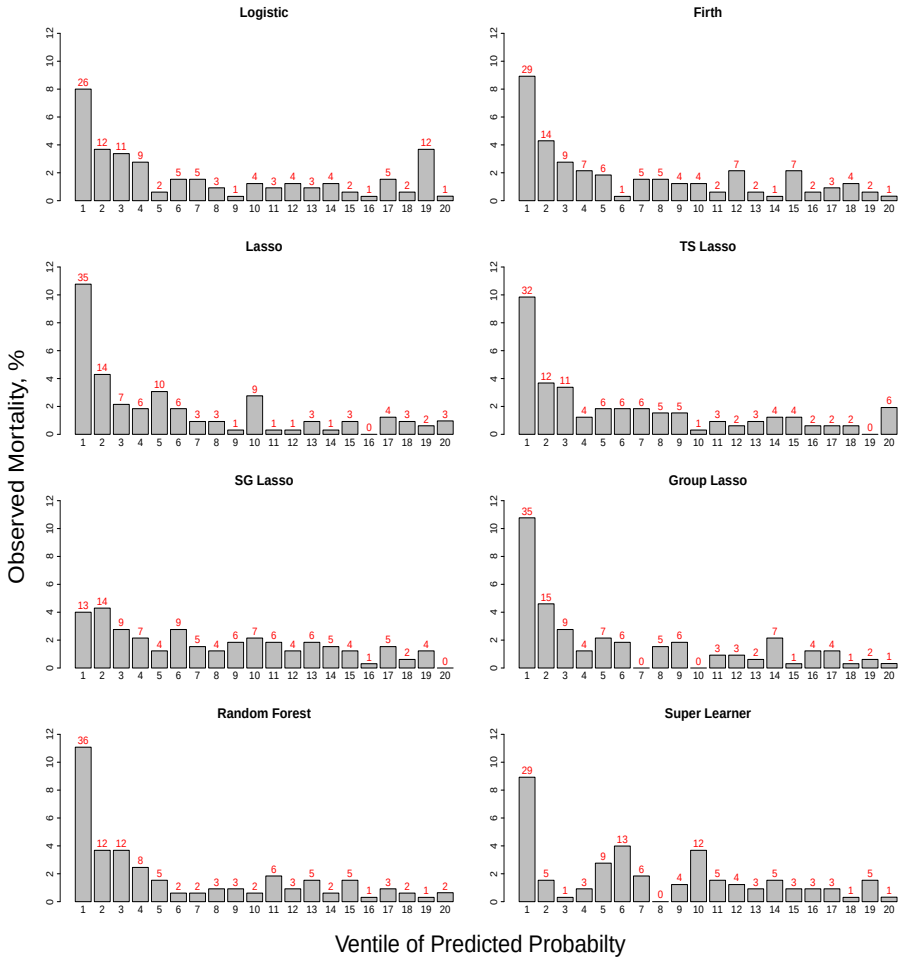

Figure 1. Data Analysis: Rate of Observed 30-Day Mortality within each Ventile of Predicted Mortality Risk for Different Algorithms in Isolated AVR Cohort. *The predicted mortality risks are in decreasing order and red values are the number of events in each ventile. TS is an abbreviation for treatment-specific and SG is sparse group.*

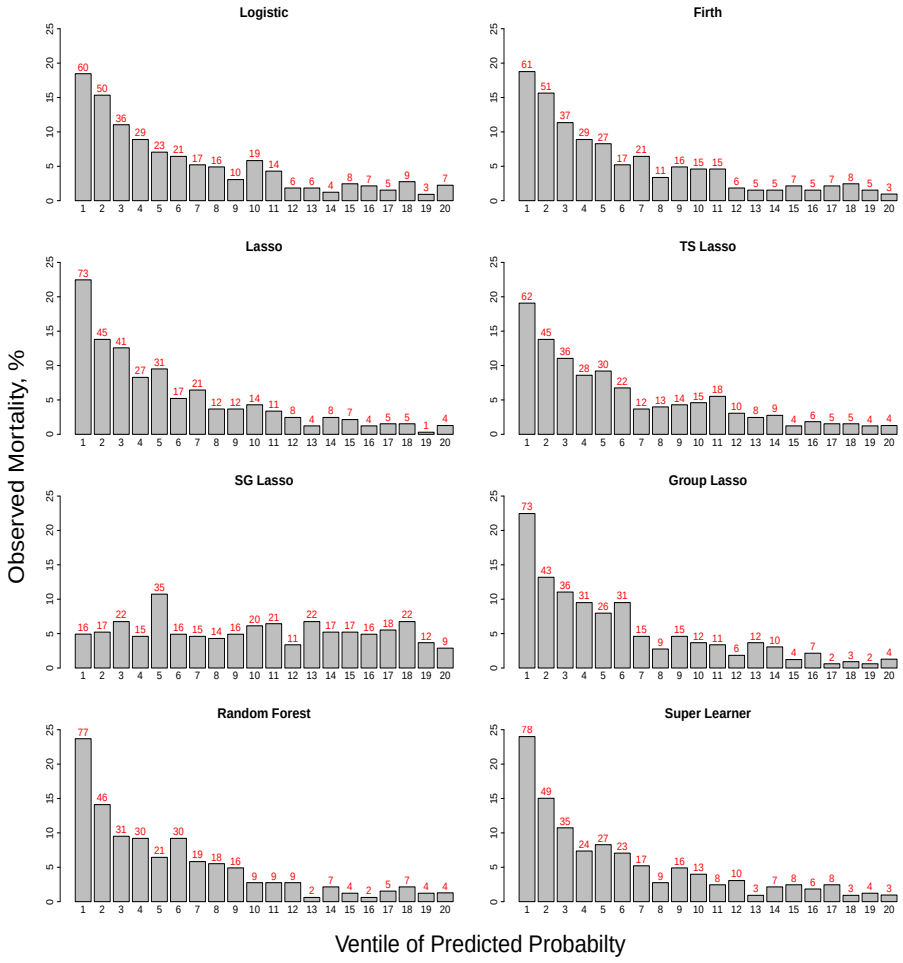

Figure 2. Data Analysis: Rate of Observed 1-Year Mortality within each Ventile of Predicted Mortality Risk for Different Algorithms in Isolated AVR Cohort. *The predicted mortality risks are in decreasing order and red values are the number of events in each ventile. TS is an abbreviation for treatment-specific and SG is sparse group.*

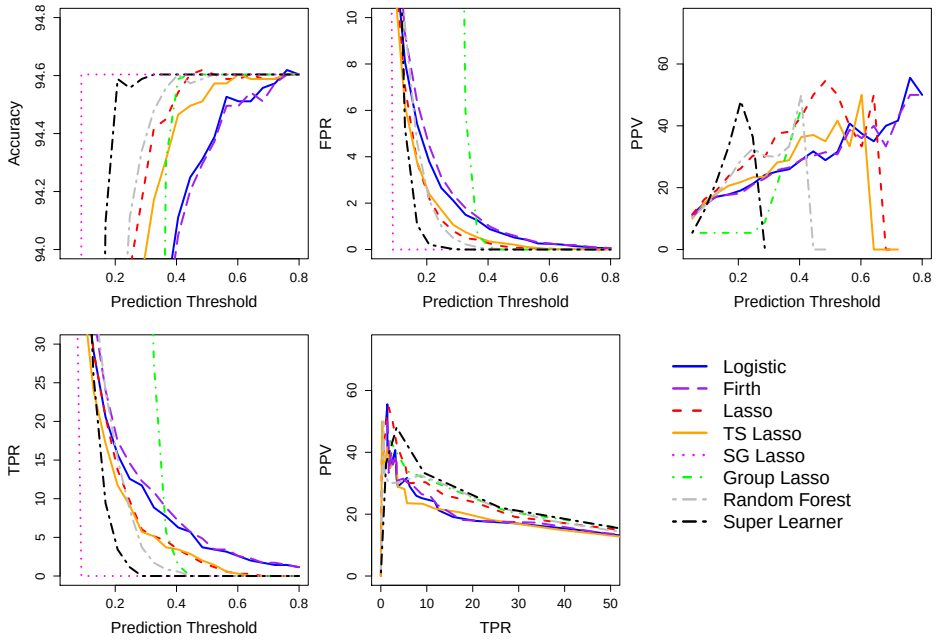

Figure 3. Data Analysis: Cross-Validated Algorithm Performance by Prediction Threshold and Precision-Recall Plot for 1-Year Mortality in Isolated AVR Cohort. *For algorithms with TPR equal to zero, PPV is undefined and not plotted. TS is an abbreviation for treatment-specific and SG is sparse group.*

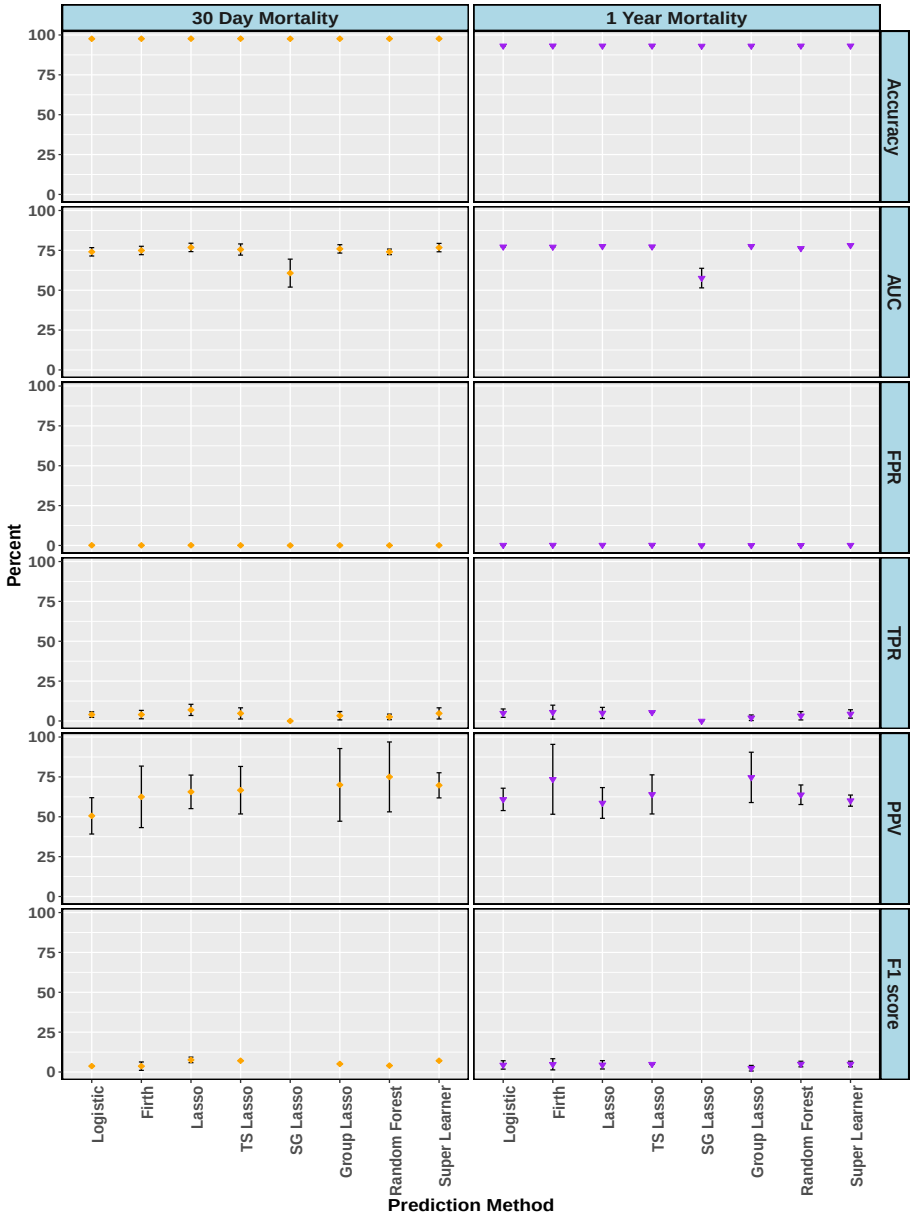

Figure 4. Data Analysis: Cross-Validated Algorithm Performance with 95% Confidence Intervals in AVR or AVR & CABG Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore* F_1 *score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. TS is an abbreviation for treatment-specific and SG is sparse group.*

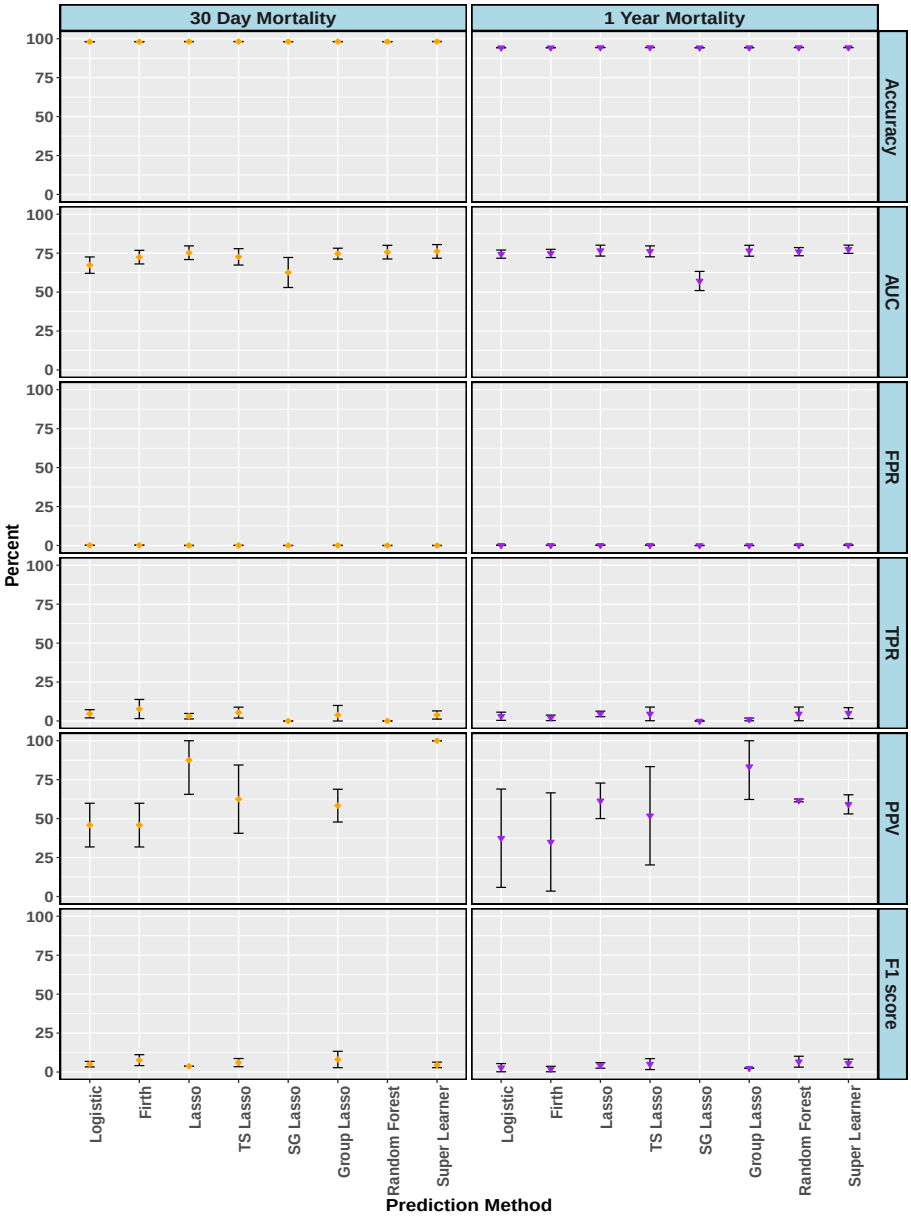

Figure 5. Data Analysis: Cross-Validated Algorithm Performance with 95% Confidence Intervals in AVR or AVR & MVR Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore* F_1 *score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. TS is an abbreviation for treatment-specific and SG is sparse group.*

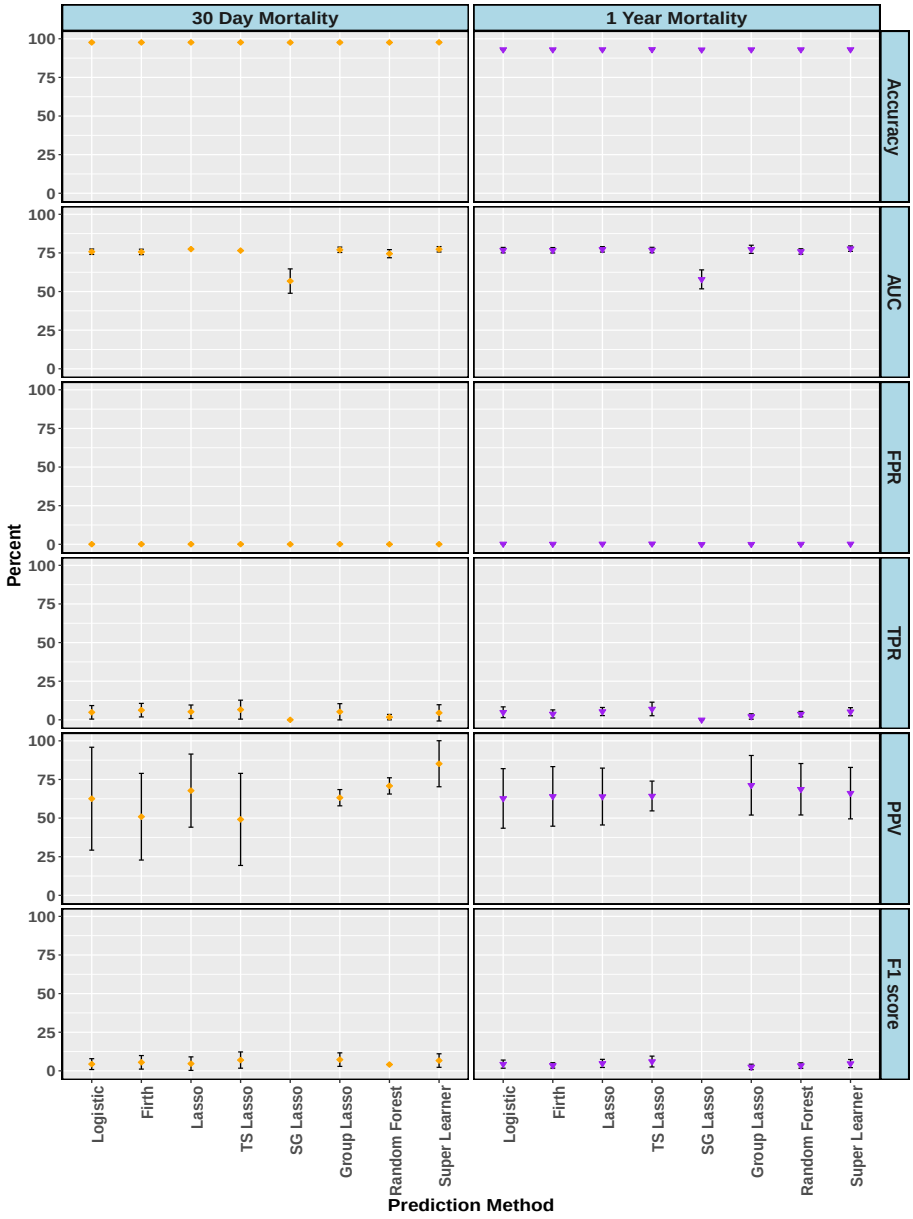

Figure 6. Data Analysis: Cross-Validated Algorithm Performance with 95% Confidence Intervals in any AVR Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. TS is an abbreviation for treatment-specific and SG is sparse group.*

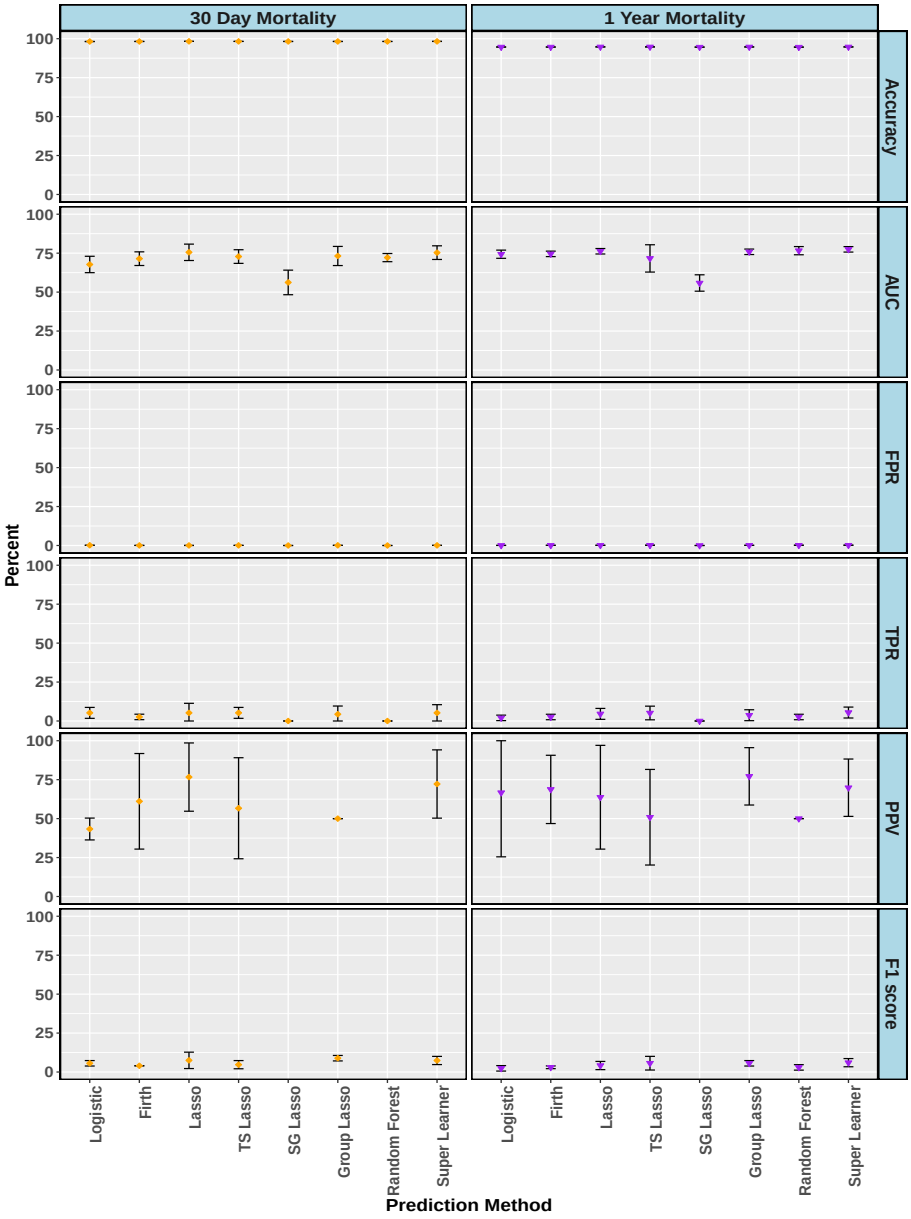

Figure 7. Data Analysis: Cross-Validated Algorithm Performance with 95% Confidence Intervals in Isolated AVR Cohort using Negative Log-Likelihood Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. TS is an abbreviation for treatment-specific and SG is group.*

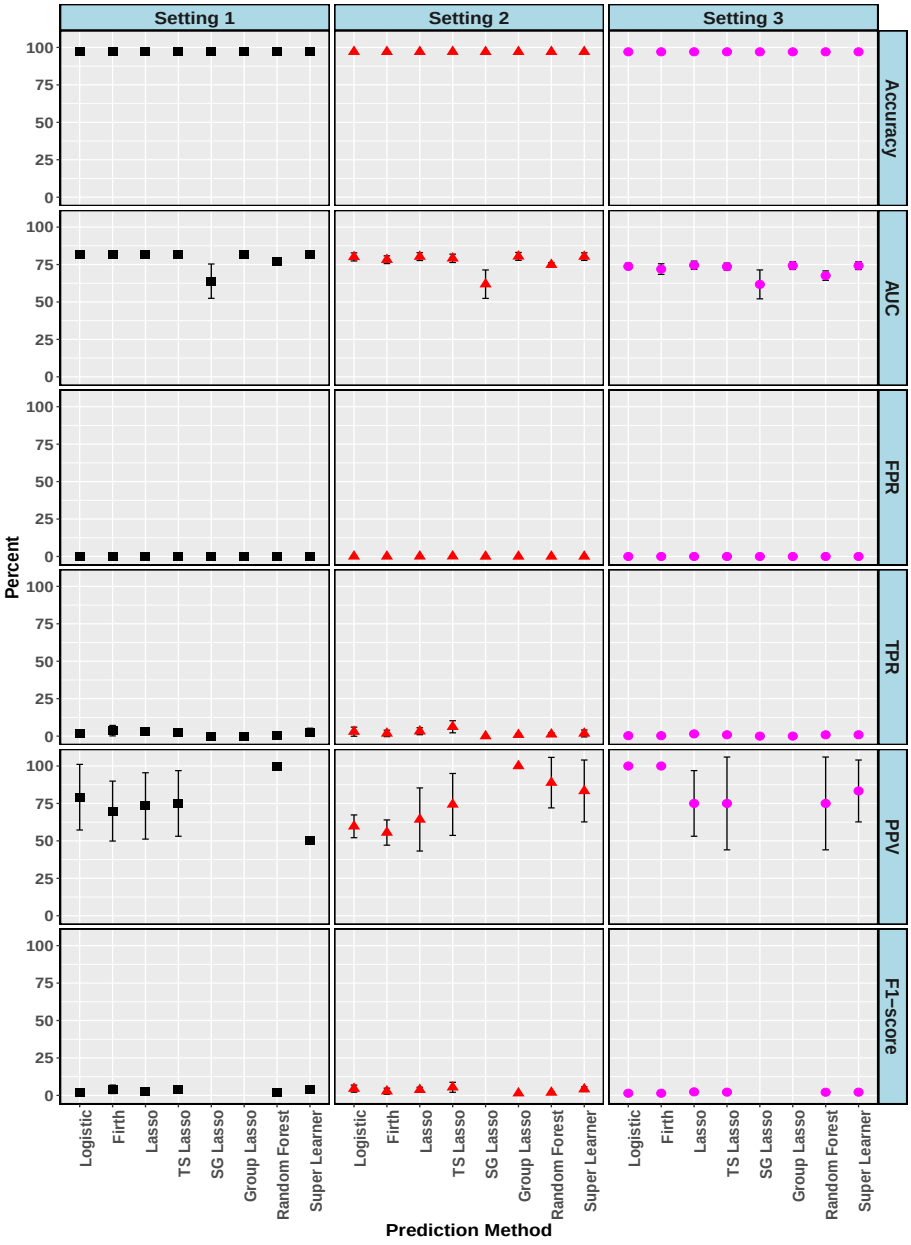

Figure 8. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 30-Day Mortality in AVR or AVR & CABG Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 84% for setting 1, 85% for setting 2, and 77% for setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

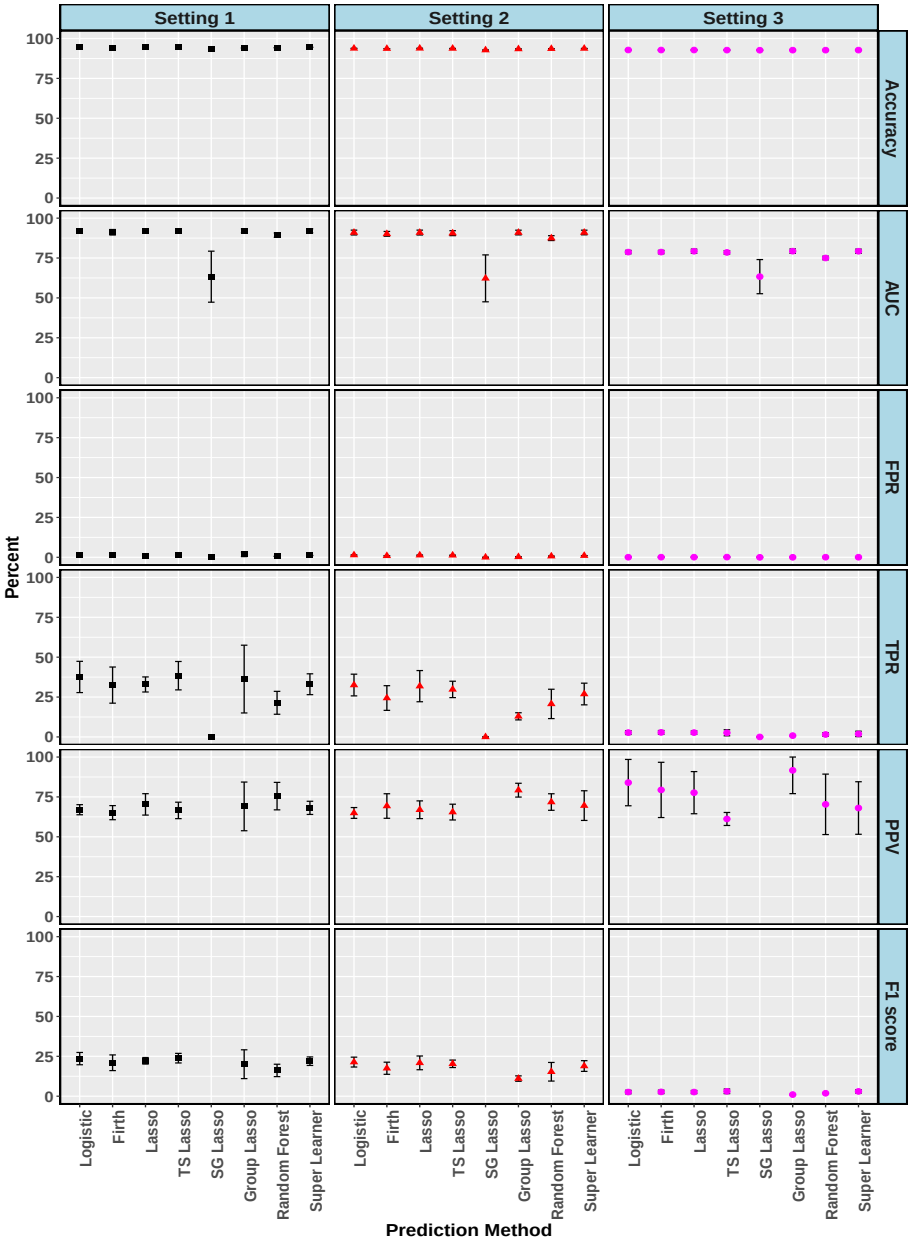

Figure 9. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 1-Year Mortality in AVR or AVR & CABG Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown.True conditional risk estimate based on AUC loss is 93% for setting 1, 91% for setting 2, and 92% for setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

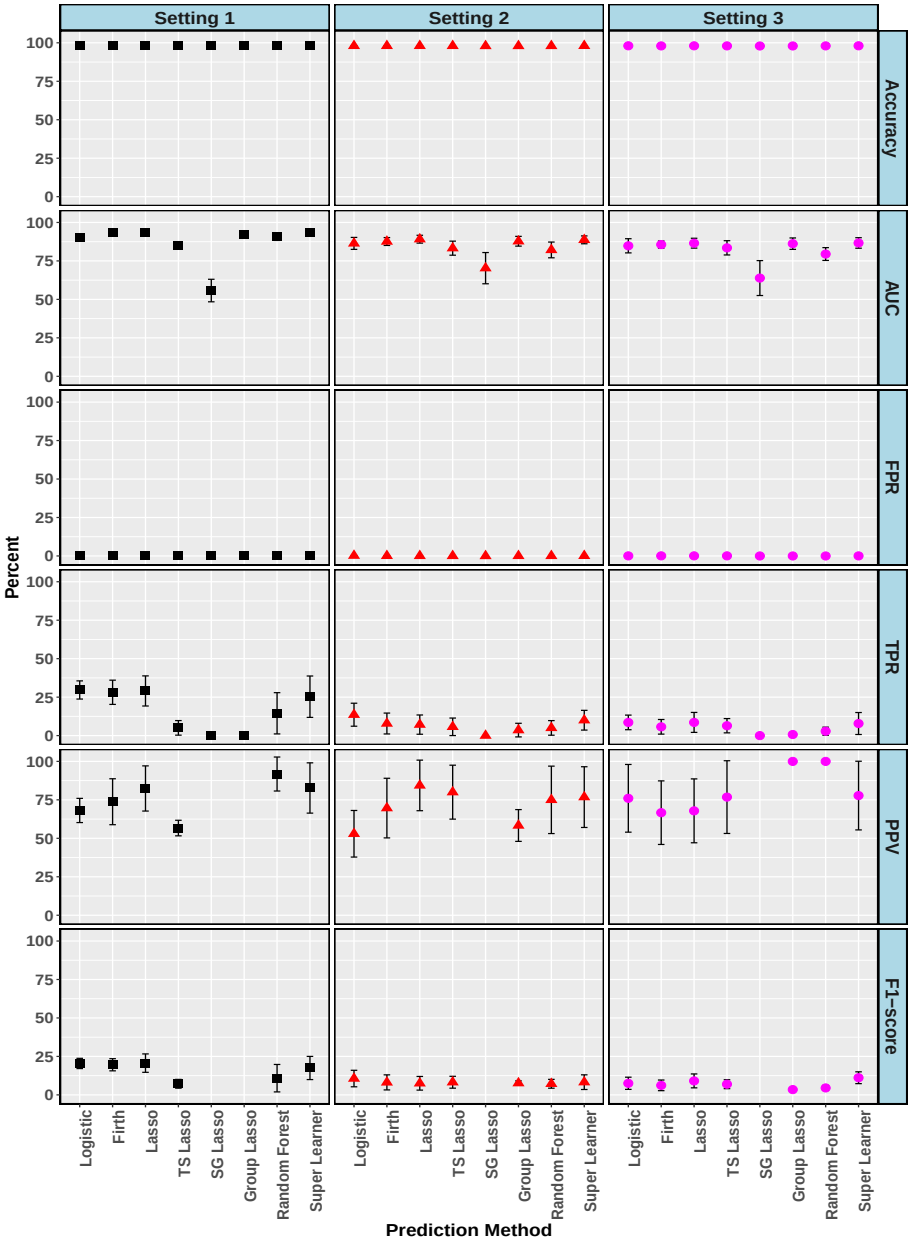

Figure 10. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 30-Day Mortality using in AVR or AVR & MVR Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 95% for setting 1, 91% for setting 2, and 94% for setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

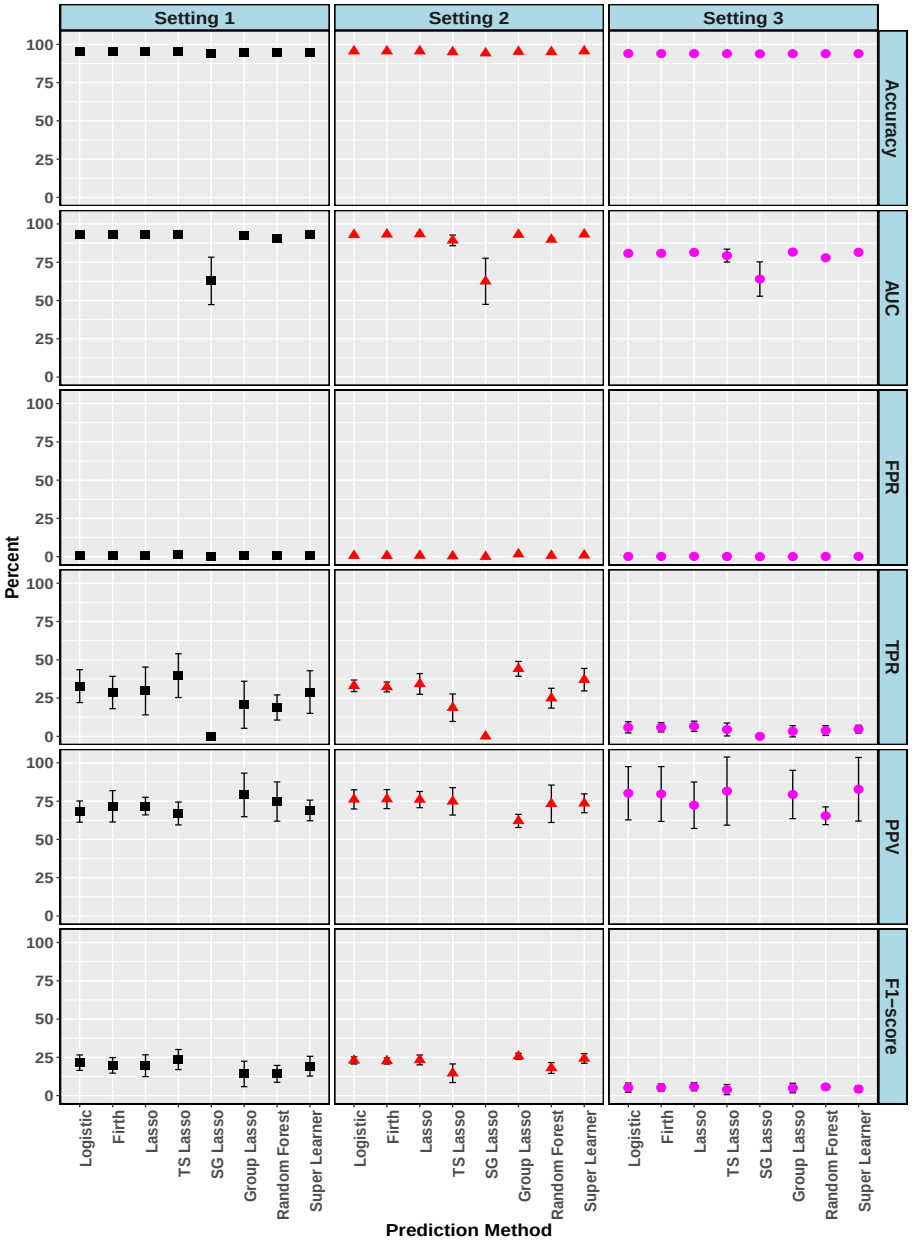

Figure 11. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 1-Year Mortality in AVR or AVR & MVR Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 94% for settings 1 and 2 and 93% for setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

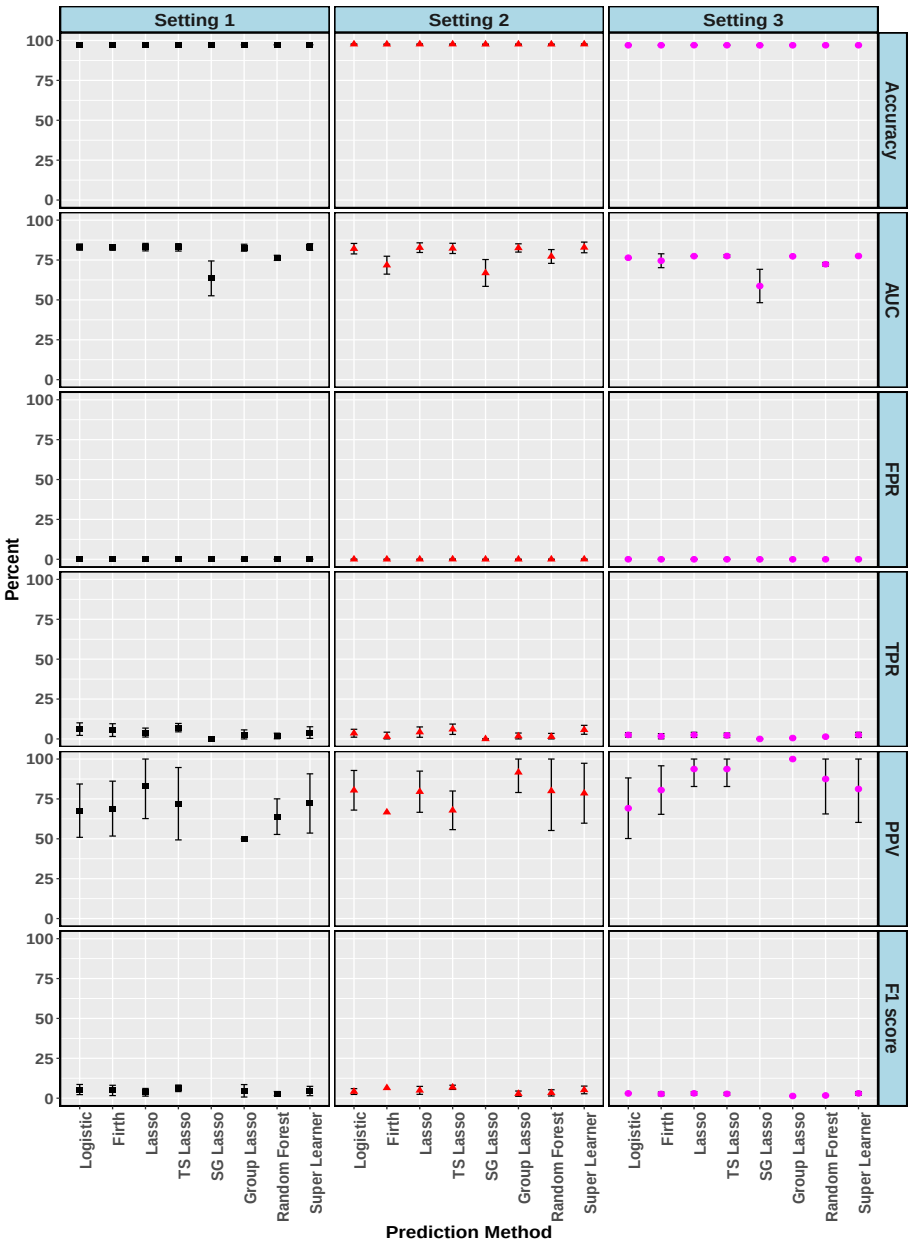

Figure 12. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 30-Day Mortality in any AVR Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 85% for setting 1, 84% for setting 2 and 79% for setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

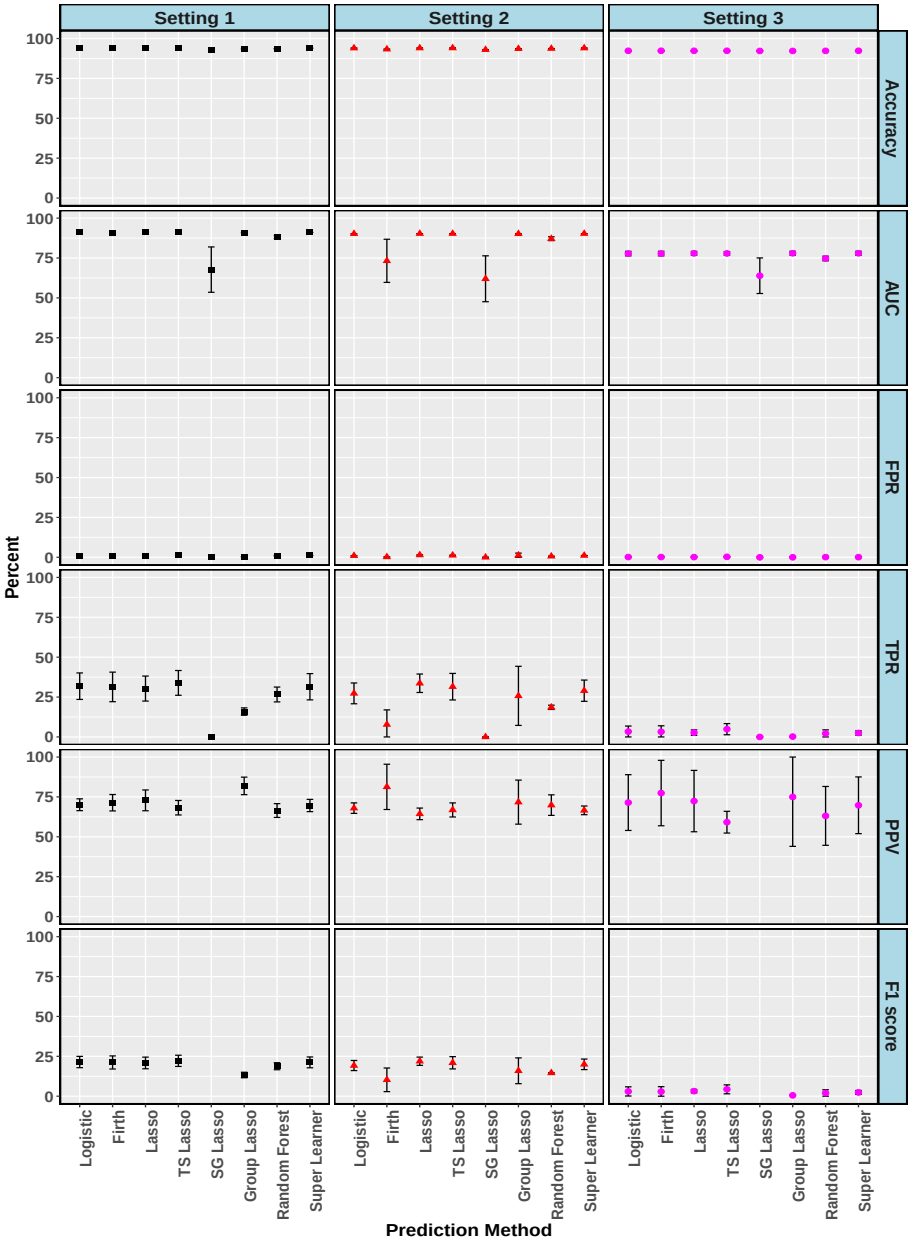

Figure 13. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 1-Year Mortality in any AVR Cohort using AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 92% for settings 1 and 3 and 91% for setting 2. TS is an abbreviation for treatment-specific and SG is sparse group.*

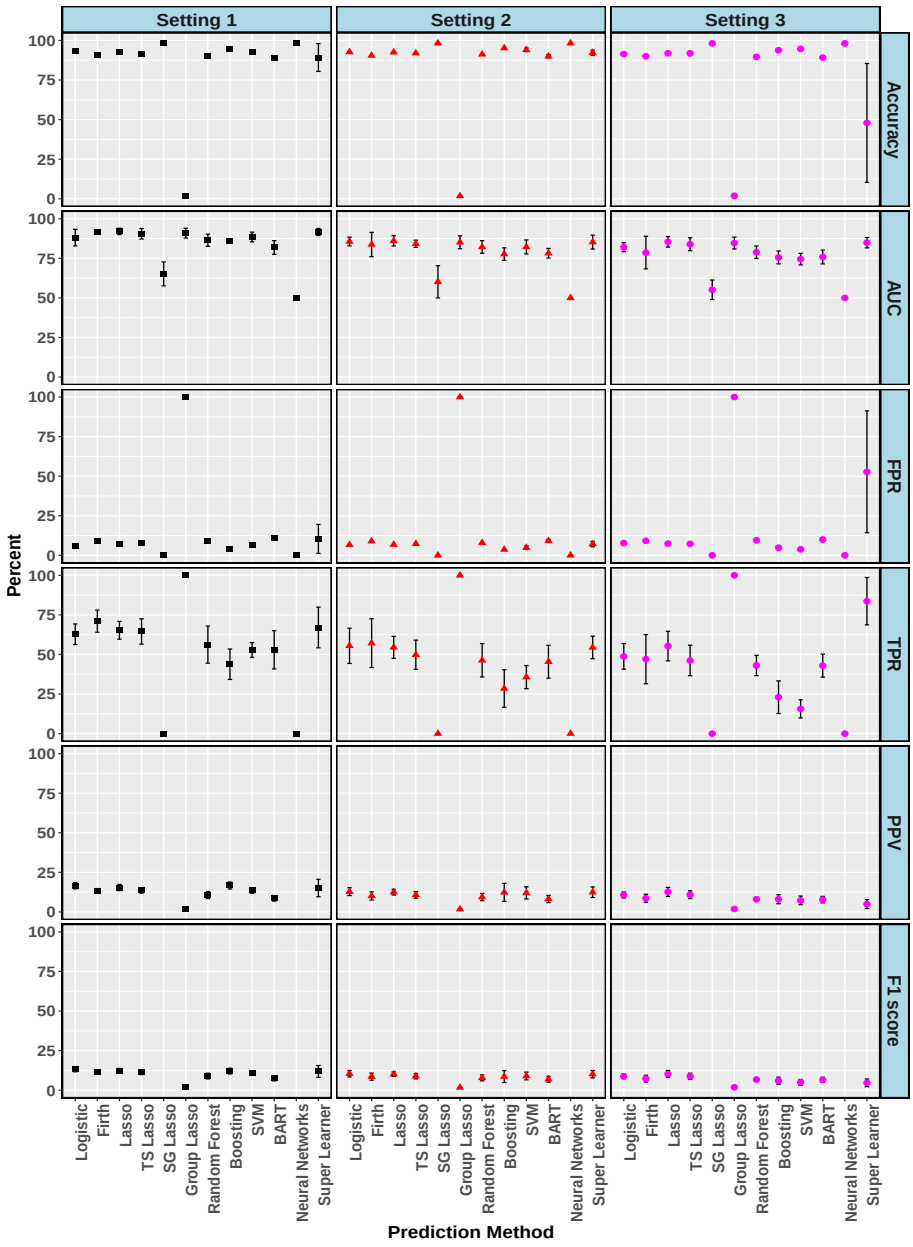

Figure 14. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 30-Day Mortality in Isolated AVR Cohort using AUC Loss Function Maximizing TPR. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 94% for setting 1, 89% for setting 2 and 95% for setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

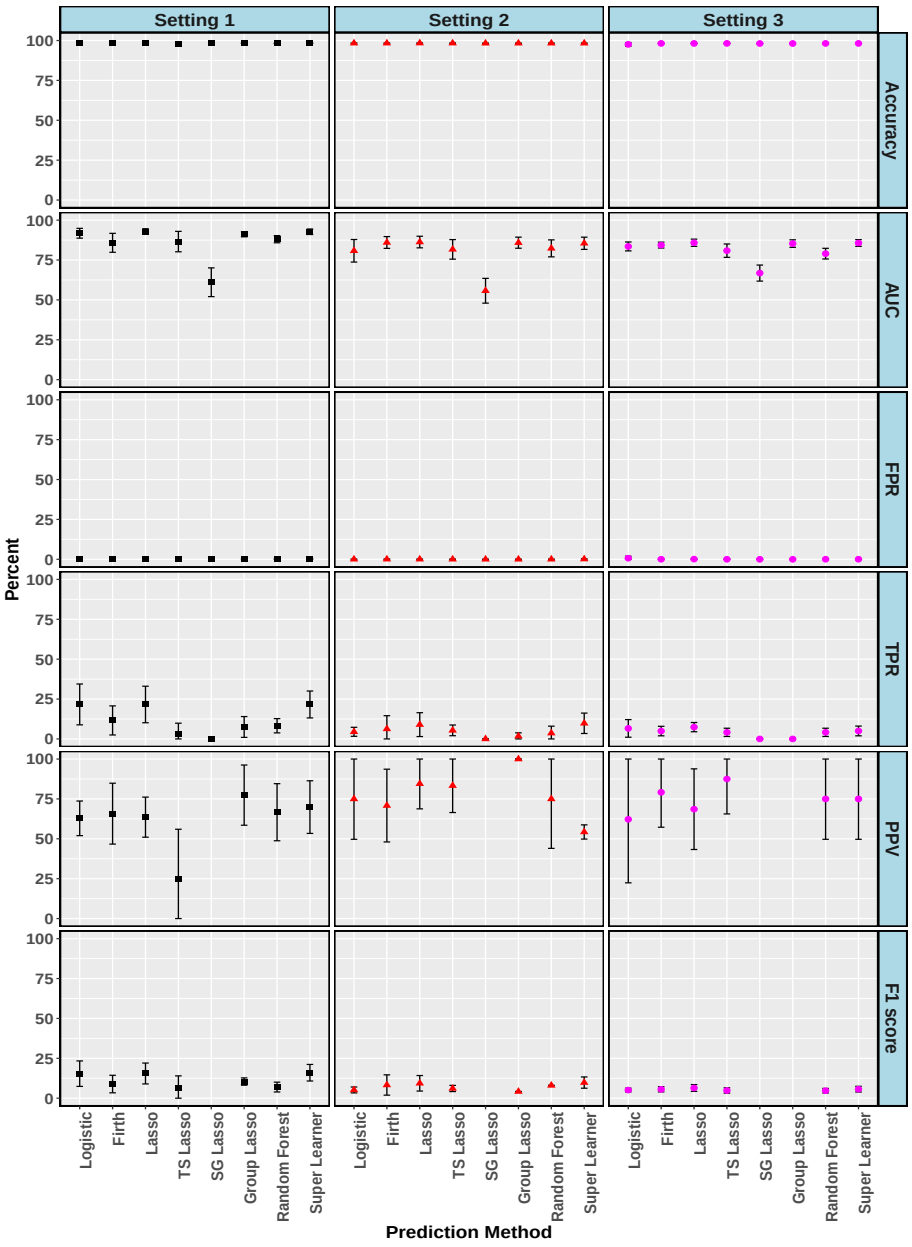

Figure 15. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 30-Day Mortality in Isolated AVR Cohort using Negative Log-Likelihood Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on average log loss (negative log likelihood divided by number of observations) is 0.05 for settings 1 and 3 and 0.07 for setting 2. For comparison, mean cross-validated log loss for super learner was 0.06 in setting 1 and 0.07 in settings 2 and 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

Prepared using sagej.cls

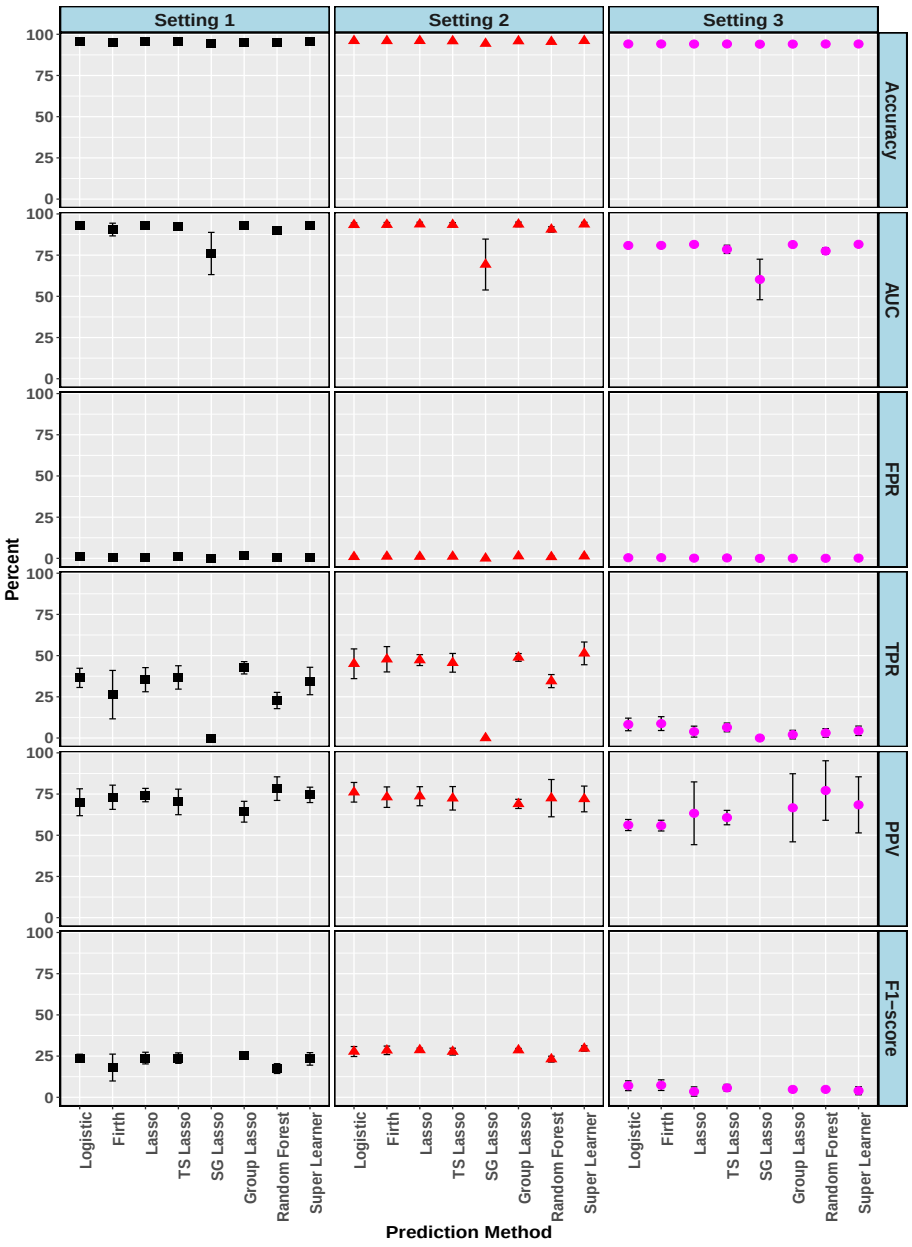

Figure 16. Simulation: Cross-validated Algorithm Performance with 95% Confidence Intervals for 1-Year Mortality in Isolated AVR using Negative Log-Likelihood Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown True conditional risk estimate based on average log loss (negative log likelihood divided by number of observations) is 0.12 for settings 1 and 3 and 0.11 for setting 2. For comparison, mean cross-validated log loss for super learner was 0.13 in setting 1, 0.12 in setting 2 and 0.19 in setting 3. TS is an abbreviation for treatment-specific and SG is sparse group.*

Prepared using sagej.cls

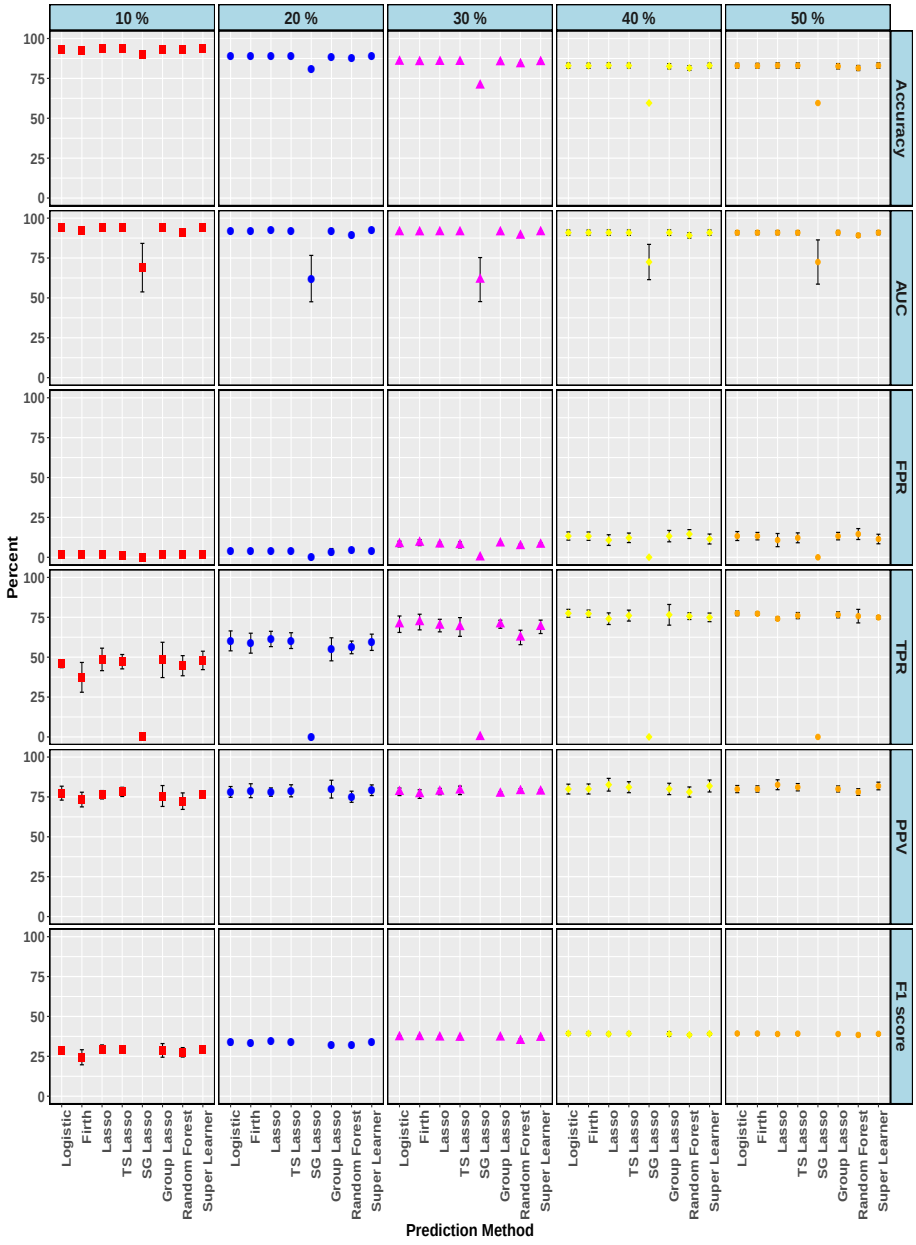

Figure 17. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for Isolated AVR and Different Mortality Rates using Simulation Setting 1 and AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 94% (for 10% event rate), 93% (for 20% and 50% event rates), 92% for (30% event rate), and 91% (for 40% event rate). TS is an abbreviation for treatment-specific and SG is sparse group.*

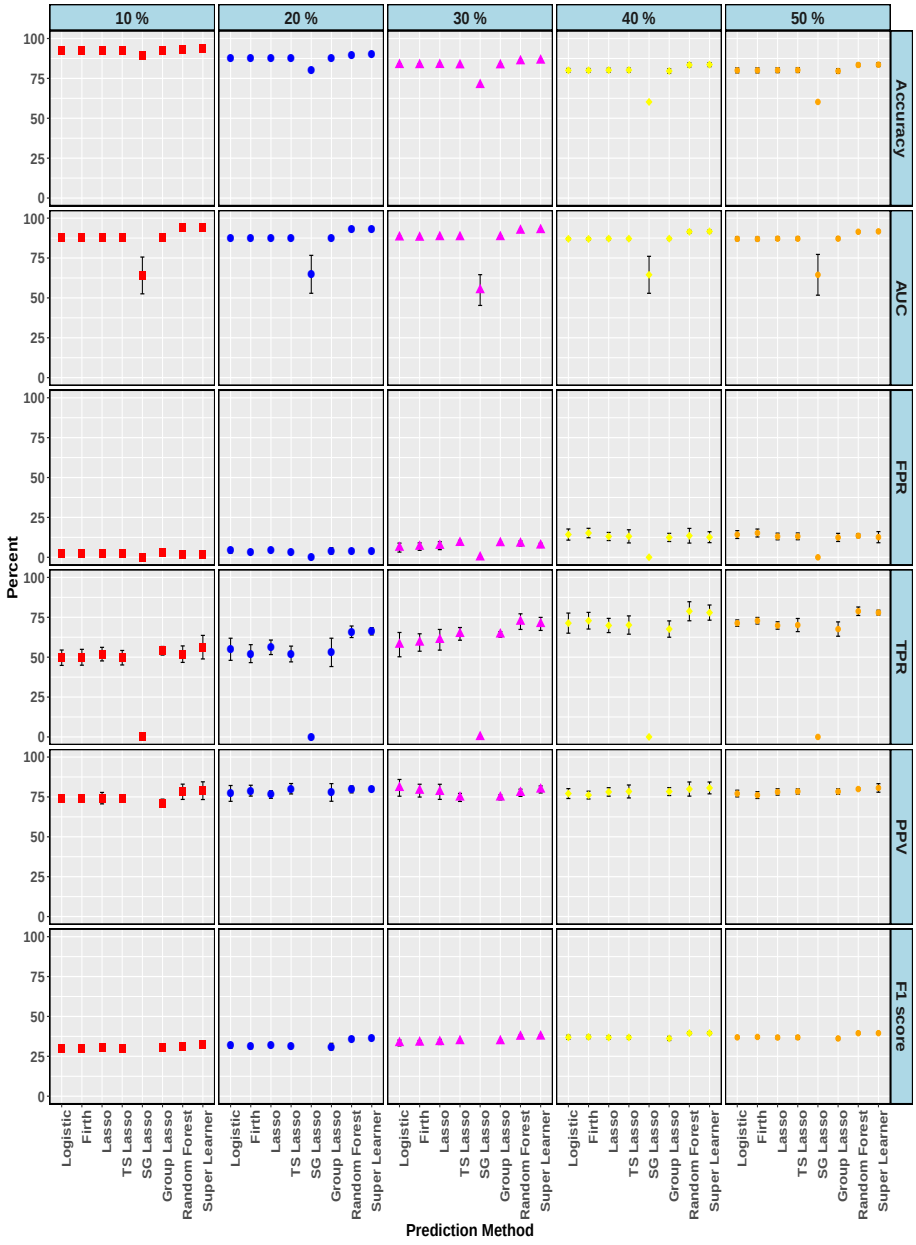

Figure 18. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for Isolated AVR and Different Mortality Rates using Simulation Setting 2 and AUC Loss Function. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. True conditional risk estimate based on AUC loss is 96% (for 10% event rate), 95% (for 20% event rate), 94% for (30% event rate), 94% (for 40% event rate), and 93% (for 50% event rate). TS is an abbreviation for treatment-specific and SG is sparse group.*

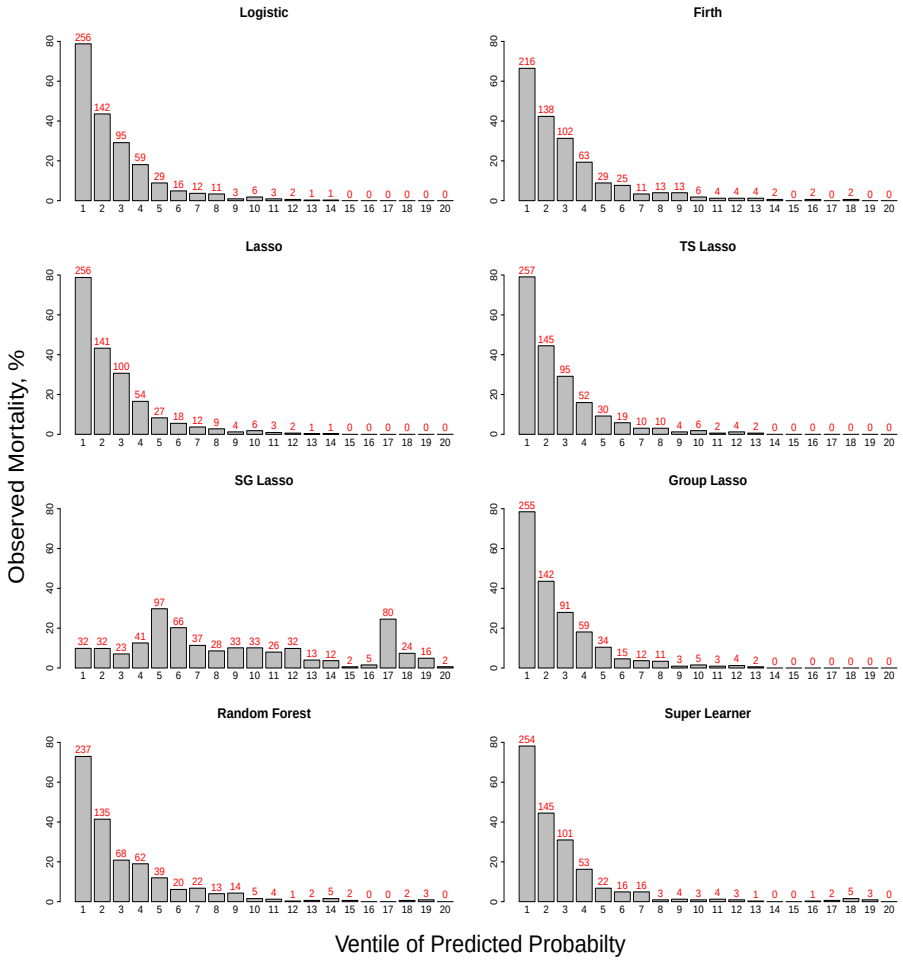

Figure 19. Mortality Rates within each Ventile of Predicted Mortality Risk for Different Algorithms in Simulated Data with 10% Mortality Rate under Simulation Setting 1. *The predicted mortality risks are in decreasing order and red values are the number of events in each ventile. TS is an abbreviation for treatment-specific and SG is sparse group.*

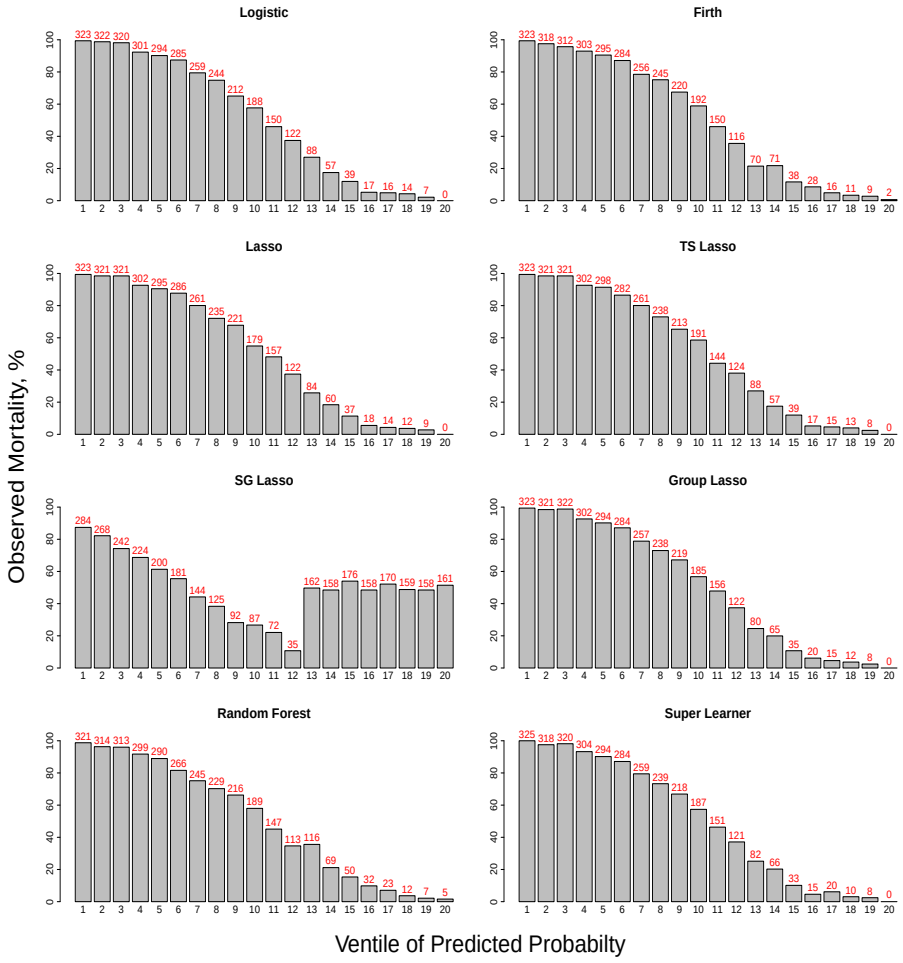

Figure 20. Mortality Rates within each Ventile of Predicted Mortality Risk for Different Algorithms in Simulated Data with 50% Mortality Rate under Simulation Setting 1. *The predicted mortality risks are in decreasing order and red values are the number of events in each ventile. TS is an abbreviation for treatment-specific and SG is sparse group.*

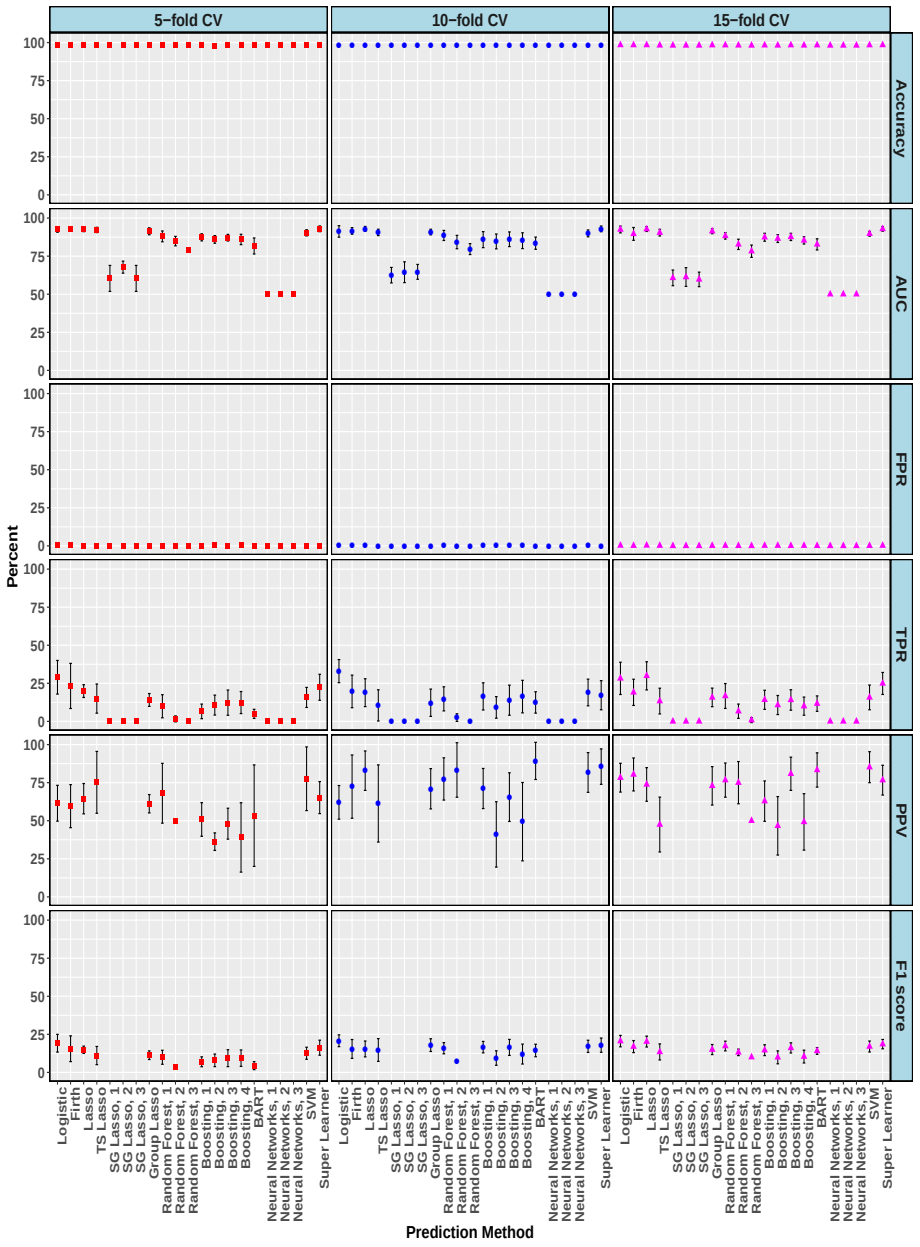

Figure 21. Simulation: Cross-Validated Algorithm Performance with 95% Confidence Intervals for 30-Day Mortality in Isolated AVR using Simulation Setting 1 and AUC Loss Function with Varied Cross-Validation Folds and Extended Algorithms with Different Hyperparameters in the Ensemble. *For algorithms with zero predicted positive values, PPV is undefined and not plotted, and therefore F_1 score is also undefined and not plotted. 95% confidence intervals for estimates with standard errors less than 1% are not shown. TS is an abbreviation for treatment-specific and SG is sparse group.*