**METHODS**

*Available variables for analysis*

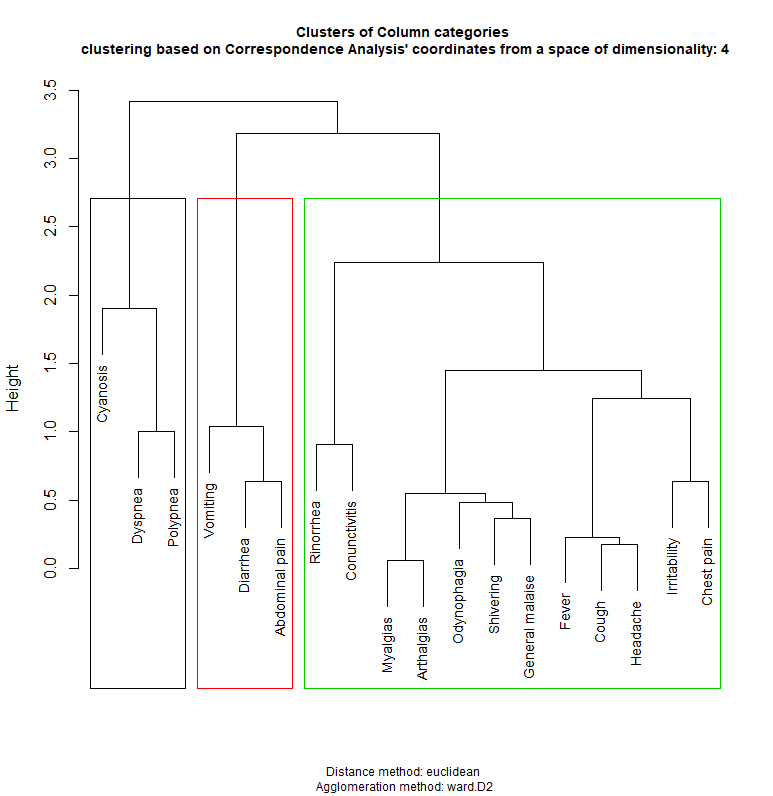
Available variables within the dataset include age, sex, occupation, nationality, and facility of case detection. Health information includes status of diabetes, obesity, chronic obstructive pulmonary disease (COPD), immunosuppression, pregnancy, arterial hypertension, cardiovascular disease (CVD), chronic kidney disease (CKD), HIV/AIDS, and asthma. Type and date of symptom onset, hospital admission, and death are available for all outpatient or hospitalized patients, as well as information regarding clinical diagnosis of pneumonia, ICU admission, and mechanic ventilatory support (MVS). Healthcare workers were defined as a composite of physicians, nurses, dentists, laboratory personnel and associated healthcare workers. Previous exposure to suspected viral cases was self-reported and was non-specific for known exposure to SARS-CoV-2 positive cases.

*Correspondence analysis and symptom clustering*

To investigate how the studied symptoms clustered in patients with confirmed SARS-CoV-2, we performed correspondence analysis on the evaluated columns for dimensionality reduction and identified the ideal number of components using the elbow method to maximize explained variance. Using these components, we performed hyerarchical cluster analysis on column categories using coordinates from a high-dimensional space corresponding to a dimensionality equal to the selected components from correspondence analysis (Dim=4). Distance within the feature space was calculated using Euclidean distance and agglomeration was conducted using Ward’s minimum variance criterion. The optimal number of symptom clusters was selected using the maximum mean silhouette method using the *CAinterprTools*, *factoextra* and *FactoMineR R packages*.

*Definition of respiratory symptoms*

For symptom clustering, we identified a 4-dimension solution for the correspondence analysis, which comprised 35.7% of the explained variance by symptoms. After applying hierarchical clustering, we identified three main cluster symptoms as shown in **Supplementary Figure 1**.



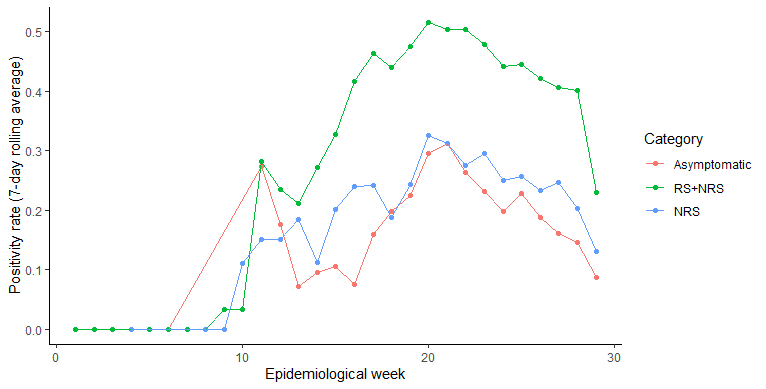
**Supplementary Figure 1.** Dendrogram for symptom clusters based on hierarchical clustering from correspondence analysis

The three symptom clusters were labeled as respiratory symptoms (RS), gastrointestinal symptoms and constitutional symptoms (Non-respiratory symptoms/NRS) [1,2]. Based on these definitions and on WHO criteria for suspected COVID-19 cases, we included cough and fever within the RS cluster [3].

*Statistical analyses*

To evaluate predictors of RS+NRS and NRS we fitted a mixed effects logistic regression models, including facility of treatment as a random effect to account for the variability in case distribution and treatment across healthcare facilities. We calculated time-to-event for hospital admission and lethality for each group and modeled these outcomes using mixed-effects Cox Proportional Risk regression, including as a frailty penalty treatment facility. All statistical analyses were performed using R version 3.6.4.

**Supplementary Figure 2.** Seven-day rolling average positivity rates for SARS-CoV-2 positivity in Mexico City, comparing RS+NRS, NRS and asymptomatic cases.



**SUPPLEMENTARY TABLES**

**Supplementary Table 1.** Cox Proportional Regression model with frailty penalty for prediction of hospitalization and mortality in pre-symptomatic/asymptomatic SARS-CoV-2 cases.

*Abbreviations*: COPD= Chronic obstructive pulmonary disease; CKD= Chronic kidney disease; CVD: Cardiovascular disease

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Parameters** | **β-coefficient** | **HR (95%CI)** | **p-value** |
| Severe COVID-19  c-stat=0.802 | Age | -0.033 | 0.97 (0.95-0.98) | <0.001 |
| CVD | 1.453 | 4.28 (1.27-14.43) | 0.019 |
| Previous viral exposure | -0.756 | 0.47 (0.26-0.86) | 0.015 |
| CKD | 1.871 | 6.49 (2.23-18.90) | <0.001 |
| Diabetes | 1.573 | 4.83 (1.90-12.26) | <0.001 |
| Mortality  c-stat=0.745 | Age >60 years | 0.040 | 1.041 (1.02-1.06) | <0.001 |
| CKD | 2.822 | 16.80 (5.53-51.01) | <0.001 |

**Supplementary Table 2.** Cox Proportional Regression model with frailty penalty for prediction of hospitalization and mortality in subjects with NRS only.

*Abbreviations*: COPD= Chronic obstructive pulmonary disease; CKD= Chronic kidney disease; NRS: Non-respiratory symptoms

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Parameters** | **β-coefficient** | **HR (95%CI)** | **p-value** |
| Severe COVID-19  c-stat=0.830 | Previous viral exposure | -1.210 | 0.30 (0.18-0.50) | <0.001 |
| Vomiting | 0.628 | 1.87 (1.13-3.11) | 0.015 |
| Rhinorrhea | -0.829 | 0.43 (0.24-0.81) | 0.009 |
| Healthcare workers | -1.419 | 0.24 (0.08-0.77) | 0.017 |
| Diabetes | 1.153 | 3.17 (1.96-5.13) | <0.001 |
| Hypertension | 0.671 | 1.96 (1.21-3.15) | 0.006 |
| Mortality  c-stat=0.887 | Previous viral exposure | -1.371 | 0.25 (0.14-0.48) | <0.001 |
| Age | 0.046 | 1.05 (1.03-1.06) | <0.001 |
| Diabetes | 0.771 | 2.16 (1.34-3.50) | 0.002 |
| COPD | 0.957 | 2.60 (1.08-6.27) | <0.001 |
| CKD | 1.294 | 3.65 (1.79-7.43) | <0.001 |

**SUPPLEMENTARY REFERENCES**

1. Han C et al. Digestive Symptoms in COVID-19 Patients With Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes. Am J Gastroenterol. 2020;115(6):916-923.
2. Jin et al. Epidemiological, Clinical and Virological Characteristics of 74 Cases of Coronavirus-Infected Disease 2019 (COVID-19) With Gastrointestinal Symptoms. Gut. 2020 Jun;69(6):1002-1009.
3. World Health Organization. Q&A On Coronaviruses (COVID-19). World Health Organization, 2020, www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses#:~:text=symptoms.