

## Article

# Inferring the Association between the Risk of COVID-19 Case Fatality and N501Y Substitution in SARS-CoV-2

Shi Zhao <sup>1,2,\*</sup> , Jingzhi Lou <sup>1</sup>, Marc K. C. Chong <sup>1,2</sup> , Lirong Cao <sup>1,2</sup> , Hong Zheng <sup>1</sup>, Zigui Chen <sup>3</sup> ,  
Renee W. Y. Chan <sup>4,5,6,7</sup> , Benny C. Y. Zee <sup>1,2</sup> , Paul K. S. Chan <sup>3</sup>  and Maggie H. Wang <sup>1,2,\*</sup>

<sup>1</sup> JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China; jzl@bethbio.com (J.L.); marc@cuhk.edu.hk (M.K.C.C.); caolr@link.cuhk.edu.hk (L.C.); hongzheng@cuhk.edu.hk (H.Z.); bzee@cuhk.edu.hk (B.C.Y.Z.)

<sup>2</sup> CUHK Shenzhen Research Institute, Shenzhen 518000, China

<sup>3</sup> Department of Microbiology, Chinese University of Hong Kong, Hong Kong, China; zigui.chen@cuhk.edu.hk (Z.C.); paulkschan@cuhk.edu.hk (P.K.S.C.)

<sup>4</sup> Department of Pediatrics, Chinese University of Hong Kong, Hong Kong, China; reneewy@cuhk.edu.hk

<sup>5</sup> Hong Kong Hub of Pediatric Excellence, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China

<sup>6</sup> CUHK-UMCU Joint Research Laboratory of Respiratory Virus & Immunobiology, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China

<sup>7</sup> Li Ka Shing Institute of Health Sciences, Faculty of Medicine, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China

\* Correspondence: zhaoshi.cmsa@gmail.com (S.Z.); maggiew@cuhk.edu.hk (M.H.W.)



**Citation:** Zhao, S.; Lou, J.; Chong, M.K.C.; Cao, L.; Zheng, H.; Chen, Z.; Chan, R.W.Y.; Zee, B.C.Y.; Chan, P.K.S.; Wang, M.H. Inferring the Association between the Risk of COVID-19 Case Fatality and N501Y Substitution in SARS-CoV-2. *Viruses* **2021**, *13*, 638. <https://doi.org/10.3390/v13040638>

Academic Editor:  
Manuel Ramos-Casals

Received: 18 March 2021  
Accepted: 6 April 2021  
Published: 8 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** As COVID-19 is posing a serious threat to global health, the emerging mutation in SARS-CoV-2 genomes, for example, N501Y substitution, is one of the major challenges against control of the pandemic. Characterizing the relationship between mutation activities and the risk of severe clinical outcomes is of public health importance for informing the healthcare decision-making process. Using a likelihood-based approach, we developed a statistical framework to reconstruct a time-varying and variant-specific case fatality ratio (CFR), and to estimate changes in CFR associated with a single mutation empirically. For illustration, the statistical framework is implemented to the COVID-19 surveillance data in the United Kingdom (UK). The reconstructed instantaneous CFR gradually increased from 1.0% in September to 2.2% in November 2020 and stabilized at this level thereafter, which monitors the mortality risk of COVID-19 on a real-time basis. We identified a link between the SARS-CoV-2 mutation activity at molecular scale and COVID-19 mortality risk at population scale, and found that the 501Y variants may slightly but not significantly increase 18% of fatality risk than the preceding 501N variants. We found no statistically significant evidence of change in COVID-19 mortality risk associated with 501Y variants, and highlighted the real-time estimating potentials of the modelling framework.

**Keywords:** COVID-19; SARS-CoV-2; N501Y substitution; B.1.1.7 lineage; case fatality; statistical modelling

## 1. Introduction

The coronavirus disease 2019 (COVID-19), whose etiological agent is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1], poses a serious threat to global health and spread to over 200 countries globally [2]. As of 28 February 2021, over 110 million COVID-19 cases were confirmed in the world with over 2.5 million associated deaths [3]. The emerging mutation in SARS-CoV-2 genomes is one of the major challenges against the control of pandemic. Around September 2020, genetic variants carrying the N501Y substitution on the spike (S) protein of SARS-CoV-2 were first detected in the United Kingdom (UK) [4], then spread to elsewhere globally, and trended to reach fixation rapidly in many places, for example, South Africa [5], Brazil [6], and the U.S. [7]. In the UK, besides other variants, the 501Y variants, as well as  $\Delta 69/\Delta 70$  deletion, were clustered into the

B.1.1.7 lineage by COVID-19 Genomics Consortium UK, which is also known as the variant of concern (VoC) 202012/01 [8]. These emerging new variants are of public health concerns to affect the epidemiological characteristics of COVID-19 [9,10], as well as the effectiveness of vaccines [11–14].

Recent analyses reported evidence that the 501Y variants were associated with an increase in transmissibility of SARS-CoV-2 at population scale [15–17], and similar findings were also reported for the D614G substitution in previous studies [18–21]. By contrast, whether the N501Y substitution leads to a difference in disease severity among SARS-CoV-2 infections remains largely unassessed. On one hand, exploring the relationship between the mutation activities and the severity of COVID-19 outcomes is of importance to assess the risks posed by the evolution of SARS-CoV-2, which may inform case management strategies and the healthcare decision-making process. On the other hand, together with some recent scientific preprints, the relevant peer-reviewed literature found either no significant evidence or a weakly positive association between clinical severity and N501Y [17,22–24]. Concerning the heterogeneities in epidemiological settings and analytical schemes of these studies, the existing evidence appears inconclusive.

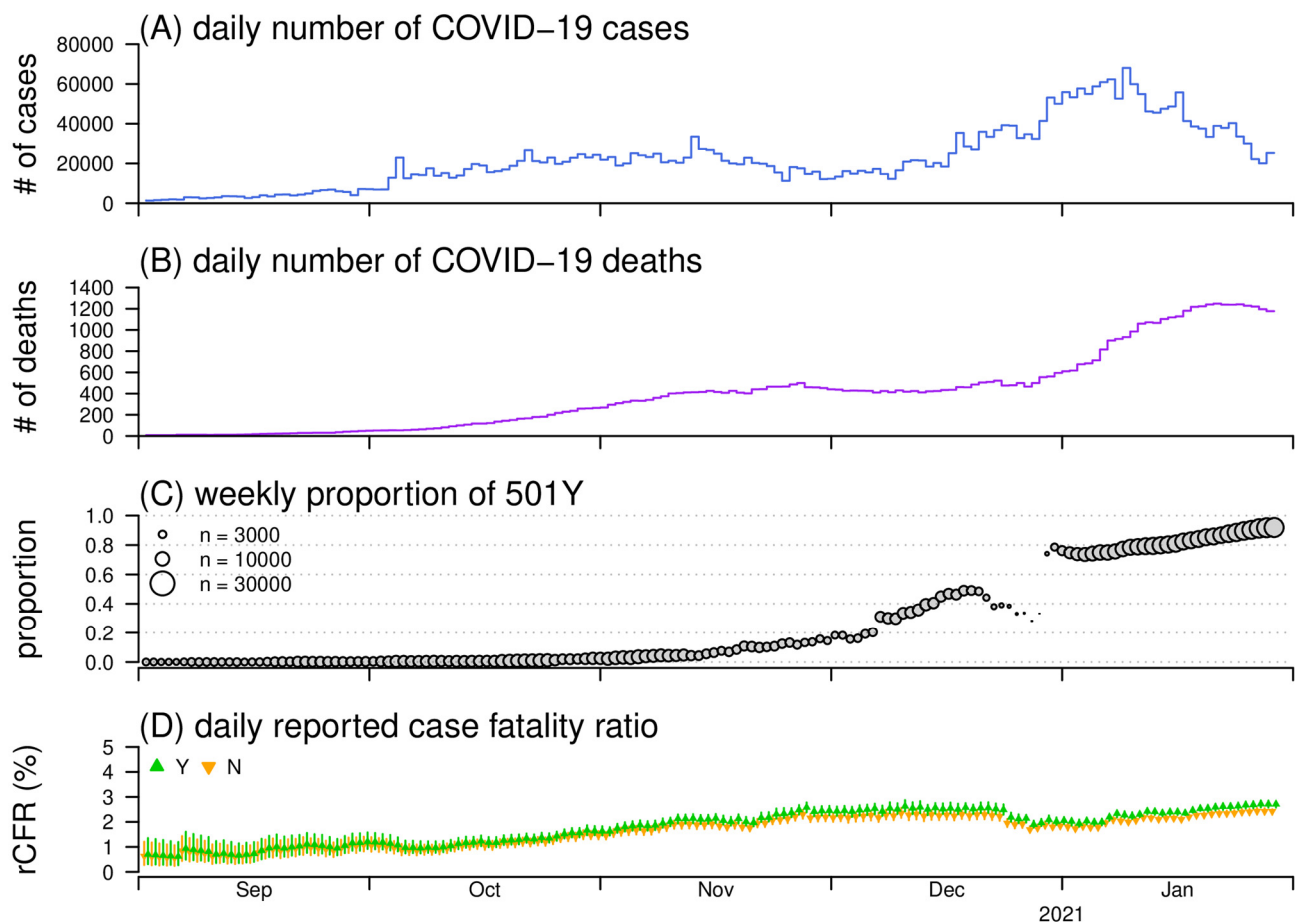
Classic epidemiological investigations, for example, cohort or case-control studies, using individual patients' information with a large sample size may explore the evidence about the association between severity of COVID-19 and N501Y directly. A recent study adopted individual patients' data from more than 2 million SARS-CoV-2 infections in the UK, and found significant positive association of B.1.1.7 lineage (which is dominant by N501Y substitution) and the fatality risk in terms of the hazard and absolute risks [25]. However, these samples of patients, as required for classic epidemiological study design, are financially expensive and difficult to collect in a short period, and commonly unavailable during the emerging stage of exposure variables, for example, the 501Y variant in this study. In this study, we adopt a statistical inference framework to assess the risk of COVID-19 case fatality using the disease surveillance data on a real-time basis. We reconstruct the real-time and variant-specific case fatality ratio (CFR) of COVID-19 empirically, and infer the change in fatality risk associated with N501Y substitution in SARS-CoV-2.

## 2. Methods

### 2.1. SARS-CoV-2 Sequencing Data and COVID-19 Surveillance Data

The SARS-CoV-2 strains were obtained via the global initiative on sharing all influenza data (GISAID) with collection dates ranging from 1 September 2020 to 31 January 2021 in the UK [26]. A total of 182,982 complete human SARS-CoV-2 strains were retrieved. All SARS-CoV-2 strains used for analysis are provided in the appendix. Multiple sequence alignment was performed using MAFFT version 7 [27], and the “Wuhan-Hu-1” genome (GenBank: “NC\_045512.2”, GISAID: “EPI\_ISL\_402125”) is considered as the reference sequence. We excluded sequences with more than 5% ambiguous amino acids during the alignment, and a total of 149,789 sequences were included for further analysis. The acknowledgment table for the sequencing data used can be found in the Supplementary Materials.

The surveillance data of COVID-19 cases and deaths in the UK were collected from the World Health Organization (WHO) COVID-19 surveillance platform [3]. Especially, to adjust for the weekly cycle in the COVID-19 deaths time series, the 7-day moving average was adopted to smooth the time series for further analysis. The COVID-19 outbreaks in the UK can be considered as two epidemics waves, and the N501Y substitution is observed during the second epidemic wave, which started in September 2020. To avoid the under-ascertainment due to reporting delays, we drop the observations in February 2021. Hence, the surveillance data of COVID-19 cases from 1 September 2020 to 31 January 2021 are included in the analysis, which matches the period of the SARS-CoV-2 sequencing data. The COVID-19 cases and deaths time series are shown in Figure 1A,B, respectively.



**Figure 1.** The daily number of COVID-19 cases (A) and deaths (B), proportion of the 501Y SARS-CoV-2 variants (C), and the reconstructed reported case fatality ratios (rCFR) (D). Panel (A) and (B) show the daily number of COVID-19 cases and deaths time series in the UK, respectively. Panel (C) shows the observed proportion of 501Y variants among all samples of SARS-CoV-2 strains, where the size of the dot indicates the weekly sample size. Panel (D) shows the estimated rCFR of 501N (in orange) and rCFR of 501Y (in green), where the dots are the point estimates and bars are the 95% CIs.

## 2.2. Statistical Parameterization

### 2.2.1. Reconstruction of the Instantaneous Case Fatality Ratio

We consider the time delay between the ascertainment of COVID-19 case and death. For an individual who deceased due to COVID-19, we used  $f$  to denote the probability distribution function (PDF) of the time interval (denoted by  $g$ ) between being reported as a case and the death. To formulate  $f(g)$ , we denoted the time interval between symptom onset and death by  $s$  following the PDF of  $h(s)$ , and the time interval between symptom onset and being reported by  $q$  following the PDF of  $\delta(q)$ . Following the previous study [28,29], the  $g$  is the difference of  $s$  minus  $q$ , i.e.,  $g = s - q$ , and we remark that  $g$  is not necessarily positive. Then, the PDF of  $g$ , that is,  $f(g)$ , is formulated as in Equation (1).

$$f(g) = \int h(g+q) \cdot \delta(q) dq. \quad (1)$$

For COVID-19, we set  $h(s)$  as a Gamma distribution with mean ( $\pm$ SD) at 20 days ( $\pm$ 10) referring to [30,31], and  $\delta(q)$  as another Gamma distribution with mean ( $\pm$ SD) at 7 days ( $\pm$ 4) referring to [28]. Although the analytical formula in Equation (1) is difficult to derive, we approach  $f(g)$  by applying Monte Carlo methods, which is also adopted in [32]. Slight changes or similar alternative settings in  $h(s)$  or  $\delta(q)$  will not affect the main results.

Since each disease-related death is diagnosed as a case in the first place, each individual case is considered as a “source” (or “pool”) of the death, that is, subjects at mortality risk.

We consider all reported cases as the pool to generate deaths, and we model this candidate pool as a time-varying function denoted by  $\Phi(t)$  at calendar time  $t$ . In other words, if one case is reported at time  $\tau$  who dies eventually, the value of  $f(g)$  is considered as the relative likelihood of death at calendar time  $(\tau + g)$ . Similar ideation was adopted to quantify the relative likelihood of transmission in [33]. Thus, the  $i$ -th case (out of all cases), who is reported at calendar time  $\tau_i$ , contributes  $f(g = t - \tau_i)$  to  $\Phi(t)$  at calendar time  $t$ . For the contribution from all reported cases, that is, for all  $i$ s, the  $\Phi(t)$  is summated as follows Equation (2).

$$\Phi(t) = \sum_i f(g = t - \tau_i). \quad (2)$$

The information of  $\tau_i$  for all  $i$ s can be obtained from the epidemic curve shown in Figure 1A. Hence, the instantaneous (or real-time) reported case fatality ratio (rCFR), that is, fatality ratio among reported cases, at calendar time  $t$  can be calculated by the ratio  $\text{rCFR}_t = d_t / \Phi_t$ . Here, the  $d_t$  is the observed number of deaths, and  $\Phi_t$  is the discretized  $\Phi(t)$  at calendar date  $t$ , which means  $\Phi_t = \int_{\text{day } t} \Phi(x) dx$ . The value of  $d_t$  can be obtained from the number of COVID-19 deaths time series as shown in Figure 1B. We remark that different SARS-CoV-2 variants are not distinguished at this stage, and thus the rCFR here is the overall (or average) case fatality ratio for all variants.

### 2.2.2. Variant-Specific Case Fatality Ratio

To incorporate the information of SARS-CoV-2 variants, we denote the proportion (or prevalence) of the variant of concern, for example, 501Y variant in this study, at calendar time  $t$  by  $\rho(t)$ , which is time-varying. The  $\rho(t)$  can be calculated by using the SARS-CoV-2 empirically, which is shown in Figure 1C. Then, the relative likelihood of death due to the  $j$ -th variant of concern at calendar time  $t$  is formulated in Equation (3).

$$\Phi_j(t) = \sum_i f(g = t - \tau_i) \cdot \rho_j(t = \tau_i). \quad (3)$$

Here, the index  $j = 0$  for 501N variant, and  $j = 1$  for 501Y variant. Straightforwardly,  $\sum_j \rho_j(t) = 1$ , and  $\sum_j \Phi_j(t) = \Phi(t)$  for all  $t$ s.

We denote the variant-specific instantaneous case fatality ratio of the  $j$ -th SARS-CoV-2 variant by  $\text{rCFR}_{j,t}$ , and then the expected number of deaths due to the  $j$ -th variant,  $E[d_{j,t}]$ , at calendar date  $t$  is modelled in Equation (4).

$$E[d_{j,t}] = \text{rCFR}_{j,t} \Phi_{j,t}. \quad (4)$$

Here, the  $\Phi_j(t)$  is discretized into  $\Phi_{j,t}$  with the same fashion as for  $\Phi(t)$ .

### 2.2.3. Risk Ratio of Case Fatality Associated with N501Y Substitution

By definition, the case fatality ratio is a measurement that quantifies the mortality risk due to SARS-CoV-2 infection. Then, the difference in the variant-specific case fatality ratios characterizing the change of mortality risk associated with N501Y substitution in SARS-CoV-2. Thus, the risk ratio (RR) of N501Y substitution, denoted by  $\zeta$  ( $\geq 0$ ), can be defined as  $\zeta = \text{rCFR}_{j=1,t} / \text{rCFR}_{j=0,t}$ . We consider  $\delta$  as a constant, which reflects the intrinsic nature of the new SARS-CoV-2 variant and thus is invariant with time. Hence, we have  $\text{rCFR}_{j=1,t} = \zeta \times \text{rCFR}_{j=0,t}$  for all calendar date  $t$ , which subject to  $0 \leq \text{rCFR}_{j=1,t} \leq 1$ . Straightforwardly,  $\zeta > 1$  indicates that N501Y substitution is associated with an increase in case fatality risk,  $\zeta < 1$  indicates that N501Y substitution is associated with a decrease in case fatality risk, and  $\zeta = 1$  indicates that N501Y substitution is not associated with any increase or decrease in case fatality risk.

Similar to Equation (4), we calculate the expected number of COVID-19 deaths,  $E[d_t]$ , at calendar date  $t$  in Equation (5).

$$E[d_t] = \text{rCFR}_{j=0,t} \Phi_{j=0,t} + \zeta \cdot \text{rCFR}_{j=0,t} \Phi_{j=1,t} = \text{rCFR}_{j=0,t} \cdot (\Phi_{j=0,t} + \zeta \cdot \Phi_{j=1,t}). \quad (5)$$

We remark that as defined, the condition for term  $\zeta \times \text{rCFR}_{j=0,t} = \text{rCFR}_{j=1,t} \leq 1$  should always be guaranteed.

### 2.3. Likelihood-Based Inference Framework

To construct the likelihood profile, we model the number of deaths due to the  $j$ -th variant,  $d_{j,t}$ , as a binomial process with sizes at  $\Phi_{j,t}$  (rounding to the closest integer) and successful probabilities at  $\text{rCFR}_{j,t}$  for each independent Bernoulli trial. Then, the number of COVID-19 deaths,  $d_t$ , follows a Poisson binomial process,  $\pi_t(\cdot)$ , which contains  $\Phi_{j=0,t}$  independent Bernoulli trials with probability of success at  $\text{rCFR}_{j=0,t}$  and  $\Phi_{j=1,t}$  independent Bernoulli trials with probability of success at  $(\text{rCFR}_{j=1,t} = \zeta \times \text{rCFR}_{j=0,t})$  for all  $t$ s. The expectation of  $d_t$  is given in Equation (5). As such, for calendar date  $t$ , we construct the loglikelihood function  $\ell_t$  of the observed daily number of COVID-19 deaths  $d_t$  using the Poisson binomial framework in Equation (6).

$$\ell_t(d_t | \zeta, \text{rCFR}_{j=0,t}, \Phi_{j=0,t}, \Phi_{j=1,t}) = \log[\pi_t(d_t | \zeta, \text{rCFR}_{j=0,t}, \Phi_{j=0,t}, \Phi_{j=1,t})]. \quad (6)$$

The detailed formulation of Poisson binomial distribution  $\pi(\cdot)$  can be found in [34] as well as elsewhere in statistical literature, which is relatively lengthy and thus is omitted in this study.

With Equation (6), we reconstruct the  $\text{rCFR}_{j=0,t}$  time series, denoted by  $\{\text{rCFR}_{j=0,t}\}$ , and estimate  $\zeta$  using the overall loglikelihood function ( $\ell$ ) defined in Equation (7).

$$\ell(\{\text{rCFR}_{j=0,t}\}, \zeta | \{d_t\}, \{\Phi_{j=0,t}\}, \{\Phi_{j=1,t}\}) = \sum_t \ell_t. \quad (7)$$

Hence, by fitting to the daily number of deaths time series, the  $\text{rCFR}$  and  $\zeta$  can be estimated by using the maximum likelihood estimation approach. The 95% confidence intervals (95% CI) are calculated by using the profile likelihood estimation framework with a cutoff threshold determined by a Chi-square quantile [35], as well as previously adopted in [36–42].

### 2.4. Sensitivity Analysis

Sensitivity analysis was carried out on the robustness and significance of the association between  $\text{rCFR}$  of COVID-19 and SARS-CoV-2 mutation activity. We examined the consistency of both the sign (or comparing with 1 for RR) and the 95%CI of the effect in case fatality risk associated with N501Y substitution under the alternative settings. We consider two sensitivity checking schemes as follows.

For the first scheme, we repeat the estimating process of  $\zeta$  with alternative PDF of and  $\delta(q)$ , which is introduced in Section 2.2.1. We consider short and long versions of the time interval between symptom onset and being reported ( $q$ ) with means at 5 and 9 days, respectively.

For the second scheme, we consider a univariate logistic regression model between  $\text{rCFR}_t$  (see Section 2.2.1) as response and  $\rho_{j=1,t}$  (see Section 2.2.2) as regressor. The regression coefficient of  $\rho$ , is evaluated as the effect size of interest.

## 3. Results and Discussion

For the second epidemic wave in the UK as shown in Figure 1A, the epidemic curve grew gradually from September to December 2020, peaked in early January 2021 with the daily number of COVID-19 cases over 60,000, and declined thereafter. The number of COVID-19 deaths time series presents a similar trend as the cases curve with a lag accounting for the progression of COVID-19 (see Figure 1B). The 501Y variants emerged around September, and maintained at a relatively low prevalence below 5% until November 2020 (see Figure 1C). Then, the 501Y variants rapidly increased and dominated with prevalence at 50% by the end of December 2020, and trended to reach fixation after January 2021.

We estimate the rCFR at 2.0% (95% CI: 1.8, 2.1) for all reported COVID-19 cases in the UK. In previous studies, the fatality ratio estimates vary across a wide range due to different concerns of “case”, which range from 0.7% to 1.3% among all SARS-CoV-2 infections [31,43], from 1.4% to 2.6% among clinically diagnosable cases [30,31,43,44], from 3.9% to 8.4% among reported cases [28,45–48], at 14% among hospitalized cases [49]. Most of these CFR estimates are calculated in the early outbreaks or during the first half of 2020, when the detecting rate of SARS-CoV-2 infections might not be as high as that after September 2020 (our study period). Thus, given that the case ascertainment efforts were largely improved in the UK, we consider that each clinically diagnosable case was likely reported. As such, the rCFR estimates in this work should be interpreted as the fatality ratio among those clinically diagnosable COVID-19 cases, which is largely in line with the existing estimates [30,31,43,44].

We reconstruct the daily instantaneous rCFR of 501N and 501Y SARS-CoV-2 variants, respectively (see Figure 1D). The overall trends of rCFR gradually increased from 1.0% in September to 2.2% in November 2020, and stabilized at this level thereafter. We suspect that the increase trend of rCFR between September and November might associate with the decreasing air temperature during the same period in the UK, about which the evidence is found in previous studies for COVID-19 [50–54]. The instantaneous CFR estimates may also have the potential to monitor the mortality risk on a real-time basis, and to further examine the associations with its potential determinants, for example, pathogenic evolution, supply of critical care resources [55] and exposure to environmental factors [56].

For the variant-specific change in CFR, we infer  $\zeta$  at 1.18 (95% CI: 0.40, 3.28) indicating that the 501Y variants may increase 18% of fatality risk among the reported cases compared with the original 501N variants. Hence, in Figure 1D, the rCFR of the 501Y variant appears higher than that of the 501N variant. By rapidly screening recent peer-reviewed literature, we identified 4 relevant studies that investigated the risk of clinical severity associated with N501Y, or alternatively, B.1.1.7, variants [17,22–24], and they detected either no statistically significant evidence (3 out of 4) or weakly positive association (1 out of 4). Together with our estimates, the concern is that the new 501Y variants might slightly increase the risks of mortality among COVID-19 patients, and more solid evidence is recently affirmed by [25]. For sensitivity checking, we find that the  $\zeta$  estimates are consistently but significantly larger than 1 in similar scales as the main estimates (data not shown), which validates our findings.

A similar phenomenon was also reported regarding the previous D614G substitution. Consistent evidence of the transmission advantage of 614G variants was reported both statistically [18,21,57] and experimentally [19,58–61], but less evidence of clinical severity linked with 614G was found [62]. The 614G variants rapidly increased since March 2020, and replaced preceding 614D variants by reaching fixation after June 2020. The D614G substitution is believed to increase the intensity of the COVID-19 pandemic during the same period, which includes increasing COVID-19 deaths globally. In the same sense, provided the transmission advantage of 501Y [15–17], the increasing intensity of COVID-19 related mortality is of public health concern as 501Y variants are replacing their predecessor. The real-time assessment of the change in mortality risk may inform the case-management strategies and healthcare planning against the consequences from mutated SARS-CoV-2 strains. As such, we highlight the importance of our analytical framework, and the public health risks related to viral mutations may be controllable with early preparedness.

This study has the following limitations. First, we presume that COVID-19 induced deaths and their time of death are correctly reported. This setting is practically reasonable since the case fatality is one of the most serious clinical outcomes, which is under more rigorous surveillance, and thus is unlikely mis-ascertained. Second, for the variant-specific rCFR in Section 2.2.2, Equation (2) holds when the COVID-19 cases and SARS-CoV-2 strains match along the same timeline. If one considers a constant reporting lag,  $q$ , the reported cases time series will have the same trends as the case time series by onset but shifted for the reporting lag. Considering the similar reporting delay also occurred for the

SARS-CoV-2 sequencing data, the effects of the two reporting lags may be counteracted. Given the  $\zeta$  estimates consistently hold in sensitivity analysis, that is, for a different distribution of  $q$ , we remark that this limitation is unlikely to affect the main conclusions in this study. Furthermore, with detailed reporting lag information of each individual case, adjustment for reporting delay can surely be carried out based on the current analytical framework. Third, the reconstruction of  $rCFR_t$  relies on the settings of  $h(s)$  and  $\delta(q)$  (see Section 2.2.1). We model both  $h(s)$  and  $\delta(q)$  of COVID-19 as two fixed Gamma distributions, which follow previous studies [28,30,31]. In the real-world situation, the distribution of  $s$  or  $q$  might be time-varying [63], which may affect the reconstruction of  $rCFR_t$ . However, the overall trends of  $rCFR_t$  estimates are unlikely changed by a slight variation in  $h(s)$  or  $\delta(q)$ , and similar situations have been studied for other epidemiological measurements, for example, reproduction number [63]. Thus, we consider the impact of this limitation on the inference of variant-specific change in mortality risk may be negligible, and our model can be extended to a more complex time-varying context about the model settings. Fourth, the distribution of  $s$  or  $q$  might be altered by the mutated variants in theory. However, by screening the literature of COVID-19, we find no evidence that  $h(s)$  or  $\delta(q)$  is varied associated with the N501Y substitution, and thus we adopted fixed distributions to govern the process. Fifth, this study focuses on exploring the effects on changing the case fatality risk associated with a single mutation, that is, N501Y, but the intrinsic biological mechanisms are commonly more complex and remain uncovered. As an example of the influenza virus, on one hand, the R384G substitution in H3N2 enhances the ability of in-host immune-escape [64], which indicates an increase in infectivity [65], but this substitution appears detrimental. On the other hand, the co-mutations of R384G in nucleoprotein (NP) could improve and compensate the viral fitness or functionality of [66,67], such that the mutated strains reached fixation rapidly in the 1993–1994 flu season. Future studies are needed for exploring the relationship of how the 501Y SARS-CoV-2 variant affects the mortality risk of COVID-19. Sixth, there are co-mutations of N501Y, for example,  $\Delta 69/\Delta 70$  deletion and E484K [4,11,15], and we remark that the independent effects of each co-mutation cannot be disentangled in this study, where the interaction among these co-mutations is unassessed. Seventh, due to the lack of individual patients' information, time-series data was used in this work, which means information loss from the data aggregation. Thus, estimating the effects of more types of variants could encounter identification issues that the samples might fail to inform each estimate. Lastly, as a data-driven study, the estimated association should be interpreted with caution. With an ecological setting, the findings in this study cannot guarantee causality, which needs further biomedical experiments in more sophisticated contexts.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/v13040638/s1>.

**Author Contributions:** M.C. Methodology, M.C., L.C., H.Z., Z.C., R.W.Y.C., B.C.Y.Z. and P.K.S.C. writing—review & editing. S.Z. conceived the study, carried out the analysis, and drafted the first manuscript. S.Z. and J.L. collected and processed the data. S.Z. and M.H.W. discussed the results. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by CUHK grant [PIEF/Ph2/COVID/06, 4054456], the Health and Medical Research Fund (HMRF) Commissioned Research on COVID-19 [INF-CUHK-1] of Hong Kong, China, and partially supported by the National Natural Science Foundation of China (NSFC) [31871340,71974165].

**Institutional Review Board Statement:** Ethical approval is waived because all datasets used in this work are publicly available.

**Informed Consent Statement:** The numbers of COVID-19 cases and deaths and SARS-CoV-2 sequencing data were collected via public domains, and thus individual consent is not applicable.

**Data Availability Statement:** All data used in this work are publicly available, and please see the data section for details.

**Acknowledgments:** This study was conducted using the resources of Alibaba Cloud Intelligence High Performance Cluster computing facilities, which is made free for COVID-19 research. The SARS-CoV-2 protein sequences were retrieved from the global initiative on sharing all influenza data at <http://platform.gisaid.org/> (accessed on 28 February 2021). The complete acknowledgment table could be found in online Supplementary Acknowledgment Tables. We thank the contributions of the submitting and the originating laboratories.

**Conflicts of Interest:** MHW is a shareholder of Beth Bioinformatics Co., Ltd. BCYZ is a shareholder of Beth Bioinformatics Co., Ltd. and Health View Bioanalytics Ltd. Other authors declared no competing interests. The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

## References

1. Hu, B.; Guo, H.; Zhou, P.; Shi, Z.L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021**, *19*, 141–154. [[CrossRef](#)] [[PubMed](#)]
2. Wu, J.T.; Leung, K.; Leung, G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* **2020**, *395*, 689–697. [[CrossRef](#)]
3. World Health Organization. Coronavirus Disease 2019 (COVID-19) Situation Reports. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (accessed on 31 March 2021).
4. Tang, J.W.; Tambyah, P.A.; Hui, D.S. Emergence of a new SARS-CoV-2 variant in the UK. *J. Infect.* **2020**, *82*, e27–e28. [[CrossRef](#)] [[PubMed](#)]
5. Tang, J.W.; Toovey, O.T.; Harvey, K.N.; Hui, D.D. Hui DDS: Introduction of the South African SARS-CoV-2 variant 501Y. V2 into the UK. *J. Infect.* **2021**, *82*, e8–e10. [[CrossRef](#)]
6. Claro, I.M.; da Silva Sales, F.C.; Ramundo, M.S.; Candido, D.S.; Silva, C.A.; de Jesus, J.G.; Manuli, E.R.; de Oliveira, C.M.; Scarpelli, L.; Campana, G.; et al. Local Transmission of SARS-CoV-2 Lineage B.1.1.7, Brazil, December 2020. *Emerg. Infect. Dis.* **2021**, *27*, 970–972. [[CrossRef](#)]
7. Galloway, S.E.; Paul, P.; MacCannell, D.R.; Johansson, M.A.; Brooks, J.T.; MacNeil, A.; Slayton, R.B.; Tong, S.; Silk, B.J.; Armstrong, G.L.; et al. Emergence of SARS-CoV-2 b. 1.1. 7 lineage—united states, December 29, 2020–January 12, 2021. *Morb. Mortal. Wkly. Rep.* **2021**, *70*, 95. [[CrossRef](#)] [[PubMed](#)]
8. Public Health England. Investigation of novel SARS-COV-2 variant Variant of Concern 202012/01: Technical briefing 1. Available online: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/959438/Technical\\_Briefing\\_VOC\\_SH\\_NJL2\\_SH2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf) (accessed on 31 March 2021).
9. Walensky, R.P.; Walke, H.T.; Fauci, A.S. SARS-CoV-2 Variants of Concern in the United States—Challenges and Opportunities. *JAMA* **2021**, *11*, 1037–1038. [[CrossRef](#)]
10. Rondinone, V.; Pace, L.; Fasanella, A.; Manzulli, V.; Parisi, A.; Capobianchi, M.R.; Ostuni, A.; Chironna, M.; Caprioli, E.; Labonia, M.; et al. VOC 202012/01 Variant Is Effectively Neutralized by Antibodies Produced by Patients Infected before Its Diffusion in Italy. *Viruses* **2021**, *13*, 276. [[CrossRef](#)] [[PubMed](#)]
11. Xie, X.; Liu, Y.; Liu, J.; Zhang, X.; Zou, J.; Fontes-Garfias, C.R.; Xia, H.; Swanson, K.A.; Cutler, M.; Cooper, D.; et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **2021**, 1–2. [[CrossRef](#)]
12. Moore, J.P.; Offit, P.A. SARS-CoV-2 Vaccines and the Growing Threat of Viral Variants. *JAMA* **2021**, *325*, 821–822. [[CrossRef](#)]
13. Muik, A.; Wallisch, A.K.; Sanger, B.; Swanson, K.A.; Muhl, J.; Chen, W.; Cai, H.; Maurus, D.; Sarkar, R.; Tureci, ˆ.; et al. Neutralization of SARS-CoV-2 lineage B. 1.1. 7 pseudovirus by BNT162b2 vaccine-elicited human sera. *Science* **2021**, *37*, 1152–1153. [[CrossRef](#)] [[PubMed](#)]
14. Supasa, P.; Zhou, D.; Dejnirattisai, W.; Liu, C.; Mentzer, A.J.; Ginn, H.M.; Zhao, Y.; Duyvesteyn, H.M.; Nutalai, R.; Tuekprakhon, A.; et al. Reduced neutralization of SARS-CoV-2 B. 1.1. 7 variant by convalescent and vaccine sera. *Cell* **2021**. [[CrossRef](#)] [[PubMed](#)]
15. Leung, K.; Shum, M.H.; Leung, G.M.; Lam, T.T.; Wu, J.T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill. Bull. Eur. Mal. Transm. = Eur. Commun. Dis. Bull.* **2021**, *26*, 2002106.
16. Zhao, S.; Lou, J.; Cao, L.; Zheng, H.; Chong, M.K.; Chen, Z.; Chan, R.W.; Zee, B.C.; Chan, P.K.; Wang, M.H. Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the UK: An early data-driven analysis. *J. Travel. Med.* **2021**, *28*, taab011. [[CrossRef](#)]
17. Davies, N.G.; Abbott, S.; Barnard, R.C.; Jarvis, C.I.; Kucharski, A.J.; Munday, J.D.; Pearson, C.A.; Russell, T.W.; Tully, D.C.; Washburne, A.D. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **2021**. [[CrossRef](#)] [[PubMed](#)]
18. Volz, E.; Hill, V.; McCrone, J.T.; Price, A.; Jorgensen, D.; O’Toole, A.; Southgate, J.; Johnson, R.; Jackson, B.; Nascimento, F.F.; et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **2021**, *184*, 64–75. [[CrossRef](#)]



19. Hou, Y.J.; Chiba, S.; Halfmann, P.; Ehre, C.; Kuroda, M.; Dinno, K.H.; Leist, S.R.; Schäfer, A.; Nakajima, N.; Takahashi, K. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **2020**, *370*, 1464–1468.
20. Zhang, L.; Jackson, C.B.; Mou, H.; Ojha, A.; Peng, H.; Quinlan, B.D.; Rangarajan, E.S.; Pan, A.; Vanderheiden, A.; Suthar, M.S.; et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **2020**, *11*, 6013. [[CrossRef](#)]
21. Zhao, S.; Lou, J.; Cao, L.; Zheng, H.; Chong, M.K.; Chen, Z.; Zee, B.C.; Chan, P.K.; Wang, M.H. Modelling the association between COVID-19 transmissibility and D614G substitution in SARS-CoV-2 spike protein: Using the surveillance data in California as an example. *Theor. Biol. Med. Model.* **2021**, *18*, 1–9. [[CrossRef](#)]
22. Iacobucci, G. Covid-19: New UK variant may be linked to increased death rate, early data indicate. *BMJ* **2021**, *372*, n230. [[CrossRef](#)]
23. Mahase, E. Covid-19: Sore throat, fatigue, and myalgia are more common with new UK variant. *BMJ* **2021**, *372*, n288. [[CrossRef](#)]
24. Brookman, S.; Cook, J.; Zucherman, M.; Broughton, S.; Harman, K.; Gupta, A. Effect of the new SARS-CoV-2 variant B.1.1.7 on children and young people. *Lancet Child. Adolesc. Health* **2021**, *5*, e9–e10. [[CrossRef](#)]
25. Davies, N.G.; Jarvis, C.I.; van Zandvoort, K.; Clifford, S.; Sun, F.Y.; Funk, S.; Medley, G.; Jafari, Y.; Meakin, S.R.; Lowe, R.; et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **2021**. [[CrossRef](#)]
26. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* **2017**, *22*, 30494. [[CrossRef](#)]
27. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)] [[PubMed](#)]
28. Jung, S.-M.; Akhmetzhanov, A.R.; Hayashi, K.; Linton, N.M.; Yang, Y.; Yuan, B.; Kobayashi, T.; Kinoshita, R.; Nishiura, H. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. *J. Clin. Med.* **2020**, *9*, 523. [[CrossRef](#)]
29. Zhao, S. A simple approach to estimate the instantaneous case fatality ratio: Using the publicly available COVID-19 surveillance data in Canada as an example. *Infect. Dis. Model.* **2020**, *5*, 575–579. [[CrossRef](#)] [[PubMed](#)]
30. Wu, J.T.; Leung, K.; Bushman, M.; Kishore, N.; Niehus, R.; De Salazar, P.M.; Cowling, B.J.; Lipsitch, M.; Leung, G.M. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **2020**, *26*, 506–510. [[CrossRef](#)] [[PubMed](#)]
31. Verity, R.; Okell, L.C.; Dorigatti, I.; Winskill, P.; Whittaker, C.; Imai, N.; Cuomo-Dannenburg, G.; Thompson, H.; Walker, P.G.T.; Fu, H.; et al. Estimates of the severity of coronavirus disease 2019: A model-based analysis. *Lancet Infect. Dis.* **2020**, *20*, 669–677. [[CrossRef](#)]
32. Ganyani, T.; Kremer, C.; Chen, D.; Torneri, A.; Faes, C.; Wallinga, J.; Hens, N. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* **2020**, *25*, 2000257. [[CrossRef](#)]
33. Wallinga, J.; Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **2004**, *160*, 509–516. [[CrossRef](#)] [[PubMed](#)]
34. Wang, Y.H. On the Number of Successes in Independent Trials. *Statistica Sinica.* **1993**, *3*, 295–312.
35. Fan, J.; Huang, T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **2005**, *11*, 1031–1057. [[CrossRef](#)]
36. Zhao, S.; Stone, L.; Gao, D.; Musa, S.S.; Chong, M.K.C.; He, D.; Wang, M.H. Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to 2020. *Ann. Transl. Med.* **2020**, *8*, 448. [[CrossRef](#)] [[PubMed](#)]
37. Zhao, S.; Lou, Y.; Chiu, A.P.; He, D. Modelling the skip-and-resurgence of Japanese encephalitis epidemics in Hong Kong. *J. Theor. Biol.* **2018**, *454*, 1–10. [[CrossRef](#)]
38. Lin, Q.; Chiu, A.P.; Zhao, S.; He, D. Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. *Stat. Methods Med Res.* **2018**, *27*, 1968–1978. [[CrossRef](#)]
39. Wang, K.; Zhao, S.; Liao, Y.; Zhao, T.; Wang, X.; Zhang, X.; Jiao, H.; Li, H.; Yin, Y.; Wang, M.H.; et al. Estimating the serial interval of the novel coronavirus disease (COVID-19) based on the public surveillance data in Shenzhen, China, from 19 January to 22 February 2020. *Transbound. Emerg. Dis.* **2020**, *67*, 2818–2822. [[CrossRef](#)]
40. Tariq, A.; Lee, Y.; Roosa, K.; Blumberg, S.; Yan, P.; Ma, S.; Chowell, G. Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. *BMC Med.* **2020**, *18*, 1–14. [[CrossRef](#)]
41. Bretó, C.; He, D.; Ionides, E.L.; King, A.A. Time series analysis via mechanistic models. *Ann. Appl. Stat.* **2009**, *3*, 319–348. [[CrossRef](#)]
42. He, D.; Ionides, E.L.; King, A.A. Plug-and-play inference for disease dynamics: Measles in large and small populations as a case study. *J. R. Soc. Interface* **2009**, *7*, 271–283. [[CrossRef](#)] [[PubMed](#)]
43. Russell, T.W.; Hellewell, J.; Jarvis, C.; Van Zandvoort, K.; Abbott, S.; Ratnayake, R.; Flasche, S.; Eggo, R.M.; Edmunds, W.J.; Kucharski, A.J.; et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance* **2020**, *25*, 2000256. [[CrossRef](#)]
44. Abdollahi, E.; Champredon, D.; Langley, J.M.; Galvani, A.P.; Moghadas, S.M. Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States. *Can. Med. Assoc. J.* **2020**, *192*, E666–E670. [[CrossRef](#)]

45. Leung, K.; Wu, J.T.; Liu, D.; Leung, G.M. 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*, 1382–1393. [[CrossRef](#)]
46. Yang, S.; Cao, P.; Du, P.; Wu, Z.; Zhuang, Z.; Yang, L.; Yu, X.; Zhou, Q.; Feng, X.; Wang, X.; et al. Early estimation of the case fatality rate of COVID-19 in mainland China: A data-driven analysis. *Ann. Transl. Med.* **2020**, *8*, 128. [[CrossRef](#)]
47. Pachetti, M.; Marini, B.; Giudici, F.; Benedetti, F.; Angeletti, S.; Ciccozzi, M.; Masciovecchio, C.; Ippodrino, R.; Zella, D. Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *J. Transl. Med.* **2020**, *18*, 1–7. [[CrossRef](#)]
48. Baud, D.; Qi, X.; Nielsen-Saines, K.; Musso, D.; Pomar, L.; Favre, G. Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **2020**, *20*, 773. [[CrossRef](#)]
49. Wu, P.; Hao, X.; Lau, E.H.Y.; Wong, J.Y.; Leung, K.S.M.; Wu, J.T.; Cowling, B.J.; Leung, G.M. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Eurosurveillance* **2020**, *25*, 2000044. [[CrossRef](#)]
50. Hassan, M.M.; El Zowalaty, M.E.; Khan, S.A.; Islam, A.; Nayem, R.K.; Järhult, J.D. Role of Environmental Temperature on the Attack rate and Case fatality rate of Coronavirus Disease 2019 (COVID-19) Pandemic. *Infect. Ecol. Epidemiol.* **2020**, *10*. [[CrossRef](#)] [[PubMed](#)]
51. Rahman, M.; Islam, M.; Shimanto, M.H.; Ferdous, J.; Rahman, A.A.-N.S.; Sagor, P.S.; Chowdhury, T. A global analysis on the effect of temperature, socio-economic and environmental factors on the spread and mortality rate of the COVID-19 pandemic. *Environ. Dev. Sustain.* **2020**, *2020*, 1–15. [[CrossRef](#)]
52. Perone, G. The determinants of COVID-19 case fatality rate (CFR) in the Italian regions and provinces: An analysis of environmental, demographic, and healthcare factors. *Sci. Total. Environ.* **2021**, *755*, 142523. [[CrossRef](#)] [[PubMed](#)]
53. Kifer, D.; Bugada, D.; Villar-Garcia, J.; Gudelj, I.; Menni, C.; Sudre, C.; Vučković, F.; Ugrina, I.; Lorini, L.F.; Posso, M.; et al. Effects of Environmental Factors on Severity and Mortality of COVID-19. *Front. Med.* **2021**, *7*, 607786. [[CrossRef](#)] [[PubMed](#)]
54. Kodera, S.; Rashed, E.A.; Hirata, A. Correlation between COVID-19 Morbidity and Mortality Rates in Japan and Local Population Density, Temperature, and Absolute Humidity. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5477. [[CrossRef](#)]
55. He, D.; Wang, X.; Gao, D.; Wang, J. Modeling the 2016–2017 Yemen cholera outbreak with the impact of limited medical resources. *J. Theor. Biol.* **2018**, *451*, 80–85. [[CrossRef](#)]
56. Yao, Y.; Pan, J.; Liu, Z.; Meng, X.; Wang, W.; Kan, H.; Wang, W. Temporal association between particulate matter pollution and case fatality rate of COVID-19 in Wuhan. *Environ. Res.* **2020**, *189*, 109941. [[CrossRef](#)] [[PubMed](#)]
57. Leung, K.; Pei, Y.; Leung, G.M.; Lam, T.T.Y.; Wu, J.T. Empirical transmission advantage of the D614G mutant strain of SARS-CoV-2. *medRxiv* **2020**. [[CrossRef](#)]
58. Weissman, D.; Alameh, M.-G.; De Silva, T.; Collini, P.; Hornsby, H.; Brown, R.; Labranche, C.C.; Edwards, R.J.; Sutherland, L.; Santra, S.; et al. D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe* **2021**, *29*, 23–31. [[CrossRef](#)]
59. Yurkovetskiy, L.; Wang, X.; Pascal, K.E.; Tomkins-Tinch, C.; Nyalile, T.P.; Wang, Y.; Baum, A.; Diehl, W.E.; Dauphin, A.; Carbone, C.; et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **2020**, *183*, 739–751. [[CrossRef](#)] [[PubMed](#)]
60. Plante, J.A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B.A.; Lokugamage, K.G.; Zhang, X.; Muruato, A.E.; Zou, J.; Fontes-Garfias, C.R.; et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **2021**, *592*, 116–121. [[CrossRef](#)]
61. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827. [[CrossRef](#)] [[PubMed](#)]
62. Groves, D.C.; Rowland-Jones, S.L.; Angyal, A. The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design. *Biochem. Biophys. Res. Commun.* **2021**, *538*, 104–107. [[CrossRef](#)]
63. Ali, S.T.; Wang, L.; Lau, E.H.Y.; Xu, X.K.; Du, Z.; Wu, Y.; Leung, G.M.; Cowling, B.J. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **2020**, *369*, 1106–1109. [[CrossRef](#)] [[PubMed](#)]
64. Berkhoff, E.G.M.; Boon, A.C.M.; Nieuwkoop, N.J.; Fouchier, R.A.M.; Sintnicolaas, K.; Osterhaus, A.D.M.E.; Rimmelzwaan, G.F. A Mutation in the HLA-B\*2705-Restricted NP383-391 Epitope Affects the Human Influenza A Virus-Specific Cytotoxic T-Lymphocyte Response in Vitro. *J. Virol.* **2004**, *78*, 5216–5222. [[CrossRef](#)] [[PubMed](#)]
65. Gog, J.R.; Rimmelzwaan, G.F.; Osterhaus, A.D.M.E.; Grenfell, B.T. Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11143–11147. [[CrossRef](#)] [[PubMed](#)]
66. Rimmelzwaan, G.F.; Berkhoff, E.G.M.; Nieuwkoop, N.J.; Fouchier, R.A.M.; Osterhaus, A. Functional compensation of a detrimental amino acid substitution in a cytotoxic-T-lymphocyte epitope of influenza A viruses by comutations. *J. Virol.* **2004**, *78*, 8946–8949. [[CrossRef](#)]
67. Rimmelzwaan, G.F.; Berkhoff, E.G.M.; Nieuwkoop, N.J.; Smith, D.J.; Fouchier, R.A.M.; Osterhaus, A.D.M.E. Full restoration of viral fitness by multiple compensatory co-mutations in the nucleoprotein of influenza A virus cytotoxic T-lymphocyte escape mutants. *J. Gen. Virol.* **2005**, *86*, 1801–1805. [[CrossRef](#)] [[PubMed](#)]