

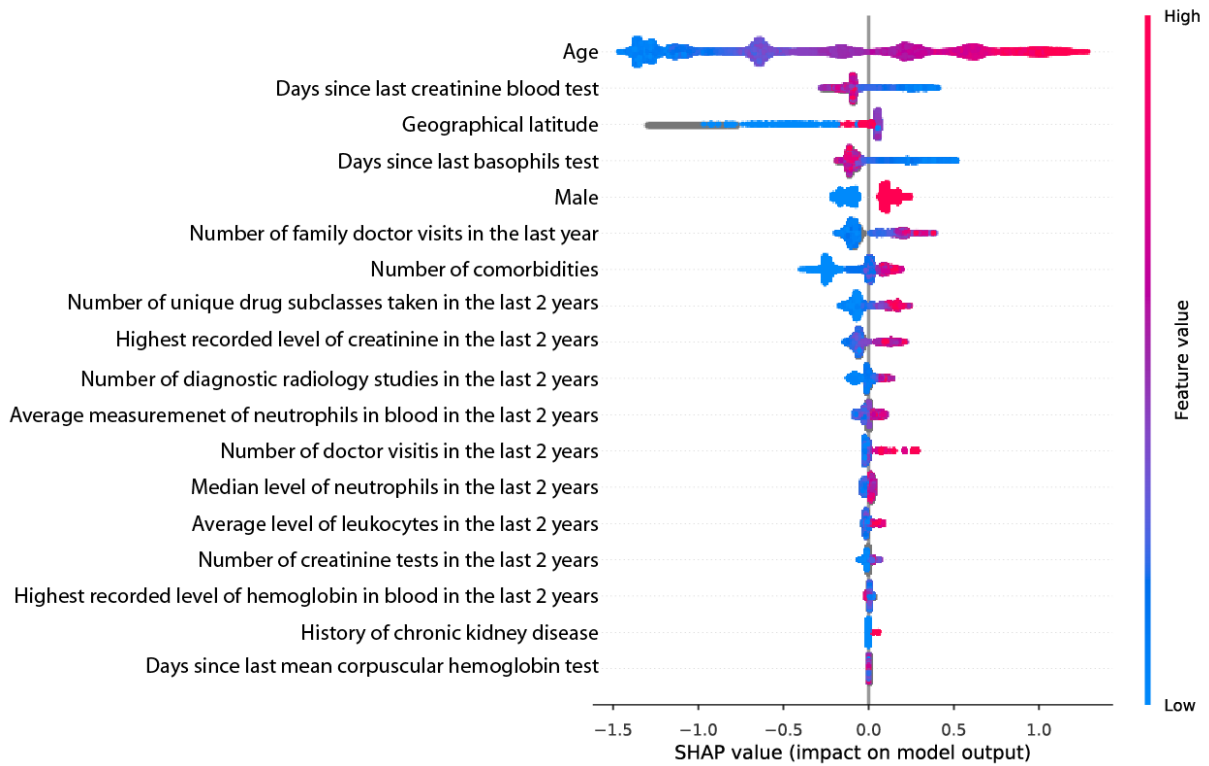
# Supplementary Material to

## COVID-19 Risk Stratification: a Multivariable Model Based on Gradient Boosting Decision Trees

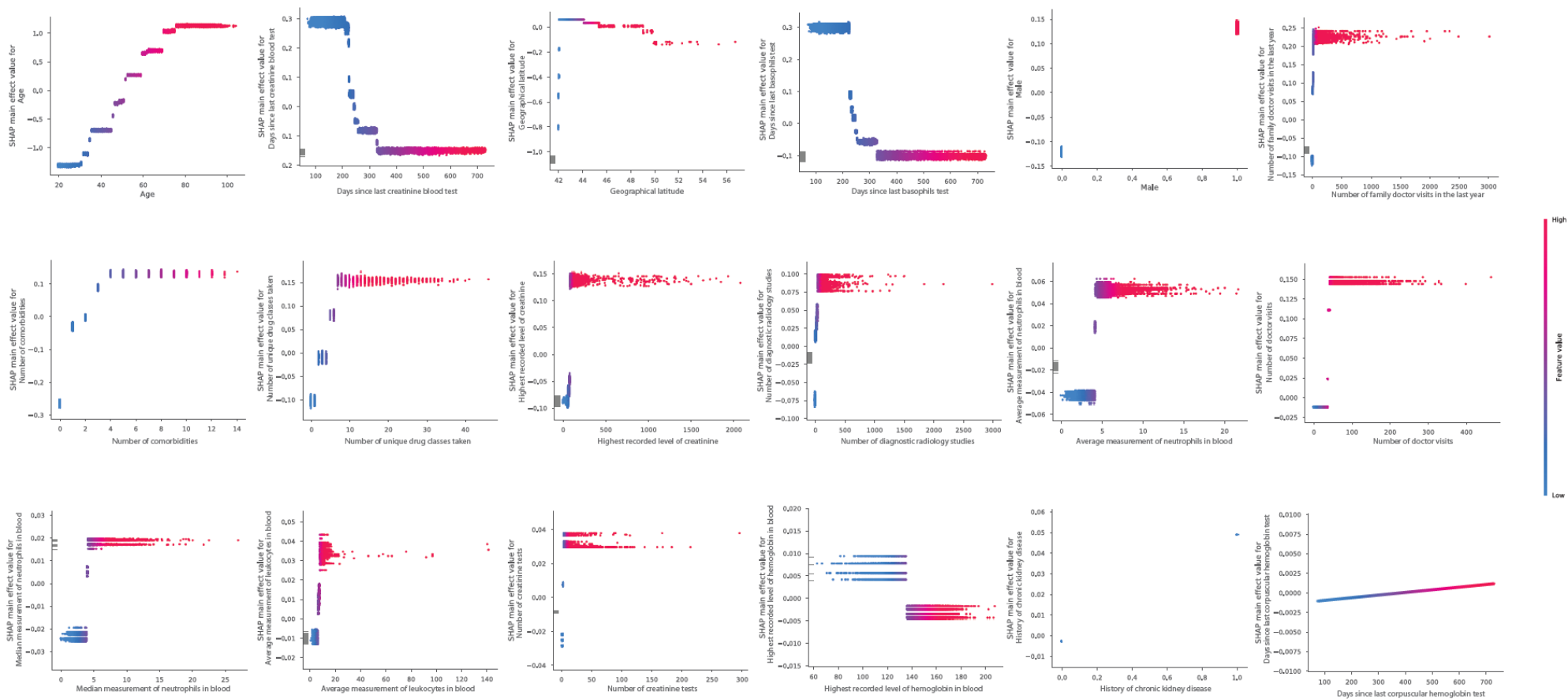
### Contents

- **Supplementary Figures**
  - Supplementary Figure 1 - SHAP value summary plot
  - Supplementary Figure 2 - Partial dependence plots
- **Supplementary Tables**
  - Supplementary Table 1 - Predictor variables considered for building XGBoost model
  - Supplementary Table 2 - Calibration curve raw data

Appendix 1, as supplied by the authors. Appendix to: Gutierrez JM, Volkovs M, Poutanen T, et al. Risk stratification for COVID-19 hospitalization: a multivariable model based on gradient-boosting decision trees. *CMAJ Open* 2021. doi: 10.9778/cmajo.20210036. Copyright © 2021 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at [cmajgroup@cmaj.ca](mailto:cmajgroup@cmaj.ca).



**Supplementary Figure 1 - Summary of SHAP values across predictor variables included in XGBoost model.** Each dot represents a data point (patient) and the color indicates magnitude of feature value (red is high, blue is low)



**Supplementary Figure 2 - Partial dependence plots for predictor variables included in XGBoost model.** In each plot, the horizontal axis is the value of the independent variable and the vertical axis represents the change in marginal SHAP value change while fixing all other variables. The color of each dot indicates the magnitude of feature value (red is high, blue is low). Variables are sorted in descending order by overall contribution (from left to right and top to bottom) in the same order as in Supplementary Figure 1.

**Supplementary Table 1** - Predictor variables considered for building XGBoost model

*Note: only variables included for model development and variables selection (N=133) are listed here.*

<b>Predictor variable</b>	<b>Fraction of patients without observations for variable</b>	<b>Decision (discard missing if fraction &gt; 0.5)</b>
1-year summary (Drug prescriptions) - unique drug subclasses	0.000000	INCLUDE
1-year summary (Doctor visits) - COUNTS	0.000000	INCLUDE
2-year summary (Drug prescriptions) - unique drug subclasses	0.000000	INCLUDE
2-year summary (Doctor visits) - COUNTS	0.000000	INCLUDE
(Demographics) - Age	0.000000	INCLUDE
(Demographics) - History of acute myocardial infarction	0.000000	INCLUDE
(Demographics) - History of arrhythmia	0.000000	INCLUDE
(Demographics) - History of arthritis	0.000000	INCLUDE
(Demographics) - History of asthma	0.000000	INCLUDE
(Demographics) - History of cancer	0.000000	INCLUDE
(Demographics) - History of chronic heart failure	0.000000	INCLUDE
(Demographics) - History of colitis	0.000000	INCLUDE
(Demographics) - History of chronic obstructive pulmonary disease	0.000000	INCLUDE
(Demographics) - History of coronary disease	0.000000	INCLUDE
(Demographics) - History of dementia	0.000000	INCLUDE
(Demographics) - History of diabetes	0.000000	INCLUDE
(Demographics) - History of hypertension	0.000000	INCLUDE
(Demographics) - Sex	0.000000	INCLUDE
(Demographics) - History of mental illness	0.000000	INCLUDE
(Demographics) - History of mood disorder	0.000000	INCLUDE
(Demographics) - Number comorbidities	0.000000	INCLUDE
(Demographics) - History of ostarthritis	0.000000	INCLUDE
(Demographics) - History of osteoporosis	0.000000	INCLUDE
(Demographics) - History of kidney disease	0.000000	INCLUDE

(Demographics) - History of stroke	0.000000	INCLUDE
(Geographical) - Education quantile	0.075655	INCLUDE
(Geographical) - Lives in Ontario eastern region	0.075655	INCLUDE
(Geographical) - Lives in Ontario GTHA region	0.075655	INCLUDE
(Geographical) - Lives in Toronto	0.075655	INCLUDE
(Demographics) - Marginalized dependencies indicator	0.075655	INCLUDE
(Demographics) - Marginalized deprivation indicator	0.075655	INCLUDE
(Demographics) - Marginalized ethnicity indicator	0.075655	INCLUDE
(Demographics) - Marginalized instability quintile	0.075655	INCLUDE
(Geographical) - Lives in Ontario western region	0.075655	INCLUDE
(Geographical) - Lives in Ontario northern region	0.075655	INCLUDE
(Geographical) - LHIN number	0.075957	INCLUDE
(Geographical) - Latitude	0.077829	INCLUDE
(Geographical) - Longitude	0.077829	INCLUDE
(Demographics) - Income quantile	0.078408	INCLUDE
(Geographical) - Lives in rural community	0.078408	INCLUDE
2-year summary (Doctor visits) - FAMILY PRACTICE AND GENERAL PRACTICE	0.079812	INCLUDE
1-year summary (Doctor visits) - FAMILY PRACTICE AND GENERAL PRACTICE	0.168103	INCLUDE
2-year summary (Laboratory tests) - number of tests for Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - number of high values of Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - highest value of Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - number of low values of Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - lowest value of Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - average value of Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - median value of Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Neutrophils	0.354541	INCLUDE
2-year summary (Laboratory tests) - number of tests for Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Eosinophils	0.354872	INCLUDE

2-year summary (Laboratory tests) - days since latest test for Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - number of high values of Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - highest value of Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - number of low values of Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - lowest value of Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - average value of Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - median value of Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Eosinophils	0.354872	INCLUDE
2-year summary (Laboratory tests) - number of tests for Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - number of high values of Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - highest value of Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - number of low values of Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - lowest value of Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - average value of Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - median value of Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Leukocytes	0.354899	INCLUDE
2-year summary (Laboratory tests) - number of tests for Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - number of high values of Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - highest value of Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - number of low values of Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - lowest value of Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - average value of Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - median value of Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Platelets	0.355119	INCLUDE
2-year summary (Laboratory tests) - number of tests for Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - number of tests for Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Hemoglobin	0.355945	INCLUDE

2-year summary (Laboratory tests) - days since earliest test for Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - number of high values of Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - number of high values of Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - highest value of Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - highest value of Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - number of low values of Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - number of low values of Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - lowest value of Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - lowest value of Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - average value of Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - average value of Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - median value of Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - median value of Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Hemoglobin	0.355945	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Monocytes	0.355945	INCLUDE
2-year summary (Laboratory tests) - number of tests for Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - number of high values of Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - highest value of Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - number of low values of Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - lowest value of Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - average value of Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - median value of Basophils	0.356110	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Basophils	0.356110	INCLUDE
2-year summary (Doctor visits) - DIAGNOSTIC RADIOLOGY	0.360240	INCLUDE
2-year summary (Laboratory tests) - number of tests for Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Creatinine	0.377750	INCLUDE

2-year summary (Laboratory tests) - number of high values of Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - highest value of Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - number of low values of Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - lowest value of Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - average value of Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - median value of Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Creatinine	0.377750	INCLUDE
2-year summary (Laboratory tests) - number of tests for Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - days since earliest test for Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - days since latest test for Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - number of high values of Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - highest value of Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - number of low values of Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - lowest value of Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - average value of Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - median value of Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE
2-year summary (Laboratory tests) - count of normal tests for Erythrocyte_mean_corpuscular_hemoglobin	0.397462	INCLUDE



**Supplementary Table 2 - Calibration curve raw data**

<b>Decile of predicted risk</b>	<b>XGBoost predicted risk score</b>	<b>Actual risk score</b>
1	0.010041	0
2	0.012682	0.005525
3	0.016369	0.005376
4	0.02382	0.019284
5	0.033734	0.027548
6	0.04851	0.027548
7	0.069461	0.054795
8	0.104906	0.09863
9	0.191218	0.176309
10	0.428636	0.426966