

## Online Supplement

SuperLearner is an ensemble machine learning method, meaning it uses a library of algorithms in tandem to predict an estimate of an expected outcome, given some set of variables. For this study, we used SuperLearner to estimate the expected cost, given probation type and background covariates (for results on SuperLearner estimation and performance for the treatment mechanism, see 14). In the end, SuperLearner assigns a weight to all of the algorithms put into the library: the more heavily an algorithm is weighted, the better it performed (according to its cross-validated mean square error), the more it is used in the prediction/estimation stage.

The algorithms (with their abbreviations) that comprised the SuperLearner library were: multivariate adaptive regression splines (earth), generalized additive models (GAM), generalized linear models (glm), lasso and elastic-net regularized generalized linear models (glmnet), improved predictors (ipred), mean, pruned recursive partitioning (rpartPrune), backward and forward stepwise AIC selection (step), random forest (randomForest), neural networks (nnet), classification and regression training (caret), multivariate adaptive polynomial spline regression (polymars).

We reviewed the SuperLearner weights assigned to the sub-algorithms used to estimate the outcome regression (i.e., the expected cost given probation type and covariates). Indeed, the algorithms did not solely rely on (i.e., did not exclusively assign higher weights to) algorithms that are more sensitive to unusual points, such as mean-based methods or variations of linear models (e.g., glm, glmnet, mean). The algorithms chosen were a combination of these algorithms, in addition to ones robust to extreme points and unusual distributions (recursive partitioning and regression trees, neural networks). See Supplementary Table 1 for coefficient

estimates on the candidate algorithms as a measure of their performance and contribution to the estimate of the outcome regression.

Supplementary Table 1

*Coefficients (weights) depicting algorithm performance for the expected cost, given probation type and background covariates. Each column is a different cost type, and each row is a candidate algorithm.*

Algorithm	Total combined	Behavioral health	ER/ Inpatient/ Res	Outpatient	Criminal justice	Supervision
earth	.04	.00	.02	.00	.00	.08
GAM	.00	.00	.00	.05	.15	.00
glm	.00	.00	.00	.00	.00	.00
glmnet	.00	.46	.00	.20	.85	.82
ipred	.18	.00	.00	.25	.00	.00
mean	.00	.12	.00	.27	.00	.00
rpartPrune	.00	.00	.00	.00	.00	.10
step	.00	.32	.00	.12	.00	.00
randomForest	.00	.00	.00	.00	.00	.00
nnet	.61	.09	.98	.00	.00	.00
caret	.00	.00	.00	.00	.00	.00
polymars	.17	.00	.00	.11	.00	.00