



## Research paper

## Epidemiology, clinical characteristics, and virologic features of COVID-19 patients in Kazakhstan: A nation-wide retrospective cohort study

Sergey Yegorov<sup>a,i,1,\*</sup>, Maiya Goremykina<sup>b,1</sup>, Raifa Ivanova<sup>b,1</sup>, Sara V. Good<sup>c</sup>, Dmitriy Babenko<sup>c,d,1</sup>, Alexandr Shevtsov<sup>e</sup>, on behalf of the COVID-19 Genomics Research Group, Kelly S. MacDonald<sup>f,g</sup>, Yersin Zhunussov<sup>h</sup>, on behalf of the Semey COVID-19 Epidemiology Research Group

<sup>a</sup> School of Sciences and Humanities, Nazarbayev University, 53 Kabanbay Batyr Ave, Nur-Sultan 010000, Kazakhstan

<sup>b</sup> Department of Rheumatology and Non-Infectious Diseases, Semey Medical University 103 Abai street, Semey 071400, Kazakhstan

<sup>c</sup> Department of Biology, University of Winnipeg, 599 Portage Avenue, Winnipeg, Manitoba R3B2E9, Canada

<sup>d</sup> Research Centre, Karaganda Medical University, 40 Gogol St, Karaganda, 100008 Kazakhstan

<sup>e</sup> National Centre for Biotechnology, 13/5 Kurgalzhynskoye road, Nur-Sultan 010000, Kazakhstan

<sup>f</sup> Departments of Medicine, Microbiology & Immunology, Max Rady College of Medicine, University of Manitoba, 745 Bannatyne Avenue, Winnipeg, Manitoba R3E 0J9, Canada

<sup>g</sup> JC Wilt Infectious Diseases Research Centre, Public Health Agency of Canada, 745 Logan Avenue, Winnipeg, Manitoba R3E3L5, Canada

<sup>h</sup> Department of Public Health, Semey Medical University, 103 Abai street, Semey 071400, Kazakhstan

<sup>i</sup> Michael G. DeGroot Institute for Infectious Disease Research, Faculty of Health Sciences, McMaster University, 1280 Main St. West, Hamilton, Ontario, L8S 4L8, Canada

## ARTICLE INFO

## Article History:

Received 8 January 2021

Revised 4 March 2021

Accepted 22 March 2021

Available online 16 April 2021

The members of the COVID-19 Genomics Research Group (CGRG) collaborative group are listed in the appendix pp. 25–26  
The members of the Semey COVID-19 Epidemiology Research Group (SCERG) collaborative group are listed in the appendix pp. 25–26

## Keywords:

COVID-19

SARS-CoV-2

Clinical characteristics

Disease risk factors

Molecular epidemiology

Central Asia

Kazakhstan

SARS-CoV-2 genomics

Disease severity

## ABSTRACT

**Background:** The earliest coronavirus disease-2019 (COVID-19) cases in Central Asia were announced in March 2020 by Kazakhstan. Despite the implementation of aggressive measures to curb infection spread, gaps remain in the understanding of the clinical and epidemiologic features of the regional pandemic.

**Methods:** We did a retrospective, observational cohort study of patients with laboratory-confirmed COVID-19 hospitalized in Kazakhstan between February and April 2020. We compared demographic, clinical, laboratory and radiological data of patients with different COVID-19 severities on admission. Logistic regression was used to assess factors associated with disease severity and in-hospital death. Whole-genome SARS-CoV-2 analysis was performed in 53 patients.

**Findings:** Of the 1072 patients with laboratory-confirmed COVID-19 in March–April 2020, the median age was 36 years (IQR 24–50) and 484 (45%) were male. On admission, 683 (64%) participants had asymptomatic/mild, 341 (32%) moderate, and 47 (4%) severe-to-critical COVID-19 manifestation; 20 in-hospital deaths (1.87%) were reported by 5 May 2020. Multivariable regression indicated increasing odds of severe disease associated with older age (odds ratio 1.05, 95% CI 1.03–1.07, per year increase;  $p < 0.001$ ), the presence of comorbidities (2.34, 95% CI 1.18–4.85;  $p = 0.017$ ) and elevated white blood cell count (WBC, 1.13, 95% CI 1.00–1.27;  $p = 0.044$ ) on admission, while older age (1.09, 95% CI 1.06–1.13, per year increase;  $p < 0.001$ ) and male sex (5.63, 95% CI 2.06–17.57;  $p = 0.001$ ) were associated with increased odds of in-hospital death. The SARS-CoV-2 isolates grouped into seven phylogenetic lineages, O/B.4.1, S/A.2, S/B.1.1, G/B.1, GH/B.1.255, GH/B.1.3 and GR/B.1.1.10; 87% of the isolates were O and S sub-types descending from early Asian lineages, while the G, GH and GR isolates were related to lineages from Europe and the Americas.

**Interpretation:** Older age, comorbidities, increased WBC count, and male sex were risk factors for COVID-19 disease severity and mortality in Kazakhstan. The broad SARS-CoV-2 diversity suggests multiple importations and community-level amplification predating travel restriction.

**Funding:** Ministry of Education and Science of the Republic of Kazakhstan.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\* Corresponding author.

E-mail addresses: [yegorovsrg@gmail.com](mailto:yegorovsrg@gmail.com), [yegorovs@mcmaster.ca](mailto:yegorovs@mcmaster.ca) (S. Yegorov).

<sup>1</sup> These authors contributed equally to this work

## Research in context

### Evidence before this study

There is currently a paucity of high-quality data on the clinical features of and risk factors associated with COVID-19 severity and death, and molecular characteristics of the viral strains in Central Asia. Conflicting data exist on the regional epidemiologic attributes of COVID-19, with some countries having denied the existence of the pandemic and/or limited transparency over case reporting. We searched PubMed, bioRxiv and medRxiv and queried the "COVID-19 Research Explorer" application on 27 December 2020 for articles describing the COVID-19 clinical features, risk factors for severe disease and death, and SARS-CoV-2 genomic diversity in Kazakhstan and other Central Asian countries with no restriction for language or date of publication. Our search terms included "COVID-19" OR "SARS-CoV-2" OR "2019nCoV" OR "novel coronavirus" AND "disease severity" OR "death" OR "mortality" OR "genetic variants" OR "whole genome sequencing" AND "Kazakhstan" or "Kyrgyzstan" OR "Tajikistan" OR "Turkmenistan" OR "Uzbekistan". We found one study from Kazakhstan describing clinical characteristics of COVID-19 patients <19 years old, showing a relatively mild disease course for paediatric COVID-19. One study from Uzbekistan characterized demographic and clinical features of hospitalized COVID-19 patients and categorized by admission to intensive care units (ICU); this study found rates of ICU admission among laboratory-confirmed COVID-19 patients approaching 17% and symptomatic presentation of COVID-19 consistent with that reported globally. Several papers were irrelevant because the authors presented modelling studies using publicly available COVID-19 data, or the studies did not have a clinical or epidemiological focus, or the studies consisted of scoping reviews or opinions. We did not find any studies based on primary research data detailing factors associated with disease severity and/or mortality across varied ages and ethnic groups or studies of genomic diversity of SARS-CoV-2 in Central Asia.

### Added value of this study

In this retrospective, cohort study, we report demographic, clinical, laboratory, and radiological findings on admission for 1074 patients with laboratory-confirmed severe acute respiratory distress syndrome coronavirus 2 (SARS-CoV-2) infection and a genomic study of SARS-CoV-2 isolates from 53 patients in Kazakhstan. The observed mortality rate (1.9%) exceeded the figures officially reported for the same period (~0.5%). Older age, comorbidities, increased WBC count, and male sex were risk factors for COVID-19 severity and/or death in Kazakhstan. COVID-19 severity was elevated in non-Kazakh patients, likely due to other, ethnicity-linked, disease risk factors. The broad genomic SARS-CoV-2 diversity, and a virus lineage unique to Kazakhstan, suggest multiple COVID-19 importations that occurred earlier than currently thought.

### Implications of all the available evidence

By characterizing for the first time in detail the clinical and epidemiologic features of COVID-19 in an ethnically diverse Central Asian nation, we provide public health policy makers and clinicians with information critical for both patient management and regional COVID-19 response planning.

## 1. Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the cause of the coronavirus disease-2019 (COVID-19), within months of emergence from Wuhan, China, rapidly spread exacting a devastating human toll across the globe [1]. While the search for effective treatments continues and vaccines have commenced early implementation [1], it is imperative that up-to-date information be available from diverse populations on the disease epidemiology, clinical presentation, and population-specific characteristics influencing COVID-19 prevention, treatment, and vaccine strategies. This is especially critical in low- and middle-income countries (LMIC), where epidemiologic surveillance is often constrained due to resource shortages [2].

Kazakhstan was the first among the Central Asian LMICs to initiate COVID-19 screening in early 2020, with the first confirmed cases identified in mid-March, and a country-wide emergency state declared on 16 March 2020 [3]. Kazakhstan neighbours Russia, China, and the Central Asian states, and harbours an extensive ground and airway transit network, setting it in a vulnerable position for both COVID-19 importation and rapid community spread. Despite the initial swift public health response, Kazakhstan has encountered major barriers regarding case reporting and attempts have been made by both the government and public to increase transparency over COVID-19-related morbidity and mortality [3,4]. Thus a study from Kazakhstan has described the clinical characteristics of COVID-19 in children [5], corroborating a relatively mild disease course for paediatric COVID-19. However, little is still known about the clinical attributes and molecular epidemiology of COVID-19 across varied age and ethnic groups, information that is urgently needed to guide the public health authorities and clinicians in the wake of the on-going pandemic.

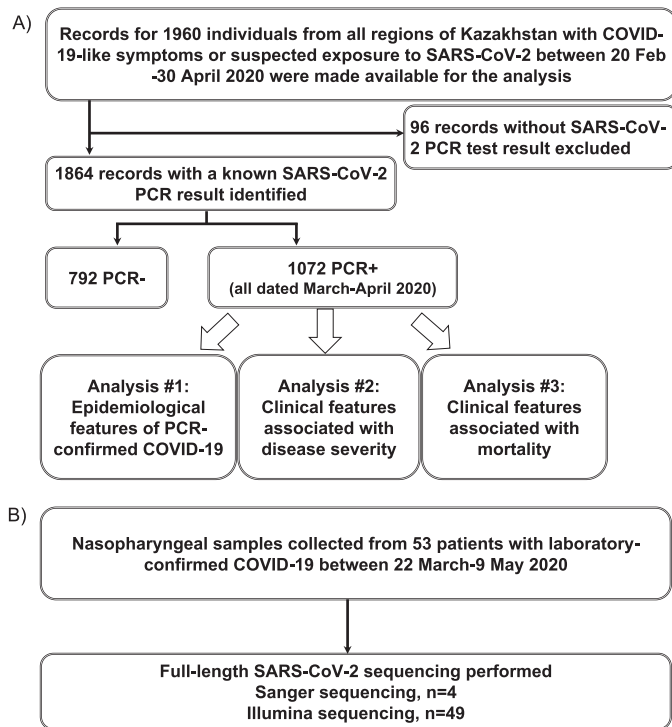
Here, we examined data from hospitalized patients with laboratory-confirmed COVID-19 during the first months of the pandemic to explore demographic, clinical and laboratory features and factors associated with COVID-19 disease severity and in-hospital death in Kazakhstan. We also analysed whole-genome SARS-CoV-2 data to characterize the regional community-level virus diversity.

## 2. Methods

### 2.1. Study design and participants

Study investigators were granted access to a subset of medical records generated between 20 February and 30 April 2020 and pertaining to 1960 patients, who presented with COVID-19-like symptoms or had a suspected exposure to SARS-CoV-2, as reported to the Republican Centre for Health Development by hospitals in 14 regions and 3 major cities (Fig 1A). All available patient data were included in the analyses. We used findings from earlier studies [6–9] to estimate that having >900 participants would provide ~85% power to detect associations of factors such as age, sex, and comorbidities with COVID-19 severity and related mortality. Patients were followed until the date of their discharge from hospital or death date; 5 May 2020 was chosen as the study end date to accommodate for delay in medical record data reporting. During the study period all subjects with suspected or confirmed COVID-19 infection were hospitalized in specialized "provisional" clinics. Clinical outcome assessment was performed as part of routine health care without involvement of study investigators.

The retrospective cohort study was approved by the Research Ethics Board of Semey Medical University as an anonymized epidemiological study, for which the requirement for informed consent was



**Figure 1.** A. The retrospective cohort study profile, B. The SARS-CoV-2 genomic study profile.

waived due to the pandemic state and urgent need to collect and analyse data. Virological samples were anonymized, and all genomic study procedures were approved by the Ethics Committee of the National Centre for Biotechnology.

## 2.2. Data collection

Epidemiological, demographic, clinical, laboratory, treatment, and outcome data were extracted from de-identified electronic medical records using a standardised data collection form. Radiological data were only available in the form of a radiologist's final diagnosis; a detailed description of imaging scans was unavailable. All data were entered into a computerized database and cross-checked by two physicians (MG and RI), and subsequently by a third researcher (SY).

## 2.3. Laboratory procedures

Laboratory confirmation of SARS-CoV-2 infection was done using real-time quantitative PCR on nasopharyngeal swabs by the regional National Centre of Expertise (NCE) laboratories using the Beijing Genomics Institute (BGI) kit (Shenzhen, China) targeting the Orf1ab locus A. Laboratory examinations included complete blood counts, blood chemical analyses, coagulation testing, liver and renal function assessment, and measurements of electrolytes, C-reactive protein, creatinine, D-dimer and lactate.

## 2.4. Virus genome sequencing, bioinformatics and phylogenetics

Full description of the sample preparation, sequencing, bioinformatic and phylogenetic methods is provided in appendix pp 2-5. Briefly, nasopharyngeal swabs were collected from randomly selected symptomatic patients, who had a positive SARS-CoV-2 PCR test and were hospitalized in the capital city, Nur-Sultan between 22 March and 9 May (Fig 1B). The number of processed samples (n=53) was determined based on the study budget and laboratory reagents

available at the time of patient recruitment. Travel history was available for 49 patients, none of whom had a history of recent international travel. SARS-CoV-2 whole genome sequencing was done using Sanger and Illumina platforms. The global SARS-CoV2 phylogeny and associated metadata were downloaded from Nextstrain[10] and the tree was reconstructed using the Nextstrain/ncov pipeline. To visualize the phylogenetic relationship of the Kazakhstan isolates in the context of possible origins of importation, we created a sub-tree using 165 sequences (see appendix p. 2) from the Global Initiative on Sharing All Influenza Data (GISAID) [11].

## 2.5. Definitions

A patient was assigned a laboratory-confirmed COVID-19 diagnosis if their medical record contained at least one positive SARS-CoV-2 PCR test result. Fever was defined as axillary temperature of at least 37.3°C. The degree of COVID-19 severity was defined by the clinical teams on the ground following the interim Ministry of Healthcare and WHO guidelines [12]. To avoid inconsistencies arising from varied terminology in electronic records, patients described as having "asymptomatic", "presymptomatic" and "mild" COVID-19 severity were combined into one "asymptomatic/mild COVID-19 disease" category. To facilitate comparisons with the literature, the term "non-severe" COVID-19 was used for both "asymptomatic/mild" and "moderate" disease, while the term "severe" COVID-19 was reserved for the pooled "severe" and "critical" groups. Recent international travel history was defined as travel outside of Kazakhstan within two weeks prior to hospitalization or symptom onset, whichever was first.

## 2.6. Statistical analysis

We hypothesised that COVID-19 severity and related mortality in Kazakhstan are associated with specific demographic, clinical, and laboratory-derived patient characteristics, and that the genetic diversity of circulating SARS-CoV-2 reflects the patterns of international travel into the country. The primary study outcomes were severe-critical COVID-19 and COVID-19-related in-hospital death. All other clinical and genomic findings were exploratory outcomes.

Variables were excluded prior to analysis if data were available for less than 40% of patients in any comparison group. We used the two-sided Mann-Whitney U,  $\chi^2$ , or Fisher's exact tests to compare differences between groups, as appropriate. Correlations among variables were explored using the Spearman rank test and principal component analyses (see appendix, pp. 6-9). For uni- and multivariable logistic regression analyses, patients with asymptomatic/mild and moderate disease severity were pooled into one "non-severe" category and compared with the "severe" (severe-critical) patient group. Risk factors associated with disease severity and death and their odds ratios (OR) were first analysed by univariable logistic regression. The linearity of continuous predictors and the logit of the outcome variables was assessed using the Box-Tidwell test. In the multivariable logistic regression analyses, to avoid model overfitting due to the limited size of endpoint events, four variables were chosen for the analysis of disease severity (n=47 for severe-critical patients) and two variables were ultimately included in the non-survivor analysis (n=20 for non-survivors). Variables significant in the univariable model ( $p < 0.05$ ) were then selected for multivariable analyses; priority was given to variables with fewest missing values, low collinearity (tolerance  $> 0.1$ , variance inflation factor  $< 10$ ) and those supported by earlier findings. Earlier studies have found older age, male sex, comorbidities, ethnicity and white blood cell count changes to be associated with adverse COVID-19 outcomes [6,7,9,13-16]. Although factors such as age and ethnicity may be considered effect modifiers in the context of COVID-19, given the small sample size of our endpoint events, we did not pursue stratification and assumed that all

variables in our models represented main exposures of interest. Hence, in our multivariable regression model of disease severity we included age, ethnicity, comorbidities, and WBC as the four variables. In our multivariable regression model of mortality, we used stepwise regression on age, sex, ethnicity, comorbidities, and WBC to select the two variables producing the best fit model, which ultimately were age and sex. Associations of age with disease severity and mortality were further explored using cubic spline models (appendix p.10). We compared the distribution of patient categories across clinical sites clustered by region and adjusted for potential effects of different clinical sites in a generalised linear model (appendix p.11). To model the association of body mass index (BMI) with disease severity and mortality, we used imputed BMI data (appendix pp.12-13). Analyses were performed using IBM SPSS V.23 and R V.4.0.3.

### 2.7. Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and the lead authors (SY, MG, RI) had final responsibility for the decision to submit manuscript for publication.

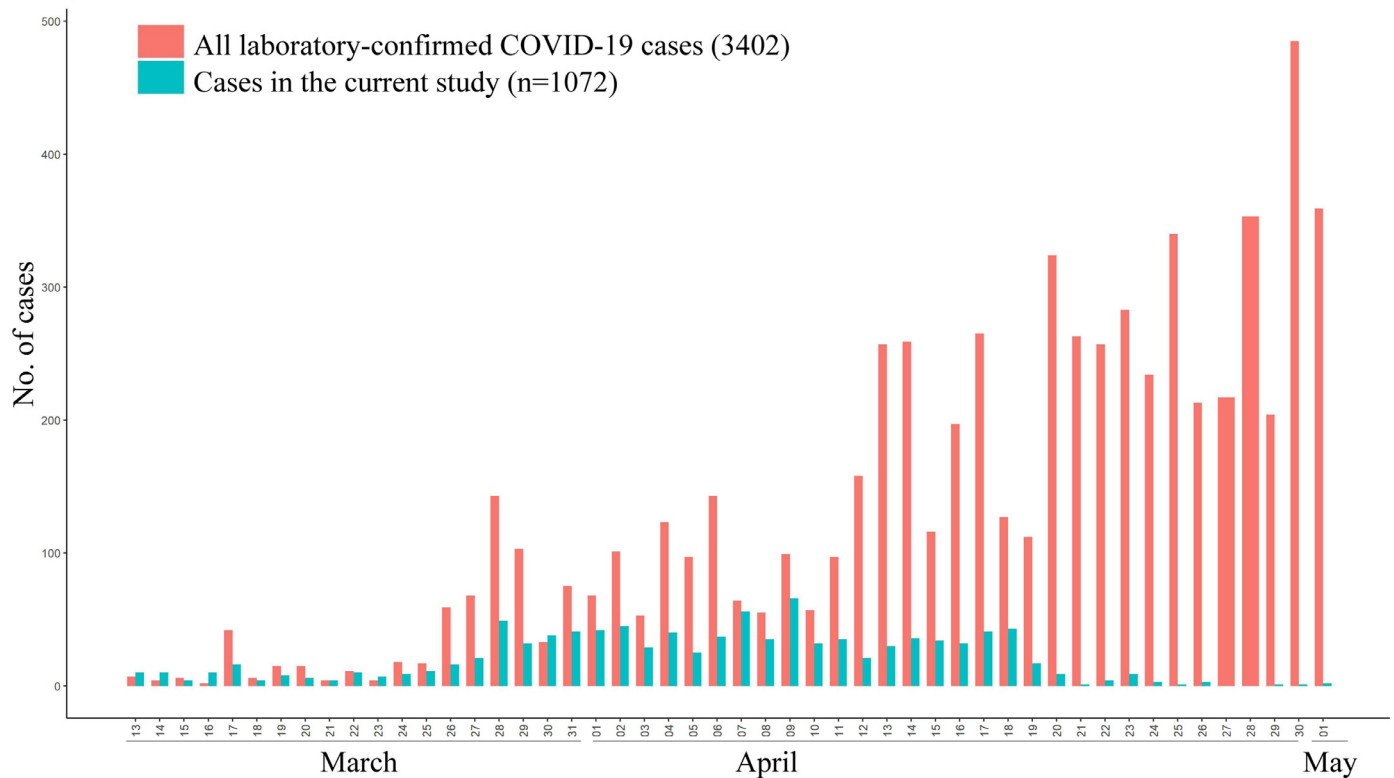
## 3. Results

After excluding 96 patients, for whom a SARS-CoV-2 PCR result was unavailable, there were 1864 patients with a known PCR result (Fig 1A). Our analysis focused on the 1072 patients, who had a confirmed SARS-CoV-2 PCR+ diagnosis, representing 32% of all (n=3402) patients with laboratory confirmed COVID-19 (Fig 2), who had been hospitalized in Kazakhstan as of April 30, 2020 [17]. 344 (32%) and 728 (68%) patients were hospitalized in March and April, respectively, and all patients had been discharged or died by 5 May 2020. 439

(41%) of all PCR+ cases were identified through screening of inbound travellers and via tracing activities, while 633 (59%) patients were admitted based on the presence of COVID-19-like symptomatology. Clinical site was known for 573 patients, most of whom (207, 36%) were admitted in the capital city and affiliated municipality (Nur-Sultan, Akmola region), followed by Almaty city and region (127, 22%) (Fig 3, appendix p. 11). Professional occupation was known for 350 patients, of whom 91 (26%) were health care workers. 170 (16%) patients reported recent international travel, predominantly from Russia (58%) and Europe (30%), via airplane (59%) and ground transportation (41%) (Fig 4). The comparisons of patient demographic and clinical characteristics on admission grouped by disease severity and death are shown in Tables 1 and 2, respectively, and in appendix pp 14-19. Variables with missing data are listed in appendix p. 20.

The median age of the cohort was 36 years, and most patients (80%) were Kazakh. Of all 1072 patients, 63.8% had asymptomatic/mild, 31.8% moderate, 3.83% severe and 0.56% critical severity of disease on admission. A total of 20 in-hospital deaths were reported: 1 (0.1%) in a patient with mild symptoms, 4 (1.2%) in patients with moderate disease, and 15 (31%) among patients with severe-critical disease on admission. There was no difference across the clinical sites in the distribution of patients by disease severity or in-hospital death (appendix p. 11).

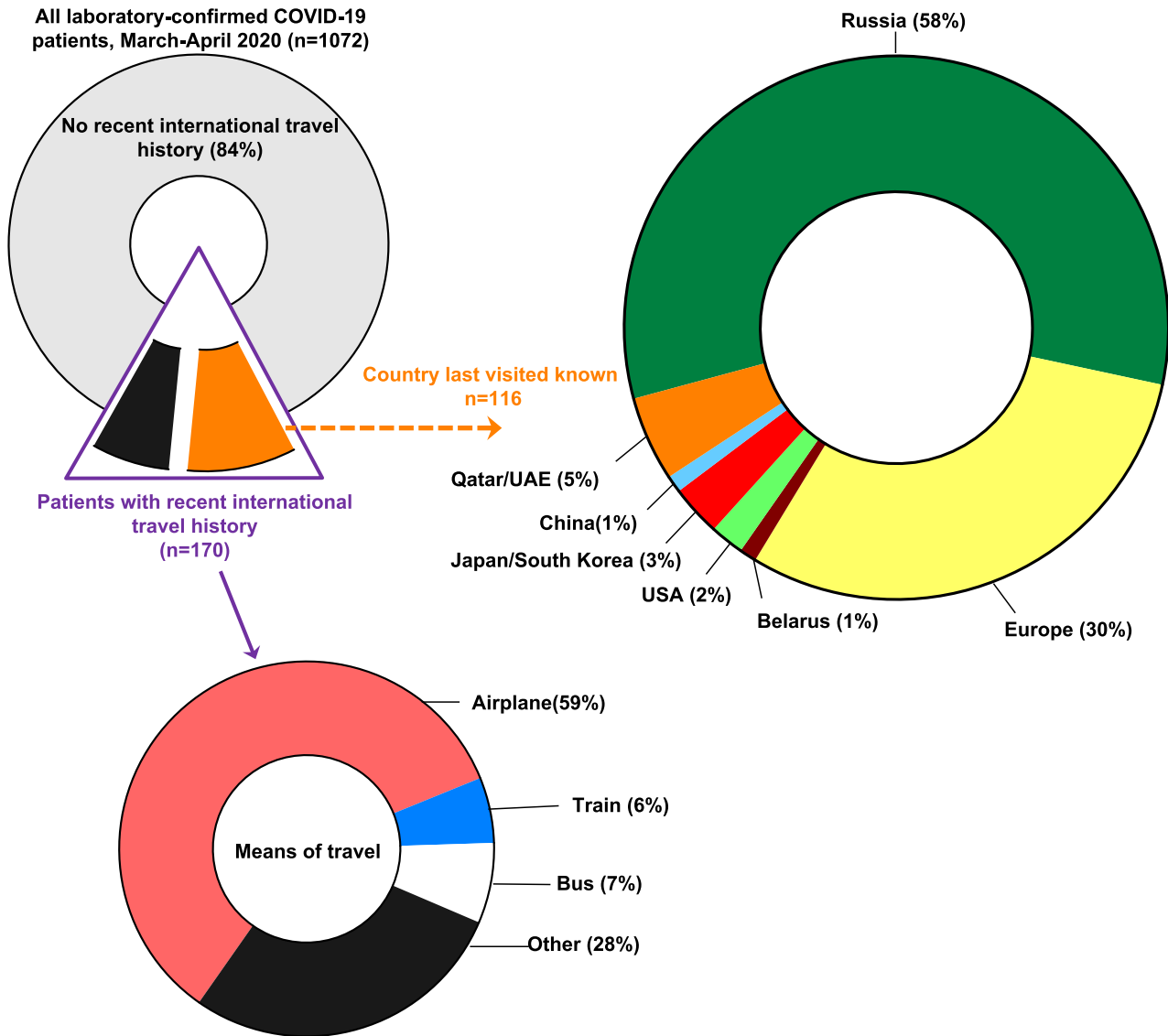
In the univariable analysis of COVID-19 severity predictors, older age, non-Kazakh ethnicity, comorbidities, elevated white blood cells (WBC), high neutrophil-to-lymphocyte ratio (NLR), lower haemoglobin, lower albumin and elevated creatinine were all associated with severe disease. In the multivariable analysis of 1072 patients (1024 non-severe and 47 severe-to-critical patients) older age (OR 1.05 [95% CI 1.03-1.07],  $p < 0.001$ ), comorbidities (OR 2.13 [95% CI 1.07-4.23],  $p < 0.031$ ) and elevated WBC (OR 1.14 [95% CI 1.01-1.28],  $p < 0.032$ ) were associated with increased odds of severe disease (Table 3). Similar results were seen in a generalized model adjusting



**Figure 2.** Epidemic curve of the confirmed COVID-19 cases in the current study compared to the official statistics on confirmed COVID-19 cases in Kazakhstan in March-April 2020 according to the Republican Centre for Health Development and World Health Organization (WHO). Official statistics were obtained from the WHO website [17].







**Figure 4.** Regions of travel and transportation used by laboratory-confirmed COVID-19 patients with a recent history of international travel.

classification guidelines [12]. Thus, the proportion of severe-critical cases in our study was 4%: ~5-fold lower than that reported elsewhere [13,14,23]. Nevertheless, the overall mortality rate in Kazakhstan (1.9%) is consistent with the estimates in COVID-19 patients presenting a wide spectrum of disease manifestations [8], though is 10-fold lower than in hospitalized patient cohorts [15,25]—likely due to symptom-unguided hospitalization of all laboratory-confirmed COVID-19 patients in Kazakhstan during this early pandemic phase.

As expected, mortality correlated with disease severity, being lowest in patients with mild disease, similar to that typically reported (~0.1%) [13], but escalating 4-fold in patients with severe-critical disease compared to severe COVID-19 elsewhere (e.g. 8% in [13]). The high mortality associated with severe-critical COVID-19 in this cohort could be due to a combination of mis-categorization of patients manifesting severe COVID-19 symptoms into non-severe disease categories, and subsequent “saturation” of the severe-critical disease category by patients with unfavourable prognosis, and a shortage of healthcare resources in a rapidly escalating pandemic [2,3].

Interestingly, non-Kazakh ethnicity was associated with both COVID-19 severity and death in univariable analyses, an effect abrogated in multivariable models, suggesting that being non-Kazakh is linked to other risk factors, consistent with augmented risk of COVID-19 infection and more severe disease among ethnic minority groups in other countries [16].

Both our clinical and laboratory findings were overall in line with earlier research. Notwithstanding, there are differences that may have arisen due to the heterogeneity in data collection by facilities across the country. For example, unlike in other studies [9,13], thrombocytopenia in our cohort was uncommon, but, consistently, lower absolute platelet counts were associated with mortality.

Based on radiological assessment, bronchitis was a common diagnosis in non-severe COVID-19, consistent with bronchial wall thickening in ~20% of COVID-19 cases, indicating an effect of SARS-CoV-2 with or without a respiratory coinfection [26]. While the prevalence of radiology-confirmed pneumonia increased with disease severity, the presence of pneumonia in 16% of “mild COVID-19” patients was

**Table 1**

Demographic and clinical characteristics of laboratory-confirmed COVID-19 patients categorized on admission by disease severity.

	All (1072)	Asymptomatic/Mild (683)	Moderate (341)	Severe/critical (47)	P value
Age, years	36.0 (24.0-50.0)	32.0 (23.0-47.0)	38.0 (28.0-53.0)	58.0 (45.5-71.0)	<0.001
0-14	94 (8.77%)	76 (11.1%)	17 (4.99%)	1 (2.13%)	<0.001
15-49	695 (64.8%)	458 (67.1%)	222 (65.1%)	15 (31.9%)	
50-64	202 (18.8%)	113 (16.5%)	76 (22.3%)	13 (27.7%)	
≥ 65	81 (7.56%)	36 (5.27%)	26 (7.62%)	18 (38.3%)	
Sex (Men)	484 (45.1%)	301 (44.1%)	155 (45.5%)	27 (57.4%)	0.157
Kazakh ethnicity	852 (79.5%)	551 (80.7%)	271 (79.5%)	30 (62.5%)	0.011
BMI, kg/m <sup>2</sup> *	23.9 (21.4-28.7)	23.0 (20.0-26.2)	25.8 (23.1-30.7)	27.5 (25.8-29.2)	0.012
Days from symptom onset to admission	1 (1-3)	1 (1-2)	2 (1-6)	4 (2-7)	<0.001
Days in hospital	16.0 (14.0-17.0)	16.0 (14.0-17.0)	16.0 (15.0-18.0)	15.0 (3.00-17.0)	0.075
Deaths	20 (1.9%)	1 (0.1%)	4 (1.2%)	15 (31.3%)	<0.001
Any comorbidities <sup>^</sup>	431 (40.2%)	235 (34.4%)	163 (47.8%)	33 (68.8%)	<0.001
<b>Select signs and symptoms<sup>^</sup></b>					
Fever (ax temp ≥ 37.3C)	131 (12.2%)	31 (4.54%)	82 (24.0%)	18 (38.3%)	<0.001
Cough	294 (27.4%)	104 (15.2%)	163 (47.8%)	27 (57.4%)	<0.001
Systolic pressure *	110 (110-120)	110 (110-120)	120 (110-130)	118 (108-126)	<0.001
<90 mmHg	11 (1.03%)	8 (1.17%)	0 (0.00%)	3 (6.38%)	<0.001
Respiratory rate *	18.0 (18.0-20.0)	18.0 (18.0-19.0)	18.0 (18.0-20.0)	22.5 (20.0-28.5)	<0.001
>24 breaths/min	25 (2.33%)	8 (1.17%)	8 (2.35%)	9 (19.1%)	<0.001
SpO <sub>2</sub> *	97.0 (96.0-98.0)	98.0 (96.0-98.0)	97.0 (96.0-98.0)	91.5 (84.5-95.0)	<0.001
<b>Select laboratory findings<sup>^</sup></b>					
White blood cells, x10 <sup>9</sup> /L, median	6.30 (5.10-7.90)	6.47 (5.10-8.00)	6.10 (5.00-7.68)	6.55 (5.50-9.05)	0.067
< 4 × 10 <sup>9</sup> /L	90 (8.8%)	54 (8.3%)	34 (10.4%)	2 (4.4%)	0.056
> 10 × 10 <sup>9</sup> /L	79 (7.8%)	51 (7.9%)	20 (6.1%)	8 (17.8%)	
NLR *	2.06 (1.36-3.16)	2.00 (1.38-3.14)	2.10 (1.45-3.05)	2.58 (0.91-6.48)	0.581
> 4.0	76 (16.3%)	45 (15.7%)	22 (14.4%)	9 (33.3%)	0.044
Median haemoglobin, g/L	136 (124-150)	137 (124-149)	137 (124-152)	127 (114-142)	0.007
Median prothrombin time, s *	13.0 (11.7-16.0)	12.8 (11.4-16)	13 (11.85-15.75)	14.75 (12.48-19.6)	0.018
Median aspartate aminotransferase, U/L *	18.5 (12.5-26.5)	17.71 (12-23.2)	20.04 (13.78-29.55)	38.69 (17.55-50.28)	<0.001
Total bilirubin, mmol/L, median *	11.2 (8.00-16.0)	11.5 (8.2-16.73)	10.45 (7.58-15)	13.2 (10.20-1)	0.002
Median creatinine, uM*	73.1 (60.3-88.0)	70.1 (57.08-86.2)	75.5 (62.45-92.8)	84 (67-117)	<0.001
Median C-reactive protein, mg/L *	2.30 (0.09-8.10)	2.00 (0.00-6.00)	2.69 (0.12-10.0)	6.00 (0.65-45.1)	0.022
<b>Radiologic findings *</b>					
Pneumonia	230 (31.2%)	69 (15.6%)	130 (49.8%)	31 (91.2%)	<0.001
Bronchitis	188 (25.5%)	113 (25.6%)	73 (28.0%)	2 (5.88%)	
<b>Treatments<sup>^</sup></b>					
Antiviral medications <sup>^</sup>	749 (69.9%)	461 (67.5%)	261 (76.5%)	27 (57.4%)	0.001
Antibiotics <sup>†</sup>	308 (28.7%)	114 (16.7%)	168 (49.3%)	26 (55.3%)	<0.001
Anticoagulant/antiplatelet therapy	29 (2.71%)	6 (0.88%)	12 (3.52%)	11 (23.4%)	<0.001
Corticosteroids	16 (1.49%)	7 (1.02%)	7 (2.05%)	2 (4.26%)	0.13
Oxygen therapy	64 (6%)	10 (1.5%)	23 (6.7)	31 (64.6%)	<0.001
Mechanical ventilation	27 (2.5%)	2 (0.3%)	6 (1.8%)	19 (39.6%)	<0.001

Data are median (IQR) or n/N (%), where N is the number of patients with available data. BMI=body mass index. NLR=neutrophil/lymphocyte ratio. SpO<sub>2</sub>=oxygen saturation. \*Data were available for <1072 patients, see appendix p. 20 for details on available sample size. ^ see appendix pp 14-19 for detailed list † see appendix p 21 for a detailed list of prescribed antibiotics.

unexpected. To facilitate data analysis, we chose to adopt the disease severity classification used by the clinical teams on the ground and therefore did not exclude “mild” patients with pneumonia from the primary analysis, although exclusion of these subjects in a sub-analysis did not have any impact on the study outcomes. Upon closer examination of this patient subset, we did not identify any potential sources of bias, such as shared clinical sites or periods of diagnosis. Therefore, we believe that the observed discrepancy arose primarily due to communication delays between clinicians and radiologists in charge, although other causes such as incomplete understanding and/or adherence of clinical teams to the WHO guidelines and erroneous data entry into the electronic medical record system may have contributed.

Unexpectedly, some asymptomatic/mild patients in our cohort received therapeutic interventions typically indicated for severe disease manifestations. It is plausible that cases classified as asymptomatic/mild on admission exhibited symptom worsening during hospitalization, resulting in indication of specialized treatments such as anticoagulants or oxygen therapy. Additionally, empiric antiviral therapies may have been administered to patients due to a lack of consistent messaging and stable clinical guidelines during this early

phase of the pandemic. Notably, some medications administered to patients in Kazakhstan, such as the antivirals imported from Russia, have limited efficacy and safety data available and the evidence for their use in COVID-19 patients should be investigated. The high rates of anticoagulant and antiplatelet therapy in patients with severe disease and non-survivors are a likely result of indication bias. However, given the decreased platelet counts in non-survivors, studies involving platelet function assays and autopsies are urgently needed to better understand the outcomes of haemostatic and thrombotic events and their treatment in COVID-19 patients.

We identified representatives of five of the eight global SARS-CoV-2 clades circulating in Kazakhstan early in the pandemic. Most isolates (46/53) belonged to clades O or S, descendants of the early Asian lineages globally prevalent prior to the emergence of the D614-G614 mutation that swept through Europe, reaching ~67% frequency among European sequences by mid-March [27]. Since our sampling post-dated the appearance of the European G clade mutation, the high prevalence of O and S sub-types in Kazakhstan suggests an early importation, perhaps weeks prior to the international travel ban declaration, and subsequent amplification through community spread. Notably, the Kazakhstan S lineage isolates clustered with viruses of

**Table 2**

Demographic and clinical characteristics of laboratory-confirmed COVID-19 patients, who had survived (survivors) or died (non-survivors) by April 30, 2020.

	All (1072)	Survivors (1052)	Non-survivors (N=20)	p value
Age, years	36.0 (24.0-50.0)	35.0 (24.0-50.0)	65.0 (57.8-77.8)	<0.001
0-14	94 (8.77%)	93 (8.84%)	1 (5.00%)	
15-49	695 (64.8%)	693 (65.9%)	2 (10.0%)	
50-64	202 (18.8%)	194 (18.4%)	8 (40.0%)	
≥ 65	81 (7.56%)	72 (6.84%)	9 (45.0%)	
Sex (Men)	484 (45.1%)	470 (44.7%)	14 (70.0%)	0.043
Kazakh ethnicity	852 (79.5%)	842 (80.0%)	10 (50.0%)	0.001
BMI, kg/m <sup>2</sup> *	23.9 (21.4-28.7)	23.9 (21.4-28.7)	43.5 (33.8-53.1)	0.213
Days from symptom onset to admission	1 (1-3)	1 (1-3)	4 (3-10)	<0.001
Days in Hospital	16.0 (14.0-17.0)	16.0 (14.0-17.0)	4 (2.0-15.0)	<0.001
Any comorbidities <sup>^</sup>	431 (40.2%)	415 (39.4%)	16 (80.0%)	<0.001
<b>Select signs and symptoms<sup>^</sup></b>				
Fever (ax temp ≥ 37.3C)	131 (12.2%)	124 (11.8%)	7 (35.0%)	<0.001
Cough	294 (27.4%)	286 (27.2%)	8 (40.0%)	0.308
Systolic pressure *	110 (110-120)	112 (110-120)	105 (100-130)	0.252
<90 mmHg	11 (1.03%)	8 (0.76%)	3 (15.0%)	0.001
Respiratory rate *	18.0 (18.0-20.0)	18.0 (18.0-20.0)	23.0 (22.0-30.0)	<0.001
>24 breaths/min	25 (2.33%)	20 (1.90%)	5 (25.0%)	<0.001
SpO <sub>2</sub> *	97.0 (96.0-98.0)	98.0 (96.0-98.0)	86.0 (74.0-95.0)	<0.001
<b>Select laboratory findings<sup>^</sup></b>				
White blood cells, x10 <sup>9</sup> /L, median	6.30 (5.10-7.90)	6.30 (5.07-7.90)	7.05 (5.69-9.30)	0.086
< 4 × 10 <sup>9</sup> /L	90 (8.8%)	90 (9.0%)	0 (0%)	0.036
> 10 × 10 <sup>9</sup> /L	79 (7.8%)	75 (7.5%)	4 (22.2%)	
NLR *	2.06 (1.36-3.16)	2.06 (1.37-3.14)	2.37 (1.15-6.11)	0.751
> 4.0	76 (16.3%)	72 (15.8%)	4 (36.4%)	0.068
Median haemoglobin, g/L	136 (124-150)	137 (124-150)	131 (119-142)	0.200
Median prothrombin time, s *	13.0 (11.7-16.0)	13.0 (11.7-16.0)	15.7 (12.6-18.5)	0.063
Median aspartate aminotransferase, U/L *	18.5 (12.5-26.5)	18.3 (12.4-25.9)	42.0 (26.0-70.0)	<0.001
Direct bilirubin, mmol/L, median *	2.70 (1.79-4.30)	2.62 (1.75-4.30)	10.0 (4.00-10.9)	0.012
Median creatinine, uM*	73.1 (60.3-88.0)	73.0 (60.0-88.0)	87.0 (75.8-163)	0.002
Median C-reactive protein, mg/L *	2.30 (0.09-8.10)	2.00 (0.02-7.90)	8.30 (4.14-42.0)	0.077
<b>Radiologic findings *</b>				
Pneumonia	230 (31.2%)	218 (30.2%)	12 (80.0%)	0.001
Bronchitis	188 (25.5%)	187 (25.9%)	1 (6.67%)	
<b>Treatments<sup>^</sup></b>				
Antiviral medications <sup>^</sup>	749 (69.9%)	739 (70.2%)	10 (50.0%)	0.087
Antibiotics <sup>†</sup>	308 (28.7%)	299 (28.4%)	9 (45.0%)	0.170
Anticoagulant/antiplatelet therapy	29 (2.71%)	24 (2.28%)	5 (25.0%)	<0.001
Corticosteroids	16 (1.49%)	15 (1.43%)	1 (5.00%)	0.262
Oxygen therapy	64 (6%)	44 (4.2%)	20 (100%)	<0.001
Mechanical ventilation	27 (2.5%)	9 (0.9%)	18 (90%)	<0.001

Data are median (IQR) or n/N (%), where N is the number of patients with available data. BMI=body mass index. NLR=neutrophil/lymphocyte ratio. SpO<sub>2</sub>=oxygen saturation. \*Data were available for <1072 patients, see appendix p. 20 for details on available sample size. ^ see appendix pp 14-19 for detailed list † see appendix p 21 for a detailed list of prescribed antibiotics.

**Table 3**

Bivariate logistic regression of factors associated with the odds of severe COVID-19 disease in Kazakhstan.

	Univariable OR (95% CI)	p value	Multivariable OR (95% CI)	p value
Age, years*	1.06 (1.04-1.08)	<0.001	1.05 (1.03-1.07)	<0.001
Male sex (vs. female)	1.68 (0.93-3.07)	0.086	..	..
Kazakh ethnicity (vs. other ethnicities)	0.40 (0.22-0.74)	0.003	0.73 (0.37-1.51)	0.380
Comorbidities	3.71 (2.00-7.25)	<0.001	2.34 (1.18-4.85)	0.017
White blood cells*	1.12 (0.99-1.25)	0.052	1.13 (1.00-1.27)	0.044
NLR > 4.0	1.15 (1.05-1.25)	0.001	..	..
Haemoglobin, g/L*	0.98 (0.97-0.99)	0.005	..	..
Prothrombin time, s*	1.00 (NA-1.00)	0.871	..	..
Fibrinogen, g/L*	1.00 (NA-1.00)	0.725	..	..
Albumin, g/L*	0.95 (0.91-0.99)	0.004	..	..
Aspartate aminotransferase, >40 U/L	11.05 (5.63-21.21)	<0.001	..	..
Total bilirubin, >17 mmol/L	1.67 (0.77-3.30)	0.165	..	..
Glucose, mmol/L*	1.04 (0.98-1.09)	0.094	..	..
Blood urea nitrogen, mmol/L*	1.00 (0.98-1.00)	0.965	..	..
Creatinine, >118 uM	5.98 (2.44-13.22)	<0.001	..	..
C-reactive protein, mg/L*	1.00 (NA-1.00)	0.847	..	..
Potassium, mmol/L*	0.85 (0.55-1.18)	0.438	..	..
Calcium, mmol/L*	1.04 (1.00-NA)	0.208	..	..

The two binary dependent variables in the model were patients with "severe" (47 patients) versus "non-severe" (1024 patients) COVID-19 disease. OR=odds ratio. NLR=neutrophil/lymphocyte ratio. \*For each additional unit.



**Table 4**  
Bivariate logistic regression of factors associated with COVID-19 mortality in Kazakhstan.

	Univariable OR (95% CI)	p value	Multivariable OR (95% CI)	p value
Age, years*	1.08 (1.05-1.11)	<0.001	1.09 (1.06-1.13)	<0.001
Sex (Men)	2.89 (1.15-8.21)	0.031	5.63 (2.06-17.57)	0.001
Kazakh ethnicity	0.25 (0.10-0.62)	0.002	..	..
Comorbidities	6.14 (2.23-21.55)	0.001	..	..
White blood cells, $\times 10^9/L^*$	1.22 (1.03-1.42)	0.014	..	..
Platelet count, $\times 10^9/L^*$	0.99 (0.98-1.00)	0.002	..	..
Fibrinogen, g/L*	1.00 (NA-1.00)	0.758	..	..
Albumin, g/L*	0.93 (0.88-0.99)	0.004	..	..
Aspartate aminotransferase, >40 U/L	21.74 (8.66-56.03)	<0.001	..	..
Direct bilirubin, >5.1 mmol/L	2.82 (0.65-8.67)	0.105	..	..
Glucose, mmol/L*	1.04 (0.97-1.09)	0.132	..	..
Blood urea nitrogen, mmol/L*	1.00 (NA-1.00)	0.889	..	..
Creatinine, >118 uM	9.14 (2.85-25.05)	<0.001	..	..
Sodium, mmol/L*	1.00 (0.98-1.06)	0.884	..	..

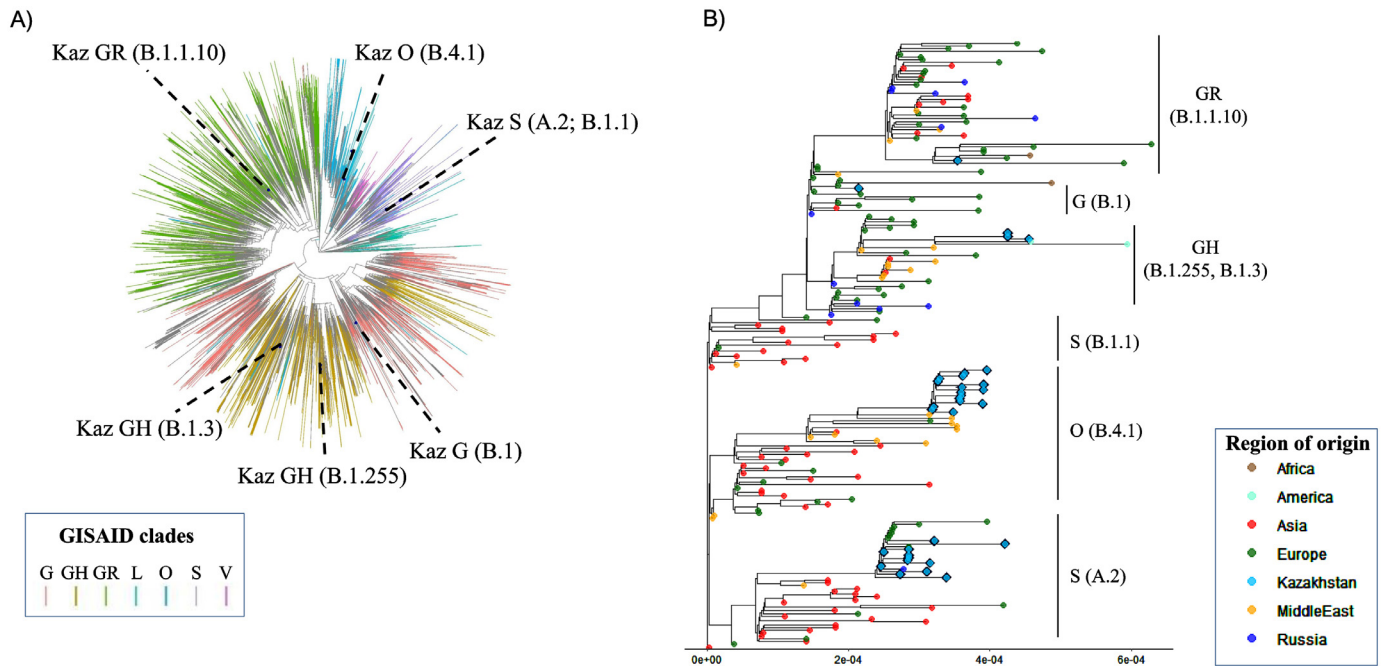
The two binary dependent variables in the model were “non-survivors” (20 patients) versus “survivors” (1052 patients). OR=odds ratio. NLR=neutrophil/lymphocyte ratio. \*For each additional unit.

from Europe, such as Spain in the case of lineage S/A.2, while the Kazakhstan O clade isolates, clustering with Middle Eastern (particularly, Iranian) strains, all belonged to lineage B.4.1, which arose uniquely in Kazakhstan. These observations are consistent with the evidence that areas of southern Europe and the Middle East represented early points of SARS-CoV-2 introduction and spread [28]. Additionally, the presence of three clade G lineages in Kazakhstan at diverse timepoints, between 25 March and 9 May, indicates multiple independent importations from Europe and the Americas.

Patient travel histories suggest that during the study period COVID-19 infections were imported mainly from the neighbouring Russia, and from Europe. Only 1% of the patients reported travel from China, where case numbers had substantially decreased by March 2020 [29]. Remarkably, 41% of the international travellers used train or bus, and other means of ground transportation, which has implications for public health policies aimed to curtail COVID-19 spread on public transit.

Our study is limited by the incomplete nature of the dataset, which included only a third of all laboratory-confirmed COVID-19 cases in March–April in Kazakhstan, heterogeneous quality of the medical record data, limited availability and missing data for some variables, potential unaccounted delays in clinical outcome reporting, lack of data on the performance characteristics of COVID-19 diagnostic tests across the country, small sample size and catchment area of the genomic study, under-powered to assess associations of viral diversity with clinical outcomes.

To conclude, despite the limitations, our study is the first to assess COVID-19 severity and related mortality predictors and describe the genetic diversity of circulating SARS-CoV-2 in Kazakhstan. These findings can help shape public health policies and therapeutic interventions both in Kazakhstan and in the neighbouring countries with similar ethnic, socio-cultural and healthcare profiles, thus meriting consideration by a broad spectrum of international public health authorities and policy makers.



**Figure 5.** **A.** Phylogenetic tree depicting the Kazakhstan (“Kaz”) virus isolates in the context of globally circulating SARS-CoV-2 lineages. Each clade is denoted by a corresponding Global Initiative on Sharing All Influenza Data (GISAID) clade name; the Pangolin lineage names are given in brackets. Branch lengths measured in units of substitutions per site. Tree is coloured based on the GISAID nomenclature (see legend). **B.** Phylodynamic analysis of the Kazakhstan SARS-CoV-2 sequences in the international context. Maximum likelihood tree of Kazakhstan viral sequences and a subset of international sequences (see Methods), coloured by region of origin.

## Funding

Genomic studies were funded by the Ministry of Education and Science of the Republic of Kazakhstan (Grant# AP08052352).

## Author Contributions

SY, MG and RI: conceived, designed, and implemented the study, drafted the manuscript. SY, MG, RI: analysed the patient data. SY, DB: performed statistical analyses. SY: contributed to genomic analyses. SVG, DB: performed genomic analyses. AS and the CGRG group performed sequencing experiments and did initial genomic data analysis. KSM provided laboratory advice, and clinical guidance and contributed to data analysis. the SCERG group extracted data from electronic medical records. RI, YZ: facilitated data acquisition, and supervised the study. All authors contributed to data interpretation, critically reviewed the manuscript draft, and approved the final version for submission.

## Data sharing

The source data and R code used for all analyses are available through <https://github.com/dimbage/COVID-19-in-KZ/find/main>. SARS-CoV-2 genomic data have been submitted to GISAID under accession IDs# EPI\_ISL\_435045-435048; EPI\_ISL\_454497-454520; EPI\_ISL\_454571-454604. Any additional data from this study will be made available, wherever possible, on appropriate request to the corresponding author.

## Supplementary information

All supplementary information can be found in the Appendix.

## Declaration of Competing Interest

The authors declare that they have no competing interests.

## Acknowledgements

We thank all the participants, public health, and clinical and laboratory staff, who have been involved in COVID-19 diagnostic testing, clinical care, and case management in Kazakhstan. The authors acknowledge the contributions from other laboratories to GISAID (see appendix pp. 52-64).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.lanepe.2021.100096](https://doi.org/10.1016/j.lanepe.2021.100096).

## References

- [1] Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 2020; published online Oct 6. DOI:10.1038/s41579-020-00459-7.
- [2] Hopman J, Allegranzi B, Mehtar S. Managing COVID-19 in low- and middle-income countries. *JAMA* 2020;323:1549-50.

- [3] Balakrishnan VS. COVID-19 response in central Asia. *The Lancet Microbe* 2020;1:e281.
- [4] Semenova Y, Glushkova N, Pivina L, et al. Epidemiological Characteristics and Forecast of COVID-19 outbreak in the Republic of Kazakhstan. *J Korean Med Sci* 2020;35. doi: 10.3346/jkms.2020.35.e227.
- [5] Bayesheva D, Boranbayeva R, Turdalina B, et al. COVID-19 in the paediatric population of Kazakhstan. *Paediatr Int Child Health* 2020; 1-7.
- [6] Liu D, Cui P, Zeng S, et al. Risk factors for developing into critical COVID-19 patients in Wuhan, China: a multicenter, retrospective, cohort study. *EclinicalMedicine* 2020;25. doi: 10.1016/j.eclinm.2020.100471.
- [7] Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430-6.
- [8] Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 2020;26:506-10.
- [9] Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054-62.
- [10] Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34:4121-3.
- [11] Bogner P, Capua I, Lipman DJ, Cox NJ. A global initiative on sharing avian flu data. *Nature* 2006;442:981.
- [12] Clinical management of COVID-19. <https://www.who.int/publications-detail/clinical-management-of-covid-19> (accessed Dec 14, 2020).
- [13] Guan W-J, Ni Z-Y, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;382:1708-20.
- [14] Tabata S, Imai K, Kawano S, et al. Clinical characteristics of COVID-19 in 104 people with SARS-CoV-2 infection on the diamond princess cruise ship: a retrospective analysis. *Lancet Infect Dis* 2020;20:1043-50.
- [15] Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA* 2020;323:2052-9.
- [16] Sze S, Pan D, Nevill CR, et al. Ethnicity and clinical outcomes in COVID-19: a systematic review and meta-analysis. *EclinicalMedicine* 2020; 100630.
- [17] Kazakhstan: WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int> (accessed Dec 14, 2020).
- [18] Seemann T, Lane CR, Sherry NL, et al. Tracking the COVID-19 pandemic in Australia using genomics. *Nat Commun* 2020;11:4376.
- [19] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565-74.
- [20] Munblit D, Nekliudov NA, Bugaeva P, et al. StopCOVID cohort: An observational study of 3,480 patients admitted to the Sechenov University hospital network in Moscow city for suspected COVID-19 infection. *Clin Infect Dis* 2020; published online Oct 9. DOI:10.1093/cid/ciaa1535.
- [21] Demkina AE, Morozov SP, Vladzimirsky AV, et al. Risk factors for outcomes of COVID-19 patients: an observational study of 795 572 patients in Russia. *medRxiv* 2020;: 2020.11.02.20224253.
- [22] Moiseev S, Avdeev S, Brovko M, Bulanov N, Tao E, Fomin V. Outcomes of intensive care unit patients with COVID-19: a nationwide analysis in Russia. *Anaesthesia* 2020; published online Oct 5. DOI:10.1111/anae.15265.
- [23] Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020;323:1239-42.
- [24] Popkin BM, Du S, Green WD, et al. Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships. *Obes Rev* 2020;21:e13128.
- [25] RECOVERY Collaborative Group, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with Covid-19 - preliminary report. *N Engl J Med* 2020; published online July 17. DOI:10.1056/NEJMoa2021436.
- [26] Carotti M, Salaffi F, Sarzi-Puttini P, et al. Chest CT features of coronavirus disease 2019 (COVID-19) pneumonia: key points for radiologists. *Radiol Med* 2020;: 1-11.
- [27] Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182:812-27 e19.
- [28] Lemey P, Hong SL, Hill V, et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat Commun* 2020;11:5110.
- [29] Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* 2020;395:1382-93.