# How closely is COVID-19 related to HCoV, SARS, and MERS? : Clinical comparison of coronavirus infections and identification of risk factors influencing the COVID-19 severity using common data model (CDM)

Yeon Hee Kim[1], YeHee Ko[1], Soo Young Kim[1], Kwangsoo Kim[2]

1. Biomedical Research Institute, Seoul National University Hospital
2. Transdisciplinary Department of Medicine & Advanced Technology, Seoul National University Hospital

## Abstract

South Korea was one of the epicenters for both the 2015 MERS and 2019 COVID-19 outbreaks. However, there has been a lack of published literature, especially using the EMR records, that provides a comparative summary of the prognostic factors present in the coronavirus-derived diseases patients. Therefore, in this study, we aimed to compare and evaluate the distinct clinical traits between the patients of different coronaviruses, including the lesser pathogenic HCoV strains, SARS-CoV, MERS-CoV, and SARS-CoV-2. We also conducted observed the risk factors by the COVID severity to investigate the extent of resemblance in clinical features between the disease groups and to identify unique factor that may influence the prognosis of the COVID-19 patients. Here, we utilize the common data model (CDM), which is the database that houses the EMR records transformed into the common format to be used by the multiple institutions. For the comparative analyses between the disease groups, we used independent t-test, Scheffe post-hoc test, and Games-howell post-hoc test and for the continuous variables, chi-square test and Fisher's exact test. Based on the analyses, we selected the variables with p-values less than 0.05 to predict COVID-19 severity by nominal logistic regression with adjustments to age and gender. From the study, we observed diabetes, cardio and cerebrovascular diseases, cancer, pulmonary disease, gastrointestinal disease, and renal disease in all patient groups. Of all, the proportions of cancer patients were highest in all groups with no statistical significance. Most interestingly, we observed a high degree of clinical similarity between the COVID-19 and SARS patients with more than 50% of measured clinical variables to show statistical similarities between two groups. Our research reflects the great significance within the bioinformatics field that we were able to effectively utilize the integrated CDM to reflect real-world challenges in the context of coronavirus. We expect the results from our study to provide clinical insights that can serve as predicator of risk factors from the future coronavirus outbreak as well as the prospective guidelines for the clinical treatments.

## Background

COVID-19 is a global pandemic that has caused more than a million deaths and nearly 50 million cases since its outbreak[1]. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also known as the coronavirus disease 2019 (COVID-19) is a novel respiratory viral disease caused by the SARS-CoV-2 virus that was first discovered in Wuhan, China, in December of 2019[2]. Current evidence suggests the virus transmits via respiratory droplets, the aerosols, from cough, sneeze, speech, and heavy breaths yet we still don't know whether these aerosols persist in the air for a prolonged period[2-3]. However, it is not the first time the world is experiencing the health threats from the coronaviruses. Coronavirus infections in humans were first reported in the 1960s. They are the RNA viruses that are commonly present in bats and belong to the family *Coronaviridae.* The family comprises four genera, *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*, of which the two, *Alphacoronavirus* and *Betacoronavirus* are known to cause respiratory infections in humans. Including the COVID-19, there are seven human coronaviruses, including HCoV229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV, and MERS-CoV. While the HCoV strains cause mild upper respiratory diseases, more pathogenic strains including the SARS-CoV, MERS-CoV, and SARS-CoV-2 cause severe respiratory symptoms and complications that may lead to death[4-6]. South Korea was no exception in facing the pandemic as one of the epicenters for both the 2015 MERS and 2019 COVID-19 outbreaks. However, there has been a lack of published literature that provides a comprehensive and comparative summary of the prognostic factors present in the coronavirus-derived diseases patients in South Korea. In the light of emerging and re-emerging viral transmissions, it is utterly important to collect a comprehensive set of data upon which the robust control measures can be established. Therefore, in this study, we aim to compare and evaluate the distinct clinical traits between the patients of different coronaviruses, including the HCoV strains, SARS-CoV, MERS-CoV, and SARS-CoV-2. Further, we aim to conduct clinical characterization of the COVID-19 patients to observe the risk factors by severity of the disease with an attempt to compare them with the identified risk factors of all other coronavirus infections in order to see whether any common or distinct trait observed from the other coronavirus diseases significantly influence the prognosis of the COVID-19 patients. The study utilizes the common data model (CDM), which is the database that houses the EMR records transformed into the common format to be used by the multiple institutions for the research purposes[7]. With the use of real-world data, we expect to present valuable insights on clinical variables of importance on the COVID-19 as well as the degree of resemblance between the coronaviruses.

## Methods

### I. Data collection and definition

We utilized the CDM within the Seoul National University Hospital located in Seoul, South Korea. The data collection period is from October 15th of 2004 to July 31st of 2020. Without any restriction on age and gender, we collected the records of symptoms, comorbidities, and laboratory test results of patients diagnosed with HCoV229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1 (HCoVs), SARS-CoV, MERS-CoV, and SARS-CoV-2 infections. During the

data mining process, we categorized all individual diagnoses by the disease types with references to the ICD codes and selected the laboratory measurement variables primarily based on the literature review. We excluded any variable that had null or sparse patient records. We divided the COVID-19 patients by disease severity based on the criteria from the published World Health Organization (WHO)[7]. Among the COVID-19 confirmed patients, those who experienced mild cold-like symptoms with no pneumonia were classified as mild, whereas the patients who showed any additional clinical presentation of pneumonia were classified as non-mild.

### II. Statistical analysis

Continuous variables were compared by the independent t-test, Scheffe post-hoc test in case of homogeneity within the variance, and Games-howell post-hoc test. They were expressed by the mean values with ± standard deviations. For categorical variables, we conducted chi-square test or Fisher's exact test to indicate whether presence of conditions differed across disease groups. Based on the analyses, we selected the variables with p-values less than 0.05 to predict COVID-19 severity by logistic regression with adjustments to age and gender[8-12]. All statistical analyses were carried using R version 3.6.2.

## Results

### I. Comparative analyses between the coronavirus infections

*Demographic and clinical characteristics*

For the study, we collected the records of 2840 COVID patients, 67 MERS patients, 39 SARS patients, and 81 other HCoV positive patients. Table 1 shows the summary of clinical characteristics of the COVID, MERS, SARS, and HCoV patients. Among the COVID patients, the mean age was 51.8 ± 26.0 and 1457 (51.3%) were males. Among the MERS patients, the mean age was 36.9 ± 23.4 and 35 (52.2%) were males whereas among the SARS and HCoV patients, the mean age of patients was 22.8 ± 28.9 and 7.2 ± 11.9 and 27 (69.2%) and 42 (51.9%) were males, respectively. For comorbidities, diabetes, cancer, pulmonary disease, gastrointestinal disease, and renal disease were present in all groups. From the comparison between the COVID and MERS patients, the COVID patients experienced significantly more of the diabetes (4.4% vs. 1.1%, $p<0.001$), cancer (29.5% vs. 3.0%, $p<0.001$), and gastrointestinal diseases (14.4% vs. 1.5%, $p < 0.01$) compared to the MERS group. However, SARS patients, compared to the COVID patients, were observed to experience more cerebrovascular (12.8% vs. 1.1%, $p < 0.001$), pulmonary (61.5% vs. 8.4%, $p < 0.001$), and renal conditions (17.9% vs. 6.4%, $p < 0.01$), whereas the HCoV group had significantly higher proportion of patients with pulmonary diseases (48.1% vs. 8.4%, $p < 0.001$) and musculoskeletal diseases (9.9% vs. 2.2%, $p < 0.001$) than the COVID group. For symptoms, patients in all disease groups reported fever, cough, dyspnea, gastrointestinal symptoms, and upper respiratory infections. Compared to the COVID patients, MERS and HCoV patients presented fever (79.1% vs. 33%, $p < 0.001$ and 60.5% vs. 33%, $p < 0.001$, respectively) and upper respiratory infections (28.4% vs 0.8%, $p < 0.001$ and 37% vs 0.8%, $p < 0.001$) more frequently. Except for dyspnea and sore throat, SARS patients compared to the

COVID patients did not show any statistically significant difference in their symptom presentations.

| Characteristic | COVID-19 (n=2840) | MERS (n=67) | p-val | SARS (n=39) | p-val | HCoV (n=81) | p-val |
|---|---|---|---|---|---|---|---|
| **Age** | 51.8 ± 26.0 | 36.9 ± 23.4 | <.001 | 22.8 ± 28.9 | 0.325 | 7.2 ± 11.9 | <.001 |
| 0-19 | 451 (15.9) | 19 (28.4) | <0.01 | 13 (33.3) | <0.01 | 72 (88.9) | <0.001 |
| 20-39 | 569 (20.0) | 8 (11.9) | 0.101 | 3 (7.7) | 0.067 | 8 (9.9) | <0.05 |
| 40-59 | 593 (20.9) | 3 (4.5) | <0.001 | 8 (20.5) | 0.955 | 2 (2.5) | <0.001 |
| 60-79 | 920 (32.4) | 5 (7.5) | <0.001 | 14 (35.9) | 0.643 | 5 (6.2) | <0.001 |
| 80+ | 307 (10.8) | 0 (0.0) | NA | 5 (12.8) | 0.688 | 0 (0.0) | NA |
| **Gender** | | | | | | | |
| Male | 1457 (51.3) | 35 (52.2) | 0.879 | 27 (69.2) | 0.026 | 42 (51.9) | 0.922 |
| **Comorbidities** | | | | | | | |
| Total | 1567 (55.2) | 9 (13.4) | <0.001 | 34 (87.2) | <0.001 | 52 (64.2) | 0.107 |
| Diabetes | 124 (4.4) | 1 (1.5) | <0.001 | 3 (7.7) | 0.246 | 1 (1.2) | 0.262 |
| Cardiovascular disease | 226 (8.0) | 0 (0.0) | NA | 6 (15.4) | 0.091 | 12 (14.8) | 0.037 |
| Cerebrovascular disease | 33 (1.1) | 0 (0.0) | NA | 5 (12.8) | <0.001 | 2 (2.5) | 0.253 |
| Malignancy | 837 (29.5) | 2 (3.0) | <0.001 | 8 (20.5) | 0.222 | 9 (11.1) | <0.001 |
| Hepatic disase | 132 (4.6) | 0 (0.0) | NA | 5 (12.8) | 0.017 | 1 (1.2) | 0.181 |
| Pulmonary disease | 238 (8.4) | 5 (7.5) | 0.789 | 24 (61.5) | <0.001 | 39 (48.1) | <0.001 |
| Gastrointestinal diseases | 410 (14.4) | 1 (1.5) | <0.01 | 9 (23.1) | 0.1286 | 6 (7.4) | 0.074 |
| Renal disease | 182 (6.4) | 4 (6.0) | 0.885 | 7 (17.9) | <0.01 | 3 (3.7) | 0.485 |
| Thyroidal disease | 30 (1.1) | 0 (0.0) | NA | 1 (2.6) | 0.346 | 2 (2.5) | 0.222 |
| Musculoskeletal disease | 62 (2.2) | 0 (0.0) | NA | 1 (2.6) | 0.581 | 8 (9.9) | <0.001 |
| **Symptoms** | | | | | | | |
| Total | 1290 (45.4) | 60 (90.0) | <0.001 | 15 (38.5) | 0.386 | 58 (71.6) | <0.001 |
| Fever | 939 (33) | 53 (79.1) | <0.001 | 9 (23.1) | 0.036 | 49 (60.5) | <0.001 |
| Cough | 102 (3.6) | 8 (11.9) | <0.001 | 3 (7.7) | 0.168 | 6 (7.4) | 0.107 |
| Dyspnea | 130 (4.6) | 2 (3.0) | 0.768 | 6 (15.4) | <0.01 | 5 (6.2) | 0.500 |
| Chest pain | 63 (2.2) | 0 (0.0) | NA | 1 (2.6) | 0.586 | 0 (0.0) | NA |
| Sore throat | 51 (1.8) | 3 (4.5) | 0.127 | 0 (0.0) | NA | 1 (1.2) | 1.00 |
| Gastrointestinal symptoms | 193 (6.8) | 1 (1.5) | 0.129 | 2 (5.1) | 1.00 | 6 (7.4) | 0.829 |
| Upper respiratory infection | 23 (0.8) | 19 (28.4) | <0.001 | 2 (5.1) | <0.05 | 30 (37.0) | <0.001 |

*Table 1. Summary of clinical characteristics between the HCoVs, SARS, MERS, COVID-19 groups. Data are presented as the mean ± standard deviation for the continuous variables, otherwise the number of patients and percentage.*

## *Laboratory findings*

Laboratory findings of each disease group are summarized in Table 2. In complete blood cell counts, lymphocyte count (38.90 ± 27.00 vs. 18.40 ± 17.10, $p < 0.001$) and eosinophil count (4.38 ± 6.09 vs. 1.37 ± 2.61, $p < 0.01$) were higher in HCoV group than the COVID group whereas hemoglobin (13.20 ± 2.12 vs. 10.90 ± 2.47, $p < 0.001$) and hematocrit (39.70 ± 5.59 vs. 33.00 ± 7.24, $p < 0.001$) levels were higher in the MERS group. In liver function measurements, COVID patients had a higher level of total bilirubin (1.39 ± 2.34 vs. 0.69 ±

0.47, $p < 0.001$ and 1.39 ± 2.34 vs. 0.67 ± 0.33, $p < 0.001$, respectively) compared to the HCoV and MERS patients, whereas the COVID patients had a higher level of albumin compared to the SARS group (3.53 ± 0.74 vs. 2.94 ± 0.83, $p < 0.001$) but lower level than the MERS group (4.17 ± 0.52 vs. 3.53 ± 0.74, $p < 0.001$). Within the kidney function, levels of blood urea nitrogen and creatinine were significantly higher among the COVID patients compared to the HCoV group (23.30 ± 20.20 vs. 12.90 ± 11.80, $p< 0.001$ and 1.33 ± 1.77 vs. 0.58 ± 0.64, $p< 0.1$, respectively). There were no observed statistical differences between the groups in the coagulation function.

| Prognostic factors (unit) | COVID | HCoV | p-val | MERS | p-val | SARS | p-val | F | p-val |
|---|---|---|---|---|---|---|---|---|---|
| **Complete blood count** | | | | | | | | | |
| Hb (g/dL) | 10.90 ± 2.47 | 11.10 ± 2.19 | 0.91 | 13.20 ± 2.12 | <.001 | 10.80 ± 2.92 | 1.00 | F(3, 3443.00) = 13.51 | <.001 |
| Hct (%) | 33.00 ± 7.24 | 32.40 ± 6.56 | 0.85 | 39.70 ± 5.59 | <.001 | 31.50 ± 8.37 | 0.65 | F(3, 87.93) = 21.62 | <.001 |
| Lymphocyte (%) | 18.40 ± 17.10 | 38.90 ± 27.00 | <.001 | 18.90 ± 11.70 | 1.00 | 19.30 ± 15.10 | 0.98 | F(3, 83.41) = 12.83 | <.001 |
| RBC ($10^6$/μL) | 3.58 ± 0.85 | 3.80 ± 0.75 | <0.05 | 4.54 ± 0.73 | <.001 | 3.46 ± 0.99 | 0.87 | F(3, 87.71) = 27.86 | <.001 |
| WBC ($10^3$/μL) | 10.12 ± 15.21 | 5.64 ± 6.02 | 0.06 | 8.10 ± 4.11 | 0.84 | 9.19 ± 6.94 | 0.98 | F(3,3416.00) = 2.82 | <0.05 |
| Eosinophil (%) | 1.37 ± 2.61 | 4.38 ± 6.09 | <.01 | 1.20 ± 1.32 | 0.85 | 2.13 ± 4.71 | 0.76 | F(3, 76.25) = 5.27 | <.01 |
| PCT (%) | 0.20 ± 0.11 | 0.19 ± 0.14 | 0.83 | 0.20 ± 0.09 | 1.00 | 0.17 ± 0.09 | 0.14 | F(3, 82.11) = 1.81 | 0.15 |
| PLT ($10^3$/μL) | 192.00 ± 120.00 | 179.00 ± 153.00 | 0.85 | 217.00 ± 100.00 | 0.38 | 176.00 ± 114.00 | 0.82 | F(3, 86.59) = 1.35 | 0.26 |
| **Liver function** | | | | | | | | | |
| Albumin (g/dL) | 3.53 ± 0.74 | 3.76 ± 0.60 | 0.06 | 4.17 ± 0.52 | <.001 | 2.94 ± 0.83 | <.001 | F(3, 80.81) = 32.98 | <.001 |
| Alkaline phosphatase (U/L) | 152.30 ± 191.20 | 132.40 ± 88.80 | 0.92 | 90.00 ± 56.60 | 0.21 | 90.00 ± 76.20 | 0.46 | F(3, 3023.00) = 2.49 | 0.06 |
| Bilirubin, total (mg/dL) | 1.39 ± 2.34 | 0.69 ± 0.47 | <.001 | 0.67 ± 0.33 | <.001 | 3.15 ± 6.90 | 0.24 | F(3, 123.52) = 50.32 | <.001 |
| AST (U/L) | 85.50 ± 432.40 | 38.40 ± 29.90 | 0.88 | 33.50 ± 34.90 | 0.89 | 108.20 ± 261.10 | 0.99 | F(3, 3050.00) = 0.46 | 0.71 |
| ALT (U/L) | 55.10 ± 167.70 | 34.00 ± 31.30 | | 27.70 ± 24.80 | | 92.80 ± 169.50 | | | |
| **Kidney function** | | | | | | | | | |
| BUN (mg/dL) | 23.30 ± 20.20 | 12.90 ± 11.80 | <.001 | 15.80 ± 13.80 | <.01 | 28.30 ± 33.00 | 0.6 | F(3, 97.99) = 19.39 | <.001 |
| Creatinine (mg/dL) | 1.33 ± 1.77 | 0.58 ± 0.64 | <.01 | 1.07 ± 1.12 | 0.8 | 1.30 ± 0.94 | 1.00 | F(3, 3390.00) = 4.16 | <.01 |
| eGFR | 78.60 ± 42.70 | 70.80 ± 36.90 | 0.93 | 105.50 ± 61.10 | <.01 | 107.50 ± 48.10 | <0.05 | F(3, 2967.00) = 8.34 | <.001 |
| **Coagulation function** | | | | | | | | | |
| aPTT (sec) | 31.10 ± 12.68 | 31.20 ± 2.86 | 1.00 | 31.90 ± 4.65 | 1.00 | 33.70 ± 7.58 | 0.88 | F(3, 2219.00) = 0.24 | 0.87 |
| Fibrinogen (mg/dL) | 400.00 ± 134.00 | 408.00 ± 132.00 | 1.00 | 454.00 ± 123.00 | 0.39 | 459.00 ± 152.00 | 0.39 | F(3, 2186.00) = 2.00 | 0.11 |
| PT (sec) | 47.40 ± 36.50 | 49.30 ± 38.00 | 1.00 | 49.00 ± 38.00 | 1.00 | 49.00 ± 38.10 | 1.00 | F(3, 4502.00) = 0.06 | 0.98 |

*Table 2. Summary of laboratory findings between the disease groups. Data are presented as the mean ± standard deviation for the continuous variables, otherwise the number of patients and percentage. Hb: Hemoglobain, Hct: Hematocrit, RBC: Red blood cell, WBC: White blood cell, PLT: Platelet, PCT: Procalcitonin, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, BUN: Blood urea nitrogen, eGFR: Estimated glomerular filtration rate, aPTT: Partial thromboplastin time, PT: Prothrombin time.*

## II. Comparative analyses of the COVID-19 severity

*Demographic and clinical characteristics*

After dividing the COVID patients by the WHO disease severity criteria (ref), there were 2596 patients in the mild group and 159 patients in the non-mild group. Table 3 shows the summary of clinical characteristics of the non-mild and non-mild patients. Within the COVID positive patients, 2596 people belonged to the mild group (91.4%) and 244 (8.6%) belonged to the non-mild group. In the mild group, the mean age was 47.9 ± 25.6 and 1298 were males. In the non-mild group, the mean age was 67.0 ± 21.7 and 159 were males. There were 1567 COVID patients with the comorbidities, of which 1327 and 240 patients belonged to the mild and non-mild group, respectively. By comparison, non-mild COVID patients presented more comorbidities upon their hospital admissions (98.3% vs. 51.1%, $p < 0.001$). As shown in the Table 3, patients in the non-mild group reported to have more cerebrovascular disease (4.9% vs. 0.8%, $p < 0.001$), pulmonary disease (36.9% vs. 5.7%, $p < 0.001$), renal disease (16.0% vs. 5.5 %, $p < 0.001$), whereas more patients in the mild-group experienced hepatic disease than the non-mild group (5.0% vs. 0.8 %, $p < 0.01$). For symptoms, non-mild group had the higher proportion of people reporting dyspnea (20.1% vs. 3.1%, $p < 0.001$), chest pain (5.7% vs. 1.9%, $p < 0.001$), and upper respiratory infections (2.5% vs. 0.7%, $p < 0.001$), whereas the mild group experienced more fever (34.5% vs. 18.0%, $p < 0.001$) compared to the non-mild group.

| Characteristic | COVID-19 (n=2840) | Mild (n=2596) | Non-mild (n=244) | p-val |
|---|---|---|---|---|
| **Gender** | | | | |
| Male | 1457 (51.3) | 1298 (50.0) | 159 (65.2) | <0.001 |
| **Age** | 51.8 ± 26.0 | 47.9 ± 25.6 | 67.0 ± 21.7 | <.001 |
| 0-19 | 451 (15.9) | 432 (16.6) | 19 (7.8) | <0.001 |
| 20-39 | 569 (20.0) | 561 (21.6) | 8 (3.3) | <0.001 |
| 40-59 | 593 (20.9) | 565 (21.8) | 28 (11.5) | <0.001 |
| 60-79 | 920 (32.4) | 802 (30.9) | 118 (48.4) | <0.001 |
| 80+ | 307 (10.8) | 236 (9.1) | 71 (29.1) | <0.001 |
| **Comorbidities** | | | | |
| Total | 1567 | 1327 | 240 | <0.001 |
| Diabetes | 124 (4.4) | 115 (4.4) | 9 (3.7) | 0.588 |
| Cardiovascular disease | 226 (8.0) | 194 (7.5) | 32 (13.1) | <0.01 |
| Cerebrovascular disease | 33 (1.1) | 21 (0.8) | 12 (4.9) | <0.001 |
| Malignancy | 837 (29.5) | 758 (29.2) | 79 (32.4) | 0.298 |
| Hepatic disedase | 132 (4.6) | 130 (5.0) | 2 (0.8) | <0.01 |
| Pulmonary disease | 238 (8.4) | 148 (5.7) | 90 (36.9) | <0.001 |
| Gastrointestinal diseases | 410 (14.4) | 390 (15.0) | 30 (12.3) | 0.251 |
| Renal disease | 182 (6.4) | 143 (5.5) | 39 (16.0) | <0.001 |
| Thyroidal disease | 30 (1.1) | 30 (1.2) | 0 (0.0) | NA |
| Musculoskeletal disease | 62 (2.2) | 55 (2.1) | 7 (2.9) | 0.443 |
| **Symptoms** | | | | |
| Total | 1290 (45.4) | 1226 (47.2) | 64 (26.2) | <0.001 |

| | | | | |
|---|---|---|---|---|
| Fever | 939 (33) | 895 (34.5) | 44 (18.0) | <0.001 |
| Cough | 102 (3.6) | 97 (3.7) | 5 (2.0) | 0.176 |
| Dyspnea | 130 (4.6) | 81 (3.1) | 49 (20.1) | <0.001 |
| Chest pain | 63 (2.2) | 49 (1.9) | 14 (5.7) | <0.001 |
| Sore throat | 51 (1.8) | 51 (2.0) | 0 (0.0) | NA |
| Gastrointestinal symptoms | 193 (6.8) | 185 (7.1) | 8 (3.3) | <0.05 |
| Upper respiratory infection | 23 (0.8) | 17 (0.7) | 6 (2.5) | <0.001 |

*Table 3. Summary of clinical characteristics between the COVID-19 severity groups. Data are presented as the mean ± standard deviation for the continuous variables, otherwise the number of patients and percentage.*

## *Laboratory findings*

Laboratory findings of each severity group are summarized in Table 4. Compared to the mild group, non-mild group showed statistically higher levels of white blood cell count (11.13 ± 9.44 vs. 9.99 ± 15.83, $p < 0.05$), eosinophil count (39.70 ± 46.42 vs. 34.65 ± 45.64, $p < 0.05$), procalcitonin (24.14 ± 14.18 vs. 22.54 ± 14.65, $p < 0.05$), and platelet (208.26 ± 120.78 vs., 190.18 ± 120.26, $p < 0.001$). While in liver function measurements, the mild group showed higher levels of albumin (3.56 ± 0.74 vs. 3.30 ± 0.64, $p < 0.001$) and total bilirubin (1.45 ± 2.46 vs. 0.97 ± 1.20 , $p < 0.001$), they had lower levels of blood nitrogen urea (45.01 ± 21.11 vs. 50.31 ± 30.10, $p < 0.001$), creatinine (91.86 ± 86.73 vs. 123.59 ± 117.71, $p < 0.001$), and fibrinogen (392.81 ± 135.42 vs. 449.46 ± 117.01, $p < 0.001$) compared to the non-mild group.

| **Prognostic factors (unit)** | **Total** | **Mild** | **Non-mild** | **p-val** |
|---|---|---|---|---|
| **Complete blood count** | | | | |
| Hb (g/dL) | 10.87 ± 2.47 | 10.91 ± 2.47 | 10.58 ± 2.40 | <0.05 |
| Hct (%) | 33.03 ± 7.24 | 33.10 ± 7.25 | 32.58 ± 7.19 | 0.181 |
| Lymphocyte (%) | 262.28 ± 186.72 | 262.44 ± 186.67 | 261.19 ± 187.30 | 0.904 |
| RBC ($10^6$/μL) | 3.58 ± 0.85 | 3.59 ± 0.85 | 3.47 ± 0.82 | <0.01 |
| WBC ($10^3$/μL) | 10.12 ± 15.21 | 9.99 ± 15.83 | 11.13 ± 9.44 | <0.05 |
| Eosinophil (%) | 35.29 ± 45.77 | 34.65 ± 45.64 | 39.70 ± 46.42 | <0.05 |
| PCT (%) | 22.74 ± 14.60 | 22.54 ± 14.65 | 24.14 ± 14.18 | <0.05 |
| PLT ($10^3$/μL) | 192.36 ± 120.45 | 190.18 ± 120.26 | 208.26 ± 120.78 | <0.01 |
| **Liver function** | | | | |
| Albumin (g/dL) | 3.53 ± 0.74 | 3.56 ± 0.74 | 3.30 ± 0.64 | <0.001 |
| Alkaline phosphatase (U/L) | 278.64 ± 165.86 | 277.97 ± 163.85 | 283.24 ± 179.08 | 0.593 |
| Bilirubin, total (mg/dL) | 1.39 ± 2.34 | 1.45 ± 2.46 | 0.97 ± 1.20 | <0.001 |
| AST (U/L) | 142.33 ± 72.62 | 142.09 ± 73.01 | 144.01 ± 70.01 | 0.622 |
| ALT (U/L) | 123.83 ± 76.56 | 123.94 ± 76.44 | 123.11 ± 77.52 | 0.846 |
| **Kidney function** | | | | |
| BUN (mg/dL) | 45.69 ± 31.02 | 45.01 ± 31.11 | 50.31 ± 30.10 | <0.001 |
| Creatinine (mg/dL) | 95.94 ± 91.90 | 91.86 ± 86.73 | 123.59 ± 117.71 | <0.001 |
| eGFR | 688.37 ± 385.65 | 692.25 ± 387.57 | 663.38 ± 372.58 | 0.158 |
| **Coagulation function** | | | | |
| aPTT (sec) | 31.14 ± 12.68 | 31.12 ± 12.46 | 31.22 ± 14.06 | 0.909 |
| Fibrinogen (mg/dL) | 400.28 ± 134.49 | 392.81 ± 135.42 | 449.46 ± 117.01 | <0.001 |
| PT (sec) | 47.44 ± 36.49 | 47.70 ± 36.86 | 45.71 ± 33.97 | 0.192 |

*Table 4. Summary of laboratory findings between the COVID severity groups. Data are presented as the mean ± standard deviation for the continuous variables, otherwise the number of patients and percentage. Hb: Hemoglobain, Hct: Hematocrit, RBC: Red blood cell, WBC: White blood cell, PLT: Platelet, PCT: Procalcitonin, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, BUN: Blood urea nitrogen, eGFR: Estimated glomerular filtration rate, aPTT: Partial thromboplastin time, PT: Prothrombin time*

## III. Logistic regression

In order to observe the extent of common clinical characteristics among disease groups that may uniquely affect the prognosis of COVID-19 patients, we conducted nominal logistic regression with the variables that showed statistical significance with p values less than 0.05. Table 5 below shows the results from the regression that is adjusted for both age and gender. The model below demonstrated that cerebrovascular disease (OR: 5.34, 95% CI: 1.06-26.27, $p < 0.05$), pulmonary disease (OR: 925.00, 95% CI: 426.15-2351.15, $p < 0.001$), renal disease (OR: 7.70, 95% CI: 3.35-19.28, $p < 0.001$) , and increased eosinophil count (OR: 4.11, 95% CI: 1.21, 14.91, $p < 0.05$) were statistically associated with the COVID patients progressing towards the more severe stage whereas the increased level of bilirubin showed statistical association with the lesser severe form of COVID-19 (OR: 0.57, 95% CI: 0.33-0.96, $p < 0.05$).

| **Prognostic factors** | **OR (95% CI)** | **p-val** |
|---|---|---|
| Age | 1.03 (0.98, 1.08) | 0.224 |
| 0-19 | 2.20 (0.05, 89.33) | 0.677 |
| 20-39 | 1.10 (0.06, 18.59) | 0.948 |
| 40-59 | 0.45 (0.09, 2.25) | 0.333 |
| 60-79 | 0.41 (0.17, 1.01) | <0.01 |
| **Male, Sex** | 1.54 (0.93, 2.54) | 0.09 |
| **Cardiovascular disease** | 1.61 (0.81, 3.29) | 0.18 |
| **Cerebrovascular disease** | 5.34 (1.06, 26.27) | <0.05 |
| **Hepatic disease** | 0.13 (0.03, 0.51) | <0.01 |
| **Pulmonary disease** | 925.00 (426.15, 2351.15) | <0.001 |
| **Renal disease** | 7.70 (3.35, 19.28) | <0.001 |
| **Albumin (Normal range)** | Reference | |
| Low | 1.58 (0.93, 2.69) | 0.088 |
| **Total bilirubin (Normal range)** | Reference | |
| High | 0.57 (0.33, 0.96) | <0.05 |
| **Hb (Normal range)** | Reference | |
| High | 0.48 (0.06, 4.42) | 0.49 |
| Low | 0.75 (0.34, 1.64) | 0.471 |
| **RBC (Normal range)** | Reference | |
| High | 0.18 (0.01, 2.43) | 0.222 |
| Low | 1.19 (0.54, 2.63) | 0.665 |
| **WBC (Normal range)** | Reference | |
| High | 1.55 (0.94, 2.59) | 0.087 |
| Low | 2.11 (1.00, 4.55) | 0.053 |
| **Eosinophil (Normal range)** | Reference | |

| | | |
|---|---|---|
| High | 4.11 (1.21, 14.91) | <0.05 |
| Low | 1.05 (0.65, 1.68) | 0.844 |
| **BUN (Normal range)** | Reference | |
| High | 0.83 (0.43, 1.59) | 0.573 |
| Low | 1.03 (0.24, 4.69) | 0.964 |
| **Creatinine (Normal range)** | Reference | |
| High | 1.03 (0.54, 1.97) | 0.92 |
| Low | 1.89 (0.93, 3.92) | 0.083 |

*Table 5. Nominal logistic regression adjusted for age and gender. Hb: Hemoglobain, RBC: Red blood cell, WBC: White blood cell, BUN: Blood urea nitrogen*

## Discussion

COVID-19 is the third coronavirus-derived disease that has fatally affected the global population. As such, it is noteworthy that the coronavirus may continue to be the viral source for a number of morbidities and mortality. With our aim to then comprehensively understand the impact of coronavirus, we examined the characteristics and associations between the COVID-19, HCoV, SARS and MERS patients within South Korea, using common data model (CDM).

From all disease groups, we observed patients with diabetes, cardio and cerebrovascular diseases, cancer, pulmonary disease, gastrointestinal disease, and renal disease. Among the comorbidities, proportion of cancer patients were highest in all groups with no statistical significance. Interestingly, we observed the most similarities in clinical features between the COVID-19 and SARS patients. Out of 17 conditions including the comorbidities and symptoms, the two groups showed no statistical differences in 12 conditions (71%). Further, within the laboratory findings, both groups presented statistical similarities in 17 out of 19 measurements (89%). Such similarity in clinical characteristics between the SARS-CoV-2 and SARS-CoV-1, which also been supported by Petrosillo et al.[5], maybe explained by their common ancestor, the bat coronavirus HKU9-1[13]. From both the comparative analyses between the disease groups and COVID-19 severity groups, cerebrovascular disease, hepatic disease, pulmonary disease, and renal disease were showed statistical significance. Thus, when applying those factors along with the selected measurement variables, we saw that cerebrovascular disease, pulmonary disease, renal disease, and increased eosinophil count were associated with the worse prognosis of COVID-19. Studies have found that COVID-19 infection can accelerate the development of cerebrovascular disease; previously published autopsy results of COVID-19 patients showed hyperemic and edematous brain tissue with some degenerated neurons[14-15]. With such findings, Avula et al. suggested the possibility of hypercoagulation leading to macro and micro thrombi formation in the vessels during the COVID-19 infection[16]. Further, our observations of patients' existing conditions, especially cancer, reflect the susceptibility of immunosuppressed patients. Other coexisting conditions in multiple organs including kidney and GI tract may be attributed to the particular pathway— upon its entry to human body cells, the spike protein of SARS-CoV-2 bind to the angiotensin-converting enzyme 2 (ACE2) receptors, which are also expressed in heart, lungs,

kidneys, and other organs[17-18]. Thus, our observations of a wide range of comorbidities among the COVID-19 patients may be explained by the aforementioned mechanism.

Our research has a great significance that we effectively utilized the big data from the integrated model, the CDM, within one of the biggest national hospitals in South Korea. CDM research is conducted worldwide for the vitalization of medical research[19]. With the use of commonly formatted EMR records, our research successfully reflects the real-world challenges with the coronavirus. Thus, we expect the results from our study to provide clinical insights that can be used as basis for predicting the prospective clinical representations by yet another coronavirus-derived illness in near future.

## References


1. World Health Organization - WHO Coronavirus Disease (COVID-19) Dashboard [Website]. Accessed on October 15th, 2020. https://covid19.who.int/

2. Chaw, S. M., Tai, J. H., Chen, S. L et al. (2020) The origin and underlying driving forces of the SARS-CoV-2 outbreak. Journal of Biomedical Science, 27(73). https://doi.org/10.1186/s12929-020-00665-8

3. Cui, J., Li, F., & Shi, Z. L. (2019) Origin and evolution of pathogenic coronaviruses. Nature reviews. Microbiology, 17(3), 181-192. https://doi.org/10.1038/s41579-018-0118-9 Jang, J., G., Hur, J., Choi, E., Y., Hong, K., S., Lee, W., & Ahn, J., H. (2020) Prognostic Factors for Severe Coronavirus Disease 2019 in Daegu, Korea. Journal of Korean Medical Science, 35(23), e209. https://doi.org/10.3346/jkms.2020.35.e209

4. J. V. J. Silva Ju´nior, T. R. R. Lopes, P. S. B. de Oliveira, R. Weiblen, & E. F. Flores. (2020) Issues on Coronavirus Disease 2019 (COVID-19) Pathogenesis. Viral Immunology, Online Ahead of Print. https://doi.org/10.1089/vim.2020.0059

5. Krishnamoorthy, S., Swain, B., Verma, R.S., and Gunthe, S., S. (2020) SARS-CoV, MERS-CoV, and 2019-nCoV viruses: an overview of origin, evolution, and genetic variations. VirusDisease. https://doi.org/10.1007/s13337-020-00632-9

6. N. Petrosillo, G. Viceconte, O. Ergonul, G. Ippolito, & E. Petersen. (2020) COVID-19, SARS and MERS: are they closely related? Clinical Microbiology and Infection, 26(6), 729-734. https://doi.org/10.1016/j.cmi.2020.03.026

7. OHDSI (Observational Health Data Sciences and Informatics): OMOP Common Data Model[Website] Accessed on October 20th, 2020. https://www.ohdsi.org/data-standardization/the-common-data-model/

8. Qaddoumi M, Al-Khamis Y, Channanath A, Tuomilehto J, Badawi D. (2019). The Status of Metabolic Control in Patients With Type 2 Diabetes Attending Dasman Diabetes Institute, Kuwait. *Frontiers in Endocrinology*. 10. doi:10.3389/fendo.2019.00412.

9. Brady SM, Burow M, Busch W, et al (2015). Reassess the t Test: Interact with All Your Data via ANOVA. *Plant Cell*. 27(8):2088-2094. doi:10.1105/tpc.15.00238

10. Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. (2019). Application of student's *t*-test, analysis of variance, and covariance. *Ann Card Anaesth*. 22(4):407-411. doi:10.4103/aca.ACA_94_19

11. Lamattina D, Berrozpe PE, Casas N, et al. (2019). Twice upon a time: The progression of canine visceral leishmaniasis in an argentinean city. *PLoS One*. 14(7). http://dx.doi.org.access.yonsei.ac.kr:8080/10.1371/journal.pone.0219395.

12. Obeso JM, García P, Martínez B, Arroyo-López FN, Garrido-Fernández A, Rodriguez A (2010). Use of logistic regression for prediction of the fate of Staphylococcus aureus in pasteurized milk in the presence of two lytic phages. *Appl Environ Microbiol*. 76(18):6038-6046. doi:10.1128/AEM.00613-10

13. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*. 63(3):457–460.

14. Mao L., Jin H., Wang M., Hu Y., Chen S., He Q., Chang J., Hong C., Zhou Y., Wang D., Miao X., Li Y., Hu B. (2019). Neurologic Manifestations of Hospitalized Patients with Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurol*. 77:683–690. doi: 10.1001/jamaneurol.2020.1127.

15. Altable, M., & de la Serna, J. M. (2020). Cerebrovascular disease in COVID-19: Is there a higher risk of stroke?. *Brain, behavior, & immunity – health*. 6(100092). https://doi.org/10.1016/j.bbih.2020.100092

16. Avula A., Nalleballe K., Narula N., Sapozhnikov S., Dandu V., Toom S., Glaser A., Elsayegh D. (2020). COVID-19 presenting as stroke. *Brain Behav. Immun*. 2020;87:115–119. doi: 10.1016/j.bbi.2020.04.077.

17. Na KR, Kim HR, Ham Y, Choi DE, Lee KW, Moon JY, Kim YS, Cheon S, Sohn KM, Kim J, Kim S, Jeong H, Jeon JW. (2020). Acute Kidney Injury and Kidney Damage in COVID-19 Patients. *J Korean Med Sci*. 35(28):e257. https://doi.org/10.3346/jkms.2020.35.e257

18. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. (2020). Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med* 14(2):185–192. doi: 10.1007/s11684-020-0754-0

19. Choi SA, Kim H, Kim S, Yoo S, Yi SY, Jeon YH, Hwang H, Kim KJ. (2020). Analysis of antiseizure drug-related adverse reactions from the electronic health record using the common data model. *Epilepsia*, *61(4)*, 610-616. https://doi.org/10.1111/epi.16472

# Appendix 1. Normal Range

| | Normal Range | | | Unit |
|---|---|---|---|---|
| | Pediatrics | | Adults | |
| **Albumin** | 3.5-5.2 | | | g/dL |
| **Alkaline phosphatase** | M: 40-130<br>F: 35-105 | | | U/L |
| **Bilirubin, total** | <1.2 | | | mg/dL |
| **GOT (AST)** | M: <40<br>F: <33 | | | U/L |
| **GPT (ALT)** | M: <41<br>F: <33 | | | U/L |
| **Hb** | 1-2 months | M: 10.2-12.7<br>F: 11.1-13.7 | M: 13-17<br>F: 12-16 | g/dL |
| | 2-6 months | M: 10.5-13.0<br>F: 10.7-13.4 | | |
| | 6 months – 2 yrs | M: 10.4-12.5<br>F: 10.8-12.6 | | |
| | 2 yrs – 6 yrs | M: 11.0-12.8<br>F: 11.1-12.9 | | |
| | 6 rs – 12 yrs | M: 11.0-13.3<br>F: 11.3-13.4 | | |
| | 12 yrs – 18 yrs | M: 11.0-14.3<br>F: 11.3-13.4 | | |
| **Hct** | 1-2 months | M: 29.1-36.6<br>F: 32.0-39.9 | M: 39-52<br>F: 36-48 | % |
| | 2-6 months | M: 30.5-37.7<br>F: 30.5-38.6 | | |
| | 6 months – 2 yrs | M: 30.5-36.4<br>F: 30.9-36.4 | | |
| | 2 yrs – 6 yrs | M: 31.5-36.8<br>F: 31.8-37.0 | | |
| | 6 rs – 12 yrs | M: 31.5-38.0<br>F: 32.3-38.3 | | |
| | 12 yrs – 18 yrs | M: 31.4-41.0<br>F: 32.1-38.7 | | |
| **Lymphocyte** | 1-2 months | M: 12.0-68.0<br>F: 16.0-68.0 | 17.0-46.0 | % |
| | 2-6 months | M: 16.0-68.0<br>F: 15.0-68.0 | | |
| | 6 months – 2 yrs | M: 15.0-67.0<br>F: 13.0-70.0 | | |
| | 2 yrs – 6 yrs | M: 11.0-54.0<br>F: 11.0-59.0 | | |
| | 6 rs – 12 yrs | M: 8.0-45.0<br>F: 10.0-47.0 | | |
| | 12 yrs – 18 yrs | M: 8.0-41.0<br>F: 8.0-39.0 | | |
| **RBC** | 1-2 months | M: 3.24-4.08<br>F: 3.55-4.57 | M: 4.2-6.3 F: 4.0-5.4 | *10^6/µL |
| | 2-6 months | M: 3.67-4.61<br>F: 3.63-4.61 | | |
| | 6 months – 2 yrs | M: 3.81-4.74<br>F: 3.83-4.67 | | |
| | 2 yrs – 6 yrs | M: 3.92-4.72<br>F: 3.83-4.67 | | |

| | | | | |
|---|---|---|---|---|
| | 6 rs – 12 yrs | M: 3.85-4.75 F: 3.88-4.72 | | |
| | 12 yrs – 18 yrs | M: 3.74-4.93 F: 3.79-4.61 | | |
| **WBC** | 1-2 months | M: 8.36-13.66 F: 7.34-12.31 | 4-10 | *10^3/µL |
| | 2-6 months | M: 7.91-13.41 F: 6.85-12.84 | | |
| | 6 months – 2 yrs | M: 7.73-13.12 F: 7.05-12.98 | | |
| | 2 yrs – 6 yrs | M: 5.97-10.49 F: 5.98-10.08 | | |
| | 6 rs – 12 yrs | M: 5.69-9.88 F: 5.41-9.70 | | |
| | 12 yrs – 18 yrs | M: 5.24-9.74 F: 35.52-9.29 | | |
| **Eosinophil** | 1-2 months | M: 2.0-5.0 F: 2.0-6.0 | 1-6 | % |
| | 2-6 months | M: 2.0-6.0 F: 2.0-5.0 | | |
| | 6 months – 2 yrs | M: 1.0-5.0 F: 1.0-4.0 | | |
| | 2 yrs – 6 yrs | M: 1.0-4.0 F: 1.0-4.0 | | |
| | 6 rs – 12 yrs | M: 2.0-4.0 F: 1.0-4.0 | | |
| | 12 yrs – 18 yrs | M: 2.0-4.0 F: 1.0-3.0 | | |
| **PCT** | M: 0.19-0.39 F: 0.19-0.41 | | | % |
| **PLT** | 1-2 months | M: 221-471 F: 184-430 | 130-400 | *10^3/µL |
| | 2-6 months | M: 215-448 F: 147-423 | | |
| | 6 months – 2 yrs | M: 185-399 F: 211-408 | | |
| | 2 yrs – 6 yrs | M: 211-370 F: 190-365 | | |
| | 6 rs – 12 yrs | M: 227-350 F: 210-339 | | |
| | 12 yrs – 18 yrs | M: 180-299 F: 192-307 | | |
| **BUN** | < 1y r | 4-19 | 18-60 yrs: 6-20 | mg/dL |
| | < 18 yrs | 5-18 | 61-90 yrs: 8-23 | |
| **Creatinine** | M: 0.7-1.20 F: 0.5-0.9 | | | mg/dL |
| **eGFR** | > 60 | | | |

| | | |
|---|---|---|
| **CK (CPK)** | M: 39-308<br>F: 26-192 | U/L |
| **aPTT** | 28.5-45.4 | sec |
| **Fibrinogen** | 200-400 | mg/dL |
| **PT** | 11.8-15.3 | sec |