Viral dynamics of SARS-CoV-2 infection and the predictive value of repeat testing

Stephen M. Kissler*[1], Joseph R. Fauver*[2], Christina Mack*[3,4], Caroline Tai[3], Kristin Y. Shiue[3,4], Chaney C. Kalinich[2], Sarah Jednak[5], Isabel M. Ott[2], Chantal B.F. Vogels[2], Jay Wohlgemuth[6], James Weisberger[7], John DiFiori[8], Deverick J. Anderson[9], Jimmie Mancell[10], David D. Ho[11], 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] Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, MI
[6] Quest Diagnostics, San Juan Capistrano, CA
[7] Bioreference Laboratories, Elmwood Park, NJ
[8] Hospital for Special Surgery, and the National Basketball Association, New York, NY
[9] Duke Center for Antimicrobial Stewardship and Infection Prevention, Durham, NC
[10] Department of Medicine, University of Tennessee Health Science Center, Memphis, TN
[11] Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY

* denotes equal contribution
† denotes co-senior authorship

Abstract

SARS-CoV-2 diagnostics that report viral RNA concentrations can be used to determine a patient's stage of infection, but this potential has not yet been realized due to a lack of prospective longitudinal data to calibrate such inferences. Here, we report the viral RNA trajectories for 68 individuals using quantitative PCR testing. On average, symptomatic and asymptomatic individuals reached similar peak viral RNA concentrations (22.2 Ct, 95% credible interval [19.1, 25.1] vs. 22.4 Ct [20.2, 24.5]) within similar amounts of time (2.9 days [0.7, 4.7] vs. 3.0 days [1.3, 4.3]), but acute shedding lasted longer for symptomatic individuals (10.5 days [6.5, 14.0] vs. 6.7 days [3.2, 9.2]). A second test within 2 days after an initial positive PCR result reliably indicated whether viral RNA concentration was increasing, decreasing, or in a low-level persistent phase. Quantitative viral RNA assessment, informed by viral trajectory, can improve algorithms for clinical and public health management.

Main text.

As mortality from the COVID-19 pandemic surpasses one million, SARS-CoV-2 continues to cause hundreds of thousands of daily new infections *(1)*. A critical strategy to curb the spread of the virus without imposing widespread lockdowns is to rapidly identify and isolate infectious individuals. Since symptoms are an unreliable indicator of infectiousness and infections are frequently asymptomatic *(2)*, diagnostic tests are key to determining whether a person is infected and may be contagious.

Real time quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) tests are the gold standard for detecting SARS-CoV-2 infection. Normally, these tests yield a binary positive/negative diagnosis based on detection of viral RNA. However, they can also inform on the viral titer via the cycle threshold (Ct). The Ct is the number of thermal cycles needed to amplify sampled viral RNA to a detectable level: the higher the sampled viral RNA concentration, the lower the Ct. This inverse correlation between Ct and viral concentration makes RT-qPCR tests far more valuable than a binary diagnostic, as they can be used to reveal a person's progress through key stages of infection *(3)*, assisting with clinical and public health decision-making. However, this potential has not yet been realized due to a lack of data describing complete Ct trajectories for individuals infected with SARS-CoV-2. The dynamics of the Ct during the earliest stages of infection, when contagiousness is rapidly increasing, have been especially unclear since viral testing is usually performed after the onset of symptoms, after viral RNA concentration has peaked and already begun to decline, and using only a single specimen *(4,5)*.

Here, we present the findings from an analysis of prospective longitudinal SARS-CoV-2 testing performed for players, staff and vendors participating in the occupational health program implemented as part of the resumption of the 2019-20 National Basketball Association (NBA) season. We report the results of 2,411 RT-qPCR tests from 68 individuals who provided at least one sample with a Ct value that was within the limit of detection.

Using the Ct values and associated clinical metadata, we identified 46 individuals with probable new SARS-CoV-2 infections. The remaining individuals were assumed to be persistently shedding SARS-CoV-2 RNA due to an infection that occurred prior to the study period. This persistent RNA shedding can last for weeks after an acute infection and seems more likely to represent non-infectious RNA than infectious virus *(6)*. For the 46 acute infections, we estimated the peak Ct value, the time from onset of infection to peak, and the time from peak to conclusion of acute viral shedding using a Bayesian statistical model. Based on these inferences, we used the full pool of 68 individuals to estimate the probability that a given Ct value was associated with the acute *vs*. persistent stages of viral shedding. Within the acute stage, we also estimated whether a given Ct value was associated with the initial period of viral proliferation (increasing viral titers) or the subsequent period of viral clearance (decreasing viral titers) (**Figure 1**). We found that a second test within two days of an initial positive can help determine whether a person is acutely shedding viral RNA and can substantially clarify whether the person is in the proliferation or the clearance stage. These findings underscore the potential value of integrating viral Ct trajectory data into viral diagnostics and surveillance algorithms to inform clinical and public health decision-making.

The study population consisted of NBA players, staff, and vendors associated with the NBA's season restart. The study period began in teams' local cities from June 23rd through July 9th, 2020, and testing continued for all teams as they transitioned to Orlando, Florida through September 7th, 2020. We report data from 68 individuals (90% male) who were tested at least five times during the study period and recorded at least one positive Ct value (<40). Due to a lack of new infections among players and team staff after clearing quarantine in Orlando, all players and team staff included in the results pre-date the Orlando phase of the restart. Of the individuals included in the study, 27 of the 46 with active infections and 40 of the 68 overall were from staff and vendors. The median number of tests administered to each of the 68 individuals was 41 (IQR [14, 51]; Range [5, 70]). The median number of positive Ct values recorded for each person was 3 (IQR [2, 4]; Range [1, 9]). A diagnosis of "acute" or "persistent" infection was abstracted from physician records. "Acute" denoted a likely new infection. "Persistent" indicated the presence of virus in a clinically recovered individual, likely due to infection that developed prior to the onset of the study. The Ct trajectories for all 68 individuals included in the analysis with their designations of acute or persistent infection are depicted in **Supplemental Figures 1–4**. If an individual reported symptoms on the day of the initial positive test, these were recorded.

Clinical samples were obtained by combined swabs of the anterior nares and oropharynx administered by a trained provider. The samples were initially tested by either Quest Diagnostics (while teams were in local markets using the Quest SARS-CoV-2 RT-qPCR *(7)*) or BioReference Laboratories (while teams were in Orlando using the cobas SARS-CoV-2 test *(8)*). Viral transport media from positive samples were sent to Yale University for subsequent RT-qPCR testing using a multiplexed version of the assay from the US Centers for Disease Control and Prevention *(9)* to normalize Ct values across testing platforms. A total of 234 samples from BioReference and 128 from Quest were tested at Yale; 49 positive samples had Ct values assigned on first testing but did not undergo repeat testing at the Yale laboratory. To account for the different calibration of the testing instruments, we used a linear conversion (**Supplemental Figures 5-7**, **Supplemental Methods**) to adjust these samples to the Yale laboratory scale. Subsequent analysis is based on the N1 Ct value from the Yale multiplex assay and on the adjusted Roche cobas target 1 assay.

We used a Bayesian statistical model to infer the peak Ct value and the durations of the proliferation and clearance stages for the 46 acute infections with at least on Ct value below 35 (**Figure 1**). We assumed that the proliferation and clearance stages could respectively be characterized by periods of exponential growth and exponential decay in viral RNA concentration *(10)*. The Ct is roughly proportional to the negative logarithm of the viral RNA concentration *(3)*, and thus Ct values are expected to decrease linearly to a peak and then increase linearly until the resolution of the acute infection (**Supplemental Figure 8**). Resolution may be followed by intermittent positive Ct values near the limit of detection *(6)*, corresponding to persistent shedding of virus or viral fragments. We used piecewise-linear regression to fit the Ct trajectories, accounting for rare false negatives during the acute infection period. Full details are provided in the **Supplemental Methods** and code may be accessed online [https://github.com/gradlab/CtTrajectories].

Of the 46 individuals with acute infections, 13 reported symptoms at the time of diagnosis. The mean peak Ct for symptomatic individuals was 22.2 (95% credible interval [19.1, 25.1]), the mean duration of the proliferation phase was 2.9 days [0.7, 4.7], and the mean duration of clearance

was 10.5 days [6.5, 14.0] (**Figure 2**). This compares with 22.4 Ct [20.2, 24.5], 3.0 days [1.3, 4.3], and 6.7 days [3.2, 9.2], respectively, for individuals who did not report symptoms at the time of diagnosis (**Figure 2**). In summary, there was little difference in mean Ct and proliferation duration between individuals with known symptoms *vs.* no known symptoms, but individuals with symptoms had a relatively longer clearance phase than individuals with no known symptoms. This yielded a slightly longer overall duration of acute infection for individuals who reported symptoms (13.4 days [9.3, 17.1]) *vs.* those who did not (9.7 days [6.0, 12.5]) (**Figure 2D-E**). For all individuals regardless of symptoms, the mean peak Ct value, proliferation duration, clearance duration, and duration of acute shedding were 22.4 Ct [20.6, 24.1], 2.7 days [1.2, 3.8], 7.4 days [3.9, 9.6], and 10.1 days [6.5, 12.6] (**Supplemental Figure 9**). There was a substantial amount of individual-level variation in the peak Ct and the proliferation and clearance stage durations (**Supplemental Figures 10–13**). Combined posterior distributions for these quantities at the individual level are depicted in **Supplemental Figure 14** with best-fit normal and gamma distributions as appropriate.

Using the full dataset of 68 individuals, we estimated the probability that a positive Ct was associated with an acute infection (*i.e.,* the proliferation or clearance phase, but not the persistent phase), and if so, the probability that it was associated with just the proliferation stage. First, we assigned to each positive sample the probability that it was collected during each of the three stages of infection. To do so, we began with the positive samples from the 46 individuals with acute infections and calculated the frequency with which each sample sat within the proliferation stage, the clearance stage, or the persistent stage (*i.e.,* neither of the previous two stages) across 10,000 posterior parameter draws for that person. For the remaining 22 individuals, all positive samples were assigned to the persistent stage. Next, we calculated the probability that a Ct value falling within a 5-unit window corresponded to an active infection (*i.e.,* either the proliferation or the clearance stage) by summing the proliferation and clearance probabilities for all positive samples with that window and dividing by the total number of positive samples in the window. We considered windows with midpoints spanning from Ct = 37.5 to Ct = 15.5 (**Figure 3A**). We performed a similar calculation to determine the probability that a Ct falling within a given 5-unit window corresponded to just the proliferation phase, assuming it had already been determined that the sample fell within an active infection (**Figure 3B**). The probability of active infection increased rapidly with decreasing Ct (increasing viral concentration), but the Ct provided little information about whether the infection was in the proliferation or the clearance stage.

Next, we assessed whether a second test within two days of the first could improve these predictions. We collected all positive samples with a subsequent sample (positive or negative) that was taken within two days and repeated the above calculations, calculating the probability that a Ct value within a five-unit window, followed by a second test with higher/lower Ct, corresponded to an active *vs.* persistent infection and to the proliferation *vs.* clearance stages. A positive test followed by a second test with lower Ct (higher viral RNA concentration) was slightly more likely to be associated with an active infection than a positive test alone (**Figure 3A**). Similarly, a positive test followed by a second test with lower Ct (higher viral RNA concentration) was much more likely to be associated with the proliferation phase than with the clearance phase (**Figure 3B**).

This report provides first comprehensive data on the early-infection RT-qPCR Ct dynamics associated with SARS-CoV-2 infection. Our findings highlight that repeated PCR tests can be used to infer the stage of a patient's infection. While a single test can inform on whether a patient is in the acute or persistent viral RNA shedding stages, a subsequent test can help identify whether viral RNA concentrations are increasing or decreasing, thus informing clinical care. Moreover, as contagiousness varies over the course of an infection *(3)*, our results also indicate how Ct values and dynamics can inform on the nature and duration of interventions needed to reduce the risk of onward transmission.

Our findings on the duration of SARS-CoV-2 viral RNA shedding expand on and agree with previous studies *(11)* and with observations that peak Ct does not differ substantially between symptomatic and asymptomatic individuals *(4)*. While previous studies have largely relied on serial sampling of admitted hospital patients, our study used prospective sampling of ambulatory infected individuals to characterize complete viral dynamics for the presymptomatic stage and for individuals who did not report symptoms. This allowed us to assess differences between the viral RNA proliferation and clearance stages for individuals with and without reported symptoms. The similarity in the early-infection viral RNA dynamics for both symptomatic and asymptomatic individuals underscores the need for SARS-CoV-2 screening regardless of symptoms. The rapid progression from a negative test to a peak Ct value 2-4 days later provides empirical support for screening and surveillance strategies that employ frequent rapid testing to identify potentially infectious individuals *(12,13)*. Taken together, the dynamics of viral RNA shedding substantiate the need for frequent population-level SARS-CoV-2 screening and a greater availability of diagnostic tests.

Our findings are limited for several reasons. The cohort does not constitute a representative sample from the population, as it was predominantly male and included professional athletes. Some of the trajectories were sparsely sampled, limiting the precision of our posterior estimates. The reporting of symptoms was imperfect, as follow-up to identify individuals who developed symptoms after initial evaluation was not systematic. As with all predictive tests, the probabilities that link Ct values with infection stages (**Figure 3**) pertain to the population from which they were calibrated and do not necessarily generalize to other populations for which the prevalence of infection and testing protocols may differ. Still, we anticipate that the central patterns will hold across populations: first, that low Cts (<30) strongly predict acute infection, and second, that a follow-up test collected within two days of an initial positive test can substantially help to discern whether a person is closer to the beginning or the end of their infection. Our study did not test for the presence of infectious virus, though previous studies have documented a close inverse correlation between Ct values and culturable virus *(14)*.

Prospective longitudinal studies using larger samples of representative populations will be valuable in affirming the generalizability of our findings. We demonstrated the value of Ct values from paired PCR tests, focusing on two tests because of the real-world challenges to performing multiple tests; future studies could further specify the precise predictive probabilities of a given sequence of Ct values, establishing the number and sequence of testing that would make for a clinically useful decision tool. Such studies should be carried out in demographically diverse populations with varying levels of SARS-CoV-2 prevalence, as the predictive probabilities of Ct sequences will vary according to individual characteristics (*e.g.* age) and the level of circulation in the population *(3)*.

To manage the spread of SARS-CoV-2, we must develop novel technologies and find new ways to extract more value from the tools that are already available. Our results, building on the first reports of the dynamics of the proliferation stage, suggest that integrating the quantitative viral RNA trajectory into algorithms for clinical management could offer benefits. The ability to chart a patient's progress through their infection underpins our ability to provide appropriate clinical care and to institute effective measures to reduce the risk of onward transmission. Marginally more sophisticated diagnostic and screening algorithms may greatly enhance our ability to manage the spread of SARS-CoV-2 using tests that are already available.

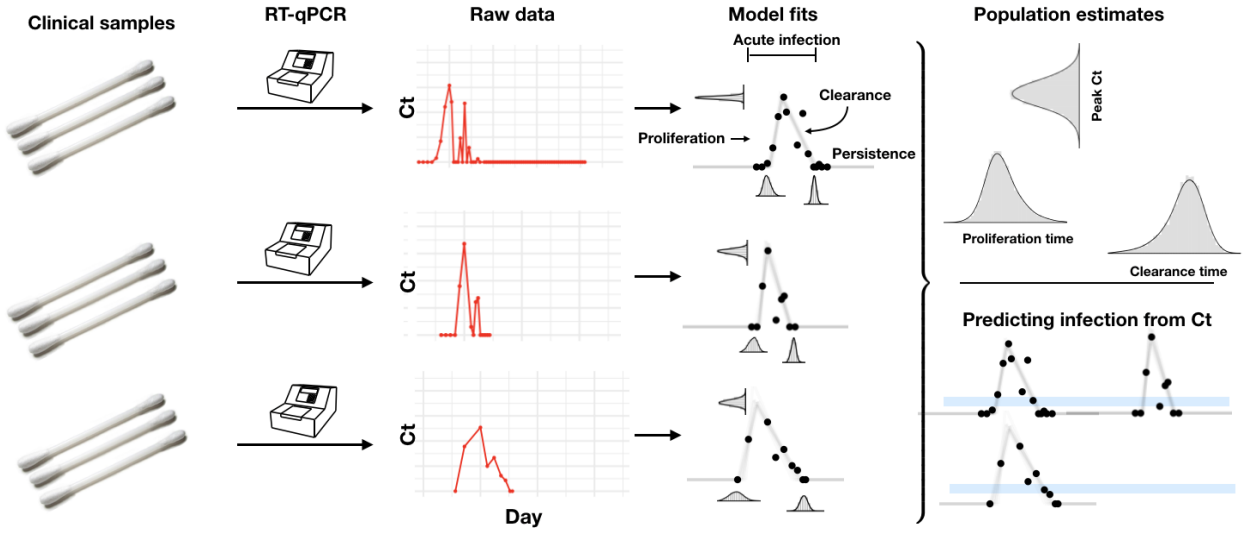


Figure 1. Illustration of the analysis pipeline. Combined anterior nares and oropharyngeal swabs were tested using a RT-qPCR assay to generate longitudinal Ct values ('Raw data', red points) for each person. Using a statistical model, we estimated Ct trajectories consistent with the data, represented by the thin lines under the 'Model fits' heading. These produced posterior probability distributions for the peak Ct, the duration of the proliferation phase (infection onset to peak Ct), and the duration of the clearance phase (peak Ct to resolution of acute infection) for each person. We estimated population means for these quantities. The model fits also allowed us to determine how frequently a given Ct value or pair of Ct values within a five-unit window (blue bars, bottom-right pane) was associated with the proliferation phase, the clearance phase, or a persistent infection.

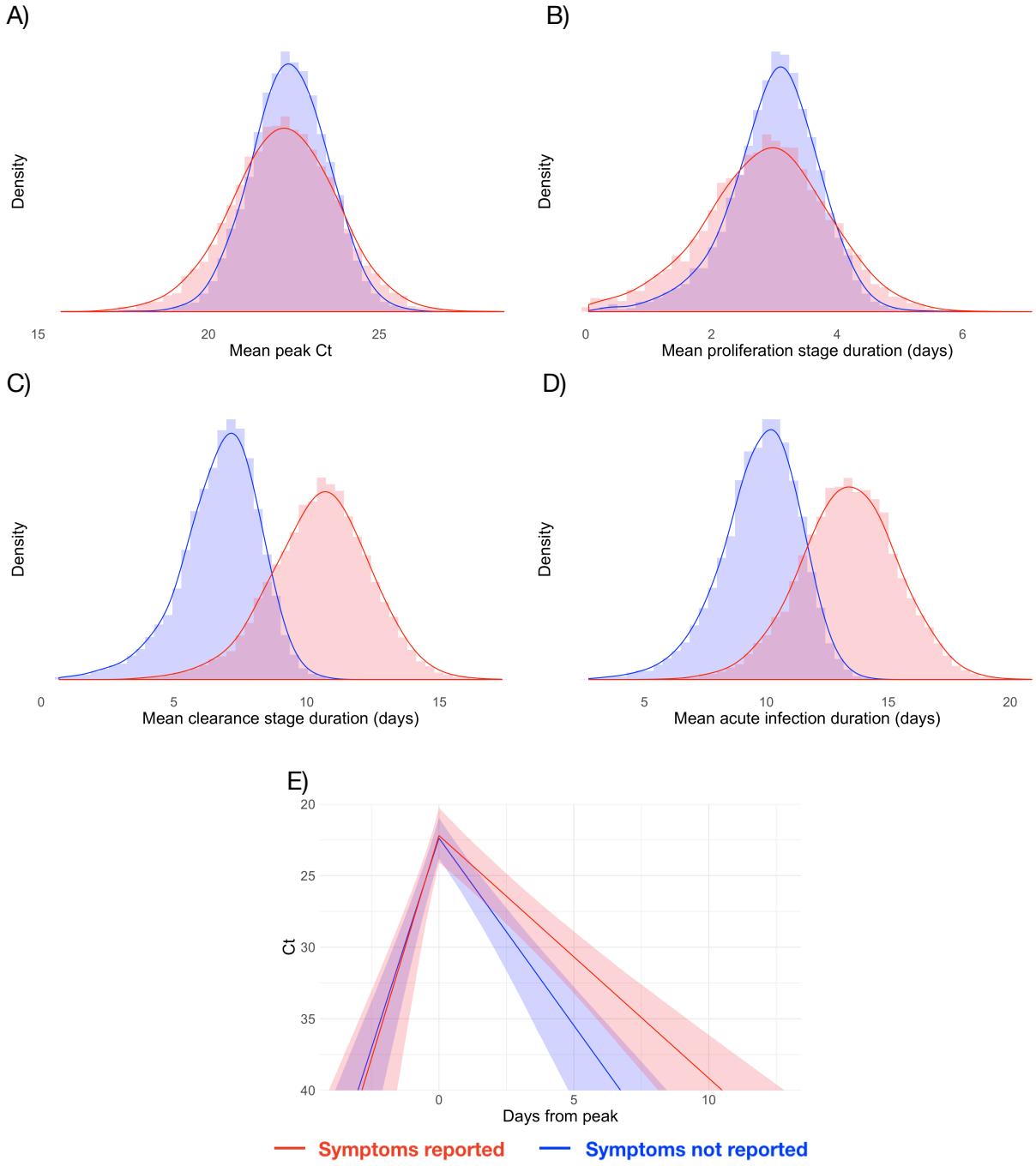


Figure 2. Peak Ct and infection stage duration distributions according to symptoms reported at time of diagnosis. Histograms (colored bars) of 10,000 simulated draws from the posterior distributions for mean peak Ct value (A), mean duration of the proliferation stage (infection detection to peak Ct, B), mean duration of the clearance stage (peak Ct to resolution of acute RNA shedding, C), and total duration of acute shedding (D) across the 46 individuals with an acute infection. The histograms are separated according to whether the person reported symptoms (red, 13 individuals) or did not report symptoms (blue, 33 individuals). The red and blue curves are kernel density estimators for the histograms to assist with visualizing the shapes of the histograms. The mean Ct trajectory corresponding to the mean values for peak Ct, proliferation duration, and clearance duration for symptomatic *vs.* asymptomatic individuals is depicted in (E) (solid lines), where shading depicts the 90% credible intervals.

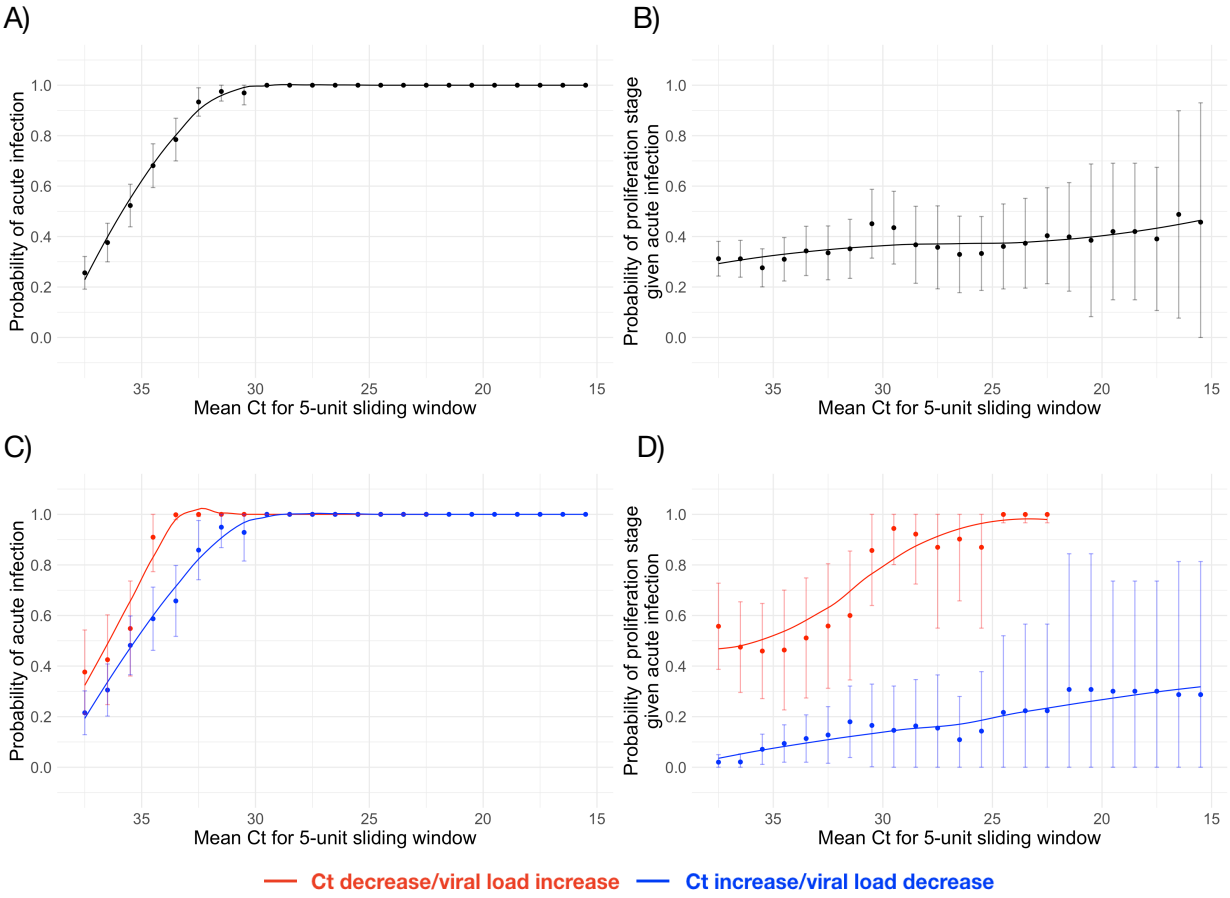


Figure 3. Relationship between single/paired Ct values and infection stage. Probability that a given Ct value lying within a 5-unit window (horizontal axis) corresponds to an acute infection (A, C) or to the proliferation phase of infection assuming an acute infection (B, D). Sub-figures A and B depict the predictive probabilities for a single positive Ct, while sub-figures C and D depict the predictive probabilities for a positive Ct paired with a subsequent test with either lower (red) or higher (blue) Ct. The curves are LOESS smoothing curves to better visualize the trends. Error bars represent the 90% Wald confidence interval.

References

1. World Health Organization. 2020 Coronavirus disease (COVID-19) Situation Report.
2. Furukawa NW, Brooks JT, Sobel J. 2020 Evidence Supporting Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 While Presymptomatic or Asymptomatic. *Emerg. Infect. Dis.* **26**. (doi:10.3201/eid2607.201595)
3. Tom MR, Mina MJ. 2020 To Interpret the SARS-CoV-2 Test, Consider the Cycle Threshold Value. *Clin. Infect. Dis.* **02115**, 1–3. (doi:10.1093/cid/ciaa619)
4. Walsh KA *et al.* 2020 SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *J. Infect.* **81**, 357–371. (doi:10.1016/j.jinf.2020.06.067)
5. Wyllie AL *et al.* 2020 Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. *N. Engl. J. Med.* **February**, NEJMc2016359. (doi:10.1056/NEJMc2016359)
6. Xiao AT, Tong YX, Zhang S. 2020 Profile of RT-PCR for SARS-CoV-2: A Preliminary Study From 56 COVID-19 Patients. *Clin. Infect. Dis.* (doi:10.1093/cid/ciaa460)
7. U.S. Food and Drug Administration. 2020 Quest Diagnostics Infectious Disease, Inc. ("Quest Diagnostics") SARS-CoV-2 RNA Qualitative Real-Time RT-PCR Emergency Use Authorization.
8. U.S. Food and Drug Administration. 2020 Roche Molecular Systems, Inc. cobas SARS-CoV-2 Emergency Use Authorization.
9. Kudo E *et al.* 2020 Detection of SARS-CoV-2 RNA by multiplex RT-qPCR. *PLOS Biol.* **18**, e3000867. (doi:10.1371/journal.pbio.3000867)
10. Cleary B, Hay JA, Blumenstiel B, Gabriel S, Regev A, Mina MJ. 2020 Efficient prevalence estimation and infected sample identification with group testing for SARS-CoV-2. *medRxiv*
11. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. 2020 SARS-CoV-2, SARS-CoV-1 and MERS-CoV viral load dynamics, duration of viral shedding and infectiousness: a living systematic review and meta-analysis. *medRxiv*
12. Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, Tambe M, Mina MJ, Parker R. 2020 Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance. *medRxiv* , 2020.06.22.20136309. (doi:10.1101/2020.06.22.20136309)
13. Paltiel AD, Zheng A, Walensky RP. 2020 Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. *JAMA Netw. Open* **3**, e2016818. (doi:10.1001/jamanetworkopen.2020.16818)
14. Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, Ladhani S, Zambon M, Gopal R. 2020 Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill.* **25**, 1–5. (doi:10.2807/1560-7917.ES.2020.25.32.2001483)
15. Ott IM, Vogels C, Grubaugh N, Wyllie AL. 2020 Saliva Collection and RNA Extraction for SARS-CoV-2 Detection V.2.
16. Carpenter B *et al.* 2017 Stan : A Probabilistic Programming Language. *J. Stat. Softw.* **76**. (doi:10.18637/jss.v076.i01)
17. R Development Core Team R. 2011 R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* **1**, 409. (doi:10.1007/978-3-540-74686-7)

Acknowledgements. We thank the NBA, National Basketball Players Association (NBPA), and all of the study participants who are committed to applying what they learned from sports towards enhancing public health. In particular, we thank D. Weiss of the NBA for his continuous support and leadership. We are appreciative of the discussions from the COVID-19 Sports and Society Working Group. We also thank J. Hay and R. Niehus for helpful suggestions on the statistical approach and P. Jack and S. Taylor for laboratory support.

Funding. This study was funded by the NWO Rubicon 019.181EN.004 (CBFV), a clinical research agreement with the NBA and NBPA (NDG), the Huffman Family Donor Advised Fund (NDG), Fast Grant funding support from the Emergent Ventures at the Mercatus Center, George Mason University (NDG), and the Morris-Singer Fund for the Center for Communicable Disease Dynamics at the Harvard T.H. Chan School of Public Health (YHG).

Author contributions. SMK conceived of the study, conducted the statistical analysis, and wrote the manuscript. JRF conceived of the study, conducted the laboratory analysis, and wrote the manuscript. CM conceived of the study, collected the data, and wrote the manuscript. CT analyzed the data and edited the manuscript. KYS analyzed the data and edited the manuscript. CCK conducted the laboratory analysis and edited the manuscript. SJ conducted the laboratory analysis and edited the manuscript. IMO conducted the laboratory analysis. CBFV conducted the laboratory analysis. JW conducted laboratory analysis and edited the manuscript. JW conducted laboratory analysis and edited the manuscript. JD conceived of the study and edited the manuscript. DJA contributed to data analysis and edited the manuscript. JM contributed to data analysis and edited the manuscript. DDH conceived of the study and edited the manuscript. NDG conceived of the study, oversaw the study, and wrote the manuscript. YHG conceived of the study, oversaw the study, and wrote the manuscript.

Competing interests.

JW is an employee of Quest Diagnostics. JW is an employee of Bioreference Laboratories.

Supplement.

Ethics.

Residual de-identified viral transport media from anterior nares and oropharyngeal swabs collected NBA players, staff, and vendors were obtained from Quest Diagnostics or 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).

Additional testing protocol details.

Residual viral transport media (VTM) from Quest Diagnostics or BioReference Laboratories were shipped overnight to Yale on dry ice. VTM was thawed on ice and 300 µL was used for RNA extraction using the MagMAX Viral/Pathogen Nucleic Acid Isolation Kit and the KingFisher Flex robot (Thermo Fisher Scientific, Waltham, MA *(15)*). Total nucleic acid was eluted into 75ul of elution buffer and SARS-CoV-2 RNA was quantified from 5 µL of extracted total RNA using a multiplexed version of the CDC RT-qPCR assay that contains the 2019-nCoV_N1 (N1), 2019-nCoV-N2 (N2), and human RNase P (RP) primer-probe sets *(9)*. The RT-qPCR was performed using the Luna Universal Probe One-Step RT-qPCR Kit (New England Biolabs, Ipswich, MA, US) and the following thermocycler conditions: (1) reverse transcription for 10 minutes at 55°C, (2) initial denaturation for 1 min at 95°C, and PCR for 45 cycles of 10 seconds at 95°C and 30 seconds at 55°C on the CFX96 qPCR machine (Bio-Rad, Hercules, CA, US).

Converting Ct values. Most (n = 226) of the 312 positive samples in the raw dataset underwent RT-qPCR at the Yale laboratory. We used the Yale Ct value whenever it was available. Still, 86 samples underwent initial diagnostic testing at BioReference Laboratories but not confirmatory testing at the Yale laboratory. Both platforms rely on a multiplex RT-qPCR strategy. The two testing platforms yield slightly different Ct values, as evidenced by the 94 samples the underwent RT-qPCR at both facilities (**Supplemental Figure 5**). For comparison between platforms, target 1 from the Roche cobas assay, which is specific to SARS-CoV-2, and the N1 target from the Yale multiplex assay were used. For the 86 samples that were not processed at the Yale laboratory, we adjusted the Ct values using the best-fit (minimum sum of squares) linear regression between the initial Ct value and the Yale Ct value for the samples that were processed in both facilities. To do so, we estimated the coefficients β_0 and β_1 in the following regression equation:

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

Here, y_i denotes the i^{th} Ct value from Yale, x_i denotes the i^{th} Ct value from the initial test, and ε_i is an error term with mean 0 and constant variance across all samples. The resulting fit (**Supplemental Figure 5**) was strong (R^2 = 0.86) with homoscedastic residuals (**Supplemental Figure 6**) that are approximately normally distributed, as evidenced by a Q-Q plot (**Supplemental Figure 7**).

Data parsing. The raw dataset included 3,207 test results for 102 individuals. We excluded 21 individuals who had 5 or fewer tests, since the data for these individuals were too sparse to reliably infer a Ct trajectory. We also excluded 13 individuals who did not record any Ct values that surpassed the RT-qPCR limit of detection (40). We removed 146 entries for which the test result was recorded as 'positive' but there was no associated Ct value; these tests were initially

conducted on an instrument that provided only a binary diagnosis and the samples were not available for confirmatory testing. This left 2,411 total tests for 68 individuals for the main analysis. We trivially shifted the date indices so that date 0 corresponded to the time of the minimum Ct. We set the Ct value for negative tests equal to the limit of detection. For the statistical analysis, we removed any sequences of 3 or more consecutive negative tests to avoid overfitting to these trivial values.

Model fitting. 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. Since Ct values are roughly proportional to the negative logarithm of viral concentration, 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, 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). This idealized trajectory is depicted in **Supplemental Figure 8.** 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 $0 \leq \omega_p \leq 14$ days and $0 \leq \omega_p \leq 30$ days to prevent inferring unrealistically 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 *σ(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.

For the 86 samples that were not tested in the Yale laboratory, we included additional uncertainty in the observed Ct value by inflating *σ(t)*, such that

$$\sigma(t) = (\tilde{\sigma}^2 + \epsilon^2 I_{adj})^{1/2}$$

Here, *σ*(tilde) is a constant, *ε* is the standard deviation of the residuals from the linear fit between the initial test and the Yale laboratory test, and I_{adj} is an indicator variable that is 1 if the sample at time *t* was adjusted and 0 otherwise.

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 (μ.) for symptomatic and asymptomatic individuals. We used a Hamiltonian Monte Carlo fitting procedure implemented in Stan (version 2.24) *(16)* and R (version 3.6.2) *(17)* to estimate the individual-level parameters ω_p, ω_r, δ, and t_p as well as the population-level parameters σ(tilde), $\mu_{\omega p}$, $\mu_{\omega r}$, μ_δ, $\sigma_{\omega p}$, $\sigma_{\omega r}$, and σ_δ. We used the following priors:

Hyperparameters:

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

$\mu_{\omega p} \sim \text{Normal}(14/2, 14/6)\ [0, 14]$
$\mu_{\omega r} \sim \text{Normal}(30/2, 30/6)\ [0, 30]$
$\mu_\delta \sim \text{Normal}(40/2, 40/6)\ [0, 40]$

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

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

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

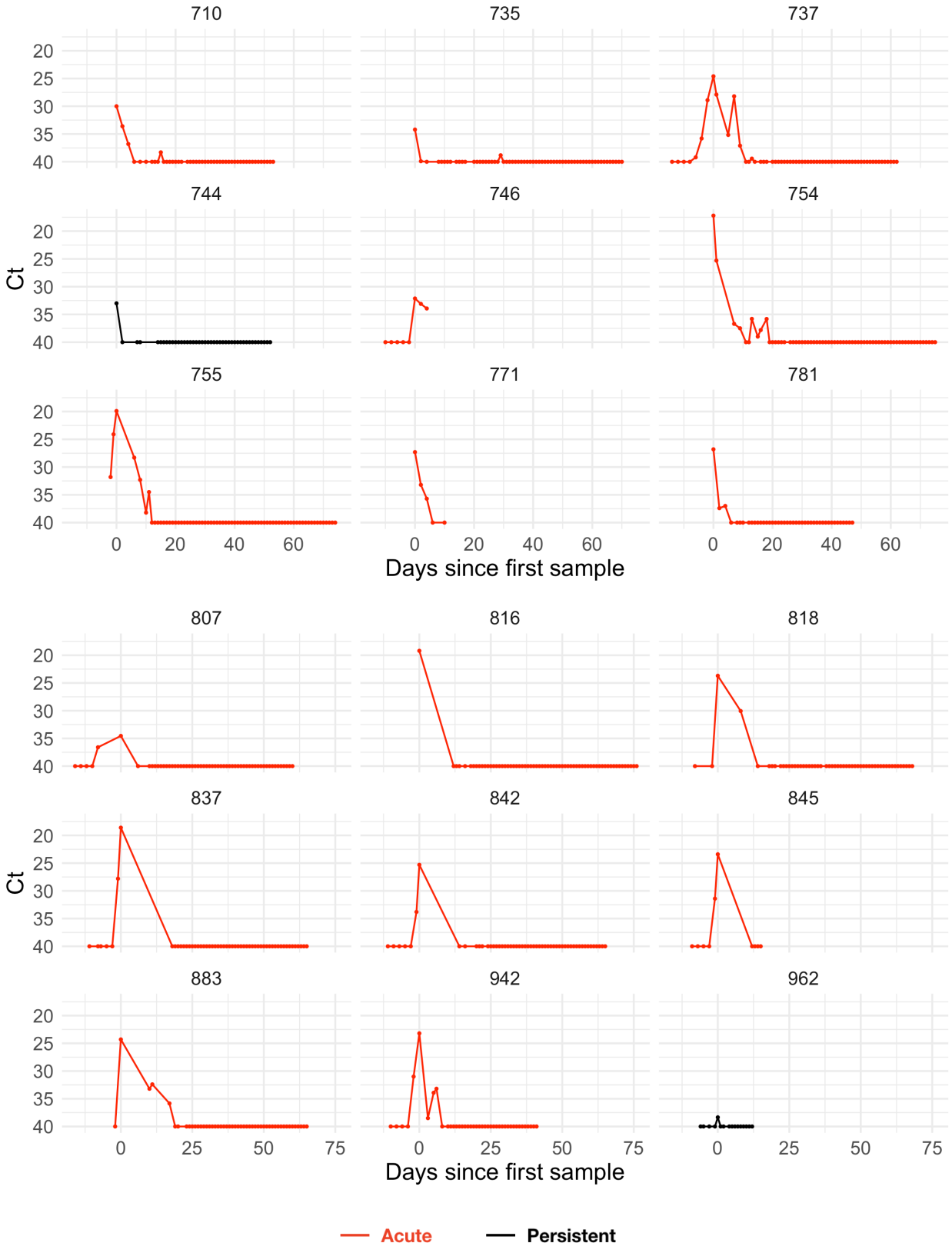
Individual-level parameters:

ω_p ~ Normal($\mu_{\omega p}$, $\sigma_{\omega p}$) [0,14]
ω_r ~ Normal($\mu_{\omega r}$, $\sigma_{\omega r}$) [0,30]
δ ~ Normal(μ_δ, σ_δ) [0,40]
t_p ~ Normal(0, 2)

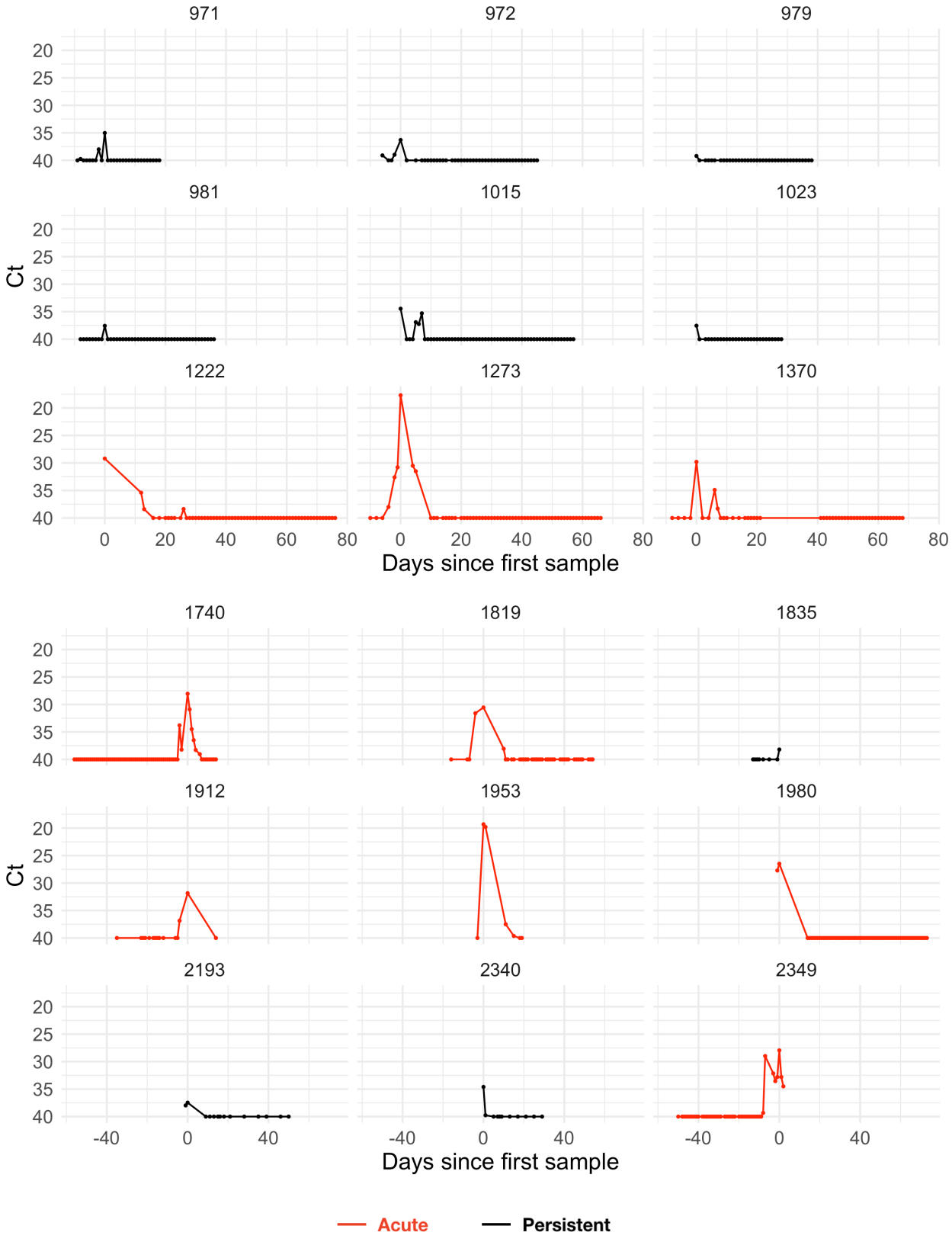
The values in square brackets denote truncation bounds for the distributions. We chose a vague half-Cauchy prior with scale 5 for the observation variance σ(tilde). The priors for the population mean values (μ.) are normally-distributed priors spanning the range of allowable values for that parameter; 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 (σ.) 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 5,000 iterations each with a target average proposal acceptance probability of 0.99. 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 except for the t_p and ω_r associated with individual 1370, as the posterior distributions for those parameters were multi-modal (see **Supplemental Figures 12-13**). This indicates good overall mixing of the chains. There were fewer than 10 divergent iterations (<0.1% of the transitions after warm-up), indicating good exploration of the parameter space.

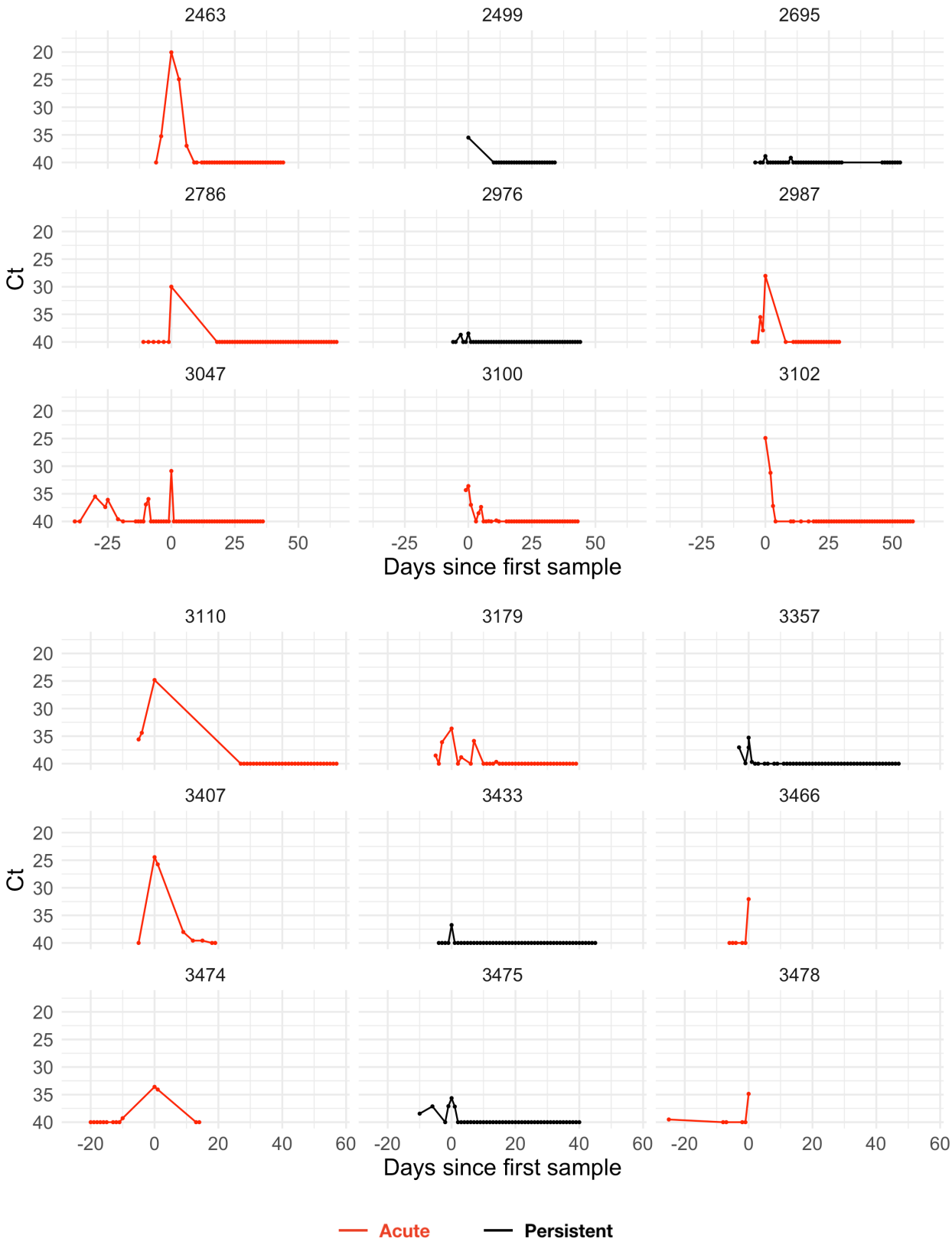
The posterior distributions for μ_δ, $\mu_{\omega p}$, and $\mu_{\omega r}$, estimated separately for symptomatic and asymptomatic individuals, are reported in **Figure 2** (main text). We fit a second model that did not distinguish between symptomatic and asymptomatic individuals. The posterior distributions for these same parameters under this model are depicted in **Supplemental Figure S9**. The posterior distributions for the individual-level parameters ω_p, ω, and δ are depicted in **Supplemental Figures S10-12,** with 500 sampled trajectories from these posterior distributions for each individual depicted in **Supplemental Figure 13.** The overall combined posterior distributions for the individual-level parameters ω_p, ω_r, and δ are depicted in **Supplemental Figure 14.** We estimated the best-fit normal (for δ) and gamma (for ω_p and ω_r) distributions using the 'fitdistrplus' package implemented in R (version 3.6.2) *(17)*.



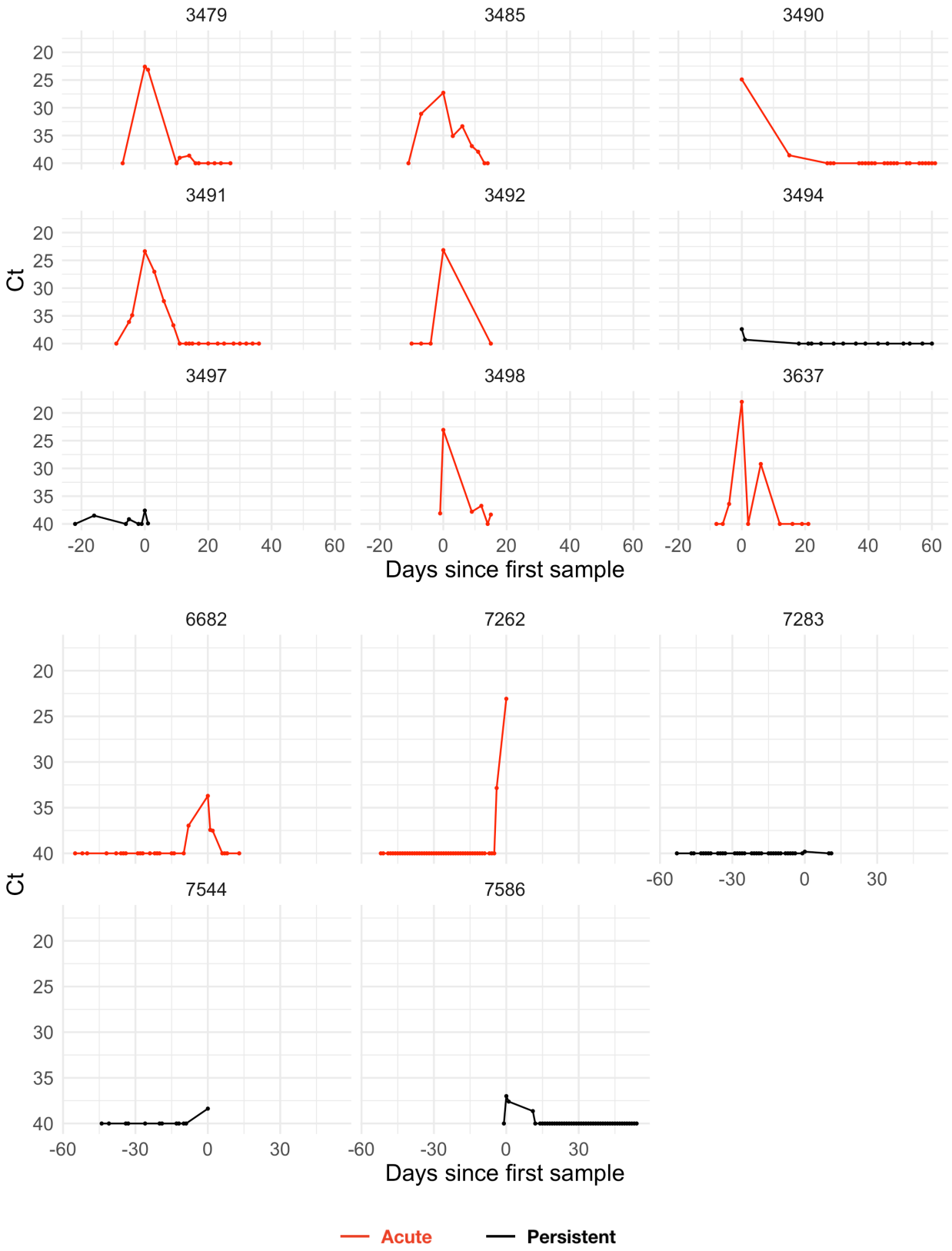
Supplemental Figure 1. Observed Ct values from the study participants (1/4). Points depict observed Ct values, which are connected with lines to better visualize trends. Individuals with presumed acute infections are marked in red. All others are in black.



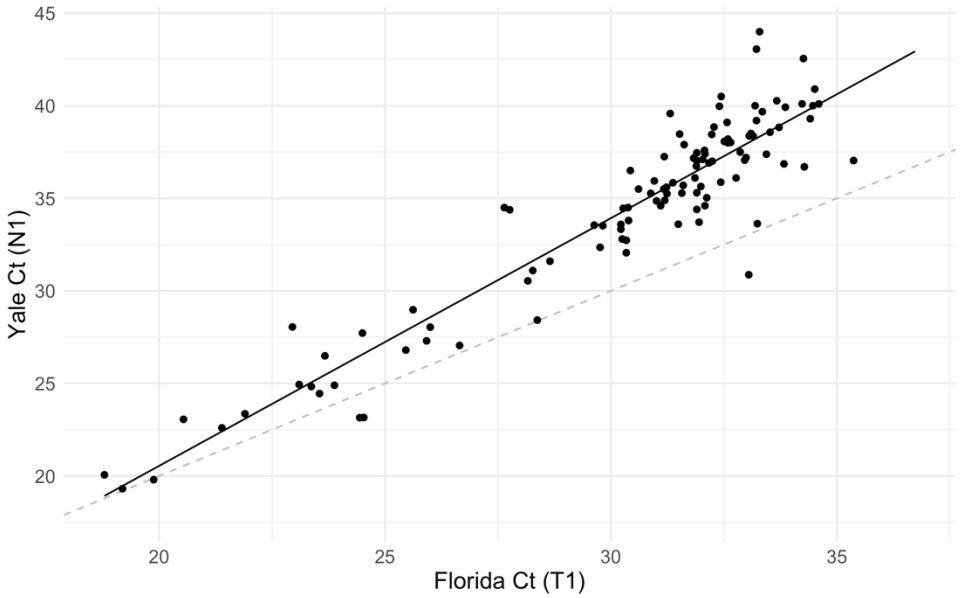
Supplemental Figure 2. Observed Ct values from the study participants (2/4). Points depict observed Ct values, which are connected with lines to better visualize trends. Individuals with presumed acute infections are marked in red. All others are in black.



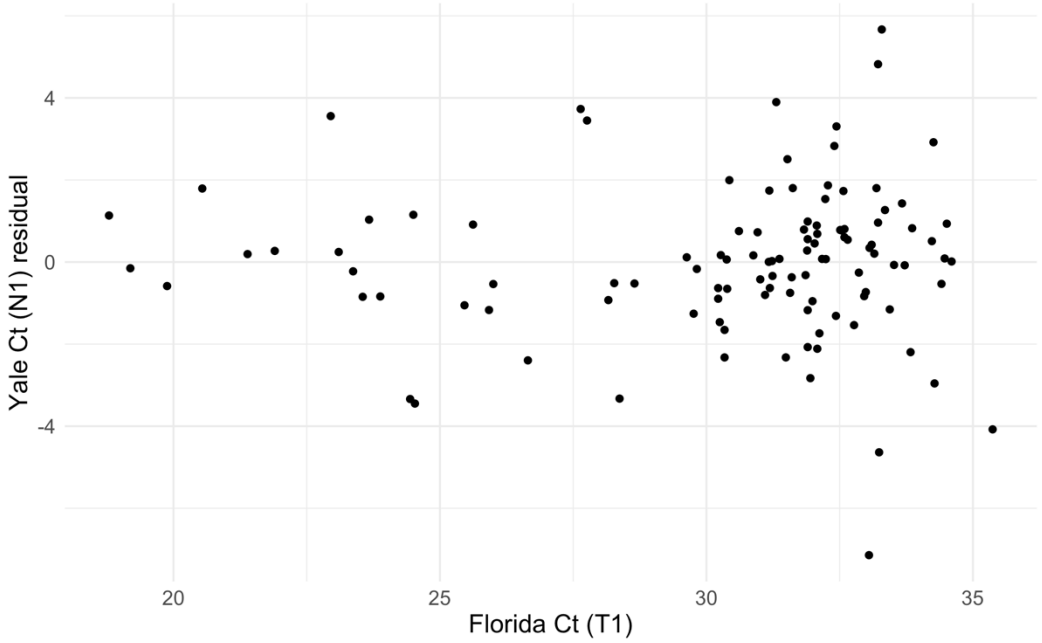
Supplemental Figure 3. Observed Ct values from the study participants (3/4). Points depict observed Ct values, which are connected with lines to better visualize trends. Individuals with presumed acute infections are marked in red. All others are in black.



Supplemental Figure 4. Observed Ct values from the study participants (4/4). Points depict observed Ct values, which are connected with lines to better visualize trends. Individuals with presumed acute infections are marked in red. All others are in black.



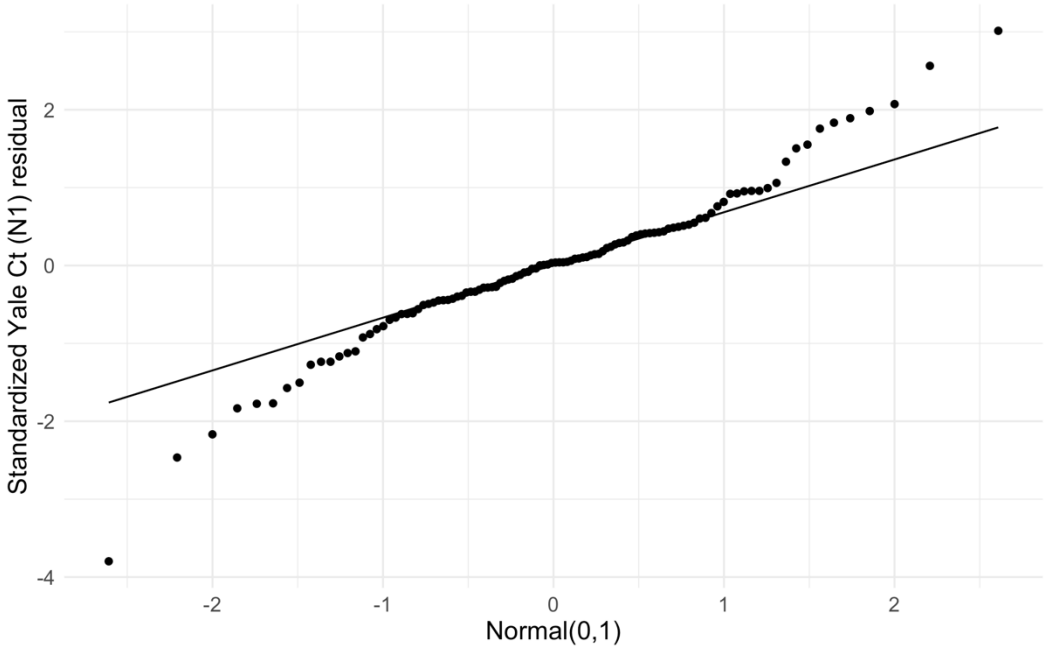
Supplemental Figure 5. Ct values from the Yale and Florida labs. Points depict the Ct values for SARS-CoV-2 nasal swab samples that were tested in both Florida and Yale labs. Ct values from Florida represent Target 1 (ORF1ab) on the Roche cobas system, and Ct values from Yale represent N1 in the Yale multiplex assay. The solid black line depicts the best-fit linear regression (intercept = –6.25, slope = 1.34, R^2 = 0.86). The dashed black line marks the 1-1 line where the points would be expected to fall if the two labs were identical.



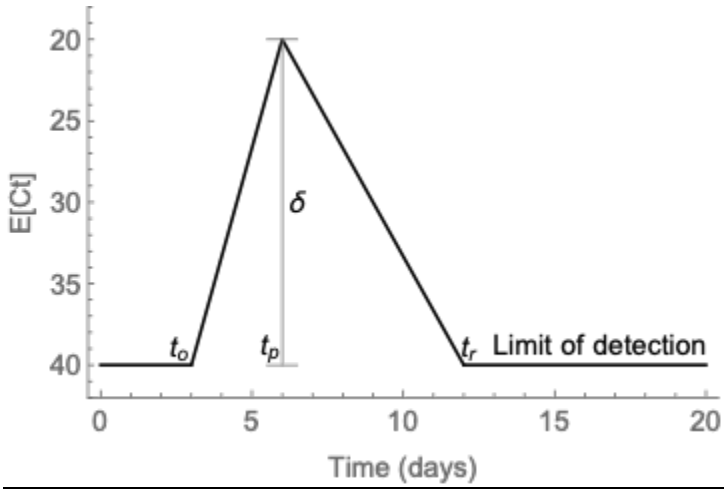
546

547 **Supplemental Figure 6. Residuals from the Yale/Florida Ct regression.** Points depict the residual after removing
548 the best-fit linear trend in the relationship between the Yale and Florida Ct values.

549



Supplemental Figure 7. QQ plot of the residuals from the Yale/Florida Ct regression. The residuals were standardized (subtracted the mean and divided by the standard deviation) before comparing with the theoretical quantiles of a normal distribution with mean 0 and standard deviation 1. The points depict the empirical quantiles of the data points and the line depicts the where the points would be expected to fall if they were drawn from a standard normal distribution.



Supplemental Figure 8. A theoretical Ct trajectory. E[Ct] is the expected Ct value on a given day. The Ct begins at the limit of detection, then declines from the time of infection (t_o) to the peak at δ cycles below the limit of detection at time t_p. The Ct then rises again to the limit of detection after t_r days. The model incorporating these parameter values used to generate this piecewise curve is given in Equation S1 (**Supplemental Methods**).

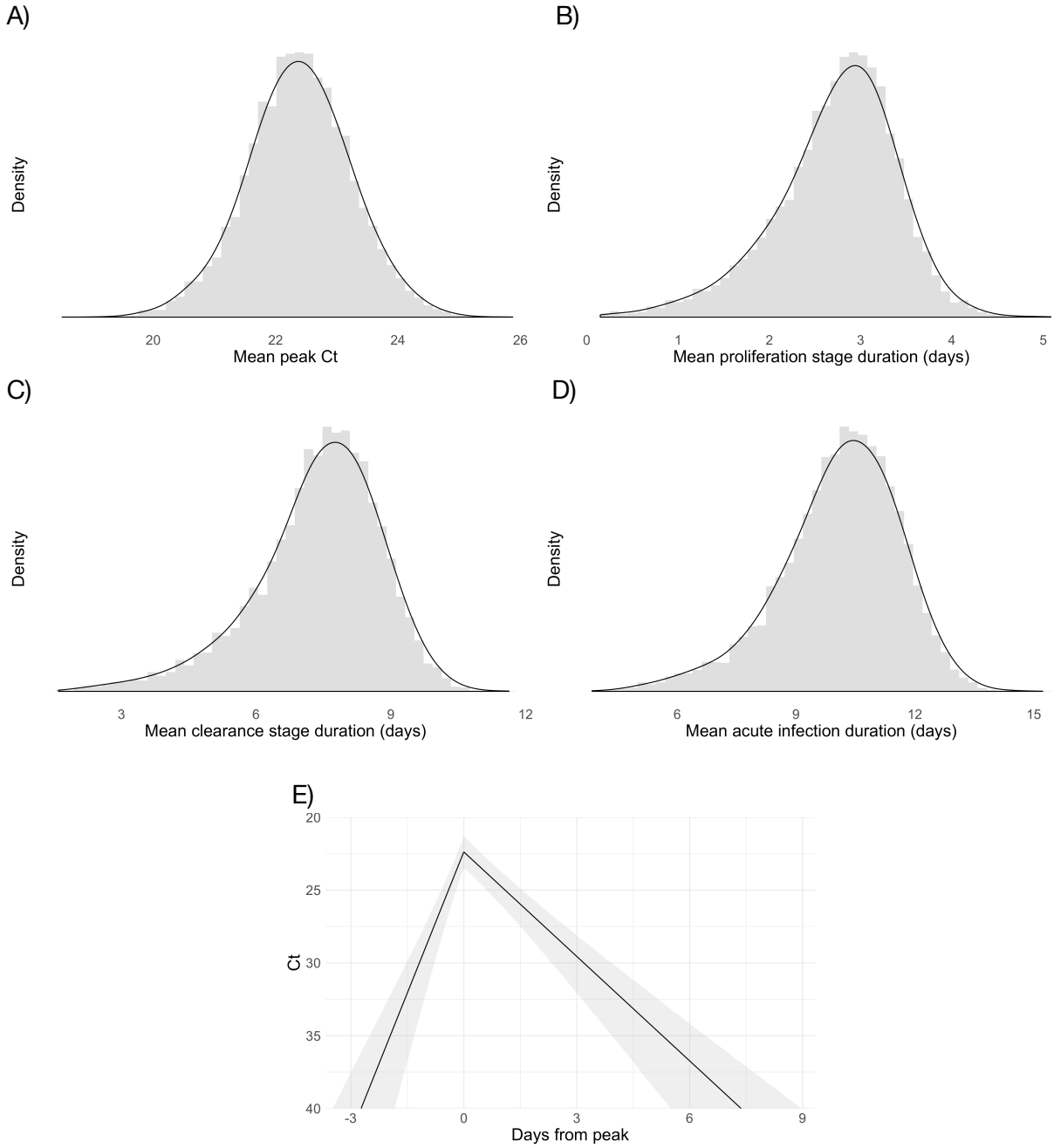
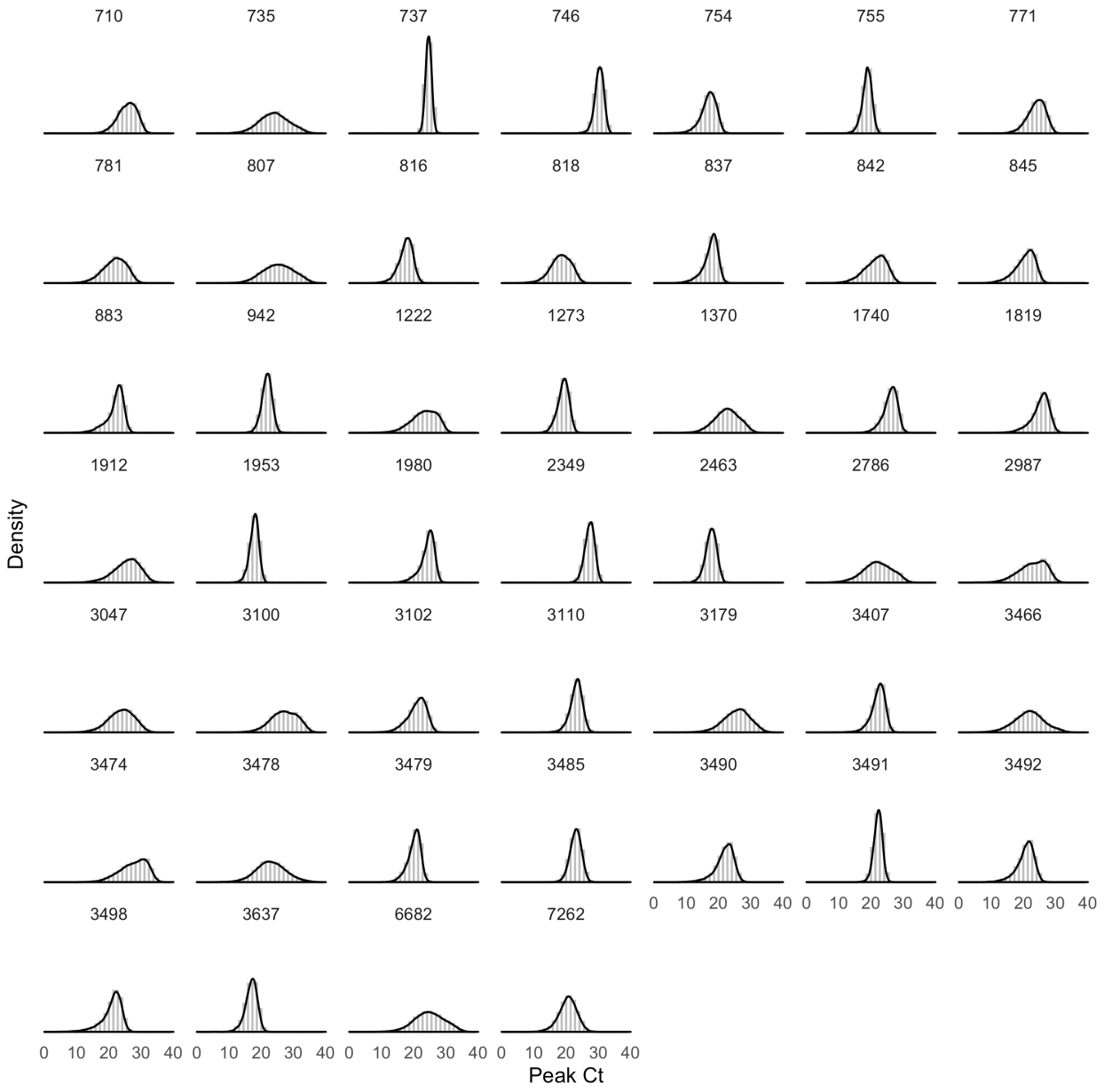
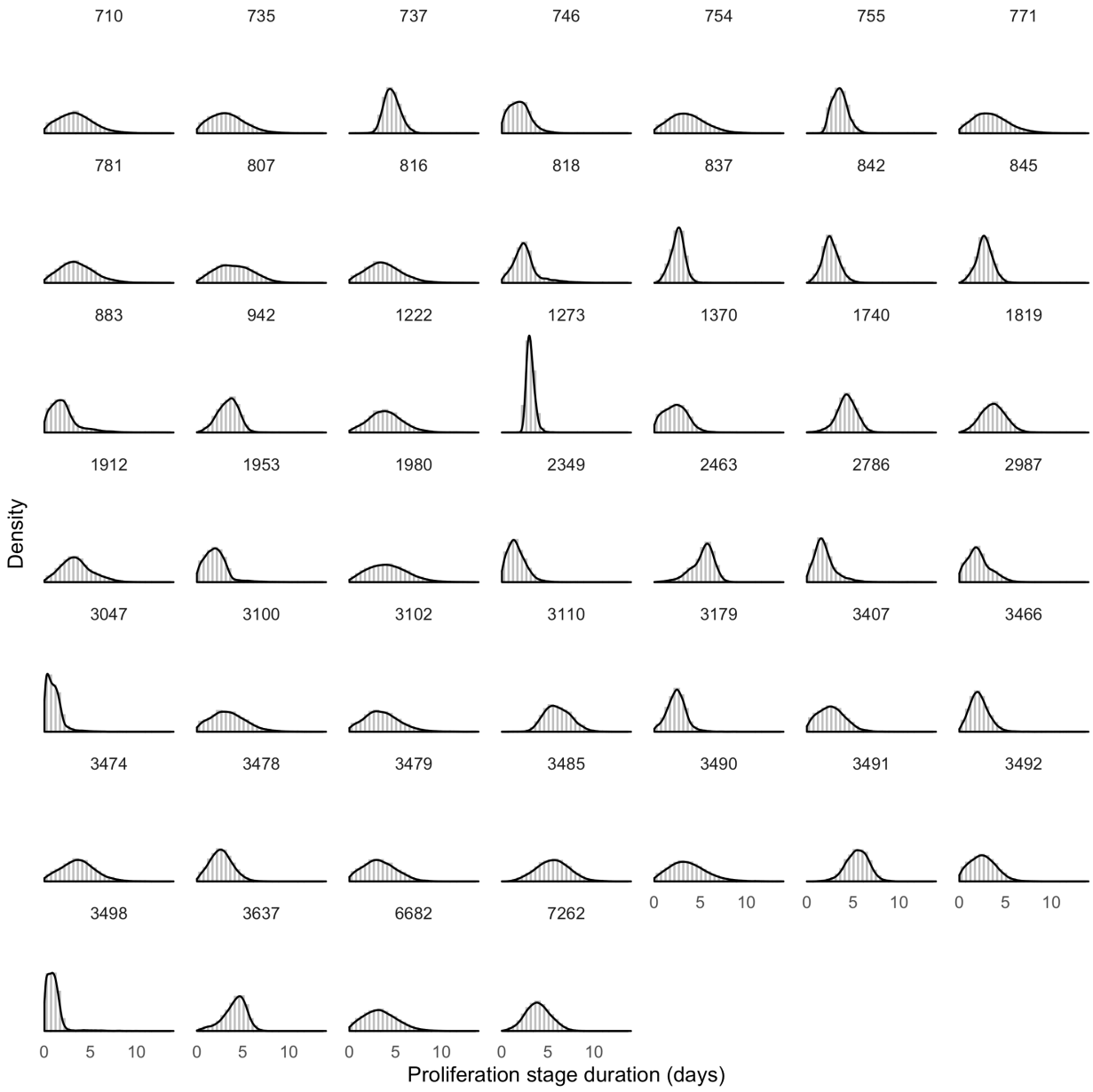


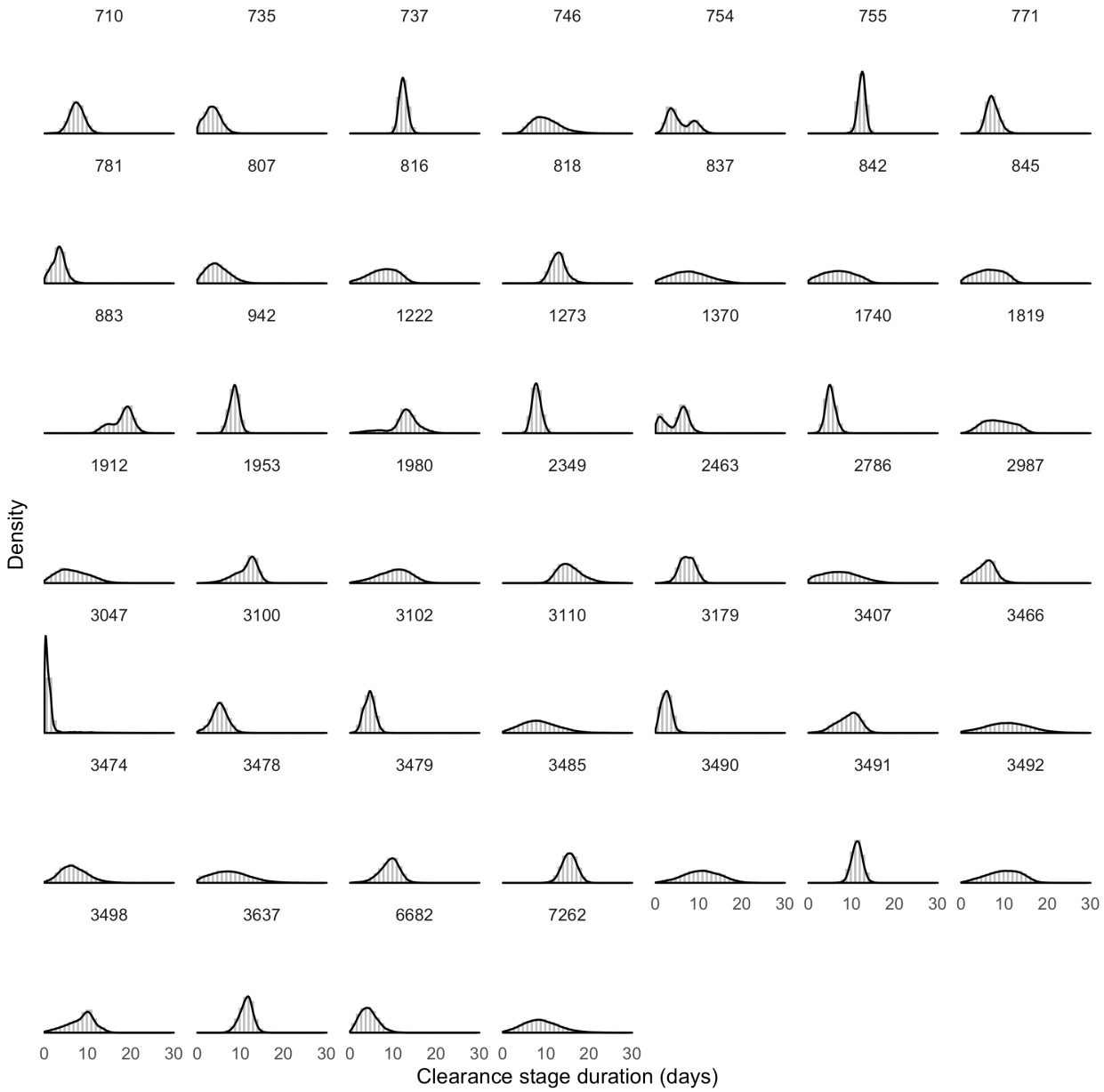
Figure S9. Mean peak Ct and waiting time distributions for individuals with acute infections. Histograms (colored bars) of 10,000 posterior draws from the distributions for peak Ct value (A), duration of the proliferation stage (infection detection to peak Ct, B), duration of the clearance stage (peak Ct to resolution of acute RNA shedding, C), and total duration of acute shedding (D) across the 46 individuals with a verified infection. The curves are kernel density estimators for the histograms to assist with visualizing the shapes of the histograms. The mean Ct trajectory corresponding to the mean values for peak Ct, proliferation duration, and clearance duration is depicted in (E) (solid lines), where shading depicts the 90% credible interval.



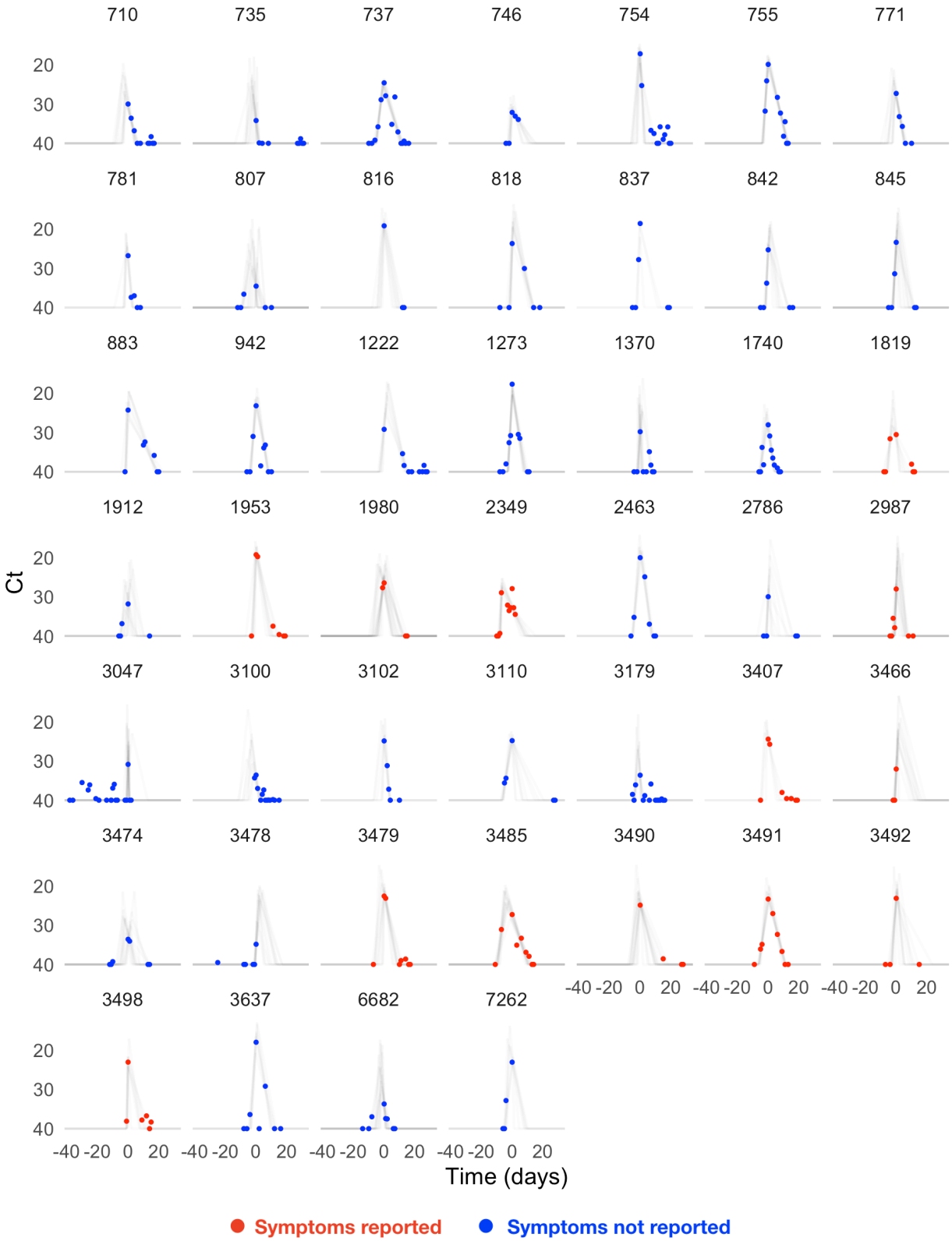
Supplemental Figure 10. Posterior peak Ct distributions for the 46 individuals with acute infections.



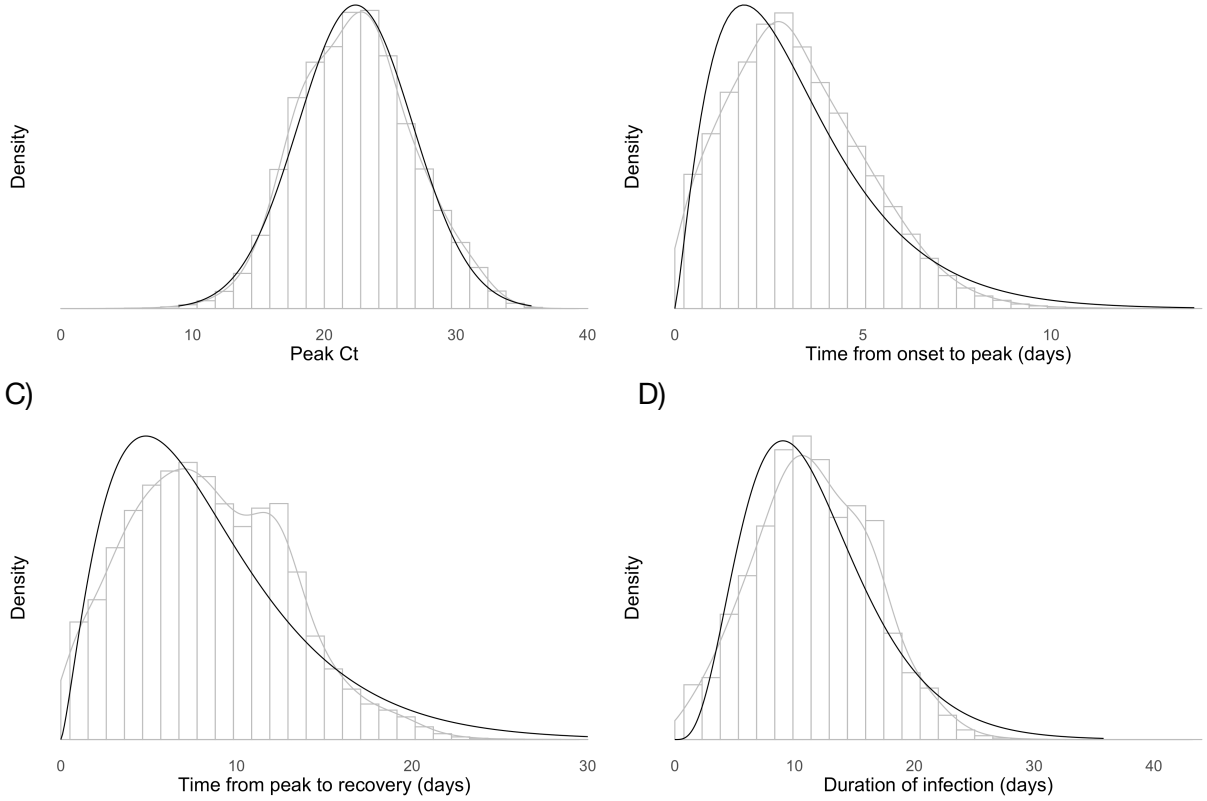
Supplemental Figure 11. Posterior distributions for the duration of the proliferation stage for 46 individuals with acute infections.



Supplemental Figure 12. Posterior distributions for the clearance stage duration for 46 individuals with acute infections.



Supplemental Figure 13. Best-fit Ct trajectories for the 46 individuals with acute infections. Thin grey lines depict 500 sampled trajectories. Points represent the observed data, with symptomatic individuals represented in red and asymptomatic individuals in blue.



Supplemental Figure 14. Individual-level peak Ct and waiting time distributions. Histograms (grey bars) of 10,000 posterior draws from the distributions for peak Ct value (A), time from onset to peak (B), time from peak to recovery (C), and total duration of infection (D) across the 46 individuals with an acute infection. Grey curves are kernel density estimators to more clearly exhibit the shape of the histogram. Black curves represent the best-fit normal (A) or gamma (B, C, D) distributions to the histograms. The duration of infection is the sum of the time from onset to peak and the time from peak to recovery. The best-fit normal distribution to the posterior peak Ct distribution had mean 22.3 and standard deviation 4.2; the best-fit gamma distribution to the proliferation stage duration had shape parameter 2.3 and inverse scale parameter 0.7; the best-fit gamma distribution to the clearance stage duration had shape parameter 2.4 and inverse scale parameter 0.3; and the best-fit gamma distribution to the total duration of infection had shape parameter 4.3 and inverse scale parameter 0.4.