Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Infectious Diseases 116 (2022) 418–425

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Persistence of clinically relevant levels of SARS-CoV2 envelope gene subgenomic RNAs in non-immunocompromised individuals

Merlin Davies[1], Laura R Bramwell[1], Nicola Jeffery[1], Ben Bunce[2], Ben P Lee[1], Bridget Knight[2], Cressida Auckland[3], Jane AH Masoli[1,4], Lorna W Harries[1,*]

[1] Institute of Clinical and Biomedical Sciences, University of Exeter, RILD building, Royal Devon and Exeter Hospital (Wonford), Barrack Road, Exeter, UK
[2] NIHR Exeter Clinical Research Facility, Royal Devon and Exeter NHS Foundation Trust, Royal Devon and Exeter Hospital, Barrack Road, Exeter, UK
[3] Microbiology Department, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[4] Healthcare for Older People, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

ARTICLE INFO

Article history:
Received 2 September 2021
Revised 29 November 2021
Accepted 2 December 2021

Keywords:
SARS-COV-2
sgRNA
transmission
persistence

ABSTRACT

Objectives: This study aimed to evaluate the associations between COVID-19 severity and active viral load, and to characterize the dynamics of active SARS-CoV-2 clearance in a series of archival samples taken from patients in the first wave of COVID-19 infection in the South West of the UK.
Methods: Subgenomic RNA (sgRNA) and E-gene genomic sequences were measured in a retrospective collection of PCR-confirmed SARS-CoV-2-positive samples from 176 individuals, and related to disease severity. Viral clearance dynamics were then assessed in relation to symptom onset and last positive test.
Results: Whilst E-gene sgRNAs declined before E-gene genomic sequences, some individuals retained sgRNA positivity for up to 68 days. 13% of sgRNA-positive cases still exhibited clinically relevant levels of virus after 10 days, with no clinical features previously associated with prolonged viral clearance times.
Conclusions: Our results suggest that potentially active virus can sometimes persist beyond a 10-day period, and could pose a potential risk of onward transmission. Where this would pose a serious public health threat, additional mitigation strategies may be necessary to reduce the risk of secondary cases in vulnerable settings.

Introduction

Ongoing transmission of SARS-CoV-2 remains a problem worldwide. Patchy vaccination coverage in multiple parts of the world, vaccine hesitancy in vulnerable groups, and differences in the efficiency of available vaccines mean that virus transmission remains a problem in community and clinical settings, and in long-term care of older people. While opportunities for unfettered transmission remain, selection pressures on the evolving virus raise the risk of producing variants that significantly increase transmissibility, morbidity, or mortality, or have the ability to evade the host immune response produced by previous infection or vaccination. The impact of more transmissible variants of SARS-CoV-2 has been observed: the B.1.1.7 Kent variant, which has elevated infectivity (Leung et al., 2021), was implicated as a crucial factor in the second wave of infections in the UK (Volz et al., 2021), while variants carrying similar genetic changes were linked with the emergence of a third wave in Europe and South America (Claro et al., 2021). Regions of Brazil retained very high rates of new infection, despite their existing population-wide seropositivity of over 76% (Sabino et al., 2021). Similarly, India experienced a catastrophic second wave, despite an estimated 50% of citizens in the largest cities having had prior exposure (Mallapaty, 2021). Reaching a 'zero Covid' situation globally in the near term seems unlikely, so there remains a need for strategies to contain transmission of SARS-CoV-2 for the foreseeable future.

A key component of the prevention of SARS-CoV-2 transmission is the identification and isolation of infected individuals. A better understanding of who is likely to present an extended infectious period, and for how long, would greatly inform strategies to more effectively limit transmission. Limited findings suggest that most COVID-19 transmission occurs in the early stages of the disease, or before the onset of symptoms (He et al., 2020; Lauer et al., 2020; Shrestha et al., 2020; Xu et al., 2020; Yang et al., 2020). Most

sgRNA, subgenomic RNA.

* **Corresponding author:** Professor Lorna W. Harries, Institute of Biomedical and Clinical Sciences, University of Exeter Medical School, Barrack Road, Exeter, EX2 5DW, UK
E-mail address: L.W.Harries@exeter.ac.uk (L.W. Harries).

https://doi.org/10.1016/j.ijid.2021.12.312

evidence suggests that replication-competent virus cannot usually be recovered from individuals with mild-to-moderate COVID-19 disease beyond 10 days of symptom onset (Arons et al., 2020; Bullard et al., 2020). Other studies have suggested that whilst the shedding of viral fragments may be prolonged, the duration of viable virus is relatively short-lived (Cevik et al., 2021) and prolonged transmission potential may only be evident in individuals with severe disease (van Kampen et al., 2021) or in severely immunocompromised patients (Aydillo et al., 2020; Avanzato et al., 2020; Choi et al., 2020). However, assessments of longer-term shedding of *infectious* virus in otherwise clinically unremarkable people are lacking because the widely available tests assess only the presence of viral fragments, not replicating viral genomes. Individuals frequently test positive for viral fragments beyond 6 weeks of onset of symptoms, in a fluctuating positive/negative pattern (Jang et al., 2021). Moreover, a positive result as detected by quantitative real-time PCR (RT-qPCR) for viral genomic sequences does not represent the presence of replication-competent virus, since viral fragments may remain after viral clearance (Arons et al., 2020; Bullard et al., 2020; Lu et al., 2020). The gold standard for determining infectious potential is viral culture. However, this is not a process suitable for rapid upscaling and high throughput. Moreover, it carries an obvious infective risk to staff and few diagnostics laboratories have the facilities to undertake the necessary culture.

A better measure of infectious potential may be the assessment of subgenomic RNAs (Wolfel et al., 2020). Subgenomic RNAs are produced by the discontinuous transcription of virion structural genes during active replication, and result in the formation of rearranged template sequences that are not found in juxtaposition in the native RNA genome of the virus (Wu and White, 2007). Although the correlation between infectivity and E-gene-derived sgRNA positivity is not always well conserved (van Kampen et al., 2021; Alexandersen et al., 2020), it may be a better proxy than E-gene viral load for replication, and better suited for assessment of archival clinical samples than other techniques, such as viral culture. Our study aimed to evaluate the relationship between genomic SARS-CoV-2 E-gene sequences, E-gene-derived sgRNA SARS-CoV-2 viral sequences, disease severity, and duration of infectious period in a retrospective collection of swab RNA samples from 265 clinically confirmed COVID-19 cases from the South West of the UK during the first wave of infection, between March 17, 2020 and November 29, 2020.

Methods

Patient samples

Our sample set consisted of RNA extracted from 265 archival nasopharyngeal swab samples derived from individuals who had previously tested positive for SARS-CoV-2 sequences at the Royal Devon and Exeter Hospital between March 17, 2020 and November 29, 2020. This sample set comprised viral genomes from the A.2, B.1, B.1.1, B.1.1.10, B.1.1.162, B.1.1.206, B.1.1.261, B.1.1.52, B.1.12, B.1.275, B.1.5, B.1.91, B.1.98, B.1.28, B.1.3, B.1.35, B.1.139, B.1.48, B.1.177.9, B.1.1.315, B.1.117, and B.1.160.6 strains circulating in the UK population at the time of testing. As described previously (Ladhani et al., 2020), participants were identified during routine clinical practice, and retrospective samples were obtained through the Royal Devon and Exeter Tissue Bank (RDETB — ethical approval REC No. 11/SW/0018). The majority of individuals provided a single sample, but serial samples were available for 17 individuals, taken as part of routine clinical care. Patient characteristics are shown in Table 1. Patients were classified as asymptomatic, mild, moderate, severe, or critical disease categories according to NIH criteria (Clinical Spectrum | COVID-19 Treatment Guidelines (nih.gov).

Assessment of E-gene and E-sgRNA viral load

RNA was extracted using the Chemagic Viral DNA/RNA Kit on the Chemagic 360D platform. Viral loads were measured by quantitative real-time PCR protocols adapted from Corman et al. (Corman et al., 2020) and Wölfel et al., (Wolfel et al., 2020) for E-gene and E-gene-derived sgRNA sequences, respectively. Reactions were performed using the SuperScript™ One-Step RT-PCR System with Platinum™ *Taq* DNA Polymerase (Thermo Fisher, Waltham, USA), according to manufacturer's instructions. E-gene reactions contained 5 μL of reaction buffer, 0.4 μL of enzyme mix, 400 nM of each primer, 200 nM of probe, and 4 μL of sample in a total volume of 10 μL. Subgenomic RNA reactions contained 5 μL of reaction buffer, 0.4 μL of enzyme mix, 0.16 μL of $MgSO_4$, 400 nM of each primer, 200 nM of probe, and 4 μL of sample in a total volume of 10 μL. PCR conditions were one cycle of 50°C for 10 minutes and one cycle of 95°C for 3 minutes, followed by 45 cycles of 95°C for 10 seconds, 56°C for 15 seconds, and 72°C for 5 seconds. Reactions were carried out in 384 well plates on the QuantStudio™ 12K Flex Real-Time PCR System (Applied Biosystems, Foster City, USA). Probes and primers were obtained from Thermo Fisher (Waltham, USA), and the sequences taken from Corman et al. (Corman et al., 2020) (E gene) and Wölfel et al., (Wolfel et al., 2020) (E sgRNA). The probes in our study were labelled at the 5′ end with 6-FAM labels and at the 3′ end with QSY quenchers. E-gene and E-sgRNA viral loads were determined in relation to known quantities of synthetic standards by standard curve analysis. These synthetic standards consisted of the probe and primer binding sites in a single long oligonucleotide for each of the E-gene and E-gene-derived sgRNA sequences, and were obtained from Invitrogen (Waltham, USA). Oligonucleotide sequences are provided in Supplementary File S1. Working stock solutions of synthetic standards at 10^{10} copies per mL in TE buffer were then created, and standard curves were prepared from this stock. For quantification of sample viral load, standard curves consisting of a 10-fold serial dilution of synthetic standards from 10^8–10^1 copies/µL were then run on each sample plate. Representative standard curves are given in Supplementary Figure S1. Viral loads were then determined from test samples by absolute quantification, as determined from the standard curves.

Association of E-gene and E-sgRNA viral load in relation to disease severity

The relationship between E-gene or E-sgRNA viral load and disease severity was assessed by linear regression analysis, with adjustment for age, sex, smoking history, and number of comorbidities. Association between E-gene or E-sgRNA viral load and presence or absence of symptoms was assessed using ANOVA. Statistical analyses were carried out using the STATA SE 16 package (StataCorp, College Station, TX, USA).

Analysis of serial samples

Access to serial samples for analysis was available for 17 individuals. These samples were taken as part of routine clinical care and, as such, there was no control over the timing of repeat samples. Nevertheless, these samples provided an opportunity to explore the dynamics of viral loads in single individuals over time. E-gene and E-gene-derived sgRNA viral loads were measured in each serial sample as described above, with an observational study of the presence or absence of E-gene or E-gene-derived sgRNA positivity carried out over successive samples.

Table 1
Clinical characteristics of the study population. The anthropometric and demographic characteristics of individuals in our sample cohort of 176 PCR-confirmed SARS-CoV-2 cases are shown. SD = standard deviation, IQR = interquartile range. Where individuals had multiple samples, the initial sample in the series was used for this analysis. Where available, duration of disease was defined as the interval between onset of symptoms and a positive test, where symptom onset pre-dated testing.

Patient characteristics (*N* = 265)	
Number E-gene positive	176
Number E-sgRNA positive	72
% male	44.9
Mean age (SD)	53.9 (SD = 17.7)
Median age (IQR)	55 (IQR = 21.75)
Number of asymptomatic patients	42.0%
Number of patients categorized as having mild disease	20.0%
Number of patients categorized as having moderate disease	12.5%
Number of patients categorized as having severe disease	18.75%
Number of patients categorized as having critical disease	6.25%
Mean duration of disease (days, SD)	7.3 (8.58)
Median duration of disease (days, IQR)	7 (8)

Assessment of duration of 'replicative' virus

The dynamics of E-gene or E-gene-derived sgRNA positivity in relation to disease duration was then explored. The majority of individuals provided a single sample at diagnosis for this analysis; for those who had provided serial samples, the initial sample was included. To provide a concrete assessment of duration of disease, the number of days between onset of symptoms and positive test was measured where symptom onset pre-dated testing. A viral load of $> 10^3$ was considered clinically relevant (Jacot et al., 2020; Zheng et al., 2020). Next, the numbers of individuals who were still testing positive for E-gene or E-gene-derived sgRNA after 10 days, 11–20 days, 21–30 days, and 31+ days were assessed. Differences in clinical traits between individuals who had a disease duration of $\leq$ 10 days and those who had a disease duration of $>$ 10 days were also investigated.

The clinical traits studied were: age, sex, initial viral load, severity (asymptomatic/mild/moderate vs severe/critical), comorbidities ($\leq$ 3 of any of diabetes, hypertension, dementia, chronic lung disease, chronic kidney disease, or cancer vs $>$ 3 of the above), and immunocompromised status or current treatment with immunosuppressive drugs (steroids, or other immunosuppressive drugs). Associations between duration of RT-qPCR positivity and continuous traits (age, viral load) were assessed by ANOVA, while associations with discrete traits (sex, immunocompromised status, concurrent treatment with immunosuppressive drugs, comorbidities, severity) was assessed by χ^2 analysis.

Results

SARS-CoV-2 patient samples

E-gene SARS-CoV-2 sequences were detected in 176/265 RNA samples taken during the first wave of infection, between March 17, 2020 and November 29, 2020, of which 72 were also E-gene-derived sgRNA positive. 74 patients were classified as asymptomatic. 36 were classified as having mild disease (cough, fever, myalgia, headache, gastrointestinal symptoms, anosmia). 22 were classified as having moderate disease (shortness of breath or dyspnoea but oxygen saturation $\geq$ 94%). 33 were classified as having severe disease (respiratory frequency $>$ 30 breaths per minute, oxygen saturation $\leq$ 94%) and 11 were classified as deemed critical (respiratory failure, assisted ventilation). The strain of SARS-CoV-2 present in each sample was confirmed by cDNA sequencing, to identify the A.2, B.1, B.1.1, B.1.1.10, B.1.1.162, B.1.1.206, B.1.1.261, B.1.1.52, B.1.12, B.1.275, B.1.5, B.1.91, B.1.98, B.1.28, B.1.3, B.1.35, B.1.139, B.1.48, B.1.177.9, B.1.1.315, B.1.117, or B.1.160.6 strains of SARS-CoV-2. No SARS-CoV-2 variant of concern (VOC) sequences were found in our sample set. The mean viral loads in the sample set were 1.9×10^7 copies/mL (range 2.1 copies to 1.2×10^9 copies/mL) for the E-gene and 1.7×10^6 copies/mL (range 12.5 copies to 7.9×10^7 copies/mL) for E-gene-derived sgRNA.

RT-qPCR detection of SARS-CoV-2 E sgRNAs

Using the previously described methods (Corman et al., 2020; Wolfel et al., 2020) it was possible to reliably detect as few as 10 copies/mL of both E-gene and E-gene-derived sgRNA viral genomes. In spite of this low detection limit, a linear relationship between crossing point and SARS-CoV-2 input was only evident at quantifications of over 100 copies (Supplementary Figure S1). Both assays performed well over a linear range of 10^3 to 10^8 copies. The PCR efficiencies for the E-gene and E-gene-derived sgRNA were 100% and 99%, respectively. Reproducibility over the linear range of the assays was good, with r^2 values between replicates of 0.998 for E-gene and 0.996 for E-gene-derived sgRNA.

Both E-gene and E-sgRNA viral loads correlate with the presence of symptoms

E-gene sequences were identified in samples derived from 176 individuals, 72 of which were also positive for E-sgRNA. Viral loads were found to be higher in symptomatic individuals than in asymptomatic individuals for both E-gene and E-sgRNA fragments: 7.98×10^4 compared with 5.46×10^4 copies/mL, respectively, for E-gene sequences ($p = 0.006$) and 4.52×10^3 compared with 1.48×10^3 copies/mL, respectively, for sgRNA sequences ($p = 0.03$) (Figure 1). A linear relationship between E-gene viral load and severity was identified, as defined by NIH criteria (β coefficient = 0.56, SE = 0.27, $p = 0.04$; Figure 2). Although a similar trend was observed between E-sgRNA viral load and severity, this did not reach statistical significance. No associations were identified between age and viral load in the initial sample for either E-gene or E-sgRNA levels, following adjustment for the presence or absence of symptoms ($p = 0.382$ and 0.314, respectively). Similarly, no associations were identified between sex and viral load for either E-gene or E-sgRNA. Individuals with $>$ 2 comorbidities had significantly higher E-gene viral loads than those with $<$ 2 comorbidities ($p = 0.02$), but this relationship was not observed for E-sgRNA.

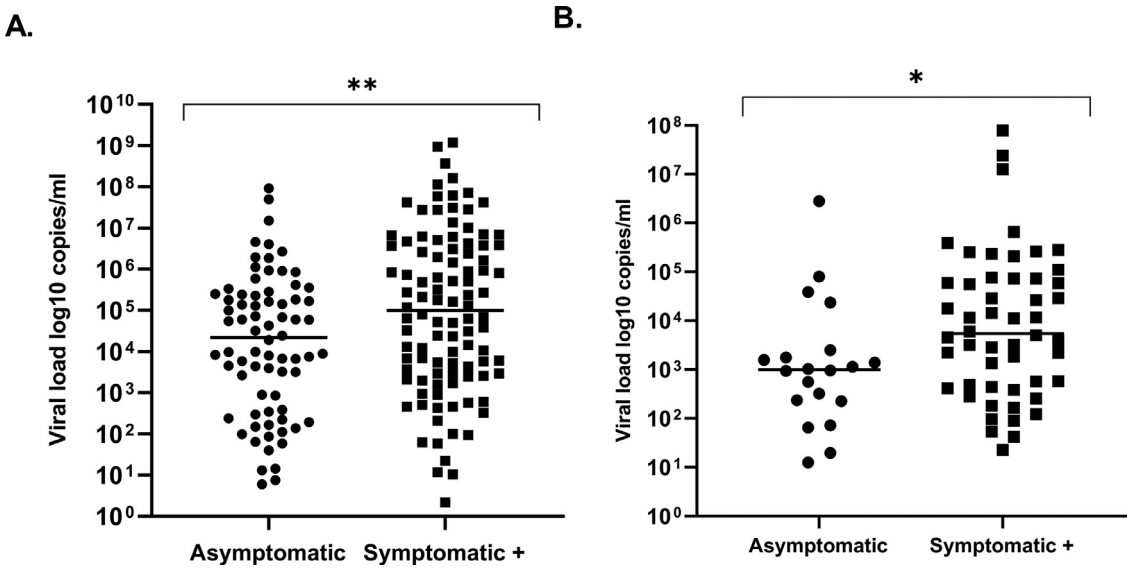

Figure 1. The association between E-gene and E-gene-derived sgRNA viral loads and the presence of COVID-19 symptoms following SARS-CoV-2 infection. **A.** Viral load assessed using assay targeted to the E-gene. **B.** Viral load assessed by assay targeted to E-gene-derived sgRNA. The presence or absence of symptoms is indicated on the *x*-axis, and the viral load (expressed in copies/mL) is given on the *y*-axis. $^{*}p < 0.05$; $^{**}p < 0.01$. Statistical significance was determined by one-way ANOVA.

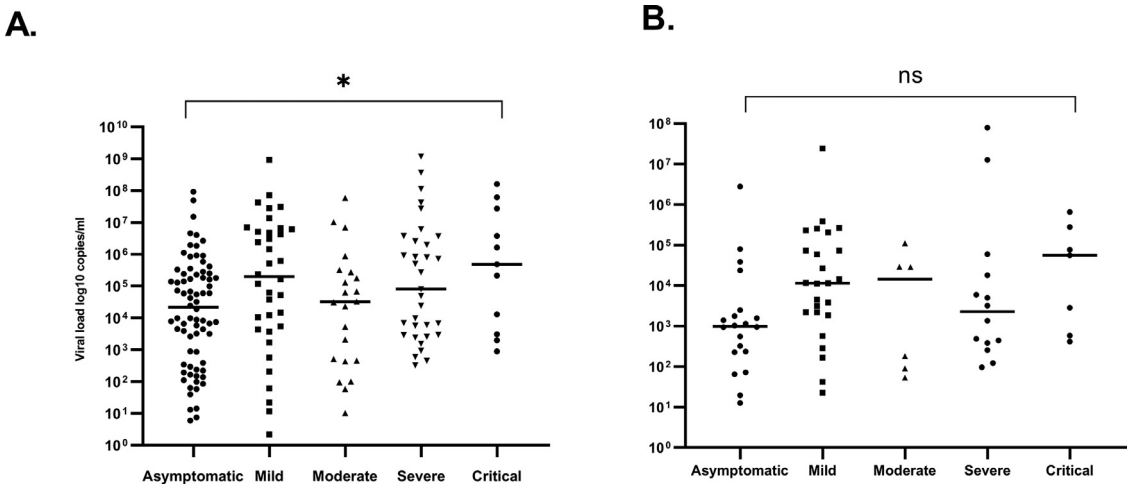

Figure 2. The association between E-gene and E-gene-derived sgRNA viral loads and disease severity following SARS-CoV-2 infection. **A.** Viral load assessed using assay targeted to the E-gene. **B.** Viral load assessed by assay targeted to E-gene-derived sgRNA. Patients were classified according to NIH criteria: i) asymptomatic, ii) mild disease (cough, fever, myalgia, headache, gastrointestinal symptoms, anosmia, shortness of breath, or abnormal chest imaging), iii) moderate disease (shortness of breath or dyspnoea, but with oxygen saturation $\geq$ 94%), iv) severe illness (respiratory frequency $>$ 30 breaths per minute, oxygen saturation $\leq$ 94%), or v) critical (respiratory failure, assisted ventilation); $n = 11$ for E-gene and $n = 7$ for E-sgRNA). The viral load expressed in copies/mL is shown on the *y*-axis. $^{*}p < 0.05$. Statistical significance was determined by one-way ANOVA.

Persistence of clinically relevant levels of SARS-CoV-2 E-sgRNAs in some individuals more than 10 days after symptom onset

For a small subset of our cohort (17 individuals), a set of serial samples taken as part of routine clinical testing was available. In 7/17 (42%) of these samples, no amplification of E-gene-derived sgRNA was identified at any timepoint, although all were positive for the E-gene sequences. SARS-CoV-2 E-sgRNAs were identified only at the time of the initial positive sample in 5/17 (29%) cases, while in a further 5/17 (29%) samples, E-gene-derived sgRNA positivity was evident in subsequent samples for up to 68 days from PCR-confirmed infection (Figure 3). Conversely, 16/17 (94%) of the samples were positive for E-gene sequences in subsequent samples (Figure 3). The higher the initial E-gene-derived sgRNA viral load, the more likely it was to observe E-gene-derived sgRNA amplification in subsequent samples (Figure 3).

To provide a more definitive assessment of variation in duration of E-gene-derived sgRNA infection, the relationships between E-gene and E-gene-derived sgRNA positive status and viral load were assessed, based on the times between onset of symptoms and the last positive PCR test. This information was available for 65/176 individuals testing positive for E-gene and 32/72 individuals testing positive for E-sgRNA. E-gene fragments were detected in 17/65 (26%) of samples between day 11 and day 31+, of which 14/65 (22%) were present at levels of $< 10^3$ copies/mL. E-sgRNA was detected in 9/32 (28%) of samples between day 11 and day 31+, of which 4/32 (13%) were present at clinically relevant levels of $> 10^3$ copies/mL (Figure 4). None of the patients in whom E-sgRNA was detected at a viral load of $> 10^3$ copies/mL had clinical features that would lead clinicians to suspect prolonged infectivity. Patient 1 was a 54-year-old male, classified as having community-acquired severe disease, who had an E-sgRNA vi-

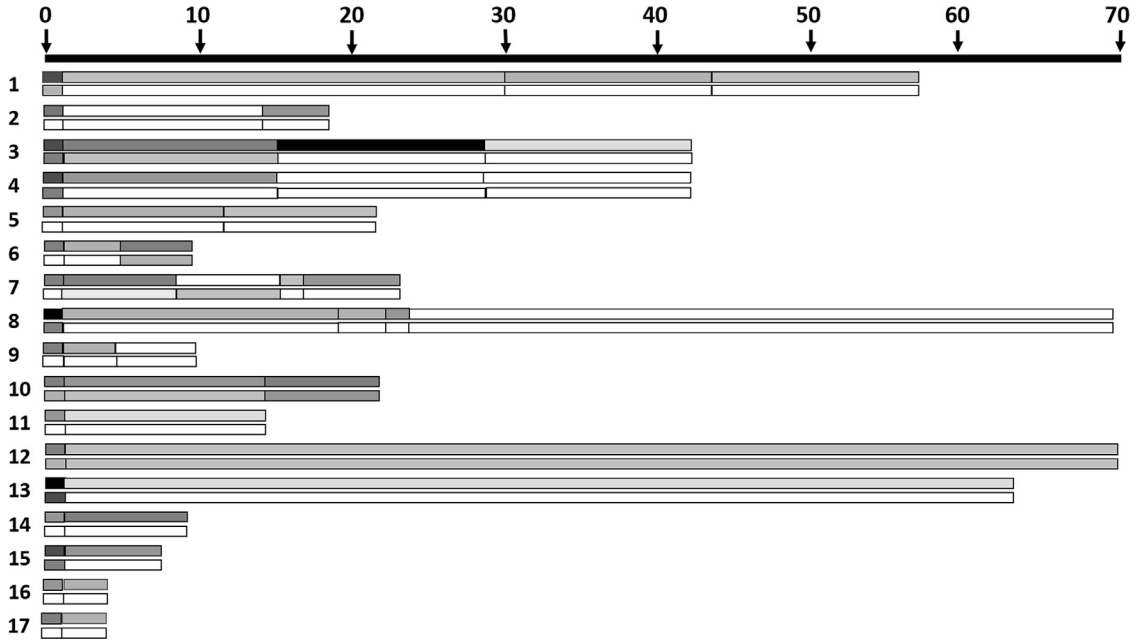

Figure 3. The kinetics of viral load decline in serial samples. The decline in SARS-CoV-2 E-gene and E-gene-derived sgRNA viral load over time in a set of 17 samples is represented. Each number corresponds to an individual patient. The '*y*-axis' bar at the top indicates the number of days. For each patient, the top bar indicates E-gene viral load and the bottom bar indicates E-gene-derived sgRNA viral load. Time of testing is indicated in days on the horizontal axis. The intensity of gray colour corresponds to the viral load, with dark gray indicating the highest viral load and light gray indicating the lowest. White bars represent timepoints at which no SARS-CoV-2 sequences were identified.

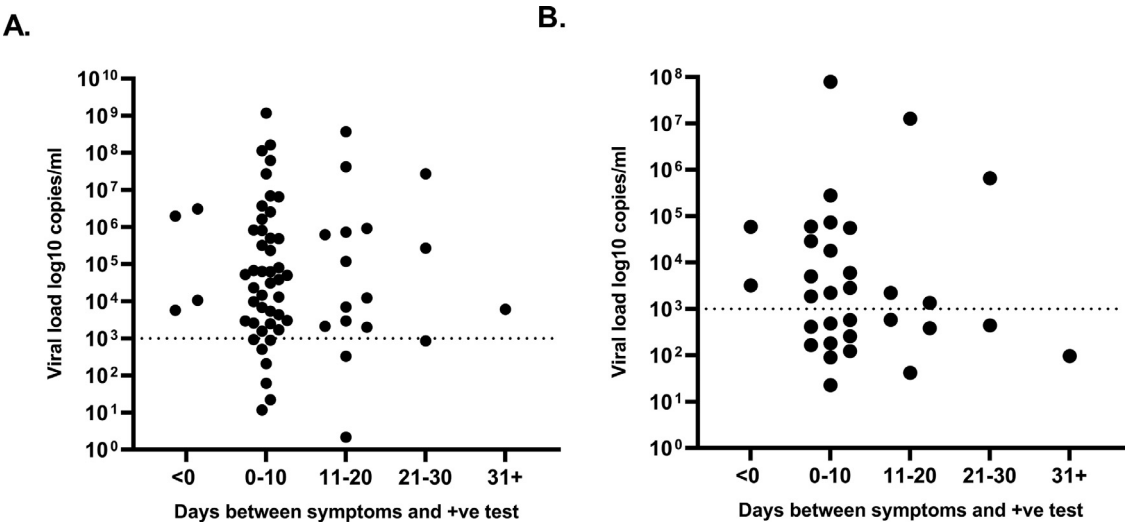

Figure 4. Cross-sectional analysis of viral clearance. In each graph, the E-gene and E-gene-derived sgRNA viral loads at diagnosis, and at up to 10 days, 11–20 days, 21–30 days, and 31+ days between symptom onset and positive test result, are given on the *x*-axis, and the viral load in copies/mL is given on the *y*-axis. The dotted line indicates a viral load of 10^3 copies/mL, which is generally taken as being clinically relevant. **A.** E-gene viral load. **B.** E-gene-derived sgRNA viral load.

ral load of 6.6×10^5 copies/mL in a sample taken 26 days after symptom onset. Patient 2 was a 53-year-old female, with mild disease, who had an E- sgRNA viral load of 2.2×10^3 copies/mL in a sample taken 20 days after symptom onset. Patient 3 was a 57-year-old male, classified as having moderate disease, who had an E-sgRNA viral load of 1.2×10^7 copies/mL in a sample taken 12 days after symptom onset. Patient 4 was a 67-year-old female, classified as having moderate disease, who had an E-sgRNA viral load of 1.4×10^3 copies/mL in a sample taken 20 days after symptom onset. There was no difference in age, viral load, sex, severity (assessed by % hospitalized), % immunocompromised patients, % patients treated with immunosuppressant drugs, or the number of comorbidities between patients with delayed viral clearance for the sample set as a whole, or for the sub-category in which viral load was present at clinically relevant levels of $> 10^3$ copies/mL (Table 2).

Discussion

COVID-19 remains a major public health issue worldwide. Our study aimed to assess the association between RT-qPCR-confirmed E-gene and E-gene-derived sgRNA viral load with disease severity in infected individuals, and to assess the kinetics and extent of E-gene and E-gene-derived sgRNA viral clearance in a population of patients from the South West of the UK in the first wave of infection. Both E-gene and E-sgRNA viral loads were found to be as-

Table 2
Clinical profiles of patients with persistent E-gene or E-sgRNA positivity (> 10 days) compared with those without (< 10 days). The association between patient characteristics and duration of detection of viral fragments is shown for E-gene-positive patients and E-sgRNA-positive patients. 'Immunocompromised' refers to patients with autoimmune disease or other immune compromise. 'Immunosuppressants' refers to patients on steroids or other immunosuppressive drugs. 'Comorbidities' refers to patients with more than one of: diabetes, hypertension, cancer, chronic lung disease, chronic kidney disease, or dementia.

	All samples			Samples > 10^3 copies/ml		
E-gene positive						
	Duration < 10 days (*n* = 46)	**Duration > 10 days (*n* = 20)**	***p*-value**	**Duration < 10 days (*n* = 39)**	**Duration > 10 days (*n* = 16)**	***p*-value**
Age; mean (SD)	58.48 (21.7)	59.89 (13.9)	0.71	59.2 (21.4)	59.4 (13.5)	0.91
Viral load; copies/mL (SD)	3.4×10^7 (1.76×10^8)	2.3×10^7 (8.48×10^7)	0.81	4.1×10^7 (1.91×10^8)	2.8×10^7 (9.22×10^7)	0.94
Sex (% male)	50%	47.4%	0.84	48.7%	50%	0.94
Severity (% hospitalised)	70%	65%	0.99	74%	75%	0.82
No. immunocompromised (%)	7%	0%	0.29	8%	0%	0.30
No. immunosuppressants (%)	15%	10%	0.35	15%	13%	0.52
Comorbidities (> 3; %)	59%	45%	0.42	64%	50%	0.38
E-sgRNA positive						
	Duration < 10 days (*n* = 23)	**Duration > 10 days (*n* = 9)**	***p*-value**	**Duration < 10 days (*n* = 14)**	**Duration > 10 days (*n* = 4)**	***p*-value**
Age; mean (SD)	57.1 (23.3)	58.4(13.6)	0.30	56.6 (20.0)	57.8 (6.4)	0.91
Viral load; copies/mL (SD)	3.5×10^6 (1.65×10^7)	1.5×10^6 (4.20×10^6)	0.44	5.7×10^6 (2.12×10^7)	3.3×10^6 (6.23×10^6)	0.61
Sex (% male)	52%	33%	0.33	50%	42.9%	0.80
Severity (% hospitalised)	70%	78%	0.64	71%	75%	0.89
No. immunocompromised (%)	9%	0%	0.36	0	0	0.58
No. immunosuppressants (%)	17%	33%	0.61	14%	25%	0.16
Comorbidities (> 3; %)	61%	33%	0.16	71%	50%	0.42

sociated with the presence of symptoms, with E-gene viral load associated with severity of disease. Previous studies have linked SARS-CoV-2 viral load with disease severity (Kwon et al., 2020; Zheng et al., 2020). Although both E-gene and E-gene-derived sgRNA viral loads were found to be associated with the presence of symptoms in our study, a linear relationship between viral load and severity was only evident for E-gene sequences. E-sgRNA viral load demonstrated a generalized upwards trend with increasing severity; however, this did not reach statistical significance, and probably represented a type-I error arising from the reduced power yielded by the small number of E-sgRNA-positive patients in our sample set.

Several parameters have been previously reported as risk factors for the persistence of viral sequences, including male sex, old age, concomitant hypertension, severe illness at admission, invasive ventilation, and corticosteroid treatment (Xu et al., 2020). However, our data provide evidence that a relatively substantial proportion of COVID-19 patients demonstrate persistence of viral sequences beyond the 10-day mark (which is currently used for infection control measures). Approximately two-thirds of patients remained positive for E-gene sequences, while one-third remained positive for E-sgRNA sequences. None of the clinical or anthropometric features that were tested (age, sex, immunocompromised status, treatment with immunosuppressive drugs, viral load) were associated with the persistence of viral sequences, suggesting that it may be difficult to predict who might exhibit delayed clearance of potentially replication-competent virus based on these factors alone.

Our study did have limitations. For example, previous studies have suggested that the presence of E-sgRNAs does not directly infer infective potential of the detected virus (van Kampen et al., 2021; Alexandersen et al., 2020). Although E-sgRNAs are produced only when the virus is actively replicating, these RNAs may be protected from degradation within the cytosol by their presence in double-membrane-enclosed vesicles, and thus may be present after the virus has ceased to replicate. Moreover, evidence has shown that E-gene-derived sgRNAs may be less abundant than sgRNAs derived from other SARS-CoV-2 genes, which may mean that levels of sgRNA detected here were underestimated (Dagotto et al., 2021). However, when choosing a suitable proxy for viral replication, it was considered prudent to use a previously validated sgRNA assay that had proven utility in clinical samples (Wolfel et al., 2020). The gold standard for determining if the presence of virus equates to the potential for infection is viral culture. However, this is a specialist technique, which carries a potential infection risk to staff. Moreover, it is not open to all diagnostic or research laboratories, and is not suitable for retrospective or archival sample collections. Conventional RT-qPCR assays for the SARS-CoV-2 viral genome do not distinguish persistent viral fragments following immune inactivation from intact and replication-competent virus, and thus do not give any indication of infectious potential. Similarly new techniques, such as reverse-transcription loop-mediated isothermal amplification (RT-LAMP) testing, may be capable of detecting infectious virus, but this technique requires longer RNA fragments that might not be available in archival samples, and has markedly reduced sensitivity compared with RT-qPCR (Dao Thi et al., 2020). Our study also did not address the duration of viral clearance or the association with severity of any of the novel variants of concern (VOC), which continue to evolve and produce successive waves of strains of SARS-CoV-2 in the population. Our sample set derived from the first wave of infection, when VOCs were not circulating at high levels in the population (Peñarrubia et al., 2020). However, *in silico* testing showed no variation within our primer and probe sequences, and our assays remained accurate and effective against the original strain, as well as the alpha and beta variants, when tested *in vitro*. Unfortunately, material for testing

performance against the delta strain was not available. However, should new variants arise that harbour nucleotide changes in the probe or primer sequences, it would be a simple matter of modifying the assay to identify these newly emerging variants.

Conclusion

Our results suggest that, in some individuals, E-sgRNAs — a proxy for active virus — can be detected for extended durations following infection, and that these individuals may be clinically unremarkable. Although this does not indicate absolutely the presence of active and replication-competent virus, it suggests that, in some cases, the infectious period may extend beyond the 10-day quarantine period currently imposed. Given the obvious potential for onward transmission that these cases may possess, more targeted studies to detect and examine secondary cases with transmission beyond 10 days should now be undertaken in these populations. These results also suggest that, in situations such as hospital inpatient care or patients returning to long-term care facilities following hospital discharge, where onward transmission would be especially problematic, it may be prudent to obtain molecular evidence of remission to protect vulnerable populations.

Conflicts of Interest

LWH and BPL declare an interest in SENISCA Ltd, as founders, directors, and as CSO and CTO, respectively.

Funding source

This work was funded by an Animal Free Research UK (AFRUK) grant (AFR20-COVID) to LWH. AFRUK is the UK's leading non-animal biomedical research charity, which exclusively funds and promotes human-relevant research that replaces the use of animals.

Ethical approval

Ethical approval for sample usage was granted by the Royal Devon and Exeter Tissue Bank (RDETB — ethical approval REC No. 11/SW/0018).

Acknowledgements

This project was supported by the National Institute for Health Research (NIHR) Exeter Clinical Research Facility, which is a partnership between the University of Exeter Medical School College of Medicine and Health, and Royal Devon and Exeter NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2021.12.312.

References

ALEXANDERSEN S, CHAMINGS A, BHATTA TR. SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. Nat Commun 2020;11:6059.

ARONS MM, HATFIELD KM, REDDY SC, KIMBALL A, JAMES A, JACOBS JR, TAYLOR J, SPICER K, BARDOSSY AC, OAKLEY LP, TANWAR S, DYAL JW, HARNEY J, CHISTY Z, BELL JM, METHNER M, PAUL P, CARLSON CM, MCLAUGHLIN HP, THORNBURG N, TONG S, TAMIN A, TAO Y, UEHARA A, HARCOURT J, CLARK S, BROSTROM–SMITH C, PAGE LC, KAY M, LEWIS J, MONTGOMERY P, STONE ND, CLARK TA, HONEIN MA, DUCHIN JS, JERNIGAN JA. Presymptomatic SARS-CoV-2 Infections and transmission in a skilled nursing facility. New England Journal of Medicine 2020;382:2081–90.

AVANZATO VA, MATSON MJ, SEIFERT SN, PRYCE R, WILLIAMSON BN, ANZICK SL, BARBIAN K, JUDSON SD, FISCHER ER, MARTENS C, BOWDEN TA, DE WIT E, RIEDO FX, MUNSTER VJ. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. Cell 2020;183:1901–12 e9.

AYDILLO T, GONZALEZ-REICHE AS, ASLAM S, VAN DE GUCHTE A, KHAN Z, OBLA A, DUTTA J, VAN BAKEL H, ABERG J, GARCIA-SASTRE A, SHAH G, HOHL T, PAPANICOLAOU G, PERALES MA, SEPKOWITZ K, BABADY NE, KAMBOJ M. Shedding of viable SARS-CoV-2 after immunosuppressive therapy for cancer. N Engl J Med 2020;383:2586–8.

BULLARD J, DUST K, FUNK D, STRONG JE, ALEXANDER D, GARNETT L, BOODMAN C, BELLO A, HEDLEY A, SCHIFFMAN Z, DOAN K, BASTIEN N, LI Y, VAN CAESEELE PG, POLIQUIN G. Predicting infectious severe acute respiratory syndrome coronavirus 2 from diagnostic samples. Clinical Infectious Diseases 2020;71:2663–6.

CEVIK M, TATE M, LLOYD O, MARAOLO AE, SCHAFERS J, HO A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. Lancet Microbe 2021;2:e13–22.

CHOI B, CHOUDHARY MC, REGAN J, SPARKS JA, PADERA RF, QIU X, SOLOMON IH, KUO HH, BOUCAU J, BOWMAN K, ADHIKARI UD, WINKLER ML, MUELLER AA, HSU TY, DESJARDINS M, BADEN LR, CHAN BT, WALKER BD, LICHTERFELD M, BRIGL M, KWON DS, KANJILAL S, RICHARDSON ET, JONSSON AH, ALTER G, BARCZAK AK, HANAGE WP, YU XG, GAIHA GD, SEAMAN MS, CERNADAS M, LI JZ. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N Engl J Med 2020;383:2291–3.

CLARO IM, DA SILVA SALES FC, RAMUNDO MS, CANDIDO DS, SILVA CAM, DE JESUS JG, MANULI ER, DE OLIVEIRA CM, SCARPELLI L, CAMPANA G, PYBUS OG, SABINO EC, FARIA NR, LEVI JE. Local transmission of SARS-CoV-2 lineage B.1.1.7, Brazil, December 2020. Emerg Infect Dis 2021;27:970–2.

CORMAN VM, LANDT O, KAISER M, MOLENKAMP R, MEIJER A, CHU DK, BLEICKER T, BRUNINK S, SCHNEIDER J. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 2020;25.

DAGOTTO G, MERCADO NA-O, MARTINEZ DR, HOU YJ, NKOLOLA JP. Comparison of subgenomic and total RNA in SARS-CoV-2 challenged rhesus macaques. LID - JVI 2021:e02370–420 02370-20 [pii] LID - 10.1128/JVI.02370-20 [doi] LID -.

DAO THI VL, HERBST K, BOERNER K, MEURER M, KREMER LP, KIRRMAIER D, FREISTAEDTER A, PAPAGIANNIDIS D, GALMOZZI C, STANIFER ML, BOULANT S, KLEIN S, CHLANDA P, KHALID D, BARRETO MIRANDA I, SCHNITZLER P, KRAUSSLICH HG, KNOP M, ANDERS S. A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. Sci Transl Med 2020;12.

HE, X., LAU, E. A.-O., WU, P., DENG, X., WANG, J., HAO, X., LAU, Y. C., WONG, J. Y., GUAN, Y., TAN, X., MO, X., CHEN, Y., LIAO, B., CHEN, W., HU, F., ZHANG, Q., ZHONG, M., WU, Y., ZHAO, L., ZHANG, F., COWLING, B. A.-O., LI, F. & LEUNG, G. A.-O. Temporal dynamics in viral shedding and transmissibility of COVID-19. 2020.

JACOT, D., GREUB, G., JATON, K. & OPOTA, O. Viral load of SARS-CoV-2 across patients and compared to other respiratory viruses. 2020.

JANG, S., RHEE, J. Y., WI, Y. M. & JUNG, B. K. Viral kinetics of SARS-CoV-2 over the preclinical, clinical, and postclinical period. 2021.

KWON J-S, KIM JY, KIM M-C, PARK SY, KIM B-N, BAE S, CHA HH, JUNG J, KIM M-J, LEE MJ, CHOI S-H, CHUNG J-W, SHIN E-C, KIM S-H. Factors of severity in patients with COVID-19: cytokine/chemokine concentrations, viral load, and antibody responses. The American Journal of Tropical Medicine and Hygiene 2020;103:2412–18.

LADHANI SN, CHOW JY, JANARTHANAN R, FOK J, CRAWLEY-BOEVEY E, VUSIRIKALA A, FERNANDEZ E, PEREZ MS, TANG S, DUN-CAMPBELL K, WYNNE-EVANS E, BELL A, PATEL B, AMIN-CHOWDHURY Z, AIANO F, PARANTHAMAN K, MA T, SAAVEDRA-CAMPOS M, MYERS R, ELLIS J, LACKENBY A, GOPAL R, PATEL M, CHAND M, BROWN K, HOPKINS S, CONSORTIUM C, SHETTY N, ZAMBON M, RAMSAY ME. Increased risk of SARS-CoV-2 infection in staff working across different care homes: enhanced COVID-19 outbreak investigations in London care Homes. J Infect 2020;81:621–4.

LAUER, S. A., GRANTZ, K. H., BI, Q., JONES, F. K., ZHENG, Q., MEREDITH, H. R., AZMAN, A. S., REICH, N. G. & LESSLER, J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. 2020.

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. LID - 10.2807/1560-7917.ES.2020.26.1.2002106 [doi] LID - 2002106. 2021.

LU J, PENG J, XIONG Q, LIU Z, LIN H, TAN X, KANG M, YUAN R, ZENG L, ZHOU P, LIANG C, YI L, DU PLESSIS L, SONG T, MA W, SUN J, PYBUS OG, KE C. Clinical, immunological and virological characterization of COVID-19 patients that test re-positive for SARS-CoV-2 by RT-PCR. EBioMedicine 2020;59.

MALLAPATY S. India's massive COVID surge puzzles scientists. Nature 2021;592:667–8.

PEÑARRUBIA, L., RUIZ, M., PORCO, R., RAO, S. N., JUANOLA-FALGARONA, M., MANISSERO, D., LÓPEZ-FONTANALS, M. & PAREJA, J. Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the COVID-19 outbreak. 2020.

SABINO, E. C., BUSS, L. F., CARVALHO, M. P. S., PRETE, C. A., JR., CRISPIM, M. A. E., FRAIJI, N. A., PEREIRA, R. H. M., PARAG, K. V., DA SILVA PEIXOTO, P., KRAEMER, M. U. G., OIKAWA, M. K., SALOMON, T., CUCUNUBA, Z. M., CASTRO, M. C., DE SOUZA SANTOS, A. A., NASCIMENTO, V. H., PEREIRA, H. S., FERGUSON, N. M., PYBUS, O. G., KUCHARSKI, A., BUSCH, M. P., DYE, C. & FARIA, N. R. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. 2021.

SHRESTHA, N. K., MARCO CANOSA, F., NOWACKI, A. S., PROCOP, G. W., VOGEL, S., FRASER, T. G., ERZURUM, S. C., TERPELUK, P. & GORDON, S. M. Distribution of transmission potential during nonsevere COVID-19 illness. 2020.

VAN KAMPEN JJA, VAN DE VIJVER D, FRAAIJ PLA, HAAGMANS BL, LAMERS MM, OKBA N, VAN DEN AKKER JPC, ENDEMAN H, GOMMERS D, CORNELISSEN JJ, HOEK RAS, VAN DER EERDEN MM, HESSELINK DA, METSELAAR HJ, VERBON A, DE STEENWINKEL JEM, ARON GI, VAN GORP ECM, VAN BOHEEMEN S, VOERMANS JC, BOUCHER CAB, MOLENKAMP R, KOOPMANS MPG, GEURTSVANKESSEL C, VAN DER EIJK AA. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). Nat Commun 2021;12:267.

VOLZ E, MISHRA S, CHAND M, BARRETT JC, JOHNSON R, GEIDELBERG L, HINSLEY WR, LAYDON DJ, DABRERA G, O'TOOLE A, AMATO R, RAGONNET-CRONIN M, HARRISON I, JACKSON B, ARIANI CV, BOYD O, LOMAN NJ, MCCRONE JT, GONCALVES S, JORGENSEN D, MYERS R, HILL V, JACKSON DK, GAYTHORPE K, GROVES N, SILLITOE J, KWIATKOWSKI DP, CONSORTIUM C-GU, FLAXMAN S, RATMANN O, BHATT S, HOPKINS S, GANDY A, RAMBAUT A, FERGUSON NM. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 2021.

WOLFEL R, CORMAN VM, GUGGEMOS W, SEILMAIER M, ZANGE S, MULLER MA, NIEMEYER D, JONES TC, VOLLMAR P, ROTHE C, HOELSCHER M, BLEICKER T, BRUNINK S, SCHNEIDER J, EHMANN R, ZWIRGLMAIER K, DROSTEN C, WENDTNER C. Virological assessment of hospitalized patients with COVID-2019. Nature 2020;581:465–9.

WU B, WHITE KA. Uncoupling RNA virus replication from transcription via the polymerase: functional and evolutionary insights. EMBO J 2007;26:5120–30.

XU K, CHEN Y, YUAN J, YI P, DING C, WU W, LI Y, NI Q, ZOU R, LI X, XU M, ZHANG Y, ZHAO H, ZHANG X, YU L, SU J, LANG G, LIU J, WU X, GUO Y, TAO J, SHI D, YU L, CAO Q, RUAN B, LIU L, WANG Z, XU Y, LIU Y, SHENG J, LI L. Factors associated with prolonged viral rna shedding in patients with coronavirus disease 2019 (COVID-19). Clin Infect Dis 2020;71:799–806.

YANG, L., DAI, J., ZHAO, J. A.-O. X., WANG, Y., DENG, P. & WANG, J. Estimation of incubation period and serial interval of COVID-19: analysis of 178 cases and 131 transmission chains in Hubei province, China. 2020.

ZHENG, S., FAN, J., YU, F., FENG, B., LOU, B., ZOU, Q., XIE, G., LIN, S., WANG, R., YANG, X., CHEN, W., WANG, Q., ZHANG, D., LIU, Y., GONG, R., MA, Z., LU, S., XIAO, Y., GU, Y., ZHANG, J., YAO, H., XU, K., LU, X., WEI, G., ZHOU, J., FANG, Q., CAI, H., QIU, Y., SHENG, J., CHEN, Y. & LIANG, T. A.-O. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January–March 2020: retrospective cohort study. 2020.