**Development of a data-driven COVID-19 prognostication tool to inform triage and step-down care for hospitalised patients in Hong Kong: A population-based cohort study**

**Protocol**

Methods

*Study population, data collection and definition*

All 1,037 confirmed cases as of 30 April 2020 were included in this study, and their observational data were traced up to 10 May. Data sources come from internal systems, CMS and eNID of NDORS. “NDORS” is an electronic platform for both HA and the Department of Health, to digitally report all suspected and confirmed statutory notifiable diseases and other infectious diseases of public health concern. A designated “eNID” module for reporting COVID-19 cases and clinical management was specifically built and interfaced to NDORS [8, 9]. This study was a retrospective data analysis on de-identified patient-based electronic medical records from CMS and eNID. Person-based information on history of chronic diseases was retrieved from an established HA’s chronic disease virtual registry that contains 25 pre-defined chronic diseases (Table 1). The registry was electronically built and based on all past CMS’ medical records using some operational counting and classification rules specific to each non-cancer disease, together with cancer cases sourced from Hong Kong’s Cancer Registry [11].

All 16 HA hospitals that treated COIVD-19 cases adopted a unified classification scheme on clinical conditions. The in-charge physicians would continuously update the condition status whenever the patient deteriorated or improved. The four clinical conditions are (1) *critical*: require intubation, or extracorporeal membrane oxygenation (ECMO) or in shock; (2) *serious*: require oxygen supplement of 3 litres or more per minute; (3) *stable*: with mild influenza-like illness (ILI) symptoms; (4) *satisfactory*: progressing well and likely to be discharged soon. Based on their clinical condition(s) along the entire clinical course, all cases were further amalgamated into three distinct outcome groups to delineate a grading for disease severity. The groups are “critical/serious”, “stable” and “satisfactory”. The patients must have ever been assessed as either “critical” or “serious” clinical condition for one or more days in the “critical/serious” group; otherwise being classified into the second group if ever assessed as “stable”. The remaining third group must be entirely assessed as “satisfactory” along the clinical course.

*Statistical analysis*

Descriptive statistical analyses were performed for the entire cohort with respect to epidemiological, clinical and laboratory data. In addition, chi-square test for categorical variables and Kruskal-Wallis test for continuous variables were performed to evaluate if there were any differences in a host of prognostic factors on day 1 and day 5 of hospital admission among the three outcome groups.  For development as well as evaluation of the model, the entire 1,037 study subjects were, proportional to outcome distribution, randomly split into a training dataset comprising of 829 subjects and a testing dataset of the remaining 208 subjects. The Extreme Gradient Boosting (XGBoost) model, which is a boosting decision tree machine learning framework allowing missing values for individual predictor variables, was developed to classify the training data into one of the three outcome groups, after taking into account a host of 30 predictors which included age, gender, chronic disease(s) history, 11 presenting symptoms as well as the worst clinical condition status, 15 biomarkers’ readings and Ct value of RT-PCR tests (based on E-Gene of the TIB MIBIOL kit) on day 1 and day 5 of admission. These predictors were chosen with reference to studies on COVID-19 [12–23] and SARS [24–26].

The XGBoost classifiers were trained and tuned using a 5-fold cross-validation approach with the training data to obtain the optimal hyperparameters [27]. All 30 features were ranked according to their relative importance using F-score, which guided a variable selection process to reduce the full model into a simpler one for practical application. The model output for each subject a probability across each of the three outcome groups, summing up to one, and with the highest probability group as the predicted outcome class. After applying the trained model to the testing dataset, it was then analysed in a 3X3 confusion matrix, based on which the model’s overall accuracy rate was computed. In view of the imbalanced outcome group distribution, macro-averaged and micro-averaged sensitivity and specificity of the three outcome groups were derived to evaluate the model performance. Decision tree of the simplified classifiers was also output for each outcome group. Partial dependency plots were output to depict the marginal effect of each model feature on the predicted outcome. (Appendix Figure 1) The XGBoost models were carried out by using Python’s XGboost version 1.10 whereas other statistical analyses by SAS version 9.4 software.