

Supplementary Table S1.

GISAID Patient Status Metadata Encodings. Mapping scheme used for raw patient status metadata. The mapping is to a patient status which is then mapped to a category and classification according to Supplementary Table S2 as detailed in the Methods section.

Original	Map1
18	Unknown
28	Unknown
37	Unknown
38	Unknown
42	Unknown
43	Unknown
44	Unknown
45	Unknown
51	Unknown
54	Unknown
55	Unknown
58	Unknown
59	Unknown
61	Unknown
67	Unknown
68	Unknown
71	Unknown
103	Unknown
Active	Active
ACTIVE	Active
Acute bronchitis	Acute infection
Acute upper respiratory infection, unspecified	Acute infection
Acute upper respiratory infections	Acute infection
Acute upper respiratory tract infection, unspecified	Acute infection
acute viral respiratory infection	Acute infection
Acute viral respiratory infection	Acute infection
Admitted for 6 Days in Bharatpur Hospital	Hospitalized
Airport surveillance	Screening
airport surveillance	Screening
Alive	Alive
alive	Alive
Ambulant	Ambulatory
Ambulatoriale	Ambulatory
Ambulatory	Ambulatory

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
AMBULATORY - UNDER TREATMENT	Ambulatory / Under treatment
Ambulatory - Under Treatment	Ambulatory / Under treatment
Ambulatory symptomatic	Mild
Ambulatory treatment	Ambulatory
Amublant	Ambulatory
Animal showing generalized sickness	Unknown
ARV clinic	Unknown
Asympomatic	Asymptomatic
asymptomatic	Asymptomatic
ASYMPTOMATIC	Asymptomatic
Asymptomatic	Asymptomatic
Asymptomatic - Ambulatory	Asymptomatic
Asymptomatic ambulatory	Asymptomatic
Asymptomatic and Ambulatory	Asymptomatic
Asymptomatic, identified as positive during preoperation investigation	Asymptomatic
Asymptomatic/Mild	Asymptomatic/Mild
Asymptomatic/Released	Asymptomatic
Asymptomatyc	Asymptomatic
Asymtomatic	Asymptomatic
Aymptomatic	Asymptomatic
B34.2 Coronavirus infection, unspecified site	Unknown
Blank	Unknown
Bronchitis	Bronchitis
Bronchoalveolar lavage	Unknown
casualty	Casualty
Casualty	Casualty
Chronic dialysis	Unknown
Clinic	Unknown
Clinical signs	Symptomatic
Cluster	Unknown
Community screening	Screening
Community Surveillance	Screening
Confirmed	Unknown
Confirmed SAD	Unknown
Confirmed SAD-SAD	Unknown
Confirmed SCTE	Unknown
Conjunctivitis	Mild
Contact	Screening

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Contact and possible infection with other infectious diseases	Screening
Contact surveillance	Screening
Contact with and exposure to other communicable diseases	Screening
Contact-SAD	Screening
Contained	Unknown
Coronavirus infection	Unknown
Cough	Mild
COVID-19	Unknown
Critical	Critical
CRITICAL	Critical
critical care admitted	Severe
Culled for pelting	Unknown
Cured	Recovered
DAMA	Unknown
dead	Dead
Dead	Dead
dead or alive, hospitalised or release	Unknown
Dead, hospitalized	Dead
Death	Dead
Decease	Dead
Deceased	Dead
deceased	Dead
Demise	Dead
Demised	Dead
Demised patient	Dead
Demised patient (Vaccinated first dose)	Dead
Desceased	Dead
Died	Dead
DIED	Dead
Discharged	Released
Discharged after recovery	Released
Diseased	Symptomatic
Domestic infection	Unknown
EHPAD	Unknown
EHPAD_IRA	Unknown
Emergency Care	Hospitalized
Encounter for general adult medical examination	Screening

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Encounter for general adult medical examination (Z.00)	Screening
Encounter for observation for other suspected diseases and conditions ruled out	Symptomatic
Encounter for other specified special examinations (Z01.8)	Unknown
Epidemiology Study	Screening
Etude	Unknown
Exitus	Dead
Facility quarantine	Screening
Family medical ward	Hospitalized
Female	Unknown
Fever	Mild
Flu-like symptoms	Flu-like
General clinic	Unknown
General Medical Screening	Screening
HCW - Mild Covid	Mild
Headache	Mild
Headache, Nasal Congestions and generalised body ache	Mild
Healthcare worker	Unknown
Healthcare Worker	Unknown
Healthworker	Unknown
Helen Joseph ward	Unknown
High temperature	Mild
Home	Home
home	Home
Hospitalized	Hospitalized
Hospitalized	Hospitalized
Hospitalized	Hospitalized
Hospitalized	Hospitalized
Hospitalized	Hospitalized
Hospital discharge	Released
Hospitalised	Hospitalized
Hospitalised oncology urology	Hospitalized
HOSPITALIZADO	Hospitalized
Hospitalized	Hospitalized
hospitalized	Hospitalized
Hospitalized - Non-Serious Case	Hospitalized / Not Serious
HOSPITALIZED - NON-SERIOUS CASE	Hospitalized / Serious

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
HOSPITALIZED - SERIOUS CASE	Hospitalized / Serious
Hospitalized - Serious Case	Hospitalized / Serious
Hospitalized (Critical)	Hospitalized / Serious
Hospitalized (Deceased)	Dead
Hospitalized (deceased)	Dead
Hospitalized (ICU)	Severe
Hospitalized (Intensive care unit and Released)	Severe
Hospitalized (Intensive care unit)	Severe
Hospitalized (Live)	Hospitalized
Hospitalized (Mild)	Hospitalized / Mild
Hospitalized (Moderate)	Hospitalized / Moderate
Hospitalized (possible MIS-C)	Hospitalized / Serious
Hospitalized (Released)	Released
Hospitalized (Severe)	Severe
Hospitalized (Stable)	Hospitalized / Stable
Hospitalized / ICU	Severe
Hospitalized Asymptomatic	Asymptomatic / Hospitalized
Hospitalized in intensive care	Severe
Hospitalized Symptomatic	Hospitalized / Symptomatic
Hospitalized(ICU)	Severe
Hospitalized, Asymptomatic, Alive	Asymptomatic / Hospitalized
Hospitalized, Deceased	Dead
Hospitalized, Discharged, Alive	Released
Hospitalized, ICU	Severe
Hospitalized, ICU, deseased	Dead
Hospitalized, ICU, Died	Dead
Hospitalized, Live	Hospitalized
Hospitalized, Live.	Hospitalized
Hospitalized, no clinical signs	Asymptomatic / Hospitalized
Hospitalized, noninvasive ventilation	Hospitalized
Hospitalized, oxygenotherapy, diarrhea	Hospitalized
Hospitalized, referred to Nepalganj Teaching hospital, stayed for 10 days and discharged.	Hospitalized
Hospitalizede	Hospitalized
Hospitalizes	Hospitalized
Hospitallized	Hospitalized
Hospitarized	Hospitalized
Hospitilized	Hospitalized
Hospitalized	Hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Hospitalized	Hospitalized
House assistance, Symptomatic	House Assistance
Human	Unknown
ICD-10 Disease: B34.2 Coronavirus infection, unspecified	Unknown
ICD-10 Disease: B34.2 Coronavirus infection, unspecified site	Unknown
ICD-10 Disease: J00- Acute nasopharyngitis [common cold]	Mild
ICD-10 Disease: J00-J06 Acute upper respiratory infections	Acute infection
ICD-10 Disease: J02.9 Acute pharyngitis, unspecified	Acute infection
ICD-10 Disease: J06.8 Other acute upper respiratory infections of multiple sites	Acute infection
ICD-10 Disease: J06.9 Acute upper respiratory infection, unspecified	Acute infection
ICD-10 Disease: J18.1 Lobar pneumonia, unspecified organism	Pneumonia
ICD-10 Disease: J18.9 Pneumonia, unspecified organism	Pneumonia
ICD-10 Disease: J40 Bronchitis, not specified as acute or chronic	Bronchitis
ICD-10 Disease: U07.1 COVID-19, virus identified	Unknown
ICD-10 Disease: U07.2 COVID-19, virus not identified	Unknown
ICD-10 Disease: U07.2 Virus not identified	Unknown
ICD-10 Disease: Z00.0 Encounter for general adult medical examination	Unknown
ICD-10 Disease: Z00.0 Encounter for general adult medical examination without abnormal findings	Symptomatic
ICD-10 Disease: Z03.8 Encounter for observation for other suspected diseases and conditions ruled out	Symptomatic
ICD-10 Disease: Z03.8 Observation for other suspected diseases and conditions	Symptomatic
ICD-10 Disease: Z04 Encounter for examination and observation for other reasons	Unknown

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
ICD-10 Disease: Z04.8 Encounter for examination and observation for request for expert evidence	Unknown
ICD-10 Disease: Z10.8 Routine general health check-up of other defined subpopulations	Unknown
ICD-10 Disease: Z11.5 Encounter for screening for other viral diseases	Unknown
ICD-10 Disease: Z20.8 Contact with and exposure to other communicable diseases	Screening
ICD-10 Disease: J06.9 Acute upper respiratory infection, unspecified	Acute infection
ICD-10 Disease: Pneumonia	Pneumonia
ICD-10 Disease: Z00.00 Encounter for general adult medical examination without abnormal findings	Asymptomatic
ICU	ICU
ICU, Deceased	Dead
ICU, Live	ICU
ICU; Serious	ICU
Immunocompromised	Unknown
Infection after vaccine/Live	Unknown / Vaccinated
Influenza Like Illness	Flu-like
In-hospital	Hospitalized
Initially hospitalized, but now improved and discharged	Released
inknown	Unknown
Inpatient	Hospitalized
inpatient	Hospitalized
Intensive Care	ICU
Intensive Care Unit	ICU
Isolation	Unknown
J03.9 Acute tonsillitis, unspecified	Acute infection
J04.1 Acute tracheitis	Acute infection
J06.8 Other acute upper respiratory infections of multiple sites	Acute infection
J06.9 Acute upper respiratory infection, unspecified	Acute infection
J12.9 Viral pneumonia, unspecified	Pneumonia
J18 Bronchopneumonia, unspecified	Pneumonia
J18.8 Other pneumonia, organism unspecified	Pneumonia

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
J18.9 Pneumonia, unspecified	Pneumonia
J22. Unspecified acute lower respiratory infection	Acute infection
Live	Alive
Live	Alive
live	Alive
Live, acute respiratory infection	Acute infection
Live, ambulatory care	Ambulatory
Live, asymptomatic	Asymptomatic
Live, hospitalized	Hospitalized
Live, mild fever and weakness	Mild
Live, mild symptoms, at home	Mild
Live, not hospitalized	Not hospitalized
Live, Not Hospitalized.	Not hospitalized
Live, physical examination	Screening
Live, Recovered	Recovered
Live, symptomatic	Symptomatic
Live.	Alive
Live. Ambulatory	Ambulatory
Live. Not hospitalized	Not hospitalized
Lived	Alive
lLive	Alive
M62.9 Disorder of muscle, unspecified, high temperature	Symptomatic
Male	Unknown
MILD	Mild
Mild	Mild
Mild / Contact exposure / Asymptomatic	Asymptomatic / Mild
Mild clinical signs	Mild
Mild clinical signs without hospitalization	Mild
Mild clinical signs without hospitalization. Diarrhoea	Mild
Mild clinical signs without hospitalization. Distorted ability to smell	Mild
Mild clinical signs without hospitalization. Distorted ability to smell and taste	Mild
Mild clinical signs without hospitalization. Distorted ability to smell and taste.	Mild
Mild clinical signs without hospitalization. Distorted ability to smell.	Mild

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Mild clinical signs without hospitalization. Distorted ability to taste	Mild
Mild Covid	Mild
Mild disease	Mild
Mild disease, full recovery, just rhinorrhea	Mild
Mild disease, persistent infection	Mild
Mild infection	Mild
Mild Symptoms	Mild
Mild symptomatology	Mild
Mild Symptoms	Mild
Mild symptoms (fever, cardiovascular disorders)	Mild
Mild symptoms, not-hospitalized	Mild
Mild symptomatic case	Mild
Mild, at home.	Mild
Mild, live	Mild
mild, unhospitalized	Mild
Mild/Contact exposure/Asymptomatic	Asymptomatic / Mild
Mild/Live	Mild
Mild/Moderate	Mild / Moderate
Moderate	Moderate
MODERATE	Moderate
Moderate / Outpatient	Moderate
Moderate, Live	Moderate
Moderate, live	Moderate
Moderate/Severe	Moderate / Severe
Multidisciplinary ICU	ICU
Nanopore	Unknown
Nasal swab	Unknown
Nasal Swab	Unknown
Nasopharyngeal and oropharyngeal swab	Unknown
Nasopharyngeal and oropharyngeal swabs	Unknown
Nasopharyngeal swab	Unknown
Neurology ward, Hospitalised	Hospitalized
New York	Unknown
NICD outbreak investigation	Unknown
No clinical signs	Asymptomatic
No Data	Unknown
No Hospitalized	Not hospitalized
Non critical hospitalization	Not critically hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Non Hospitalized	Not hospitalized
Non hospitalized	Not hospitalized
non hospitalized	Not hospitalized
Non Severe	Not severe
Non-hospitalized	Not hospitalized
Non-hospitalized, symptoms: cough	Mild
Non-hospitalized, symptoms: sore throat	Mild
Non-hospitalized, symptoms: runny nose	Mild
Not hospitalized	Not hospitalized
Not Hospitalized	Not hospitalized
Not Quarantined	Unknown
Not vaccinated	Not vaccinated
not vaccinated	Not vaccinated
Not Vaccinated	Not vaccinated
Nursing Home	Unknown
Nursing home	Unknown
Nursing house	Unknown
Obit - Death due to lack of oxygen in the health unit.	Dead
Oligosymptomatic	Mild
Oronpharyngeal swab	Unknown
Oropharyngeal swab	Unknown
Oro-pharyngeal swab	Unknown
Other	Unknown
Other acute upper respiratory infections of multiple sites	Acute infection
Other acute upper respiratory tract infections of multiple localization	Acute infection
Other unidentified viral infections	Unknown
ou	Mild
Oupatient	Not hospitalized
Outaptient	Not hospitalized
Outgoing traveler	Unknown
outpatient	Not hospitalized
Outpatient	Not hospitalized
out-patient	Not hospitalized
Out-patient	Not hospitalized
OUTPATIENT - HOME MONITORING	Home
Outpatient - Home Monitoring	Home
Outpatient mild disease	Mild

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Outpatient, Mild disease	Mild
Outpatient, mild disease	Mild
Outre mer	Unknown
Overseas inflow	Unknown
Paediatric casualty	Casualty
Paeds casualty	Casualty
Paucisymptomatic	Paucisymptomatic
Pauci-Symptomatic	Paucisymptomatic
Paucisymptomatyc	Paucisymptomatic
Paucisymtomatic	Paucisymptomatic
Paucisymtpmatic	Paucisymptomatic
Persistent infection	Chronic infection
Pneumonia	Pneumonia
Pneumonia (chest X-ray)	Pneumonia
Pneumonia (chest X-ray), not critical	Pneumonia
Pneumonia, unspecified organism	Pneumonia
Positive	Unknown
Post mortem	Dead
Quarantine	Unknown
Quarantine isolation	Unknown
Quarantine Isolation	Unknown
Quarantined	Unknown
random surveillance	Screening
Random surveillance	Screening
Rebased	Unknown
Recovered	Recovered
RECOVERED	Recovered
recovered	Recovered
Recovering	Recovered
Recovery	Recovered
Recovery 2021-04-18	Recovered
Referido	Unknown
Reinfected	Reinfection
Reinfection	Reinfection
Re-infection/moderate/live	Moderate / Reinfection
Release	Released
Released	Released
released	Released
Released, Live	Released
Respiratory disease	Symptomatic

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Retesting	Unknown
Routine	Screening
Routine general health check-up of other defined subpopulations	Screening
Sance Asymptomatic	Asymptomatic
Screen	Screening
Screened for travel	Screening
screening	Screening
SCREENING	Screening
Screening	Screening
Serious	Serious
Sever/hospitalised	Severe / Hospitalized
Severa	Severe
Severe	Severe
SEVERE	Severe / Hospitalized
Severe (Hospitalized)	Severe / Hospitalized
Severe / ICU	Severe / Hospitalized
Severe clinical signs	Severe
Severe infection	Severe
Severe symptoms with underlying asthma and hypertension	Severe
Severe, hospitalized	Severe
Severe, live	Severe
Severo	Severe
Sick	Symptomatic
sore throat, sweating, T-norm	Mild
Stable	Stable
Stable in quarantine	Mild
Sudden death	Dead
Surveillance	Screening
Suspect	Unknown
Suspect/SAD/SAD	Unknown
Suspect/SAD/SAD/SAD	Unknown
Suspected	Unknown
Suspected Corona	Unknown
Suspected coronavirus infection	Unknown
Suspect-SAD	Unknown
Symptomatic	Symptomatic
Symptomatic	Symptomatic
symptomatic	Symptomatic

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Symptomatic - Ambulatory	Ambulatory
Symptomatic mild disease	Mild
Symptomatic ambulatory	Ambulatory
Symptomatic and Ambulatory	Ambulatory
Symptomatic and Hospitalized	Hospitalized
Symptomatic, Outpatient	Outpatient
Symptomatic/mild illness/Live	Mild
Symptomatic/Reinfection	Symptomatic / Reinfection
Symptoms indicative of upper respiratory infection	Symptomatic
Throat	Mild
Trauma	Symptomatic
U07.1 Coronavirus infection caused by the COVID-19 virus, the virus has been identified (confirmed by laboratory testing, regardless of the severity of clinical signs or symptoms)	Unknown
U07.1 COVID-19, virus identified	Unknown
U07.2 Coronavirus infection caused by COVID-19 virus, virus not identified	Unknown
U07.2 COVID-19, virus not identified	Unknown
UCI	Unknown
unknown	Unknown
Umknown	Unknown
Undetlying DM and Ischemic Heart Disease	Unknown
unknkown	Unknown
unknoww	Unknown
Unknow	Unknown
unknow	Unknown
unknwon	Unknown
unkonown	Unknown
unkonwn	Unknown
Unkown	Unknown
unkown	Unknown
vaccinated	Vaccinated
Vaccinated	Vaccinated
Vaccinated & Demise	Dead / Vaccinated
Vaccine escape/ Live	Vaccinated
Vaccine escape/Live	Vaccinated
Vaccine escape/mild/Live	Vaccinated / Mild
Voyageur	Unknown

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
x	Unknown
Z00 General medical examination	Unknown
Z00.0 General medical examination	Unknown
Z01 Other special examinations and investigations of persons without complaint or reported diagnosis	Unknown
Z01.7 Laboratory examination	Unknown
Z04.9 Examination and observation for unspecified reason	Unknown
Z11.5	Unknown
Z11.5 Special screening examination for other viral diseases	Unknown
Z13.9 Special screening examination, unspecified	Unknown
Z20.0 Contact with and exposure to intestinal infectious diseases	Screening
Z20.8 Contact with and exposure to other communicable diseases	Screening
Active case	Active
ALIVE	Alive
DEAD	Dead
Deceased (BID)	Dead
Deceased (Brought in Dead)	Dead
Decreased	Dead
Deseased	Dead
Emergency care	Hospitalized
Expired	Dead
Fatal	Dead
Home Isolation	Home
Hospitalised, Reinfection	Hospitalized
Hospitalised-live	Hospitalized
Hospitalized / Rawat Inap	Hospitalized
Hospitalized -ICU-critical	Severe
Hospitalized, Pregnant	Hospitalized
In-patient	Hospitalized
J02.9 Acute pharyngitis, unspecified	Acute Infection
J04.2 Acute laryngotracheitis	Acute Infection
J06 Acute upper respiratory infections of multiple and unspecified sites	Acute Infection
J41 Simple chronic bronchitis	Acute Infection

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
J81 Pulmonary oedema	Pneumonia
Live / Hidup	Alive
Mild - not requiring supplementary oxygen	Mild
Mild Disease	Mild
Non-hospitalized	Not hospitalized
Nursing-home	Unknown
Outpatient Care, Alive	Not hospitalized
Outpatient-live	Not hospitalized
R05 Cough	Mild
R06.0 Dyspnoea, Orthopnoea, Shortness of breath	Symptomatic
R43.0 Anosmia	Mild
R50 Fever of other and unknown origin	Mild
R50.9 Fever, unspecified	Mild
R51 Headache	Mild
R53 Malaise and fatigue	Mild
Recover	Recovered
Released / Lepas Rawat	Released
Severe, Live	Severe
Symptomatic - Hospitalized	Hospitalized
Symptomatic COVID-19	Symptomatic
Symptoms: fever, general discomfort, headache, cough, rhinorrhea.	Mild
Symptoms: fever, general discomfort, muscle or joint pain, headache, cough, odynophagia, anosmia, rhinorrhea.	Mild
Symptoms: fever, general discomfort, muscle or joint pain, headache, cough, odynophagia, vomiting and diarrhea.	Mild
Symptoms: fever, general discomfort, muscle or joint pain, headache, rhinorrhea, vomiting or diarrhea.	Mild
Symptoms: fever, headache, loss of smell.	Mild
Symptoms: general discomfort, muscle or joint pain, headache, cough, odynophagia, rhinorrhea.	Mild
UNK	Unknown
Vaccine escape	Vaccinated
Very Severe	Severe
Admitted to ER	Hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
alive/not hospitalized	Not hospitalized
Criticalli III with SLE Encephalitis	Unknown
Deceased (brought in dead)	Dead
Deceased (in-hospital death)	Dead
Deceased (In-Hospital Death)	Dead
Discharged alive	Released
Hospitalizim	Hospitalized
Incarcerated	Unknown
J12.8 Other viral pneumonia	Pneumonia
live hospitalized	Hospitalized
Live non hospitalized	Not hospitalized
live non hospitalized	Not hospitalized
live, outpatient	Not hospitalized
mild	Mild
Mild COVID	Mild
Out patient	Not hospitalized
outpatient , Live	Not hospitalized
Outpatients	Not hospitalized
Released then hospitalized. Improved	Released
Sever	Severe
severe	Severe
Transfer to other facility	Hospitalized
U07.1 COVID-19	Symptomatic
U07.2 COVID-19	Symptomatic
Z00.0 General examination and investigation of persons without complaint and reported diagnosis	Asymptomatic
-	Unknown
,	Unknown
0.0	Unknown
1.0	Unknown
10.0	Unknown
11.0	Unknown
12.0	Unknown
13.0	Unknown
14.0	Unknown
15.0	Unknown
16.0	Unknown
17.0	Unknown
18.0	Unknown

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
19.0	Unknown
2.0	Unknown
20 - 29	Unknown
20.0	Unknown
21.0	Unknown
22.0	Unknown
23.0	Unknown
24.0	Unknown
25.0	Unknown
26.0	Unknown
27.0	Unknown
28.0	Unknown
29.0	Unknown
3.0	Unknown
30.0	Unknown
31.0	Unknown
32.0	Unknown
33.0	Unknown
34.0	Unknown
35.0	Unknown
36.0	Unknown
37.0	Unknown
38.0	Unknown
39.0	Unknown
4.0	Unknown
40.0	Unknown
41.0	Unknown
42.0	Unknown
43.0	Unknown
44.0	Unknown
45.0	Unknown
46.0	Unknown
47.0	Unknown
48.0	Unknown
49.0	Unknown
5.0	Unknown
50.0	Unknown
51.0	Unknown
52.0	Unknown
53.0	Unknown

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
54.0	Unknown
55.0	Unknown
56.0	Unknown
57.0	Unknown
58.0	Unknown
59.0	Unknown
6.0	Unknown
60.0	Unknown
62.0	Unknown
63.0	Unknown
64.0	Unknown
65.0	Unknown
66.0	Unknown
67.0	Unknown
68.0	Unknown
69.0	Unknown
7.0	Unknown
72.0	Unknown
73.0	Unknown
74.0	Unknown
75.0	Unknown
76.0	Unknown
78.0	Unknown
82.0	Unknown
83.0	Unknown
84.0	Unknown
85.0	Unknown
86.0	Unknown
87.0	Unknown
92.0	Unknown
94.0	Unknown
Accident And Emergency	Unknown
ACCIDENT AND EMERGENCY	Unknown
Admitted to Emergency Room	Hospitalized
Admitted to the Emergency Room	Hospitalized
Admitted	Hospitalized
Admitted for splenectomy, fever for 3 days	Hospitalized
Admitted to the ER	Hospitalized
Aftercare	Symptomatic
alive, hospitalized	Hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Alive, hospitalized	Hospitalized
Alive, not hospitalized	Not hospitalized
alive, not hospitalized	Not hospitalized
Allmn vervakning	Unknown
AMBULATORIO - EN TRATAMIENTO	Ambulatory
AMBULATORIO - SEGUIMIENTO DOMICILIARIO	Ambulatory
AMBULATORIO - SEGUIMIENTO TERMINADO	Ambulatory
Ambulatory - under treatment	Ambulatory
Asintomatico	Asymptomatic
Aspirado	Severe
Asymptomatic - Hospitalized	Asymptomatic
Asymptomatic, exposure +ve	Asymptomatic
Asymptomtic	Asymptomatic
ASYMTOMATIC	Asymptomatic
BID	Unknown
CAS4 MAIN CASUALTY	Unknown
Case follow-up	Unknown
CASUALTY	Unknown
Casualty Main Casualty	Unknown
CBC, FER, DID, LDH normal	Unknown
CBC, FER, DID, LDH normal, CRP mildly raised	Unknown
CBC, FER, DID, LDH normal, CRP mildly raised, CXR normal	Unknown
CBC, FER, DID, LDH not too bad	Unknown
cmjah	Unknown
Cold	Mild
Community Screen And Test	Screening
Community surveillance	Screening
Community surveillance halfway house	Screening
Confirm SAD	Unknown
Contact Case	Screening
Control	Screening
cough cold	Mild
cough, cold, SOB	Moderate
Critically ill	Severe
CXR, CBC, FER, DID, LDH, managed in Outpatient	Mild
Dead Body	Dead
death	Dead
DECEASED	Dead

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Deceased (brought in death)	Dead
Deceased (brought-in dead)	Dead
Deceased (brought-in death)	Dead
Deceased postmortem	Dead
Deceased, severe, Hospitalized	Dead
Deceased(brought in dead)	Dead
Decesead	Dead
Diagnosed	Dead
Discharge	Released
EN TRATAMIENTO	Symptomatic
Fallecido	Dead
FAMILY MEDICINE CLINIC	Mild
fever	Mild
Fever and Headache	Mild
fever, cold	Mild
Fever, cough	Mild
fever, cough cold sore throat	Mild
fever, cough cold sore throat,	Mild
fever, cough cold sore throat, SOB	Moderate
fever, cough sore throat, mild	Mild
Fever, cough, 3 days	Mild
fever, cough, cold	Mild
fever, runny nose	Mild
fever, sore throat	Mild
First Contact	Unknown
Hisopado	Unknown
Home isolated	Mild
Home Isolated	Mild
Home isolation	Mild
hospitalised	Hospitalized
Hospitalised - serious case	Serious
HOSPITALIZADO - CASO GRAVE -	Serious
HOSPITALIZADO - CASO NO GRAVE	Hospitalized / Moderate
Hospitalization	Hospitalized
HOSPITALIZED	Hospitalized
Hospitalized - death	Dead
Hospitalized - high improvement	Hospitalized
Hospitalized - non-serious case	Hospitalized / Moderate
Hospitalized - under treatment	Hospitalized
Hospitalized- serious case	Hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Hospitalized, Release	Released
Hospitalized, Released	Released
Hospitalized, severe	Severe
hospitalized	Hospitalized
Hotel isolation	Mild
hunter-harvested	Unknown
Imported	Unknown
In isolation	Unknown
In patient Labour ward	Unknown
In patient Peadiatric Surgery	Unknown
In-patient ARV Clinic	Unknown
In-patient Cardiology	Unknown
In-patient chronic dialysis	Unknown
In-patient ENT	Unknown
In-patient general surgery	Unknown
In-patient Labour	Unknown
In-patient Neurology	Unknown
In-patient Nuclear Medicine Renal	Unknown
In-patient Oncology	Unknown
In-patient Peadiatric Medical	Unknown
In-patient Peadiatric Renal	Unknown
In-patient Peadiatric Surgical	Unknown
In-patient Urology	Unknown
In-patint ENT	Unknown
increased CRP, other labs are normal	Mild
inpatient ,stable	Hospitalized / Stable
INT	Unknown
J00 Acute nasopharyngitis	Mild
J00 Acute nasopharyngitis Z03.8 Observation for other suspected diseases and conditions	Mild
J12 Viral pneumonia, unspecified	Pneumonia
J18.1 Lobar pneumonia, unspecified	Pneumonia
J20.9 Acute bronchitis, unspecified	Bronchitis
J30 Vasomotor rhinitis	Mild
J30.0 Vasomotor rhinitis	Mild
Leve	Alive
LIVE	Alive
Live - Intensive Care Unit	Severe
Live, Mild	Mild
Live, recovered	Recovered

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Local	Unknown
M42 .1 Adult osteochondrosis of spine	Unknown
Medical Casualty	Unknown
Mild cold symptoms	Mild
Mild cough	Mild
MILD COVID	Mild
Mild covid	Mild
MILD DISEASE	Mild
Mild fever and cough	Mild
Mild fever and cough, SOB 3 days	Moderate
Mild illness	Mild
mild symptomatic	Mild
mild symptoms	Mild
Mild symptoms	Mild
Mild symptoms (fever, cough)	Mild
Mild symptoms (fever, sore throat)	Mild
Mild symptoms fever, cough, sore throat, anosmia	Mild
Mild symptoms, fever, cough	Mild
Mild symptoms, fever, cough, sore throat	Mild
Missing	Unknown
Moderate covid	Moderate
Moderate COVID	Moderate
MODERATE COVID	Moderate
MODERATE DISEASE	Moderate
Moderate Disease	Moderate
MODERATE DISESE	Moderate
moderate symptomatic, hospitalized with oxygen did not require ICU, improvement and then discharge	Moderate / Severe
Moderate to Severe	Moderate / Severe
MODERTE DISEASE	Moderate
Nasal congestion, cold, dry cough	Mild
Nasopharyngeal Swab	Unknown
Nasopharyngeal and Oropharyngeal Swabs	Unknown
Nasopharyngeal swab and Oropharyngeal swab	Unknown
Nasopharyngeal Swabs	Unknown
No signs and symptoms	Asymptomatic
non Hospitalized	Not hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
non-hospitalized	Not hospitalized
Normal biomarkers	Mild
NOT STATED	Unknown
Nurse-house	Unknown
OPD patient	Not hospitalized
OUT PATIENT DEPARTMENT	Not hospitalized
Out-patient ARV clinic	Not hospitalized
Out-patient ARV Clinic	Not hospitalized
Out-patient Medical	Not hospitalized
Out-patient Staff	Not hospitalized
Out-patient Staff clinic	Not hospitalized
Outpatient - death	Not hospitalized
Outpatient - follow-up completed	Not hospitalized
Outpatient - Follow-up completed	Not hospitalized
Outpatient - home monitoring	Not hospitalized
outpatient / mild	Not hospitalized
outpatient care	Not hospitalized
outpatient, stable	Not hospitalized
paucisymptomatic	Paucisymptomatic
Pregnant 17 weeks+ 6 days	Unknown
Pregnant 26 weeks	Unknown
Pregnant, lymphocytosis on CBC	Unknown
Probable Case	Unknown
Quarantine facility	Unknown
quarantined	Unknown
R50 Fever, unspecified	Mild
Received first dose of vaccine Astrazeneca 31/03/2021	Vaccinated
Reinfection, Mild disease	Mild / Reinfection
Reinfection, Paucisymptomatic	Paucisymptomatic / Reinfection
Runny nose	Mild
SEGUIMIENTO DOMICILIARIO	Unknown
Severe case	Severe
SEVERE COVID	Severe
Severe covid	Severe
Severe COVID pneumonia in ICU	Severe
severe COVID pneumonia, recovered	Severe
SEVERE DISEASE	Severe
Severe Disease	Severe

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
severe symptomatic, required ICU, improvement after discharge	Severe
Severe, deceased	Severe
Severe, Hospitalized	Severe
Sintomatico	Symptomatic
Stayed at an accommodation facility, Released	Mild
Stayed at home, Released	Mild
Survived	Alive
Suspected Case	Symptomatic
Symptomatic (mild)	Mild
Symptomatic mildly for 2 days	Mild
Symptomatic Throat	Mild
Symptomatic, fever	Mild
Travel	Unknown
U07.2 COVID virus not identified	Unknown
U07.2 COVID-19, virus not identified J06.8 Other acute upper respiratory infections of multiple sites	Unknown
unkmown	Unknown
unknwn	Unknown
Vaccinated-1 dose-Astrazeneca	Vaccinated
Vaccinated-2 dose-Astrazeneca	Vaccinated
Vaccinated-2 dose-Pfizer	Vaccinated
Vaccinationsgenombrott	Vaccinated
Voyageur Sortant	Unknown
WARD 3 PAEDIATRIC	Unknown
z.008 Other general examinations	Unknown
Z00.0 General Medical Examination	Unknown
Z00.8 Other general examinations	Unknown
Z03.8 Observation for other suspected diseases and conditions	Unknown
Ano	Unknown
Assymptomatic	Asymptomatic
asymtomatic	Asymptomatic
Brought In Dead	Dead
Con cardiopatia.	Symptomatic
deceased 20/8	Dead
deceased 14/8	Dead
Home treatment	Mild
Hospitalizado	Hospitalized

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S1

Original	Map1
Hospitalized ICU	Severe
Hospitalized severe	Severe
Hospitalized-severe	Severe
Hospitalized,hospital transfer	Hospitalized
In-patient cardiac	Unknown
In-patient Cardio	Unknown
In-patient labour	Unknown
In-patient neurology	Unknown
In-patient paediatric	Unknown
In-patient paediatrics renal	Unknown
In-patient peritoneal dialysis	Unknown
In-patient staff	Unknown
In-patient surgery	Unknown
In-patient surgical	Unknown
Incoming Traveler	Unknown
J06.9 Acute upper respiratory infections of multiple and unspecified sites	Acute infection
Manipur	Unknown
Mild disease	Mild
NO	Unknown
No	Unknown
Not Collected	Unknown
Out-patient casualty	Not hospitalized
Out-patient ophthalmology	Not hospitalized
Outgoing Traveler	Unknown
referred to another hospital	Hospitalized
Self Isolation	Screening
symptomatic	Symptomatic
symptomatic. Diabetes	Symptomatic
U07.1 COVID-19 , virus identified	Unknown
U07.2 COVID-19 virus not identified	Unknown
Voyageur Entrant	Unknown
Yes	Unknown