Respiratory disease transmission is highly context dependent and difficult to quantify or predict at the individual level. This is especially the case when transmission from pre-symptomatic, asymptomatic, and mildly-symptomatic (PAMS) subjects is frequent, as with SARS-CoV-2 (1–8).

Transmission is therefore typically inferred from population-level information and summarized as a single overall average, known as the basic reproductive number, R0. While R0 is an essential and critical parameter for understanding and managing population-level disease dynamics, it is a resultant, downstream characterization of transmission. With regard to SARS-CoV-2, many finer-grained upstream questions regarding infectiousness remain unresolved or unaddressed. Three categories of uncertainty are 1) differences in infectiousness among individuals or groups such as PAMS subjects, according to age, gender, vaccination status, etc., 2) timing and degree of peak infectiousness, timing of loss of infectiousness, rates of infectiousness increase and decrease, and how these relate to onset of symptoms (when present), and 3) differences in infectiousness due to inherent properties of virus variants.

These interrelated issues can all be addressed via the combined study of two clinical virological parameters: the viral load (viral RNA concentration) in patient samples and virus isolation success in cell culture trials. While viral load and cell culture infectivity cannot be translated directly to in vivo infectiousness, and the impact of social context and behavior on transmission is very high, these quantifiable parameters can generally be expected to be those most closely associated with transmission likelihood. A strong relationship between SARS-CoV-2 viral load and transmission has been reported (9), comparing favorably with the situation with influenza virus, where the association is less clear (10, 11).
The emergence of more transmissible SARS-CoV-2 variants, such as the B.1.1.7 lineage (UK variant of concern, 202012/01), emphasizes the importance of correlates of shedding and transmission. The scarcity of viral load data in those with recent variants and PAMS subjects of all ages (12) is a blind spot of key importance because many outbreaks have clearly been triggered and fuelled by these subjects (2, 13–17). Viral load data from PAMS cases are rarely available, greatly reducing the number of studies with information from both symptomatic and PAMS subjects and that span the course of infections (12, 18). Making matters worse, it is not possible to place positive RT-PCR results from asymptomatic subjects in time relative to a nonexistent day of symptom onset, so these cases cannot be included in studies focused on incubation period. Additionally, viral load time courses relative to the day of symptom onset rely on patient recall, a suboptimal measure subject to human error and which overlooks infections from presymptomatic or asymptomatic contacts (12). An alternative and more fundamental parameter, the day of peak viral load, can be estimated from dated viral load time series data, drawn from the entire period of viral load rise and fall and the full range of symptomatic statuses.

To better understand SARS-CoV-2 infectiousness we analyzed viral load, cell culture isolation, and genome sequencing data from a diagnostic laboratory in Berlin (Charité – Universitätsmedizin Berlin Institute of Virology and Labor Berlin). We first address a set of questions regarding infectiousness at the moment of disease detection, especially in PAMS subjects whose infections were detected at walk-in community test centres. Because these people are circulating in the general community prior to the detection of their infections, and are healthy enough to present at such centres, their prevalence and shedding are of key importance to the understanding and prevention of transmission. As well as PAMS subjects, we consider the infectiousness suggested by first-positive tests from hospitalised patients, and differences according to age, virus variant, and gender. A further set of temporal questions are then addressed by studying how infectiousness changes during the infection course. Using viral load measurements from patients with at least three RT-PCR tests, we estimate the onset of infectious viral shedding, peak viral load, and the rates of viral load increase and decline. Knowledge of these parameters enables fundamental comparisons between groups of subjects and between virus strains, and highlights the misleading impression created by viral loads from first-positive RT-PCR tests if the time of testing in the infection course is not considered.

**Study composition**

We examined 936,423 SARS-CoV-2 routine diagnostic RT-PCR results from 415,935 subjects aged 0-100 years from February 24, 2020 to April 2, 2021. Samples were collected at test centres and medical practices mostly in and around Berlin, Germany, and analyzed with LightCycler 480 and cobas 6800/8800 systems from Roche. Of all tested subjects, 25,381 (6.1%) had at least one positive RT-PCR test (Table 1). Positive subjects had a mean age of 51.7 years with high standard deviation (sd) of 22.7 years, and a mean of 4.5 RT-PCR tests (sd 5.7), of which 1.7 (sd 1.4) were positive. Of the positive subjects, 4344 had tests on at least three days (with at least two tests positive), and were included in a time series analysis.

We divided the 25,381 positive subjects into three groups (Fig. 1). Hospitalised: 9519 (37.5%) subjects, includes all those who tested positive in an in-patient hospitalised context at any point in their infection; PAMS: 6110 (24.1%) subjects whose first positive sample was obtained in any of 24 Berlin COVID-19 walk-in community test centres, provided they were not in the Hospitalised category; and Other: 9752 (38.4%) subjects not in the first two categories (table S1). As Fig. 1 shows, there were very few elderly PAMS subjects, and relatively low numbers of young subjects in all three groups. The validity of the PAMS classification is supported by the fact that of the overall 6159 infections detected at walk-in test centres, only 49 (0.8%) subjects were later hospitalised. Subjects testing positive at these centres are almost certainly receiving their first positive test, because they are instructed to immediately self-isolate and our data confirms that such subjects are rarely re-tested: only 4.6% of people with at least three test results had their first test at a walk-in test center. Of the 9519 subjects who were ever hospitalised, 6835 were already in hospital at the time of their first positive test. PAMS subjects had a mean age of 38.0 years (sd 13.7), typically younger than Other subjects (mean 49.1 years, sd 23.5), with Hospitalised the oldest group (mean 63.2 years, sd 20.7). Typing RT-PCR indicated that 1533 subjects were infected with a strain belonging to the B.1.1.7 lineage, as confirmed by full genomes from next-generation sequencing (see materials and methods).

**First-positive viral load**

Across all subjects, the mean viral load (herein given as \(\log_{10}\) RNA copies per swab) in the first positive-testing sample was 6.39 (sd 1.83). PAMS subjects had viral loads higher than the Hospitalised for ages up to 70 years, as exemplified by a 6.9 mean for PAMS compared to a 6.0 mean in Hospitalised adult subjects of 20-65 years. Crude comparisons of viral loads in age groups show no substantial difference in first-positive viral load between groups of people aged over 20 years (Table 1), and that children and adolescents have mean first-positive viral loads differences ranging between -0.49 (-0.69, -0.29) and -0.16 (-0.31, -0.01) compared to

**Page numbers not final at time of first release**

First release: 25 May 2021

www.sciencemag.org

Downloaded from http://science.sciencemag.org on June 21, 2021
adults aged 20-65 (Table 2). Here and below, parameter differences between age groups show the younger value minus the older, so a negative difference indicates a lower value in the younger group. Ranges given in parentheses are 90% credible intervals.

We used a Bayesian thin-plate spline regression to estimate the relationship between age, clinical status, and viral load from the first positive RT-PCR of each subject, adjusting for gender, type of test center, and PCR system used. The Bayesian model well represents the observed data (Fig. 1B, Table 2, and fig. S1). The raw data and the Bayesian estimation (Fig. 2A), suggest considering subjects in three age categories: young (ages 0-20 years, grouped into five-year brackets), adult (20-65 years), and elderly (over 65 years). We estimated an average first-positive viral load of 6.40 (6.37, 6.42) for adults and a similar mean of 6.35 (6.32, 6.39) for the elderly (Fig. 2A). Younger age groups had lower mean viral loads than adults, with the difference falling steadily from -0.50 (-0.62, -0.37) for the very youngest (0-5 years) to -0.18 (-0.23, -0.12) for older adolescents (15-20 years) (Table 2). Young age groups of PAMS subjects have lower estimated viral loads than older PAMS subjects, with differences ranging from -0.18 (-0.29, -0.07) to -0.63 (-0.96, -0.32). Among Hospitalised subjects these differences are smaller, ranging from -0.18 (-0.45, 0.07) to -0.11 (-0.22, 0.01) (Table 2 and Fig. 2B). Viral loads of subjects younger than 65 years were around 0.75 higher for PAMS than for Hospitalised subjects (Fig. 2A), likely due to a systematic difference in RT-PCR test timing, discussed below.

**Associating viral load with cell culture infectivity**
We estimated the association between viral load and successful cell culture isolation probability (hereafter “culture probability”) by combining the Bayesian regression estimation with cell culture isolation data from our own laboratory (19) and from Perera et al. (20) (Fig. 2C). Across all ages, the average estimated culture probability at the time of first positive RT-PCR was 0.35 (0.01, 0.94). The mean culture probability is higher for PAMS cases, at 0.44 (0.01, 0.98), than Hospitalised cases, at 0.32 (0.00, 0.92) (Fig. 2D). Comparing PAMS cases, we found differences, in particular for children aged 0-5 compared to adults aged 20-65, with average culture probabilities of 0.329 (0.003, 0.950) and 0.441 (0.008, 0.981) respectively, and a difference of -0.112 (-0.279, -0.003). Age group differences in Hospitalised cases range from -0.028 (-0.104, 0.009) to -0.018 (-0.055, 0) (Table 2).

First-positive viral loads are weakly bimodally distributed (Figs. 1A and 2A), which is not reflected in age-specific means. The resultant distribution of culture probability includes a majority of subjects with relatively low, and a minority with very high culture probability (Fig. 2E and fig. S2). The highly-infectious subset includes 2228 of 25,381 positive subjects (8.78%) with a first-positive viral load of at least 9.0 log_{10}, corresponding to an estimated culture probability of -0.92 to 1.0. Of these 2228 subjects, 804 (36.09%) were PAMS at the time of testing, with a mean (median) age of 37.6 (34.0) and sd of 13.4 years. PAMS subjects are over-represented in this highly-infectious group among those aged 20-80 years, and Hospitalised subjects are over-represented in those aged 80-100 years (fig. S3).

**Estimating B.1.1.7 infectiousness at first-positive test**
The 1533 subjects infected with a B.1.1.7 virus in our dataset had an observed mean first-positive viral load of 7.38 (sd 1.54), which is 1.05 log_{10} higher (0.97, 1.13) than non-B.1.1.7 subjects in the full dataset. To increase specificity, we compared 1453 B.1.1.7 cases with 977 non-B.1.1.7 cases using viral loads only from centres with B.1.1.7 and non-B.1.1.7 cases, and only from the same day or one day before or after the B.1.1.7 sample was taken. This analysis adjusted for clinical status, gender, RT-PCR system, subject age, and also modeled random test center effects. The results show that B.1.1.7 cases are associated with a 1.0 (0.9, 1.1) higher viral load (Fig. 3 and table S2). This results in a mean estimated B.1.1.7 subject culture probability of 0.50 (0.03, 0.97), considerably higher than the overall figure of 0.31 (0.00, 0.94) for the non-B.1.1.7 subjects in the comparison, corresponding to a median 2.6 (50% credible interval: 1.4, 5.1) times higher culture probability for samples from B.1.1.7 cases. To investigate whether there might be a difference in cell culture infectivity due to a factor other than viral load, we isolated virus from 105 samples (22 B.1.1.7, 83 B.1.177) in Caco-2 cells from a collection of 223 samples with matched viral loads. While no statistical difference was seen in the distribution of viral loads that resulted in successful isolation (fig. S4), uncertainty due to the routine diagnostic laboratory context, including uncontrolled pre-analytical parameters such as transportation time and temperature, together with the small isolation-positive sample sizes are insufficient to support a conclusion that the distributions do not differ (see materials and methods).

**Estimating infectiousness over time**
To investigate viral load over the course of the infection, we estimated the slopes of a model of linear increase and then decline of log_{10} viral load using a Bayesian hierarchical model. The analysis used the time series of the 4344 subjects who had RT-PCR results on at least three days (with at least two tests being positive). The number of subjects with multiple test results skewed heavily toward older subjects, with very few below the age of 20 meeting the criterion (Fig. 4A). We estimated time from onset of shedding to peak viral load of 4.31 (4.04, 4.60) days, mean peak viral load of 8.1 (8.0, 8.3), and mean decreasing viral load slope of -0.168
(-0.171, -0.165) log_{10} per day (fig. S5). Figure S6 shows that while Hospitalised patients are estimated to be uniformly highly infectious at peak viral load, the infectiousness of PAMS subjects at peak load is more variable.

The temporal placement of the full 18,136 RT-PCR results from these 4344 subjects (80% of whom were hospitalised with COVID-19 at some point in their infections) is shown in fig. S7. Per-subject trajectories can differ considerably from that described by the mean parameters (Fig. 4B and fig. S8). Across all subjects, PAMS cases were on average detected 5.1 (4.5, 5.7) days after peak load, 2.4 (1.7, 3.0) days before non-PAMS cases, which were on average detected 7.4 (7.2, 7.6) days after peak load. We estimate that 962 (914, 1010) of the 4344 subjects (22.14% (21.04, 23.25)) had a first positive test before the time of their peak viral load, with a mean of 1.4 (1.3, 1.5) days before reaching peak viral load. Among the infections detected after peak viral load, the timing of the first positive RT-PCR test is estimated at 9.8 (9.6, 10.0) days after peak viral load, with sd of 6.9 (6.8, 7.0) days, reflecting a broad time range of infection detection. Estimated peak viral loads were higher in Hospitalised subjects than Other, and higher in Other than PAMS, with differences of 0.68 (0.83, 0.52) and 0.96 (0.33, 1.53) respectively (fig. S9 and table S3). No differences were seen according to gender. Viral load time courses are similar across age groups, though younger subjects have lower peak viral load than adults aged 45-55 (Fig. 5, A and C, fig. S10, and table S4). Model parameters suggest slightly longer time to peak, higher peak, and more rapid decline in viral load when the analysis is restricted to subjects with successively higher numbers of RT-PCR results (fig. S11 and table S5), with an increasing percentage of hospitalised subjects. Differences in model parameters according to the number of tests in subjects may reflect increased parameter accuracy due to additional data, though other factors associated with being tested more frequently may be responsible. The Bayesian estimation of the model agrees well with a separate second implementation based on simulated annealing (fig. S12, table S5, and supplementary text).

We estimate that the rise from near-zero to peak culture probability takes 1.8 (1.3, 2.6) days, with a mean peak culture probability of 0.74 (0.61, 0.85). Mean culture probability then declines to 0.52 (0.40, 0.64) at five days and to 0.29 (0.19, 0.40) at ten days after peak viral load. Subject-level time courses can deviate substantially from these mean estimates (Fig. 4C). Peak culture probabilities for age groups range from a low of 0.54 (0.39, 0.71) for 0-5 year olds to 0.80 (0.67, 0.90) for subjects over 65 years. The least infectious youngest children have 78% (61, 94) of the peak culture probability of adults aged 45 to 55 (Fig. 5, B and D, and table S4). Insufficient data precludes a reliable B.1.1.7 viral load time series analysis at this point.

Discussion

Limitations

Our analysis attempted to account for effects of gender, PCR system, and test center type. Although we could not incorporate inter-run variability or the variability in the sample pre-analytic, such as type of swab or initial sample volume in our conversion of RT-PCR cycle threshold values to log_{10} viral load values, these variabilities apply to all age groups and do not affect the interpretation of data for the purpose of the present study. If the proportion of subjects with a certain clinical status differs between age groups in the study sample, this could lead to over- or underestimation of differences in viral load between age groups. However, as our study compares viral load between age groups stratified by clinical status, it appears unlikely that differential testing biases our results.

Interpreting first-positive viral loads

Viral loads and their differences are not easy to interpret, absent knowledge of when in the disease course the samples were taken and the correspondence between viral load and shedding. The higher first-positive viral loads in PAMS subjects than Hospitalised subjects are likely due to time of detection. This is suggested in the first place by the estimated 2.4 (1.7, 3.0) day difference in test timing, which would produce a viral load difference of ~0.4 using the -0.168 daily viral load decline gradient from the (mainly hospitalised) time series subjects. Additionally, the time series of PAMS, Other, and Hospitalised subjects estimates that, throughout the infection course, the Hospitalised group have higher viral loads than Other, who are in turn higher than PAMS (fig. S9 and table S3). This relationship holds across age groups (fig. S13) and also in a fine-grained split of test centres by clinical severity (fig. S14). Similarly, the lower first positive viral loads in elderly PAMS subjects may be due to these subjects being less likely to be tested as early due to being more likely to be house-bound, less likely to be employed, less mobile, more cautious and inclined to get tested with only mild symptoms, etc. The impact on infectiousness of differences in viral load must be informed by where the viral loads fall on the viral load / infectivity curve. In our data, the viral loads involved in the difference between the means in children and adults and the difference between means in B.1.1.7 and non-B.1.1.7 subjects result in quite different corresponding culture probabilities (see below).

A highly-infectious minority and over-dispersion

The bimodal distribution of culture probabilities (Fig. 2, D and E) shows a small group of 8.78% of highly-infectious subjects. This qualitatively agrees with a model (21) and a study (22) concluding that 10% and 15% of index cases, respectively, may be responsible for 80% of transmission. Oth-
er studies reported that 8-9% of individuals harboured 90% of total viral load (23), that in cases from India (24) and Hong Kong (6) ~70% of index cases had no secondary cases. The risk posed by PAMS subjects is highlighted by the fact that 36.1% of the highly-infectious subjects in our study were PAMS at the time of the detection of their infection, that their mean age was 37.6 years with a high standard deviation of 13.4 years (figs. S2 and S3), and our estimate that infectiousness peaks 1-3 days before onset of symptoms (if any).

**Comparison with influenza virus**

Absent direct knowledge from a large number of SARS-CoV-2 transmission events, we could try to draw conclusions regarding infectiousness from studies of other respiratory viruses, such as influenza. However, it has become clear that there are important differences and uncertainties that would cast doubt on such a comparison. Influenza may have later onset of viral shedding, shedding finishes earlier, there may be a lower secondary attack rate, viral loads are much lower, there is variation between virus subtypes, the role of asymptomatic subjects in transmission is uncertain or thought to be reduced, and the frequency of asymptomatic infections is uncertain, especially in children (10, 11, 25–29). Age-specific behavioral differences do however make a large contribution to the established higher shedding of children compared to adults in influenza. This should be an important consideration for SARS-CoV-2, as shown by studies indicating higher transmission between children of similar ages (6, 24) and high transmission heterogeneity (22). Despite many decades of close study of influenza virus, the relationship between viral load and transmission is unclear (10, 11). The situation with respiratory syncytial virus is even less clear (30). Understanding SARS-CoV-2 transmission will likely be at least as challenging, given the high frequency of transmission from PAMS subjects (1–8), suggesting an important role for clinical parameters, given the apparently strong association between viral load and transmission, independent of symptoms (9).

**Estimated infectiousness in the young**

The differences we observe in first-positive RT-PCR viral load between groups based on age are minor, as in other studies (31–35) and the viral loads in question, in the range of 5.9 to 6.6 (Table 1), are in a region of the viral load / culture probability association where changes in viral load have relatively little impact on estimated culture probability (Fig. 2C). Comparisons between adult viral loads and those of children and the relative infectious risks they pose are difficult due to the likely influence of non-viral factors. Nasopharyngeal swab samples, which often carry higher viral loads, are rarely taken from young children due to pain and lack of cooperation, and the sample volume carried by smaller pediatric swab devices is lower than in larger swabs used for adults (36). Infections in mildly-symptomatic children may be initially missed and only detected later (37), resulting in lower first-positive viral loads. Our results of similar viral load trajectories for children and adults (Fig. 5), and the numeric range of the viral load values in question (Fig. 2C), suggest that viral load differences between children and adults are too small to alone produce large differences in infectiousness. The relative impact on transmission of general age-related physiological differences, such as different innate immune responses (38), may be small as compared to the impact of large differences in frequency of close contacts and transmission opportunities.

**Timing of estimated peak infectiousness relative to onset of symptoms**

We estimated the time from onset of shedding to peak viral load at 4.3 days. Previous studies and reviews of COVID-19 report mean incubation times of 4.8 to 6.7 days (4, 39–44), which suggests that, on average, a period of high infectivity can start several days before symptoms onset. Viral load rise may vary between individuals, and limitations of the available data suggest that our analysis may underestimate inter-individual variation in viral load increase. The failure to isolate virus in cell culture beyond 10 days from symptom onset (19, 20, 35, 45, 46) together with our estimated slope of viral load decline also suggests peak viral load occurs 1-3 days before symptom onset (supplementary text). Data from 171 hospitalised patients from a Charité – Universitätsmedizin cohort suggest a figure of 4.3 days (fig. S15 and supplementary text).

**Estimated infectiousness of the B.1.1.7 variant**

We found an approximately 1 log\(_{10}\) higher first-positive viral load in people infected with a B.1.1.7 virus than people infected with a wild-type. The scale of the viral load difference and the fact that it is also present in the comparison between B.1.1.7 infected subjects and non-B.1.1.7 infected subjects drawn from the same test centres at the same times, argue that the difference is not due to a systematic difference in time of sampling. The 1 log\(_{10}\) higher B.1.1.7 viral load can be compared to implied 5-10x higher B.1.1.7 viral loads in two large and closely-controlled UK studies, a vaccine trial (47) and a mortality study (48), based on RT-PCR cycle threshold differences of ~3 and 2.3 respectively. Several other studies also appear to point to a higher B.1.1.7 viral load (49–52) (supplementary text). Importantly, the mean B.1.1.7 viral load value in our study falls in a region of the viral load / culture probability curve with steep gradient (Fig. 2C), resulting in an estimated culture probability considerably higher than for non-B.1.1.7 subjects. Although a strong
correlation has been observed between SARS-CoV-2 viral load and transmission (9), here we are estimating infectivity probability from cell culture trials. Any impact of a change in viral load on transmission will be highly dependent on context, so the large difference in estimated culture probability in our data is only a proxy indication of potentially higher transmissibility of the B.1.1.7 strain. We estimate B.1.1.7 infected subjects having a 2.6 times higher mean culture probability than non-B.1.1.7 infected subjects. This range can be compared to a UK study that found a 1.3 relative increase in secondary attack rates for B.1.1.7 index cases in ~60,000 household contacts (53), a UK study estimating a 1.7 to 1.8 increase in transmission (54), and to an estimate of a 43% to 90% higher reproductive number (55).

**Summary**

Our results indicate that PAMS subjects in apparently-healthy groups can be expected to be as infectious as hospitalised patients at the time of detection. The relative levels of expected infectious virus shedding of PAMS subjects (including children) is of high importance because these people are circulating in the community and it is clear that they can trigger and fuel outbreaks (56). The results from our time series analysis, and their generally good agreement with results from studies based on other metrics (often epidemiological), show that accurate estimations can be directly obtained from two easily-measured virological parameters, viral load and sample cell culture infectivity. Such results can be put to many uses: to estimate transmission risk from different groups (by age, gender, clinical status, etc.), quantify variance, show differences in virus variants, highlight and quantify over-dispersion, and to inform quarantine, containment, and elimination strategies. Our understanding of the timing and magnitude of change in viral load and infectiousness, including the impact of influencing factors, will continue to improve as data from large studies accumulate and are analyzed. A major ongoing challenge is to connect what we learn about estimated infectiousness from these clinical parameters to highly context-dependent in vivo transmission. Based on our estimates of infectiousness of PAMS subjects and the higher viral load found in subjects infected with the B.1.1.7 variant, we can safely assume that non-pharmaceutical interventions such as social distancing and mask wearing have been key in preventing many additional outbreaks. Such measures should be employed in all social settings and across all age groups, wherever the virus is present.

**Materials and methods**

**Age ranges**

Age categories for the analysis of the first-positive test results mentioned in the text indicate mathematically open-closed ranges of years (e.g., 0-5 signifies (0-5] years). We group subjects up to 20 years old into age categories spanning five years, subjects from 20 to 65 years into an adult group, and elderly subjects into a 65+ category. This categorisation is motivated by the observed data and the Bayesian estimation of viral load differences between children of different ages and adults. The age groupings used in the viral load time series analysis are broader in the younger categories to increase the cardinality of those groups, due to the fact that few young people have at least three RT-PCR tests (Fig. 4A).

**Viral loads**

Viral load is semiquantitative, estimating RNA copies per entire swab sample, while only a fraction of the volume can reach the test tube. The quantification is based on a standard preparation tested in multiple diluted replicates to generate a standard curve and derive a formula upon which RT-PCR cycle threshold values are converted to viral loads. This approach does not reflect inter-run variability or the variability in the sample pre-analytic, such as type of swab or initial sample volume (varying between 2.0 and 4.3 mL). However, these variabilities apply to all age groups and do not affect the interpretation of data for the purpose of the present study.

Viral load figures are given as the logarithm base 10. Viral load is estimated from the cycle threshold (Ct) value using the empirical formulae 14.159 - (Ct * 0.297) for the Roche Light Cycler 480 system and 15.043 - (Ct * 0.296) for the Roche cobas 6800/8800 systems. The formulae are derived from testing standard curves and cannot be transferred to calculate viral load in other laboratory settings. Calibration of the systems and chemistries in actual use is required.

**B.1.1.7 viral load analysis**

No assignment regarding symptomatic status was made for B.1.1.7 subjects due to uncertainties regarding exact operational protocols at outbreak hospitals. B.1.1.7 assignment to samples was initially made according to typing-RT-PCR tests that detect the N501Y and 69/70 deletion in the amino acid sequence of the virus spike protein. Examination of the complete viral genome of 49 samples confirmed that the subjects were in fact infected with the B.1.1.7 variant, with all variant-defining substitutions and deletions (57) found in all cases. No consistent additional mutations or deletions/insertions were found in the sequences.

Sequencing read mapping was performed with Bowtie, with alignment using MAFFT, and visual inspection using Geneious Prime (all version numbers given below). For the statistical comparison of B.1.1.7 and non-B.1.1.7 subjects, we identified test centres (hospital departments or wards, or
organisations outside hospitals) that reported B.1.1.7 cases, and chose as comparison groups non-B.1.1.7 cases that were detected in these test centres on the same day or one day earlier or later. By modeling random effects for test centres, we estimate the expected viral load difference as the average of the within-test center differences. The consistent effect of B.1.1.7 throughout a range of comparison scenarios is shown in table S2.

**Sample type**

An estimated 3% of our samples were from the lower respiratory tract. These were not removed from the dataset because of their low frequency and the fact that the first samples for patients are almost universally swab samples. Samples from the lower respiratory tract are generally taken from patients only after intubation, by which point viral loads have typically fallen.

**PAMS status**

Metadata needed to discriminate patients into sub-cohorts based on underlying diseases, outcome, or indications for diagnostic test application, including symptomatic status, were not always available. In the absence of subject-level data, we inferred PAMS status using the type of submitting test center as an indicator, classifying subjects as PAMS at the time of testing if their first-positive sample was taken from a walk-in COVID-19 test center and the subject had no later RT-PCR test done in a hospitalised context (e.g., in a ward or an intensive care unit). The correspondence between viral load and PAMS status derived herein may therefore be less accurate than in studies with subject-level symptom data. However, we make no formal claims regarding symptomatic status, and instead emphasize the fact that these PAMS subjects were healthy enough to be presenting at walk-in COVID-19 test centres, and were therefore capable to some extent, at that time, of circulating in the general community.

**Bayesian analysis of age - viral load associations**

We estimated associations of viral load and age with a thin-plate spline regression using the brms package (58, 59) in R (60). Spline coefficients were allowed to vary between groups determined by the type of the test center and clinical status (PAMS, Hospitalised, or Other), and random intercepts captured effects of test centres. To reduce the impact of outliers we used Student-t distributed error terms. The analysis additionally accounted for baseline differences between subject groups, B.1.1.7 status, gender, and for the effect of the RT-PCR system. We also estimated the association between viral load and culture probability in order to calculate the expected culture probability at different age levels. This analysis used weakly-informative priors and was estimated using four chains with 1000 warm-up samples and 2000 post-warm-up samples. Convergence of MCMC chains was examined by checking that Potential Scale Reduction Factors (R-hat) values were below 1.1. All calculations of age averages and group differences are based on posterior predictions generated from estimated model parameters. Expected probabilities of positive cultures (and their differences) were calculated by applying the posterior distribution of model parameters from the culture probability model to posterior predictions from the age association model.

**Combining culture probability data**

To estimate the association between viral load and culture probability, we used data previously described by Wölfel (19) and Perera (20). Four other data sets could not be included because Ct values were not converted to viral loads (35, 46, 61, 62). The data from the study by van Kampen et al. (63) were not included because they differed (by viral load of ~1.0) from the data used for the current analysis, likely due to a combination of factors including many patients who were in critical or immunocompromised condition, a high proportion of samples obtained from the lower respiratory tract including late in the infectious course, and likely differences in cell culture trials. It is unsurprising that these data result in a shifted viral load / culture probability curve, and we excluded them because our focus was largely on first positive RT-PCR results from the upper respiratory tract, including from many subjects who were PAMS. The Digital Supplement shows the plot of the van Kampen data set compared to the two we used. To calculate the expected culture probability, by age (as in Fig. 2D) or by day from peak viral load (as in Fig. 4C), we combined the viral loads (Figs. 2A and 4B) with the results of the regression of culture probability shown in Fig. 2C. We used posterior predictions from the age regression model, which reflect the variation of log10 viral load within age groups, to estimate culture probabilities by age. For instance, to obtain the culture probability for a specific age and group, we look up the estimated (expected) viral load for that group, add an error term, and, using the association shown in Fig. 2C, and determine the expected culture probability. We used expected time courses, i.e., the model’s best guess for a time course, to estimate culture probability time courses.

**B.1.1.7 isolation data**

The Institute of Virology at Charité – Universitätsmedizin Berlin routinely receives SARS-CoV-2 positive samples for confirmatory testing and sequencing. For this study we used anonymized remainder samples from a large laboratory in northern Germany, that were all stored in phosphate-buffered saline (PBS) and therefore suitable for cell culture...
isolation trials. Sample transport to the originating lab and later to Berlin was unrefrigerated, via road. As part of the routine testing, these samples were classified by typing RT-PCR and complete genome sequencing (64). 113 B.1.1.7 lineage samples and 110 B.1.177 lineage samples were selected, with approximately matched (pre-inoculation) SARS-CoV-2 RNA concentrations. Caco-2 (human colon carcinoma) cell cultures (65) were inoculated twice from each sample, once with undiluted material and once with a 1:10 dilution. The diluted inoculant was used to reduce the probability of culturing failure due to the possible presence of host immune factors (antibodies, cytokines, etc) that might negatively impact isolation success, and to reduce the possibility of other unrelated agents (bacteria, fungi, etc) resulting in cytopathic effect in the culture system. For cell culture isolation trials, 1.6x10^5 cells were seeded per well in a 24-well plate. Cells were inoculated with swab suspensions for one hour at 37°C, subsequently rinsed with PBS, and fed with 1 mL fresh Dulbecco’s modified Eagle's minimum essential medium (DMEM; ThermoFisher Scientific) supplemented with 2% fetal bovine serum (FBS; Gibco), 100 U/mL penicillin, and 100 μg/mL streptomycin (P/S; ThermoFisher Scientific), and 2.5 μg/mL Amphotericin B (biomol) then incubated for five days before harvesting supernatant for RT-PCR testing. Positive cell culture isolation was defined by a minimum 10x higher SARS-CoV-2 RNA load in the supernatant compared to the inoculant and signs of a typical SARS-CoV-2 cytopathic effect. Culture isolation was successful for 22 B.1.1.7 and 61 B.1.177 samples. Due to uncertainty regarding sample handling before arrival at the originating diagnostic laboratory and the unrefrigerated transport, it was not possible to determine whether isolation failures were due to samples containing no infectious particles (due to sample degradation) or for other reasons. Such reasons could include systematic handling differences according to variant type or a difference in virion stability and durability regarding environmental factors such as temperature. Therefore, negative isolation outcome samples were excluded from analysis. The strong likelihood of many cases of complete sample degradation is evident from the isolation failure of many samples with high pre-inoculation viral load, with the viral load in these cases merely indicating the presence of non-infectious SARS-CoV-2 RNA (fig. S4). Given this context, we were reduced to questioning whether there might be a difference in the range of viral loads that were able to result in isolation between B.1.1.7 and non-B.1.17 variants. Such a difference could result from a difference in the ratio of viral RNA to infectious particles produced by the variants, or from a non-viral load difference in the variants. We examined the distribution of pre-inoculation viral loads from isolation-positive samples from both variants for a difference. No statistically significant difference was found, but in the converse, the isolation-positive sample sizes are too low to support the assertion that the distributions do not differ.

**Estimating viral load time course**

Each RT-PCR test in our data set has a date, but no information regarding the suspected date of subject infection or onset of symptoms (if any). Although determining the day of peak viral load for a single person based on a series of dated RT-PCR results would not in general be feasible due to individual variation, with data from a large enough set of people, a clear and consistent model of viral load change over time can be inferred with very few assumptions.

We included a single leading and/or trailing negative RT-PCR result, if dated within seven days of the closest positive RT-PCR. To produce a model of typical viral load decline on a reasonable single-infection timescale we excluded subjects whose full time series contains positive RT-PCRs spread over a period exceeding 30 days. Such time series may be due, for example, to contamination, to later swabbing that picks up residual RNA fragments in tonsillar tissue (66), to re-infection (67–69), or may represent atypical infection courses (such as in immunocompromised or severely ill elderly patients) (70). We excluded data from subjects with an infection delimited by both an initial and a trailing negative test when there was only a single positive RT-PCR result between.

We estimated the slopes for a model of linear increase and then decline of log_{10} viral load. To compensate for the absence of information regarding time of infection, we also estimated the number of days from infection to the first positive test for each participant, to position the observed time series relative to the day of peak viral load. The analysis was implemented in two ways. Initially, simulated annealing was used to find an optimized fit of the parameters, minimizing a least squares error function. Secondly, a Bayesian hierarchical model estimated subject-specific time courses, imputed the viral load assigned to each initial or trailing negative test, and modeled associations of age, gender, clinical status, and RT-PCR system with model parameters. We tested both methods on data subsets ranging from subjects with at least three to at least nine RT-PCR results. The two methods produced results that were in generally good agreement (table S5). The finer-grained Bayesian approach appears more sensitive than the simulated annealing and its results, for subjects with at least three RT-PCR results, are those described in the main text.

*Simulated annealing approach:* A simulated annealing optimization algorithm (71) was used to adjust the time series for each subject slightly earlier or later in time, by amounts drawn from a Normal distribution with mean 0.0 and standard deviation 0.1 days. The error function was the
sum of squares of distances of each viral load from a viral load decline line whose slope was also adjusted as part of the annealing process. In the error calculation, negative test results were assigned a viral load of 2.0, in accordance with our SARS-CoV-2 assay limit of detection and sample dilution (19). The initial slope of the decline line was set to -2.0 and was varied using N(0, 0.01). A second, optional, increase line initialized with a slope of 2.0, adjusted using an N(0, 0.01) random variable, was included in the error computation if the day of a RT-PCR test was moved earlier than day zero (the modeled day of peak viral load). The height of the intercept (i.e., the estimated peak viral load) between the increase line (if any) and the decline line was also allowed to vary randomly (starting value 10.0, varied using N(0, 0.1)). The full time series for each subject was initialised to a vary randomly (starting value 10.0, varied using N(0, 0.1)). A second, optional, increase line (the modeled day of peak viral load). The height of the intercept heights as well as seven different methods for the initial placement of time series. In general, maximum viral load and decline slopes were robust to data subset and initial placement, while fig. S17 shows the 49 participants with the highest R-hat values. While the high R-hat values could be removed by using a mixture approach to model shift for these participants, in light of their low frequency we retained the simpler model to avoid additional complexity. Alternatively, constraining the shift parameter to negative numbers would also improve R-hat values for these subjects, at the cost of the additional assumption that infections are generally not detected weeks after infection.

Sensitivity analysis: In addition to examining the viral load time series of subjects with RT-PCR results on at least three days, we tested both approaches on data from subjects with results from a minimum of four to nine days. Given the degree of temporal viral load variation seen in other studies (18–20, 35, 41, 46, 63, 73, 74), and in our own data, our expectation was that a relatively high minimum number of results might be required before reliable parameter estimates with small variance would be obtained, but this proved not to be the case. The simulated annealing approach was tested with a wide range of initial slopes and intercept heights as well as seven different methods for the initial placement of time series. In general, maximum viral load and decline slopes were robust to data subset and initial time series location, though there was variation in the length of the time to peak viral load, depending on how early in time the time series were initially positioned, the initial slope of the increase line and height of the maximum viral load time course implements the same basic model, and additionally estimates associations of model parameters with covariates age, sex, B.1.1.7 status, and clinical status, estimates subject-level parameters (slope of log_{10} load increase, peak viral load, slope of log_{10} load decrease) as random effects, and accounts for effects of PCR system and test center types with random effects. To estimate the number of days from infection to the first test (henceforth ‘shift’) we constrained the possible shift values from -10 to 20 days and used a uniform prior on the support. In contrast to the other subject-level parameters, we estimated subject-level shifts independently, i.e., without a hierarchical structure. Fig. S7 shows the placement in time of individual viral loads after shifting for subjects with RT-PCR results from at least three days. Model parameters changed gradually when subsets of subjects with an increasing minimum number of RT-PCR results, from three to nine, were examined (fig. S11 and table S5). The viral load assigned to negative test results (which may include viral loads below the level of detection) is estimated with a uniform prior on the support from -Inf to 3 (see also the caption of fig. S7). Using prior predictive distributions we specified (weakly) informative priors for this analysis. This analysis was implemented in Stan (72). Full details and R and Stan code for the Bayesian analysis, as well as comparison of priors and posteriors, are given in the supplementary materials.

Bayesian approach: The Bayesian analysis of viral load time course implements the same basic model, and additionally estimates associations of model parameters with covariates age, sex, B.1.1.7 status, and clinical status, estimates subject-level parameters (slope of log_{10} load increase, peak viral load, slope of log_{10} load decrease) as random effects, and accounts for effects of PCR system and test center types with random effects. To estimate the number of days from infection to the first test (henceforth ‘shift’) we constrained the possible shift values from -10 to 20 days and used a uniform prior on the support. In contrast to the other subject-level parameters, we estimated subject-level shifts independently, i.e., without a hierarchical structure. Fig. S7 shows the placement in time of individual viral loads after shifting for subjects with RT-PCR results from at least three days. Model parameters changed gradually when subsets of subjects with an increasing minimum number of RT-PCR results, from three to nine, were examined (fig. S11 and table S5). The viral load assigned to negative test results (which may include viral loads below the level of detection) is estimated with a uniform prior on the support from -Inf to 3 (see also the caption of fig. S7). Using prior predictive distributions we specified (weakly) informative priors for this analysis. This analysis was implemented in Stan (72). Full details and R and Stan code for the Bayesian analysis, as well as comparison of priors and posteriors, are given in the supplementary materials.

Checking convergence of the model parameters showed that while 99.3% of all parameters converged with an R-hat value below 1.1, some subject-level parameters of 118 subjects (among 4344 subjects with at least 3 RT-PCR results) showed R-hat values between 1.1 and 1.74. Inspection of these parameters showed that these convergence difficulties were due to observed time courses that could arguably be placed equally well at the beginning or a later stage of the infection. Figure S16 shows a set of 81 randomly-selected posterior predictions, to give an impression of time series placement, while fig. S17 shows the 49 participants with the parameters with the highest R-hat values. While the high R-hat values could be removed by using a mixture approach to model shift for these participants, in light of their low frequency we retained the simpler model to avoid additional complexity. Alternatively, constraining the shift parameter to negative numbers would also improve R-hat values for these subjects, at the cost of the additional assumption that infections are generally not detected weeks after infection.
load, etc. This is as expected as the settings of these parameters can be used to bias the probability that a time series is initially positioned early or late in time and how difficult it is for it to subsequently move to the other side of the peak viral load at day zero. Table S5 shows parameter values for both approaches on the various data subsets.

Day of infection: We define the moment of infection as the time point at which the increasing viral load crosses zero of the log_{10} y-axis, i.e., when just one viral particle was estimated to be present. Because the time of infection depends on the estimated peak viral load and the slope with which viral load increases, the data should optimally include multiple pre-peak viral load test results for each individual. If, as in the current data set, only a subset of subjects have test results from pre-peak viral load, a hierarchical modeling approach still allows calculating subject-level estimates. Intuitively, this approach uses data from all subjects to calculate an average slope parameter for increasing viral load. In addition, it models subject-level parameters as varying around the group level parameter. To further refine the estimation of slope parameters the model also uses the covariates age (see fig. S10), gender, and clinical status. Because negative test results could be false negatives, viral loads for these tests are imputed (with an upper bound of 3). Subject-level peak viral load and declining slope are modeled with the same approach. More generally, using a hierarchical model and shrinkage priors for covariates effects results in more accurate predictions in terms of estimation of slope parameters the model also uses the covariates age (see fig. S10), gender, and clinical status. This is especially the case if, as in the current data set, subject-level data are sparse.

Onset of symptoms: The 317 onset of symptoms dates for hospitalised patients were collected as part of the Project-level data are sparse. This is especially the case if, as in the current data set, sub-Charité – Universitätsmedizin Berlin (version 0.5.0) (numpy (1.18.3) (80), seaborn_sinaplot (version 0.11.1) (81), numpys (1.18.3) (82), seaborn_sinaplot (version 0.5.0) (71), and seaborn (version 0.10.1) (84). Sequence analysis used Bowtie2 (2.4.1) (85), bcftools and samtools (1.9) (86, 87), Geneious Prime (2021.0.3) (88), ivar (1.2.2) (89), and MAFFT (4.475) (90). Analyses in R (4.0.2) (60) were conducted using the following main packages: brms (2.13.9) (58, 59), rstanarm (2.21.1) (91), rstan (2.21.2) (92), data.table (1.13.3) (93), and ggplot2 (3.3.2) (94). Bayesian analysis in R was based on Stan (2.25) (72). Parallel execution was performed with GNU Parallel (20201122 (‘Biden’) (95)).

Data curation and anonymisation

Research clearance for the use of routine data from anonymized subjects is provided under paragraph 25 of the Berlin Landeskrankenhausgesetz. All data are anonymized before processing to ensure that it is not possible to infer patient identity from any processing result. All patient information is securely combined into a token that is then replaced with a value from a strong one-way hash function prior to the distribution of data for analysis. Viral loads are calculated from RT-PCR cycle threshold values that have only one decimal place of precision.

REFERENCES AND NOTES


95. O. Tange, GNU Parallel 20201122 ('Biden') (2020); www.gnu.org/software/parallel/.


97. Additional statistical information and the R code and data to reproduce the results, figures, and tables are available at https://doi.org/10.5281/zenodo.4774226.


ACKNOWLEDGMENTS

Computation has been performed on the HPC for Research/Clinic cluster of the Berlin Institute of Health, supported by Dieter Beule, Manuel Holtgrewe, and Oliver Stolpe. Thanks to Udo Gieraths and Leonie Meiners for careful commentary on the manuscript, to the Charité – Universitätsmedizin P. COVID-19 collaborative study group for providing additional onset of symptoms data, and to Stephen Kessler for providing additional details regarding their NPA study. The conditions allowing the work to be done with no need for consent are given at https://gesetze.berlin.de/bsbe/document/jlr-KHGBE2011V4P25. Funding: Work at Charité – Universitätsmedizin Institute of Virology is funded by European Commission via project ReCoVer. German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) through projects 5ZIF (301-4-7-01.703) to CD, VARIPadh (01K2021) to VMC, PROVID (FKZ 01K20160C) to CD, VMC, and LES. and (NaFoUniMedCovid19 (NUM) – COVIM, FKZ: 01K2021) to CD, VMC, and LES. The Pa-COVID 19 Study is supported by grants from the Berlin Institute of Health (BIH). This study was supported in part by the German Ministry of Health (Konsilariinlb für Coronaviren and SeCoV) to CD and VMC. TCJ is in part funded through NIAID-NIH CEIRS contract HHSN272201400008C. Author contributions: TCJ, GB, BM. Bioinformatic processing, statistical analysis, interpretation of results, writing original draft and final text: TV. Statistical analysis, interpretation of results, writing original draft and final text, next-generation sequencing: JS, JBS, TB, JT, MLS. sample preparation, virus isolation and culturing. RT-PCR, next-generation sequencing: LES, FK: collection of symptom onset data; PM, RS, MZ, JH, AK, AS, AE: diagnostic work and collection of raw data; VMC: diagnostic data collection, viral load calibration, supervision of laboratory work, interpretation of results; CD: project concept, interpretation of results, writing original draft and final text. Competing interests: Authors declare that they have no competing interests. Data and materials availability: Additional statistical information and the R code and data to reproduce the results, figures, and tables are available (97). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.eabi5273/DC1
Supplementary Text Figs. S1 to S5
Tables S1 to S5
References (98–102)
MDAR Reproducibility Checklist
15 March 2021; accepted 21 May 2021
Published online 25 May 2021
10.1126/science.abi5273
Table 1. Age stratification of first-positive RT-PCR tests and viral load. Column name abbreviations are as follows: N: number of subjects with a positive test result; Pos. %: percentage of positive subjects; Load (sd): mean log_{10} viral load and standard deviation. \geq 3 tests: number of subjects with at least three RT-PCR test results, as used in the viral load time course analysis. Age ranges (in years) are open-closed intervals.

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>Pos. %</th>
<th>Load (sd)</th>
<th>\geq 3 tests</th>
<th>N</th>
<th>Pos. %</th>
<th>Load (sd)</th>
<th>N</th>
<th>Pos. %</th>
<th>Load (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>330</td>
<td>1.8</td>
<td>5.9 (1.84)</td>
<td>16</td>
<td>36</td>
<td>5.1</td>
<td>6.6 (1.87)</td>
<td>32</td>
<td>0.9</td>
<td>5.6 (2.22)</td>
</tr>
<tr>
<td>5-10</td>
<td>185</td>
<td>1.8</td>
<td>6.0 (1.73)</td>
<td>12</td>
<td>39</td>
<td>6.2</td>
<td>6.1 (1.83)</td>
<td>18</td>
<td>1.4</td>
<td>5.8 (1.97)</td>
</tr>
<tr>
<td>10-15</td>
<td>227</td>
<td>2.2</td>
<td>6.0 (1.76)</td>
<td>8</td>
<td>51</td>
<td>6.9</td>
<td>6.4 (1.92)</td>
<td>22</td>
<td>1.4</td>
<td>6.0 (2.02)</td>
</tr>
<tr>
<td>15-20</td>
<td>643</td>
<td>3.0</td>
<td>6.3 (1.87)</td>
<td>39</td>
<td>192</td>
<td>5.1</td>
<td>6.7 (1.77)</td>
<td>121</td>
<td>2.5</td>
<td>6.1 (1.95)</td>
</tr>
<tr>
<td>20-25</td>
<td>1637</td>
<td>3.2</td>
<td>6.5 (1.89)</td>
<td>110</td>
<td>696</td>
<td>4.0</td>
<td>6.9 (1.86)</td>
<td>246</td>
<td>2.7</td>
<td>5.9 (1.92)</td>
</tr>
<tr>
<td>25-35</td>
<td>4452</td>
<td>3.0</td>
<td>6.6 (1.90)</td>
<td>320</td>
<td>1988</td>
<td>3.9</td>
<td>7.0 (1.83)</td>
<td>614</td>
<td>2.2</td>
<td>6.0 (1.88)</td>
</tr>
<tr>
<td>35-45</td>
<td>3393</td>
<td>2.7</td>
<td>6.4 (1.84)</td>
<td>323</td>
<td>1277</td>
<td>3.5</td>
<td>6.9 (1.79)</td>
<td>576</td>
<td>2.0</td>
<td>6.0 (1.90)</td>
</tr>
<tr>
<td>45-55</td>
<td>3341</td>
<td>3.1</td>
<td>6.4 (1.81)</td>
<td>401</td>
<td>1012</td>
<td>3.4</td>
<td>6.9 (1.83)</td>
<td>733</td>
<td>2.3</td>
<td>5.9 (1.77)</td>
</tr>
<tr>
<td>55-65</td>
<td>3322</td>
<td>2.7</td>
<td>6.3 (1.78)</td>
<td>623</td>
<td>674</td>
<td>3.0</td>
<td>6.8 (1.82)</td>
<td>1039</td>
<td>2.1</td>
<td>5.9 (1.80)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>7851</td>
<td>3.0</td>
<td>6.4 (1.79)</td>
<td>2492</td>
<td>145</td>
<td>5.8</td>
<td>6.8 (1.87)</td>
<td>3434</td>
<td>2.3</td>
<td>6.2 (1.86)</td>
</tr>
</tbody>
</table>

Table 2. Pairwise age comparisons of first-positive RT-PCR viral load and estimated culture probability calculated from spline regression or raw data. Only the spline-based regression adjusts for effects of the test center and RT-PCR system. Differences are mean differences, with 90% credible intervals or confidence intervals from null hypothesis significance testing given in parentheses. p MW: p-value from Mann-Whitney U test (96).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Comparison</th>
<th>Spline-based regression (adjusted)</th>
<th>Raw data (unadjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Culture probability difference</td>
<td>log_{10} load difference</td>
</tr>
<tr>
<td>All</td>
<td>0-5 vs 20-65</td>
<td>-0.067 (-0.167, -0.002)</td>
<td>-0.50 (-0.62, -0.37)</td>
</tr>
<tr>
<td>All</td>
<td>5-10 vs 20-65</td>
<td>-0.054 (-0.132, -0.002)</td>
<td>-0.40 (-0.50, -0.30)</td>
</tr>
<tr>
<td>All</td>
<td>10-15 vs 20-65</td>
<td>-0.045 (-0.111, -0.002)</td>
<td>-0.30 (-0.39, -0.22)</td>
</tr>
<tr>
<td>All</td>
<td>15-20 vs 20-65</td>
<td>-0.033 (-0.076, -0.001)</td>
<td>-0.18 (-0.23, -0.12)</td>
</tr>
<tr>
<td>PAMS</td>
<td>0-5 vs 20-65</td>
<td>-0.067 (-0.167, -0.002)</td>
<td>-0.50 (-0.62, -0.37)</td>
</tr>
<tr>
<td>PAMS</td>
<td>5-10 vs 20-65</td>
<td>-0.112 (-0.279, -0.003)</td>
<td>-0.63 (-0.96, -0.32)</td>
</tr>
<tr>
<td>PAMS</td>
<td>10-15 vs 20-65</td>
<td>-0.092 (-0.228, -0.003)</td>
<td>-0.51 (-0.77, -0.26)</td>
</tr>
<tr>
<td>PAMS</td>
<td>15-20 vs 20-65</td>
<td>-0.064 (-0.162, -0.002)</td>
<td>-0.35 (-0.54, -0.17)</td>
</tr>
<tr>
<td>Hospitalised</td>
<td>0-5 vs 20-65</td>
<td>-0.033 (-0.087, -0.001)</td>
<td>-0.18 (-0.29, -0.07)</td>
</tr>
<tr>
<td>Hospitalised</td>
<td>5-10 vs 20-65</td>
<td>-0.028 (-0.104, 0.009)</td>
<td>-0.18 (-0.45, 0.07)</td>
</tr>
<tr>
<td>Hospitalised</td>
<td>10-15 vs 20-65</td>
<td>-0.025 (-0.084, 0.003)</td>
<td>-0.16 (-0.36, 0.03)</td>
</tr>
<tr>
<td>Hospitalised</td>
<td>15-20 vs 20-65</td>
<td>-0.022 (-0.071, 0.001)</td>
<td>-0.14 (-0.29, 0.02)</td>
</tr>
<tr>
<td>Other</td>
<td>0-5 vs 20-65</td>
<td>-0.018 (-0.055, 0.000)</td>
<td>-0.11 (-0.22, 0.01)</td>
</tr>
<tr>
<td>Other</td>
<td>5-10 vs 20-65</td>
<td>-0.058 (-0.148, -0.001)</td>
<td>-0.36 (-0.51, -0.20)</td>
</tr>
<tr>
<td>Other</td>
<td>10-15 vs 20-65</td>
<td>-0.044 (-0.110, -0.001)</td>
<td>-0.27 (-0.39, -0.15)</td>
</tr>
<tr>
<td>Other</td>
<td>15-20 vs 20-65</td>
<td>-0.026 (-0.072, -0.001)</td>
<td>-0.16 (-0.27, -0.06)</td>
</tr>
</tbody>
</table>
Fig. 1. Distribution of age and first-positive viral load in PAMS, Hospitalised, and Other subjects. (A) Distribution of observed first-positive viral loads for 25,381 subjects according to clinical status (6110 PAMS, 9519 Hospitalised, 9752 Other) and age group. (B) Age-viral load association with observed viral loads and confidence intervals as circles (with size indicating subject count) with vertical lines, and model-predicted viral loads and credible intervals as a black roughly-horizontal line with grey shading. (C) Overlapping age histograms according to subject clinical status. Because inclusion in the study required a positive RT-PCR test result, and testing is in many cases symptom-dependent, the study may have a proportion of PAMS cases that differs from the proportion in the general population.
Fig. 2. Estimated viral load and culture probability at time of first positive RT-PCR test. Shaded regions show 90% credible intervals in all panels. To indicate change within each 90% region, shading decreases in intensity from a narrow 50% credibility interval level to the full 90%. (A) Estimated mean viral load in first-positive RT-PCR test according to age and status. The stacked histogram (right) shows the observed viral load distribution. Small age-year to age-year variations in the proportion of subjects groups cause the fluctuations in the estimated viral loads for the total sample. Because the shaded region shows the 90% credible interval for the mean, it does not include the higher values shown in the histogram on the right. (B) Differences in estimated first-positive viral load according to age and status. Each colored line is specific to a particular subset of subjects (PAMS, Hospitalised, Other). The line shows how viral load differs by age for subjects of the corresponding status from that of 50-year old (rounded age) subjects of the same status. The comparison against those of age 50 avoids comparing any subset of the subjects against a value (such as the overall mean) that is computed in part based on that subset, thereby partially comparing data to itself. The mean first-positive viral load for PAMS and Hospitalised subjects of age 50 are 7.2 and 6.2, respectively, allowing relative y-axis differences to be translated to approximate viral loads. (C) Estimation of the association between viral load and cell culture isolation success rate based on data from our own laboratory (19) and the study of Perera et al. (20). Viral load differences in the range ~6 to ~9 have a large impact on culture probability, while the impact is negligible for differences outside that range. The vertical lines indicate the observed mean first-positive viral load for different subject groups and the horizontal lines the corresponding expected probability of a positive culture. (D) Estimated culture probability at time of first-positive RT-PCR according to age and status, obtained by combining the results in (A) and (C). Culture probability is calculated from posterior predictions, that is the posterior means shown in (A) plus error variance. The histogram on the right shows that mean culture probabilities calculated from observed log_{10} viral load values are not well-matched by credible intervals, which do not include the most-probable estimated culture probabilities. (E) Same y-axis as (D). Culture probability with highest posterior density regions, which do include the most-probable estimated culture probabilities and match the histograms in (D) well. (F) Differences of estimated expected culture probability at time of first-positive RT-PCR for age groups, with plot elements as described for (B).
Fig. 3. Posterior distributions of estimated viral loads and culture probabilities for B.1.1.7 and non-B.1.1.7 subjects, and their differences. The viral loads and estimated culture probability of 1387 B.1.1.7 and 977 non-B.1.1.7 subjects. To select a comparable subset of non-B.1.1.7 viral loads for the comparison, non-B.1.1.7 subjects were included only from test centres that had detected a B.1.1.7 variant as well as at least one non-B.1.1.7 subject, and only if the non-B.1.1.7 infection was detected on the same day as a B.1.1.7 infection was detected, plus or minus one day. Similar differences exist when viral loads from larger, less restrictive, subsets of non-B.1.1.7 subjects are used in the comparison (table S2; see materials and methods). (A) Posterior distribution of log_{10} viral load. (B) Difference of average viral load between B.1.1.7 and non-B.1.1.7 cases. (C) Posterior distribution of the estimated culture probability. See also fig. S2. (D) Difference of mean culture probability between B.1.1.7 and non-B.1.1.7 cases. Horizontal lines in (A), (B), and (D) indicate 90% credible intervals, and the highest posterior density intervals in (C).
Fig. 4. Viral load and estimated infectious virus shedding time series. Of 25,381 positive subjects, 4344 had three or more RT-PCR test results available and these were used in a viral load time series analysis. Subjects with only one result cannot be placed in time due to inherent ambiguity from the model having both an increasing and a decreasing phase, and those with only two test results were excluded from the time series analysis due to insufficient data for temporal placement (their number of data points is less than number of model parameters being estimated). (A) The number of subjects with three or more RT-PCR test results available, at least two of which were positive, according to age. (B) Estimated time course of viral load for 18,136 RT-PCR results from the 4344 subjects with at least three RT-PCR results. Blue lines are expected complete time courses for individual cases. The sample mean is shown in red, with its 90% credible interval as a shaded area. The small histogram on the right shows the distribution of all observed viral loads. The histogram values at zero correspond to the initial and trailing negative tests in subject timelines. Raw viral load time series, per subject and split by number of RT-PCR tests, are shown in fig. S8. (C) Estimated time course of positive cell culture probability, calculated by applying the results shown in Fig. 2C to the estimated viral load time courses in (B). Blue lines are expected time courses for individual subjects. The sample average is shown in red, with its 90% credible interval as a shaded area. The small histogram to the right shows the distribution of culture probability in the sample, and was obtained by applying the curve in Fig. 2C to the data in the histogram in (B).
Fig. 5. Estimated expected viral load and culture probability for age groups by time. (A) Change in estimated viral load over time according to age group for 4344 subjects with at least three RT-PCR tests, at least two of which were positive. The age coloring, range, and number of subjects in each category is given in the figure legend. Shading indicates the 90% credible interval of the mean. (B) Change in estimated culture probability over time according to age. Age groups, coloring, and shading are as in (A). (C) Estimated age group differences in mean peak viral load, corresponding to the values at day zero in (A). (D) Estimated age group differences in mean peak culture probability, corresponding to the values at day zero in (B). In (C) and (D), adjusted differences account for variations by age in clinical status and gender.
Estimating infectiousness throughout SARS-CoV-2 infection course

Terry C. Jones, Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, Florian Kurth, Peter Menzel, Rolf Schwarzer, Marta Zuchowski, Jörg Hofmann, Andi Krumbholz, Angela Stein, Anke Edelmann, Victor Max Corman and Christian Drosten

published online May 25, 2021

http://science.sciencemag.org/content/early/2021/05/24/science.abi5273

http://science.sciencemag.org/content/suppl/2021/05/24/science.abi5273.DC1

This article cites 86 articles, 10 of which you can access for free

http://science.sciencemag.org/content/early/2021/05/24/science.abi5273#BIBL

http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).