Bayesian modelling of repeated cross-sectional epidemic prevalence survey data

Nicholas Steyn[1], Marc Chadeau-Hyam[2], Paul Elliott[3], and Christl A. Donnelly[1,4]

[1]Department of Statistics, University of Oxford, Oxford, United Kingdom
[2]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom
[3]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom
[4]Pandemic Science Institute, University of Oxford, Oxford, United Kingdom

Abstract

Large-scale epidemic prevalence surveys are highly informative but resource-intensive tools for monitoring the spread of infectious diseases. Even in large studies, the generated data are often noisy, presenting challenges in both interpretation and inference. State-of-the-art Bayesian methods for analysing epidemic survey data were constructed independently and under pressure during the COVID-19 pandemic. In this paper, we compare two existing approaches (leveraging Bayesian P-splines and approximate Gaussian processes) and a novel approach using sequential Monte Carlo for smoothing and performing inference on epidemic survey data. We use our simpler approach to investigate the impact of survey design and underlying epidemic dynamics on the quality of estimates. We then incorporate these considerations into the existing methods and compare all three methods on simulated data and on real-world data from the SARS-CoV-2 REACT-1 prevalence study in England. All three methods, once appropriate considerations are made, produce similar estimates of infection prevalence; however, estimates of the growth rate and instantaneous reproduction number are more sensitive to underlying assumptions. Notebooks demonstrating all three methods are also provided alongside recommendations on hyperparameter selection and other practical guidance, with some cases resulting in orders-of-magnitude faster runtime.

1 Introduction

The aims of infectious disease surveillance include describing disease burden, monitoring trends, and identifying outbreaks and novel pathogens through the collection, analysis, and interpretation of health data [1]. A range of passive systems, such as automated reporting from healthcare facilities, and active systems, such as field investigations, are used to gather these data. Recently, novel surveillance methods, such as wastewater testing [2], mobile phone data [3, 4], and social media monitoring [5], have been developed to complement traditional surveillance methods. While these methods provide an unprecedented volume of data, they are often subject to biases and limitations that make reliable inference and interpretation difficult [6].

Large-scale prevalence surveys are another tool for the surveillance of infectious diseases. These surveys typically use random sampling methods to produce estimates of the prevalence of infection in a given population. The quality of data collected also allows for the robust estimation of epidemiological quantities such as prevalence P_t, the growth rate r_t, and the instantaneous reproduction number R_t [7]. Established during the COVID-19 pandemic, the REeal-time Assessment of Community Transmission (REACT-1) study in England [8] processed over 2.5 million self-administered throat and nose samples between May 2020 and March 2022. The ONS Coronavirus Infection Survey [9] monitored the spread of SARS-CoV-2 in the United Kingdom, also processing millions of samples over the course of the pandemic. These surveys have been instrumental in understanding, for example, the spread of SARS-CoV-2 [8], the dynamics of infection hospitalisation and infection fatality ratios [10, 11], and the impact of vaccination [12].

Implementing such large-scale surveys is expensive; for example, the ONS Coronavirus Infection Survey cost £988.5 million (until September 2023) [13]. It is therefore critical to maximise the information extracted from the data — not only to justify cost, but also to support timely decision-making and improve policy relevance. This is made challenging by the substantial noise inherent in point prevalence estimates. Even with an ideal survey design, large sample size, and only ignorable non-response, the coefficient of variation of common binomial proportion estimators can be large [14], particularly when infection prevalence is low.

Even when the goal is to perform "model-free" inference, noise in these data often necessitates the use of smoothing methods to improve the quality of the outputs. Smoothing methods impose constraints on day-to-day variability, allowing data from multiple days to inform point estimates. Smoothing allows the researcher to make more confident statements, increasing information yield or reducing the sample size needed for a given confidence level. However, all smoothing methods

make assumptions about the underlying process, whether explicitly or implicitly. Even seemingly simple smoothing methods can introduce substantial bias, turning "model-light" estimates (that rely only minimally on modelling assumptions) into "model-heavy" estimates (that are strongly shaped by modelling assumptions) that may not accurately reflect reality.

Here we compare three Bayesian methods for simultaneously smoothing and performing inference on epidemic survey data. Two existing methods are those of Eales, Ainslie, Walters, et al. [15] and Abbott and Funk [7], henceforth referred to as the Eales method and the Abbott method, which were developed for the REACT-1 study and ONS Coronavirus Infection Survey, respectively. The third is a novel method based on hidden-state models and sequential Monte Carlo (SMC) [16, 17], offering an approach that has the flexibility to incorporate key assumptions of both the Eales and Abbott methods. We refer to this method as the SIMPLE (Survey Inference Method for Prevalence in Longitudinal Epidemiology) method.

The Eales method fits a series of penalised Bayesian splines (B-splines) from which quantities such as the growth rate, prevalence, and instantaneous reproduction number (R_t) are calculated. The Abbott method models the latent infection incidence using a Gaussian process. Explicitly modelling infection incidence (instead of only prevalence) allows for more comprehensive modelling of the epidemic, such as potentially more accurate estimation of R_t, but also requires assumptions to be made about the relationship between population incidence and prevalence (often through an infection-to-positivity curve). An additional method in the literature is that by McCabe, Danelian, Panovska-Griffiths, and Donnelly [18], which is not considered here as it is primarily an adaptation of the Eales method to ONS Community Infection Survey data.

The results of this paper are structured in three parts. First, we use the SIMPLE method to investigate the impact of survey design (such as the number of samples and individual response bias) and underlying epidemic dynamics (such as the variability in the growth rate) on the quality of estimates, highlighting key modelling decisions that should be made when analysing epidemic survey data. Second, we demonstrate how the methods of Eales and Abbott can be adapted to account for these considerations, and compare all three (improved) methods on simulated data to highlight similarities and differences in their performance. Third, we apply all three methods to real-world data from the REACT-1 study, comparing the quality of estimates and computational requirements of each method. We provide recommendations for future modelling and survey design based on our findings.

Our goal is to provide a framework for understanding the impact of modelling decisions on the

quality of estimates, to consolidate and improve state-of-the-art methods, and to make recommendations for future modelling and survey design. We also provide documented source code for all three methods, notebooks demonstrating their use, and recommendations for hyperparameter selection and other practical considerations.

2 Materials and Methods

2.1 The SIMPLE method

We introduce a collection of state-space models for simultaneously smoothing and performing inference on epidemic survey data. Each state-space model consists of an explicitly defined epidemic and observation model. The epidemic model encodes our assumptions about the unobserved dynamics of the epidemic, while the observation model describes how the observed data are generated from the underlying epidemic.

Two epidemic models are considered: one that models the daily exponential growth rate r_t in prevalence P_t and one that models the instantaneous reproduction number R_t, infection incidence I_t and prevalence P_t. Three observation models are considered: a basic model that assumes the number of positive swabs (diagnostic tests) follows a binomial distribution, an extra-binomial model that accounts for overdispersion in the data, and a weighted model that accounts for survey weights.

We refer to the unknown time-varying quantities of interest, such as r_t and P_t, as hidden states. Other unknown parameters that are not time-varying, such as the level of overdispersion in the extra-binomial model, are referred to as fixed parameters.

Throughout this paper, we use the term *prevalence* to describe the proportion of the population that would test positive. This quantity may differ from the proportion of the population with an active infection (due to imperfect test sensitivity/specificity) and from the proportion who are infectious (as test positivity can outlast infectiousness).

2.1.1 Growth rate epidemic model

This model assumes that the daily growth rate r_t in prevalence follows a Gaussian random walk, encoding an assumption that r_t varies smoothly over time. Prevalence P_t on day $t \in \{1, 2, \ldots, T\}$ then varies exponentially with rate r_t.

$$r_t = r_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2) \tag{1}$$

$$P_t = P_{t-1} e^{r_t} \tag{2}$$

Coupled with initial distributions for r_0 and P_0, and a prior distribution for σ, equations 1 and 2 define the entire epidemic model. Parameter σ controls the smoothness of our estimates and is inferred from the data.

We use default initial distributions of $r_0 \sim \mathrm{Uniform}(-0.3, 0.3)$ and $P_0 \sim \mathrm{N}(\hat{p_1}, \hat{p_1}(1-\hat{p_1})/n_1)$, where $\hat{p_1} = n_1^+/n_1$ (the Wald interval), and default prior distribution $\sigma \sim \mathrm{Uniform}(0, 0.2)$, although inferences are insensitive to the choice of these prior distributions (supplementary section 1).

As this model makes minimal assumptions about the underlying epidemic, we use it as the default epidemic model unless otherwise stated.

2.1.2 Reproduction number epidemic model

Alternatively, we may want to estimate the instantaneous reproduction number R_t, a popular alternative measure to r_t for characterising the rate of epidemic spread [19]. R_t is the average number of secondary infections generated by a primary infection at time t if an individual were to undergo their entire infectious period at this time. As with r_t, we employ a Gaussian random walk to smooth estimates, now on $\log R_t$ to ensure positivity:

$$\log R_t = \log R_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma_R^2) \tag{3}$$

We then employ the renewal model [17, 20], which relates past infection incidence $I_{1:t-1}$ to current infection incidence I_t through the instantaneous reproduction number R_t and a generation time distribution (representing the time from infection of an infector and their infectee, described by probability mass function w_u):

$$I_t = R_t \sum_{u=1}^{t} I_{t-u} w_u \tag{4}$$

Finally, we relate prevalence P_t to infection incidence I_t through a test-sensitivity function h_u that describes how likely an individual is to test positive u days after infection. We do

not consider imperfect test specificity, although this could be included with the addition of a constant term to P_t:

$$P_t = \sum_{u=1}^{t} I_{t-u} h_u \tag{5}$$

When applying the reproduction number epidemic model to data from the REACT-1 study, the generation time is assumed to follow a gamma distribution with mean 6.4 days and standard deviation 4.2 days [7], discretised by evaluating the gamma density function at integer time steps and normalising so $\sum w_u = 1$. In the absence of a REACT-1-specific test-sensitivity function, we use the pointwise central estimate of h_u from [21], reflecting the type of test used in REACT-1 (reverse transcription polymerase chain reaction (RT-PCR)), but not other study-specific factors that may impact sensitivity. This is the same mean test-sensitivity function as used in the Abbott method.

Estimating R_t requires many additional assumptions about the underlying epidemic, and thus has more potential to bias estimates of P_t. We focus on the growth rate epidemic model for much of this paper, although demonstrate the R_t estimator on real-world REACT-1 data.

2.1.3 Basic observation model

Given n_t swabs conducted on day t, the observed number of positive swabs n_t^+ is modelled as a binomial random variable with probability P_t:

$$n_t^+ \sim \text{Binomial}(n_t, P_t) \tag{6}$$

This is also the observation model used in the Eales method as originally published.

2.1.4 Extra-binomial observation model

Prevalence data are often overdispersed relative to the binomial distribution, exhibiting what is known as extra-binomial variation [22]. If we assume that the "observable" prevalence at time t is beta-distributed with mean equal to prevalence P_t and variance $\rho P_t(1 - P_t)$, our observation distribution becomes:

$$n_t^+ \sim \text{Beta-binomial}(n_t, \alpha_t, \beta_t) \tag{7}$$

where $\alpha_t = P_t \left(\frac{1}{\rho} - 1\right)$ and $\beta_t = (1 - P_t) \left(\frac{1}{\rho} - 1\right)$.

The additional parameter $\rho \in (0, 1)$ controls the modelled level of overdispersion in the data. Larger values of ρ indicate greater levels of overdispersion while letting $\rho \to 0$ recovers the binomial model. By default, we use a Uniform$(0, 0.01)$ prior distribution for this parameter.

This is our default observation model and is used in all subsequent analyses unless otherwise stated.

2.1.5 Weighted observation model

Survey weights are used to account for bias in survey data arising from unequal probabilities of selection and/or response. Letting $w_{t,i}$ be the (normalised) weight assigned to the ith sample taken at time t, and $x_{t,i}$ be the corresponding result (where $x_{t,i} = 1$ is individual i on day t tests positive, and zero otherwise), the weighted swab positivity is:

$$\hat{p}_t = \sum_{i=1}^{n_t} x_{t,i} w_{t,i} \tag{8}$$

If the weights are uncorrelated with swab positivity, then $Var(\hat{p}_t) = P_t(1 - P_t) \sum_{i=1}^{n_t} w_i^2$. However, if the weights are correlated with the individual probability of testing positive, the variance of $\hat{p}_t$ may be over or underestimated by this expression. The difference depends on the specific relationship between weights and outcome, which is generally unknown. Thus, we model $\hat{p}_t$ using a normal distribution with mean P_t and variance $cP_t(1 - P_t) \sum_{i=1}^{n_t} w_{t,i}^2$, where c is an estimated scalar parameter:

$$\hat{p}_t \sim \mathrm{N}\left(P_t, cP_t(1 - P_t) \sum_{i=1}^{n_t} w_{t,i}^2\right) \tag{9}$$

The observation distribution is no longer exact, in that we are approximating the unknown distribution of $\hat{p}_t$ with a normal distribution. This is similar to the approach used in the Abbott method. By default, we use a Uniform$(0.1, 10)$ prior distribution for c.

2.1.6 Sequential Monte Carlo (SMC)

We use an SMC algorithm (also known as a particle filter), the bootstrap filter, to estimate the posterior distributions of the hidden states (such as r_t, P_t, and R_t) at each time step t, given the observed data [17]. We also use this algorithm to estimate the posterior predictive distribution of n_t^+, which is similar to the frequentist's predictive distribution, and can be used to assess the quality of fit of the model when fitting to real-world data. Particle marginal Metropolis Hastings (PMMH) is used to estimate the fixed parameters, the uncertainty of which is then marginalised

over in the final inference.

We run three chains of the PMMH algorithm, each for 100 iterations at a time, until the maximum Gelman-Rubin diagnostic ($\hat{R}$) [23] is less than 1.05 and the minimum effective sample size (ESS) is greater than 100. These algorithms are detailed in full in supplementary section 2. We make the full source code (written natively in Julia [24]) available on GitHub, along with notebooks, to facilitate the application of these methods to other datasets.

When fitting the model for the reproduction number, an additional "wind-in" period (default 10 days) is required to account for infection history prior to the first observation. After sampling from the standard prior distribution for P_0, assumed values of $I_{-9:0}$ are set to $P_0/\sum h_u$, ensuring the distribution of implied prevalence at time 0 is consistent with the chosen prior distribution. As 10 days may truncate the generation time distribution and test-sensitivity function, we rescale the generation time distribution to sum to 1 and the test-sensitivity function to sum to its original total over the combined length of past data and the wind-in period.

2.2 The Eales method

The Eales method [15] models logit-transformed P_t using Bayesian p-splines:

$$\text{logit}P_t = \sum_{i=1}^{N} b_i B_{i,n}(t), \quad \text{logit}(x) = \log \frac{x}{1-x} \tag{10}$$

where $B_{i,n}(t)$ are basis functions and b_i are estimated spline coefficients. The coefficients give the value of the spline at the corresponding "knots" t_i while the basis functions allow for interpolation. By default, fourth-order basis functions $B_{i,4}(t)$ are used. We refer the reader to the original paper for the full construction [15].

The smoothness of these splines is controlled by the spacing of the spline knots (set by the user), and a second-order Gaussian random walk prior distribution on the spline coefficients, with standard deviation σ_{Eales} (estimated from the data):

$$b_i - b_{i-1} = b_{i-1} - b_{i-2} + u_i, \quad u_i \sim \text{N}(0, \sigma^2_{Eales}) \tag{11}$$

As $\text{logit}P_t - \text{logit}P_{t-1} \approx r_t$ for small P_t, this is approximately equivalent to modelling the growth rate using a Gaussian random walk, a similar smoothing assumption to our growth rate epidemic model.

By default, spline knots are placed approximately every 5 days (exactly 5 days when the duration of the data is divisible by 5) to balance flexibility with computational efficiency. As we generally work in integer time, there is no benefit to knot-spacing shorter than 1 day. If knots are placed on each day, we only need to evaluate the splines at the knot locations, and thus can work with b_i directly (i.e. we do not need to employ any splines). In this case, except for modelling the change in logitP_t instead of $\log P_t$, the model is equivalent to the SIMPLE (extra-binomial) model. We examine this equivalence further in supplementary section 3.

The original Eales method modelled positive swabs n_t^+ with a binomial distribution, equivalent to our basic observation model. We update this to use a beta-binomial distribution with overdispersion parameter ρ, equivalent to our extra-binomial observation model, improving both convergence of the algorithm and the quality-of-fit of the model (supplementary sections 3 and 4). We do not provide a weighted implementation of the Eales method, although this could be achieved by using the weighted observation model described above.

Growth rates (in prevalence) are back-calculated from the fitted splines by setting $r_t = \log(P_t/P_{t-1})$, where $P_{1:T}$ are sampled from the posterior distribution of the splines. A weakly informative inverse-gamma prior distribution ($\alpha = \beta = 0.0001$) is used for σ_{Eales}, a uniform prior distribution is used for ρ, and a uniform prior distribution is used for the first two spline coefficients.

Rather than explicitly model incidence, the Eales method makes a series of simplifying assumptions when estimating the reproduction number. Specifically, it assumes that at each time step t, R_t has been fixed for the past τ days (a trailing window approach akin to that used in EpiEstim [20]). If τ is chosen to be larger than both the maximum generation time and maximum duration of swab positivity, then the renewal equation (4) can be used to estimate R_t directly. In practice, $\tau = 14$ days is used to prevent oversmoothing of R_t, despite this likely being less than the maximum duration of test sensitivity for SARS-CoV-2 (which, according to [21], remains above 10% even after 21 days after infection). The trailing window approach is known to produce biased estimates of R_t [25], which can be partially accounted for by reporting estimates shifted by $\tau/2$ days [26].

Eales, Ainslie, Walters, et al. [15] provide source code to fit this model using the R programming language [27] and Rstan. We adapt their code to fit the model using the actively developed interface to Stan, cmdstanr [28]. As with Rstan, inference is performed using Hamiltonian Monte Carlo (HMC) with the adaptive No-U-Turn Sampler (NUTS). We also adjust the hyperparameters of their algorithm (increasing the maximum tree-depth to 15 and decreasing the number of

warm-up/sampling iterations to 200/300) to substantially improve convergence times, from at least 30 hours on the full REACT-1 dataset to less than 1 hour (supplementary section 4).

As in the SIMPLE method, convergence is measured by running three chains and checking that maximum $\hat{R} < 1.05$ and minimum $ESS > 100$. In this case, "hidden states" such as the growth rate are estimated alongside fixed parameters, so also feature in the convergence checks. This means the maximum $\hat{R}$ and minimum ESS diagnostics are not directly comparable to those of the SIMPLE method, although they do reflect the computationally expensive aspects of each method (estimating the fixed parameters in the SIMPLE method is much more expensive than estimating the hidden states).

2.3 The Abbott method

The Abbott method [7] models the first-order difference (other order differences are possible but not considered here) of logit-transformed daily infection incidence using a Laplace eigenfunction approximation [29, 30] to a zero-mean Gaussian process:

$$\text{logit}(I_t) = I_0 + \sum_{s=1}^{t} GP_s \tag{12}$$

where I_0 is the initial logit-transformed incidence and GP_s is the value of the approximated Gaussian process at time s. At small incidence values I_t, this is approximately equivalent to modelling the growth rate in incidence r_t^{inc} as a Gaussian process:

$$r_t^{inc} = \log I_t - \log I_{t-1} \approx \text{logit} I_t - \text{logit} I_{t-1} = GP_t \tag{13}$$

A squared exponential kernel, parameterised by variance σ^2_{Abbott} and lengthscale ℓ, is used for the Gaussian process:

$$k(GP_t, GP_s) = \sigma^2_{Abbott} \exp\left(-\frac{(t-s)^2}{2\ell^2}\right) \tag{14}$$

If the lengthscale ℓ is large relative to unit time-steps, this Gaussian process can be locally approximated by an AR(1) process: $GP_{t+1} = \phi GP_t + \epsilon_t$ where $\phi = \exp(-1/2\ell^2)$ and $\epsilon_t \sim N(0, \sigma^2_{Abbott}(1-\phi^2))$. This approximation highlights a mild similarity to our basic model: the Abbott model can be (very loosely) viewed as a Gaussian random walk, except on r_t^{inc} and with a drift towards zero growth. The strength of this drift is controlled by ℓ with larger values of ℓ

indicating a slower drift towards zero growth.

Prevalence P_t is modelled as a convolution of past incidence with a test-sensitivity function that describes how likely an individual is to test positive u days after infection (equation 5). During the COVID-19 pandemic, curves from [21] were used. We use the same curves in our implementation (including in simulated data where applicable), although these curves depend on the pathogen (including variant, in the case of SARS-CoV-2 [31]), testing procedures (e.g. swabbing technique [32] and storage during transport [12]), and the population being tested (e.g. age structure and vaccination status [33]). As a result, these curves are not universal, and should be separately estimated for each scenario. We compare multiple estimates of these curves, and the sensitivity of model results to these curves, in supplementary section 5.

The original Abbott model used a normal likelihood for observed swab positivity with mean P_t and an estimated variance, a necessary simplification given the format of their data. We employ the beta-binomial observation model instead, although the binomial or weighted observation models are also possible and can be implemented by changing only a few lines of the Stan code. The original model also leveraged antibody data, accounting for vaccination status, in their model. We focus on swab positivity data so do not consider this here. Their antibody model could be included in any of the three methods discussed in this paper. We leave this for future work.

Growth rates are back-calculated from the estimated prevalence by setting $r_t = \log(P_t/P_{t-1})$. A weakly informative inverse-gamma prior distribution is used for ℓ, a zero-mean normal distribution with standard deviation 0.1 is used as the prior distribution for σ_{Abbott}, a uniform prior distribution is used for ρ, a normal prior distribution is used for the initial logit-incidence (at the beginning of the 50-day wind-in period) with mean -4.6 and standard deviation 2, and a zero-mean normal prior distribution with standard deviation 0.25 is used for the initial growth rate in incidence (at the beginning of the 50-day wind-in period).

As the Abbott method models infection incidence directly, the reproduction number can be estimated by directly applying the renewal model (equation 4) to the estimated incidence curves.

Abbott and Funk provide source code in R and cmdstanr to fit their model. When fitting their model, we use 200 warm-up and 300 sampling iterations. As in the Eales method, convergence is measured by running three chains and checking that maximum$\hat{R} < 1.05$ and minimum$ESS > 100$. This source code assumes data of a different format to ours (their inputs are daily means and credible intervals from a different Bayesian model), so we adapt their code to fit to raw

data. We also remove code relating to their antibody model. The modified code is provided on GitHub.

We summarise the three methods considered in table 1.

Table 1: An overview of key differences in the three methods considered in this paper. Computational complexity is with respect to the precise models considered in this manuscript. Each method has benefits and drawbacks depending on the specific scenario.

	SIMPLE	Eales	Abbott
Language	Julia	R & Stan (C++)	R & Stan (C++)
Fitting method	SMC & PMMH	HMC with NUTS	HMC with NUTS
Smoothing model	Random walk on r_t or R_t	B-splines on P_t	Gaussian process on I_t
Mathematical complexity	Lowest	Moderate	Highest
Computational complexity	Moderate	High	Moderate
Assumptions made	Fewest (r_t) or many (R_t)	Few	Many
Exactness	Exact*	Exact*	Approximate

*All methods involve Monte Carlo methods and thus are subject to Monte Carlo error.

2.4 Simulated data

Simulated datasets are generated by sampling from the prior predictive distribution of a given model at fixed parameter values. In general, T days (default 100) of an epidemic are simulated with a fixed initial prevalence P_0 (default 1%) and growth rate r_0 (default 0). As the Abbott method models incidence and requires a wind-in period (default 50 days when simulating), we instead fix infection incidence at the start of the wind-in period at 0.1%. Other model-specific parameters are set to default values as outlined in table 2. These are chosen to reflect estimates from the REACT-1 study, thus representing plausible values for SARS-CoV-2-like epidemics. To prevent unrealistic scenarios, we bound simulated prevalence in the range $(0.1\%, 30\%)$ by resampling if a simulation exceeds these bounds.

Weighted survey data are generated by sampling artificial survey weights $w_{t,i}$ from a log-normal distribution (default log-mean 0 and scale $\xi = 0.5$) and normalising so $\sum_{i=1}^{n_t} w_{t,i} = 1$. To introduce bias, the probability that individual $i \in \{1, ..., n_t\}$ tests positive is set to:

$$P_{t,i} = P_t \frac{w_{t,i}}{\sum_{j=1}^{n_t} w_{t,j}^2} \tag{15}$$

This results in individuals with higher survey weights (i.e. individuals that are underrepresented in the sample) being more likely to test positive, while the weighted average of $P_{t,i}$ is equal to P_t. The simulated dataset then consists of $\{x_{t,i}, w_{t,i}\}$ where each $x_{t,i}$ is a realisation of a Bernoulli($P_{t,i}$) random variable. This imposes a linear relationship between survey weight and

Table 2: Default parameter values for simulated data, unless otherwise stated. These are chosen to reflect estimates from fitting each model to the REACT-1 study. Prevalence and incidence are as a proportion of the population.

Parameter	Default value	Description
General parameters		
T	100 days	Duration of the simulated epidemic
P_0	0.01	Initial prevalence
r_0	0	Initial growth rate (per day)
$n_t, \quad t \in \{1, \ldots, T\}$	5,000	Daily swabs
SIMPLE method (growth rate epidemic model)		
σ	0.01	Standard deviation of random walk on r_t (per day)
ρ	2×10^{-4}	Overdispersion parameter (beta-binomial observation distribution)
ξ	0.5	Scale of log-normal distribution for simulated weights
Eales method		
σ_{Eales}	0.1	Standard deviation of random walk on spline knots (per knot)
ρ	2×10^{-4}	Overdispersion parameter (beta-binomial observation distribution)
Abbott method		
σ_{Abbott}	0.08	Gaussian process kernel amplitude
ℓ	10	Gaussian process kernel lengthscale
ρ	2×10^{-4}	Overdispersion parameter (beta-binomial observation distribution)
I_{-init}	0.001	Initial incidence at the start of the wind-in period
r_{-init}	0	Initial growth rate in incidence (per day) at the start of the wind-in period

individual swab positivity, so represents a high level of bias for any assumed scale ξ. Default $\xi = 0.5$ was chosen to reflect the observed distribution of per-round survey weights in the REACT-1 study. An empirical analysis of these weights is provided in supplementary section 6.

To consider the effect of survey design and epidemic dynamics, we fit each of the SIMPLE observation models (basic, extra-binomial, and weighted) to simulated datasets of duration $T = 100$ days from each observation model (with the growth rate epidemic model). Five values of the daily sample size n_t are considered (10, 100, 1,000, 10,000, 100,000) and two values of σ (0.008, 0.016). When simulating from the extra-binomial model we assume $\rho = 2 \times 10^{-4}$, and when simulating from the weighted model we assume $\xi = 0.5$. The choice of these values were guided by estimates from the REACT-1 study (see the REACT-1 results for ρ and supplementary section 6 for ξ). Each simulation is repeated 10 times to average over stochasticity. To evaluate model performance, we consider the coverage and average width of the 95% credible intervals for estimated prevalence P_t. Equivalent results for the growth rate in swab positivity r_t are presented in supplementary section 7.1.

2.5 The REACT-1 study

The REACT-1 study was an infection prevalence survey that tested for SARS-CoV-2 infection in England between May 2020 and March 2022 [8]. Conducted over 19 rounds, a total of 2.5

million self-administered throat and nose samples were processed using RT-PCR. Daily swab positivity and sample sizes for all 19 rounds of the REACT-1 study are reported in Figure 1. An average of 6,236 samples were taken on each of the 400 days that sampling occurred, or an average of 3,564 samples per day over the 700 days spanned by the study (from 1 May 2020 to 31 March 2022 inclusive).

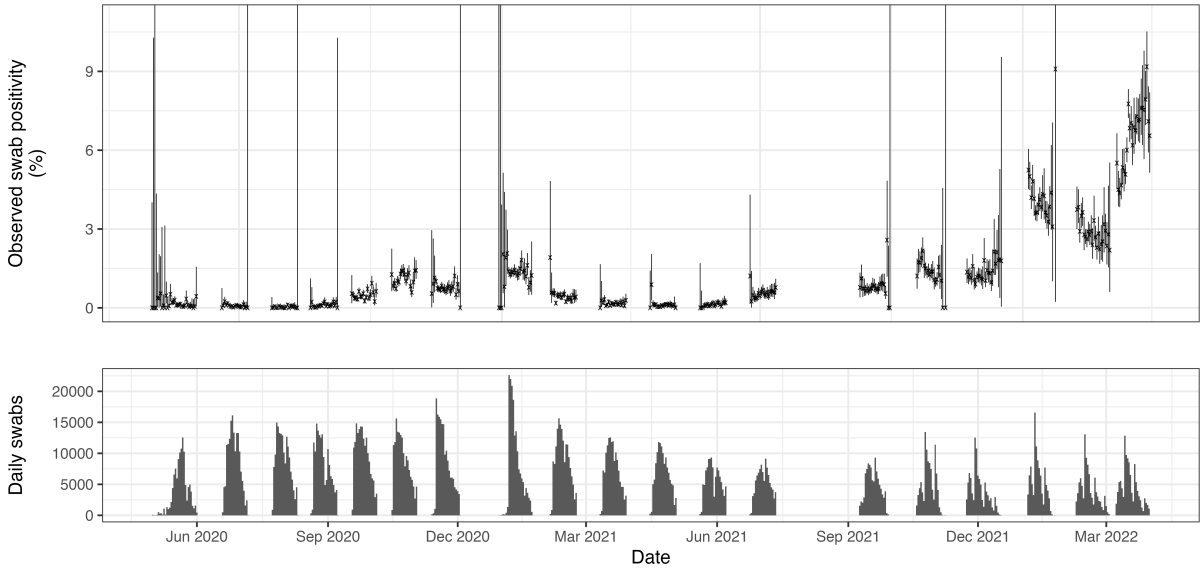

Figure 1: Daily SARS-CoV-2 swab positivity in England from the REACT-1 survey (upper) and corresponding daily sample sizes (lower). Daily 95% confidence intervals (vertical lines) for prevalence were calculated using the Agresti and Coull method [14] in the *binconf* function of the *Hmisc* package in R [34].

3 Results

3.1 Survey design and epidemic dynamics

All three observation models produce well-calibrated and similarly narrow credible intervals when fit to simulated data from the basic observation model (i.e., a survey with binomial-distributed observations), despite the extra-binomial and weighted observation models featuring an additional (and in this case, unnecessary) parameter (figure 2, column A). The width of these credible intervals decreases as n_t increases, with 1000 daily samples generally sufficient to produce credible intervals less than 1 percentage point in width, although this is highly simulation-dependent. Larger sample sizes may also allow for the estimation of prevalence by region and/or demographic, as seen in the REACT-1 study. When fit to simulated data with more variable growth rates (higher σ, dashed lines), the width of the credible intervals increases slightly, suggesting that a minor increase in daily sample size may be required when the growth rate is changing faster (to achieve the same level of precision). Finally, the weighted

observation model produces poor coverage at low sample sizes, due to the breakdown of the normal approximation to the binomial distribution.

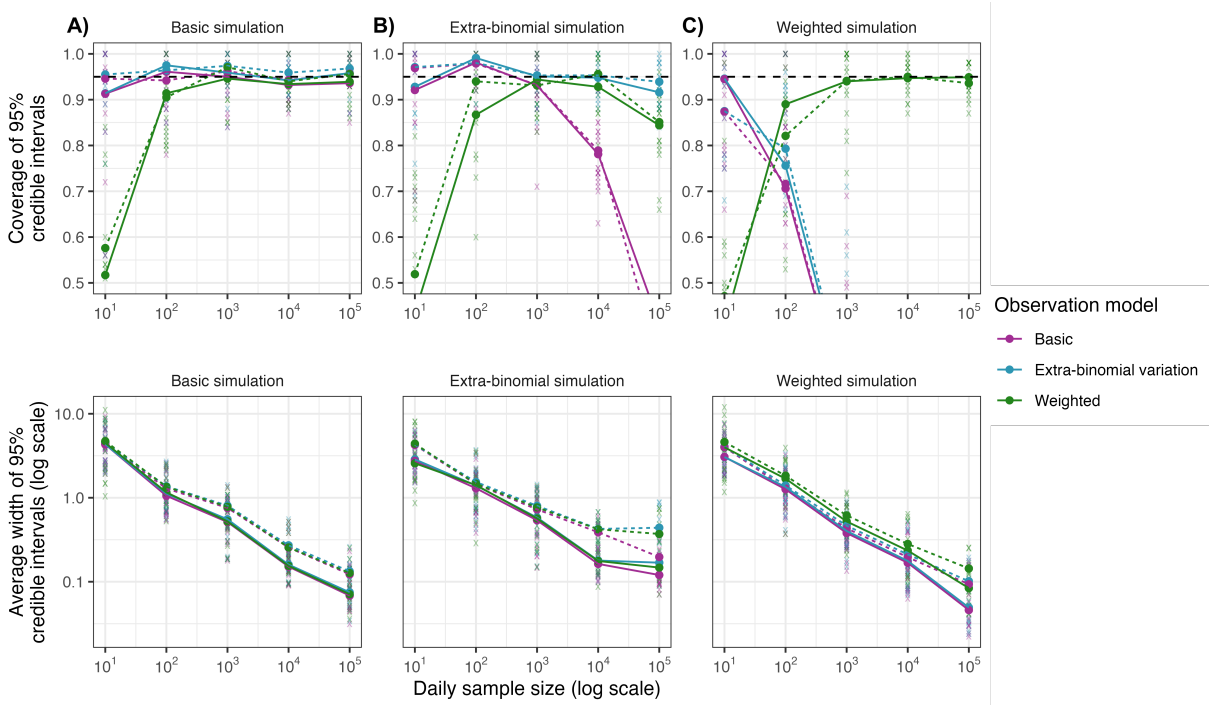

Figure 2: Coverage and average width of 95% credible intervals for prevalence P_t (%) from fitting all three observation models (purple: basic, blue: extra-binomial, green: weighted) to simulated data from each model (column A: basic, column B: extra-binomial, column C: weighted). Results from individual simulations are shown as semi-transparent crosses, with averages over 10 simulations shown as points connected by solid lines (for assumed $\sigma = 0.008$) and dashed lines (for assumed $\sigma = 0.016$). A range of assumed daily sample sizes n_t are considered (x-axis). The horizontal black dashed line indicates the target coverage of 0.95. The y-axis for coverage is truncated to $(0.5, 1.0)$, although the coverage in some cases falls outside this range: reaching a minimum average of 0.33 for the basic model fit to the extra-binomial simulations and a minimum average of 0 for the basic and extra-binomial models fit to the weighted simulations (all when $n_t = 10^5$).

When fit to simulated data featuring extra-binomial variation, only the extra-binomial model consistently produces well-calibrated estimates of P_t (figure 2, column B). At larger values of n_t, when the normal approximation to the binomial distribution is valid, the weighted model also produces well-calibrated estimates of P_t, as the additional parameter c allows the model to account for the extra-binomial variation. The basic model, while well calibrated at smaller values of n_t, produces poorly calibrated estimates as n_t increases. This is due to the basic model assuming that the observation variance is inversely proportional to n_t, forcing the credible intervals on P_t to decrease in width as n_t increases, even if additional observation noise is present. The basic model initially accounts for this by artificially increasing the estimated value of σ, allowing variation in r_t to capture the additional variation in n_t^+ (at the cost of biased estimates of σ and r_t—see supplementary section 7.2), although this breaks down as n_t gets very large.

This has real-world implications as seen on the REACT-1 dataset, where the original Eales method (using a binomial observation model) produces noticeably different estimates of r_t and P_t compared to the modified method using an extra-binomial observation model (supplementary section 4).

Finally, both the basic and extra-binomial models perform poorly when fit to simulated weighted data, with only the weighted model being able to recover the true prevalence (figure 2, column C). Unlike in the other models, estimates of growth rates are less prone to weighting-associated bias (supplementary figure S8). We emphasise caution when translating these results to real-world datasets: the simulated weighted model represents an extreme scenario where weights and individual swab positivity are perfectly correlated and the weights are known. In practice, these weights must be estimated through techniques such as random iterative method (RIM) weighting [35], which introduces additional uncertainty that is not accounted for by this model.

Given a predetermined model structure, uncertainty about r_t and P_t arises from two sources: (1) process and observation noise in the epidemic and observation model, and (2) uncertainty about the fixed parameters governing these models. While uncertainty associated with the former can only practically be reduced by collecting more data, uncertainty about fixed parameters can be decreased by using more informative prior distributions. At smaller values of n_t, a wide uniform prior distribution can result in the posterior distribution assigning probability mass to implausibly large values of σ and ρ, leading to overestimation of the width of credible intervals for r_t and P_t. Using more informative prior distributions that place less prior mass on implausibly large values can help to reduce the width of the credible intervals for r_t, P_t, and n_t^+, particularly at smaller values of n_t.

3.2 Comparing methods

We compare the SIMPLE, Eales, and Abbott methods on 10 simulated datasets from each model with default parameterisations (table 2), assuming a beta-binomial observation process. All three models produce very similar posterior predictive distributions for swab positivity, and similar posterior distributions for prevalence, suggesting that, given the same observation model, the estimation of P_t and n_t^+ is robust to the minor differences in smoothing assumptions between models (figure 3).

The Eales and Abbott models both enforce a fixed minimum amount of smoothness in the data. In the case of the Eales model, this is via spline knots being placed less frequently than the

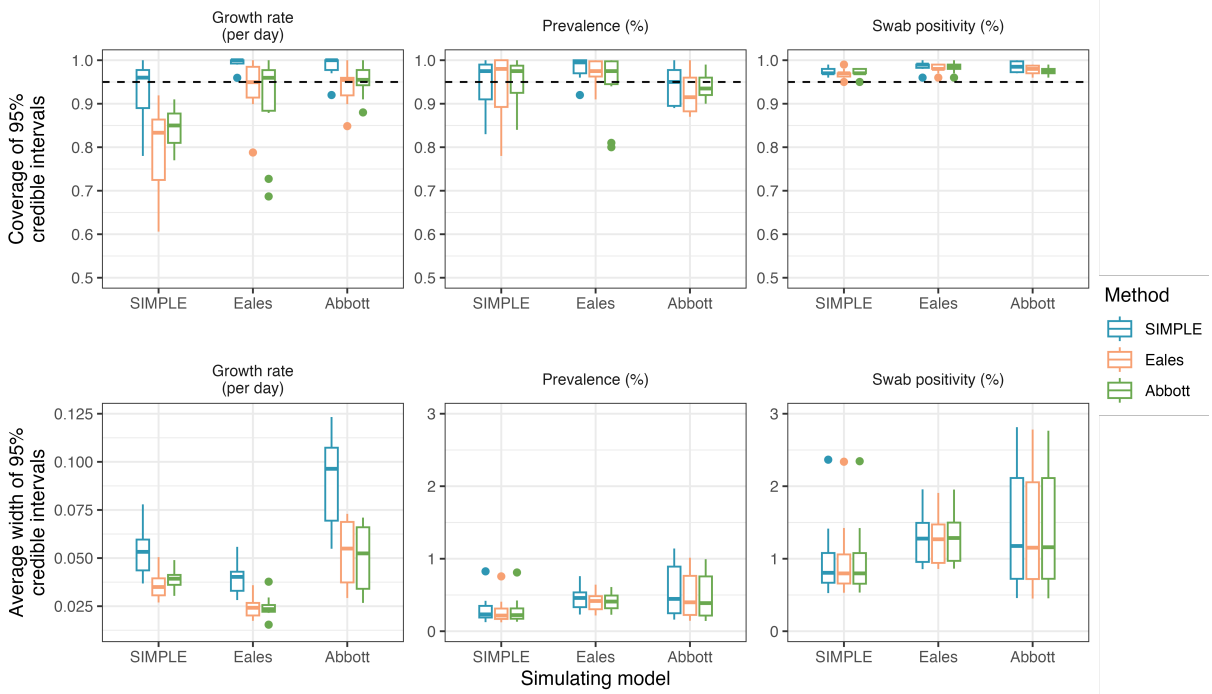

Figure 3: Coverage and average width of 95% credible intervals for the growth rate and prevalence, and coverage and average width of 95% predictive credible intervals for swab positivity. Each method (SIMPLE, Eales, and Abbott) is fit to 10 simulated datasets from each model. The horizontal black dashed line indicates the target coverage of 0.95. Boxes present the interquartile range of the results with the median shown as a horizontal line. Whiskers extend to the most extreme data point within 1.5 times the interquartile range from the box. Outliers are shown as points.

observed data; while in the case of the Abbott model, this results from modelling prevalence as a convolution of smooth incidence and the test-sensitivity function. This minimum smoothness results in the Eales and Abbott methods producing narrower credible intervals on r_t than the SIMPLE method, helpful when the true growth rate is smooth (the SIMPLE method overcovers r_t in simulations from the Eales and Abbott models), but results in undercoverage when the true growth rate is more variable (figure 3).

The total time taken to fit the models to all 30 simulations was 18m 40s for the SIMPLE (extra-binomial) method, 54m 58s for the Abbott method, and 47m 44s for the Eales method. A successful iteration of the Abbott method takes less time than the Eales method, however, the Abbott method sometimes fails to converge, requiring refitting of the model. Further refinement of the code and/or better selection of prior distributions may improve convergence times for the Abbott method. A more comprehensive comparison of the runtime of each method is provided in supplementary section 8.

3.3 The REACT-1 study

All methods (when using beta-binomial observation distributions) produce similar estimates of the growth rate r_t, prevalence P_t, and predictive swab positivity n_t^+/n_t, when fit to the REACT-1 dataset (figure 4), with an average width of 95% credible intervals for P_t of 0.26 to 0.27 percentage points. Minor differences, largely arising from different prior assumptions about the initial growth rate r_0, are observed in estimates at t near zero. All methods also produce very similar estimates of the overdispersion parameter ρ, with posterior mean values of 1.9 to 2.0 (table 3). Other parameters are not directly comparable. Finally, all methods produced posterior predictive distributions for n_t^+ that slightly overcovered the observed data (table 3).

Assuming fixed values of the smoothing parameter σ and overdispersion ρ over the entire study period of 700 days may not be appropriate [36]. To assess temporal variation in these parameters, we fit the models separately to four time periods: 1 May 2020 to 3 December 2020 (REACT-1 study rounds 1-to-7), 30 December 2020 to 12 July 2021 (REACT-1 study rounds 8-to-13), 9 September 2021 to 17 December 2021 (REACT-1 study rounds 14-to-16), and 5 January 2022 to 31 March 2022 (REACT-1 study rounds 17-to-19). These periods align approximately with changes in the dominant SARS-CoV-2 variant in England (Wildtype, Alpha, Delta, and Omicron) and large gaps in sampling.

Results from the SIMPLE method (with the growth rate epidemic model) suggest that σ was higher in study rounds 14-to-16 and 17-to-19 (central estimates of $\sigma = 0.015$) than in study rounds 1-to-7 (central estimate of $\sigma = 0.011$) and rounds 8-to-13 (central estimate of $\sigma = 0.0083$), indicating greater variability in the growth rate in later study rounds. The results also suggest that ρ was higher in study rounds 1-to-7 and 17-to-19 (central estimates of 2.6 and 2.8, respectively) than in study rounds 8-to-13 and 14-to-16 (central estimates of 1.3 and 1.1), suggesting greater observation noise at the study's start and end. These trends are consistent with results from the SIMPLE method with the reproduction number epidemic model, and the Eales and Abbott methods (table 3). While differences in parameter estimates are not statistically significant, estimates of the growth rates r_t do show some sensitivity to whether the models are fit separately to each time period or all together (supplementary section 9.1).

The SIMPLE method (with the growth rate epidemic model) and Abbott method exhibit comparable runtimes, taking 4m 12s and 6m 43s on the full dataset (all study rounds), or 14s and 29s on the smallest dataset (study rounds 17-to-19), respectively. This is verified in supplementary section 8, where a more extensive comparison of runtimes is provided. The Abbott method

can sometimes fail to converge, requiring refitting of the model, thus increasing the total time taken to fit the model. The Eales method is slower than both of these methods, particularly on larger datasets, taking 38m 54s to fit to the full dataset and 35s to fit to the smallest dataset. This slowdown is partially due to the need to increase the maximum tree-depth HMC hyperparameter from 15 to 16 for convergence on larger datasets. Finally, the SIMPLE method with the reproduction number epidemic model is the slowest on smaller datasets, but remains faster than the Eales method on larger ones (taking 23m 8s on the full dataset and 1m 23s on the smallest dataset). The runtimes reported in table 3 are for a single run of each method. A more extensive analysis of runtimes is included in supplementary section 8.

Despite each method producing similar estimates of r_t and P_t and a similar predictive distribution for n_t^+ (figure 4), a number of arbitrary decisions were made when fitting these models. For example, we present estimates of r_t and P_t from fitting to all 19 study rounds simultaneously with fixed parameters, rather than grouping the data into shorter time periods. We also assume the extra-binomial observation model is appropriate, and we do not consider survey weights (as daily-applicable weights were not available). We test these specific modelling decisions in supplementary section 9.1, 9.2, and 9.3 respectively. Furthermore, we find evidence for variant-specific parameter values in supplementary section 10.

The reproduction number

While all methods produce very similar estimates of P_t, and similar estimates of r_t when fit to the REACT-1 data, estimates of R_t differ more substantially (figure 5). All models in this section are fit using the extra-binomial observation model.

The SIMPLE and Abbott methods produce similar central estimates of R_t, with the Abbott method producing wider credible intervals on average, reflecting the incorporation of uncertainty in the test-sensitivity function (where the SIMPLE method uses the mean of the test-sensitivity function). Adapting the SIMPLE method to allow for uncertainty in this function is possible and is an avenue for future work.

The Eales method produces estimates of R_t that are delayed and oversmoothed in comparison to the SIMPLE and Abbott methods (most apparent in figure 5-B), a direct result of the trailing window approach to estimating R_t. The delay can be partially accounted for by shifting estimates by $\tau/2$ days [26] (figure 5-C), but the oversmoothing is inherent to the method. Furthermore, as a trailing window of length $\tau = 14$ days does not include the entire generation time distribution

Table 3: Results from fitting the three methods to data from the REACT-1 prevalence study. Runtimes were measured once for each dataset considered and can vary considerably. Convergence diagnostics of maximum $\hat{R}$ and minimum ESS are reported, although these are not directly comparable between the SIMPLE and Eales/Abbott methods. Measures of fit are reported as coverage of the posterior predictive distribution and average width of 95% credible intervals on P_t (in terms of percentage points). Parameter estimates are shown as posterior means with 95% credible intervals in parentheses. Note that σ_{Eales} depends on the knot-spacing, which varies slightly between study rounds, so these estimates are not directly comparable even within the same model.

Study round	**All rounds**	**1-to-7**	**8-to-13**	**14-to-16**	**17-to-19**
Observations	400	147	119	67	67
Duration (days)	700	217	195	100	86
SIMPLE method					
Stopping criteria: max $\hat{R} < 1.05$ and min ESS > 100					
Runtime	4m 12s	42s	46s	25s	14s
Max $\hat{R}$/Min ESS	1.01/113	1.02/110	1.02/100	1.04/110	1.04/129
Coverage of 95% Cr.I. n_t^+	97.2%	97.3%	97.5%	98.5%	98.5%
Avg. width of 95% Cr.I. P_t	0.27	0.16	0.17	0.37	0.84
Parameter σ	0.010	0.011	0.0083	0.015	0.015
	(0.0071, 0.011)	(0.0055, 0.022)	(0.0041, 0.016)	(0.0068, 0.028)	(0.0075, 0.028)
Parameter ρ	1.9	2.6	1.3	1.1	2.8
	(1.3, 2.7)	(1.6, 4.5)	(0.61, 2.2)	(0.15, 3.1)	(1.2, 5.0)
SIMPLE method (reproduction number epidemic model)					
Stopping criteria: max $\hat{R} < 1.05$ and min ESS > 100					
Runtime	23m 38s	3m 9s	1m 26s	59s	1m 23s
Max $\hat{R}$/Min ESS	1.02/101	1.04/106	1.02/101	1.03/109	1.04/127
Coverage of 95% Cr.I. n_t^+	97.2%	96.6%	97.5%	98.5%	98.5%
Avg. width of 95% Cr.I. P_t	0.26	0.14	0.16	0.36	0.79
Parameter σ_R	0.069	0.063	0.056	0.12	0.11
	(0.050, 0.097)	(0.034, 0.12)	(0.023, 0.11)	(0.046, 0.23)	(0.050, 0.22)
Parameter ρ	2.0	2.6	1.3	1.2	3.4
	(1.4, 2.6)	(1.8, 3.8)	(0.55, 2.2)	(0.12, 2.8)	(1.6, 6.0)
Eales method*					
Stopping criteria: 500 samples (200 warm-up and 300 sampling)					
Runtime	38m 54s*	2m 58s	1m 14s	43s	35s
Max $\hat{R}$/Min ESS	1.04/152*	1.02/380	1.02/259	1.01/245	1.01/260
Coverage of 95% Cr.I. n_t^+	96.5%	96.6%	97.5%	97.0%	98.5%
Avg. width of 95% Cr.I. P_t	0.26	0.14	0.15	0.34	0.74
Parameter σ_{Eales}	0.11	0.097	0.074	0.15	0.15
	(0.084, 0.16)	(0.051, 0.17)	(0.033, 0.15)	(0.060, 0.30)	(0.079, 0.28)
Parameter ρ	1.9	2.7	1.3	0.98	2.9
	(1.4, 2.5)	(1.7, 4.1)	(0.59, 2.3)	(0.075, 2.6)	(1.2, 5.4)
Abbott method					
Stopping criteria: 500 samples (200 warm-up and 300 sampling)					
Runtime	6m 43s	47s	49s	40s	29s
Max $\hat{R}$/Min ESS	1.03/223	1.02/163	1.02/124	1.02/144	1.02/164
Coverage of 95% Cr.I. n_t^+	97.5%	98.0%	97.5%	95.5%	97.0%
Avg. width of 95% Cr.I. P_t	0.27	0.15	0.17	0.33	0.75
Parameter σ_{Abbott}	0.074	0.069	0.077	0.10	0.081
	(0.033, 0.21)	(0.026, 0.19)	(0.024, 0.22)	(0.031, 0.23)	(0.026, 0.22)
Parameter ℓ	8.9	22	19	3.9	7.3
	(0.38, 22)	(0.58, 79)	(0.29, 100)	(0.18, 20)	(0.17, 27)
Parameter ρ	1.9	2.7	1.3	1.0	3.1
	(1.3, 2.5)	(1.7, 4.1)	(0.54, 2.3)	(0.16, 2.5)	(1.5, 5.8)

*In order to obtain convergence of the Eales method on the full REACT-1 dataset, the "maximum tree-depth" HMC hyperparameter was increased from 15 to 16, resulting in an increase in runtime.

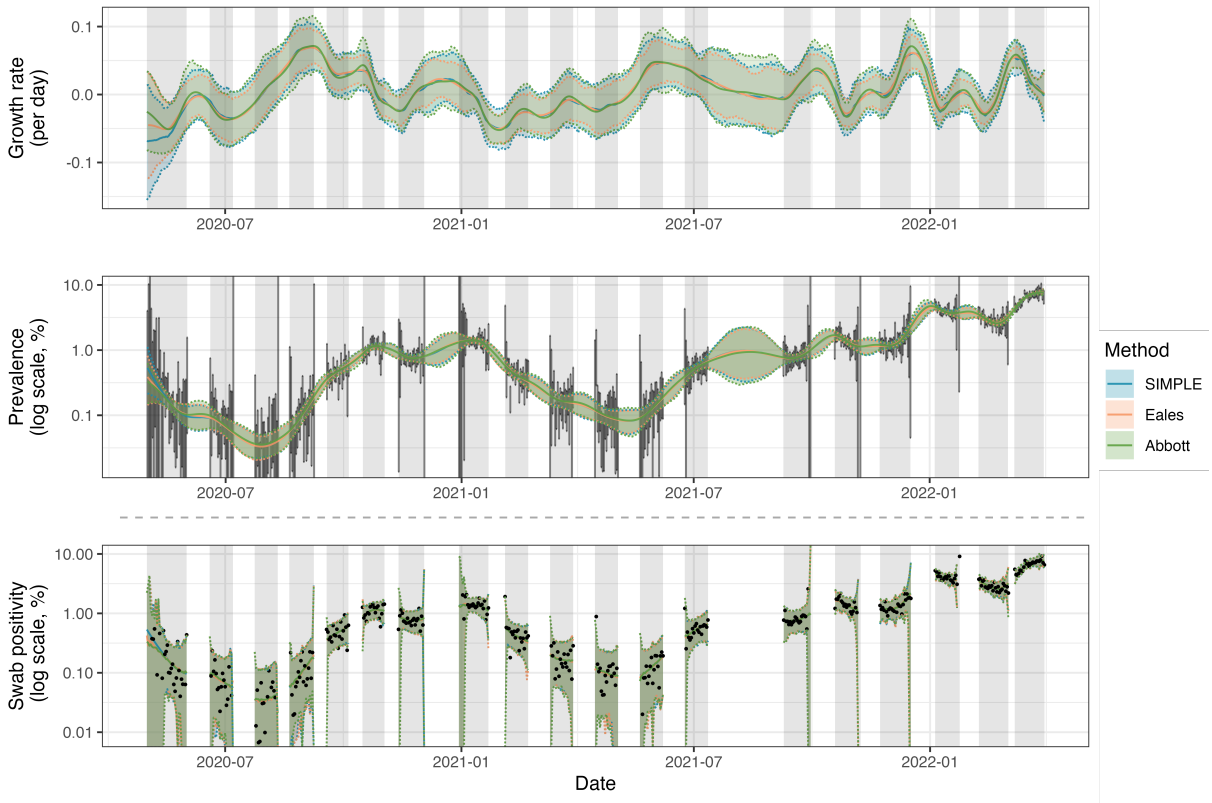

Figure 4: Estimates of the growth rate r_t, prevalence P_t, and predictive swab positivity n_t^+/n_t, for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022 using data from the REACT-1 study. All three methods are fit assuming a beta-binomial observation distribution. Solid coloured lines show the posterior means while shading and dashed lines show 95% credible intervals (of the posterior distribution for r_t and P_t, and of the posterior predictive distribution for n_t^+/n_t). Independent daily confidence intervals from the Agresti-Coull method [14] for P_t are shown in vertical grey lines. The data, daily observed swab positivity n_t^+/n_t, are shown in black points. Grey shading indicates the periods in which sampling was conducted.

or test-sensitivity duration, additional biases are introduced. Note that, while the Eales method was used to estimate R_t in the published results from the REACT-1 study, our estimates may not align with these due to our use of a beta-binomial observation model instead of the original binomial model.

Despite these differences, there is no clear best approach. Most notably, estimates from the SIMPLE and Abbott methods depend on the assumed test-sensitivity function, which can vary significantly between settings. This is explored further in supplementary section 5, where we compare three different curves estimated for SARS-CoV-2 RT-PCR tests, and demonstrate how the choice of curve impacts estimates of R_t.

As the reproduction number epidemic model in the SIMPLE method is no longer Markovian, we must use fixed-lag resampling during the parameter inference stage [17], leading to an increase in computation time (23m 38s for the full dataset or 1m 23s on the smallest dataset). The faster

runtime of the SIMPLE method with the growth rate epidemic model could be obtained by fitting a Markovian model (instead of the renewal model), such as a compartmental susceptible-infectious-recovered-type model, which has the added benefit of not requiring a test-sensitivity function. However, this type of model places additional assumptions on the underlying epidemic dynamics, which can be difficult to validate. We leave this for future work.

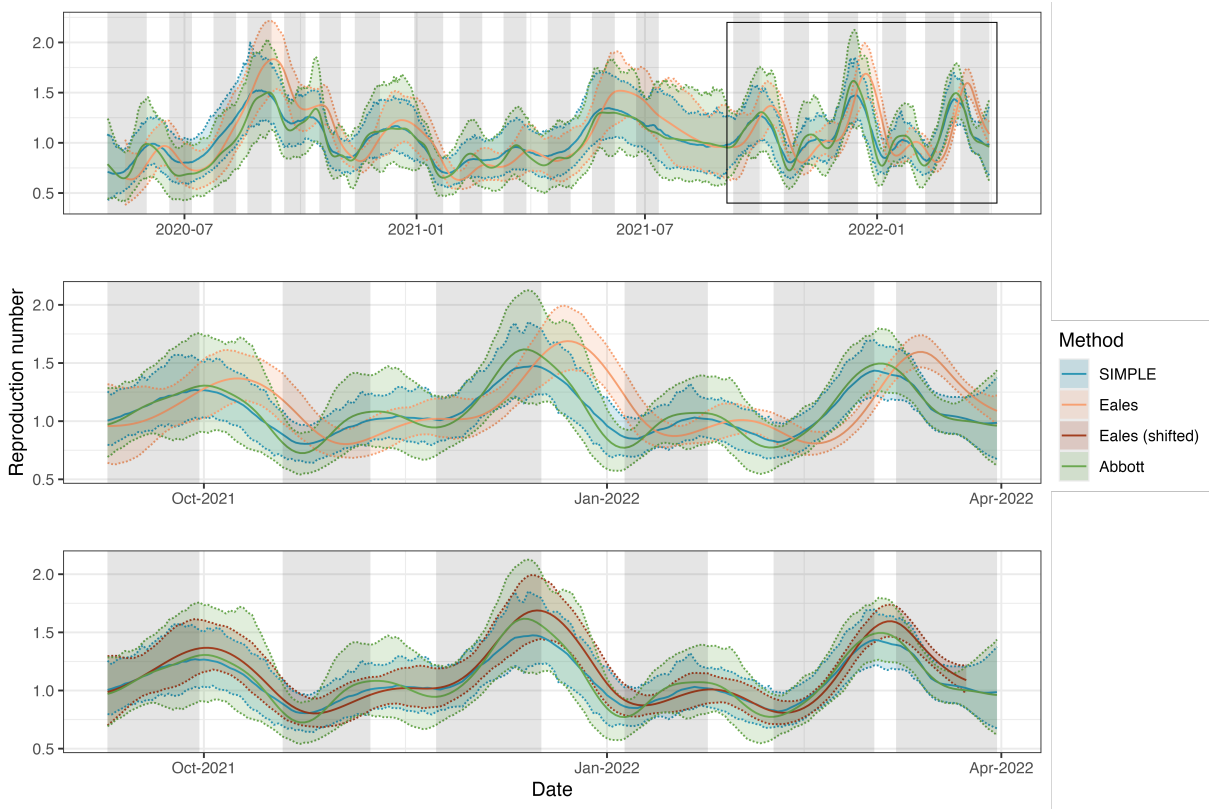

Figure 5: Estimates of the instantaneous reproduction number R_t for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022 using data from the REACT-1 study. All methods are fit assuming a beta-binomial observation distribution. Solid coloured lines show central estimates while shading and dashed lines show 95% credible intervals. Grey shading indicates the periods in which sampling was conducted. The second panel shows the same estimates for a shorter period (9 September 2021 to 31 March 2022), emphasising the differences between the three methods. The third panel shows the same estimates, with estimates from the Eales method shifted by $\tau/2 = 7$ days to partially account for bias induced by the trailing window smoothing assumption.

4 Discussion

Smoothing prevalence data allows multiple days to inform point estimates, reducing noise and increasing the confidence in estimates. However, all smoothing methods make assumptions, whether explicitly or implicitly, about the underlying process, and results can be sensitive to these assumptions.

A key concern highlighted in this paper is the presence of overdispersion in the observed data.

If not properly accounted for, this can lead to biased prevalence estimates and credible intervals with poor coverage. Growth rate estimates also become more variable, as the model attempts to explain extra-binomial noise through artificial variability in r_t. In addition to the simulated results presented in this paper, we also find that this impacts real-world estimates of P_t and r_t in the REACT-1 study. Once overdispersion is appropriately handled, all methods produce accurate prevalence estimates, even when fit to data generated by other models. This is because prevalence estimates are largely driven by the observed data, not by the smoothing assumptions.

Even when the data are not generated by a beta-binomial sampling process, the overdispersion parameter ρ serves as a general error term. The similarity of estimates of ρ between the methods (when fit to real-world data from the REACT-1 study) suggests that this parameter is estimable from observed data independently of the assumed smoothing mechanism. Growth rate estimates (and to a lesser extent, prevalence estimates) in the REACT-1 study were sensitive to the inclusion of this parameter. As the allowance for overdispersion, even when simulated data were generated from a binomial model, had little impact on the accuracy of estimates, we recommend including this parameter in all models. If survey weights are available and the daily sample size is sufficiently large (typically $n_t > 100$, although low-prevalence scenarios may require more), then the weighted model is able to account for both overdispersion and sampling bias, and thus should be the first choice.

In addition to appropriately smoothing prevalence data, the methods presented in this paper allow for the estimation of the growth rate in prevalence r_t. Unlike estimates of prevalence, assumed smoothing dynamics have a substantial impact on estimates of r_t. Only the SIMPLE method produced 95% credible intervals with consistently good coverage of this quantity, at the cost of substantially wider intervals. However, as prevalence is naturally a convolution of past incidence, the additional smoothness implied by the SIMPLE method with the reproduction number epidemic model, or the Abbott method, may provide an appropriate way to reduce estimated uncertainty about r_t.

Estimating R_t from prevalence data requires assumptions about the relationship between incidence and swab positivity. This can be achieved by assuming R_t is constant over a sufficiently long time period (Eales method), by incorporating a test-sensitivity function (SIMPLE and Abbott methods), or by fitting a fully mechanistic model (not considered here). Each approach introduces different sources of potential bias. The trailing window assumption in the Eales method guarantees some degree of misspecification [25], which can only be partially mitigated

by shifting estimates by $\tau/2$ days [26]. Test-sensitivity functions offer an alternative but also vary substantially: several distinct estimates of this function for SARS-CoV-2 RT-PCR tests were produced during the COVID-19 pandemic [21, 32, 37, 38], with meaningful differences in both duration and peak sensitivity. Ideally this function would be estimated for each epidemiological setting being modelled, which would require repeat swabs from a subset of survey participants, although collecting sufficient data may be infeasible in low-prevalence scenarios.

An alternative way to relate incidence to prevalence is to model individual-level cycle threshold (Ct) values [33]. Instead of relying on an assumed test-sensitivity function, this approach fits a model of Ct values as a function of time since infection, replacing the observation model for aggregated data with one for individual Ct data. Models for individual Ct value trajectories have previously been estimated from cross-sectional data, without needing repeated sampling from individuals. This method makes fuller use of the information contained in Ct values, rather than reducing test results to binary outcomes, but requires access to individual-level data, is computationally intensive, and depends on correctly specifying the Ct model.

Our SIMPLE method is mathematically simpler than the Eales and Abbott methods, requires no external software (e.g. Stan), is more computationally efficient than the Eales method, and avoids the approximations used in the Abbott method. However, there are two main limitations. First, the resampling step in the PMMH algorithm is computationally expensive. For Markovian models, past-state resampling can be disabled during PMMH, enabling fast inference. For non-Markovian models, past-state resampling is necessary, slowing down the PMMH algorithm, as seen in the SIMPLE model for the reproduction number. In these cases, the Abbott method is more computationally efficient. Second, PMMH relies on stochastic likelihood estimates, whose variance increases with data length and model complexity. This results in $O(T^2)$ time complexity for parameter inference, which was not a bottleneck for the datasets used here ($T \leq 700$) but will become important for longer time series. While we use relatively simple models in this paper, more complex models would benefit from more advanced algorithms [39].

While we make several improvements to the Eales and Abbott methods—including using beta-binomial observation distributions and modifying HMC hyperparameters—further optimisation is likely possible which could improve their relative performance. In particular, the Abbott method sometimes fails to converge. Reparameterising the model, or using better-specified initial values, could reduce or prevent the occurrence of this.

Estimate precision can be improved by leveraging additional data sources. The original Abbott

method, for example, included an observation model for antibody data [7]. If individual-level data are available, the approach of Pouwels, House, Pritchard, et al. [9], later refined in [40], models individual probabilities of testing positive. This approach was used to produce official estimates of prevalence in the ONS Coronavirus Infection Survey. In this method, population-level estimates are produced by poststratifying the individual-level estimates by demographic-geographical response types. This has several advantages: multilevel regression and poststratification has been shown to outperform classical survey weighting [41], and partial pooling improves estimation of demographic effects (whereas the methods presented in this paper produce demographic-specific estimates by fitting separate models). However, this approach introduces structural assumptions on the relationship between individual positivity and covariates, and requires access to individual-level data.

Our findings indicate that while all three methods provide robust estimates of prevalence and observed positive swabs, minor differences in their smoothing assumptions can impact growth rate estimates. Applying these methods to data from the REACT-1 study, we observed that all three methods produced similar estimates of prevalence and growth rates, with the SIMPLE and Abbott methods demonstrating greater computational efficiency. This paper offers a suite of validated tools that can be readily adapted for future epidemic surveys.

Data availability statement

All source code and data (simulated and from the REACT-1 survey) used in this study are available at: https://github.com/nicsteyn2/EpidemicSurveySmoothing.

Financial disclosure statement

N.S. acknowledges support from the Oxford-Radcliffe Scholarship from University College, Oxford, the EPSRC CDT in Modern Statistics and Statistical Machine Learning (Imperial College London and University of Oxford), and A. Maslov for studentship support. M.C-H. acknowledges support from Cancer Research UK, Population Research Committee Project grant "Mechanomics" (grant 22184), the H2020-EXPANSE (Horizon 2020 grant 874627) and H2020-LongITools (Horizon 2020 grant 874739). P.E. is director of the MRC Centre for Environment and Health (MR/L01341X/1 and MR/S019669/1). P.E. acknowledges support from the Department of Health & Social Care in England (NIHR and UKRI, REACT-LC, COV-LT-0040), the Medical Research Council (MR/V030841/1), and the UK Health Security Agency

for the REACT-3 study (2024-2025). C.A.D. acknowledges support from the MRC Centre for Global Infectious Disease Analysis, the NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, the NIHR funded Vaccine Efficacy Evaluation for Priority Emerging Diseases (PR-OD-1017-20007), and the Oxford Martin Programme in Digital Pandemic Preparedness. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

References

[1] Jillian Murray and Adam L. Cohen. "Infectious Disease Surveillance". In: *International Encyclopedia of Public Health* (2017), pp. 222–229. DOI: 10.1016/B978-0-12-803678-5.00517-8.

[2] Harsh Dutta, Geetanjali Kaushik, and Venkatesh Dutta. "Wastewater-Based Epidemiology: A New Frontier for Tracking Environmental Persistence and Community Transmission of COVID-19". In: *Environmental Science and Pollution Research* 29.57 (2022), pp. 85688–85699. ISSN: 1614-7499. DOI: 10.1007/s11356-021-17419-0.

[3] Linus Bengtsson, Jean Gaudart, Xin Lu, et al. "Using Mobile Phone Data to Predict the Spatial Spread of Cholera". In: *Scientific Reports* 5.1 (2015), p. 8923. ISSN: 2045-2322. DOI: 10.1038/srep08923.

[4] Google. *COVID-19 Community Mobility Report*. URL: https://www.google.com/covid19/mobility.

[5] David A. Broniatowski, Michael J. Paul, and Mark Dredze. "National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic". In: *PLOS ONE* 8.12 (2013), e83672. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0083672.

[6] Valerie C. Bradley, Shiro Kuriwaki, Michael Isakov, et al. "Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake". In: *Nature* 600.7890 (2021), pp. 695–700. ISSN: 1476-4687. DOI: 10.1038/s41586-021-04198-4.

[7] Sam Abbott and Sebastian Funk. *Estimating Epidemiological Quantities from Repeated Cross-Sectional Prevalence Measurements*. Preprint. Epidemiology, 2022. DOI: 10.1101/2022.03.29.22273101.

[8] Paul Elliott, Matthew Whitaker, David Tang, et al. "Design and Implementation of a National SARS-CoV-2 Monitoring Program in England: REACT-1 Study". In: *American Journal of Public Health* 113.5 (2023), pp. 545–554. ISSN: 1541-0048. DOI: 10.2105/AJPH.2023.307230.

[9] Koen B Pouwels, Thomas House, Emma Pritchard, et al. "Community Prevalence of SARS-CoV-2 in England from April to November, 2020: Results from the ONS Coronavirus Infection Survey". In: *The Lancet Public Health* 6.1 (2021), e30–e38. ISSN: 24682667. DOI: 10.1016/S2468-2667(20)30282-6.

[10] Thomas Ward, Martyn Fyles, Alex Glaser, et al. "The Real-Time Infection Hospitalisation and Fatality Risk across the COVID-19 Pandemic in England". In: *Nature Communications* 15.1 (2024), p. 4633. ISSN: 2041-1723. DOI: 10.1038/s41467-024-47199-3.

[11] Oliver Eales, David Haw, Haowei Wang, et al. "Dynamics of SARS-CoV-2 Infection Hospitalisation and Infection Fatality Ratios over 23 Months in England". In: *PLOS Biology* 21.5 (2023), e3002118. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.3002118.

[12] Marc Chadeau-Hyam, Haowei Wang, Oliver Eales, et al. "SARS-CoV-2 Infection and Vaccine Effectiveness in England (REACT-1): A Series of Cross-Sectional Random Community Surveys". In: *The Lancet Respiratory Medicine* 10.4 (2022), pp. 355–366. ISSN: 2213-2600, 2213-2619. DOI: 10.1016/S2213-2600(21)00542-7.

[13] *Total Cost of the COVID-19 Infection Survey - Office for National Statistics*. https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/totalcostofthecovid19infectionsurvey.

[14] Alan Agresti and Brent A. Coull. "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions". In: *The American Statistician* 52.2 (1998), pp. 119–126. ISSN: 0003-1305. DOI: 10.2307/2685469. JSTOR: 2685469.

[15] Oliver Eales, Kylie E. C. Ainslie, Caroline E. Walters, et al. "Appropriately Smoothing Prevalence Data to Inform Estimates of Growth Rate and Reproduction Number". In: *Epidemics* 40 (2022), p. 100604. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2022.100604.

[16] Arnaud Doucet, Nando De Freitas, and Neil Gordon, eds. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. New York: Springer, 2001. ISBN: 978-0-387-95146-1.

[17] Nicholas Steyn, Kris V. Parag, Robin N. Thompson, and Christl A. Donnelly. "A Primer on Inference and Prediction with Epidemic Renewal Models and Sequential Monte Carlo". In: *arXiv* arXiv:2503.18875 (2025). DOI: 10.48550/arXiv.2503.18875. eprint: 2503.18875.

[18] Ruth McCabe, Gabriel Danelian, Jasmina Panovska-Griffiths, and Christl A. Donnelly. "Inferring Community Transmission of SARS-CoV-2 in the United Kingdom Using the ONS COVID-19 Infection Survey". In: *Infectious Disease Modelling* 9.2 (2024), pp. 299–313. ISSN: 2468-0427. DOI: 10.1016/j.idm.2024.01.011.

[19] Kris V. Parag, Robin N. Thompson, and Christl A. Donnelly. "Are Epidemic Growth Rates More Informative than Reproduction Numbers?" In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 185.Supplement_1 (2022), S5–S15. ISSN: 0964-1998. DOI: 10.1111/rssa.12867.

[20] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics". In: *American Journal of Epidemiology* 178.9 (2013), pp. 1505–1512. ISSN: 0002-9262. DOI: 10.1093/aje/kwt133.

[21] Joel Hellewell, Timothy W. Russell, Rebecca Matthews, et al. "Estimating the Effectiveness of Routine Asymptomatic PCR Testing at Different Frequencies for the Detection of SARS-CoV-2 Infections". In: *BMC Medicine* 19.1 (2021), p. 106. ISSN: 1741-7015. DOI: 10.1186/s12916-021-01982-x.

[22] Alberto Luceño and Federico De Ceballos. "Describing Extra-Binomial Variation with Partially Correlated Models". In: *Communications in Statistics - Theory and Methods* 24.6 (1995), pp. 1637–1653. ISSN: 0361-0926. DOI: 10.1080/03610929508831576.

[23] Andrew Gelman and Donald B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (1992), pp. 457–472. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177011136.

[24] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. "Julia: A Fresh Approach to Numerical Computing". In: *SIAM Review* 59.1 (2017), pp. 65–98. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/141000671.

[25] Nicholas Steyn and Kris V. Parag. *Robust Uncertainty Quantification in Popular Estimators of the Instantaneous Reproduction Number*. 2025. DOI: 10.1101/2024.10.22.24315918.

[26] Katelyn M. Gostic, Lauren McGough, Edward B. Baskerville, et al. "Practical Considerations for Measuring the Effective Reproductive Number, Rt". In: *PLOS Computational Biology* 16.12 (2020), e1008409. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008409.

[27] R Core Team. *R: A Language and Environment for Statistical Computing*. Manual. Vienna, Austria, 2021.

[28] Jonah Gabry, Rok Češnovar, Andrew Johnson, and Steve Bronder. *Cmdstanr: R Interface to 'CmdStan'*. Manual. 2024.

[29] Arno Solin and Simo Särkkä. "Hilbert Space Methods for Reduced-Rank Gaussian Process Regression". In: *Statistics and Computing* 30.2 (2020), pp. 419–446. ISSN: 0960-3174, 1573-1375. DOI: 10.1007/s11222-019-09886-w.

[30] Gabriel Riutort-Mayol, Paul-Christian Bürkner, Michael R. Andersen, Arno Solin, and Aki Vehtari. *Practical Hilbert Space Approximate Bayesian Gaussian Processes for Probabilistic Programming*. 2022. DOI: 10.48550/arXiv.2004.11408. arXiv: 2004.11408 [stat].

[31] Paolo Calistri, Laura Amato, Ilaria Puglia, et al. "Infection Sustained by Lineage B.1.1.7 of SARS-CoV-2 Is Characterised by Longer Persistence and Higher Viral RNA Loads in Nasopharyngeal Swabs". In: *International Journal of Infectious Diseases* 105 (2021), pp. 753–755. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2021.03.005.

[32] Rachelle N Binny, Patricia Priest, Nigel P French, et al. "Sensitivity of Reverse Transcription Polymerase Chain Reaction Tests for Severe Acute Respiratory Syndrome Coronavirus 2 Through Time". In: *The Journal of Infectious Diseases* 227.1 (2023), pp. 9–17. ISSN: 0022-1899. DOI: 10.1093/infdis/jiac317.

[33] James A. Hay, Lee Kennedy-Shaffer, Sanjat Kanjilal, et al. "Estimating Epidemiologic Dynamics from Cross-Sectional Viral Load Distributions". In: *Science* 373.6552 (2021), eabh0635. DOI: 10.1126/science.abh0635.

[34] Frank E Jr Harrell. *Hmisc*. 2024. DOI: 10.32614/CRAN.package.Hmisc.

[35] Trevor Sharot. "Weighting Survey Results". In: *Journal of the Market Research Society* 28.3 (1986), pp. 269–284.

[36] Bernard Cazelles, Clara Champagne, and Joseph Dureau. "Accounting for Non-Stationarity in Epidemiology by Embedding Time-Varying Parameters in Stochastic Models". In: *PLOS Computational Biology* 14.8 (2018), e1006211. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006211.

[37] Lauren M. Kucirka, Stephen A. Lauer, Oliver Laeyendecker, Denali Boon, and Justin Lessler. "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure". In: *Annals of Internal Medicine* 173.4 (2020), pp. 262–267. ISSN: 0003-4819. DOI: 10.7326/M20-1495.

[38] Zhen Zhang, Qifang Bi, Shisong Fang, et al. "Insight into the Practical Performance of RT-PCR Testing for SARS-CoV-2 Using Serological Data: A Cohort Study". In: *The Lancet Microbe* 2.2 (2021), e79–e87. ISSN: 2666-5247. DOI: 10.1016/S2666-5247(20)30200-7.

[39] Nikolas Kantas, Arnaud Doucet, Sumeetpal S. Singh, Jan Maciejowski, and Nicolas Chopin. "On Particle Methods for Parameter Estimation in State-Space Models". In: *Statistical Science* 30.3 (2015), pp. 328–351. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/14-STS511.

[40] Koen B. Pouwels, David W. Eyre, Thomas House, et al. "Improving the Representativeness of UK's National COVID-19 Infection Survey through Spatio-Temporal Regression and Post-Stratification". In: *Nature Communications* 15.1 (2024), p. 5340. ISSN: 2041-1723. DOI: 10.1038/s41467-024-49201-4.

[41] Andrew Gelman, Jeffrey Lax, Justin Phillips, Jonah Gabry, and Robert Trangucci. "Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion". In: (2018). URL: https://sites.stat.columbia.edu/gelman/research/unpublished/MRT(1).pdf.