

Appendix A

Table A1 Predictors Used in the Statistical Analysis

Variable	Label	Code
Race	American Indian	0
	Asian	1
	Black	2
	Hispanic	3
	Other	4
	Hawaiian	5
	Unknown	6
	White	7
Gender	Female	0
	Male	1
	Unknown	2
Age	0-9 years	0
	10-19 years	1
	20-29 years	2
	30-39 years	3
	40-49 years	4
	50-59 years	5
	60-69 years	6
	70-79 years	7
	80+ years	8
	Unknown	9
Hosp_yn (Was the patient hospitalized?)	Yes	0

Icu_yn
(Was the patient admitted to an intensive care unit (ICU)?)

Death_yn
(Did the patient die as a result of this illness?)

Hc_work_yn
(Is the patient a health care worker in the United States?)

Pna_yn
(Did the patient develop pneumonia?)

Abxchest_yn
(Did the patient have an abnormal chest X-ray?)

Acuterespdistress_yn
(Did the patient have acute respiratory distress syndrome?)

Mechvent_yn No 1
(Did the patient receive mechanical ventilation (MV)/intubation?)

Fever_yn
Fever >100.4F (38C)

Sfever_yn (Subjective fever)

Chills_yn (Chills)

Myalgia_yn (Muscle aches)

Runnose_yn (Running nose)

Sthroat_yn (Sore Throat)

Unknown

2

Cough_yn (Cough)

Sob_yn (Shortness of breath
(dyspnea))

Nauseavomit_yn (Nausea or
Vomiting)

Headache_yn (Headache)

Abdom_yn (Abdominal pain)

Diarrhea_yn (Diarrhea)

Medcond_yn (Pre-existing
medical conditions?)

Table A2 Adjusted pooled sampling given the best cut-off values

Method	Cutoff	Group size	Pooled samples	Train data	Baseline
CART	0.65	100	15223	Original Imbalanced	15423
CART	0.9	50	14221	Original Imbalanced	14573
CART	0.65	20	10418	Original Imbalanced	10899
CART	0.55	15	9058	Original Imbalanced	9358
CART	0.9	10	7747	Original Imbalanced	7907
CART	0.6	5	6579	Original Imbalanced	6629
GB	0.05	100	13513	Original Imbalanced	15423
GB	0.05	50	10928	Original Imbalanced	14573
GB	0.05	20	9268	Original Imbalanced	10899
GB	0.05	15	9022	Original Imbalanced	9358
GB	0.15	10	7660	Original Imbalanced	7907
GB	0.15	5	6543	Original Imbalanced	6629
LR	0.05	100	13931	Original Imbalanced	15423
LR	0.05	50	11573	Original Imbalanced	14573
LR	0.05	20	9716	Original Imbalanced	10899
LR	0.1	15	9015	Original Imbalanced	9358
LR	0.75	10	7657	Original Imbalanced	7907
LR	0.5	5	6539	Original Imbalanced	6629
RF	0.05	100	12521	Original Imbalanced	15423
RF	0.05	50	10021	Original Imbalanced	14573
RF	0.05	20	7822	Original Imbalanced	10899
RF	0.1	15	7363	Original Imbalanced	9358
RF	0.15	10	6629	Original Imbalanced	7907

RF	0.25	5	6361	Original Imbalanced	6629
SVM	0.05	100	12038	Original Imbalanced	15423
SVM	0.05	50	9222	Original Imbalanced	14573
SVM	0.05	20	6696	Original Imbalanced	10899
SVM	0.05	15	6435	Original Imbalanced	9358
SVM	0.05	10	6350	Original Imbalanced	7907
SVM	0.25	5	6254	Original Imbalanced	6629
LCA	0.65	100	15322	Original Imbalanced	15423
LCA	0	50	14573	Original Imbalanced	14573
LCA	0	20	10899	Original Imbalanced	10899
LCA	0	15	9358	Original Imbalanced	9358
LCA	0	10	7907	Original Imbalanced	7907
LCA	0	5	6629	Original Imbalanced	6629
CART	0.5	100	13425	Oversampled Balanced	15423
CART	0.25	50	11332	Oversampled Balanced	14573
CART	0.6	20	9393	Oversampled Balanced	10899
CART	0.55	15	9047	Oversampled Balanced	9358
CART	0.95	10	7677	Oversampled Balanced	7907
CART	0.85	5	6579	Oversampled Balanced	6629
GB	0.5	100	13214	Oversampled Balanced	15423
GB	0.5	50	10631	Oversampled Balanced	14573
GB	0.5	20	9234	Oversampled Balanced	10899
GB	0.5	15	8968	Oversampled Balanced	9358
GB	0.75	10	7667	Oversampled Balanced	7907
GB	0.95	5	6559	Oversampled Balanced	6629
LR	0.3	100	13308	Oversampled	15423

					Balanced	
					Oversampled	
LR	0.45	50	11524		Balanced	14573
					Oversampled	
LR	0.55	20	9415		Balanced	10899
					Oversampled	
LR	0.55	15	8845		Balanced	9358
					Oversampled	
LR	0.9	10	7755		Balanced	7907
					Oversampled	
LR	0.95	5	6589		Balanced	6629
					Oversampled	
RF	0.05	100	12118		Balanced	15423
					Oversampled	
RF	0.05	50	10619		Balanced	14573
					Oversampled	
RF	0.15	20	8559		Balanced	10899
					Oversampled	
RF	0.15	15	8006		Balanced	9358
					Oversampled	
RF	0.7	10	7167		Balanced	7907
					Oversampled	
RF	0.85	5	6373		Balanced	6629
					Oversampled	
SVM	0.5	100	15224		Balanced	15423
					Oversampled	
SVM	0.75	50	14222		Balanced	14573
					Oversampled	
SVM	0.2	20	10437		Balanced	10899
					Oversampled	
SVM	0.2	15	9014		Balanced	9358
					Oversampled	
SVM	0.2	10	7600		Balanced	7907
					Oversampled	
SVM	0.3	5	6444		Balanced	6629
					Oversampled	
LCA	0.85	100	15322		Balanced	15423
					Oversampled	
LCA	0	50	14573		Balanced	14573
					Oversampled	
LCA	0	20	10899		Balanced	10899
					Oversampled	
LCA	0	15	9358		Balanced	9358
					Oversampled	
LCA	0	10	7907		Balanced	7907
					Oversampled	
LCA	0	5	6629		Balanced	6629

Note: CART; classification and regression tree, GB; gradient boosting, LCA; latent class

analysis, LR; logistic regression, RF; random forest, SVM; support vector machine