Densely sampled viral trajectories suggest longer duration of acute infection with B.1.1.7 variant relative to non-B.1.1.7 SARS-CoV-2

Stephen M. Kissler*[1], Joseph R. Fauver*[2], Christina Mack*[3,4], Caroline G. Tai[3], Mallery I. Breban[2], Anne E. Watkins[2], Radhika M. Samant[3], Deverick J. Anderson[5], David D. Ho[6], Nathan D. Grubaugh†[2], Yonatan H. Grad†[1]

[1] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA
[2] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT
[3] IQVIA, Real World Solutions, Durham, NC
[4] Department of Epidemiology, University of North Carolina-Chapel Hill, Chapel Hill, NC
[5] Duke Center for Antimicrobial Stewardship and Infection Prevention, Durham, NC
[6] Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY

* denotes equal contribution
† denotes co-senior authorship

Correspondence and requests for materials should be addressed to:
Email: ygrad@hsph.harvard.edu
Telephone: 617.432.2275

Abstract.

To test whether acute infection with B.1.1.7 is associated with higher or more sustained nasopharyngeal viral concentrations, we assessed longitudinal PCR tests performed in a cohort of 65 individuals infected with SARS-CoV-2 undergoing daily surveillance testing, including seven infected with B.1.1.7. For individuals infected with B.1.1.7, the mean duration of the proliferation phase was 5.3 days (90% credible interval [2.7, 7.8]), the mean duration of the clearance phase was 8.0 days [6.1, 9.9], and the mean overall duration of infection (proliferation plus clearance) was 13.3 days [10.1, 16.5]. These compare to a mean proliferation phase of 2.0 days [0.7, 3.3], a mean clearance phase of 6.2 days [5.1, 7.1], and a mean duration of infection of 8.2 days [6.5, 9.7] for non-B.1.1.7 virus. The peak viral concentration for B.1.1.7 was 19.0 Ct [15.8, 22.0] compared to 20.2 Ct [19.0, 21.4] for non-B.1.1.7. This converts to 8.5 $\log_{10}$ RNA copies/ml [7.6, 9.4] for B.1.1.7 and 8.2 $\log_{10}$ RNA copies/ml [7.8, 8.5] for non-B.1.1.7. These data offer evidence that SARS-CoV-2 variant B.1.1.7 may cause longer infections with similar peak viral concentration compared to non-B.1.1.7 SARS-CoV-2. This extended duration may contribute to B.1.1.7 SARS-CoV-2's increased transmissibility.

Main text.

The reasons for the enhanced transmissibility of SARS-CoV-2 variant B.1.1.7 are unclear. B.1.1.7 features multiple mutations in the spike protein receptor binding domain[1] that may enhance ACE-2 binding[2], thus increasing the efficiency of virus transmission. A higher or more persistent viral burden in the nasopharynx could also increase transmissibility. To test whether acute infection with B.1.1.7 is associated with higher or more sustained nasopharyngeal viral concentrations, we assessed longitudinal PCR tests performed in a cohort of 65 individuals infected with SARS-CoV-2 undergoing daily surveillance testing, including seven infected with B.1.1.7, as confirmed by whole genome sequencing.

We estimated (1) the time from first detectable virus to peak viral concentration (proliferation time), (2) the time from peak viral concentration to initial return to the limit of detection (clearance time), and (3) the peak viral concentration for each individual (**Supplementary Appendix**).[3] We estimated the means of these quantities separately for individuals infected with B.1.1.7 and non-B.1.1.7 SARS-CoV-2 (**Figure 1**). For individuals infected with B.1.1.7, the mean duration of the proliferation phase was 5.3 days (90% credible interval [2.7, 7.8]), the mean duration of the clearance phase was 8.0 days [6.1, 9.9], and the mean overall duration of infection (proliferation plus clearance) was 13.3 days [10.1, 16.5]. These compare to a mean proliferation phase of 2.0 days [0.7, 3.3], a mean clearance phase of 6.2 days [5.1, 7.1], and a mean duration of infection of 8.2 days [6.5, 9.7] for non-B.1.1.7 virus. The peak viral concentration for B.1.1.7 was 19.0 Ct [15.8, 22.0] compared to 20.2 Ct [19.0, 21.4] for non-B.1.1.7. This converts to 8.5 $\log_{10}$ RNA copies/ml [7.6, 9.4] for B.1.1.7 and 8.2 $\log_{10}$ RNA copies/ml [7.8, 8.5] for non-B.1.1.7. Data and code are available online.[4]

These data offer evidence that SARS-CoV-2 variant B.1.1.7 may cause longer infections with similar peak viral concentration compared to non-B.1.1.7 SARS-CoV-2, and this extended duration may contribute to B.1.1.7 SARS-CoV-2's increased transmissibility. The findings are preliminary, as they are based on seven B.1.1.7 cases. However, if borne out by additional data, a longer isolation period than the currently recommended 10 days after symptom onset[5] may be needed to effectively interrupt secondary infections by this variant. Collection of longitudinal PCR and test positivity data in larger and more diverse cohorts is needed to clarify the viral trajectory of variant B.1.1.7. Similar analyses should be performed for other SARS-CoV-2 variants such as B.1.351 and P.1.

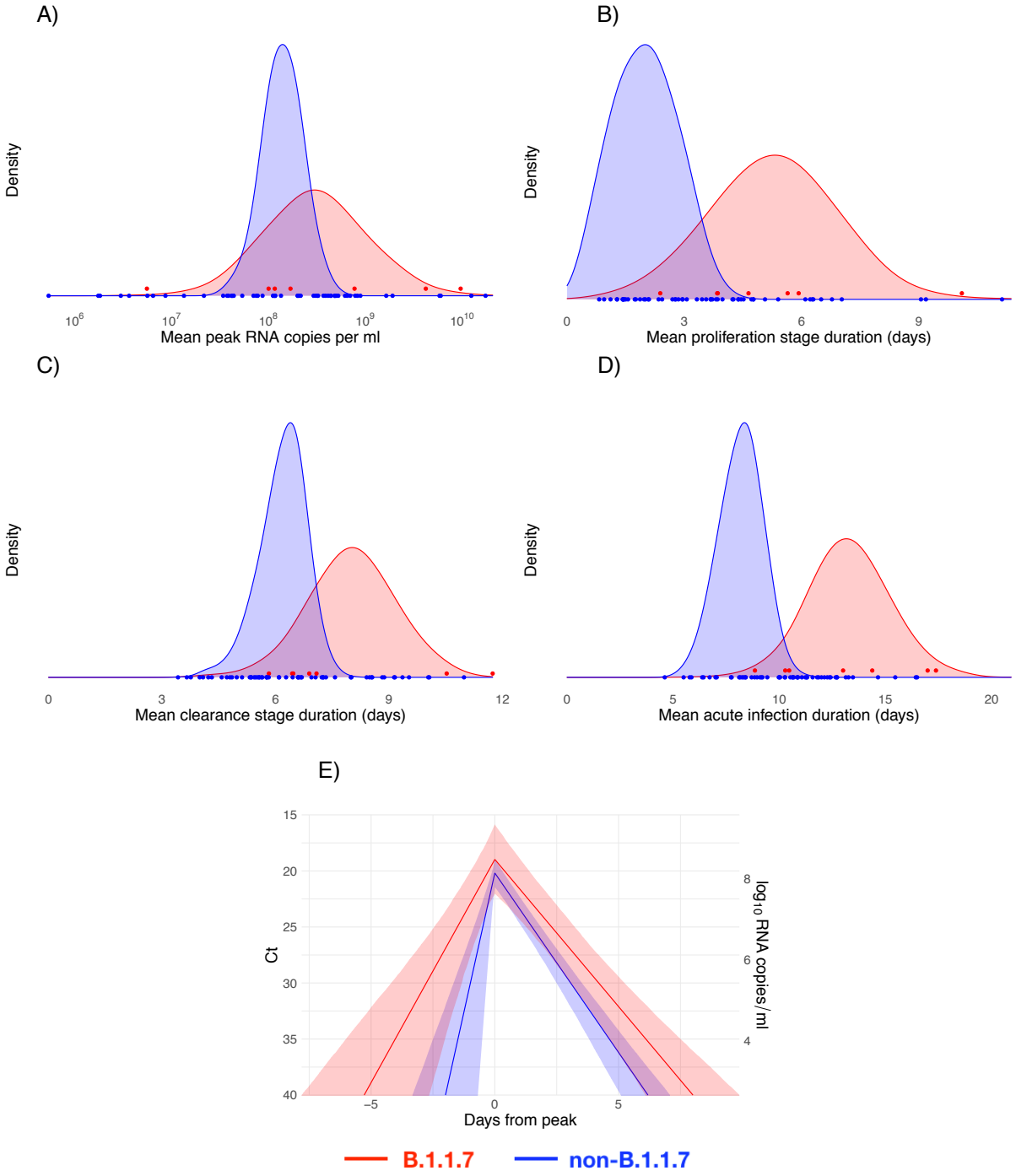


Figure 1. Estimated viral trajectories for B.1.1.7 and non-B.1.1.7 SARS-CoV-2. Posterior distributions for the mean peak viral concentration (A), mean proliferation duration (B), mean clearance duration (C), mean total duration of acute infection (D), and mean posterior viral concentration trajectory (E) for the B.1.1.7 variant (red) and non-B.1.1.7 SARS-CoV-2 (blue). In (A)–(D), distributions depict kernel density estimates obtained from 2,000 draws from the posterior distributions for each statistic. Points depict the individual-level posterior means for each statistic. In (E), solid lines depict the estimated mean viral trajectory. Shaded bands depict the 90% credible intervals for the mean viral trajectory.

References

1. Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(3):95-99. doi:10.15585/mmwr.mm7003e2
2. Yi C, Sun X, Ye J, Ding L, Liu M, Yang Z, et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol*. 2020;17(6):621-630. doi:10.1038/s41423-020-0458-z
3. Kissler SM, Fauver JR, Mack C, Olesen SW, Tai C, Shiue KY, et al. SARS-CoV-2 viral dynamics in acute infections. *medRxiv*. Published online 2020:1-13. doi:10.1101/2020.10.21.20217042
4. Kissler S. Github Repository: CtTrajectories_B117. Published 2020. Accessed February 8, 2020. https://github.com/skissler/CtTrajectories_B117
5. Centers for Disease Control and Prevention. Duration of Isolation and Precautions for Adults with COVID-19. COVID-19. Published 2020. Accessed February 8, 2020. https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html
6. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, et al. Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell*. 2020;181(5):990-996.e5. doi:10.1016/j.cell.2020.04.021
7. Loman N, Rowe W, Rambaut A. nCoV-2019 novel coronavirus bioinformatics protocol.
8. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;5(11):1403-1407. doi:10.1038/s41564-020-0770-5
9. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations.*; 2020.
10. Kudo E, Israelow B, Vogels CBF, Lu P, Wyllie AL, Tokuyama M, et al. Detection of SARS-CoV-2 RNA by multiplex RT-qPCR. Sugden B, ed. *PLOS Biol*. 2020;18(10):e3000867. doi:10.1371/journal.pbio.3000867
11. Vogels C, Fauver J, Ott IM, Grubaugh N. *Generation of SARS-COV-2 RNA Transcript Standards for QRT-PCR Detection Assays.*; 2020. doi:10.17504/protocols.io.bdv6i69e
12. Cleary B, Hay JA, Blumenstiel B, Gabriel S, Regev A, Mina MJ. Efficient prevalence estimation and infected sample identification with group testing for SARS-CoV-2. *medRxiv*. Published online 2020.
13. Tom MR, Mina MJ. To Interpret the SARS-CoV-2 Test, Consider the Cycle Threshold Value. *Clin Infect Dis*. 2020;02115(Xx):1-3. doi:10.1093/cid/ciaa619
14. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan : A Probabilistic Programming Language. *J Stat Softw*. 2017;76(1). doi:10.18637/jss.v076.i01
15. R Development Core Team R. R: A Language and Environment for Statistical Computing. Team RDC, ed. *R Found Stat Comput*. 2011;1(2.11.1):409. doi:10.1007/978-3-540-74686-7

Supplementary Appendix.

Ethics.

Residual de-identified viral transport media from anterior nares and oropharyngeal swabs collected from players, staff, vendors, and associated household members from a professional sports league were obtained from BioReference Laboratories. In accordance with the guidelines of the Yale Human Investigations Committee, this work with de-identified samples was approved for research not involving human subjects by the Yale Internal Review Board (HIC protocol # 2000028599). This project was designated exempt by the Harvard IRB (IRB20-1407).

Study population. The data reported here represent a convenience sample including team staff, players, arena staff, and other vendors (e.g., transportation, facilities maintenance, and food preparation) affiliated with a professional sports league. Clinical samples were obtained by combined swabs of the anterior nares and oropharynx administered by a trained provider. Viral concentration was measured using the cycle threshold (Ct) according to the Roche cobas target 1 assay. For an initial pool of 298 participants who first tested positive for SARS-CoV-2 infection during the study period (between November 28th, 2020 and January 20th, 2021), a diagnosis of "novel" or "persistent" infection was recorded. "Novel" denoted a likely new infection while "persistent" indicated the presence of virus in a clinically recovered individual. A total of 65 individuals (90% male) had novel infections that met our inclusion criteria: at least five positive PCR tests (Ct < 40) and at least one test with Ct < 35. Seven of these individuals were infected with the B.1.1.7 variant as confirmed by genomic sequencing.

Genome sequencing and lineage assignments: RNA was extracted from remnant nasopharyngeal diagnostic specimens and used as input for SARS-CoV-2 genomic sequencing as previously described.[6] Samples were sequenced on the Oxford Nanopore MinION. Consensus sequences were generated using the ARTIC Network analysis pipeline[7] and samples with >80% genome coverage were included in analysis. Individual SARS-CoV-2 genomes were assigned to PANGO lineages using Pangolin v.2.1.8.[8] All viral genomes assigned to the B.1.1.7 lineage were manually examined for representative mutations.[9]

Converting Ct values to viral genome equivalents. To convert Ct values to viral genome equivalents, we first converted the Roche cobas target 1 Ct values to equivalent Ct values on a

multiplexed version of the RT-qPCR assay from the US Centers for Disease Control and Prevention.[10] We did this following our previously described methods.[3] Briefly, we adjusted the Ct values using the best-fit linear regression between previously collected Roche cobas target 1 Ct values and CDC multiplex Ct values using the following regression equation:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Here, y_i denotes the i^{th} Ct value from the CDC multiplex assay, x_i denotes the i^{th} Ct value from the Roche cobas target 1 test, and ε_i is an error term with mean 0 and constant variance across all samples. The coefficient values are $\beta_0 = -6.25$ and $\beta_1 = 1.34$.

Ct values were fitted to a standard curve in order to convert Ct value data to RNA copies. Synthetic T7 RNA transcripts corresponding to a 1,363 b.p. segment of the SARS-CoV-2 nucleocapsid gene were serially diluted from 10^6-10^0 RNA copies/µl in duplicate to generate a standard curve[11] **(Supplementary Table 1)**. The average Ct value for each dilution was used to calculate the slope (-3.60971) and intercept (40.93733) of the linear regression of Ct on log-10 transformed standard RNA concentration, and Ct values from subsequent RT-qPCR runs were converted to RNA copies using the following equation:

$$\log_{10}([\mathrm{RNA}]) = (Ct - 40.93733)/(-3.60971) + \log_{10}(250)$$

Here, [RNA] represents the RNA copies /ml. The $\log_{10}(250)$ term accounts for the extraction (300 µl) and elution (75 µl) volumes associated with processing the clinical samples as well as the 1,000 µl/ml unit conversion.

Model fitting.

For the statistical analysis, we removed any sequences of 3 or more consecutive negative tests to avoid overfitting to these trivial values. Following our previously described methods,[3] we assumed that the viral concentration trajectories consisted of a proliferation phase, with exponential growth in viral RNA concentration, followed by a clearance phase characterized by exponential decay in viral RNA concentration.[12] Since Ct values are roughly proportional to the negative logarithm of viral concentration[13], this corresponds to a linear decrease in Ct followed by a linear increase. We therefore constructed a piecewise-linear regression model to estimate the

peak Ct value, the time from infection onset to peak (*i.e.* the duration of the proliferation stage), and the time from peak to infection resolution (*i.e.* the duration of the clearance stage). The trajectory may be represented by the equation

$$E[Ct(t)] = \begin{cases} \text{l.o.d.} & t \leq t_o \\ \text{l.o.d.} - \frac{\delta}{t_p - t_o}(t - t_o) & t_o < t \leq t_p \\ \text{l.o.d.} - \delta + \frac{\delta}{t_r - t_p}(t - t_p) & t_p < t \leq t_r \\ \text{l.o.d.} & t > t_r \end{cases}$$

Here, E[*Ct(t)*] represents the expected value of the Ct at time *t,* "l.o.d" represents the RT-qPCR limit of detection, δ is the absolute difference in Ct between the limit of detection and the peak (lowest) Ct, and t_o, t_p, and t_r are the onset, peak, and recovery times, respectively.

Before fitting, we re-parametrized the model using the following definitions:

- $\Delta Ct(t)$ = l.o.d. – *Ct(t)* is the difference between the limit of detection and the observed Ct value at time *t.*
- $\omega_p = t_p - t_o$ is the duration of the proliferation stage.
- $\omega_c = t_r - t_p$ is the duration of the clearance stage.

We constrained $0.25 \leq \omega_p \leq 14$ days and $2 \leq \omega_p \leq 30$ days to prevent inferring unrealistically small or large values for these parameters for trajectories that were missing data prior to the peak and after the peak, respectively. We also constrained $0 \leq \delta \leq 40$ as Ct values can only take values between 0 and the limit of detection (40).

We next assumed that the observed $\Delta Ct(t)$ could be described the following mixture model:

$$\Delta Ct(t) \sim \lambda \, \text{Normal}(E[\Delta Ct(t)], \sigma(t)) + (1 - \lambda) \, \text{Exponential}(\log(10))\Big]_0^{\text{l.o.d}}$$

where E[$\Delta Ct(t)$] = l.o.d. - E[*Ct(t)*] and λ is the sensitivity of the q-PCR test, which we fixed at 0.99. The bracket term on the right-hand side of the equation denotes that the distribution was truncated to ensure Ct values between 0 and the limit of detection. This model captures the scenario where

most observed Ct values are normally distributed around the expected trajectory with standard deviation $\sigma(t)$, yet there is a small (1%) probability of an exponentially distributed false negative near the limit of detection. The log(10) rate of the exponential distribution was chosen so that 90% of the mass of the distribution sat below 1 Ct unit and 99% of the distribution sat below 2 Ct units, ensuring that the distribution captures values distributed at or near the limit of detection. We did not estimate values for λ or the exponential rate because they were not of interest in this study; we simply needed to include them to account for some small probability mass that persisted near the limit of detection to allow for the possibility of false negatives.

We used a hierarchical structure to describe the distributions of ω_p, ω_r, and δ for each individual based on their respective population means $\mu_{\omega p}$, $\mu_{\omega r}$, and μ_δ and population standard deviations $\sigma_{\omega p}$, $\sigma_{\omega r}$, and σ_δ such that

$\omega_p \sim \text{Normal}(\mu_{\omega p}, \sigma_{\omega p})$

$\omega_r \sim \text{Normal}(\mu_{\omega r}, \sigma_{\omega r})$

$\delta \sim \text{Normal}(\mu_\delta, \sigma_\delta)$

We inferred separate population means ($\mu_\cdot$) for B.1.1.7- and non-B.1.1.7-infected individuals. We used a Hamiltonian Monte Carlo fitting procedure implemented in Stan (version 2.24)[14] and R (version 3.6.2)[15] to estimate the individual-level parameters ω_p, ω_r, δ, and t_p as well as the population-level parameters σ^*, $\mu_{\omega p}$, $\mu_{\omega r}$, μ_δ, $\sigma_{\omega p}$, $\sigma_{\omega r}$, and σ_δ. We used the following priors:

Hyperparameters:

$\sigma^* \sim \text{Cauchy}(0, 5)\ [0, \infty]$

$\mu_{\omega p} \sim \text{Normal}(14/2, 14/6)\ [0.25, 14]$

$\mu_{\omega r} \sim \text{Normal}(30/2, 30/6)\ [2, 30]$

$\mu_\delta \sim \text{Normal}(40/2, 40/6)\ [0, 40]$

$\sigma_{\omega p} \sim \text{Cauchy}(0, 14/\tan(\pi(0.95\text{-}0.5)))\ [0, \infty]$

$\sigma_{\omega r} \sim \text{Cauchy}(0, 30/\tan(\pi(0.95\text{-}0.5)))\ [0, \infty]$

$\sigma_\delta \sim \text{Cauchy}(0, 40/\tan(\pi(0.95\text{-}0.5)))\ [0, \infty]$

Individual-level parameters:

$\omega_p \sim \text{Normal}(\mu_{\omega p}, \sigma_{\omega p})$ [0.25,14]

$\omega_r \sim \text{Normal}(\mu_{\omega r}, \sigma_{\omega r})$ [2,30]

$\delta \sim \text{Normal}(\mu_{\delta}, \sigma_{\delta})$ [0,40]

$t_p \sim \text{Normal}(0, 2)$

The values in square brackets denote truncation bounds for the distributions. We chose a vague half-Cauchy prior with scale 5 for the observation variance σ^*. The priors for the population mean values ($\mu_\cdot$) are normally distributed priors spanning the range of allowable values for that parameter; this prior is vague but expresses a mild preference for values near the center of the allowable range. The priors for the population standard deviations ($\sigma_\cdot$) are half Cauchy-distributed with scale chosen so that 90% of the distribution sits below the maximum value for that parameter; this prior is vague but expresses a mild preference for standard deviations close to 0.

We ran four MCMC chains for 1,000 iterations each with a target average proposal acceptance probability of 0.8. The first half of each chain was discarded as the warm-up. The Gelman R-hat statistic was less than 1.1 for all parameters. This indicates good overall mixing of the chains. There were no divergent iterations, indicating good exploration of the parameter space. The posterior distributions for μ_δ, $\mu_{\omega p}$, and $\mu_{\omega r}$, were estimated separately for individuals infected with B.1.1.7 and non-B.1.1.7. These are depicted in **Figure 1** (main text). Draws from the individual posterior viral trajectory distributions are depicted in **Supplementary Figure 1.** The mean posterior viral trajectories for each individual are depicted in **Supplementary Figure 2.**

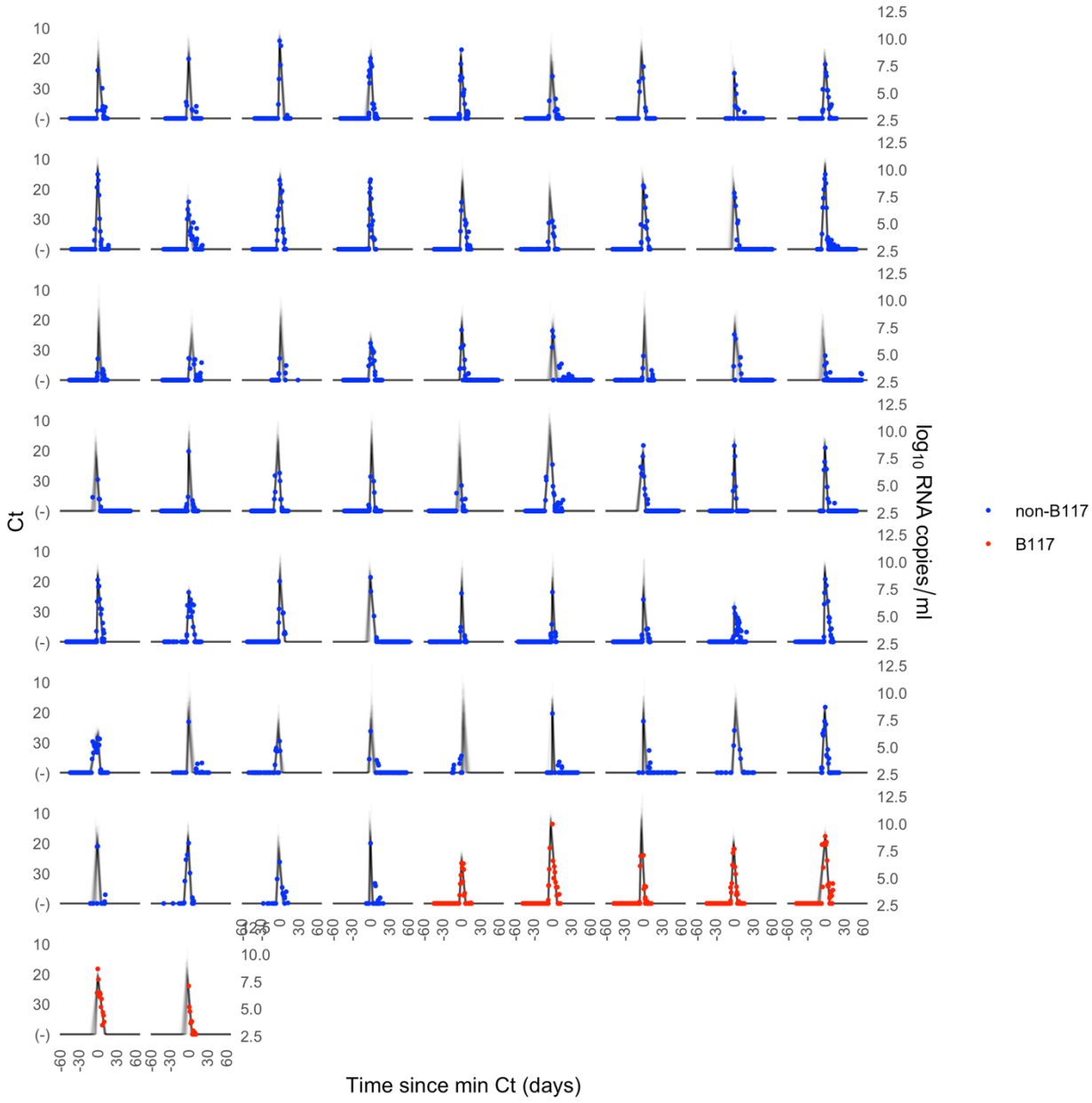
Checking for influential outliers. To examine whether the posterior distributions for the B.1.1.7-infected individuals reflected the influence of a single outlier, we re-fit the model seven times, omitting one of the B.1.1.7 trajectories each time. The inferred parameter values were fairly consistent, though omitting either of two of the B.1.1.7 cases (cases 5 and 6 in **Supplementary Table 2**). yields an infection duration with a 90% credible interval that overlaps with that of the non-B.1.1.7 90% credible interval for infection duration.

Standard (copies/ul)	Replicate 1 (Ct)	Replicate 2 (Ct)	Average Ct
10^6	19.3	19.7	19.5
10^5	23.0	21.2	22.1
10^4	26.9	26.7	26.8
10^3	30.6	30.4	30.5
10^2	34.0	34.0	34.0
10^1	37.2	36.6	36.9
10^0	N/A	39.9	39.9

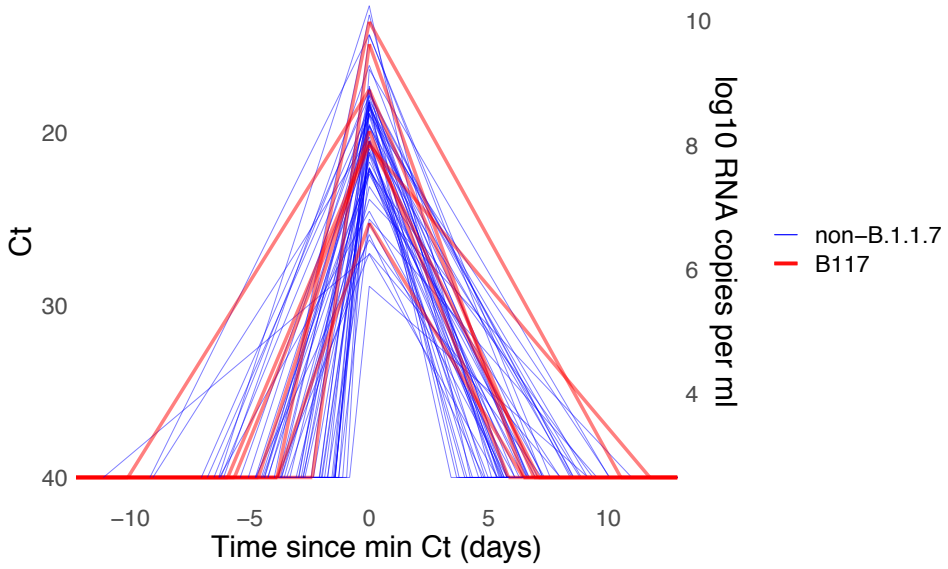
Supplementary Table 1. Standard curve relationship between virus RNA copies and Ct values. Synthetic T7 RNA transcripts corresponding to a 1,363 base pair segment of the SARS-CoV-2 nucleocapsid gene were serially diluted from 10^6-10^0 and evaluated in duplicate with RT-qPCR. The best-fit linear regression of the average Ct on the log10-transformed standard values had slope -3.60971 and intercept 40.93733 (R^2 = 0.99).

Omitted B117 Case	Proliferation duration (days) [90% CI]	Clearance duration (days) [90% CI]	Infection duration (days) [90% CI]	Peak viral concentration (log(copies/ml)) [90% CI]
None	5.3 [2.7, 7.8]	8.0 [6.1, 9.9]	13.3 [10.1, 16.5]	8.5 [7.6, 9.4]
1	5.5 [3.0, 8.1]	8.3 [6.3, 10.3]	13.9 [10.6, 17.0]	8.8 [7.9, 9.8]
2	5.7 [3.1, 8.4]	7.5 [5.1, 9.6]	13.2 [9.8, 16.5]	8.2 [7.4, 9.1]
3	5.9 [3.3, 8.6]	8.3 [6.3, 10.3]	14.2 [11.0, 17.4]	8.2 [7.4, 9.1]
4	5.4 [2.7, 7.9]	8.5 [6.3, 10.5]	13.9 [10.5, 17.0]	8.5 [7.6, 9.4]
5	4.3 [1.8, 6.9]	8.3 [6.2, 10.3]	12.6 [9.4, 15.8]	8.4 [7.5, 9.3]
6	5.4 [3.0, 7.9]	7.1 [5.1, 9.1]	12.6 [9.4, 15.6]	8.6 [7.8, 9.6]
7	5.2 [2.6, 7.7]	8.1 [6.0, 10.2]	13.3 [10.1, 16.6]	8.6 [7.8, 9.4]
Non-B.1.1.7 reference	2.0 [0.7, 3.3]	6.2 [5.1, 7.1]	8.2 [6.5, 9.7]	8.2 [7.8, 8.5]

Supplementary Table 2. Posterior population mean viral trajectory parameter values and 90% credible intervals for B.1.1.7 infections when omitting single trajectories. Each row corresponds to a model fit obtained by omitting one person who was infected with B.1.1.7, so that the parameter values are informed by six of the seven B.1.1.7 infections. The final row lists the fitted parameter values for the non-B.1.1.7 infections for reference.



Supplementary Figure 1. Ct values for 65 individuals with estimated viral trajectories. Each pane depicts the recorded Ct values (points) and derived log-10 genome equivalents per ml (log(ge/ml)) for a single person during the study period. Points along the horizontal axis represent negative tests. Time is indexed in days since the minimum recorded Ct value (maximum viral concentration). Individuals with confirmed B.1.1.7 infections are depicted in red. Non-B.1.1.7 infections are depicted in blue. Lines depict 100 draws from the posterior distribution for each person's viral trajectory.



Supplementary Figure 2. Mean posterior viral trajectories for each person in the study. Lines depict the posterior mean viral trajectory specified by the posterior mean proliferation time, mean clearance time, and mean peak Ct. Trajectories are aligned temporally to have the same peak time. B.1.1.7 trajectories are depicted in red, non-B.1.1.7 in blue.