

CALIBRATION AND DISCRIMINATION TESTING

FIGURE 1 shows the distribution of predicted values (event probabilities) from the risk-adjustment model.

Minimum = 0.0003095 = 0.031%
10th percentile = 0.0003095 = 0.031%
Median = 0.0006421 = 0.064%
Mean = 0.0030036 = 0.30%
90th percentile = 0.0035837 = 0.36%
Maximum = 0.715681 = 71.57%

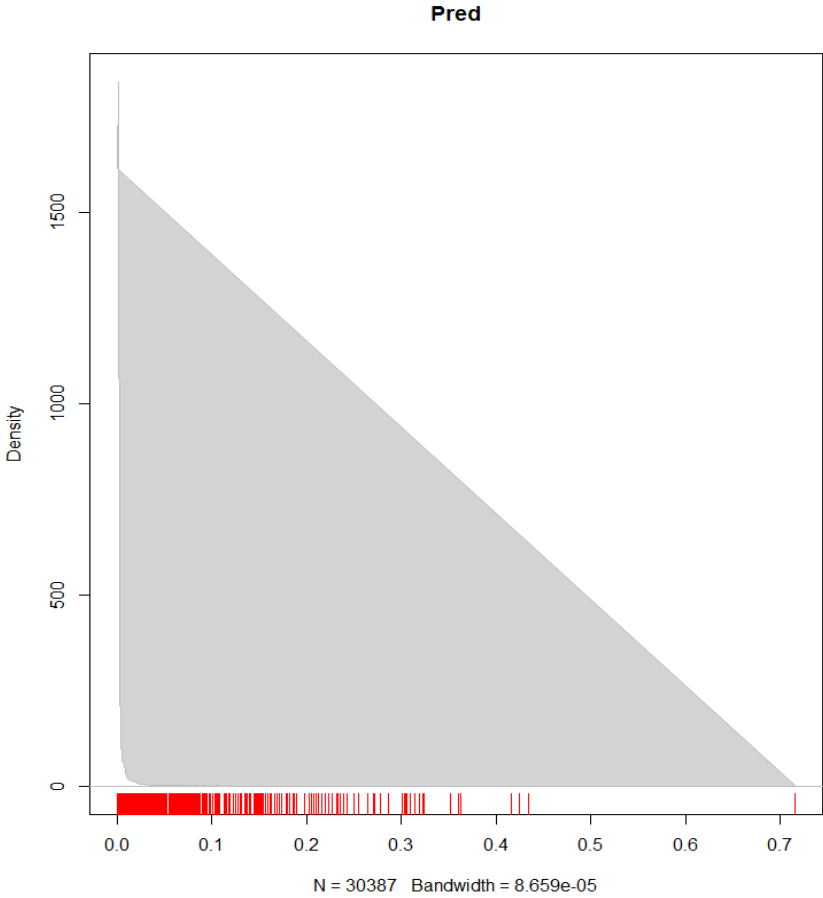


FIGURE 2 shows the receiver operating characteristic (calibration) curve from the holdout test set in feature selection (AUCs: ROC= 0.82647769; PRC= 0.09849938), where the dotted lines represent the performance of an uninformative model.

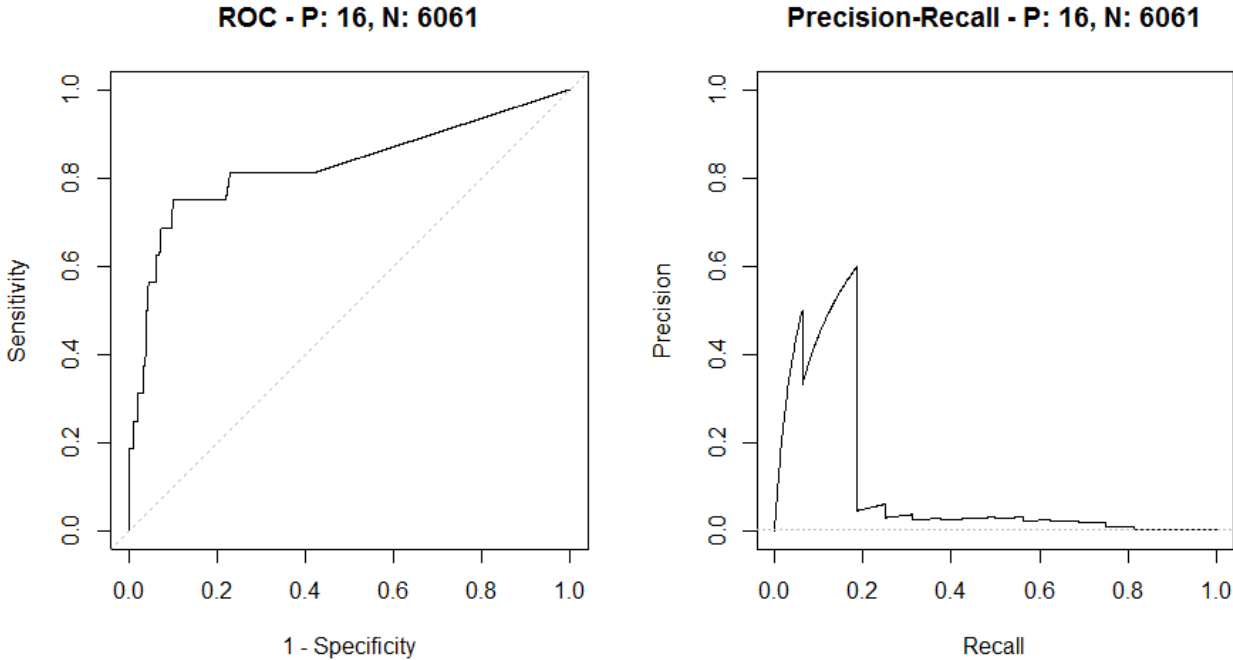
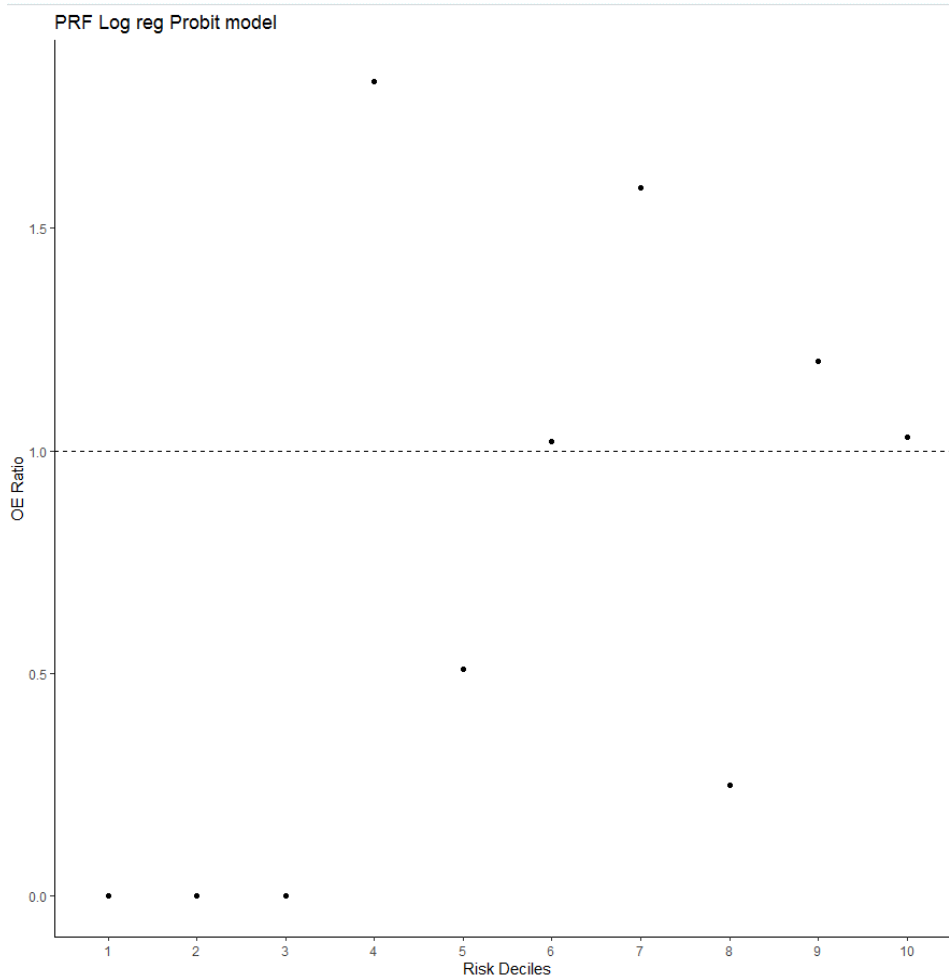


FIGURE 3 shows the Hosmer-Lemeshow decile calibration plot from the final risk-adjustment model. The results are unstable due to a small number of observed events (num_obs) and expected events (num_exp) in all of the bottom eight deciles.



decile	o_e_ratio	num_obs	num_exp	numrecs	Event rate	% of events	
1	1	0	0	0.941	<u>3039</u>	0.0000%	0.00%
2	2	0	0	0.941	<u>3039</u>	0.0000%	0.00%
3	3	0	0	0.941	<u>3039</u>	0.0000%	0.00%
4	4	1.83	3	1.64	<u>3038</u>	0.0987%	3.30%
5	5	0.51	1	1.95	<u>3039</u>	0.0329%	1.10%
6	6	1.02	2	1.97	<u>3039</u>	0.0658%	2.20%
7	7	1.59	4	2.52	<u>3038</u>	0.1317%	4.40%
8	8	0.25	1	4.05	<u>3039</u>	0.0329%	1.10%
9	9	1.2	9	7.53	<u>3039</u>	0.2962%	9.89%
10	10	1.03	71	68.8	<u>3038</u>	2.3371%	78.02%

FIGURE 4 shows the calibration band plot from the final risk-adjustment model. The null hypothesis of perfect calibration is rejected at the $p < 0.05$ level (i.e., $p = 0.049$), but the 95% confidence boundaries never cross the bisector.

