

Distribution of Transmission Potential During Nonsevere COVID-19 Illness

Nabin K. Shrestha,^{1,®} Francisco Marco Canosa,¹ Amy S. Nowacki,² Gary W. Procop,³ Sherilynn Vogel,³ Thomas G. Fraser,¹ Serpil C. Erzurum,⁴ Paul Terpeluk,⁵ and Steven M. Gordon¹

¹Department of Infectious Diseases, Cleveland Clinic, Cleveland, Ohio, USA, ²Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio, USA, ³Department of Pathology and Laboratory Medicine, Cleveland Clinic, Cleveland, Ohio, USA, ⁴Department of Pulmonary Medicine, Cleveland Clinic, Cleveland, Ohio, USA, and ⁵Department of Occupational Health, Cleveland Clinic, Cleveland, Ohio, USA, Ohio, USA, ⁴Department of Pulmonary Medicine, Cleveland Clinic, Cleveland, Ohio, USA, and ⁵Department of Occupational Health, Cleveland Clinic, Cleveland, Ohio, USA

Background. Patients recovering from coronavirus disease 2019 (COVID-19) often continue to test positive for the causative virus by polymerase chain reaction (PCR) even after clinical recovery, thereby complicating return-to-work plans. The purpose of this study was to evaluate transmission potential of COVID-19 by examining viral load with respect to time.

Methods. Health care personnel (HCP) at Cleveland Clinic diagnosed with COVID-19, who recovered without needing hospitalization, were identified. Threshold cycles (Ct) for positive PCR tests were obtained and viral loads calculated. The association of viral load with days since symptom onset was examined in a multivariable regression model, which was reduced by stepwise backward selection to only keep variables significant at a level of .05. Viral loads by day since symptom onset were predicted using the model and transmission potential evaluated by examination of a viral load-time curve.

Results. Over 6 weeks, 230 HCP had 528 tests performed. Viral loads declined by orders of magnitude within a few days of symptom onset. The only variable significantly associated with viral load was time since onset of symptoms. Of the area under the curve (AUC) spanning symptom onset to 30 days, 96.9% lay within the first 7 days, and 99.7% within 10 days. Findings were very similar when validated using split-sample and 10-fold cross-validation.

Conclusions. Among patients with nonsevere COVID-19, viral loads in upper respiratory specimens peak by 2 or 3 days from symptom onset and decrease rapidly thereafter. The vast majority of the viral load-time AUC lies within 10 days of symptom onset. **Keywords.** SARS virus; disease transmission; infectious; viral load; area under curve.

A large number of health care personnel (HCP) have contracted coronavirus disease 2019 (COVID-19) [1], the disease caused by the severe acute respiratory syndrome (SARS)–associated coronavirus 2 (SARS-CoV-2). When infected HCP recover, it becomes necessary to allow them back to work without placing their patients or coworkers at risk of acquiring the infection from them. In our institution, we adopted a policy which required at least 7 days since onset of symptoms, at least 3 days of improvement in respiratory symptoms and fever, and 2 negative SARS-CoV-2 polymerase chain reaction (PCR) tests in a 24-hour period, an approach consistent with Centers for Disease Control and Prevention (CDC) guidance for returning essential personnel to work at the time.

When HCP began to get repeat testing, however, we discovered that many continued to have positive PCR tests. Recognizing the reality of testing challenges in many parts of

Clinical Infectious Diseases[®] 2020;XX(XX):1–6

the country, the CDC updated its guidance for returning essential personnel to work, with an option of adopting either a test-based strategy requiring 2 negative tests 24 hours apart or, alternatively, a symptom-based or time-based strategy that would allow return to work if personnel met specified clinical criteria including at least 3 days since improvement in symptoms and at least 10 days since onset of symptoms, or at least 10 days since a positive test in the absence of symptoms [2].

A non-test-based strategy conveniently allows HCP back to work without having to contend with persistently positive test results. However, it would be more reassuring for healthcare organizations to embrace a non-test-based strategy if data showed that such an approach is unlikely to pose a risk of infection to others from returning HCP.

The risk of transmission of an infectious agent is not the same throughout the course of illness. Transmission risk at any instant would be expected to depend on the quantity of the infectious agent in relevant clinical sites at the time. Thus, for a respiratory viral infection, the risk of transmission at any instant depends on the viral load in respiratory sites at that instant. Hence, examination of viral load-time curves should provide good insight into transmission potential over different intervals during an episode of illness.

Received 13 May 2020; editorial decision 20 June 2020; accepted 23 June 2020; published online June 29, 2020.

Correspondence: N. K. Shrestha, 9500 Euclid Avenue, G-21, Cleveland, OH 44195 (shrestn@ ccf.org).

[©] The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com. DOI: 10.1093/cid/ciaa886

The purpose of this study was to evaluate transmission potential by examining viral load with respect to time since onset of symptoms in HCP infected with SARS-CoV-2.

METHODS

This was a cohort study examining viral loads in HCP diagnosed with COVID-19. The HCP cohort was chosen for the study because of documentation of the date of onset of illness in templated notes of monitoring calls from Occupational Health. The study was approved by the Cleveland Clinic Institutional Review Board (IRB no. 20–1631). A waiver of informed consent and waiver of Health Insurance Portability and Accountability Act authorization were approved to allow access to protected health information by the research team, with the understanding that sharing or releasing identifiable data to anyone other than the study team was not permitted without additional IRB approval.

Screening and Inclusion and Exclusion Criteria

The Cleveland Clinic began testing for SARS-CoV-2 by PCR on 12 March 2020, with a separate streamlined process to test HCP started at a similar time. All HCP tested between 16 March and 20 April 2020 were identified from the laboratory information system and screened for inclusion in the study. All those with at least 1 positive test were included. Those with severe illness (who required hospitalization during the course of their illness) and those for whom a symptom-onset date could not be ascertained were excluded.

Specimen Collection

Introduction of the SARS-CoV-2 PCR test was done in concert with a standardized work flow for collection of specimens. Because of overwhelming demand, and consequently unacceptable delays, a separate queue was created for HCP. Specimen collection consisted of obtaining a nasopharyngeal swab by trained personnel at a single drive-through location and transporting it to the laboratory in universal/viral transport medium.

Polymerase Chain Reaction Testing

To maintain consistency, a decision was made to test all HCP specimens on the same testing platform. The reverse-transcription (RT)–PCR test developed at the CDC was used, and internal validation studies were performed in-house before introduction of the test. Using commercially available plasmids that contained the nucleocapsid (N) gene, the limit of detection was found to be 20 copies/ μ L for upper respiratory specimens and 2 copies/ μ L for lower respiratory specimens.

For each specimen, 200 μ L of clinical specimen in universal/ viral transport medium was rendered noninfectious within a biological safety cabinet through the addition of 200 μ L of Bacterial Lysis Buffer (Roche Diagnostics, Indianapolis, IN). Nucleic acid extraction was performed using 200 μ L of the inactivated specimen using the MagNA Pure system (Roche). Five microliters of eluate was added to $15 \,\mu$ L of PCR mastermix for each of the PCR reactions. The test was designed to target 3 separate regions of the virus nucleocapsid (N) gene. An amplification control targeting a human gene, the RNase P gene, was assessed for each specimen. Thus, each test consisted of 4 separate PCRs. Testing was performed on an ABI 7500 or ABI 7500 Fast Dx (ThermoFisher, Waltham, MA). All 3 PCRs for N gene targets (ie, N1, N2, and N3) had to be positive for a specimen to be reported as positive. If only 1 or 2 of the targets generated amplification, then the specimen was considered indeterminate and repeated either with the CDC assay or the Simplexa COVID-19 Direct Kit (DiaSorin, Cypress, CA).

Using the Threshold Cycle to Calculate Viral Load

The threshold cycle (Ct) for a real-time PCR test is the cycle at which products of amplification first register at or above the detection threshold. The higher the load of the target in the specimen, the lower the Ct. The minimum detectable viral load (MDVL) is the smallest viral load that can be detected by the test. For this study, the Ct for a specimen containing the MDVL (Ct_{MDVI}) is 40 cycles, as any specimen registering a Ct higher than 40 is considered a negative test. Thus, Ct_{MDVL} – Ct is zero for a specimen containing the MDVL and is proportionally higher for any sample with a higher viral load. Each unit difference in Ct_{MDVL} – Ct represents a difference of 1 PCR cycle. Given that each PCR cycle results in a doubling of the amplified target, the quantity of the viral target as a multiple of the MDVL can be determined by the following formula: $2^{(Ct_{MDVL} - Ct)}$. If the efficiency of the PCR reaction is less than 100%, calculated numbers will be proportionally smaller, but relative differences between any 2 viral loads will still be the same. Thus 2^(Ct_{MDVL} - Ct) is the actual viral load in the specimen, expressed as a multiple of the MDVL.

Acquisition of Ct Values and Relevant Clinical Information

The Ct values for each PCR component (N1, N2, and N3) of the positive tests were obtained. The mean of the 3 Ct values was considered the Ct for the test. Relevant demographic information, comorbidity data, and date of onset of symptoms were obtained from the electronic medical record.

Covariates

In addition to time since onset of illness, factors that were considered to possibly influence viral load were age, gender, race, body mass index (BMI), and comorbid conditions such as chronic lung disease, smoking status, immunosuppressed status, chronic heart disease, chronic kidney disease, and chronic liver disease. Chronic heart disease was defined as coronary artery disease or valvular heart disease with moderate to severe stenosis or regurgitation. Chronic lung disease was defined as chronic obstructive pulmonary disease, bronchiectasis, asthma, or chronic interstitial lung disease. Chronic kidney disease was defined as an estimated glomerular filtration rate of 60 mL/minute or less, or maintenance hemodialysis or peritoneal dialysis. Chronic liver disease was defined as cirrhosis from any cause. Immunocompromised state was defined as being on biologic agent therapy or other immunosuppressive medications, having a hematological malignancy, or having a history of hematological or solid-organ transplant.

Statistical Analysis

Analyses were performed by N. K. S. and A. S. N. using R version 4.0.0 [3]. Body mass index data were missing for 15 study subjects (6.5%). Examination of missing values with margin plots from the R package *VIM* [4] showed that the missing BMI values were missing completely at random. Values for these missing values were imputed using a method of multivariate imputation by chained equations using the R package *mice* [5].

Scatterplots of viral load over time were investigated. The association of viral load with time since onset of symptoms was explored using restricted cubic spline regression with the R package splines [3]. Splines allow for the modeling of nonlinear relationships among variables. On 10-fold crossvalidation, a restricted cubic spline regression model with 4 degrees of freedom yielded the lowest error rate. This also produced a viral load-time curve that appeared to be biologically plausible (ie, viral load increasing at the time of onset of symptoms). All numeric variables, diabetes mellitus, smoking status, and other categorical variables where the smaller group contained at least 10% of subjects were considered as covariates. Variables that did not remain significant at a level of .05 were dropped by backward elimination in a stepwise fashion. Viral loads on successive days since onset of symptoms were predicted using the regression model. Areas under the viral load-time curve were determined by integration of the function representing the relationship between the predicted viral load and days since onset of symptoms. Model performance was validated using both split-sample (80:20) and 10-fold cross-validation. Graphics were created using the R base and *ggplot2* packages [3, 6].

RESULTS

During the 6-week study period, 252 HCP tested positive for SARS-CoV-2. Twenty (8%) were excluded because they were hospitalized during the course of their illness and 2 (<1%) were excluded because the date of onset of symptoms could not be ascertained on review of their records. The remaining 230 were included in the study and had a total of 528 tests performed. Of these patients, 106 (46%) had a single test done, 28 (12%) had 2, 40 (17%) had 3, 39 (17%) had 4, 12 (5%) had 5, and 5 (2%) had 6 tests done. Days from onset of symptoms to first test was

Table 1. Study Subject Characteristics

Characteristics	Value (N = 230)
Demographics	
Age, y	40 (14)
Female sex	171 (74)
Body mass index, kg/m ²	30.2 (7.8)
Race	
Caucasian	160 (70)
African-American	47 (20)
Other	23 (10)
Comorbidities	
Chronic lung disease	23 (10)
Current smoker	6 (3)
Chronic heart disease	4 (2)
Chronic kidney disease	5 (2)
Liver cirrhosis	0 (0)
Diabetes mellitus	17 (7)
Immunocompromised	4 (2)
Hypertension	37 (16)
Data are expressed as n (%) unless a unit is given, ir	which case they are expressed as

Data are expressed as n (%) unless a unit is given, in which case they are expressed as mean (SD).

distributed with a mean (SD) of 4.9 (4.9) days. Characteristics of the included HCPs are shown in Table 1.

Distribution of Viral Load Over Time Since Onset of Symptoms

Figure 1 shows the distribution of viral load against days since onset of symptoms. Each point on the scatterplot demonstrates a single test. Viral loads for tests done within the first few days after onset of symptoms have a very wide distribution, with many tests having very high values. The figure shows that viral loads decline rapidly by orders of magnitude within the first few days. When viral load trends for each individual patient are examined, a rapid decline in viral loads is noted in the majority



Figure 1. Scatterplot of viral load versus days since onset of symptoms. Points on the scatterplot represent individual tests. The *y*-axis is on a logarithmic scale. Viral load is represented as number of times the minimum detectable viral load. Negative tests are assigned a viral load of 1 to avoid a log(0) error. The shape of each point corresponds to the gender of the patient tested. Points are jittered along the *x*-axis to unmask overlap.

(see Supplementary Figure 1). Fluctuations in viral load over time seem to occur once viral loads are at lower levels, but these viral loads generally remain low.

Association of Viral Load With Time

In a restricted cubic spline regression model the only variable that remained significantly associated with viral load was time since onset of symptoms. The performance of the final model is shown in Figure 2.

Estimation of Transmission Potential Using a Viral Load-time Curve

Figure 3 shows a viral load-time curve based on viral loads predicted by the regression model. The area under the curve (AUC) represents the distribution of transmission potential over time during the course of illness. Of the AUC spanning the interval from onset of symptoms to 30 days, 86.3% lie within the first 5 days, 96.9% within the first 7 days, and 99.7% within the first 10 days. Results were very similar in an 80:20 split-sample validation, with 89.1% of the 30-day AUCs within the first 5 days, 97.3% within the first 7 days, and 99.7% within the first 5 days (see Supplementary Figure 2). The corresponding mean AUC proportion values on 10-fold cross-validation were 85.6%, 96.7%, and 99.7%, respectively.

DISCUSSION

This study shows that in otherwise healthy subjects with nonsevere illness from COVID-19, the SARS-CoV-2 viral load is very high within 2–3 days of onset of symptoms and falls rapidly by orders of magnitude within a few days. It should be noted that none of these HCP were treated with hydroxychloroquine or other COVID-19–related treatments.



Figure 2. Relationship between days since onset of symptoms and the \log_{10} viral load in a restricted cubic splines regression model with 4 degrees of freedom. The boxplots show the distribution of the \log_{10} of the actual viral load each day after onset of symptoms. Viral load is represented as number of times the minimum detectable viral load. The black circles and fitted line represent the predicted \log_{10} viral loads for each day.



Figure 3. Viral load-time curve showing the proportion of the 30-day AUC that lies within various intervals. Viral loads were predicted from the final restricted cubic spline regression model. The predicted viral load for each day was plotted against days since onset of symptoms. Viral load is represented as number of times the minimum detectable viral load. The area under the viral load-time curve was calculated by integration of the function representing the relationship between the predicted viral load and days since onset of symptoms. Abbreviations: d, days; AUC, area under the curve.

The premise of this study was that viral load in the nasopharynx should be a reasonable marker of viral loads at other locations in the respiratory tract. This is a reasonable assumption. Viral loads in saliva and endotracheal aspirates decreased proportionally over time in 1 study [7]; and viral loads in tongue, nasal, and midturbinate swabs were linearly correlated with viral loads in nasopharyngeal swabs in another [8]. Modeling suggests that the contagiousness of respiratory viral infection is proportional to the quantity of microorganism at respiratory sites [9]. Such an association of viral load with transmission risk has been shown in other viral infections such as human immunodeficiency virus (HIV) and human herpesvirus 6 (HHV6) [10, 11]. Estimates from our study imply that, in individuals with nonsevere illness, transmission potential of COVID-19 is greatly diminished by 7–10 days since the onset of symptoms.

The distribution of transmission potential found in this study is consistent with related findings in other studies. A rapid decline in viral load over a few days has been demonstrated in several studies [7, 12–16]. A positive PCR test does not necessarily correlate with presence of transmissible virus. In prior studies, live virus could not be isolated after 8 days since symptom onset even with positive PCR results [13, 17], and virus could not be isolated from specimens with Ct values of 24 or higher [17], supporting the hypothesis that transmission risk is low when viral load is low. A study that estimated infectiousness based on transmission events, not on viral loads, concluded that infectiousness declined quickly within 7 days [15]. A contact-tracing assessment found no infections among those exposed to the index cases after more than 5 days since symptom onset [18]. The main strength of this study is that it demonstrates a way to quantitate transmission potential of COVID-19 over the course of illness. It is very difficult to determine the absolute risk of transmission of any viral infection at any specified viral load. However, knowledge of relative viral loads over time can provide a very good estimate of relative transmission potential over different intervals during the course of illness. Our study methodology using Ct values from PCR tests is simple enough that any healthcare organization can replicate the study internally using local data to help guide decisions in their own organizations.

A limitation of our study is that specimens were not collected in all subjects at prespecified time points. Once 2 negative tests were obtained, the subject would not have undergone subsequent tests. Thus, data points in the study could be expected to have been increasingly biased towards higher viral loads as time from onset of symptoms increased. Thus, the further out, the more likely that the viral load would be an overestimate of the true viral load, thereby explaining the upward tick at the tail end of the curve in Figure 2. However, this limitation has minimal effect on the overall conclusion that the majority of transmission potential lies early in the course of illness.

The extremely high viral loads within 2 or 3 days since onset of symptoms suggest that viral loads may be almost as high in the immediate presymptomatic period, suggesting substantial transmission potential in the presymptomatic period. Since our study did not have actual viral loads in the presymptomatic period for any patient, concerns about presymptomatic transmission based on our study alone are hypothetical. However, presymptomatic transmission has been documented elsewhere [15, 19]. A study that examined 77 infector-infectee transmission pairs estimated that 44% of secondary cases were infected during the index cases' presymptomatic stage [15]. The potential for substantial transmission in the presymptomatic stage supports the use of masks in public during the pandemic in the current situation where there is active community transmission, the majority of the population remains susceptible, and there is no available vaccine to provide protection.

Although this study examined HCP, findings can be expected to be similar in non-HCP with nonsevere illness, thereby allowing nonhealthcare organizations to make decisions on appropriate timing for returning recovering employees to work based on this study. Findings from this study should not, however, be extrapolated to individuals who have severe illness, as it is possible such patients may have more prolonged shedding of virus [20]. Similarly, these findings cannot be used to make decisions about when isolation can be discontinued in hospitalized patients.

Testing to return personnel to work after COVID-19 poses challenges. PCR tests may continue to remain positive long after clinical recovery. Our study's finding that transmission potential is largely over by 10 days from onset of symptoms supports a policy that would allow personnel with nonsevere COVID-19 who have recovered clinically to return to work after 10 days since onset of symptoms, without a need for laboratory testing. There will undoubtedly be outliers who shed virus for longer periods of time, but it is likely that the viral load in such situations is low, and it is not clear if this represents transmissible virus [13, 17]. A requirement of wearing masks for a period of time after returning to work would protect against such outliers. In our institution, our return-to-work policy for HCP infected with COVID-19 was updated to require at least 10 days since onset of symptoms, at least 3 days since clinical recovery, and a requirement to wear a mask for 2 weeks after return, without a need for testing to return to work.

In conclusion, among patients with nonsevere COVID-19, viral loads in upper respiratory specimens peak by 2 or 3 days from onset of symptoms and decrease greatly within a few days. The vast majority of the area under the viral load-time curve lies within 10 days of onset of symptoms. These findings should help inform decisions about returning employees recovering from COVID-19 to work in both healthcare and nonhealthcare organizations.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Note

Potential conflicts of interest. S. C. E. reports grants from the National Institutes of Health during the conduct of the study and was the Chair of the ABIM Pulmonary Disease Board. All other authors report no potential conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

- Burrer SL, de Perio MA, Hughes MM, et al. Characteristics of health care personnel with COVID-19—United States, February 12–April 9, 2020. MMWR Morb Mortal Wkly Rep 2020; 69:477–81.
- Centers for Disease Control and Prevention. Return-to-work criteria for healthcare workers. 2020. Available at: https://www.cdc.gov/coronavirus/2019-ncov/ hcp/return-to-work.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc. gov%2Fcoronavirus%2F2019-ncov%2Fhealthcare-facilities%2Fhcp-return-work. html. Accessed 1 May 2020.
- 3. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, **2020**.
- Kowarik A, Templ M. Imputation with the {R} Package {VIM}. J Stat Softw 2016; 74:1–16.
- van Buuren S, Groothuis-Oudshoorn K. {mice}: multivariate imputation by chained equations in R. J Stat Softw 2011; 45:1–67.
- Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag, 2016.
- To KKW, Tsang OTY, Leung WS, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect Dis 2020; 20:565–74.
- Tu Y-P, Jennings R, Hart B, et al. Swabs collected by patients or health care workers for SARS-CoV-2 testing. N Engl J Med 2020. Published online 3 June 2020. doi:10.1056/NEJMc2016321.
- Almocera AES, Nguyen VK, Hernandez-Vargas EA. Multiscale model withinhost and between-host for viral infectious diseases. J Math Biol 2018; 77:1035–57.

- Quinn TC, Wawer MJ, Sewankambo N, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 2000; 342:921–9.
- 11. Mayer BT, Krantz EM, Wald A, et al. Estimating the risk of human herpesvirus 6 and cytomegalovirus transmission to Ugandan infants from viral shedding in saliva by household contacts. Viruses **2020**; 12:171.
- Arons MM, Hatfield KM, Reddy SC, et al; Public Health–Seattle and King County and CDC COVID-19 Investigation Team. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. N Engl J Med 2020; 382:2081–90.
- Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020; 581:465–9.
- 14. Sakurai A, Sasaki T, Kato S, et al. Natural history of asymptomatic SARS-CoV-2 infection. N Engl J Med **2020**. Published online 12 June 2020. doi:10.1056/NEJMc2013020.
- He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020; 26:672–5.

- Zou L, Ruan F, Huang M, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N Engl J Med 2020; 382:1177–9.
- Bullard J, Dust K, Funk D, et al. Predicting infectious SARS-CoV-2 from diagnostic samples. Clin Infect Dis 2020 Published online 22 May 2020. doi:10.1093/ cid/ciaa638.
- Cheng H-Y, Jian S-W, Liu D-P, et al. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA Intern Med. Published online 1 May 2020. doi:10.1001/jamainternmed.2020.2020.
- Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic transmission of SARS-CoV-2—Singapore, January 23-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020; 69:411–5.
- 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.