

# 1        **Viral loads observed under competing strain dynamics**

2        **Authors:** James A. Hay<sup>1,2,3\*</sup>, Lee Kennedy-Shaffer<sup>4</sup>, Michael J. Mina<sup>1,2,3,5</sup>

## 3        **Affiliations:**

4        <sup>1</sup>Center for Communicable Disease Dynamics, Harvard T H Chan School of Public Health, Boston,  
5        MA.

6        <sup>2</sup>Department of Epidemiology, Harvard T H Chan School of Public Health, Boston, MA.

7        <sup>3</sup>Department of Immunology and Infectious Diseases, Harvard T H Chan School of Public Health,  
8        Boston, MA.

9        <sup>4</sup>Department of Mathematics and Statistics, Vassar College, Poughkeepsie, NY.

10        <sup>5</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School

11        \*Correspondence to: [jhay@hsph.harvard.edu](mailto:jhay@hsph.harvard.edu) (JAH)

## 12        **Abstract**

13        A plausible mechanism for the increased transmissibility of SARS-CoV-2 variants of concern  
14        (VOCs) results from VOC infections causing higher viral loads in infected hosts. However,  
15        investigating this hypothesis using routine RT-qPCR testing data is challenging because the  
16        population-distribution of viral loads changes depending on the epidemic growth rate; lower  
17        cycle threshold (Ct) values for a VOC lineage may simply reflect increasing incidence relative  
18        to preexisting lineages. To understand the extent to which viral loads observed under routine  
19        surveillance systems reflect viral kinetics or population dynamics, we used a mathematical  
20        model of competing strain dynamics and simulated Ct values for variants with different viral  
21        kinetics. We found that comparisons of Ct values obtained under random cross-sectional  
22        surveillance were highly biased unless samples were obtained at times when the variants had  
23        comparable growth rates. Conversely, comparing Ct values from symptom-based testing was  
24        largely unaffected by epidemic dynamics, and accounting for the time between symptom onset  
25        and sample collection date further reduced the risk of statistical errors. Finally, we show how  
26        a single cross-sectional sample of Ct values can be used to jointly estimate differences in viral  
27        kinetics and epidemic growth rates between variants. Epidemic dynamics should be  
28        accounted for when investigating strain-specific viral kinetics using virologic surveillance data,  
29        and findings should be corroborated with longitudinal viral kinetics studies.

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## 30 Introduction

31 One of the biggest threats to combating the SARS-CoV-2 pandemic, or indeed any virologic  
32 epidemic, is the emergence of novel variants that may be harder to control or exhibit increased  
33 disease severity [1–4]. Variants with increased growth rates, which may arise through  
34 mutations affecting infectivity or antigenicity, can quickly come to dominate existing lineages  
35 and generate new waves of infection [5–11]. Whereas existing non-pharmaceutical  
36 interventions and population immunity may be sufficient to suppress existing viruses, they may  
37 be insufficient for more transmissible variants, threatening the robustness of vaccine-induced  
38 herd immunity and exacerbating problems in settings with limited access to vaccines.  
39 Identifying measurable properties of variants that indicate increased transmission potential is  
40 therefore essential for controlling the SARS-CoV-2 pandemic, as in the case of seasonal  
41 influenza [12–14].

42 A hypothesized mechanism for increased transmissibility relates to improved within-host  
43 replication, which may result in higher viral loads [15–17]. If viral load predicts infectivity  
44 [17,18,19], then infections with new variants that elicit higher peak viral load, shorter  
45 incubation periods or slower clearance rates could be more transmissible and for a longer  
46 period of time. Testing this hypothesis would ideally rely on longitudinal viral kinetics studies  
47 to directly compare viral loads over the course of infection [16,20,21]. However, such data are  
48 rare, and comparisons of viral loads have therefore typically been done using routinely  
49 collected RT-qPCR surveillance data from asymptomatic or symptomatic individuals [5,22].  
50 Indeed, multiple such studies have now proposed that samples isolated from variant of  
51 concern (VOC) infections demonstrate higher viral loads than from non-VOC infections [22–  
52 26].

53 However, comparisons of viral loads, usually proxied using RT-qPCR cycle threshold (Ct)  
54 values, from surveillance samples are potentially biased depending on the epidemiological  
55 context [27–29]. It has been shown that cycle threshold (Ct) values observed through  
56 population-level surveillance are expected to change depending on the underlying epidemic  
57 growth rate: Ct values are skewed lower when the epidemic is growing due to the abundance  
58 of recent infections, and skewed higher when the epidemic is growing due to the  
59 predominance of older infections [27,30]. Comparing Ct values from different lineages which  
60 may have different growth rates at the time of sample collection therefore has the potential for  
61 bias. Higher viral loads from VOC samples may simply reflect a growing epidemic as opposed  
62 to higher peak or more sustained viral loads, making it difficult to accurately infer differences

63 in underlying viral kinetics. Methods are needed to accurately quantify the contribution of  
64 virologic changes to viral load dynamics at the within-host and population levels.

65 Here, we explore in simulation different virologic and surveillance scenarios where epidemic  
66 dynamics can confound viral load comparisons between variants. We first demonstrate how  
67 average viral loads observed at the population level from new, more transmissible variants  
68 would be expected to differ from existing viruses even with identical post-infection viral  
69 kinetics. We then demonstrate how observations of these patterns differ depending on  
70 whether samples are obtained through random cross-sectional surveillance or symptom-  
71 based testing. We show that accounting for the epidemic growth rate when comparing  
72 samples from random cross-sectional surveillance or time-since-onset when comparing  
73 samples from symptom-based surveillance can lead to robust comparisons of RT-qPCR  
74 results between variants with different epidemic dynamics. Finally, we present a method for  
75 comparing growth rates from different variants using a single cross-section of Ct values whilst  
76 accounting for potential differences in underlying viral kinetics.

## 77 Results

### 78 *Modeling framework to compare viral loads arising in a two-strain epidemic*

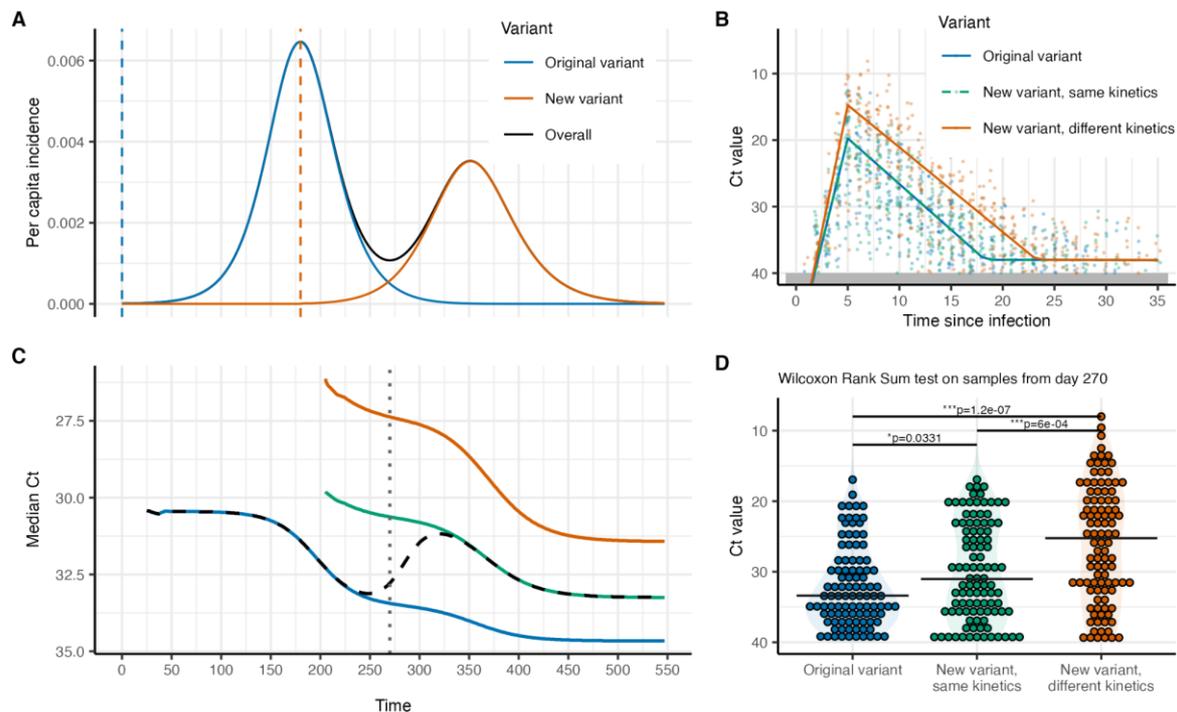
79 We developed a mathematical model to simulate viral loads, observed as RT-qPCR Ct values,  
80 in a population undergoing a two-strain epidemic (Figure 1). We implemented a two-strain  
81 SEIR model to simulate incidence curves under a scenario where a more transmissible variant  
82 (referred to as the “new variant”) is introduced into the population and outcompetes a  
83 preexisting, less transmissible lineage (referred to as the “original variant”). We assumed that  
84 these strains have identical epidemiological parameters other than their basic reproductive  
85 numbers, and that infection elicits symmetric cross-protection against the other variant. We  
86 combined these incidence curves (Figure 1A) with the viral kinetics models shown in Figure  
87 1B to simulate viral loads for infected individuals over time. We compared two scenarios for  
88 underlying viral kinetics: 1) both variants have identical viral kinetics and 2) the new variant  
89 has a higher peak viral load and a slower clearance rate. To simulate observed viral loads,  
90 individuals were randomly sampled under one of two strategies: either 1) random cross-  
91 sectional surveillance, where individuals are sampled from the population at random  
92 regardless of their infection status, or 2) symptom-based surveillance, where individuals are  
93 tested after some delay following the onset of symptoms. These scenarios underpin all  
94 analyses up to the section “*Quantifying differences in growth rate and viral kinetics of variants*  
95 *using cross-sectional Ct values*” unless stated otherwise.

### 96 *Comparing viral loads from samples with different growth rates*

97 The simulations show that the distribution of viral loads among infected individuals changes  
98 over time, reflecting the growth rate of the epidemic (Figure 1C). Because the two variants  
99 have different transmission rates and introduction dates, their viral load distributions differ at  
100 any given point in time regardless of any true difference in viral kinetics. Under random cross-  
101 sectional surveillance, this difference arises because the time-since-infection distribution  
102 changes over the epidemic: randomly sampled infections are typically more recent when the  
103 epidemic is growing than when it is declining [27,30]. As a result, statistical tests comparing  
104 Ct values from samples obtained under random cross-sectional surveillance at a single point  
105 in time will reflect differences arising both from epidemic dynamics, which dictate the recency  
106 of infection, as well as underlying viral kinetics (Figure 1D). When simulated samples were  
107 obtained in this way and the new variant samples were compared to the original variant  
108 samples using a Wilcoxon rank-sum test with significance level of 5%, we found that type 1  
109 statistical errors occurred in 25% of simulations when using 25 detectable Ct values from each

110 of the two variants, increasing to nearly 100% when comparing 500 detectable Ct values  
111 (Figure S1).

112 One approach to overcome these statistical biases and accurately detect underlying  
113 differences in viral kinetics is to compare samples taken from time points of comparable growth  
114 rates (Figure 2A). For example, comparing original variant samples taken during the first wave  
115 of infections to new variant samples taken during the second wave will lead to more accurate  
116 statistical tests for underlying differences in viral kinetics (Figure 2B). Comparing samples in  
117 this way resulted in type 1 errors in only 5% of trials (the nominal rate), with statistical power  
118 of at least 95% when 250 or more detectable Ct values per variant were sampled (Figure  
119 S2&S3).



120

121 **Figure 1. Variants with different epidemic growth rates will exhibit differences in observed**

122 **average viral loads regardless of true differences in within-host kinetics. (A)** Incidence curves

123 from a two-strain susceptible-exposed-infected-recovered (SEIR) model, where a virus introduced on

124 day 0 (vertical blue line) with  $R_0=1.5$  is outcompeted by a new variant introduced on day 180 (vertical

125 red line) with  $R_0=2.5$ , leading to two waves of infection. (B) Variants are assumed to follow either

126 identical or different viral load kinetics, with modal viral loads (solid lines) peaking within one week post

127 infection, coinciding with the typical time of symptom onset, and then declining to near the limit of

128 detection at around three weeks post infection. Substantial variation in individual Ct values (individual

129 points) are observed due to individual-level kinetics and sampling variation. (C) The median Ct value

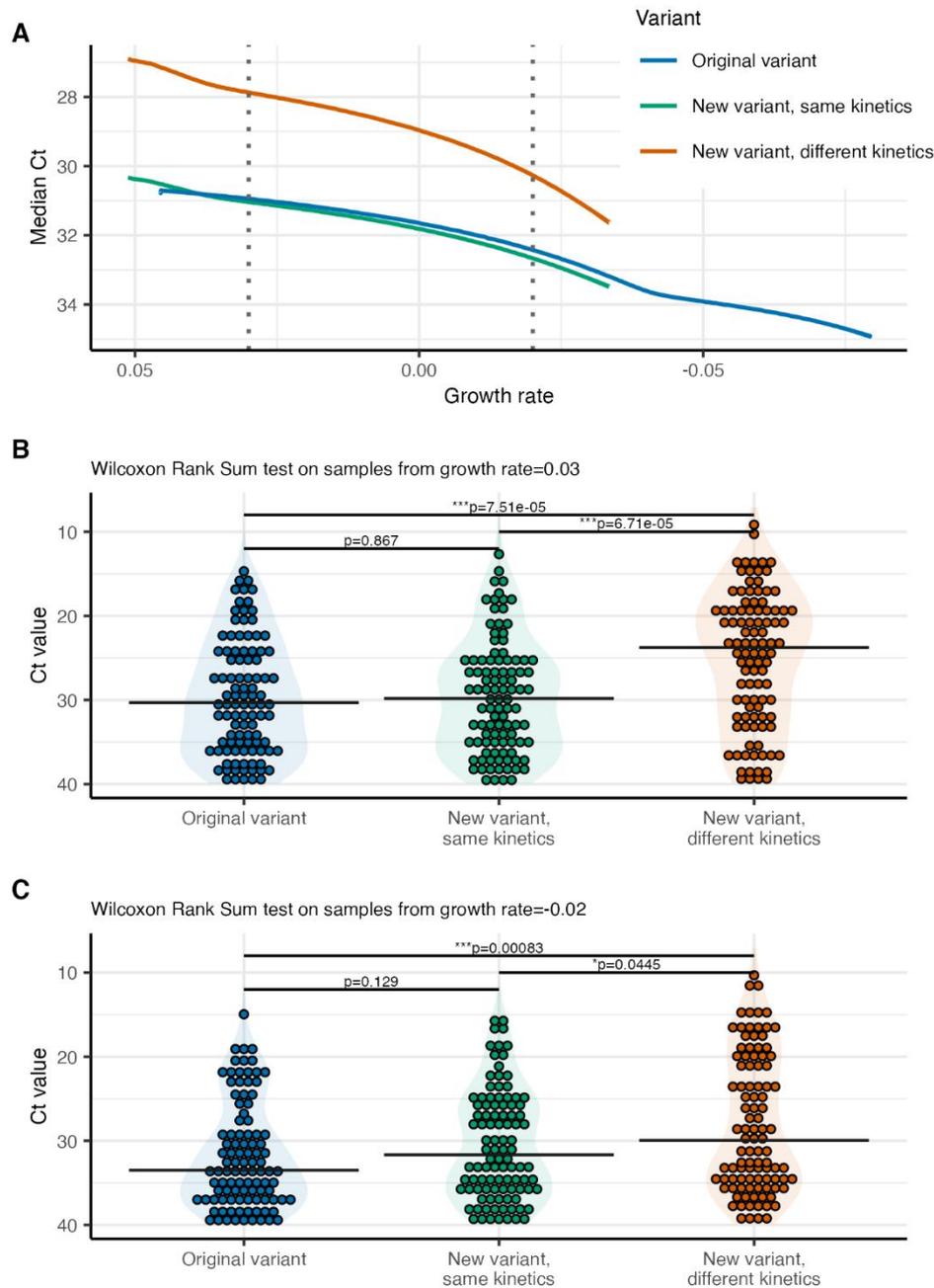
130 observed from individuals sampled entirely at random will reflect the growth rate of that variant at the

131 time of sampling. Black dashed line shows the overall median Ct value. (D) Comparison of simulated

132 Ct values obtained on day 270 using a Wilcoxon rank sum test. When the original variant is in decline

133 and the new variant is growing, the comparison reflects a significant difference in viral load between

134 the variants regardless of a true difference in underlying viral kinetics.



135

136 **Figure 2. Comparing Ct values obtained from variants at times of comparable epidemic growth**  
 137 **lead to accurate comparisons of viral loads. (A)** Using the simulations shown in Figure 1C, median  
 138 Ct values for the variants are plotted against their growth rate at time of sampling. **(B)** Comparison of  
 139 simulated Ct values obtained from the time when each variant had a growth rate of 0.03 (i.e., the log  
 140 ratio of new infections tomorrow relative to today) using a Wilcoxon Rank Sum test. **(C)** As in (B), but  
 141 when the growth rate is -0.02.

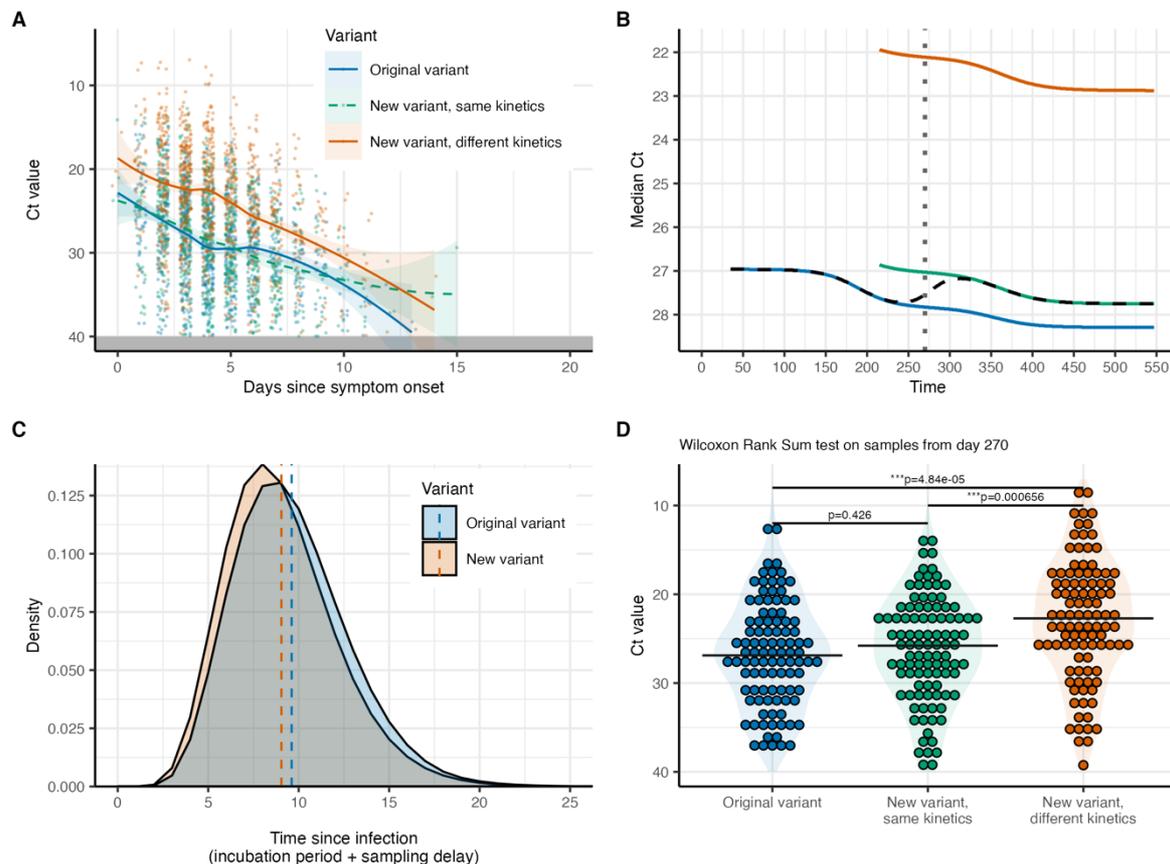
142 *Viral loads from symptom-based surveillance more accurately reflect underlying differences*  
143 *in viral kinetics*

144 In reality, most RT-qPCR results are obtained from non-random sampling. In particular,  
145 symptom-based surveillance, where individuals with recent symptom onset seek testing, is  
146 likely to comprise the majority of samples. To test whether epidemiological biases were  
147 present in Ct values obtained under symptom-based surveillance, we simulated Ct values  
148 using the same time-since-infection model in Figure 1C but assumed that individuals were  
149 sampled after some delay following the onset of symptoms. We assumed that symptomatic  
150 individuals have incubation periods drawn from a log-normal distribution with median 5.0 days  
151 and variance of 5.8 days, and are subsequently sampled after some delay drawn from a  
152 discretized gamma distribution with shape and scale parameters of 5 and 1 respectively  
153 (Figure 3A). This corresponds to a mean confirmation delay of 4.5 days and variance of 5.1  
154 days following symptom onset. Both variants were assumed to have the same incubation  
155 period and sampling delay distribution. In addition, on the individual level, the Ct value at any  
156 day was independent of the incubation period length and sampling delay. We note that the  
157 choice of these distribution parameters is arbitrary; similar patterns will be observed with  
158 different values.

159 Although some differences remained in observed Ct values over time between the two  
160 variants, the difference was small unless the new variant had truly different underlying viral  
161 kinetics (Figure 3B). This is because the time-since-infection distributions for the two strains  
162 are comparable regardless of the underlying growth rate, as individuals are always sampled  
163 at a similar time post onset and therefore post infection (Figure 3C). Therefore, comparing Ct  
164 values between variants obtained from symptomatic surveillance largely reflects true  
165 differences in underlying viral kinetics (Figure 3D).

166 However, statistical comparisons may still lead to flawed conclusions if the time-since-onset  
167 and sampling delay distributions are substantially different between the two variants. Even  
168 when the same distributions are assumed, some small difference in the observed time-since-  
169 infection distribution will arise due to differences in the underlying epidemic growth rate (see  
170 Figure S3 and S4 from [27] for further details). For example, if an epidemic is growing, then  
171 individuals with symptom onset on a given day are more likely to have been recently infected  
172 with a short incubation period simply due to the abundance of recent infections. Our  
173 simulations show that as the number of samples being compared increases, the probability of  
174 encountering a type 1 statistical error when using the Wilcoxon rank sum test increases. This  
175 is true because of the reduced variation within groups, even though the between-variant  
176 estimated median Ct difference is the same (Figure S4).

177 Using a linear regression model and including days-since-onset as an explanatory variable  
178 drastically reduces the frequency of type 1 errors (Figure S5). Indeed, accounting for days-  
179 since-onset becomes vital when the distribution of delays between symptom onset and  
180 sampling differs between the two variants. Figure S6 shows type 1 statistical errors are almost  
181 guaranteed when using a Wilcoxon rank sum test if the original variant has a different sampling  
182 delay distribution (assuming shape and scale parameters of 7 and 0.9 as opposed to 5 and  
183 1), but Figure S7 shows that these errors can be drastically reduced through using a  
184 regression model accounting for days-since-onset. We note that type 1 errors still arise at  
185 large sample sizes, as accounting for days-since-onset only partially accounts for the small  
186 differences in the time-since-infection distribution (i.e., it does not account for the incubation  
187 period).



188

189 **Figure 3. Ct values obtained through symptom-based surveillance more accurately reflect true**  
190 **viral load differences. (A)** Ct values simulated under the same time-since-infection model as in Figure  
191 1, but with samples taken after the onset of symptoms and some additional sampling delay. Time from  
192 infection to symptom onset was drawn from a log-normal distribution with median 5.0 days and variance  
193 5.8 days. Time from symptom onset to sample collection was drawn from a discretized gamma  
194 distribution with mean 4.5 days and variance 5.1 days. Solid line and ribbons show fitted smoothing  
195 spline and 95% CI. **(B)** Median Ct value observed over time from individuals sampled under symptom-  
196 based surveillance. **(C)** Distribution of incubation periods for individuals sampled on day 270 of the  
197 simulation, stratified by variant. Note that the observed distribution on day 270 differs from the log-  
198 normal distribution used for simulation through its convolution with the infection incidence curve. **(D)**  
199 Comparison of simulated Ct values obtained on day 270 under symptom-based surveillance using a  
200 Wilcoxon rank sum test.

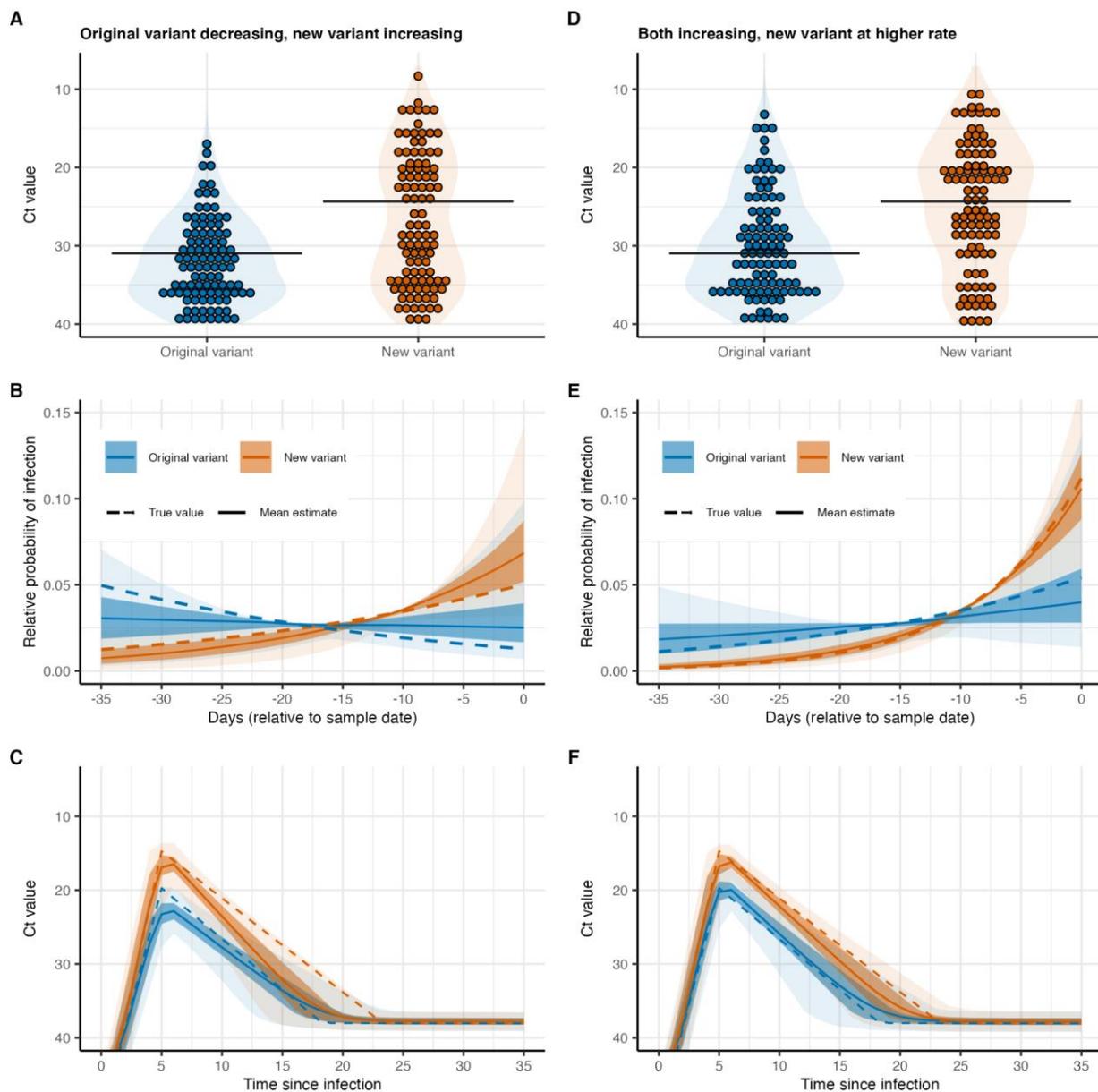
201 *Quantifying differences in growth rate and viral kinetics of variants using cross-sectional Ct*  
202 *values*

203 Finally, we approached the system from a different perspective: rather than testing for  
204 differences in viral kinetics between variants based on surveillance samples, we asked if we  
205 could use random cross-sectional samples taken from the same point in time to infer  
206 differences in variant-specific growth rates. Again, we generated a synthetic population of  
207 SARS-CoV-2 infections by simulating Ct values obtained through random cross-sectional  
208 surveillance at a single point in time assuming that the variants had different viral kinetics (as  
209 shown in Figure 1B). We then adapted a previously described method to estimate the growth  
210 rate and viral kinetics parameters based on single cross-sections of Ct values [27]. We tested  
211 two epidemiologic scenarios: 1) incidence of the original variant declines while the new variant  
212 increases and 2) both variants have positive growth rates at the same time, but the new variant  
213 has a higher basic reproductive number ( $R_0=2.5$  vs. 1.5).

214 When 100 detectable Ct values for each variant were obtained through random cross-  
215 sectional surveillance at a time when one variant was declining and the other was increasing  
216 in frequency (Figure 4A), we were generally able to accurately re-estimate the true growth  
217 rates of the two variants (Figure 4B). The model was also able to quantify the true difference  
218 in peak viral loads elicited by the new variant (Figure 4C); however, it was not able to identify  
219 the slower clearance rate. Similarly, in the scenario when infections from both variants were  
220 simultaneously increasing but at different rates (Figure 4D), we were able to identify that the  
221 new variant likely had a higher growth rate, although not unequivocally based on the 95%  
222 credible intervals (CrI) on the difference in growth rate (Figure 4E). Again, differences in peak  
223 Ct value were identifiable, but not in the clearance rate (Figure 4F). These results show that  
224 although lower median Ct values may be explained either by higher growth rates or differences  
225 in viral kinetics, the full distribution of Ct values may hold information that allows both  
226 processes to be identified.

227 The results shown in Figure 4 are from a single simulation of detectable Ct values. Therefore,  
228 to explore the identifiability of these differences in growth rates and viral kinetics based on  
229 cross-sectional samples of different sizes and stochastic draws, we repeated the analyses  
230 with varying sample sizes from 25 to 500 detectable Ct values per variant, running 100  
231 simulations for each sample size. When one variant was in decline while the other was  
232 increasing in frequency (Figure S8), we generally identified the different direction of the  
233 epidemic trajectories with as few as 25 Ct values per variant (with increasing certainty at  
234 increasing sample sizes). However, when both variants were increasing at the same time but  
235 at different rates (Figure S9), we were not able to exclude the possibility of no difference in

236 growth rate even with 500 Ct values per variant. This appeared to be largely driven by the lack  
237 of precision in growth rate estimates for the original variant, resulting in wide credible intervals.  
238 Nonetheless, posterior mean estimates were consistently identified as different between the  
239 two variants. Differences in peak Ct value were also identifiable in both epidemiologic  
240 scenarios with as few as 25 Ct values; however, even at 500 Ct values per variant, we were  
241 unable to reliably identify the true difference in viral clearance rates (Figures S10 and S11).



242

243 **Figure 4.** (A) Simulated Ct values assuming that the new variant has a higher peak viral load (-5 Ct at  
 244 peak) and a longer time to clearance (5 additional days from peak to second hinge). Samples were  
 245 assumed to be obtained under random cross-sectional surveillance at day 270 of the simulation shown  
 246 in Figure 1. Horizontal lines show median Ct. (B) Estimated epidemic trajectory inferred using the single  
 247 cross section of Ct values. Solid line, dark shaded region and light shaded region show posterior mean,  
 248 50% credible intervals (CrI) and 95% CrI respectively. Dashed line shows the true epidemic trajectory  
 249 for each variant. (C) Inferred viral kinetics from the same model fitting procedure as in (B). Shown are  
 250 the posterior estimates for the modal Ct value over time since infection. Solid line, dark shaded region  
 251 and light shaded region show posterior mean, 50% credible intervals (CrI) and 95% CrI respectively.  
 252 Dashed line shows the true modal Ct curve. (D), (E), and (F) are identical to (A), (B) and (C) respectively,  
 253 but with Ct values obtained from a simulation where both variants are seeded at the same time, but  
 254 with the new variant assumed to have a higher basic reproductive number.

## 255 Discussion

256 Given the relationship between epidemic growth rate and the distribution of viral loads in a  
257 population [27], care must be taken when directly comparing viral loads between variants  
258 using surveillance samples to avoid drawing incorrect conclusions of differences in strain  
259 virulence. Infections from variants with positive growth rates will typically be more recent than  
260 infections from co-circulating variants with negative growth rates, resulting in higher average  
261 viral loads. These findings provide a conceptually similar caution to analyses assessing the  
262 association between mutations and transmissibility [14,31]: just as mutations may be  
263 associated with increased growth rates simply due to founder effects or chance, infections  
264 with new variants may be associated with increased viral loads simply because they are  
265 experiencing higher epidemic growth rates.

266 We found that comparing Ct values between variants using samples obtained entirely at  
267 random from the population are far more likely to lead to incorrect conclusions regarding  
268 differences in underlying viral kinetics than samples obtained from recently symptomatic  
269 individuals. This is because of how these sampling strategies reflect the underlying time-since-  
270 infection distribution: cross-sectional surveillance samples individuals at a random time point  
271 in their infection, whereas symptom-based surveillance systematically tests people at a similar  
272 time after infection. Because viral loads [32], and therefore Ct values, are dictated by the time  
273 post infection, comparing Ct values from samples with similar time-since-infection distributions  
274 will reflect differences in viral kinetics and not epidemic growth rates. Accounting for  
275 differences in the time-since-infection distribution between datasets, which may be achieved  
276 using samples taken from times of comparable growth rates or by including time-since-onset  
277 or time-since-infection in a regression model, will improve the reliability of statistical tests  
278 comparing Ct values between variants. However, these results assume that the variants have  
279 similar delays between exposure and symptom onset. If, for example, newer variants have  
280 systematically shorter symptom onset delays, then new variant samples obtained under  
281 symptom-based surveillance would reflect a shorter time-since-infection distribution.

282 At the time of writing, the SARS-CoV-2 literature has a mixture of studies that either do not  
283 acknowledge the potential for biases in Ct value comparisons [6,26], discuss the potential for  
284 this bias [33,34], or take clear steps to account for differences in the time-since-onset or time-  
285 since-infection distribution when comparing Ct values [5,22]. For example, an analysis of P.1  
286 samples (Gamma variant) in Manaus, Brazil found that Ct values declined over time as the  
287 prevalence of P.1 infections increased. An initial comparison of values found a statistically  
288 significant association between P.1 infection and lower Ct value; however, after accounting

289 for the delay between symptom onset and sample collection, the significance of this  
290 relationship was lost, similar to our simulation results [5]. Another analysis comparing Ct  
291 values between two variants in Washington State, USA also initially performed a direct  
292 comparison of Ct values between a variant with a 614G substitution [22]. Again, the authors  
293 found that 614G was associated with lower Ct values than 614D. In contrast to the analysis  
294 by [5], this significant difference in Ct value remained after accounting for a number of potential  
295 confounders such as days after symptom onset and patient age. In this study, epidemic  
296 dynamics were likely not important in dictating Ct value distributions, as Ct values did not  
297 appear to be associated with time of sample collection and the time-since-onset distribution  
298 was similar between the two variants.

299 Although epidemic dynamics have the potential to bias viral load comparisons, particularly  
300 using random cross-sectional samples, many study designs are unlikely to be affected.  
301 Comparing samples from recently symptomatic individuals, particularly when the distribution  
302 of delays between symptom onset and sample collection date are similar between the variants  
303 or are included in a regression model, is likely to lead to reliable conclusions. Such  
304 comparisons will be even more reliable if the timing of symptom onset is dependent on viral  
305 load (i.e., symptom onset occurs as a result of viral loads reaching their peak as opposed to  
306 independently). It is worth noting that differences in sampling delay distributions will likely vary  
307 independently of the epidemic growth rate simply due to the logistics of sample collection,  
308 limitations on testing capacity and changes in policy dictating who is tested. Accounting for  
309 time-since-onset is therefore advisable regardless. Other instances where epidemic dynamics  
310 will not affect viral load comparisons are when samples are obtained longitudinally from the  
311 same individuals, allowing the comparison of the full viral kinetics curve [20,35], or obtained  
312 near the time of exposure such that all samples have a similar time-since-infection [36].

313 There are a number of additional factors affecting viral load distributions that we did not  
314 consider here. First, we assumed in our simulations that Ct values were comparable across  
315 all samples. In reality, Ct values from different lineages may not be comparable across primers  
316 and platforms, such that Ct value comparisons reflect technological limitations rather than  
317 differences in viral load [37]. In such cases, conversion to a common scale such as viral load  
318 using calibration curves may be advisable [17]. We also did not consider how patient-level  
319 factors affect viral loads. Some limited evidence exists to suggest that children exhibit  
320 systematically lower viral loads than adults, with observations also affected by external factors  
321 such as swab quality or viral location within the respiratory tract [16,38,39]. Clinical severity  
322 may also affect viral load kinetics, with symptomatic patients exhibiting slower clearance rates  
323 than asymptomatic infections [32,40]. If samples being compared between variants represent

324 different underlying populations, which is likely when the age distribution of cases shifts over  
325 time, then viral load differences may reflect differences in the tested population rather than  
326 viral kinetics. Such shifts in clinical severity and age distributions have been seen in countries  
327 with high vaccine coverage, as vaccination of the elderly and at-risk groups shifts the relative  
328 prevalence of infections into younger populations [41]. Additionally, vaccination not only  
329 affects the composition of the population testing positive for SARS-CoV-2, but infections in  
330 vaccinated individuals may also exhibit lower viral loads [42]. Overall, care should be taken to  
331 either compare sample sets from similar populations, or meta-data on relevant demographic  
332 factors should be accounted for when comparing Ct values between variants.

333 Finally, we adapted a previous method, which is ultimately a generalized linear regression  
334 model, to use Ct values obtained from a single point in time to simultaneously estimate  
335 differences in growth rates alongside differences in viral kinetics [27]. In reality, these two  
336 mechanisms may not be uniquely identifiable; however, understanding the combination of viral  
337 kinetics and transmissibility differences that can explain the data may still be valuable to detect  
338 variants of concern early and to quantify transmission advantages. This approach may be  
339 particularly useful in settings where sequencing capacity is limited, as sequencing-  
340 independent means of stratifying samples by lineage (e.g., variant-specific primers, single  
341 gene failure etc.) can be used to then compare Ct values between variants [43]. However, we  
342 emphasize that this approach currently requires samples obtained through random cross-  
343 sectional surveillance, or nearly random samples such as non-COVID patient hospital testing.

## 344 **Materials and methods**

### 345 *SEIR model*

346 We used a deterministic, two-strain, susceptible-exposed-infected-recovered (SEIR) model to  
347 simulate incidence curves for two competing strains in a fully susceptible population [44]. We  
348 assumed that two strains were seeded in the population: an “original variant” with  $R_0 = 1.5$ ,  
349 and a “new variant” with  $R_0 = 2.5$ . The “original variant” was seeded on day 0 of the simulation,  
350 and the “new variant” was either seeded on day 180 or on day 0 as specified in the *Results*.  
351 We assumed that both strains had a 3-day mean latent period and 7-day mean infectious  
352 period, and that individuals became permanently immune to the strain they were infected with  
353 upon recovery. We also assumed that infection with one strain elicited strong cross-immunity  
354 (75% reduced infection probability) to the other. Overall, this is a simplified implementation of  
355 the model presented by [44] with no seasonality, no waning immunity and no births or deaths.  
356 The model was solved using daily time steps, and daily growth rates were calculated as  $g(t) =$   
357  $\log\left(\frac{y(t+1)}{y(t)}\right)$  where  $y(t)$  is the incidence of new infections on day  $t$ . Full model equations and a  
358 table of parameters are shown in the *Supplementary Material: SEIR model equations* and  
359 Table S1.

### 360 *Viral kinetics model*

361 We used an existing model describing the average and distribution of viral loads as a function  
362 of time-since-infection with SARS-CoV-2 [27]. Briefly, the model assumes that the modal Ct  
363 value follows a piecewise linear function following infection,  $C_{mode}(a)$ . After an initial 1-day  
364 period of no viral growth, the modal Ct value decreases monotonically to a peak value at day  
365 5 post infection. We assumed the peak Ct value was 20 for the original variant, or 15 for the  
366 new variant in scenarios assuming different viral kinetics. Ct values then increase to a plateau  
367 at a value of 38, which occurred on day 13 post peak for the original variant and day 18 post  
368 peak for the new variant with different viral kinetics. Thereafter, individuals have a daily  
369 probability of becoming fully undetectable, modeled as a Bernoulli process with probability  
370  $p_{add}$ .

371 To capture the substantial variation across individuals and samples in Ct values observed on  
372 a given day post-infection, we assumed that Ct values follow a Gumbel distribution with a  
373 scale parameter  $\sigma(a)$  that begins at 5 but decreases to 4 towards the end of infection. This  
374 decreasing scale parameter captures the fact that samples taken early in infection are affected  
375 by variation from both individual-level heterogeneity in kinetics in addition to sampling  
376 variation, whereas samples taken towards the end of infection largely represent consistent,

377 small quantities of residual viral RNA. We assumed that the maximum Ct value was 40, and  
378 all samples from undetectable or uninfected individuals were excluded from the analyses. We  
379 also note that although we describe the model on the scale of Ct values, analogous results  
380 would hold for viral loads, as Ct values are linearly related to log viral loads. Full model  
381 equations and a table of parameters are shown in the *Supplementary Material: viral kinetics*  
382 model and Table S2.

### 383 *Simulated surveillance samples and Ct values*

384 We combined the SEIR and viral kinetics models to simulate Ct values among the infected  
385 population over time. In brief, we used the SEIR model to simulate when individuals are  
386 infected with either variant over time and used the viral kinetics model to simulate their  
387 observed Ct value when sampled on a particular day post infection. Observations were  
388 simulated using two surveillance strategies described below. For the analyses underpinning  
389 Figures S1-S7, we generated 1000 simulated datasets for each sample size of 25, 50, 100,  
390 250 or 500. We simulated datasets on day 270 of the epidemic in scenarios where the new  
391 variant had a later seed date, or day 50 when the strains were seeded on the same day.

392 First, we considered random cross-sectional sampling, where individuals are tested entirely  
393 at random regardless of their time since infection. In this case, we simulated the distribution  
394 of detectable Ct values  $X_{v,t}$  for variant  $v$  on a given day of the epidemic  $t$  as  $X_{v,t} \sim f_{v,t}(x)$ .  $f_{v,t}(x)$   
395 is the probability density function (PDF) for *detectable* Ct values on day  $t$  of the epidemic,  
396 calculated by convoluting the variant-specific incidence curve (which describes the distribution  
397 of times since infection) and viral kinetics model (which describes the distribution of observed  
398 Ct values on each day after infection):

$$399 \quad f_{v,t}(x) = \frac{\sum_{a=0}^{A_{max}} p_a(x) \phi_a \pi_{v,t-a}}{\sum_{a=0}^{A_{max}} \phi_a \pi_{v,t-a}}$$

400 where  $A_{max}$  is the maximum time-since-infection for which individuals may still be PCR  
401 detectable (set to 35 days);  $p_a(x)$  is the Gumbel probability density function scaled to only  
402 take values between 0 and  $C_{LOD}$  with location parameter  $C_{mode}(a)$  and scale parameter  $\sigma(a)$ ;  
403  $\phi_a$  is the probability of being PCR detectable on day  $a$  post infection; and  $\pi_{v,t}$  gives the  
404 probability of infection with variant  $v$  on day  $t$  of the epidemic. Individual observations were  
405 generated by simulating from this PDF. Note that this PDF can be modified to generate PCR-  
406 negative observations, but we are only interested in the distribution of PCR-positive  
407 observations here.

408 Second, we simulated Ct values under symptom-based surveillance, where individuals are  
409 tested following the onset of symptoms. Again, we simulated from the PDF for Ct values on  
410 day  $t$  of the epidemic as  $X_{v,t} \sim g_{v,t}(x)$ , but this time convoluting the incubation period and  
411 sampling delay distributions as well as the incidence and viral kinetics models. Note again that  
412 we assume that only PCR-positive Ct values were observed:

$$413 \quad g_{v,t}(x) = \frac{\sum_{d=0}^{D_{\max}} \sum_{o=0}^{O_{\max}} p_{d+o}(x) \phi_{d+o} s(d) w(o) \pi_{v,t-(d+o)}}{\sum_{d=0}^{D_{\max}} \sum_{o=0}^{O_{\max}} \phi_{d+o} s(d) w(o) \pi_{v,t-(d+o)}}$$

414 where  $d$  is the sampling delay in days;  $o$  is the incubation period in days;  $s(d)$  is the PDF for  
415 the discretized gamma sampling delay distribution; and  $w(o)$  is the PDF for the discretized  
416 log-normal incubation period distribution.

417 Simulating Ct values for the analyses shown in Figure S5 and Figure S7 is slightly more  
418 involved because each individual needs both an observed Ct value and corresponding  
419 sampling delay. Although the principle of simulating from the PDF is the same, we simulated  
420 individual-level line list data for these analyses as described in the *Supplementary Material:*  
421 *simulating samples under symptom-based surveillance*.

## 422 *Statistical methods*

423 Direct statistical comparisons of Ct distributions from the two variants were two-sided, two-  
424 sample Wilcoxon rank sum tests (Mann-Whitney test) at a significance level of 5%. In the  
425 analyses controlling for days-since-onset when comparing Ct values, we fit linear regression  
426 models of the form  $E[x_i | d_i, v_i] = \beta_0 + \beta_1 d_i + \beta_2 v_i$ , where  $d_i$  gives the days between  
427 symptom onset and sample collection for individual  $i$ , and  $v_i$  gives the variant infecting that  
428 individual. Hypothesis tests in the regression models were conducted for the null hypothesis  
429  $H_0: \beta_2 = 0$  using an asymptotic t-test. Power was defined as the proportion of simulations  
430 under a given scenario and sample size where the null hypothesis (the distribution of Ct values  
431 for both variants are equal or there is no effect of variant on Ct value) was correctly rejected  
432 in simulations where the variants have different viral kinetics. Type 1 errors were defined as  
433 tests which incorrectly rejected the null hypothesis in simulations where the variants have  
434 identical viral kinetics.

## 435 *Estimating growth rates and viral kinetics from cross-sectional samples*

436 Using the simulated random cross-sectional samples of detectable Ct values, we jointly  
437 estimated the posterior distributions of viral kinetics parameters and the exponential growth  
438 rate that are consistent with the observed data. We repeated this process for each of the 100

439 simulated datasets of sample sizes of 25, 50, 100, 250 and 500 detectable Ct values per  
440 variant. Model fitting was carried out using the *virosolver* R package [45]. In brief, model fitting  
441 involves: (1) defining the likelihood of observing Ct values conditional on the viral kinetics  
442 model parameters and incidence curve, assuming that incidence follows exponential growth  
443 with an unknown growth rate parameter; (2) defining priors for all model parameters as  
444 described in Table S2; and (3) estimating the posterior distribution of all model parameters  
445 conditional on the Ct data using Markov chain Monte Carlo (MCMC). For each model fit, three  
446 chains were run for 100,000 iterations each with the first 50,000 iterations discarded as burn  
447 in. Convergence was assessed based on the trace plots and obtaining effective sample sizes  
448  $>200$  and  $\hat{R} < 1.1$  for all estimated parameters.

449 For the most part, we followed the method exactly as described in [27]. However, because we  
450 are interested in jointly estimating the viral kinetics and epidemic growth rates for two co-  
451 circulating strains (whereas [27] is defined for only one strain), we made the following  
452 modifications: (1) the peak Ct value parameter for the new variant was defined relative to the  
453 peak Ct value for the original variant (i.e.,  $C'_p = \rho C_p$ ); (2) the second hinge point of the Ct  
454 model for the new variant was defined relative to the original variant (i.e.,  $t'_s = \eta t_s$ ); and (3)  
455 each variant has its own exponential growth rate parameter ( $\beta_v$ ).

456 **Acknowledgements:** This work was supported by the US National Institutes of Health  
457 Director's Early Independence Award DP5-OD028145 (MJM and JAH).

458 **Competing interests:** MJM is an advisor for Detect, LivePerson and COVID Signals. JAH  
459 and LKS declare no competing interests.

460 **Data availability:** All code to perform the analyses and generate the figures presented in  
461 this article are available under the GNU General Public License version 3  
462 at [https://github.com/jameshay218/variant\\_viral\\_loads](https://github.com/jameshay218/variant_viral_loads).

## 463 References

- 464 1. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 Variants - Clinical, Public Health,  
465 and Vaccine Implications. *N Engl J Med*. 2021;384: 1866–1868.
- 466 2. Gog JR, Grenfell BT. Dynamics and selection of many-strain pathogens. *Proc Natl*  
467 *Acad Sci U S A*. 2002;99: 17209–17214.
- 468 3. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, et al. Global circulation  
469 patterns of seasonal influenza viruses vary with antigenic drift. *Nature*. 2015;523: 217–  
470 220.
- 471 4. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. The global  
472 circulation of seasonal influenza A (H3N2) viruses. *Science*. 2008;320: 340–346.
- 473 5. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics  
474 and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*.  
475 2021;372: 815–821.
- 476 6. Davies NG, Jarvis CI, CMMID COVID-19 Working Group, Edmunds WJ, Jewell NP,  
477 Diaz-Ordaz K, et al. Increased mortality in community-tested cases of SARS-CoV-2  
478 lineage B.1.1.7. *Nature*. 2021;593: 270–274.
- 479 7. Threat Assessment Brief: Implications for the EU/EEA on the spread of the SARS-CoV-  
480 2 Delta (B.1.617.2) variant of concern. 23 Jun 2021 [cited 13 Jul 2021]. Available:  
481 [https://www.ecdc.europa.eu/en/publications-data/threat-assessment-emergence-and-](https://www.ecdc.europa.eu/en/publications-data/threat-assessment-emergence-and-impact-sars-cov-2-delta-variant)  
482 [impact-sars-cov-2-delta-variant](https://www.ecdc.europa.eu/en/publications-data/threat-assessment-emergence-and-impact-sars-cov-2-delta-variant)
- 483 8. [No title]. [cited 13 Jul 2021]. Available:  
484 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachme](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1001358/Variants_of_Concern_VOC_Technical_Briefing_18.pdf)  
485 [nt\\_data/file/1001358/Variants\\_of\\_Concern\\_VOC\\_Technical\\_Briefing\\_18.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1001358/Variants_of_Concern_VOC_Technical_Briefing_18.pdf)
- 486 9. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing  
487 transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. 2021;593: 266–269.
- 488 10. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al.  
489 Emergence and rapid spread of a new severe acute respiratory syndrome-related  
490 coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa.  
491 *medRxiv*. 2020; 2020.12.21.20248640.
- 492 11. Dhar MS, Marwal R, Radhakrishnan VS, Ponnusamy K, Jolly B, Bhoyar RC, et al.  
493 Genomic characterization and Epidemiology of an emerging SARS-CoV-2 variant in  
494 Delhi, India. *medRxiv*. 2021; 2021.06.02.21258076.
- 495 12. Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, et al. Predictive  
496 Modeling of Influenza Shows the Promise of Applied Evolutionary Biology. *Trends*  
497 *Microbiol*. 2018;26: 102–118.
- 498 13. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al.  
499 SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*.  
500 2021;19: 409–424.
- 501 14. MacLean OA, Orton RJ, Singer JB, Robertson DL. No evidence for distinct types in the  
502 evolution of SARS-CoV-2. *Virus Evol*. 2020;6. doi:10.1093/ve/veaa034

- 503 15. Cai Y, Zhang J, Xiao T, Lavine CL, Rawson S, Peng H, et al. Structural basis for  
504 enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science*. 2021.  
505 doi:10.1126/science.abi9745
- 506 16. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, et al.  
507 Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*.  
508 2021;373. doi:10.1126/science.abi5273
- 509 17. Lee LYW, Rozmanowski S, Pang M, Charlett A, Anderson C, Hughes GJ, et al. Severe  
510 Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infectivity by Viral Load, S  
511 Gene Variants and Demographic Factors, and the Utility of Lateral Flow Devices to  
512 Prevent Transmission. *Clin Infect Dis*. 2021 [cited 14 Jul 2021]. doi:10.1093/cid/ciab421
- 513 18. Marks M, Millat-Martinez P, Ouchi D, et al. Transmission of COVID-19 in 282 clusters in  
514 Catalonia, Spain: a cohort study. *Lancet Infect Dis* 2021; 5: 629–36.  
515 [https://doi.org/10.1016/S1473-3099\(21\)00398-4](https://doi.org/10.1016/S1473-3099(21)00398-4).
- 516 19. Bjorkman KK, Saldi TK, Lasda E, Bauer LC, Kovarik J, Gonzales PK, Fink MR, Tat KL,  
517 Hager CR, Davis JC, Ozeroff CD, Brisson GR, Larremore DB, Leinwand LA, McQueen  
518 MB, Parker R. Higher viral load drives infrequent SARS-CoV-2 transmission between  
519 asymptomatic residence hall roommates. *J Infect Dis*. 2021. doi:10.1093/infdis/jiab386.
- 520 20. Kissler SM, Fauver JR, Mack C, Tai CG, Breban MI, Watkins AE, et al. Densely  
521 sampled viral trajectories for SARS-CoV-2 variants alpha (B.1.1.7) and epsilon  
522 (B.1.429). *medRxiv*. 2021; 2021.02.16.21251535.
- 523 21. Calistri P, Amato L, Puglia I, Cito F, Di Giuseppe A, Danzetta ML, et al. Infection  
524 sustained by lineage B.1.1.7 of SARS-CoV-2 is characterised by longer persistence  
525 and higher viral RNA loads in nasopharyngeal swabs. *Int J Infect Dis*. 2021;105: 753–  
526 755.
- 527 22. Müller NF, Wagner C, Frazar CD, Roychoudhury P, Lee J, Moncla LH, et al. Viral  
528 genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State. *Sci Transl*  
529 *Med*. 2021;13. doi:10.1126/scitranslmed.abf0202
- 530 23. Roquebert B, Haim-Boukobza S, Trombert-Paolantoni S, Lecorche E, Verdurme L,  
531 Foulongne V, et al. SARS-CoV-2 variants of concern are associated with lower RT-  
532 PCR amplification cycles between January and March 2021 in France. *medRxiv*.  
533 doi:10.1101/2021.03.19.21253971
- 534 24. Borges V, Sousa C, Menezes L, Gonçalves AM, Picão M, Almeida JP, et al. Tracking  
535 SARS-CoV-2 lineage B.1.1.7 dissemination: insights from nationwide spike gene target  
536 failure (SGTF) and spike gene late detection (SGTL) data, Portugal, week 49 2020 to  
537 week 3 2021. *Eurosurveillance*. 2021. doi:10.2807/1560-7917.es.2021.26.10.2100130
- 538 25. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic  
539 characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in  
540 London, UK: a whole-genome sequencing and hospital-based cohort study. *The Lancet*  
541 *Infectious Diseases*. 2021. doi:10.1016/s1473-3099(21)00170-5
- 542 26. Kidd M, Richter A, Best A, Cumley N, Mirza J, Percival B, et al. S-Variant SARS-CoV-2  
543 Lineage B1.1.7 Is Associated With Significantly Higher Viral Load in Samples Tested by  
544 TaqPath Polymerase Chain Reaction. *J Infect Dis*. 2021;223: 1666–1670.

- 545 27. Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, et al.  
546 Estimating epidemiologic dynamics from cross-sectional viral load distributions.  
547 Science. 2021. doi:eabh0635.
- 548 28. Walker AS, Pritchard E, House T, Robotham JV, Birrell PJ, Bell I, et al. Ct threshold  
549 values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide  
550 variation across populations and over time. *Elife*. 2021;10. doi:10.7554/eLife.64683
- 551 29. Alizon S, Selinger C, Sofonea MT, Haim-Boukobza S, Giannoli J-M, Ninove L, et al.  
552 Epidemiological and clinical insights from SARS-CoV-2 RT-PCR cycle amplification  
553 values. *medRxiv*. 2021; 2021.03.15.21253653.
- 554 30. Gostic KM, Kucharski AJ, Lloyd-Smith JO. Effectiveness of traveller screening for  
555 emerging pathogens is shaped by epidemiology and natural history of infection. *Elife*.  
556 2015;4. doi:10.7554/eLife.05564
- 557 31. Grubaugh ND, Hanage WP, Rasmussen AL. Making Sense of Mutation: What D614G  
558 Means for the COVID-19 Pandemic Remains Unclear. *Cell*. 2020. pp. 794–795.
- 559 32. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral  
560 shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet*  
561 *Microbe*. 2021;2: e13–e22.
- 562 33. Eales O, Page AJ, Tang SN, Walters CE, Wang H, Haw D, et al. SARS-CoV-2 lineage  
563 dynamics in England from January to March 2021 inferred from representative  
564 community samples. doi:10.1101/2021.05.08.21256867
- 565 34. Golubchik T, Lythgoe KA, Hall M, Ferretti L, Fryer HR, MacIntyre-Cockett G, et al. Early  
566 analysis of a potential link between viral load and the N501Y mutation in the SARS-  
567 COV-2 spike protein. doi:10.1101/2021.01.12.20249080
- 568 35. Smith RL, Gibson LL, Martinez PP, Ke R, Mirza A, Conte M, et al. Longitudinal  
569 assessment of diagnostic test performance over the course of acute SARS-CoV-2  
570 infection. *J Infect Dis*. 2021 [cited 16 Jul 2021]. doi:10.1093/infdis/jiab337
- 571 36. Li B, Deng A, Li K, Hu Y, Li Z, Xiong Q, et al. Viral infection and transmission in a large  
572 well-traced outbreak caused by the Delta SARS-CoV-2 variant.  
573 doi:10.1101/2021.07.07.21260122
- 574 37. Rhoads D, Peaper DR, She RC, Nolte FS, Wojewoda CM, Anderson NW, et al. College  
575 of American Pathologists (CAP) Microbiology Committee Perspective: Caution Must Be  
576 Used in Interpreting the Cycle Threshold (Ct) Value. *Clinical infectious diseases: an*  
577 *official publication of the Infectious Diseases Society of America*. 2021. pp. e685–e686.
- 578 38. Cendejas-Bueno E, Romero-Gómez MP, Escosa-García L, Jiménez-Rodríguez S,  
579 Mingorance J, García-Rodríguez J, et al. Lower nasopharyngeal viral loads in pediatric  
580 population. The missing piece to understand SARS-CoV-2 infection in children? *J*  
581 *Infect*. 2021. doi:10.1016/j.jinf.2021.06.009
- 582 39. Borremans B, Gamble A, Prager KC, Helman SK, McClain AM, Cox C, et al.  
583 Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2  
584 infection by time since symptom onset. *Elife*. 2020;9. doi:10.7554/eLife.60122

- 585 40. Kissler SM, Fauver JR, Mack C, Olesen SW, Tai C, Shiue KY, et al. Viral dynamics of  
586 acute SARS-CoV-2 infection and applications to diagnostic and public health strategies.  
587 PLOS Biology. 2021. p. e3001333. doi:10.1371/journal.pbio.3001333
- 588 41. Riley S, Wang H, Eales O, Haw D, Walters CE, Ainslie KEC, et al. REACT-1 round 12  
589 report: resurgence of SARS-CoV-2 infections in England associated with increased  
590 frequency of the Delta variant. medRxiv. 2021; 2021.06.17.21259103.
- 591 42. Decreased infectivity following BNT162b2 vaccination: A prospective cohort study in  
592 Israel. The Lancet Regional Health - Europe. 2021;7: 100150.
- 593 43. Brown KA, Gubbay J, Hopkins J, Patel S, Buchan SA, Daneman N, et al. S-Gene  
594 Target Failure as a Marker of Variant B.1.1.7 Among SARS-CoV-2 Isolates in the  
595 Greater Toronto Area, December 2020 to March 2021. JAMA. 2021. p. 2115.  
596 doi:10.1001/jama.2021.5607
- 597 44. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission  
598 dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020;368: 860–  
599 868.
- 600 45. J. A. Hay, L. Kennedy-Shaffer, jameshay218/virosolver: Publication release, version  
601 v1.0.2, Zenodo (2021); <http://doi.org/10.5281/zenodo.4776812>.
- 602 46. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman,  
603 N. G. Reich, J. Lessler. The incubation period of coronavirus disease 2019 (COVID-19)  
604 from publicly reported confirmed cases: estimation and application. Ann. Int. Med. 172,  
605 577–582 (2020).

# 1        **Viral loads observed under competing strain dynamics**

2

3

## **Supplementary Material**

4

5        **Authors:** James A. Hay<sup>1,2,3\*</sup>, Lee Kennedy-Shaffer<sup>4</sup>, Michael J. Mina<sup>1,2,3,5</sup>

6        <sup>1</sup>Center for Communicable Disease Dynamics, Harvard T H Chan School of Public Health, Boston,  
7        MA.

8        <sup>2</sup>Department of Epidemiology, Harvard T H Chan School of Public Health, Boston, MA.

9        <sup>3</sup>Department of Immunology and Infectious Diseases, Harvard T H Chan School of Public Health,  
10        Boston, MA.

11        <sup>4</sup>Department of Mathematics and Statistics, Vassar College, Poughkeepsie, NY.

12        <sup>5</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School

13        \*Correspondence to: [jhay@hsph.harvard.edu](mailto:jhay@hsph.harvard.edu) (JAH)

## 14 1. SEIR model equations

15 Shown below are the ordinary differential equations for the two-strain susceptible-exposed-  
 16 infected-recovered model. Each equation shows the transition rate for a given state pair. For  
 17 example,  $\{S_1S_2\}$  gives the population who are susceptible to both strain 1 and 2,  $\{S_1E_2\}$  gives  
 18 the population who are susceptible to strain 1 and exposed to strain 2 etc. For concision, all  
 19 individuals infected with a given strain are denoted  $I_v$ , where  $I_1 = \{I_1S_2\} + \{I_1E_2\} + \{I_1I_2\} +$   
 20  $\{I_1R_2\}$  and  $I_2 = \{S_1I_2\} + \{E_1I_2\} + \{I_1I_2\} + \{R_1I_2\}$ . Note that epidemic seeding is also included  
 21 here, where  $|seed < t \leq seed + 7|\kappa$  indicates that exposed individuals are generated at rate  
 22  $\kappa$  within a 7-day window following seeding. Parameters are described in Table S1. Note that

$$23 R_0 = \frac{\beta}{\gamma}.$$

$$24 \frac{d\{S_1S_2\}}{dt} = -\beta_1\{S_1S_2\}I_1 - \beta_2\{S_1S_2\}I_2 - |seed_1 < t \leq seed_1 + 7|\kappa_1 - |seed_2 < t$$

$$25 \leq seed_2 + 7|\kappa_2$$

$$26 \frac{d\{E_1S_2\}}{dt} = \beta_1\{S_1S_2\}I_1 - v\{E_1S_2\} - (1 - \chi)\beta_2\{E_1S_2\}I_2 + |seed_1 < t \leq seed_1 + 7|\kappa_1$$

$$27 \frac{d\{I_1S_2\}}{dt} = v\{E_1S_2\} - \gamma\{I_1S_2\} - (1 - \chi)\beta_2\{I_1S_2\}I_2$$

$$28 \frac{d\{R_1S_2\}}{dt} = \gamma\{I_1S_2\} - (1 - \chi)\beta_2\{R_1S_2\}I_2$$

$$29 \frac{d\{S_1E_2\}}{dt} = \beta_2\{S_1E_2\}I_2 - v\{S_1E_2\} - (1 - \chi)\beta_1\{S_1E_2\}I_1 + |seed_2 < t \leq seed_2 + 7|\kappa_2$$

$$30 \frac{d\{E_1E_2\}}{dt} = -2v\{E_1E_2\} + (1 - \chi)\beta_1\{S_1E_2\}I_1 + (1 - \chi)\beta_2\{E_1S_2\}I_2$$

$$31 \frac{d\{I_1E_2\}}{dt} = v\{E_1E_2\} - v\{I_1E_2\} - \gamma\{I_1E_2\} + (1 - \chi)\beta_2\{I_1S_2\}I_2$$

$$32 \frac{d\{R_1E_2\}}{dt} = -v\{R_1E_2\} + \gamma\{I_1E_2\} + (1 - \chi)\beta_2\{R_1S_2\}I_2$$

$$33 \frac{d\{S_1I_2\}}{dt} = v\{S_1E_2\} - \gamma\{S_1I_2\} - (1 - \chi)\beta_1\{S_1I_2\}I_1$$

$$34 \frac{d\{E_1I_2\}}{dt} = v\{E_1E_2\} - v\{E_1I_2\} - \gamma\{E_1I_2\} + (1 - \chi)\beta_1\{S_1I_2\}I_1$$

$$35 \frac{d\{I_1I_2\}}{dt} = v\{E_1I_2\} + v\{I_1E_2\} - 2\gamma\{I_1I_2\}$$

$$36 \frac{d\{R_1I_2\}}{dt} = \gamma\{I_1I_2\} - \gamma\{R_1I_2\} + v\{R_1E_2\}$$

$$37 \frac{d\{S_1R_2\}}{dt} = \gamma\{S_1I_2\} - (1 - \chi)\beta_1\{S_1R_2\}I_1$$

$$38 \frac{d\{E_1R_2\}}{dt} = \gamma\{E_1I_2\} - v\{E_1R_2\} + (1 - \chi)\beta_1\{S_1R_2\}I_1$$

$$39 \frac{d\{I_1R_2\}}{dt} = v\{E_1R_2\} + \gamma\{E_1I_2\} - \gamma\{I_1R_2\}$$

$$40 \frac{d\{R_1R_2\}}{dt} = \gamma\{R_1I_2\} + \gamma\{I_1R_2\}$$

41

## 42 2. Viral kinetics model

43 To describe modal Ct values  $C_{mode}(a)$  in an individual on each day post infection  $a$ , we used  
 44 a previously described viral kinetics model (main text reference [27]). We reproduce the  
 45 equations here for completeness. Overall, the model is a two-hinge function that describes the  
 46 timing and Ct value of several switch points in the viral kinetics process. After a short period  
 47  $t_e$  of viral loads remaining unchanged at  $C_0$ , the modal Ct value decreases to a minimum of  
 48  $C_p$  over a period of  $t_p$  days. Ct values then increase over  $t_s$  days to plateau at  $C_s$ . From this  
 49 time onward the distribution of Ct values among detectable individuals remains unchanged;  
 50 however, to account for individuals fully recovering and becoming PCR negative, we modeled  
 51 a Bernoulli process with a daily probability  $p_{addl}$  of becoming undetectable on each day after  
 52 the final hinge point ( $a < t_e + t_p + t_s$ ).

$$53 \quad C_{mode}(a) = \begin{cases} C_0, & a \leq t_e \\ C_0 + \frac{C_p - C_0}{t_p}(a - t_e), & t_e < a \leq t_e + t_p \\ C_p + \frac{C_s - C_p}{t_s}(a - t_e - t_p), & t_e + t_p < a \leq t_e + t_p + t_s \\ C_s, & t_e + t_p + t_s < a \end{cases}$$

54 To capture variation in observed Ct values arising from sampling variation and individual-level  
 55 variation in viral kinetics, we assumed that observed Ct values follow a Gumbel distribution  
 56 with location parameter set by the above model:  $C(a) \sim \text{Gumbel}(C_{mode}(a), \sigma(a))$ . The scale  
 57 parameter  $\sigma(a)$  was assumed to shrink over time given by:

$$58 \quad \sigma(a) = \begin{cases} \sigma_{obs}, & a < t_e + t_p + t_s \\ \sigma_{obs} \left[ 1 - \frac{1 - S_{mod}}{t_{mod}}(a - t_e - t_p - t_s) \right], & t_e + t_p + t_s \leq a < t_e + t_p + t_s + t_{mod} \\ \sigma_{obs} S_{mod}, & t_e + t_p + t_s + t_{mod} \leq a \end{cases}$$

59 A key part of the model is the description of the proportion of individuals who remain PCR  
 60 positive on each day post infection. Individual samples can be undetectable in one of two  
 61 ways: (1) having a Ct value drawn from the Gumbel distribution above the limit of detection  
 62  $C_{LOD}$  or (2) fully clearing the infection following the Bernoulli process described above. The  
 63 probability of being detectable on day  $a$  post infection is therefore given by:

$$64 \quad \phi_a = \begin{cases} P[C(a) < C_{LOD}], & a \leq t_e + t_p + t_s \\ P[C(a) < C_{LOD}](1 - p_{addl})^{a - t_e - t_p - t_s}, & a > t_e + t_p + t_s \end{cases}$$

65

### 66 3. Simulating samples under symptom-based surveillance

67 To generate complete line lists for the simulated data underpinning Figure S5 and Figure S7,  
68 we used the SEIR model, viral kinetics model, incubation period distribution and sampling  
69 delay distribution to generate a population of size  $N$  where each individual has a binary  
70 infection state, an infection time, a binary symptomatic state, an incubation period, a sampling  
71 delay, and an observed Ct value. All time-related variables are in days. Note that the incidence  
72 curve is generated only by the SEIR model before individual symptom onset dates are  
73 simulated—individual symptomatic states do not impact transmission dynamics.

74 First, we generated an infection state (1 or 0) for each variant for each individual from a  
75 Bernoulli distribution with probability  $p = \sum_{u=0}^{t_{max}} y_v(u)/N$  where  $y_v(t)$  is the incidence of new  
76 infections with variant  $v$  at time  $t$ . We then simulated an infection time for each infected  
77 individual with each variant  $t_{v,i}$ :

$$78 \quad t_{v,i} \sim \frac{y_v(t)}{\sum_{u=0}^{t_{max}} y_v(u)}$$

79 Next, we generated a symptomatic state for each infected individual from a Bernoulli  
80 distribution with probability  $p = 0.35$ . Symptomatic individuals then have an incubation period  
81 drawn from:

$$82 \quad o_i \sim dlogNormal(\mu, \sigma)$$

83 Where *dlogNormal* is the discretized log-normal distribution with mean  $\mu$  and standard  
84 deviation  $\sigma$ . We set  $\mu = 1.621$  and  $\sigma = 0.418$  based on previous estimates (main text  
85 reference [46]). Individuals then have a sampling delay drawn from:

$$86 \quad d_i \sim dgamma(a, s)$$

87 Where *dgamma* is the discretized gamma distribution with shape parameter  $a$  and scale  
88 parameter  $s$ . We set  $a=5$  or  $7$  and  $s=1$  or  $0.9$  as described in the main text. Infected individuals  
89 have a recovery time after which they are guaranteed to be PCR negative, drawn from a  
90 negative binomial distribution:

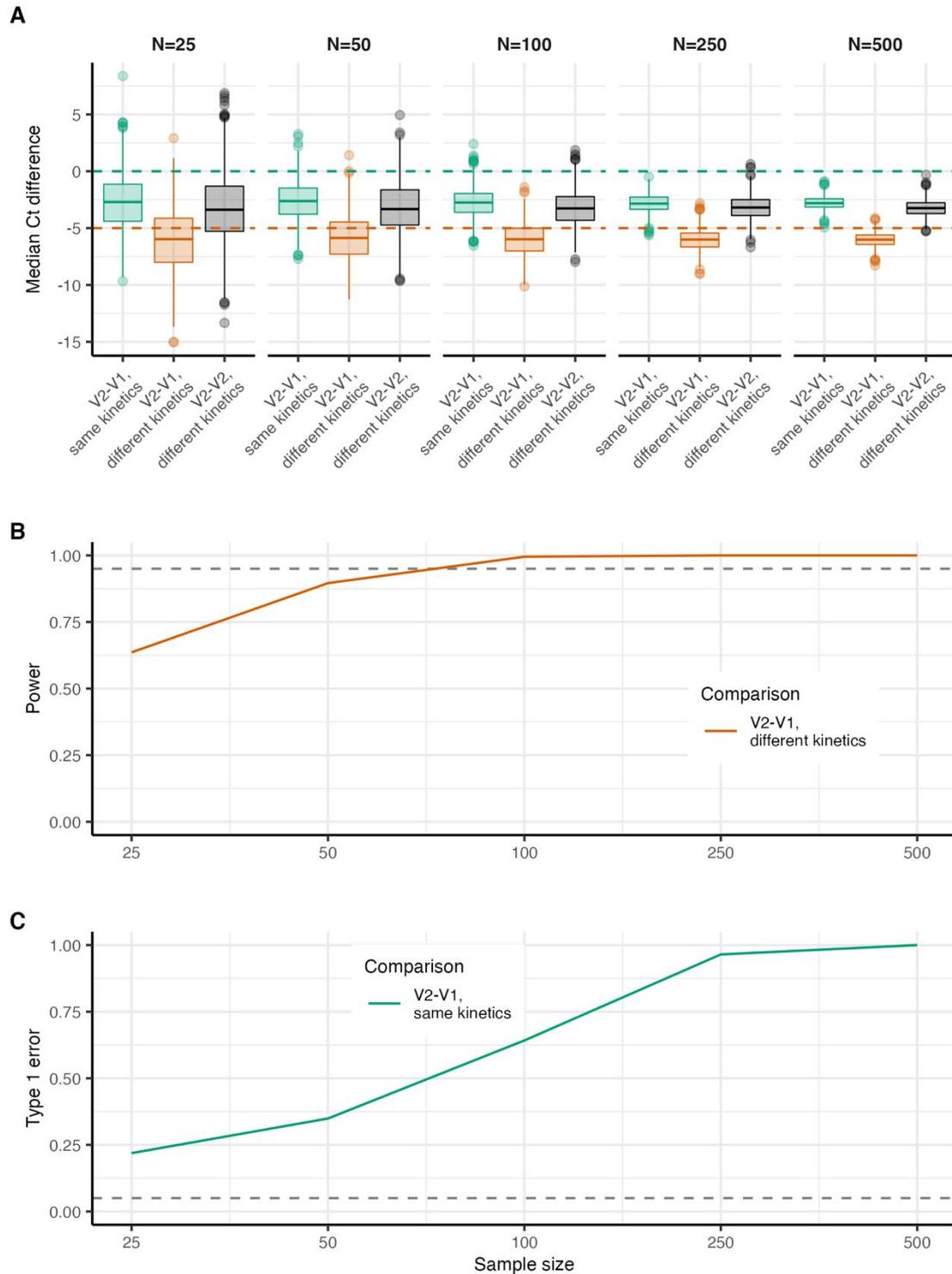
$$91 \quad t^- \sim nbinom(p_{addl}) + t_e + t_p + t_s$$

92 Where *nbinom* is the negative binomial distribution, and the other parameters are described  
93 in Table S2. Finally, we simulate an observed Ct value for each individual  $i$  observed on day  
94  $t_{v,i} + o_i + d_i$  under the model:

$$95 \quad x_i \sim \begin{cases} p_{d_i+o_i}(x), & d_i + o_i < t^- \\ C_{LOD}, & d_i + o_i \geq t^- \end{cases}$$

96 Where  $p_{d_i+o_i}(x)$  is the Gumbel probability density function with location parameter  $C_{mode}(d_i +$   
97  $o_i)$  and scale parameter  $\sigma(d_i + o_i)$ . All  $x_i > C_{LOD}$  are set to  $C_{LOD}$ .

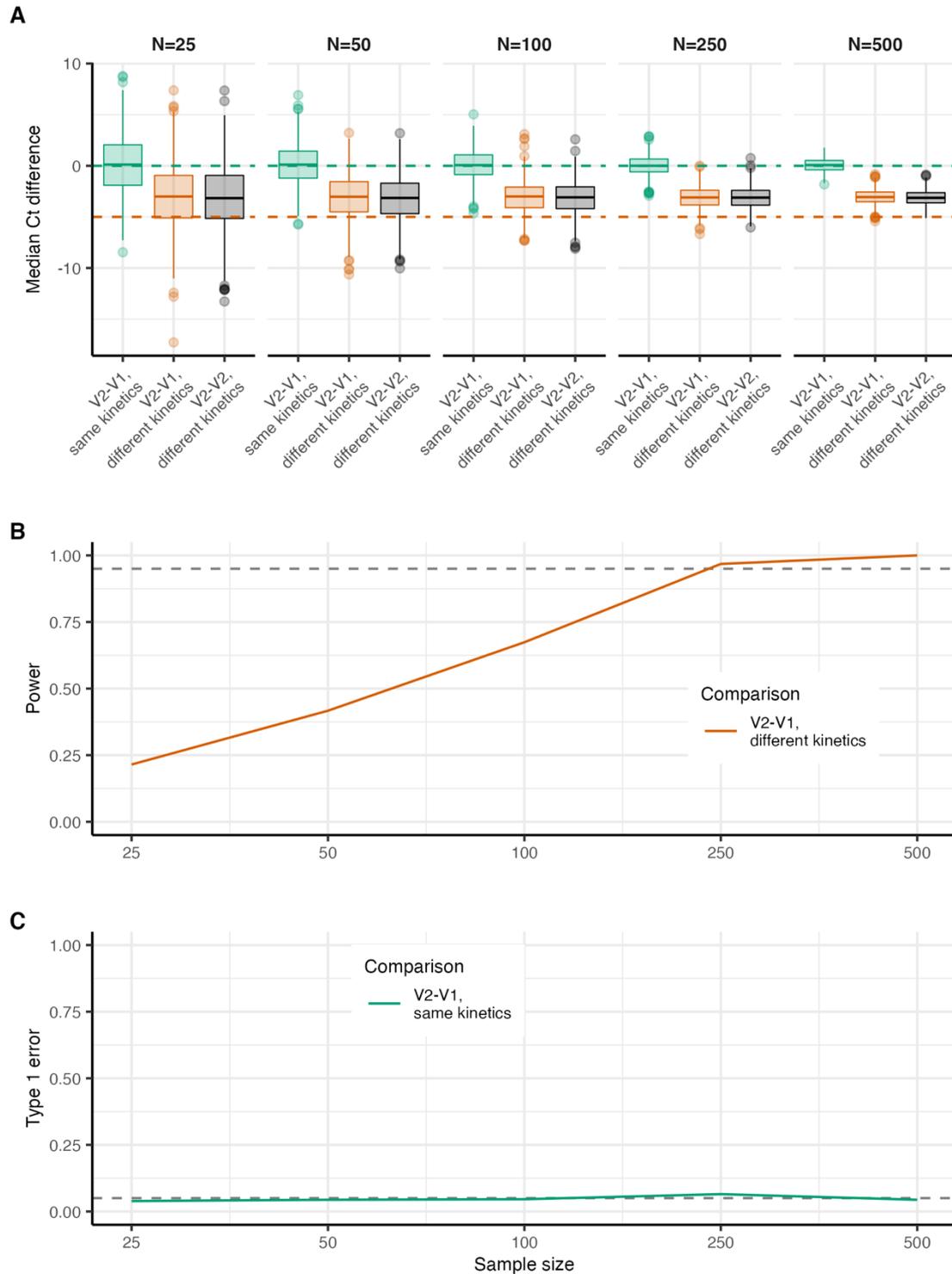
98 To compare Ct values on a given day of the simulation  $t_{samp}$ , we found all individuals in the  
99 line list with detectable Ct values where  $t_{samp} - 3.5 < t_{v,i} + o_i + d_i \leq t_{samp} + 3.5$  (i.e., all  
100 individuals who were sampled in a 7 day window around the chosen time), and then resampled  
101 from these Ct values with replacement to obtain a sample of the specified size (25, 50, 100,  
102 250 or 500).



103

104 **Figure S1. Accuracy and power when comparing Ct values from variants obtained through**  
 105 **random cross-sectional surveillance at day 270 of the simulation shown in Figure 1.** All statistical  
 106 tests are two-sided Wilcoxon rank sum tests. **(A)** Boxplots show the interquartile range (IQR, 75th to  
 107 25th percentile) and median across 1000 simulations for the difference in Ct values when comparing  
 108 the original variant (V1), the new variant (V2) with identical kinetics, or the new variant (V2) with different  
 109 kinetics. Whiskers show 1.5 times the largest and smallest values within 1.5 times the IQR. Dots show  
 110 individual simulations outside 1.5 times the IQR. Horizontal dashed green lines show the true difference

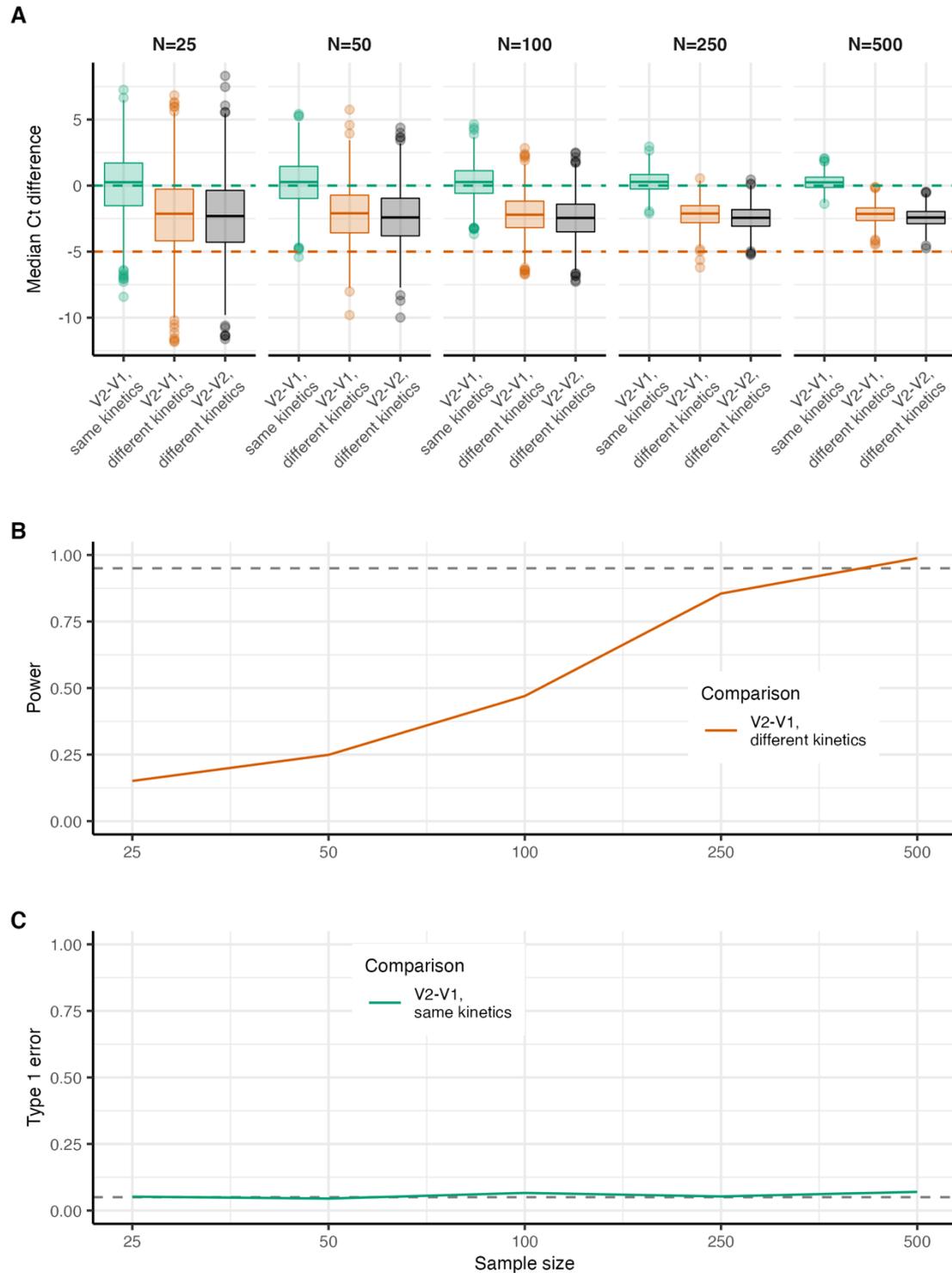
111 in median Ct values between the original variant and the new variant with identical kinetics. Horizontal  
112 dashed orange line shows the difference in peak Ct value assumed for the new variant with different  
113 kinetics relative to the original variant. For an entirely accurate test, the green dots should all align on  
114 the green horizontal line, and the orange and black dots should be identical and negative. **(B)**  
115 Assessment of empirical power to detect a true difference in Ct values at different sample sizes. **(C)**  
116 Probability of type 1 error (incorrectly infer a difference in Ct value) at different sample sizes.



117

118 **Figure S2. Accuracy and power when comparing Ct values from variants obtained through**  
 119 **random cross-sectional surveillance when samples are obtained when both variants have a**  
 120 **growth rate of 0.03 using the simulation shown in Figure 1. All statistical tests are two-sided**  
 121 **Wilcoxon rank sum tests. (A) Boxplots show the interquartile range (IQR, 75th to 25th percentile) and**  
 122 **median across 1000 simulations for the difference in Ct value when comparing the original variant (V1),**  
 123 **the new variant (V2) with identical kinetics, or the new variant (V2) with different kinetics. Whiskers**  
 124 **show 1.5 times the largest and smallest values within 1.5 times the IQR. Dots show individual**

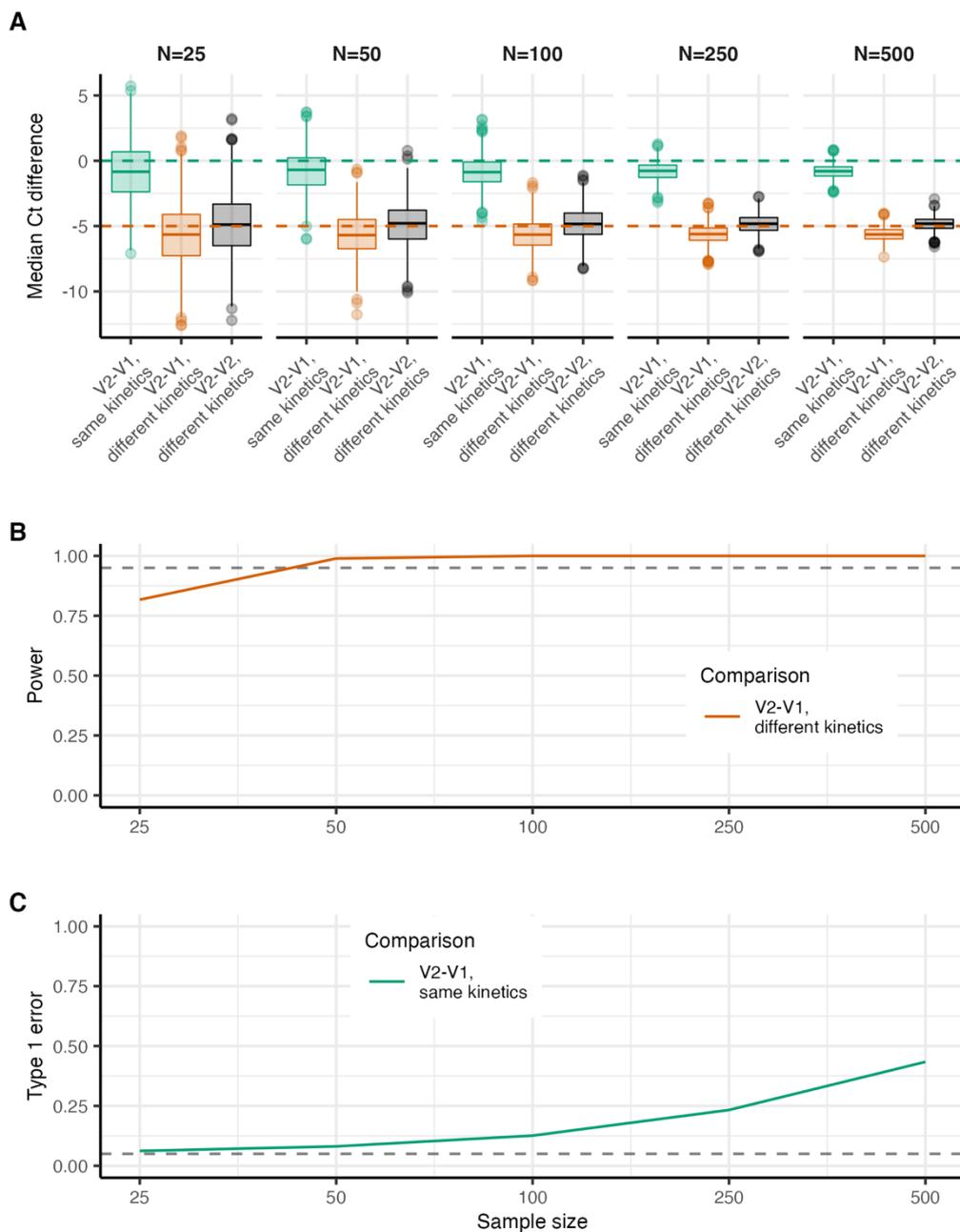
125 simulations outside 1.5 times the IQR. Horizontal dashed green lines show the true difference in median  
126 Ct values between the original variant and the new variant with identical kinetics. Horizontal dashed  
127 orange line shows the difference in peak Ct value assumed for the new variant with different kinetics  
128 relative to the original variant. For an entirely accurate test, the green dots should all align on the green  
129 horizontal line, and the orange and black dots should be identical and negative. **(B)** Assessment of  
130 empirical power to detect a true difference in Ct values at different sample sizes. **(C)** Probability of type  
131 1 error (incorrectly infer a difference in Ct value) at different sample sizes.



132

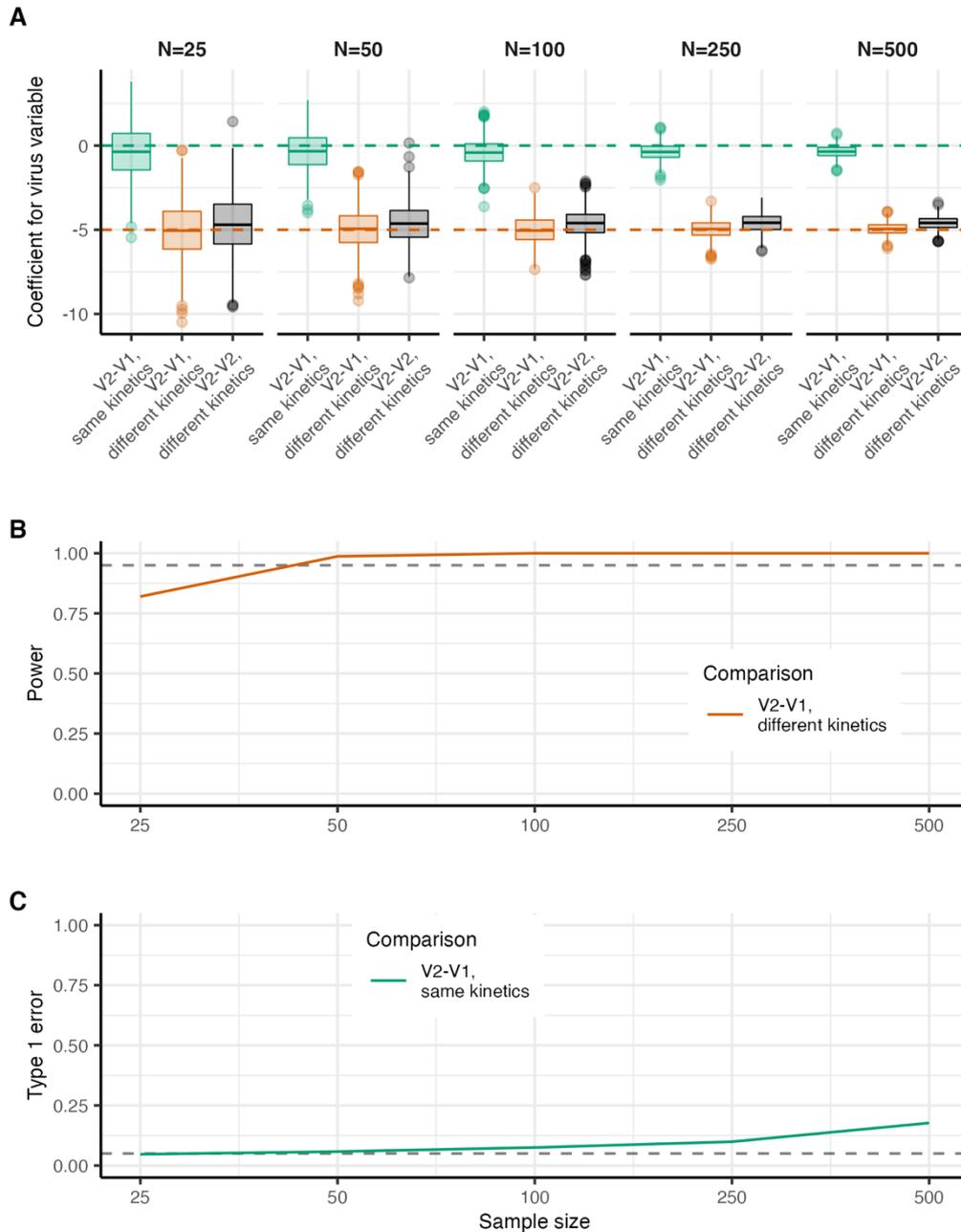
133 **Figure S3. Accuracy and power when comparing Ct values from variants obtained through**  
 134 **random cross-sectional surveillance when samples are obtained when both variants have a**  
 135 **growth rate of -0.02 using the simulation shown in Figure 1. All statistical tests are two-sided**  
 136 **Wilcoxon rank sum tests. (A) Boxplots show the interquartile range (IQR, 75th to 25th percentile) and**  
 137 **median across 1000 simulations for the difference in Ct value when comparing the original variant (V1),**  
 138 **the new variant (V2) with identical kinetics, or the new variant (V2) with different kinetics. Whiskers**  
 139 **show 1.5 times the largest and smallest values within 1.5 times the IQR. Dots show individual**

140 simulations outside 1.5 times the IQR. Horizontal dashed green lines show the true difference in median  
141 Ct values between the original variant and the new variant with identical kinetics. Horizontal dashed  
142 orange line shows the difference in peak Ct value assumed for the new variant with different kinetics  
143 relative to the original variant. For an entirely accurate test, the green dots should all align on the green  
144 horizontal line, and the orange and black dots should be identical and negative. **(B)** Assessment of  
145 empirical power to detect a true difference in Ct values at different sample sizes. **(C)** Probability of type  
146 1 error (incorrectly infer a difference in Ct value) at different sample sizes.



147

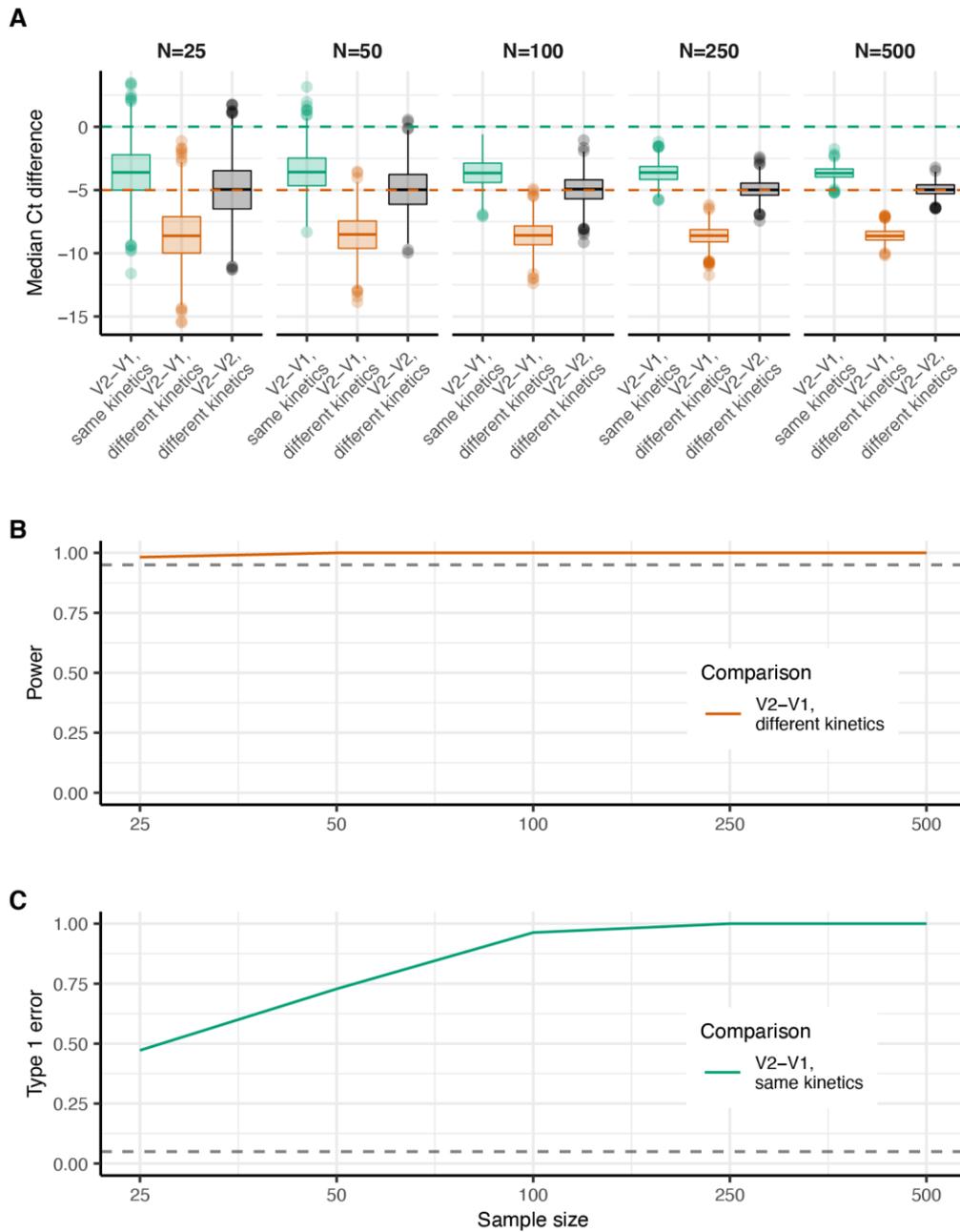
148 **Figure S4. Accuracy and power when comparing Ct values from variants obtained through**  
 149 **symptom-based surveillance when samples are obtained at day 270 of the simulation shown in**  
 150 **Figure 3.** All statistical tests are two-sided Wilcoxon rank sum tests. **(A)** Boxplots show the interquartile  
 151 range (IQR, 75th to 25th percentile) and median across 1000 simulations for the difference in Ct value  
 152 when comparing the original variant (V1), the new variant (V2) with identical kinetics, or the new variant  
 153 (V2) with different kinetics. Whiskers show 1.5 times the largest and smallest values within 1.5 times  
 154 the IQR. Dots show individual simulations outside 1.5 times the IQR. Horizontal dashed green lines  
 155 show the true difference in median Ct values between the original variant and the new variant with  
 156 identical kinetics. Horizontal dashed orange line shows the difference in peak Ct value assumed for the  
 157 new variant with different kinetics relative to the original variant. For an entirely accurate test, the green  
 158 dots should all align on the green horizontal line, and the orange and black dots should be identical and  
 159 negative. **(B)** Assessment of empirical power to detect a true difference in Ct values at different sample  
 160 sizes. **(C)** Probability of type 1 error (incorrectly infer a difference in Ct value) at different sample sizes.



161

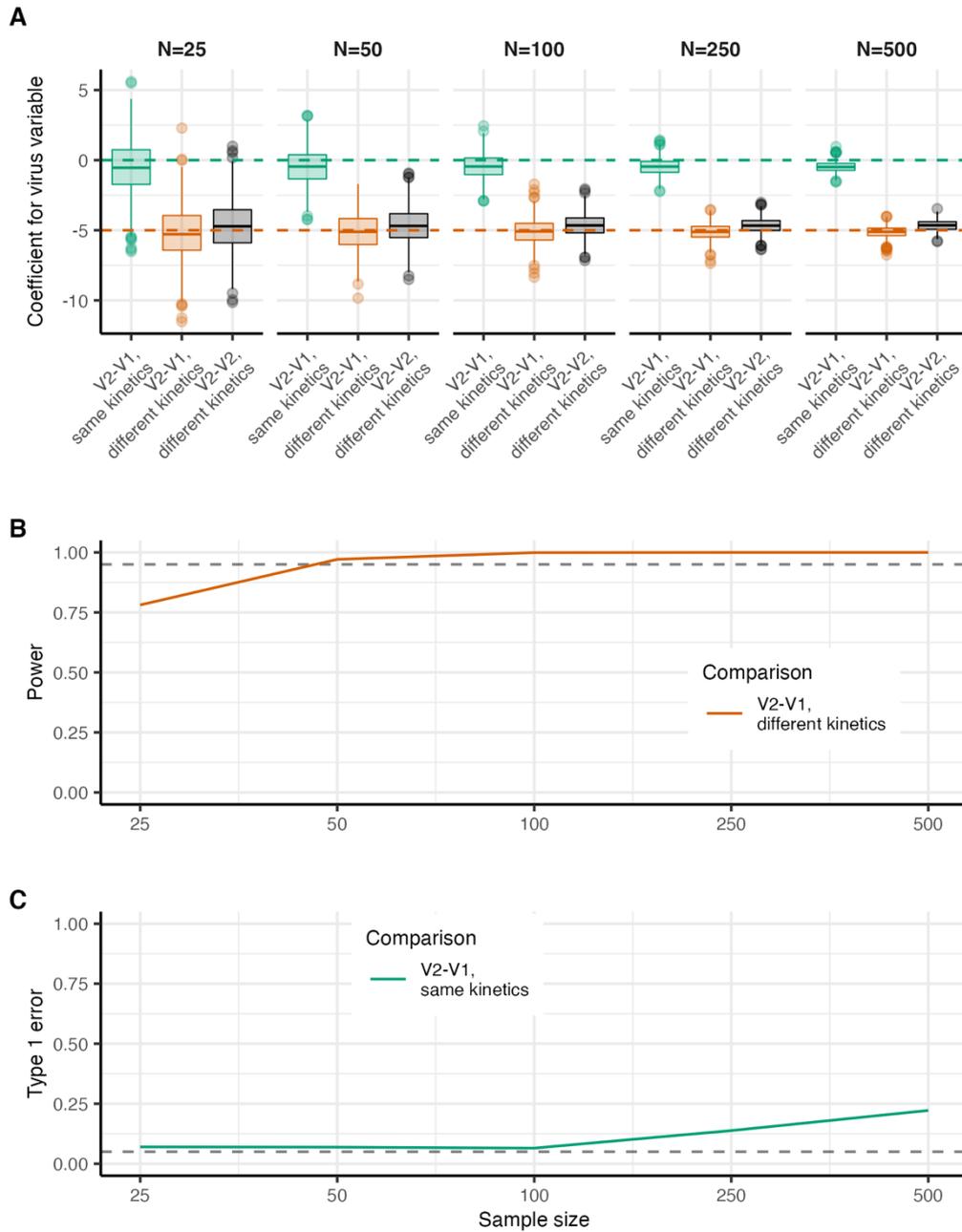
162 **Figure S5. Accuracy and power when comparing Ct values from variants obtained through**  
 163 **symptom-based surveillance when samples are obtained at day 270 of the simulation shown in**  
 164 **Figure 3.** All statistical tests are simple linear regression models as described in Materials and methods.  
 165 (A) Boxplots show the interquartile range (IQR, 75th to 25th percentile) and median across 1000  
 166 simulations of the mean regression coefficient estimate for the effect of variant on difference in Ct value  
 167 when comparing the original variant (V1), the new variant (V2) with identical kinetics, or the new variant  
 168 (V2) with different kinetics. Whiskers show 1.5 times the largest and smallest values within 1.5 times  
 169 the IQR. Dots show individual simulations outside 1.5 times the IQR. Horizontal dashed green lines  
 170 show the true difference in median Ct values between the original variant and the new variant with  
 171 identical kinetics. Horizontal dashed orange line shows the difference in peak Ct value assumed for the  
 172 new variant with different kinetics relative to the original variant. For an entirely accurate test, the green  
 173 dots should all align on the green horizontal line, and the orange and black dots should be identical and

174 negative. **(B)** Assessment of empirical power to detect a true difference in Ct values at different sample  
175 sizes. **(C)** Probability of type 1 error (incorrectly infer a difference in Ct value) at different sample sizes.



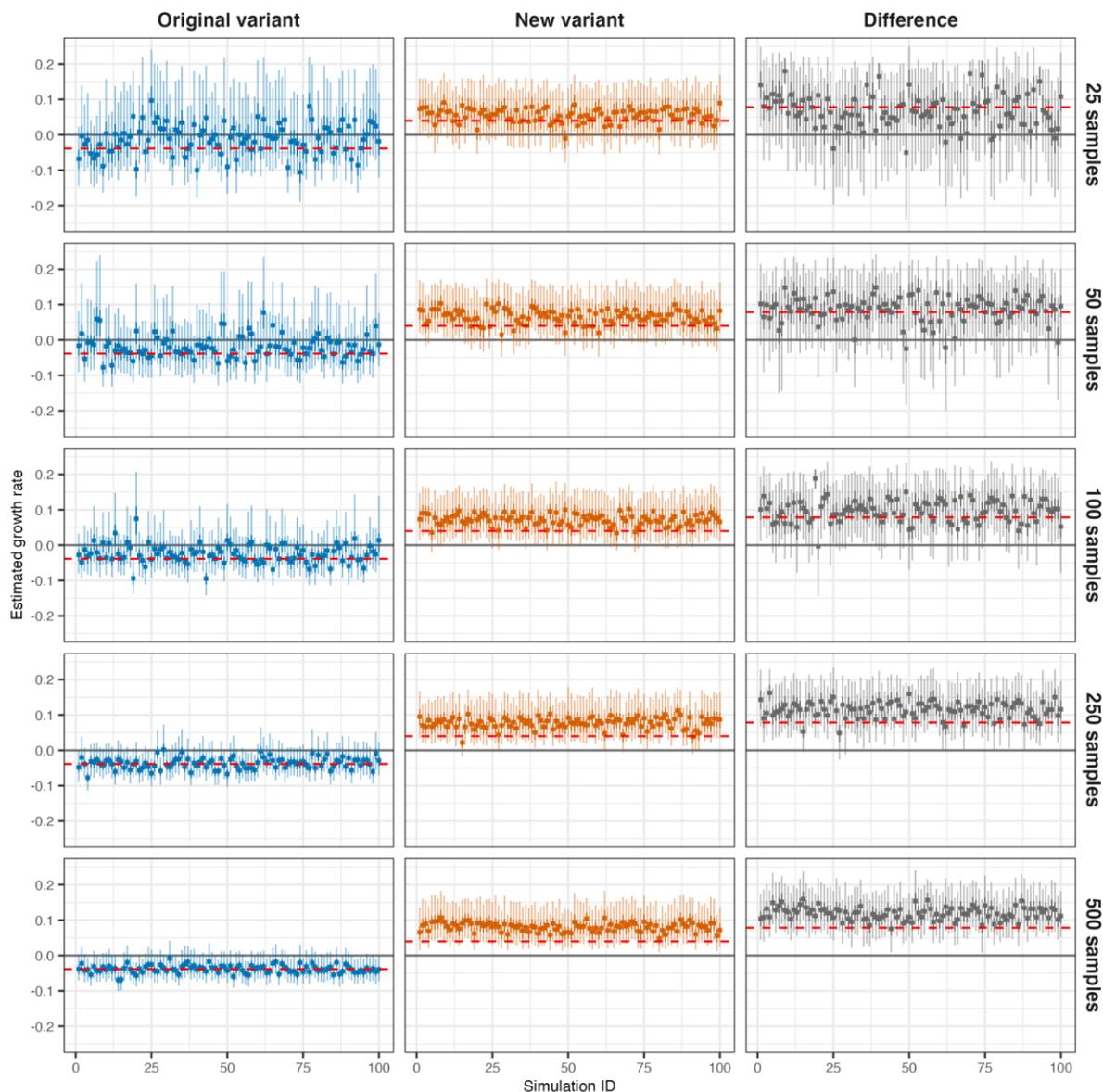
176

177 **Figure S6. Identical to Figure S4, but assuming that the sampling delay distribution of original**  
 178 **variant has a slightly higher mean and standard deviation than the new variant (shape parameter**  
 179 **and shape parameters of 7 and 0.9 respectively vs. 5).**



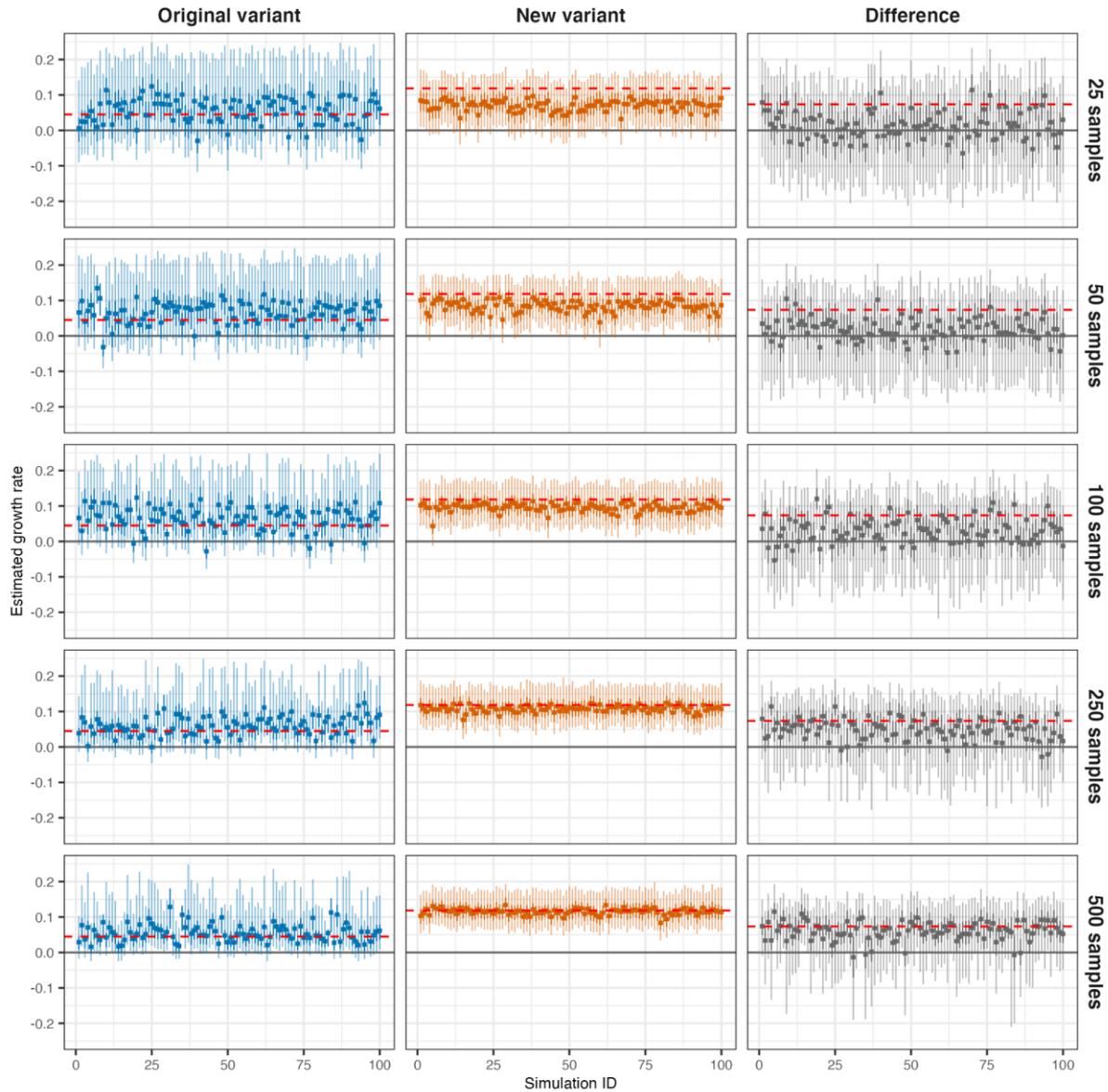
180

181 **Figure S7. Identical to Figure S5, but assuming that the sampling delay distribution of original**  
 182 **variant has a slightly higher mean and standard deviation than the new variant (shape parameter**  
 183 **and shape parameters of 7 and 0.9 respectively vs. 5).**



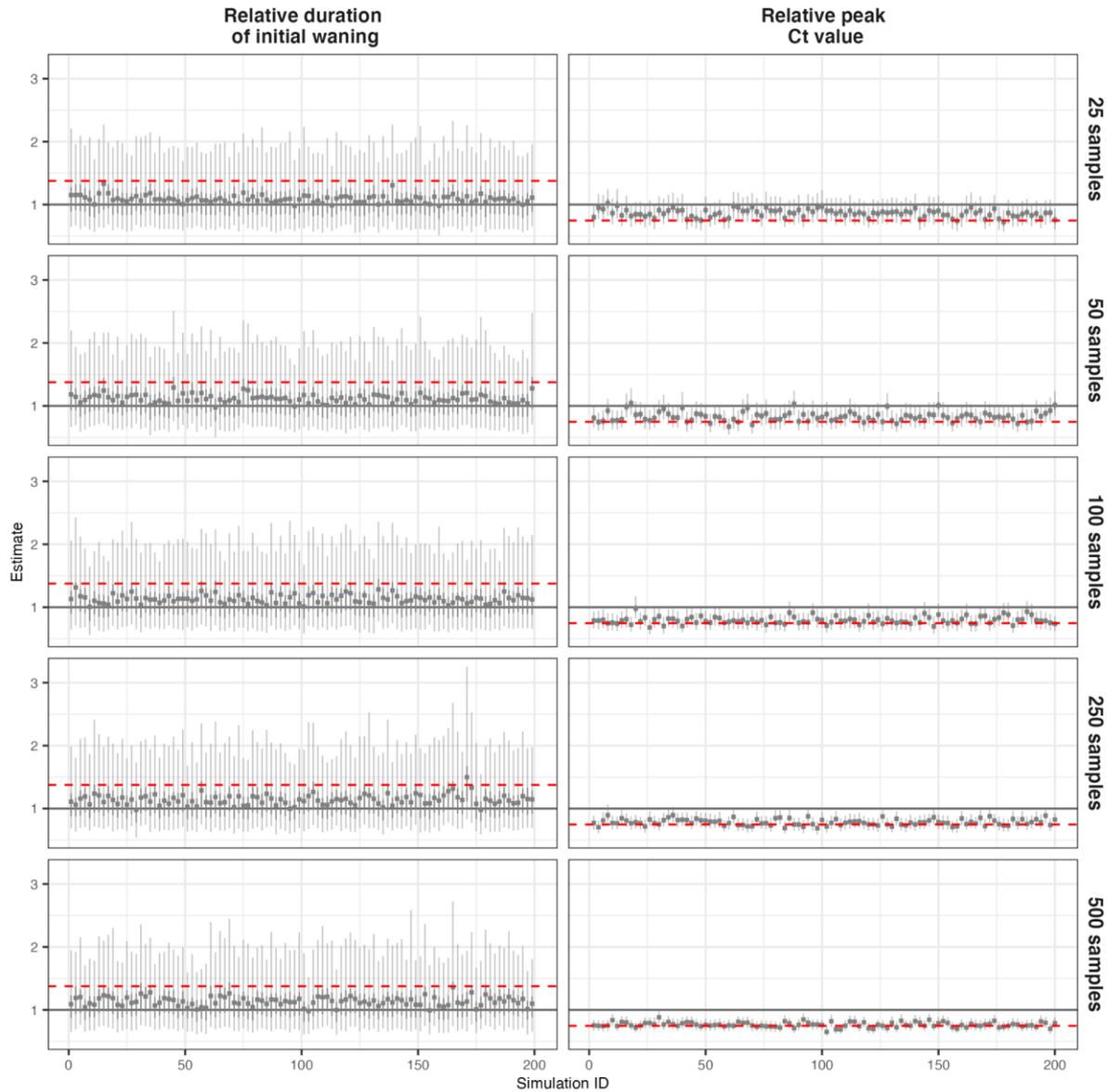
184

185 **Figure S8. Estimated growth rates and variant differences compared to true simulated values,**  
186 **assuming that samples are obtained at day 270 of the simulation shown in Figure 1. The original**  
187 **variant is in epidemic decline, whereas the new variant is in epidemic growth. Rows show**  
188 **increasing sample sizes. Each point-range plot shows the posterior mean (dot), 50% CrI (dark line) and**  
189 **95% CrI (faint line) for a distinct simulation. Dashed red line shows true value in the simulation.**



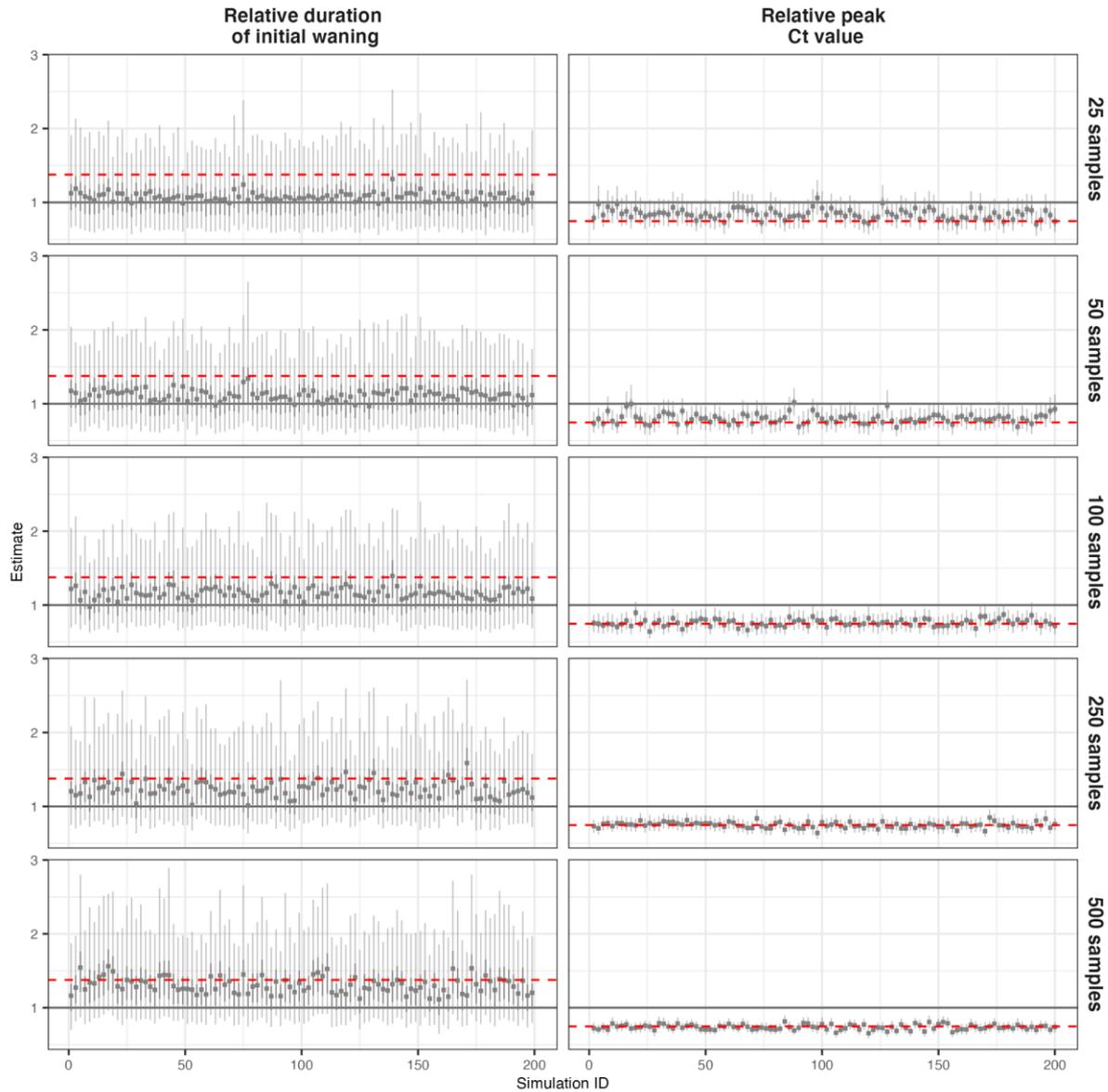
190

191 **Figure S9. Estimated growth rates and variant differences compared to true simulated values,**  
192 **assuming that samples are obtained from a simulation where both variants are increasing**  
193 **simultaneously. The original variant has  $R_0=1.5$ , whereas the new variant has  $R_0=2.5$ . Rows show**  
194 **increasing sample sizes. Each point-range plot shows the posterior mean (dot), 50% CrI (dark line) and**  
195 **95% CrI (faint line) for a distinct simulation. Dashed red line shows true value in the simulation.**



196

197 **Figure S10. Posterior estimates for differences in viral kinetics between the two variants,**  
198 **assuming that samples are obtained at day 270 of the simulation shown in Figure 1. The original**  
199 **variant is in epidemic decline, whereas the new variant is in epidemic growth. Rows show**  
200 **increasing sample sizes. Each point-range plot shows the posterior mean (dot), 50% CrI (dark line) and**  
201 **95% CrI (faint line) for a distinct simulation. Dashed red line shows true value in the simulation.**



202

203 **Figure S11. Posterior estimates for differences in viral kinetics between the two variants,**  
204 **assuming that samples are obtained at day 270 of the simulation shown in Figure 1. The original**  
205 **variant has  $R_0=1.5$ , whereas the new variant has  $R_0=2.5$ . Rows show increasing sample sizes. Each**  
206 **point-range plot shows the posterior mean (dot), 50% CrI (dark line) and 95% CrI (faint line) for a distinct**  
207 **simulation. Dashed red line shows true value in the simulation.**

208 **Table S1: Description of SEIR model parameters**

| <b>Parameter</b> | <b>Description</b>                      | <b>Assumed value</b>           |
|------------------|---|--------------------------------|
| $\beta_1$        | Original variant transmission rate      | $1.5/\gamma$                   |
| $\beta_2$        | New variant transmission rate           | $2.5/\gamma$                   |
| $\kappa_1$       | Original variant importation rate       | 1/100000                       |
| $\kappa_2$       | New variant importation rate            | 1/100000                       |
| $1/\nu$          | Latent period in days                   | 3                              |
| $1/\gamma$       | Infectious period in days               | 7                              |
| $\chi$           | Cross immunity between strains          | 0.75                           |
| $seed_1$         | Time of original variant seeding (days) | 0                              |
| $seed_2$         | Time of new variant seeding (days)      | 0 or 180                       |
| $N$              | Population size                         | 1 (model is solved per capita) |

209

210 **Table S2: Description of viral kinetics model parameters**

| Parameter                         | Description  | Point estimate            | Prior                                |
|-----------------------------------|--|---------------------------|--------------------------------------|
| <b>Viral kinetics model</b>       |  |                           |                                      |
| $t_e$                             | Time from infection to initial viral growth  | 1.00 days                 | Fixed                                |
| $C_0$                             | Ct value at time of infection  | 45.0                      | Fixed                                |
| $t_p$                             | Days from initial viral growth to peak viral load  | 5.00 days                 | Fixed                                |
| $C_p$                             | Modal Ct value at peak viral load  | 19.7                      | Normal(19.7, 2.00)                   |
| $t_s$                             | Time from peak viral load to plateau phase   | 13.3 days                 | Normal(13.3, 3.00)                   |
| $C_s$                             | Modal Ct value at $a = t_{eclipse} + t_{peak} + t_{switch}$  | 38.0                      | Normal(38.0, 1.00)                   |
| $p_{addl}$                        | Daily probability of detectability loss after $t_{switch}$   | 0.103                     | Beta(10.5, 91.2)                     |
| $C_{LOD}$                         | Limit of detection of Ct value   | 40.0                      | Fixed                                |
| $\sigma_{obs}$                    | Initial scale parameter for the Gumbel distribution until $a = t_e + t_p + t_s$  | 5.00                      | Normal(5.00, 0.50)                   |
| $S_{mod}$                         | Multiplicative factor applied to scale parameter for the Gumbel distribution starting at $a = t_{eclipse} + t_{peak} + t_{switch} + t_{scale}$ | 0.789                     | Fixed                                |
| $t_{mod}$                         | Time from secondary waning phase until Gumbel distribution reaches its minimum scale parameter   | 14.0 days                 | Fixed                                |
| <b>New variant viral kinetics</b> |  |                           |                                      |
| $\rho$                            | Peak Ct value of new variant relative to original variant  | 0.747 (5 Ct values lower) | Log-normal(0, 0.5) (on linear scale) |
| $\eta$                            | Time from peak to plateau phase for new variant relative to original variant   | 1.38 (5 days longer)      | Log-normal(0, 0.5) (on linear scale) |
| <b>Exponential growth model</b>   |  |                           |                                      |
| $\beta_1$                         | Original variant exponential growth rate   | NA                        | Normal(0, 0.1)                       |
| $\beta_2$                         | New variant exponential growth rate  | NA                        | Normal(0, 0.1)                       |

211