

# Estimating epidemiologic dynamics from cross-sectional viral load distributions

James A. Hay<sup>1,2,3\*†</sup>, Lee Kennedy-Shaffer<sup>1,2,4\*†</sup>, Sanjat Kanjilal<sup>5,6</sup>, Niall J. Lennon<sup>7</sup>, Stacey B. Gabriel<sup>7</sup>, Marc Lipsitch<sup>1,2,3</sup>, Michael J. Mina<sup>1,2,3,8\*</sup>

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

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

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

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

<sup>5</sup>Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA.

<sup>6</sup>Department of Infectious Diseases, Brigham and Women's Hospital, Boston, MA.

<sup>7</sup>Broad Institute of MIT and Harvard, Cambridge, MA.

<sup>8</sup>Department of Pathology, Brigham and Women's Hospital, Boston, MA.

\*Corresponding authors. Email: [jhay@hsph.harvard.edu](mailto:jhay@hsph.harvard.edu) (JAH), [lkennedyshaffer@vassar.edu](mailto:lkennedyshaffer@vassar.edu) (LKS), [mmina@hsph.harvard.edu](mailto:mmina@hsph.harvard.edu) (MJM).

†These authors contributed equally to this work.

1 **Abstract:** Estimating an epidemic's trajectory is crucial for developing public health responses  
2 to infectious diseases, but incidence data used for such estimation are confounded by variable  
3 testing practices. We show instead that the population distribution of viral loads observed under  
4 random or symptom-based surveillance, in the form of cycle threshold (Ct) values, changes  
5 during an epidemic and that Ct values from even limited numbers of random samples can  
6 provide improved estimates of an epidemic's trajectory. Combining multiple such samples and  
7 the fraction positive improves the precision and robustness of such estimation. We apply our  
8 methods to Ct values from surveillance conducted during the SARS-CoV-2 pandemic in a  
9 variety of settings and demonstrate new approaches for real-time estimates of epidemic  
10 trajectories for outbreak management and response.

## 11 **Main Text:**

12 Real-time tracking of infection incidence during an epidemic is fundamental for public  
13 health planning and intervention (1, 2). In the severe acute respiratory syndrome coronavirus-2  
14 (SARS-CoV-2) pandemic, key epidemiological parameters such as the time-varying effective  
15 reproductive number,  $R_t$ , have typically been estimated using the time-series of observed case  
16 counts, percent of positive tests, or deaths, usually based on reverse-transcription quantitative  
17 polymerase chain reaction (RT-qPCR) testing. However, reporting delays (3), limited testing  
18 capacities, and changes in test availability over time all impact the ability of routine testing to  
19 reliably and promptly detect underlying changes in infection incidence (4, 5). In particular,  
20 whether changes in case counts at different times reflect epidemic dynamics or simply changes in  
21 testing have been major topics of debate with important economic, health and political  
22 ramifications. Here, we describe a new method to overcome these biases and obtain accurate  
23 estimates of the epidemic trajectory, one that does not require repeat measurements and uses  
24 routinely generated but currently discarded quantitative data from RT-qPCR testing from single or  
25 successive cross-sectional samples.

26 RT-qPCR tests provide quantitative results in the form of cycle threshold (Ct) values,  
27 which are inversely correlated with  $\log_{10}$  viral loads, but they are often reported only as binary  
28 “positives” or “negatives” (6, 7). It is common when testing for other infectious diseases to use  
29 this quantification of sample viral load, for example, to identify individuals with higher clinical  
30 severity or transmissibility (8–11). For SARS-CoV-2, Ct values may be useful in clinical  
31 determinations about the need for isolation and quarantine (7, 12), identifying the phase of an  
32 individual’s infection (13, 14) and predicting disease severity (14, 15). However, individual-level  
33 decision making based on Ct values has not yet become a widespread reality due to the variability  
34 in measurements across testing platforms and samples, and limited data to understand SARS-CoV-  
35 2 viral kinetics in asymptomatic and presymptomatic infections. These concerns do not necessarily  
36 hold at the population level: whereas a single high Ct value may not necessarily guarantee a low  
37 viral load in one sample, high Ct values in many samples will indicate a population with  
38 predominantly low viral loads. Indeed, the population-level distribution of Ct values does appear  
39 to change over time. For example, a systematic incline in the distribution of quantified Ct values  
40 has been noted alongside epidemic decline (12, 14, 16).

41 We demonstrate that population-level changes in the distribution of observed Ct values can  
42 arise as an epidemiological phenomenon, and propose methods to use these quantitative values to  
43 estimate epidemic trajectories from one or more cross-sectional samples.

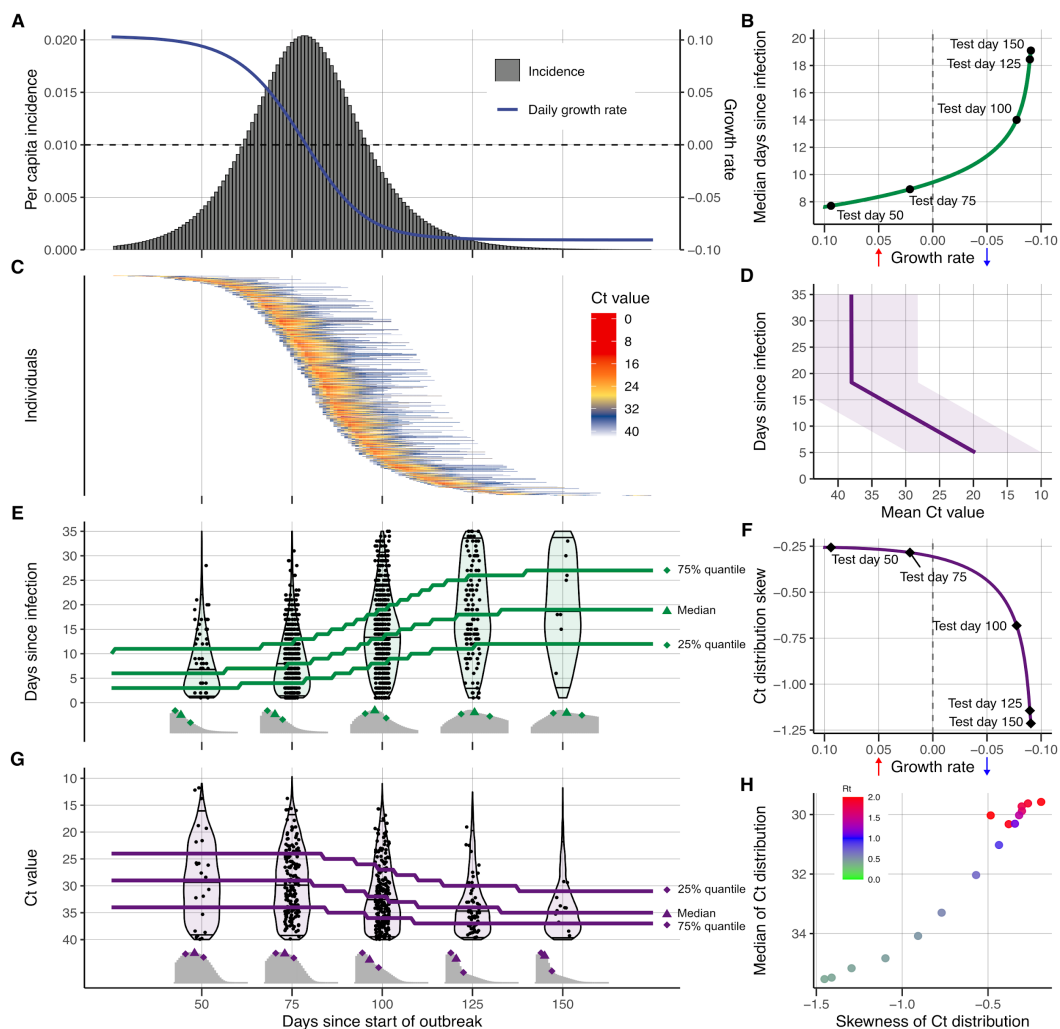
## 44 **Relationship Between Observed Ct Values and Epidemic Dynamics**

45 First, we show that the interaction of within-host viral kinetics and epidemic dynamics can  
46 drive changes in the distribution of Ct values over time without a change in the underlying  
47 pathogen kinetics. To demonstrate the epidemiological link between transmission rate and  
48 measured viral loads or Ct values, we first simulated infections arising under a deterministic  
49 susceptible-exposed-infectious-recovered (SEIR) model (Fig. 1A, *Materials and Methods*:

50 *Simulated Epidemic Transmission Models*). Parameters used are in Table S1. At selected testing  
51 days during the outbreak, simulated Ct values are observed from a random sample of the  
52 population using the Ct distribution model described in *Materials and Methods: Ct Value Model*  
53 and shown in Figs. S1 and S2. By drawing simulated samples for testing from the population at  
54 specific time points, these simulations recreate realistic cross-sectional distributions of detectable  
55 viral loads across the course of an epidemic. Throughout, we assume each individual is infected at  
56 most once, ignoring re-infections as these appear to be a negligible portion of infections in the  
57 epidemic so far (17).

58 Early in the epidemic, infection incidence grows rapidly and the typical infection is thus  
59 recent; as the epidemic wanes, however, the average time since exposure increases as the rate of  
60 new infections decreases (Fig. 1B,E) (18); this is analogous to the average age being lower in a  
61 growing vs. declining population (19). Infections are often unobserved events, but we can rely on  
62 an observable quantity, such as viral load, as a proxy for the time since infection. Since Ct values  
63 change over time within infected hosts (Fig. 1C), random sampling of individuals during epidemic  
64 growth is more likely to measure individuals who were recently infected and therefore in the acute  
65 phase of their infection with higher quantities of viral RNA. Conversely, sampling infected  
66 individuals during epidemic decline is more likely to capture individuals in the convalescent phase,  
67 typically sampling lower quantities of viral RNA (Fig. 1D). The distribution of observed Ct values  
68 therefore changes over time, as measured by the median, quartiles, and skewness (Fig. 1G). While  
69 estimates for an individual's time since infection based on a single Ct value will be highly  
70 uncertain, the population-level distribution of observed Ct values will vary with the growth rate,  
71 and therefore  $R_t$ , of new infections (Fig. 1F,H). Similar principles have been applied to serologic  
72 data to infer unobserved individual-level infection events (16, 20–22) and population-level  
73 parameters of infectious disease spread (20, 23–27).

74 This phenomenon is also present, though less pronounced, among viral loads measured  
75 under symptom-based surveillance (Fig. S3). One might imagine that the typical time since  
76 infection would not depend on the epidemic trajectory in individuals systematically sampled soon  
77 after symptom onset. However, the distribution of delays between infection date and test date is a  
78 convolution of the infection incidence curve and the confirmation delay distribution (time from  
79 infection to testing of symptomatic infections) (28). Individuals tested due to recent symptom onset  
80 are more likely to have been recently infected with a short incubation period during epidemic  
81 growth than during epidemic decline, where more onsets are from older infections with longer  
82 incubation periods. The time-since-infection distribution of individuals tested based on symptom  
83 onset, and therefore their measured viral loads, is influenced by the stage of the epidemic.



84

85 **Fig. 1. The cycle threshold (Ct) value distribution reflects epidemiological dynamics over the**  
 86 **course of an outbreak.** (A) Per capita daily incidence (histogram) and daily growth rate (blue  
 87 line) of new infections in a simulated epidemic using a susceptible-exposed-infectious-recovered  
 88 (SEIR) model. (B) Median days since infection vs. daily growth rate of new infections by epidemic  
 89 day. Labeled points here and in (E–G) show five time points in the simulated epidemic. (C)  
 90 Observed Ct value by day for 500 randomly sampled infected individuals. (D) Viral kinetics model  
 91 (increasing Ct value following peak and subsequent plateau near the limit of detection),  
 92 demonstrating the time course of Ct values (x-axis, line shows mean and ribbon shows 95%  
 93 quantile range) against days since infection (y-axis). Note that the y-axis is arranged to align with  
 94 (E). (E) Distribution of days since infection (violin plots and histograms) for randomly selected  
 95 individuals over the course of the epidemic. Median and first and third quartiles are shown as green  
 96 lines and points. (F) Skewness of observed Ct value distribution vs. daily growth rate of new  
 97 infections by epidemic day. (G) Distribution of observed Ct values (violin plots and histograms)  
 98 among sampled infected individuals by epidemic day. Median and first and third quartile are  
 99 shown as purple lines and points. (H) Time-varying effective reproductive number,  $R_t$ , derived  
 100 from the SEIR simulation, plotted against median and skewness of observed Ct value distribution.

101 By modeling the variation in observed Ct values arising from individual-level viral  
102 growth/clearance kinetics and sampling errors, the distribution of observed Ct values becomes an  
103 estimable function of the times since infection, and the expected median and skewness of Ct values  
104 at a given point in time are then predictable from the growth rate. This function can then be used  
105 to estimate the epidemic growth rate conditional on a set of observed Ct values. The relationship  
106 between observed Ct value and epidemic growth rate holds for any testing approach, though  
107 calibration is needed to define the precise mapping (i.e., using a different RT-qPCR instrument, a  
108 different Ct value threshold, or in a different lab; see Fig. S4).

## 109 **Inferring Epidemic Trajectory Using a Single Cross-Section**

110 From these relationships, we derived a method to formally infer the epidemic growth rate  
111 given a single cross-section of RT-qPCR test results. The method combines two models: (1) the  
112 likelihood of observing a Ct value or negative result conditional on having been infected on a given  
113 day; and (2) the likelihood of being infected on a given day prior to the sample date. For (1), we  
114 used a Bayesian model and defined priors for the mode and range of Ct values following infection  
115 based on the existing literature (*Materials and Methods: Ct Value Model* and *Single Cross-Section*  
116 *Model*). For (2), we initially developed two models to describe the probability of infection over  
117 time: (a) constant exponential growth of infection incidence; or (b) infections arising under an  
118 SEIR model. Both models provide estimates for the epidemic growth rate, but make different  
119 assumptions regarding the possible shape of the outbreak trajectory: the exponential growth model  
120 assumes a constant growth rate, whereas the SEIR model assumes that the growth rate changes  
121 daily depending on the remaining number of susceptible individuals.

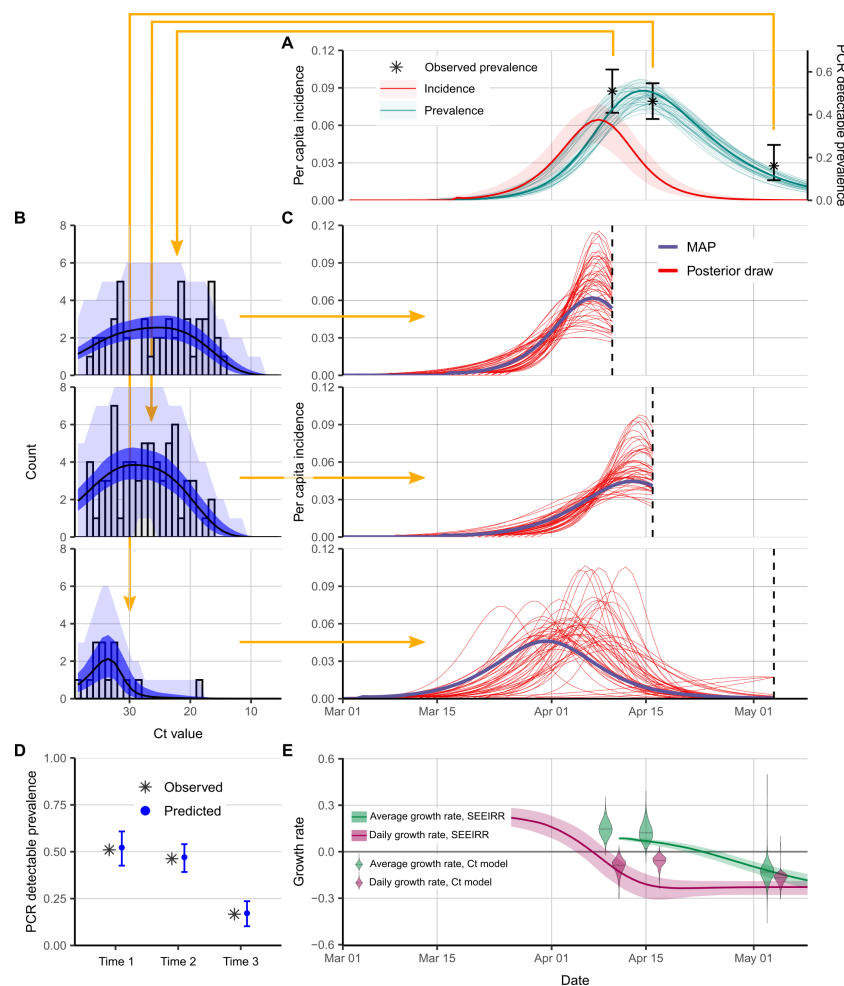
122 We first investigated how the distribution of Ct values and prevalence of PCR positivity  
123 changed over time in four well-observed Massachusetts long-term care facilities that underwent  
124 SARS-CoV-2 outbreaks in March and April 2020 (29). These facilities were relatively closed after  
125 outbreaks began, so we model the outbreak within each facility using an extended SEIR (SEEIRR)  
126 model, with additional exposed and recovered compartments to account for the duration of PCR  
127 positivity (*Materials and Methods: Simulated Epidemic Transmission Models*). In each facility,  
128 we have the results of near-universal PCR testing, including both residents and staff, from three  
129 time points after the outbreak began, including the number of positive samples, the Ct values of  
130 positive samples, and the number of negative samples (*Materials and Methods: Nursing Home*  
131 *Data*). Fig. 2 shows results for one of these facilities, while Fig. S5 shows results for the other  
132 three.

133 In Fig. 2A, we fit the SEEIRR compartmental model to the three observed point prevalence  
134 values from the facility as a benchmark. The distribution of observed Ct values at each time point  
135 (Fig. 2B) shifts higher and becomes more left-skewed at later time points. We then fit the  
136 exponential growth and simple SEIR models using the Ct likelihood to each individual cross-  
137 section to get posterior distributions for the epidemic trajectory up to that time point (Fig. 2C).  
138 Note that these fits do not use any longitudinal data; each is fit to the positive and negative Ct  
139 values from only one time point. To assess the fit, we compare the predicted Ct distribution and

140 point prevalence from each fit to the data (Fig. 2B,D) and compare the growth rates from these fits  
141 to those derived from the fits to the point prevalences. Posterior distributions of all Ct value model  
142 parameters are shown in Fig. S6.

143 While both sets of results are fitted models and so neither can be considered the truth, we  
144 find that the Ct method fit to one cross-section of data provides a similar posterior median  
145 trajectory to the compartmental model fit to three point prevalences. In particular, the Ct-based  
146 models appear to accurately discern whether the samples were taken soon or long after peak  
147 infection incidence. Both methods were in agreement over the direction of the past average and  
148 recent daily growth rates (i.e., whether the epidemic is currently growing or declining, and whether  
149 the growth rate has dropped relative to the historic average). The average growth rate estimates  
150 were very similar at most time points, though the daily growth rate appeared to decline earlier in  
151 the compartmental model. Overall, these results demonstrate that a single cross-section of Ct  
152 values can provide similar information to point prevalence estimates from three distinct sampling  
153 rounds.

154 To ensure that our method provides accurate estimates of the epidemic trajectory, we  
155 performed extensive simulation-recovery experiments using a synthetic nursing home population  
156 undergoing a stochastic SEIR epidemic. We assess performance using various models, including  
157 a version that uses only positive Ct values, and varying parameters of the simulation; details are in  
158 *Materials and Methods: Simulated Nursing Home Outbreaks* and results in Figs. S7–S9.



159

160 **Fig. 2. Single cross-sectional distributions of observed cycle threshold (Ct) values can be used**  
 161 **to reconstruct epidemic trajectories in a Massachusetts nursing home.** (A) Estimated  
 162 prevalence (faint teal lines show posterior samples, solid teal line shows posterior median, teal  
 163 ribbon shows 95% CrI) and incidence (red line shows posterior median, red ribbon shows 95%  
 164 CrI) from the standard compartmental (SEEIRR) model fit to point prevalence at three sampling  
 165 times (error bars show 95% binomial confidence intervals). (B) Model-predicted Ct distributions  
 166 (blue) fitted to the observed Ct values (grey bars) from each of three cross-sectional samples.  
 167 Shown are the posterior median (black line) and 95% CrI for the expected Ct distribution (dark  
 168 blue ribbon), and 95% prediction intervals based on simulated observations (light blue ribbon).  
 169 Note that prediction intervals are much wider than credible intervals, as they result from simulating  
 170 observations with a small sample size. (C) Each panel shows results from fitting the Ct-based SEIR  
 171 model separately to three cross-sections of virologic data. Shown are random posterior samples  
 172 (red lines) and the maximum posterior probability trajectory (purple line) for the incidence curve.  
 173 (D) Ct model-predicted median (blue point) and 95% CrI (blue error bars) for the proportion of  
 174 samples testing positive compared to the observed proportion tested positive (grey cross). (E) 35-  
 175 day (green) and 1-day (magenta) average growth rates from the Ct model estimates in part (C) at  
 176 three time points (violin plots) compared to growth rate estimates from the SEEIRR model in part  
 177 (A) (lines and shaded ribbons).

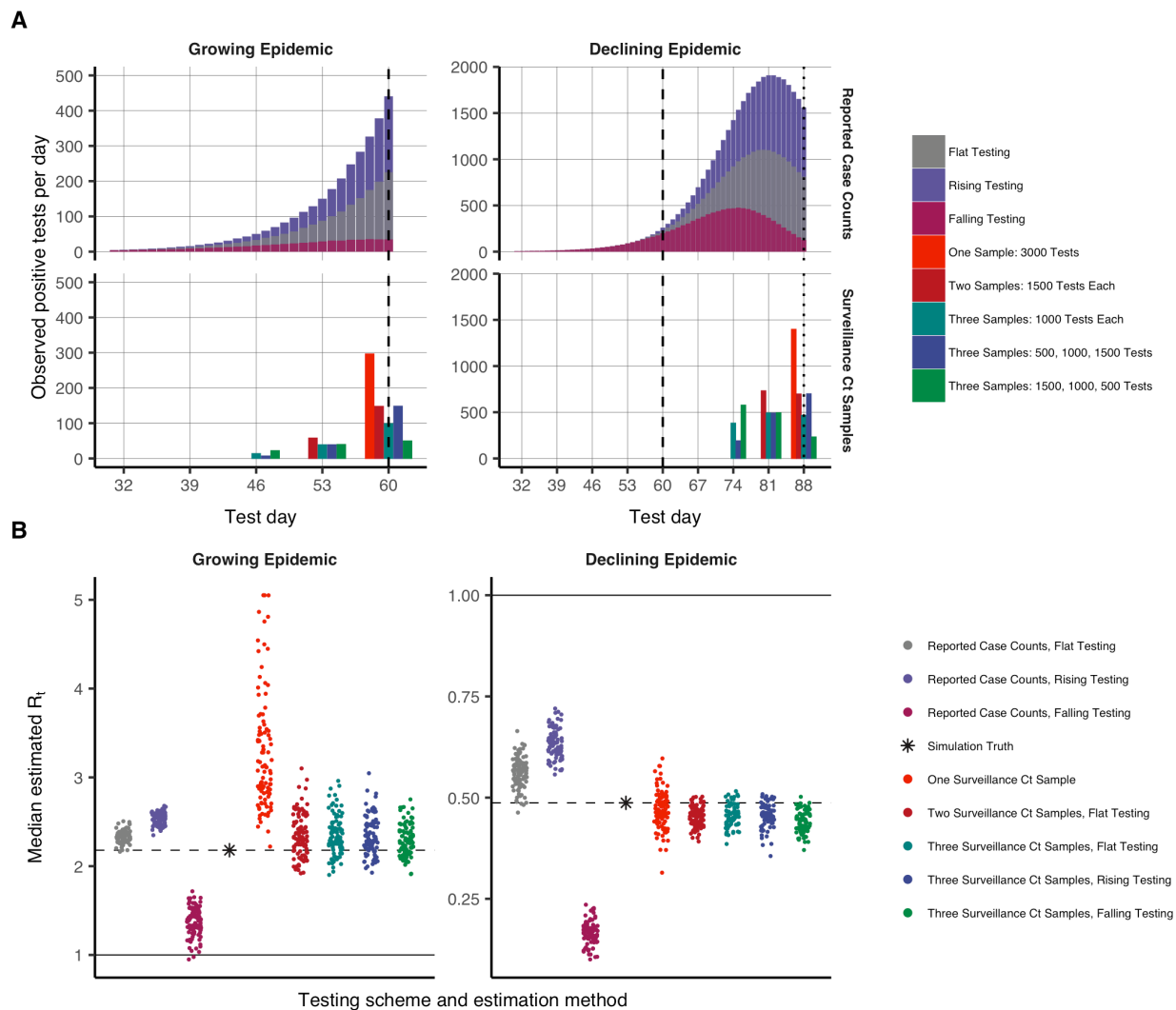
## 178 **Inferring Epidemic Trajectory Using Multiple Cross-Sections**

179 Next, we extended our method to combine data from multiple cross-sections, allowing us  
180 to more reliably estimate the epidemic trajectory (*Materials and Methods: Multiple Cross-Sections*  
181 *Model* and *Markov Chain Monte Carlo Framework*). In many settings, the epidemic trajectory is  
182 monitored using reported case counts, the definition of which can change during the epidemic (30).  
183 Limiting reported cases to positive test results, the number of new positives among the tests  
184 conducted each day can be used to calculate  $R_t$  (3). However, these data represent the growth rate  
185 of positive tests and not the incidence of infection, requiring adjustments to account for changes  
186 in testing capacity, the delay between infection and test report date, and the conversion from  
187 prevalence to incidence. When, instead, Ct values from surveillance sampling is available, our  
188 methods can overcome these limitations by providing a direct mapping between the distribution  
189 of Ct values and infection incidence. Crucially, the Ct-based methods are agnostic to changing  
190 testing rates, providing unbiased growth rate estimates where case count-based methods exhibit  
191 bias (5).

192 To demonstrate the performance of these methods, we use them to recover parameters from  
193 SEIR-based simulations under a variety of testing schemes (*Materials and Methods: Simulated*  
194 *Testing Schemes*). We compare the performance of  $R_t$  estimation using reported case counts via  
195 the R package *EpiNow2* (28, 31), where reporting depends on testing capacity and the symptom  
196 status of infected individuals, to the performance of our methods when one, two, or three  
197 surveillance samples are available with observed Ct values, with a total of about 0.3% of the  
198 population sampled (3000 tests spread among the samples).

199 Figure 3 plots the posterior median  $R_t$  from each of the 100 simulations of each method  
200 when the epidemic is growing (day 60) and declining (day 88). Except when only one sample is  
201 used, the Ct-based methods fitting to an SEIR model exhibit minimal bias, even when testing  
202 capacity changes. Methods based on reported case counts, on the other hand, exhibit noticeable  
203 upward bias when testing rates increase over the period observed and substantial downward bias  
204 when testing rates decrease. The Ct-based methods do exhibit higher variability, however. This is  
205 captured by the Bayesian inference model, as all of the Ct-based methods achieve at least nominal  
206 coverage of the 95% credible intervals among these 100 simulations. The methods based on  
207 reported case counts have coverage below 70% when testing is falling at either time point and  
208 when cases are rising while the epidemic is declining.





209

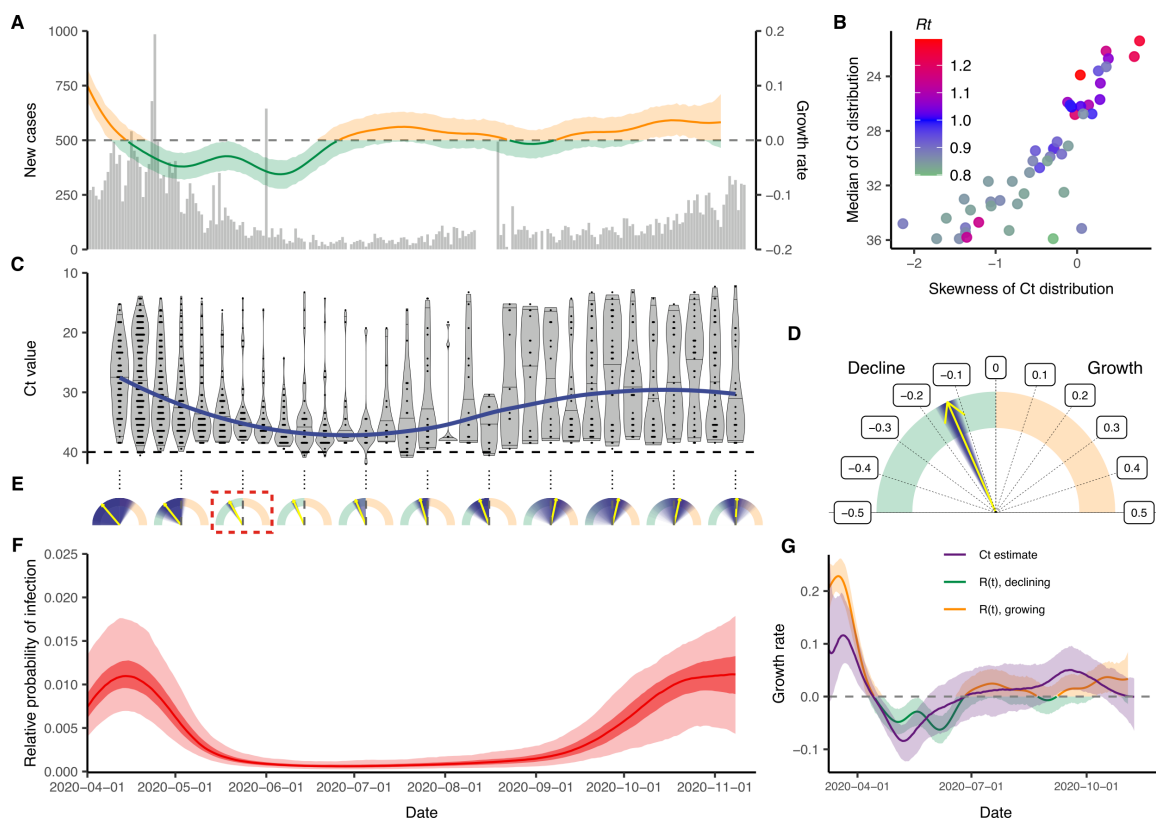
210 **Fig. 3. Inferring epidemic trajectory from cross-sectional surveillance samples with observed**  
 211 **cycle threshold (Ct) values yields nearly unbiased estimates of the time-varying effective**  
 212 **reproductive number,  $R_t$ , whereas changing testing rates lead to biased estimation using**  
 213 **reported case counts. (A) Number of positive tests per day by sampling time in epidemic and**  
 214 **testing scheme for reported case counts (top row) and surveillance Ct sampling (bottom row), from**  
 215 **a simulated susceptible-exposed-infectious-recovered (SEIR) epidemic. Observation times are**  
 216 **shown by vertical lines. (B)  $R_t$  estimates from 100 simulations for each epidemic sampling time,**  
 217 **testing scheme, and estimation method. Each point is the posterior median from a single**  
 218 **simulation.  $R_t$  estimates for reported case counts use *EpiNow2* estimation and for surveillance Ct**  
 219 **samples use the Ct-based likelihood for one or multiple cross-sections fitted to an SEIR model.**  
 220 **True model-based  $R_t$  on the sampling day is indicated by the black star and dashed horizontal line,**  
 221 **while an  $R_t$  of 1, indicating a flat outbreak, is indicated by the solid horizontal line.**

## 222 **Reconstructing Complex Incidence Curves Using Ct Values**

223 Simple epidemic models are useful to understand recent incidence trends when data are  
224 sparse or in relatively closed populations where the epidemic start time is approximately known  
225 (*Materials and Methods: Epidemic Seed Time Priors*). In reality, however, the epidemic usually  
226 follows a more complex trajectory which is difficult to model parametrically. For example, the  
227 SEIR model does not account for the implementation/relaxation of non-pharmaceutical  
228 interventions unless explicitly specified in the model. For a more flexible approach to estimating  
229 the epidemic trajectory from multiple cross-sections, we developed a third model for infection  
230 incidence, using a Gaussian Process (GP) prior for the underlying daily probabilities of infection  
231 (32). The GP method provides estimated daily infection probabilities without making strong  
232 assumptions about the epidemic trajectory, assuming only that infection probabilities on  
233 contemporaneous days are correlated, with decreasing correlation at increasing temporal distances  
234 (*Materials and Methods: Gaussian Process Model*). Movie S1 demonstrates how estimates of the  
235 full epidemic trajectory can be sequentially updated using this model as new samples become  
236 available over time.

237 With the objective of reconstructing the entire incidence curve using routinely collected  
238 RT-qPCR data, we used anonymized, Ct values from positive samples measured from nearly all  
239 hospital admissions into Brigham & Women's Hospital (BWH) in Boston, MA, between April 3  
240 and November 10, 2020 (*Materials and Methods: Brigham & Women's Hospital Data*). We  
241 aligned these with estimates for  $R_t$  based on case counts in Massachusetts (Fig. 4A–C). The median  
242 and skewness of the detectable Ct distribution was correlated with  $R_t$  (Fig. 4B), in line with our  
243 theoretical predictions. Tests taken prior to April 3 were restricted to symptomatic patients only,  
244 while those after April 15 represented near-universal testing of all hospital admissions and non-  
245 admitted ER patients. The median Ct value rose (corresponding to a decline in median viral load)  
246 and skewness of the Ct distribution fell in the late spring and early summer, as shelter-in-place  
247 orders and other non-pharmaceutical interventions were rolled out (Fig. 4C), but the median  
248 declined and skewness rose in late summer and early fall as these measures were relaxed,  
249 coinciding with an increase in observed case counts for the state (Fig. 4A).

250 Using the observed Ct values we estimated the daily growth rate of infections using the  
251 SEIR model on single cross-sections (Fig. 4D,E, Fig. S10, Fig. S11) and the daily relative  
252 probability of infection over time using the GP model (Fig. 4F, Fig. S12). Similar temporal trends  
253 were inferred under both models, and the GP model provided growth rate estimates that followed  
254 those estimated using observed case counts (Fig. 4G). While these data are not strictly a random  
255 sample of the community, and the observed case counts do not necessarily provide a ground truth  
256 for the  $R_t$  value, this demonstrates the ability of this method to re-create epidemic trajectories and  
257 estimate growth or decline of cases using only positive Ct values collected through routine testing.  
258 Interestingly, our estimated epidemic trajectory using only routinely generated Ct values from a  
259 single hospital was remarkably similar to changes in viral loads obtained from wastewater data  
260 (Fig. S13) (33).



261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

**Fig. 4. Single cross-sectional distributions of observed cycle threshold (Ct) values can estimate growth rate and multiple cross-sectional distributions can estimate the complex statewide epidemic trajectory from hospital-based surveillance at Brigham & Women's Hospital in Massachusetts.** (A) Daily confirmed new cases in Massachusetts (gray bars) and estimated time-varying effective reproductive number,  $R_t$ . (B) Estimated  $R_t$  from the case counts vs. median and skewness of observed Ct value distribution by weekly sampling times. (C) Distribution (violin plots and points) and smoothed median (blue line) of observed Ct values by sampling week. (D) Posterior median (yellow arrow) and distribution (blue shaded area) of estimated daily growth rate of incident infections from a susceptible-exposed-infectious-recovered (SEIR) model fit to a single cross-section of observed Ct value data from the week commencing 2020-05-24. Shading density is proportional to posterior density. (E) Posterior medians (yellow arrow) and full distributions (blue shaded area) of estimated daily growth rate of incident infections from SEIR models each fit to a single cross-section by sampling week used. Red box denotes the panel from (D). (F) Posterior distribution of relative probability of infection by date from a Gaussian Process (GP) model fit to all observed Ct values (ribbons show 95% and 50% credible intervals, line shows posterior median). (G) Comparison of estimated daily growth rate of incident infections from GP model (blue line and shaded ribbons show posterior median and 95% CrI) to that from  $R_t$  estimation using observed case counts (red and green line and shaded ribbons show posterior median and 95% CrI) by date. Note that the x-axis is truncated at 2020-04-01, but estimates stretch back to 2020-03-01 (Fig. S13).

## 282 Discussion

283 The usefulness of Ct values for public health decision making is currently the subject of  
284 much discussion and debate. One unexplained observation which has been consistently observed  
285 in many locations is that the distribution of observed Ct values has varied over the course of the  
286 current SARS-CoV-2 pandemic, which has led to questions over whether the fitness of the virus  
287 has changed (12, 14, 16). Our results demonstrate instead that this can be explained as a purely  
288 epidemiologic phenomenon, without any change in individual-level viral dynamics or testing  
289 practices. We find that properties of the population-level Ct distribution strongly correlate with  
290 estimates for the effective reproductive number or growth rate in real-world settings, in line with  
291 our theoretical predictions.

292 Using quantitative diagnostic test results from multiple different tests conducted in a single  
293 cross-sectional survey, Rydevik et al. (18) demonstrated that epidemic trends could be inferred  
294 from virologic data. The methods we describe here use the phenomenon observed in the present  
295 pandemic and the relationship between incidence rate, time since infection, and virologic test  
296 results to estimate a community's position in the epidemic curve, under various models of  
297 epidemic trajectories, based on data from one or more cross-sectional surveys using a single  
298 virologic test. Despite the challenges of sampling variability, individual-level differences in viral  
299 kinetics, and limitations in comparing results from different laboratories or instruments, our results  
300 demonstrate that RT-qPCR Ct values, with all of their quantitative variability for an individual,  
301 can be highly informative of population-level dynamics. This information is lost when  
302 measurements are reduced to binary classifications.

303 Our results demonstrate that this method can be used to estimate epidemic growth rates  
304 based on data collected at a single time point, and independent of assumptions about the intensity  
305 of testing. Comparisons of simulated Ct values and observed Ct values with growth rates and  $R_t$   
306 estimates validate this general approach. Results should be interpreted with caution in cases where  
307 the observed Ct values are not from a population census or a largely random sample, or when there  
308 are very few samples with detectable viral load. When testing is based primarily on the presence  
309 of symptoms or follow-up of contacts of infected individuals, people may be more likely to be  
310 sampled at specific times since infection and thus the distribution of observed Cts would not be  
311 representative of the population as a whole. This method may be most useful in settings where  
312 representative surveillance samples can be obtained independent of COVID-19 symptoms, such  
313 as the REACT study (34). These methods allow municipalities to evaluate and monitor, in real-  
314 time, the role of various epidemic mitigation interventions, for example by conducting even a  
315 single or a small number of random virologic testing samples as part of surveillance.

316 These results are sensitive to the true distribution of observed viral loads each day after  
317 infection. Different swab types, sample types, instruments, or Ct thresholds may alter the  
318 variability in the Ct distribution (15, 16, 35, 36), leading to different relationships between the  
319 specific Ct distribution and the epidemic trajectory. Setting-specific calibrations, for example  
320 based on a reference range of Ct values, will be useful to ensure accuracy. Here, we generated a

321 viral kinetics model based on observed properties of measured viral loads (proportion detectable  
322 over time following symptom onset, distribution of Ct values from true specimens), and used these  
323 results to inform priors on key parameters when estimating growth rates. The growth rate estimates  
324 can therefore be improved by choosing more precise, accurate priors relevant to the observations  
325 used during model fitting. In cases where results come from multiple testing platforms, the model  
326 should either be adjusted to account for this by specifying a different distribution for each platform  
327 based on its properties or, if possible, the Ct values should be transformed to a common scale such  
328 as log viral copies. Results could also be improved if individual-level features that may affect viral  
329 load, such as symptom status, age, and antiviral treatment, are available with the data and  
330 incorporated into the Ct value model (14–16, 37, 38). A similar approach may also be possible  
331 using serologic surveys, as an extension of work that has related time since infection to antibody  
332 titers for other infectious diseases (26, 27). If multiple types of tests (e.g., antigen and PCR) are  
333 conducted at the same time, combining information can substantially reduce uncertainty in these  
334 estimates as well (18). If variant strains are associated with different viral load kinetics and become  
335 common (39), this should be incorporated into the model as well.

336 This method has a number of limitations. While the Bayesian framework incorporates the  
337 uncertainty in viral load distributions into inference on the growth rate, parametric assumptions  
338 and reasonably strong priors on these distributions aid in identifiability. If these parametric  
339 assumptions are violated, inference may not be reliable. In addition, the methods described here  
340 and the relationship between incidence and skewness of Ct distributions become unreliable when  
341 there are very few positive cases, so results should be interpreted with caution and sample sizes  
342 increased in periods with low incidence. In some cases with one or a small number of cross-  
343 sections, the observed Ct distribution could plausibly result from all individuals very early in their  
344 infection at the start of fast epidemic growth, all during the recovery phase of their infection during  
345 epidemic decline, or a mixture of both (Fig. 4E, Fig. S11). We therefore used a parallel tempering  
346 MCMC algorithm which is able to accurately estimate these multimodal posterior distributions.  
347 Interpretation of the estimated median growth rate and credible intervals should be done with  
348 proper epidemiological context: estimated growth rates that are grossly incompatible with other  
349 data can be safely excluded.

350 This method may also overstate uncertainty in the viral load distributions if results from  
351 different machines or protocols are used to inform the prior. A more precise understanding of the  
352 viral load kinetics, and modeling those kinetics in a way that accounts for the epidemiologic and  
353 technical setting of the measurements, will help improve this approach and determine whether Ct  
354 distribution parameters from different settings are comparable. Because of this, quantitative  
355 measures from RT-qPCR should be reported regularly for SARS-CoV-2 cases and early  
356 assessment of pathogen load kinetics should be a priority for future emerging infections. The use  
357 of a control procedure in the measurements, like using the ratio of detected viral RNA to detected  
358 human RNA, could also improve the reliability and comparability of Ct measures.

359 The Ct value is a measurement with magnitude, which provides information on  
360 underlying viral dynamics. Although there are challenges to relying on single Ct values for

361 individual-level decision making, the aggregation of many such measurements from a population  
362 contains substantial information. These results demonstrate how population-level distributions of  
363 Ct values can provide information on important epidemiologic questions of interest, even from a  
364 single cross-sectional survey. Better epidemic planning and more targeted epidemiological  
365 measures can then be implemented based on this survey, or use of Ct values can be combined  
366 with repeated sampling to maximize the use of available evidence.

367 **References**

- 368 1. H. V. Fineberg, M. E. Wilson. Epidemic science in real time. *Science* **324**, 987 (2009).
- 369 2. World Health Organization. Public health surveillance for COVID-19: interim guidance.  
370 <https://www.who.int/publications/i/item/who-2019-nCoV-surveillanceguidance-2020.8>  
371 (Accessed February 12, 2021).
- 372 3. T. Jombart, K. van Zandvoort, T. W. Russell, C. I. Jarvis, A. Gimma, S. Abbott, S. Clifford,  
373 S. Funk, H. Gibbs, Y. Liu, C. A. B. Pearson, N. I. Bosse, Centre for the Mathematical  
374 Modelling of Infectious Diseases COVID-19 Working Group, R. M. Eggo, A. J. Kucharski,  
375 W. J. Edmunds. Inferring the number of COVID-19 cases from recently reported deaths  
376 [version 1; peer review: 2 approved]. *Wellcome Open Res.* **5**, 78 (2020).
- 377 4. M. Lipsitch, D. L. Swerdlow, L. Finelli. Defining the epidemiology of COVID-19—studies  
378 needed. *NEJM* **382**, 1194–1196 (2020).
- 379 5. R. A. Betensky, Y. Feng. Accounting for incomplete testing in the estimation of epidemic  
380 parameters. *Int. J. Epidemiol.* **49**, 1419–1426 (2020).
- 381 6. J. L. Vaerman, P. Saussoy, I. Ingargiola. Evaluation of real-time PCR data. *J. Biol. Regul.*  
382 *Homeost. Agents* **18**, 212–214 (2004).
- 383 7. M. R. Tom, M. J. Mina. To interpret the SARS-CoV-2 test, consider the cycle threshold value.  
384 *Clin. Inf. Dis.* **71**, 2252–2254 (2020).
- 385 8. T. C. Quinn, M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O.  
386 Meehan, T. Lutalo, R. H. Gray. Viral load and heterosexual transmission of Human  
387 Immunodeficiency Virus Type 1. *NEJM* **342**, 921–929 (2000).
- 388 9. J. A. Fuller, M. K. Njenga, G. Bigogo, B. Aura, M. O. Ope, L. Nderitu, L. Wakhule, D. D.  
389 Erdman, R. F. Breiman, D. R. Feikin. Association of the C<sub>T</sub> values of real-time PCR of viral  
390 upper respiratory tract infection with clinical severity, Kenya. *J. Med. Virol.* **85**, 924–932  
391 (2013).
- 392 10. S. Bolotin, S. L. Deeks, A. Marchand-Austin, H. Rilkoff, V. Dang, R. Walton, A. Hashim, D.  
393 Farrell, N. S. Crowcroft. Correlation of Real Time PCR Cycle Threshold Cut-Off with  
394 *Bordetella pertussis* Clinical Severity. *PLoS One* **10**, e0133209 (2015).
- 395 11. T. K. Tsang, B. J. Cowling, V. J. Fang, K. H. Chan, D. K. M. Ip, G. M. Leung, J. S. M. Peiris,  
396 S. Cauchemez. Influenza A virus shedding and infectivity in households. *J. Inf. Dis.* **212**, 1420–  
397 1428 (2015).
- 398 12. M. Moraz, D. Jacot, M. Papadimitriou-Olivgeris, L. Senn, G. Greub, K. Jatón, O. Opota.  
399 Universal admission screening strategy for COVID-19 highlighted the clinical importance of  
400 reporting SARS-CoV-2 viral loads. *New Microbes New Infect.* **38**, 100820 (2020).
- 401 13. Y. Chen, L. Li. SARS-CoV-2: virus dynamics and host response. *Lancet Inf. Dis.* **20**, 515–516  
402 (2020).
- 403 14. D. Jacot, G. Greub, K. Jatón, O. Opota. Viral load of SARS-CoV-2 across patients and  
404 compared to other respiratory viruses. *Microbes Infect.* **22**, 617–621 (2020).

- 405 15. K. A. Walsh, K. Jordan, B. Clyne, D. Rohde, L. Drummond, P. Byrne, S. Ahern, P. G. Carty,  
406 K. K. O'Brien, E. O'Murchu, M. O'Neill, S. M. Smith, M. Ryan, P. Harrington. SARS-CoV-  
407 2 detection, viral load and infectivity over the course of an infection. *J. Infect.* **81**, 357–371  
408 (2020).
- 409 16. A. S. Walker, E. Pritchard, T. House, J. V. Robotham, P. J. Birrell, I. Bell, J. I. Bell, J. N.  
410 Newton, J. Farrar, I. Diamond, R. Studley, J. Hay, K. D. Vihta, T. Peto, N. Stoesser, P. C.  
411 Matthews, D. W. Eyre, K. B. Pouwels, COVID-19 Infection Survey Team. Viral load in  
412 community SARS-CoV-2 cases varies widely and temporally.  
413 <https://www.medrxiv.org/content/10.1101/2020.10.25.20219048v1> (2020).
- 414 17. M. Gousseff, P. Penot, L. Gallay, D. Batisse, N. Benech, K. Bouiller, R. Collarino, A. Conrad,  
415 D. Slama, C. Joseph, A. Lemaigen, F.-X. Lescure, B. Levy, M. Mahevas, B. Pozzetto, N.  
416 Vignier, B. Wyplosz, D. Salmon, F. Goehringer, E. Botelho-Nevers. Clinical recurrences of  
417 COVID-19 symptoms after recovery: viral lapse, reinfection or inflammatory rebound? *J.*  
418 *Infect.* **81**, 816–817 (2020).
- 419 18. G. Rydevik, G. T. Innocent, G. Marion, R. S. Davidson, P. C. L. White, C. Billinis, P. Barrow,  
420 P. P. C. Mertens, D. Gavrier-Widén, M. R. Hutchings. Using combined diagnostic test results  
421 to hindcast trends of infection from cross-sectional data. *PLoS Comput. Biol.* **12**, e1004901  
422 (2016).
- 423 19. J. Wallinga, M. Lipsitch. How generation intervals shape the relationship between growth rates  
424 and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604 (2007).
- 425 20. B. Borremans, N. Hens, P. Beutels, H. Leirs, J. Reijnders. Estimating time of infection using  
426 prior serological and individual information can greatly improve incidence estimation of  
427 human and wildlife infections. *PLoS Comput. Biol.* **12**, e1004882 (2016).
- 428 21. H. Salje, D. A. T. Cummings, I. Rodriguez-Barraquer, L. C. Katzelnick, J. Lessler, C.  
429 Klungthong, B. Thaisomboonsuk, A. Nisalak, A. Weg, D. Ellison, L. Macareo, I. K. Yoon, R.  
430 Jarman, S. Thomas, A. L. Rothman, T. Endy, S. Cauchemez. Reconstruction of antibody  
431 dynamics and infection histories to evaluate dengue risk. *Nature* **557**, 719–723 (2018).
- 432 22. J. A. Hay, A. Minter, K. E. C. Ainslie, J. Lessler, B. Yang, D. A. T. Cummings, A. J. Kucharski,  
433 S. Riley. An open source tool to infer epidemiological and immunological dynamics from  
434 serological data: serosolver. *PLoS Comput. Biol.* **16**, e1007840 (2020).
- 435 23. H. E. de Melker, F. G. A. Versteegh, J. F. P. Schellekens, P. F. M. Teunis, M. Kretzschmar.  
436 The incidence of *Bordetella pertussis* infections estimated in the population from a  
437 combination of serological surveys. *J. Infect.* **53**, 106–113 (2006).
- 438 24. J. Simonsen, K. Mølbak, G. Falkenhorst, K. A. Krogfelt, A. Linneberg, P. F. M. Teunis.  
439 Estimation of incidences of infectious diseases based on antibody measurements. *Stat. Med.*  
440 **28**, 1882–1895 (2009).
- 441 25. P. F. M. Teunis, J. C. H. van Eijkeren, C. W. Ang, Y. T. H. P. van Duynhoven, J. B. Simonsen,  
442 M. A. Strid, W. van Pelt. Biomarker dynamics: estimating infection rates from serological data.  
443 *Stat. Med.* **31**, 2240–2248 (2012).



- 444 26. D. A. Helb, K. K. A. Tetteh, P. L. Felgner, J. Skinner, A. Hubbard, E. Arinaitwe, H. Mayanja-  
445 Kizza, I. Ssewanyana, M. R. Kanya, J. G. Beeson, J. Tappero, D. L. Smith, P. D. Crompton,  
446 P. J. Rosenthal, G. Dorsey, C. J. Drakeley, B. Greenhouse. Novel serologic biomarkers provide  
447 accurate estimates of recent *Plasmodium falciparum* exposure for individuals and  
448 communities. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4438–E4447 (2015).
- 449 27. K. M. Pepin, S. L. Kay, B. D. Golas, S. S. Shriner, A. T. Gilbert, R. S. Miller, A. L. Graham,  
450 S. Riley, P. C. Cross, M. D. Samuel, M. B. Hooten, J. A. Hoeting, J. O. Lloyd-Smith, C. T.  
451 Webb, M. G. Buhnerkempe. Inferring infection hazard in wildlife populations by linking data  
452 across individual and population scales. *Ecol. Lett.* **20**, 275–292 (2017).
- 453 28. K. M. Gostic, L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R.  
454 Niehaus, J. A. Hay, P. M. De Salazar, J. Hellewell, S. Meakin, J. D. Munday, N. I. Bosse, K.  
455 Sherratt, R. N. Thompson, L. F. White, J. S. Huisman, J. Scire, S. Bonhoeffer, T. Stadler, J.  
456 Wallinga, S. Funk, M. Lipsitch, S. Cobey. Practical considerations for measuring the effective  
457 reproductive number,  $R_t$ . *PLoS Comp. Biol.* **16**, e1008409 (2020).
- 458 29. N. J. Lennon, R. P. Bhattacharyya, M. J. Mina, H. L. Rehm, D. T. Hung, S. Smole, A. Woolley,  
459 E. S. Lander, S. B. Gabriel. Comparison of viral loads in individuals with or without symptoms  
460 at time of COVID-19 testing among 32,480 residents and staff of nursing homes and assisted  
461 living facilities in Massachusetts.  
462 <https://www.medrxiv.org/content/10.1101/2020.07.20.20157792v1> (2020).
- 463 30. T. K. Tsang, P. Wu, Y. Lin, E. H. Y. Lau, G. M. Leung, B. J. Cowling. Effect of changing case  
464 definitions for COVID-19 on the epidemic curve and transmission parameters in mainland  
465 China: a modelling study. *Lancet Public Health* **5**, e289–296 (2020).
- 466 31. S. Abbott, J. Hellewell, R. N. Thompson, K. Sherratt, H. P. Gibbs, N. I. Bosse, J. D. Munday,  
467 S. Meakin, E. L. Doughty, J. Y. Chun, Y. W. D. Chan, F. Finger, P. Campbell, A. Endo, C. A.  
468 B. Pearson, A. Gimma, T. Russell, CMMID COVID modelling group, S. Flasche, A. J.  
469 Kucharski, R. M. Eggo, S. Funk. Estimating the time-varying reproduction number of SARS-  
470 CoV-2 using national and subnational case counts [version 2; peer review: awaiting peer  
471 review]. *Wellcome Open Res.* **5**, 112 (2020).
- 472 32. X. Xu, T. Kypraios, P. D. O’Neill. Bayesian non-parametric inference for stochastic epidemic  
473 models using Gaussian processes. *Biostatistics* **17**, 619–633 (2016).
- 474 33. Massachusetts Water Resources Authority. Wastewater COVID-19 Tracking.  
475 <https://www.mwra.com/biobot/biobotdata.htm> (Accessed January 7, 2021).
- 476 34. S. Riley, C. Atchison, D. Ashby, C. A. Donnelly, W. Barclay, G. Cooke, H. Ward, A. Darzi,  
477 P. Elliott, REACT Study Group. REal-time Assessment of Community Transmission  
478 (REACT) of SARS-CoV-2 virus: study protocol [version 1; peer review: 1 approved, 1  
479 approved with reservations]. *Wellcome Open Res.* **5**, 200 (2020).
- 480 35. R. Niehus, E. van Kleef, A. Turlej-Rogacka, C. Lammens, Y. Carmeli, H. Goossens, E.  
481 Tacconelli, B. Carevic, L. Preotescu, S. Malhotra-Kumar, B. S. Cooper. Quantifying antibiotic  
482 impact on within-patient dynamics of extended-spectrum beta-lactamase resistance. *eLife* **9**,  
483 e49206 (2020).

- 484 36. D. Rhoads, D. R. Peaper, R. C. She, F. S. Nolte, C. M. Wojewoda, N. W. Anderson, B. S. Pritt.  
485 College of American Pathologists (CAP) Microbiology Committee perspective: caution must  
486 be used in interpreting the cycle threshold (Ct) value. *Clin. Infect. Dis.* 10.1093/cid/ciaa1199  
487 (2020).
- 488 37. A. T. Xiao, Y. X. Tong, S. Zhang. Profile of RT-PCR for SARS-CoV-2: a preliminary study  
489 from 56 COVID-19 patients. *Clin. Inf. Dis.* **71**, 2249–2251 (2020).
- 490 38. W. C. Ko, J. M. Rolain, N. Y. Lee, P. L. Chen, C. T. Huang, P. I. Lee, P. R. Hsueh. Arguments  
491 in favour of remdesivir for treating SARS-CoV-2 infections. *Int. J. Antimicrob. Agents* **55**,  
492 105933 (2020).
- 493 39. E. Mahase. COVID-19: what have we learnt about the new variant in the UK? *BMJ* **371**, m4944  
494 (2020).
- 495 40. D. B. Larremore, B. Wilder, E. Lester, S. Shehata, J. M. Burke, J. A. Hay, M. Tambe, M. J.  
496 Mina, R. Parker. Test sensitivity is secondary to frequency and turnaround time for COVID-  
497 19 screening. *Sci. Adv.* **7**, eabd5393 (2021).
- 498 41. A. Tahamtan, A. Ardebili. Real-time RT-PCR in COVID-19 detection: issues affecting the  
499 results. *Expert Rev. Mol. Diagn.* **20**, 453–454 (2020).
- 500 42. K. K. W. To, O. T. Y. Tsang, W. S. Leung, A. R. Tam, T. C. Wu, D. C. Lung, C. C. Y. Yip, J.  
501 P. Cai, J. M. C. Chan, T. S. H. Chik, D. P. L. Lau, C. Y. C. Choi, L. L. Chen, W. M. Chan, K.  
502 H. Chan, J. D. Ip, A. C. K. Ng, R. W. S. Poon, C. T. Luo, Vi. C. C. Cheng, J. F. W. Chan, I. F.  
503 N. Hung, Z. Chen, H. Chen, K. Y. Yuen. Temporal profiles of viral load in posterior  
504 oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-  
505 2: an observational cohort study. *Lancet Inf. Dis.* **20**, 565–574 (2020).
- 506 43. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer,  
507 T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R.  
508 Ehmann, K. Zwirgmaier, C. Drosten, C. Wendtner. Virological assessment of hospitalized  
509 patients with COVID-19. *Nature* **581**, 465–469 (2020).
- 510 44. Q. Z. Long, X. J. Tang, Q. L. Shi, Q. Li, H. J. Deng, J. Yuan, J. L. Hu, W. Xu, Y. Zhang, F. J.  
511 Lv, K. Su, F. Zhang, J. Gong, B. Wu, X. M. Liu, J. J. Li, J. F. Qiu, J. Chen, A. L. Huang.  
512 Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature*  
513 *Med.* **26**, 1200–1204 (2020).
- 514 45. Y. Liu, L. M. Yan, L. Wan, T. X. Xiang, A. Le, J. M. Liu, M. Peiris, L. L. M. Poon, W. Zhang.  
515 Viral dynamics in mild and severe cases of COVID-19. *Lancet Inf. Dis.* **20**, 656–657 (2020).
- 516 46. A. Chandrashekar, J. Liu, A. J. Martinot, K. McMahan, N. B. Mercado, L. Peter, L. H.  
517 Tostanoski, J. Yu, Z. Maliga, M. Nekorchuk, K. Busman-Sahay, M. Terry, L. M. Wrijil, S.  
518 Ducat, D. R. Martinez, C. Atyeo, S. Fischinger, J. S. Burke, M. D. Slein, L. Pessaint, A. Van  
519 Ry, J. Greenhouse, T. Taylor, K. Blade, A. Cook, B. Finneyfrock, R. Brown, E. Teow, J.  
520 Velasco, R. Zahn, F. Wegmann, P. Abbink, E. A. Bondzie, G. Dagotto, M. S. Gebre, X. He,  
521 C. Jacob-Dolan, N. Kordana, Z. Li, M. A. Lifton, S. H. Mahrokhian, L. F. Maxfield, R.  
522 Nityanandam, J. P. Nkolola, A. G. Schmidt, A. D. Miller, R. S. Baric, G. Alter, P. K. Sorger,

- 523 J. D. Estes, H. Andersen, M. G. Lewis, D. H. Barouch. SARS-CoV-2 infection protects against  
524 rechallenge in rhesus macaques. *Science* **369**, 812–817 (2020).
- 525 47. J. Yu, L. H. Tostanoski, L. Peter, N. B. Mercado, K. McMahan, S. H. Mahrokhian, J. P.  
526 Nkolola, J. Liu, Z. Li, A. Chandrashekar, D. R. Martinez, C. Loos, C. Atyeo, S. Fischinger, J.  
527 S. Burke, M. D. Slein, Y. Chen, A. Zuiani, F. J. N. Lelis, M. Travers, S. Habibi, L. Pessaint,  
528 A. Van Ry, K. Blade, R. Brown, A. Cook, B. Finneyfrock, A. Dodson, E. Teow, J. Velasco,  
529 R. Zahn, F. Wegmann, E. A. Bondzie, G. Dagotto, M. S. Gebre, X. He, C. Jacob-Dolan, M.  
530 Kirilova, N. Kordana, Z. Lin, L. F. Maxfield, F. Nampanya, R. Nityanandam, J. D. Ventura,  
531 H. Wan, Y. Cai, B. Chen, A. G. Schmidt, D. R. Wesemann, R. S. Baric, G. Alter, H. Andersen,  
532 M. G. Lewis, D. H. Barouch. DNA vaccine protection against SARS-CoV-2 in rhesus  
533 macaques. *Science* **369**, 806–811 (2020).
- 534 48. S. F. Sia, L. M. Yan, A. W. H. Chin, K. Fung, K. T. Choy, A. Y. L. Wong, P. Kaewpreedee,  
535 R. A. P. M. Perera, L. L. M. Poon, J. M. Nicholls, M. Peiris, H. L. Yen. Pathogenesis and  
536 transmission of SARS-CoV-2 in golden hamsters. *Nature* **583**, 834–838 (2020).
- 537 49. A. M. Bosco-Lauth, A. E. Hartwig, S. M. Porter, P. W. Gordy, M. Nehring, A. D. Byas, S.  
538 VandeWoude, I. K. Ragan, R. M. Maison, R. A. Bowen. Experimental infection of domestic  
539 dogs and cats with SARS-CoV-2: pathogenesis, transmission, and response to reexposure in  
540 cats. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26382–26388 (2020).
- 541 50. M. M. Arons, K. M. Hatfield, S. C. Reddy, A. Kimball, A. James, J. R. Jacobs, J. Taylor, K.  
542 Spicer, A. C. Bardossy, L. P. Oakley, S. Tanwar, J. W. Dyal, J. Harney, Z. Chisty, J. M. Bell,  
543 M. Methner, P. Paul, C. M. Carlson, H. P. McLaughlin, N. Thornburg, S. Tong, A. Tamin, Y.  
544 Tao, A. Uehara, J. Harcourt, S. Clark, C. Brostrom-Smith, L. C. Page, M. Kay, J. Lewis, P.  
545 Montgomery, N. D. Stone, T. A. Clark, M. A. Honein, J. S. Duchin, J. A. Jernigan.  
546 Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *NEJM*  
547 **382**, 2081–2090 (2020).
- 548 51. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D.  
549 Bonsall, C. Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with  
550 digital contact tracing. *Science* **368**, eabb6936 (2020).
- 551 52. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G.  
552 Reich, J. Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from  
553 publicly reported confirmed cases: estimation and application. *Ann. Int. Med.* **172**, 577–582  
554 (2020).
- 555 53. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X.  
556 Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F.  
557 Zhang, B. J. Cowling, F. Li, G. M. Leung. Temporal dynamics in viral shedding and  
558 transmissibility of COVID-19. *Nature Med.* **26**, 672–675 (2020).
- 559 54. M. Cevik, M. Tate, O. Lloyd, A. E. Maraolo, J. Schafers, A. Ho. SARS-CoV-2, SARS-CoV,  
560 and MERS-CoV viral load dynamics, duration of viral shedding and infectiousness: a  
561 systematic review and meta-analysis. *Lancet Microbe* **2**, e13–e22 (2021).

- 562 55. B. Borremans, A. Gamble, K. C. Prager, S. K. Helman, A. M. McClain, C. Cox, V. Savage, J.  
563 O. Lloyd-Smith. Quantifying antibody kinetics and RNA detection during early-phase SARS-  
564 CoV-2 infection by time since symptom onset. *eLife* **9**, e60122 (2020).
- 565 56. A. Singanayagam, M. Patel, A. Charlett, J. L. Bernal, V. Saliba, J. Ellis, S. Ladhani, M.  
566 Zambon, R. Gopal. Duration of infectiousness and correlation with RT-PCR cycle threshold  
567 values in cases of COVID-19, England, January to May 2020. *Euro Surveill.* **25**, 2001483  
568 (2020).
- 569 57. B. La Scola, M. Le Bideau, J. Andreani, V. T. Hoang, C. Grimaldier, P. Colson, P. Gautret, D.  
570 Raoult. Viral RNA load as determined by cell culture as a management tool for discharge of  
571 SARS-CoV-2 patients from infectious disease wards. *Eur. J. Clin. Microbiol. Infect. Dis.* **39**,  
572 1059–1061 (2020).
- 573 58. J. Bullard, K. Dust, D. Funk, J. E. Strong, D. Alexander, L. Garnett, C. Boodman, A. Bello, A.  
574 Hedley, Z. Schiffman, K. Doan, N. Bastien, Y. Li, P. G. Van Caesele, G. Poliquin. *Clin.*  
575 *Infect. Dis.* **71**, 2663–2666 (2020).
- 576 59. J. van Beek, Z. Igloi, T. Boelsums, E. Fanoy, H. Gotz, R. Molenkamp, J. van Kampen, C.  
577 GeurtsvanKessel, A. van der Eijk, D. van de Vijver, M. Koopmans. From more testing to smart  
578 testing: data-guided SARS-CoV-2 testing choices.  
579 <https://www.medrxiv.org/content/10.1101/2020.10.13.20211524v2> (2020).
- 580 60. J. Y. Kim, J. H. Ko, Y. Kim, Y. J. Kim, J. M. Kim, Y. S. Chung, H. M. Kim, M. G. Han, S. Y.  
581 Kim, B. S. Chin. Viral load kinetics of SARS-CoV-2 infection in first two patients in Korea.  
582 *J. Korean Med. Sci.* **35**, e86 (2020).
- 583 61. N. Li, X. Wang, T. Lv. Prolonged SARS-CoV-2 RNA shedding: not a rare phenomenon. *J.*  
584 *Med. Virol.* **92**, 2286–2287 (2020).
- 585 62. H. Kawasuji, Y. Takegoshi, M. Kaneda, A. Ueno, Y. Miyajima, K. Kawago, Y. Fukui, Y.  
586 Yoshida, M. Kimura, H. Yamada, I. Sakamaki, H. Tani, Y. Morinaga, Y. Yamamoto. Viral  
587 load dynamics in transmissible symptomatic patients with COVID-19.  
588 <https://www.medrxiv.org/content/10.1101/2020.06.02.20120014v1> (2020).
- 589 63. D. J. Earl and M. W. Deem. Parallel Tempering: Theory, Applications, and New Perspectives.  
590 *Phys. Chem. Chem. Phys.* **7**, 3910–3916 (2005).
- 591 64. M. Plummer, N. Best, K. Cowles, K. Vines. CODA: Convergence Diagnosis and Output  
592 Analysis for MCMC. *R News* **6**, 7–11 (2006).
- 593 65. R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*  
594 (Chapman and Hall/CRC Press, Boca Raton, FL, ed. 2, 2020).

595  
596

597 **Acknowledgments:** We thank Steven Riley for helpful discussions.

598

599 **Funding:**

600 U.S. National Institutes of Health Director’s Early Independence Award DP5-OD028145  
601 (MJM, JAH)

602 Morris-Singer Fund (LKS, ML)

603 U.S. Centers for Disease Control and Prevention Award U01IP001121 (LKS, ML)

604 U.S. National Institute of General Medical Sciences award U54GM088558 (ML, JAH, LKS)

605

606 **Author contributions:**

607 Conceptualization: JAH, LKS, ML, MJM

608 Methodology: JAH, LKS, ML, MJM

609 Visualization: JAH, LKS, ML, MJM

610 Investigation: JAH, LKS, SK, ML, MJM

611 Resources: SK, NJL, SBG, MJM

612 Data curation: JAH, SK, NJL, SBG, MJM

613 Software: JAH, LKS, SK, NJL, SBG

614 Funding acquisition: ML, MJM

615 Supervision: ML, MJM

616 Writing – original draft: JAH, LKS, SK, ML, MJM

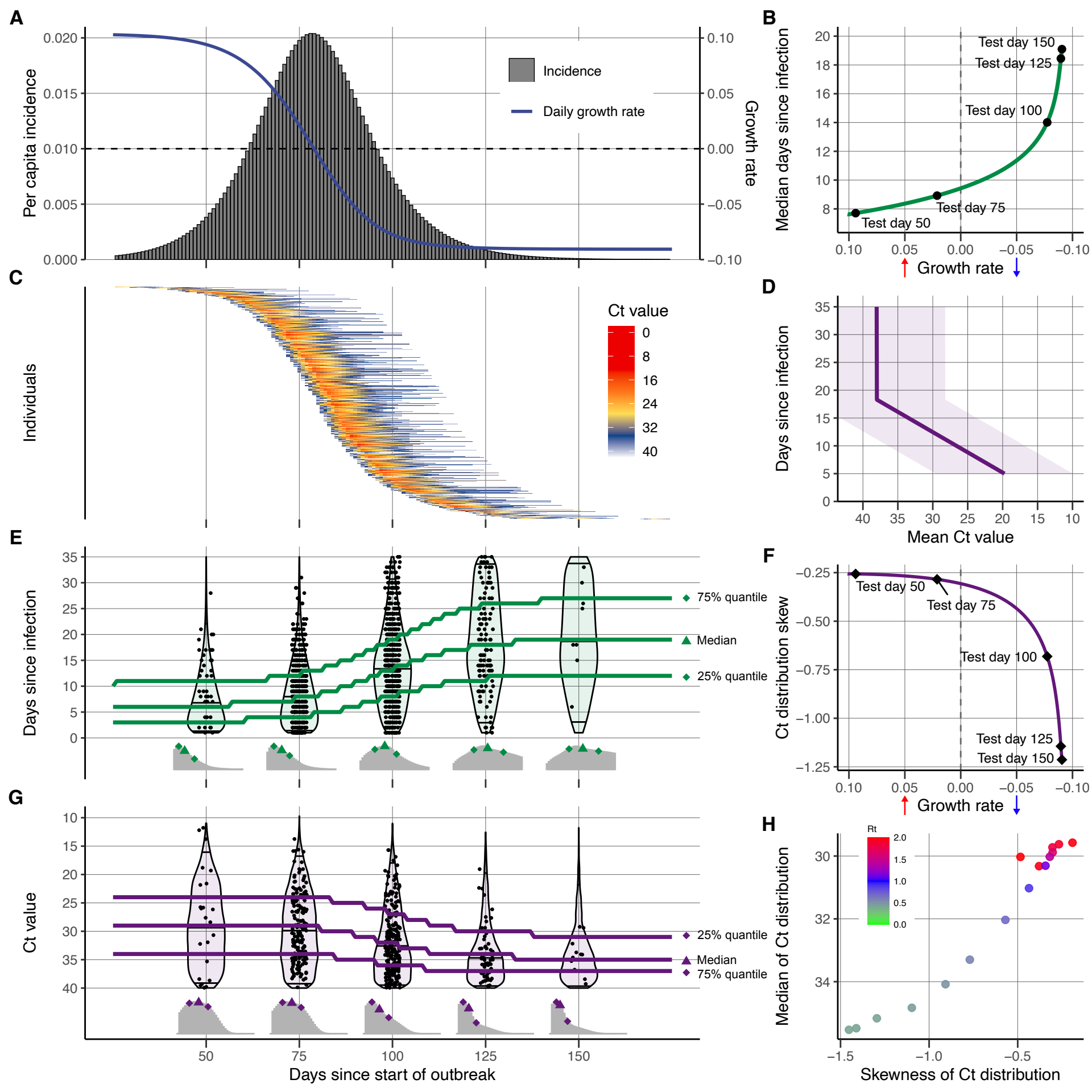
617 Writing – review & editing: JAH, LKS, SK, NJL, SBG, ML, MJM

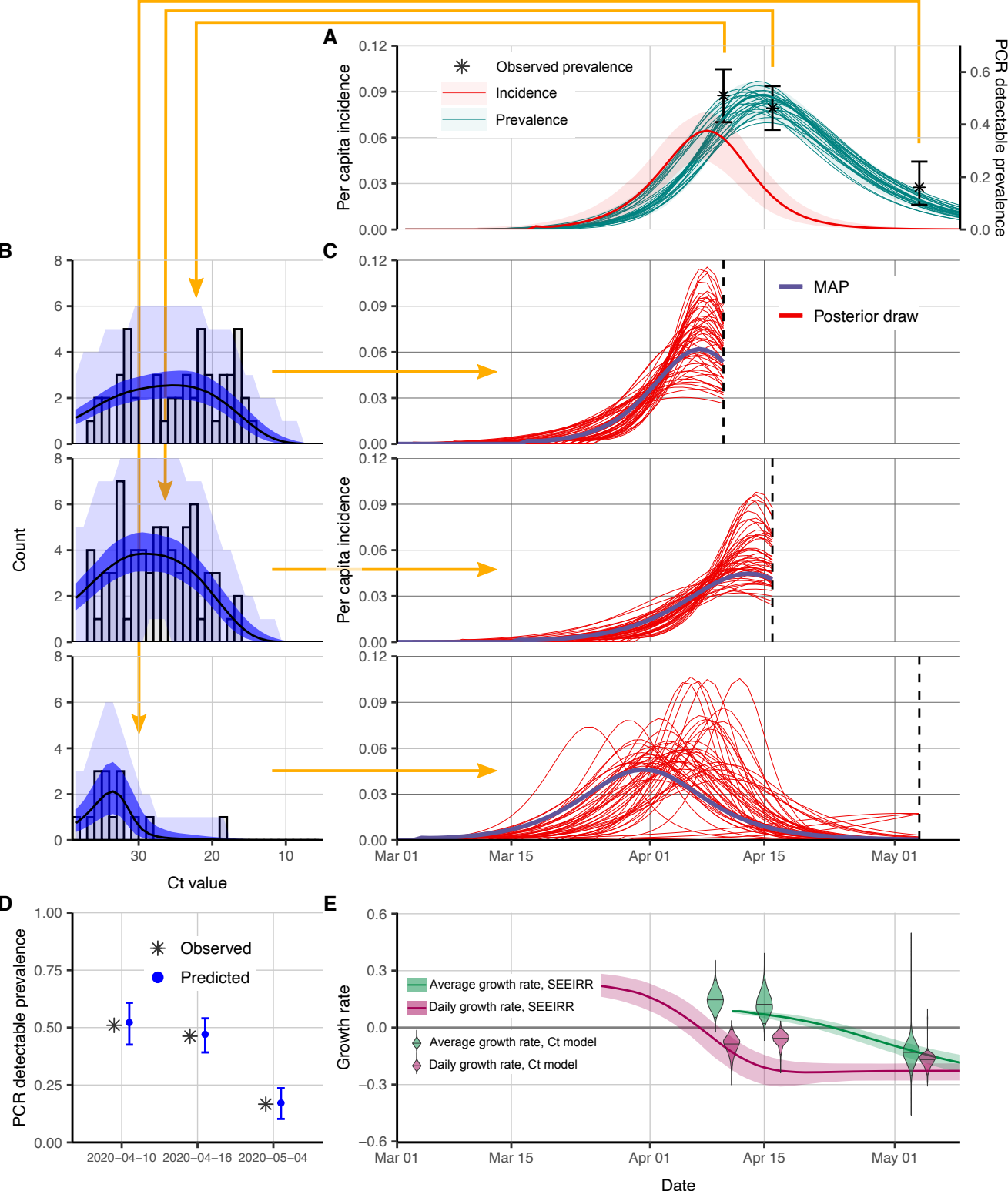
618

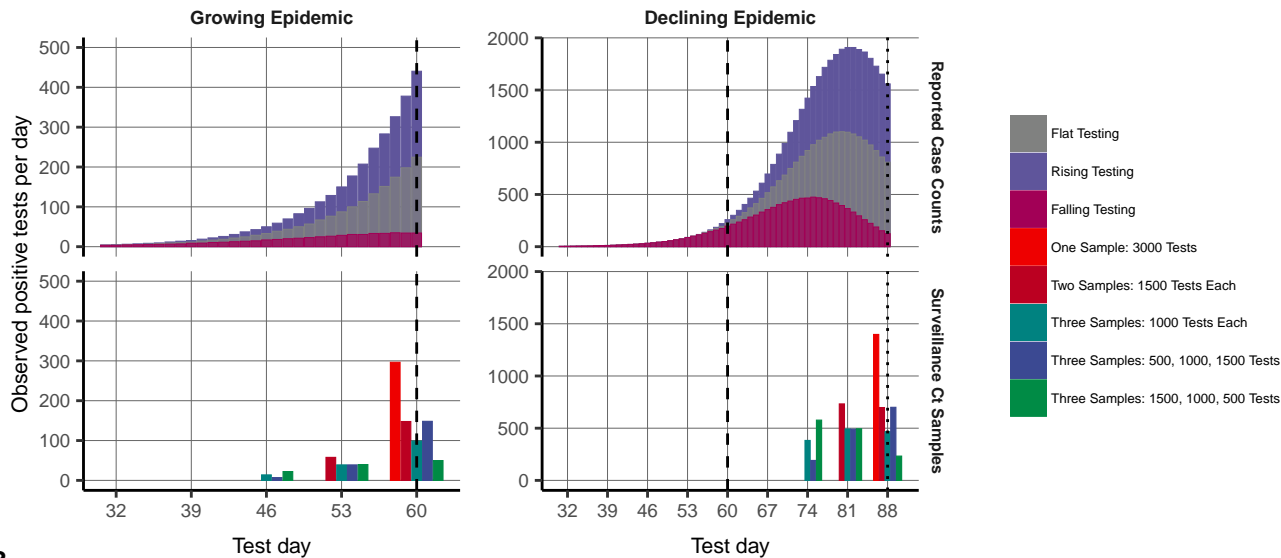
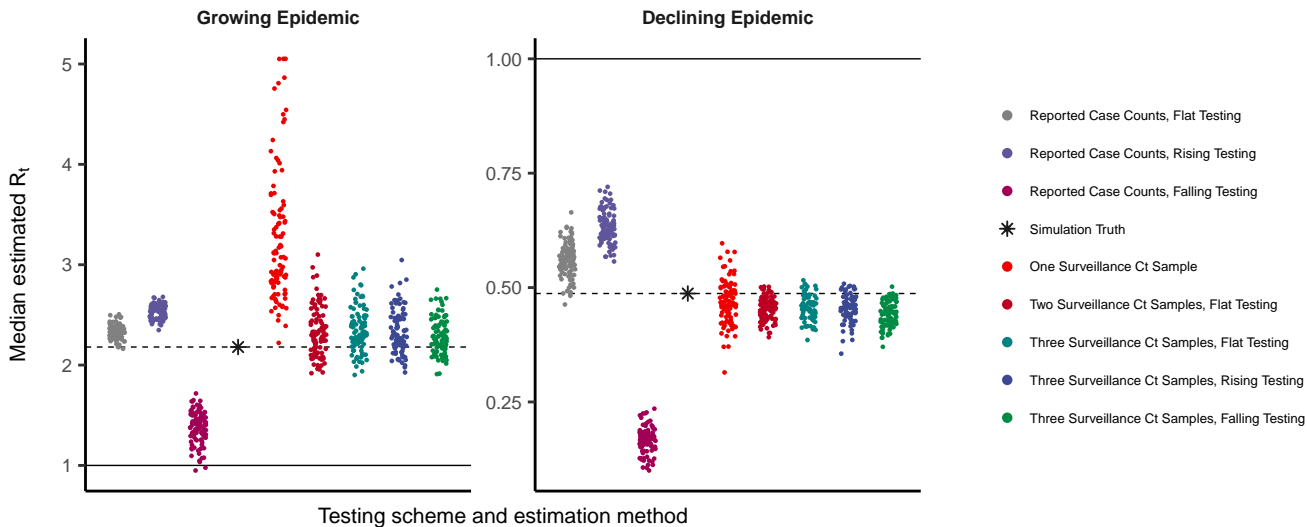
619 **Competing interests:** ML discloses honoraria/consulting from Merck, Affinivax, Sanofi-  
620 Pasteur, and Antigen Discovery; research funding (institutional) from Pfizer, and unpaid  
621 scientific advice to Janssen, Astra-Zeneca, and Covaxx (United Biomedical). MJM is a medical  
622 advisor for Detect. All other authors declare no competing interests.

623

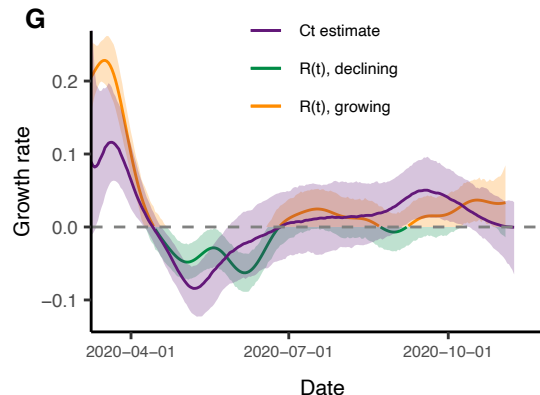
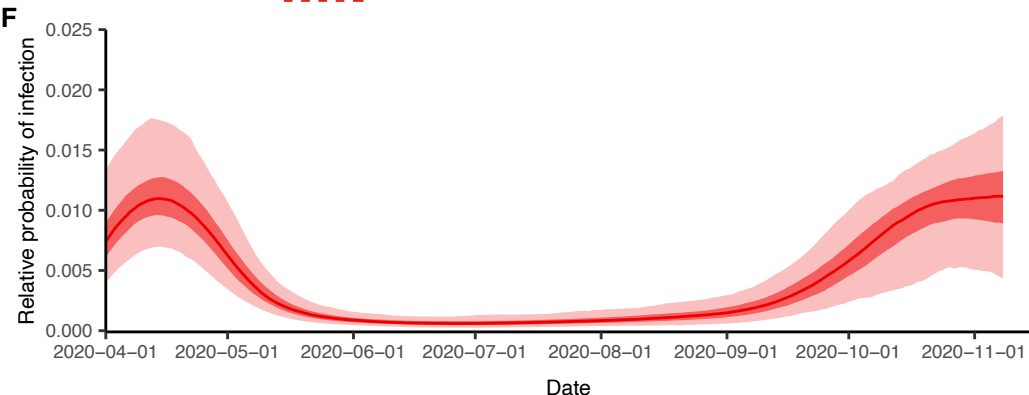
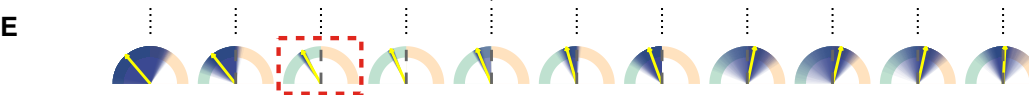
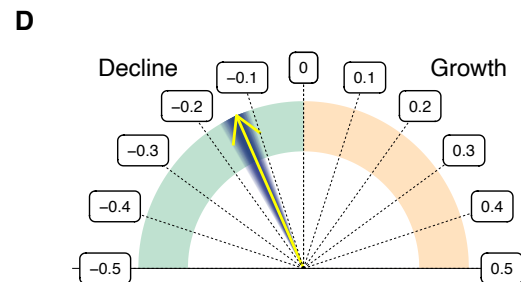
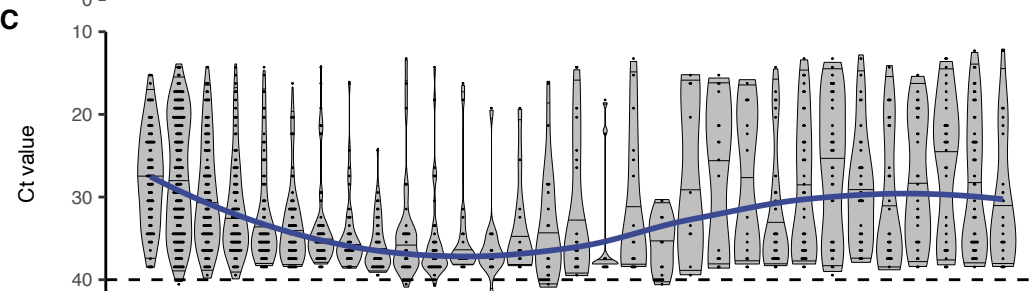
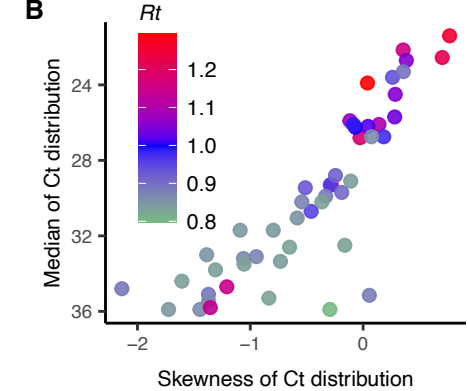
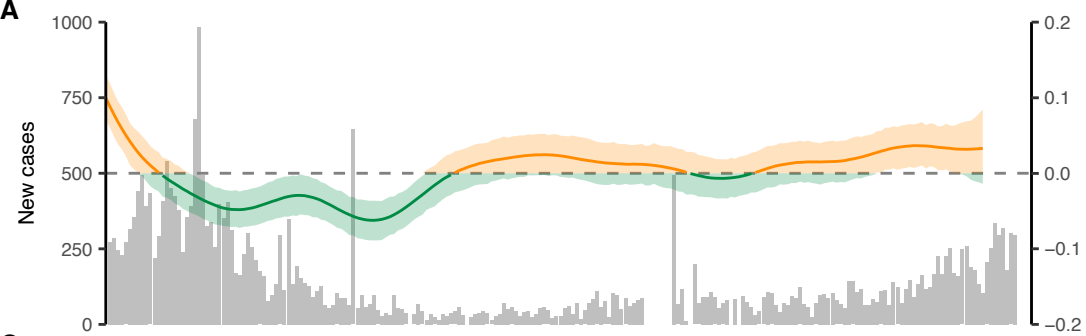
624 **Data and materials availability:** All code to perform the analyses and generate the figures  
625 presented in this article is available at [https://github.com/jameshay218/virosolver\\_paper](https://github.com/jameshay218/virosolver_paper) and  
626 <https://github.com/jameshay218/virosolver>. Simulated data and real data used in the analyses are  
627 also available at [https://github.com/jameshay218/virosolver\\_paper](https://github.com/jameshay218/virosolver_paper). For the model fitting, code  
628 for the Markov chain Monte Carlo framework is available at  
629 <https://github.com/jameshay218/lazymcmc> and  
630 [https://github.com/jameshay218/lazymcmc/tree/parallel\\_tempering](https://github.com/jameshay218/lazymcmc/tree/parallel_tempering). The authors used code  
631 developed by Abbott et al. to estimate  $R_t$  from reported case counts; this is available at  
632 <https://github.com/epiforecasts/EpiNow2>.

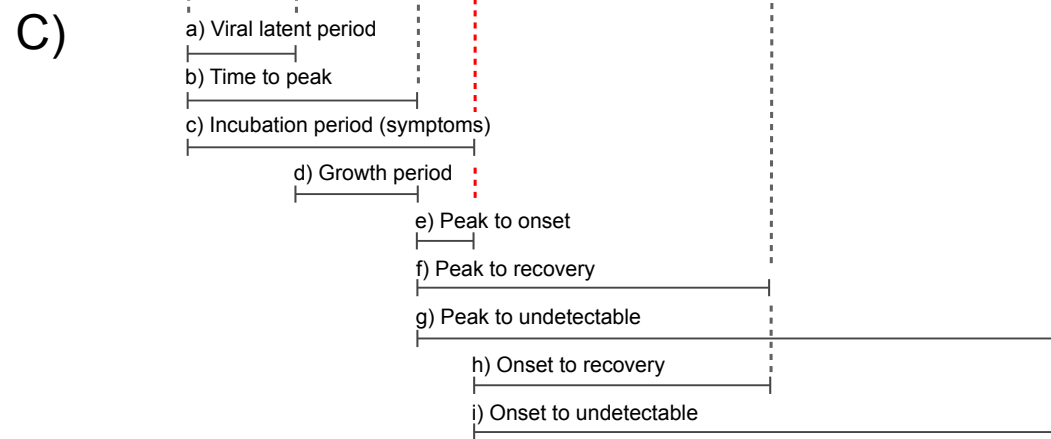
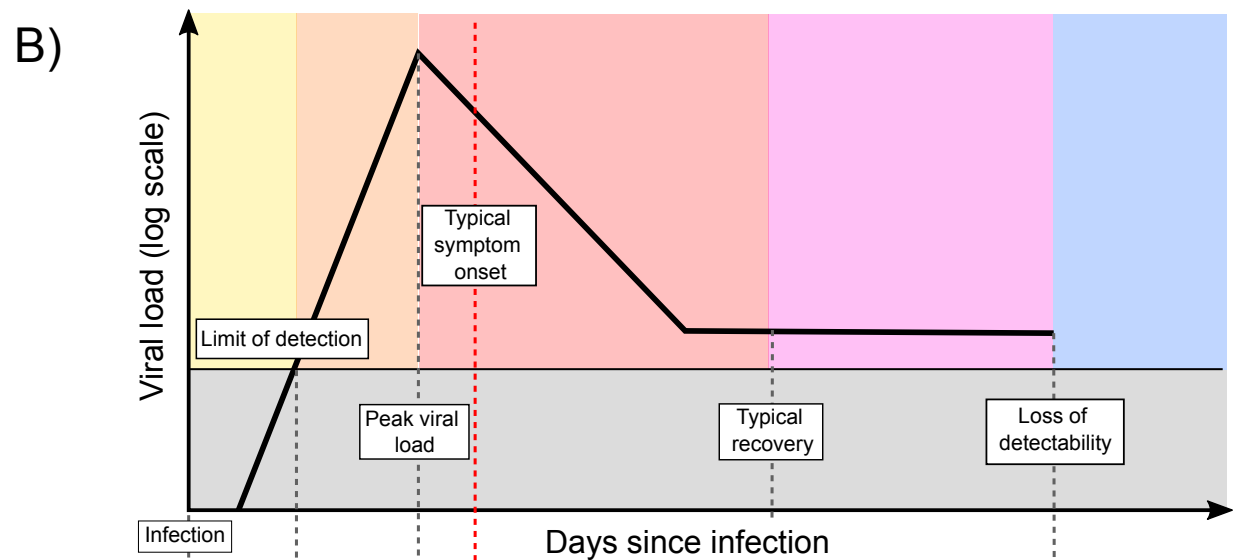
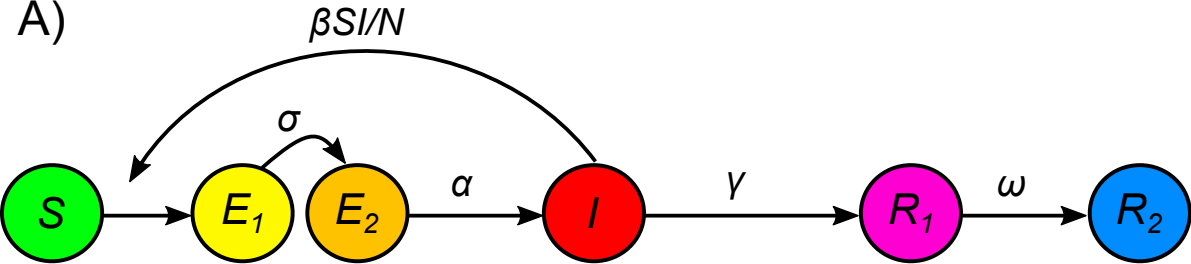




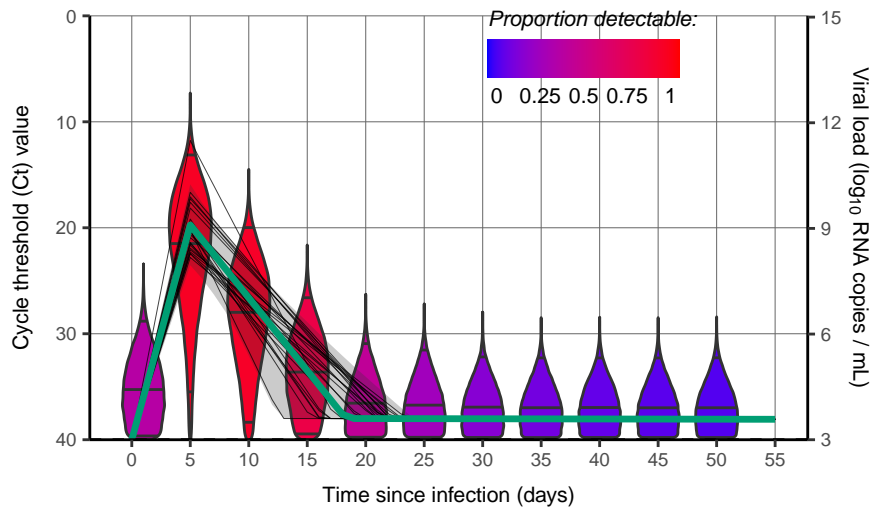
**A****B**



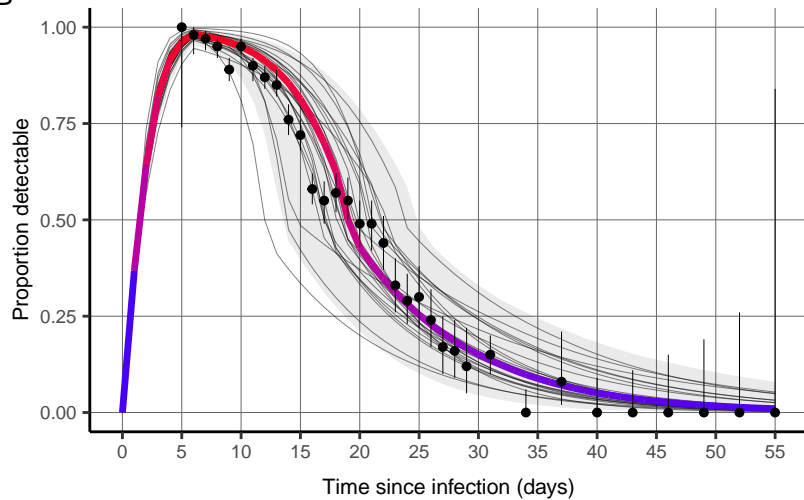




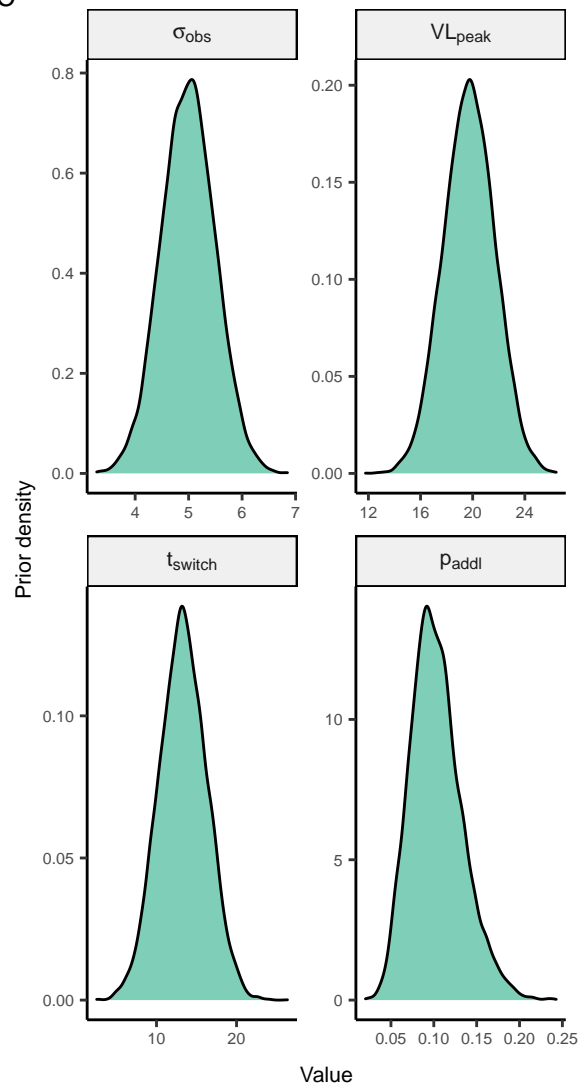
A

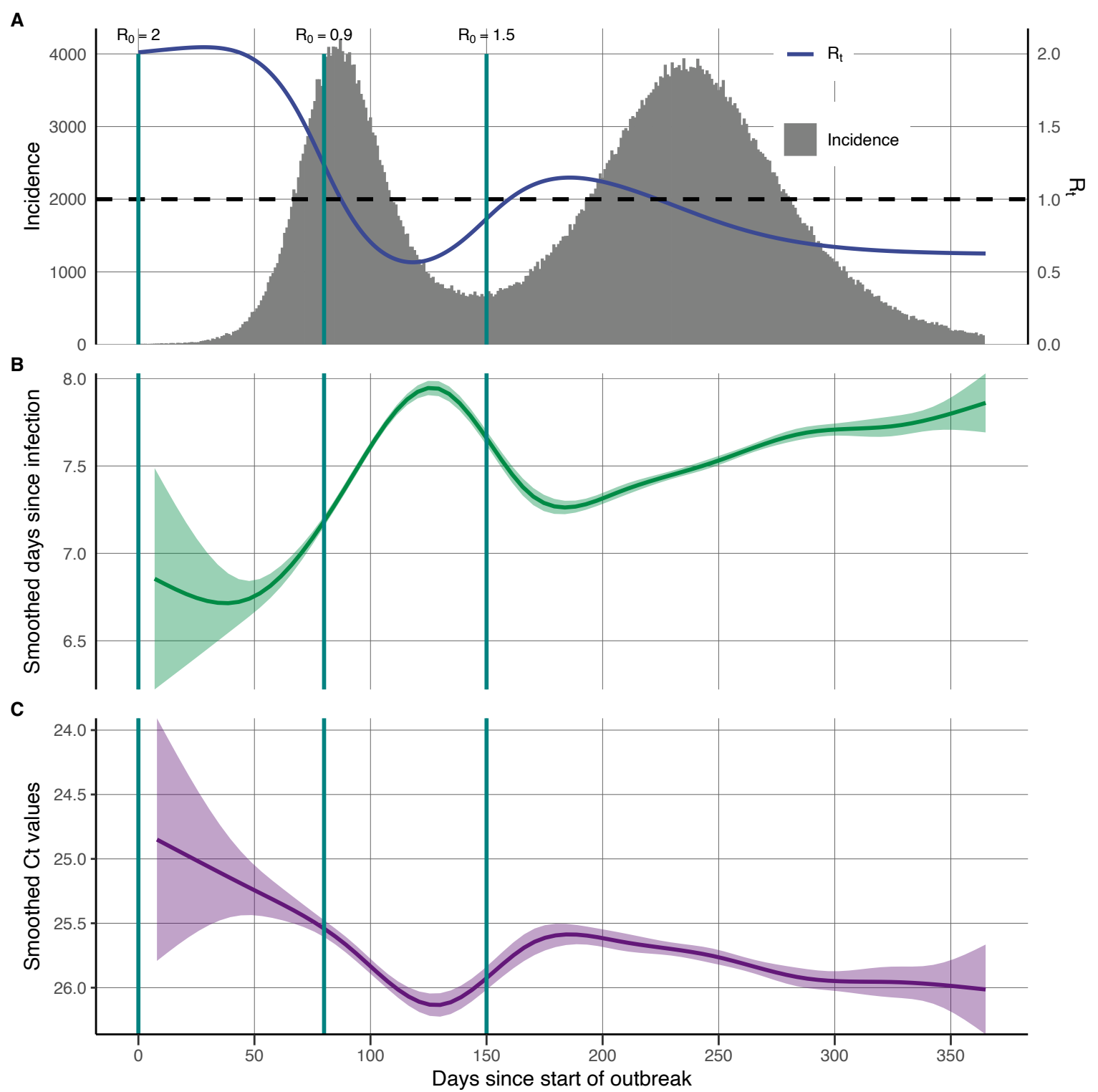


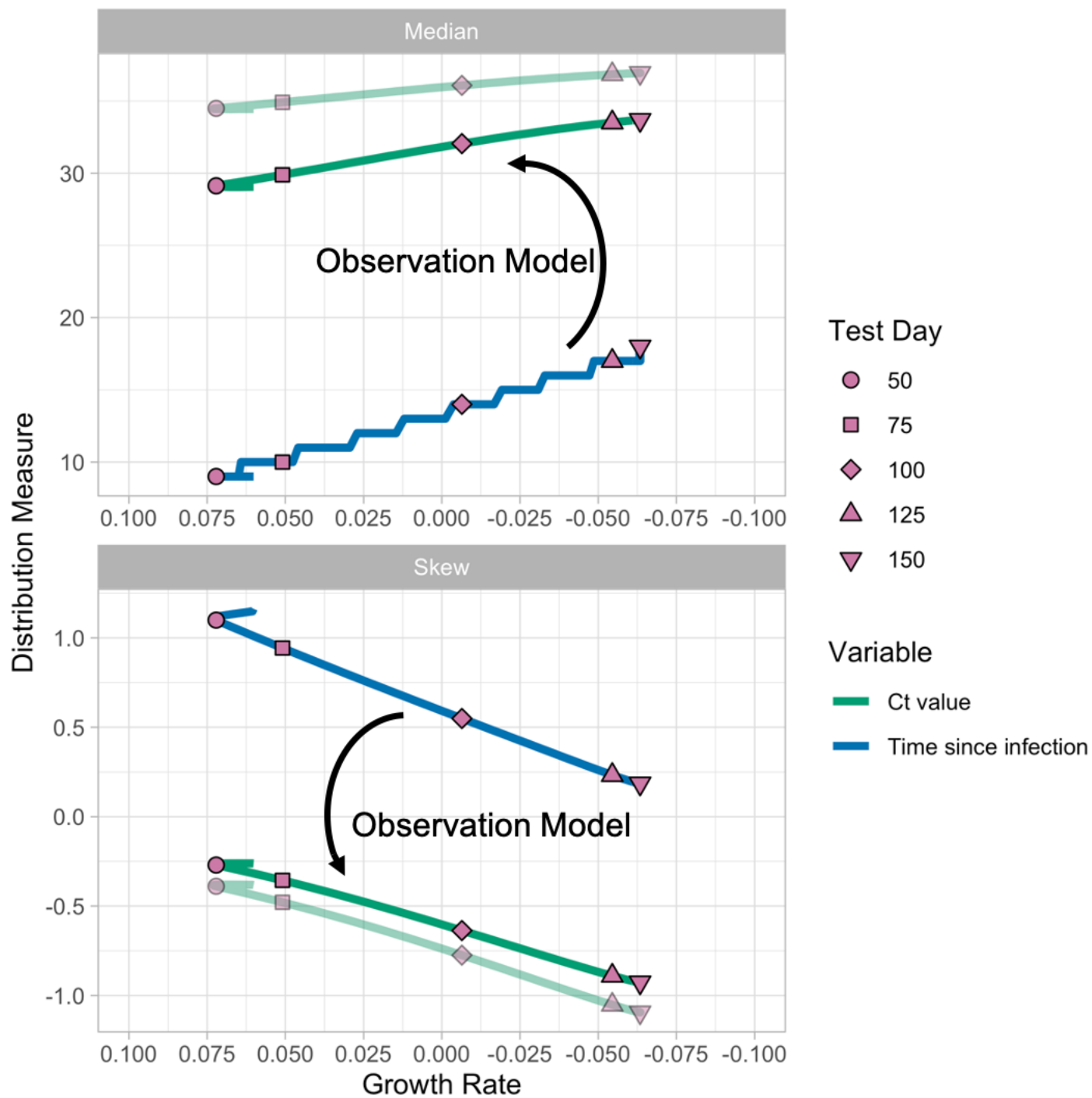
B

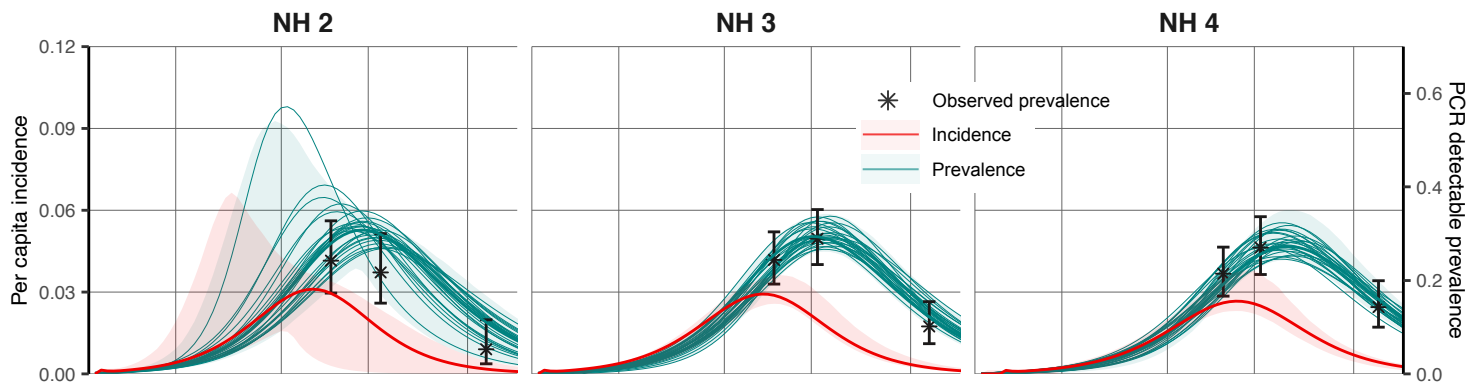
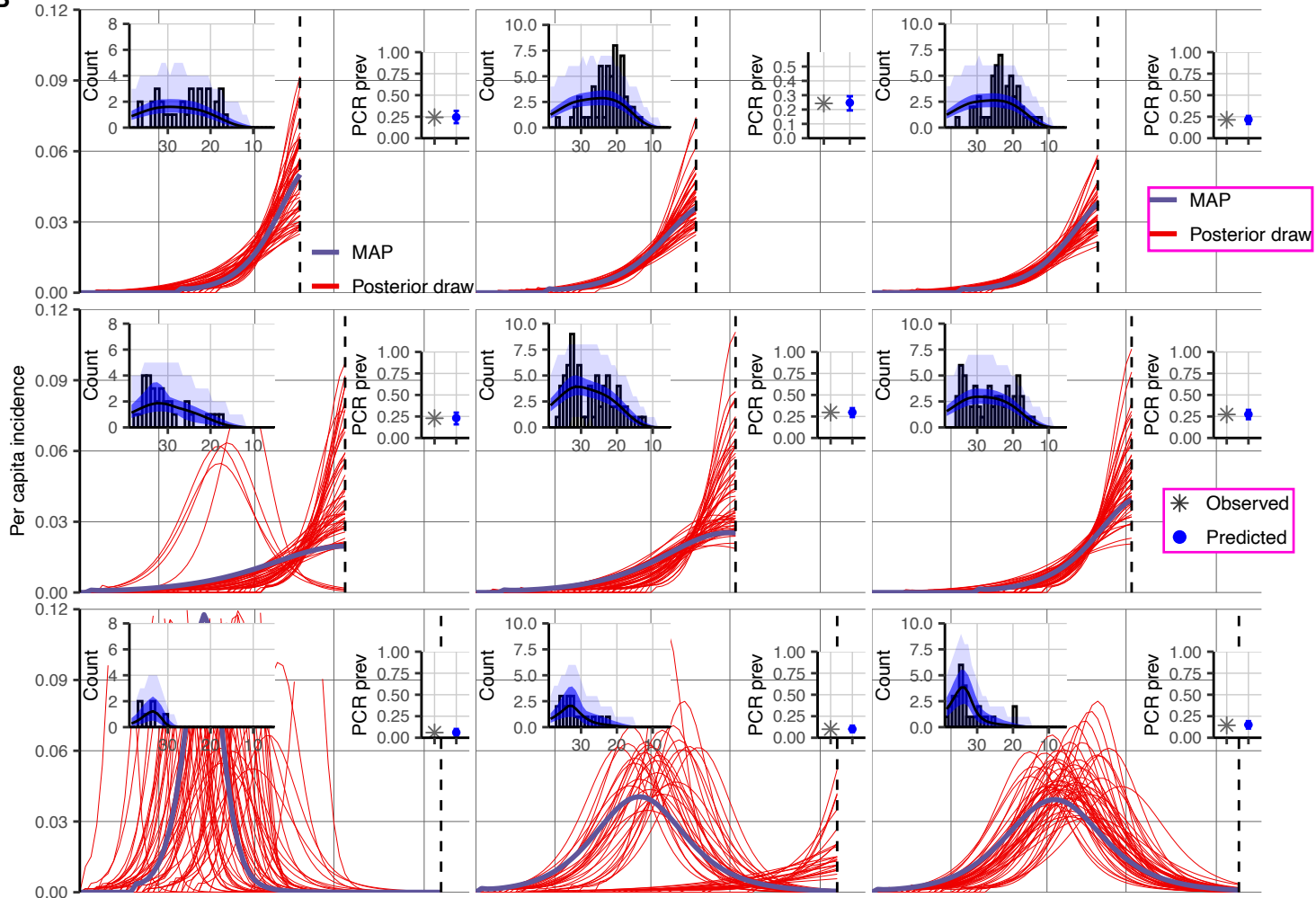
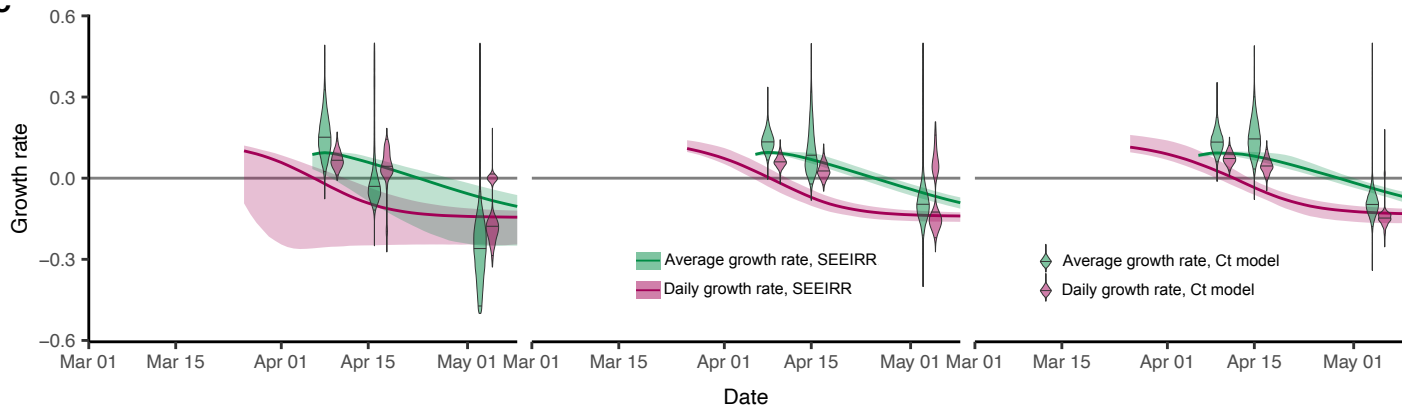


C



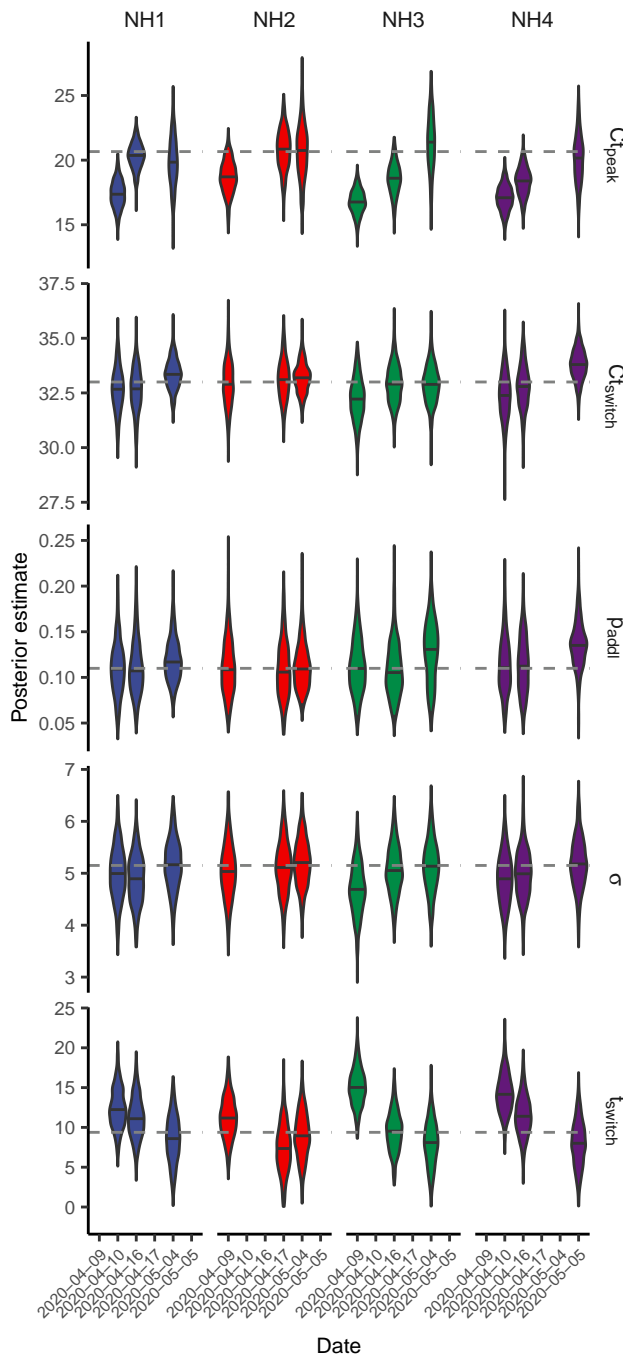




**A****B****C**

A

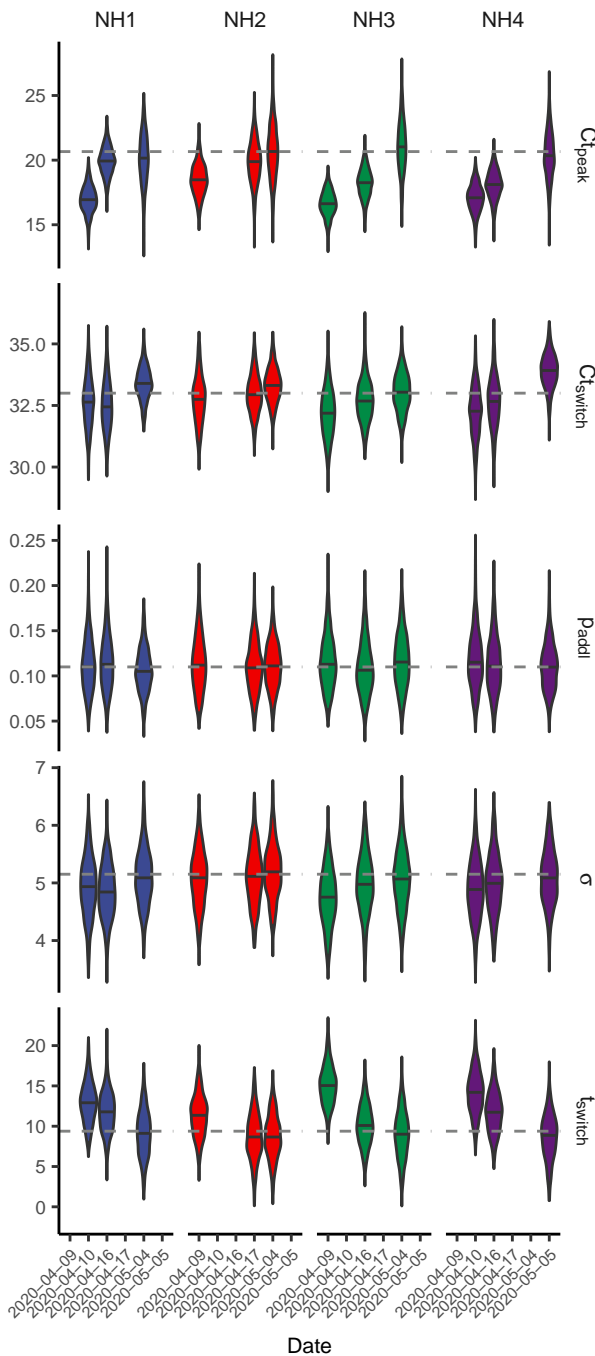
## SEIR fits



Location NH1 NH2 NH3 NH4

B

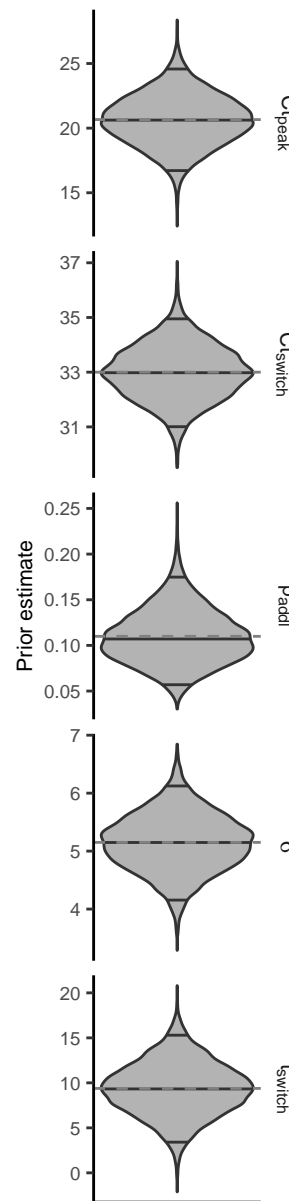
## Exponential model fits

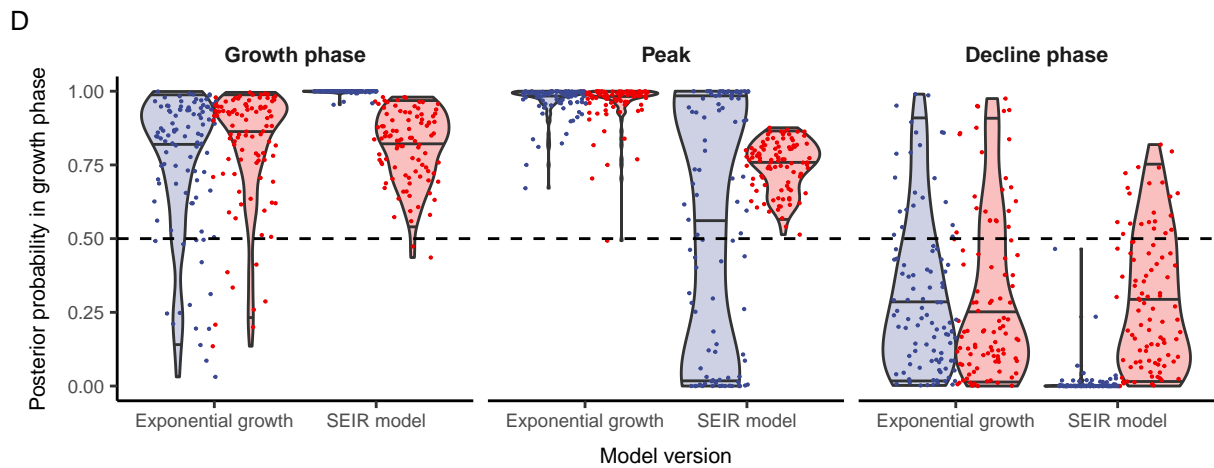
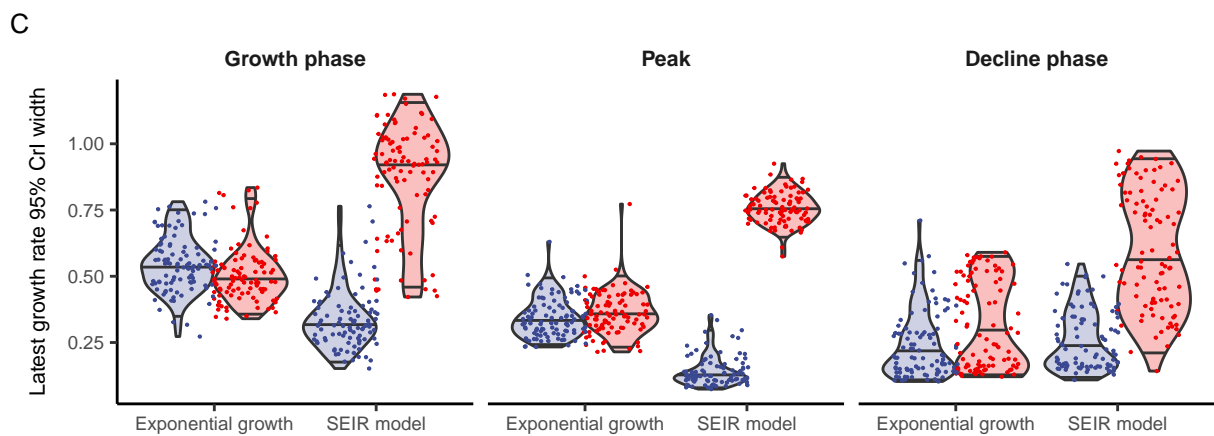
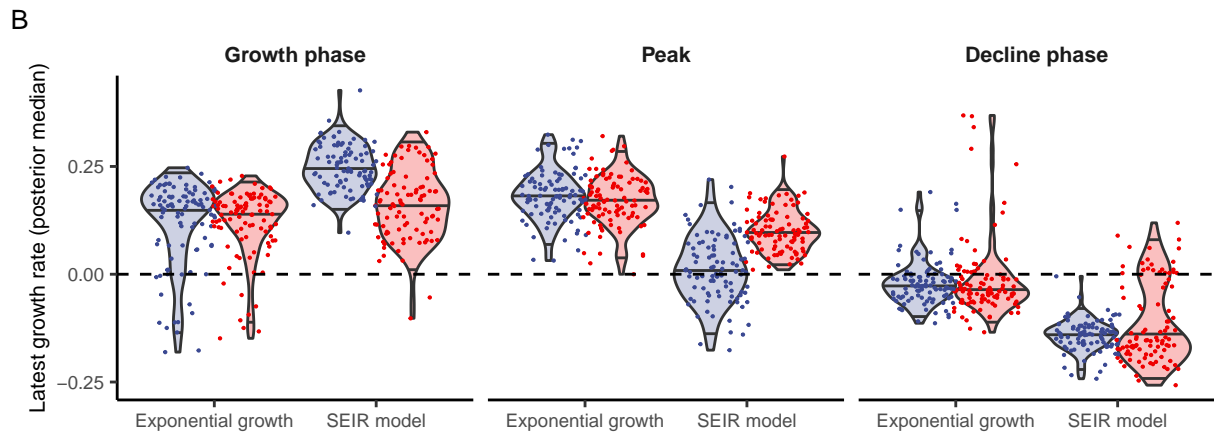
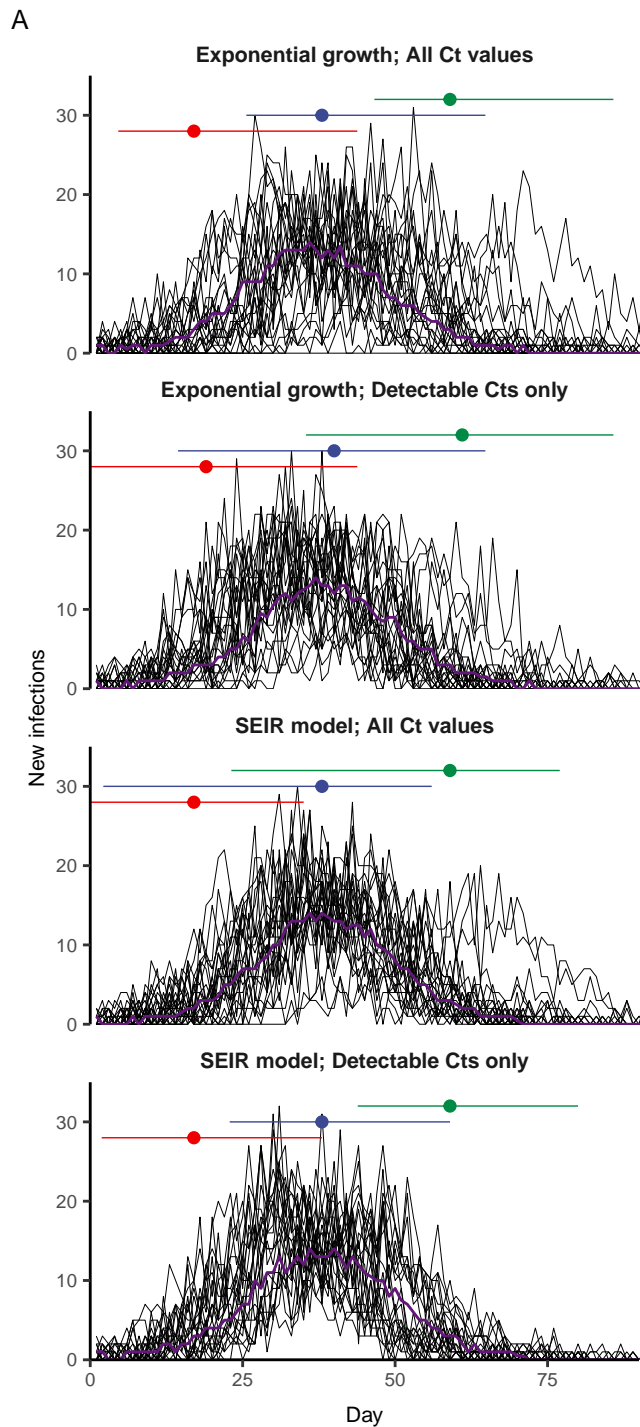


Location NH1 NH2 NH3 NH4

C

## Assumed priors



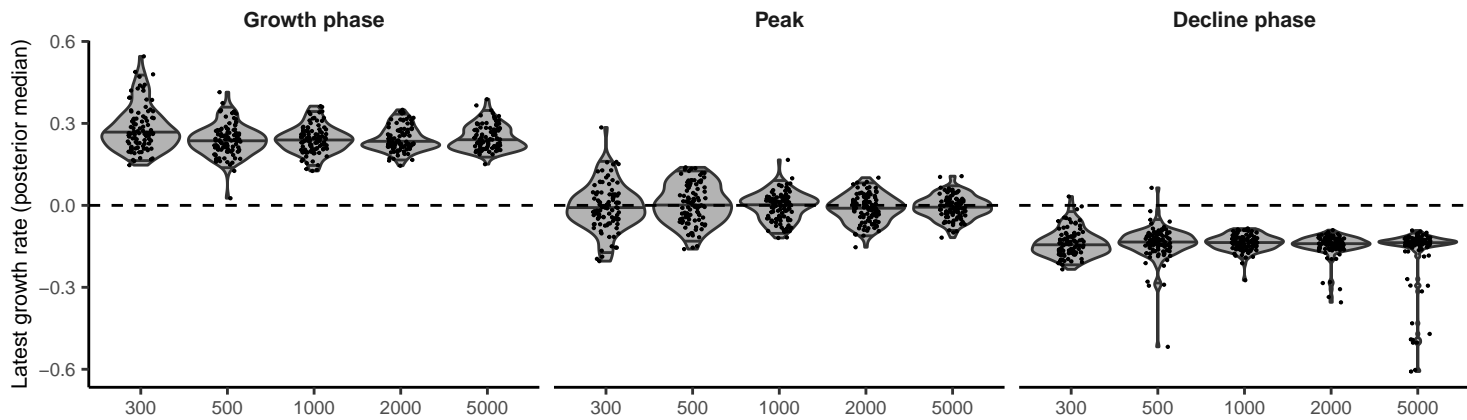


● Growth phase 
 ● Peak 
 ● Decline phase 
 ● Median

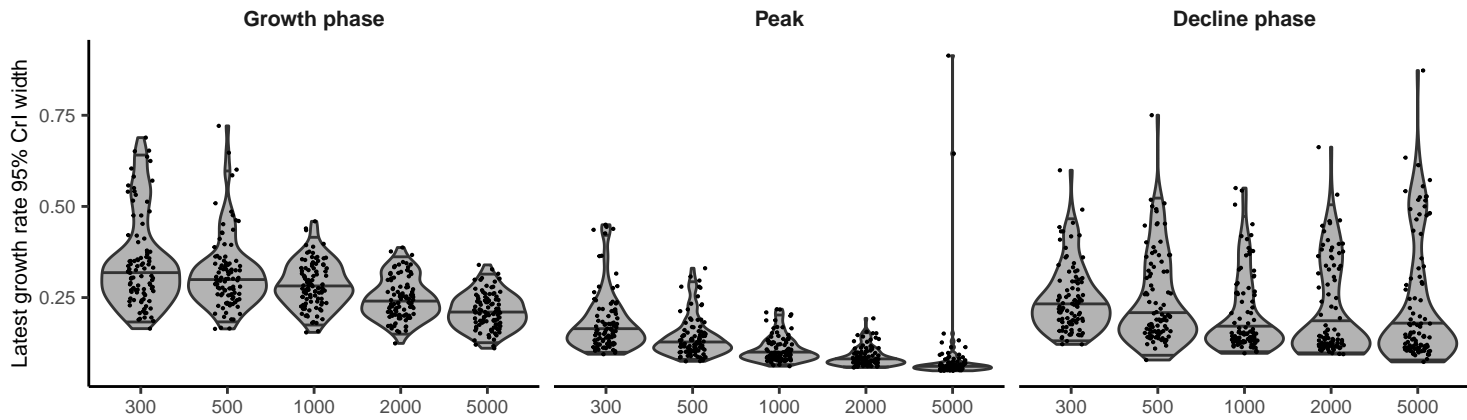
Data ● All Ct values 
 ● Detectable Cts only



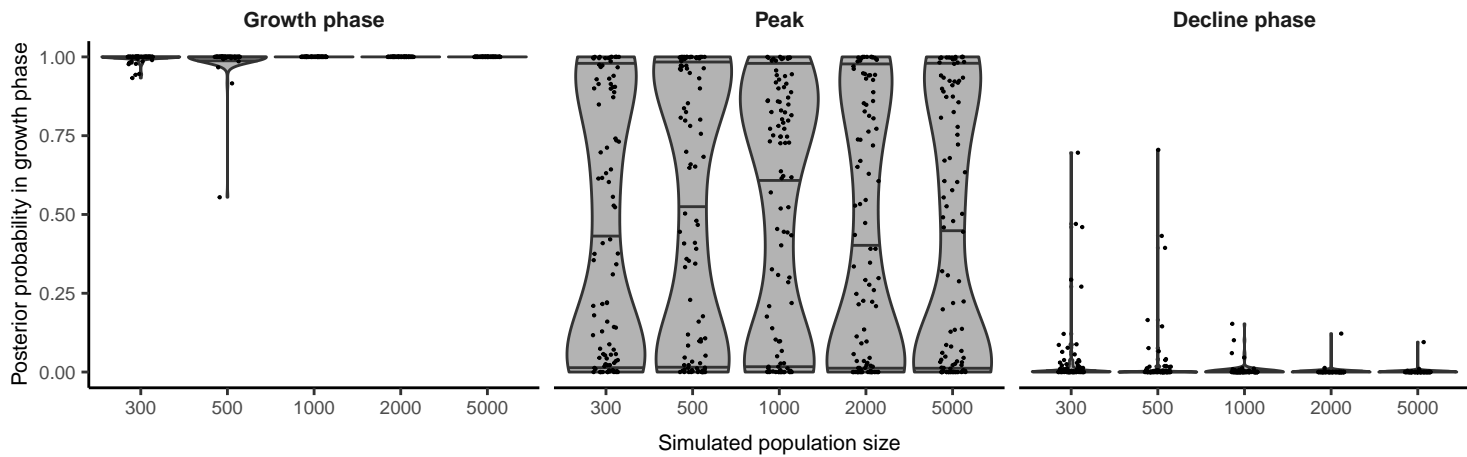
A



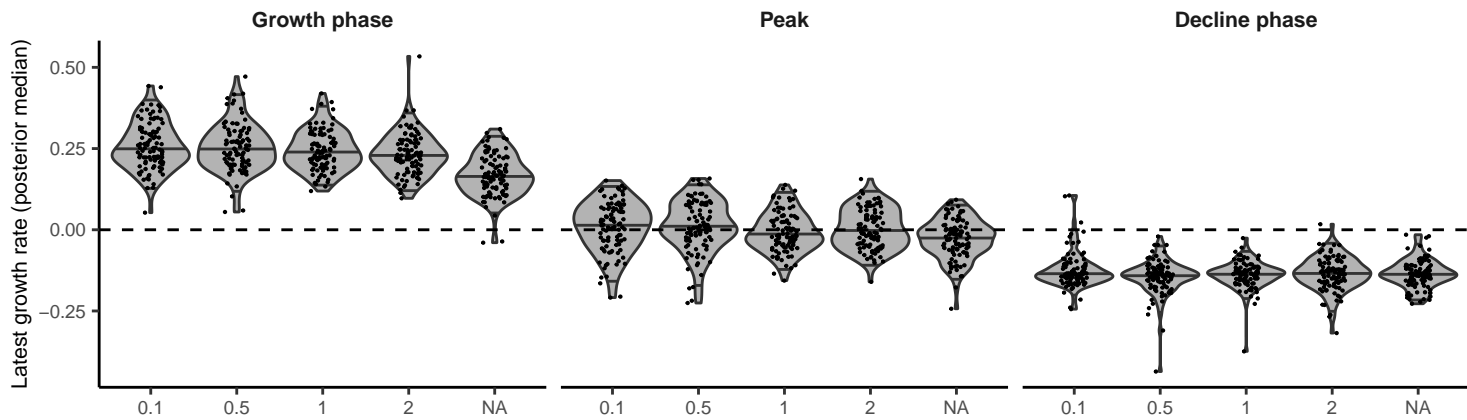
B



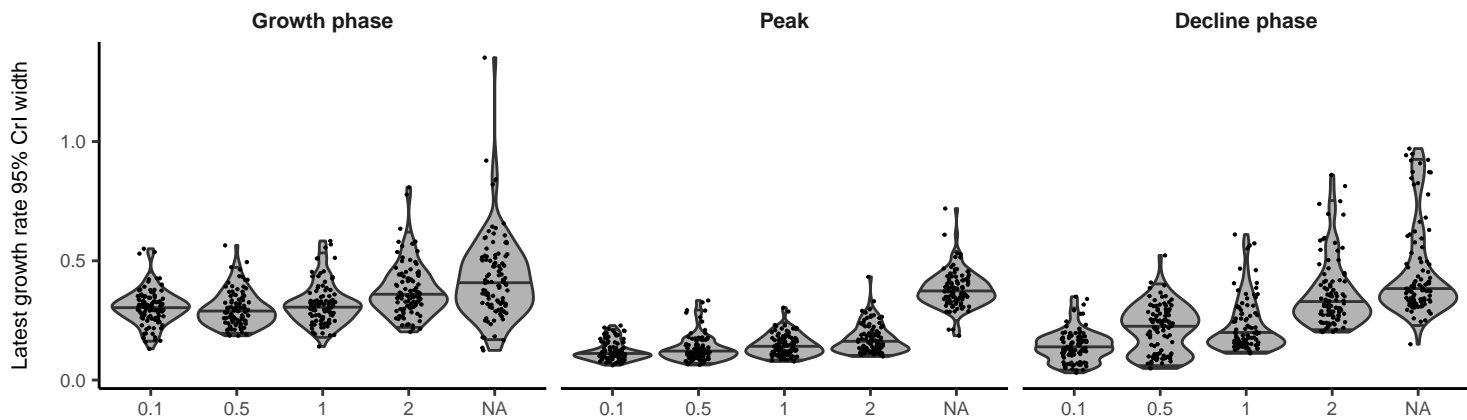
C



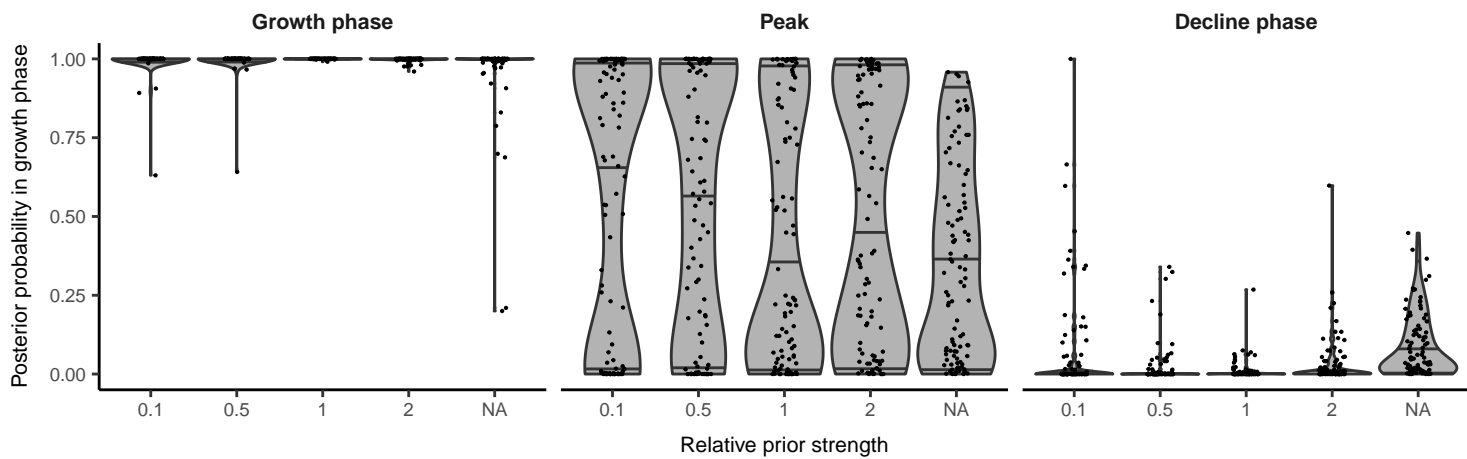
A

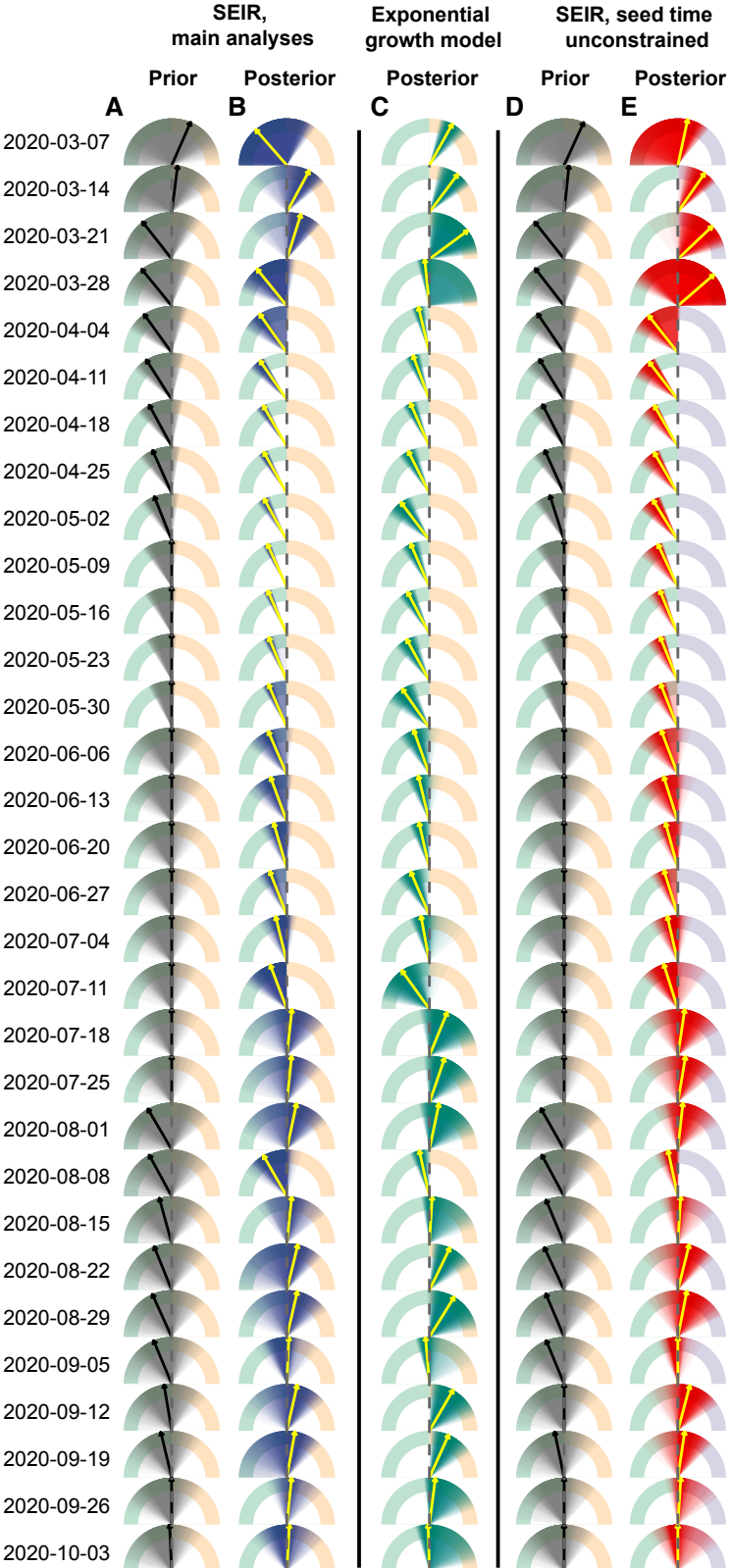


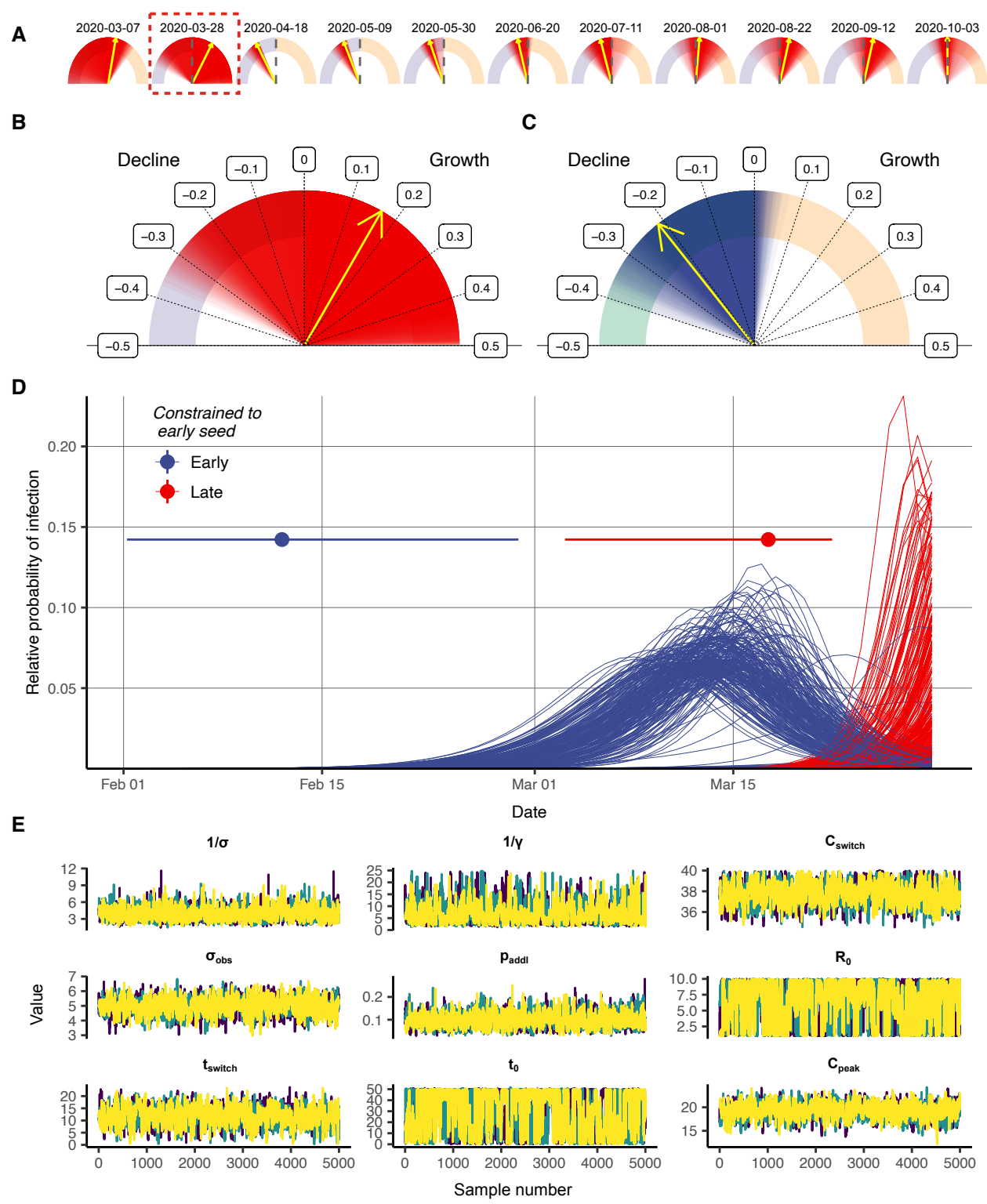
B

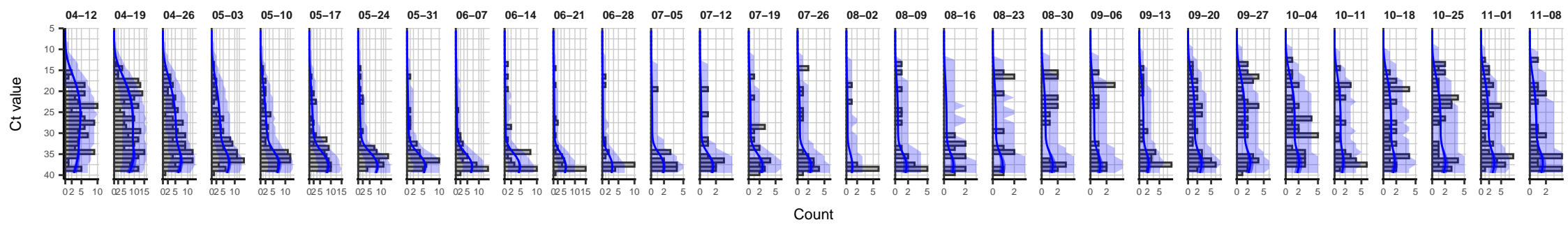
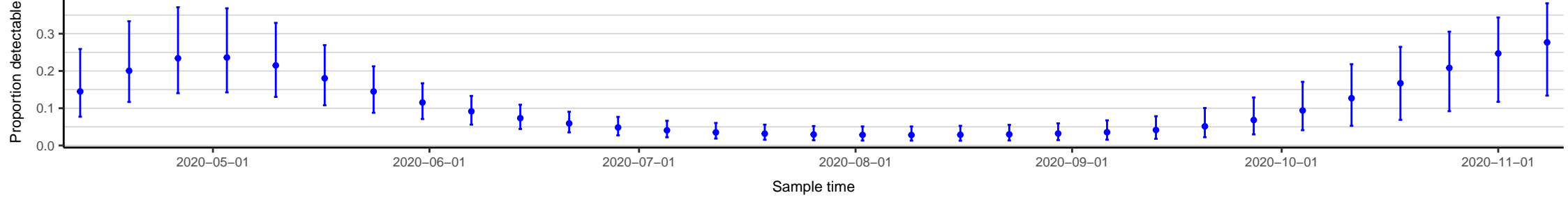
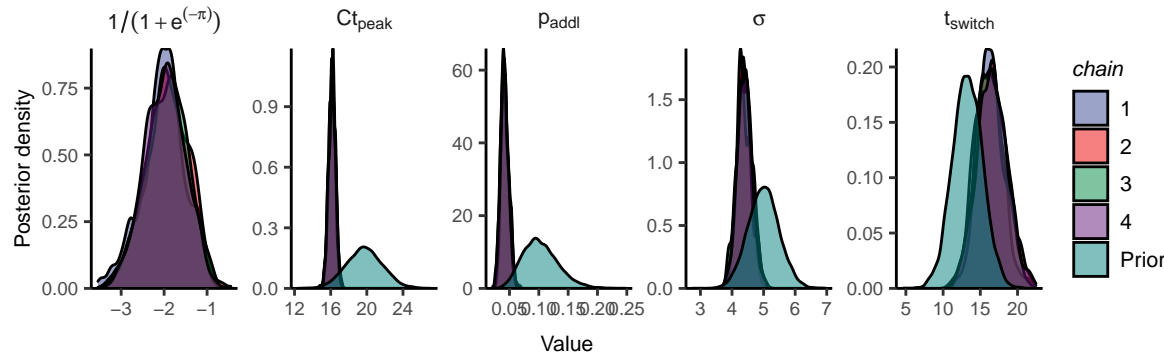
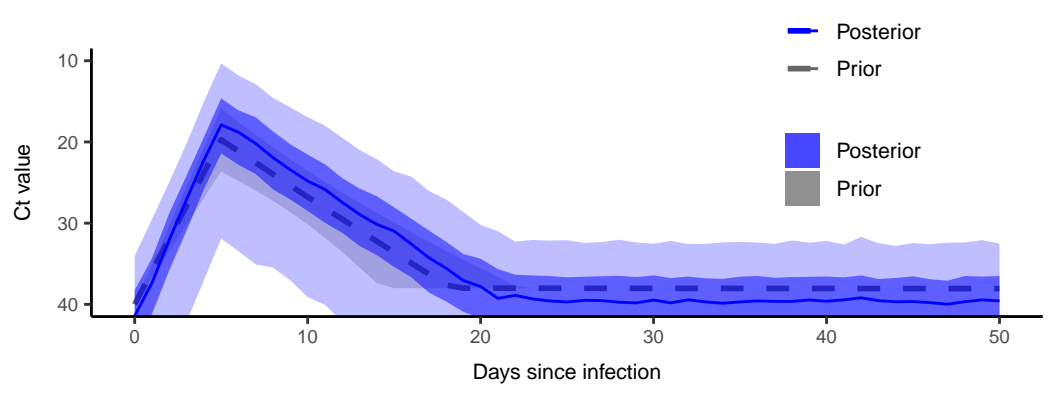
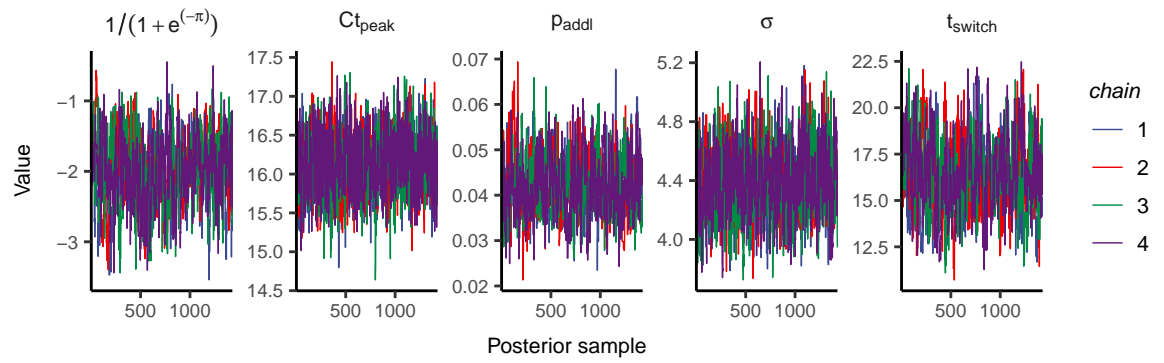


C







**A****B****C****E****D****F**