Duration of Viral Nucleic Acid Shedding and Early Reinfection with the Severe Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Health Care Workers and First Responders

Brad J. Biggerstaff[1], Lara J Akinbami[2,3], Craig Hales[2,3], Philip A. Chan[4], Lyle R. Petersen[1]

1. Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado, 80521 USA
2. National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland, 20782 USA
3. U.S. Public Health Service, Rockville, Maryland, 20852 USA
4. Rhode Island Department of Health, Providence, Rhode Island, 02908, USA

Corresponding author: Lyle R. Petersen, MD, MPH; FAX: 970-266-3502; Phone: 970-215-2968; EMAIL: LXP2@CDC.GOV

(abstract): 100/100 words

(text): 1997/2000 words

Short summary: Among frequently tested immunocompetent adults, analysis of gaps of varying length between subsequent positive and negative SARS-CoV-2 nucleic acid test results indicated that a 90-day period between positive results can reliably define reinfection, although reinfection can occur at shorter intervals.

Footnote page

None of the authors have a commercial or other association that might pose a conflict of interest (e.g., pharmaceutical stock ownership, consultancy, advisory board membership, relevant patents, or research funding).

There were no financial sources of support for this study.

The study results have not previously been presented at a public meeting or in any other forum.

Correspondence can be addressed to Lyle R. Petersen, MD, MPH, 3156 Rampart Road, Fort Collins, CO 80521, email LXP2@CDC.GOV.

Disclaimer: The content and conclusions are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

ABSTRACT

We estimated the distributions of duration of SARS-CoV-2 nucleic acid shedding and time to reinfection among 137 persons with at least two positive nucleic acid amplification test (NAAT) results from March to September 2020. We analyzed gaps of varying length between subsequent positive and negative NAAT results and estimated a mean duration of nucleic acid shedding of 30.1 (95% CI 26.3, 34.5) days. The mean time to reinfection was 89.1 (95% CI 75.3, 103.5) days. Together, these indicate that a 90-day period between positive NAAT results can reliably define reinfection in immunocompetent persons although reinfection can occur at shorter intervals.

Key Words: SARS-CoV-2; nucleic acid; reinfection; infection; viral shedding

INTRODUCTION/BACKGROUND

While the most definitive evidence for reinfection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) derives from persons with sequential infections from SARS-CoV-2 viruses with different nucleic acid sequences, most sequences are not available for analyses . Thus, many reinfections are defined by positive SARS-CoV-2 nucleic acid amplification test (NAAT) results separated by some time period . However, the minimal time interval to define reinfection is unclear given sparse extant data on viral RNA shedding duration following infection as well as how quickly reinfection can occur. While reports of immunocompromised patients indicate that viral RNA shedding can exceed 100 days , two case series of hospitalized patients indicate a median shedding duration of 20-31 days . Juxtaposed to these reports of viral nucleic acid shedding lasting weeks to months, one metanalysis suggested a mean interval of only 48 days between subsequent infections when reinfection was defined as two positive NAAT results separated by a negative NAAT result .

To estimate the distribution of duration of SARS-CoV-2 RNA shedding and time to reinfection, we examined NAAT result data from a cohort of health care workers and first responders who had participated in a SARS-CoV-2 serology survey at their workplaces in Rhode Island .

METHODS

Study Population and General Approach

The source study population consisted of 11,985 Rhode Island healthcare workers and first responders who participated in a SARS-CoV-2 serology survey in July and August 2020. Seroprevalence among nursing home personnel (13.1%) exceeded that of hospital personnel (1.3%) and first responders (3.1%) . For the current study, NAAT results from March 1 to September 15, 2020 were obtained for all serosurvey participants from mandatory statewide surveillance reporting (Supplemental Figure 1 shows

overall Rhode Island COVID-19 case counts). Statewide NAAT results were linked to the serology study's unique participant identification numbers and anonymized results were provided to CDC. The overall study population had a median of two NAAT results per person, while nursing home workers had a median of 13 due to weekly testing initially recommended and then mandated in high-incidence counties . However, specific testing schedules or criteria were not available for individual agencies. The final study population comprised 137 persons with ≥2positive NAAT results.

Each participant had positive and negative NAAT results separated by temporal gaps of varying lengths (Supplementary Figure 2). We hypothesized that a series of positive results separated by shorter gaps likely represented viral RNA shedding from a single infection; whereas positive results separated by longer gaps likely represented reinfections. We used a statistical approach to assess if there were in fact two distinct distributions of short and long gaps. We then estimated the duration of NAAT positivity from a single infection using the cumulative durations of short gaps for each person. Using the distribution of NAAT positivity and the estimated long gap distribution, we computed the distribution of times from infection to reinfection.

CDC human subjects research officials determined that the activity was public health surveillance as defined in 45 CFR 46.

Statistical methods

We analyzed the temporal gaps between test results, irrespective of testing date. We first assessed the gaps between positive tests to determine the number of apparent groupings using k-means clustering and the "gap statistic" . We identified two classes: short and long gap classes. We then fit gamma, lognormal, Weibull, and normal distributions using maximum likelihood for each class, and selected the model with the minimum Akaike Information Criterion (AIC). For the short gap class, we then defined a maximum duration cutoff for an observed short gap to be the 99th percentile of the distribution.

Using this cutoff, we then computed for each participant the duration of single-infection positivity ("shedding") as the cumulative sum of the observed short gaps from the first positive result until either the first long gap was reached or there were no subsequent positive tests (Figure 1). Because a participant might have tested positive after the last observed positive test had there been testing, the actual date of the end of shedding was censored. To account for this censoring in estimating the distribution of shedding duration, we recorded for each participant an interval with the following endpoints: the left endpoint was the computed duration of observed shedding; and the right endpoint was the duration from the last positive in the observed shedding sequence of positive tests to the next negative test or to the next positive test after a long gap, whichever was shorter. If neither of these results occurred, the right endpoint was set to infinity. The intervals thus captured for each participant the underlying time of reversion from test-positive status to post-infection test-negative status for that infection.

Next, these intervals were used to fit parametric survival models (exponential, gamma, lognormal, and Weibull) using standard maximum likelihood methods, with the final model chosen as the one with the minimum AIC. The resulting model characterized the distribution of the duration of shedding. The distribution of the duration between the initial positive results between two consecutive infections was then defined as the sum of the distribution of the duration of shedding and the long gap distribution estimated in the first step above. Its density was then computed as the convolution of the respective densities, and summary measures of interest, mean and quantiles, were computed.

We used bootstrap methods to characterize uncertainty in the estimation of the various quantities computed above (i.e., distributional parameters and summary measures) where we resampled at the participant level to capture potential within-participant dependencies. We implemented the estimation sequence outlined above for each of 5000 bootstrap resampled data sets. Final estimates presented are

means of corresponding bootstrap values for quantities of interest, and 95 percent confidence intervals (CI) were computed as the 2.5th and 97.5th percentiles of these bootstrap distributions.

All analyses were performed in R (v. 4.0.5, R Core Team [2021]).

RESULTS

The final study population comprised 137 participants with ≥2positive NAAT results; 108 (79%) were female, the median age was 44 years (range 20-68), seven (5%) reported having an immunocompromising condition or using immunocompromising medications, and seven (5%) reported previous COVID-19-related hospitalization. Ninety-two (67%) participants worked in nursing homes, 32 (23%) in hospitals, and 13 (9.5%) in other locations. Participants had a mean of 2.9 (standard deviation 1.5) and a median of two positive NAAT results (range 2-11). Seventy-nine had one temporal gap between positive NAAT results, 30 had two gaps, 11 had three gaps, and the remaining 17 had four to 10 gaps.

K-means clustering for the observed gaps between positive tests resulted in two classes being optimal using the "gap statistic", indicating that there were two distinct distributions, short and long gaps, consistent with our assumption that short gaps represented positive NAAT results from the same infection and long gaps represented NAAT results from two infections. Figure 2(A) shows the observed long and short gaps and the estimated short (gamma) and long (lognormal) bootstrap distributions (95% confidence bands) for the two groups. The 99th percentile point of the estimated short gap distribution was 31.2 (95% CI 27.1, 35.9) days, used as the cutoff of "short gaps" when computing shedding durations.

To provide robust estimation of the distribution of single infection shedding duration, we restricted analyses to the 47 participants with ≥6 tests and ≥3 positive results. For these participants, the intervals

from the last positive test of an RNA shedding duration to next-test negative are shown in Figure 2(B). The resulting estimated density (95% pointwise confidence bands) from the bootstrapped lognormal survival analyses is shown in Figure 2(C). The mean of this estimated distribution was 30.1 (95% CI 26.3, 34.5) days, and the median was 27.8 days (95% CI 24.6, 31.1). For shedding events following this distribution, 3.2% (95% CI 0.5, 8.7%) and 0.4% (95% 0.0, 1.6%) were expected to exceed 60 days and 90 days, respectively.

Combining the estimated long gap duration (shown in gray in Figure 2(A)) with the estimated duration of shedding (Figure 2(C)) gives the estimated density of the duration to reinfection (95% confidence bands) shown in Figure 2(D). For this distribution, the mean was 89.1 (95% CI 75.3, 103.5) days, and the median was 86.4 (95% CI 72.8, 100.9) days. For reinfection events following this distribution, 8.5% (95% CI 1.01, 23.7%) and 57.0% (95% CI 31.1, 79.1%) were expected to be shorter than 60 days and 90 days, respectively.

Of the 137 participants included in the analysis, 11 had NAAT test result patterns that indicated reinfection. The durations between their infections are shown in Figure 2(D), and they range from approximately the 5th (57 days) to the 90th (114 days) percentiles of this distribution. Of note is that among these 11 participants, each was tested ≥6 times (range 6-14; median 10), and all had multiple negative tests during the observed long gaps (Supplement, Figure 2).

DISCUSSION

Our analysis of this statewide cohort indicated a mean 30.1-day duration of shedding of SARS-CoV-2 nucleic acid, with expected very few durations exceeding 60 (3.2%) or 90 days (0.4%). These findings are similar to two studies of hospitalized patients demonstrating median shedding durations of 20 days among inpatients and 31 days among severely ill hospitalized patients . Together, these results indicate that a 90-day interval between positive NAAT results will reliably differentiate reinfection from

prolonged RNA shedding from a single infection, which supports current CDC recommendations for investigations for suspected reinfection . However, because immunocompromised persons may have extended periods of viral RNA shedding , our results cannot be generalized to that population. Of the three study participants who reported taking immunosuppressing medications, one had the apparent longest duration of single-infection shedding (78 days), while that of the other two were roughly in the middle of the distribution of the other participants' single-infection shedding duration. Thus, including these participants was unlikely to bias the duration estimates, while providing a robust element to the analysis and its generalizability.

In our cohort with an approximate 7-month observation period, the mean estimated interval between initial positive NAAT results of subsequent infections was 89 days, indicating that SARS-CoV-2 reinfections can occur soon after an initial infection. This finding is consistent with a study of inpatient or emergency department encounters that found that reinfections occurred after a mean of 116 days when reinfection was defined as two positive SARS-CoV-2 NAAT results $\geq$90 days apart . While our data indicate that reinfection can occur relatively quickly, the observed average time to reinfection in other populations will vary according to the force of infection and length of the follow-up period.

Negative NAAT results interspersed between positive test results occurred commonly (Supplementary Figure 2) and did not necessarily demark reinfection episodes, as many occurred within short-gap intervals. Negative results during a single infection episode could have resulted from intermittent RNA shedding or false negative test results. In a meta-analysis of 123 persons reported in 56 publications, the mean time from an initial positive NAAT result to the first positive NAAT result from a subsequent infection was only 47.9 days; however, only a single negative NAAT result between two positive NAAT results was required to define reinfection . Thus, the meta-analysis likely underestimated the time to reinfection.

This study has several limitations. False positive or false negative NAAT results could have occurred. Our study occurred early in the COVID-19 pandemic and thus the potential influence of viral variants could not be assessed, and this could alter the potential for reinfection and so of the timing between infections. We also could not assess the clinical significance of reinfections. Finally, the study population was limited to particular occupations and did not include children.

CONCLUSION

Our data can inform decisions regarding both the duration of shedding and the interval between positive NAAT results used to define reinfection. Our results indicate that a 90-day period between positive NAAT results to define reinfection among immunocompetent persons will misclassify very few persons with a single infection-related shedding as being reinfected. Conversely, our data indicate that extending this period will increase the number of reinfections misclassified as single infections.

ACKNOWLEDGMENTS

The authors acknowledge Kushal Modi for his assistance in creating the study database. We thank Maryanne Ingratta, Susan Lukacs, Lisa Mackey, Samira Sami, and Nga Vuong for their many contributions to the original serology survey and continued support. The authors report no conflicts of interest. There were no financial sources of support for this study.

REFERENCES

FIGURE LEGENDS

Figure 1. Illustration of the gap analysis for a single participant. The gaps between positive nucleic acid amplification test results were assigned to two classes—short and long gaps—using k-means clustering. Shedding duration was computed using the sum of the short gaps with an adjustment to account for the possibility that shedding may have extended beyond the date of the last positive test. Methods to make this adjustment considered the interval from the last positive test of a shedding event to the next negative test result or to the next positive test after a long gap interval, whichever was shorter.

Figure 2. (A) Estimated densities (solid lines; 95% pointwise confidence bands as dashed lines) for the short (black) and long (gray) gap distributions. Original gap data are shown as the line segments at the bottom of the graph, colored by k-means grouping (260 gaps for 137 participants). (B) Last positive to next-test negative interval data for the 47 participants with ≥6 test and ≥2 positive results used in the estimation of the shedding duration. All observations are interval censored, except those that extend to the right boundary of the graph, indicated by arrowheads, which are right censored. (C) Estimated density (95% pointwise confidence bands) of shedding duration. (D) Estimated density (95% pointwise confidence bands) of duration to reinfection. Shown as tick marks just above the x-axis are the durations between infections for the 11 participants whose test result pattern would be classified as indicating reinfection based on the definitions derived in the analysis.

Figure 1

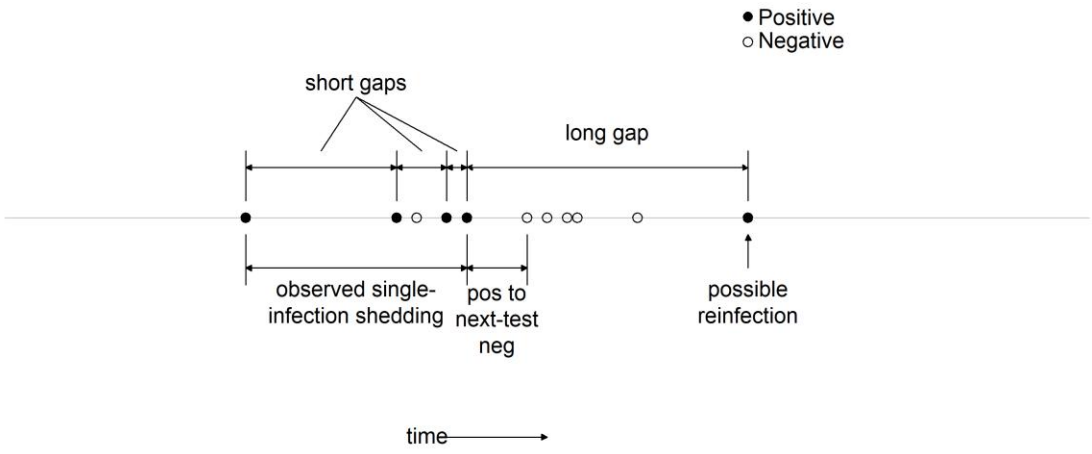

Figure 2