

Modelling serological data using reversible jump Markov chain Monte Carlo

David Hodgson¹, James Hay² and Adam Kucharski¹

¹CMMID, London School of Hygiene and Tropical Medicine, London, UK

²Pandemic Sciences Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Text version: 0.1

This document is work-in-progress, and its contents may change in future versions.

February 13, 2024

Abstract

This document outlines a method for inferring correlates of protection against infection using serological data. Many studies infer i) prevalence from serological data using seropositivity thresholds (titre above a pre-defined value) or ii) incidence rates using seroconversion thresholds (change in titre between two-time points above a pre-defined value). These threshold values are determined from empirical data from previous studies and are often used as a fixed heuristic for identifying infection across all individuals. However, seropositivity and seroconversion rates for many pathogens vary according to host factors (e.g. age, infection history, sex) and time since infection, meaning using these fixed thresholds may lead to incorrectly estimated seroprevalence and seroincidence rates. Antibody kinetic modelling can account for individual-level variability arising from host factors and time since infection. Here, a two-phase antibody kinetic function is defined: one function for the infected individuals and a function for the non-infected individuals. Using this, it is possible to probabilistically determine the likelihood of infection based on the change in titre between two bleeds. Using these inferences, we can determine a functional form for the correlate of protection: the probability of infection for a given titre value at exposure. However, as this correlate influences the probability of infection given a titre, it will also influence the probability that an individual is in the infection or non-infection state, influencing the antibody kinetics of an individual. In addition, to determine a correlate of protection, we must identify individuals who were exposed (have the potential to be infected if they had no existing immunity) but not infected, which can be challenging to determine with surveillance data. Here, we derive a reversible-jump MCMC (RJ-MCMC) algorithm, which can use serological data to infer i) the two-phase antibody kinetics, ii) the correlate of protection, and iii) the probability of infection and exposure within one statistical framework (**Figure 0.1**). We show that this algorithm can consistently recover the correlate of protection against infection on multiple simulated serological datasets.

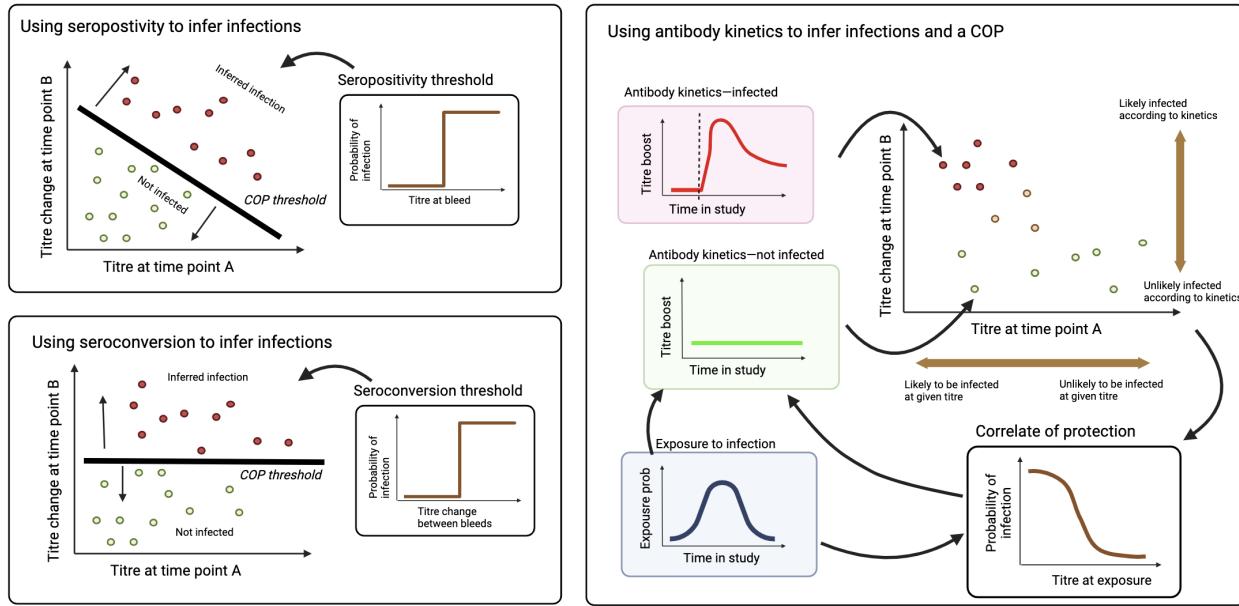


Figure 0.1: Schematic showing how to use seropositivity, seroconversion, and antibody kinetics can be used to infer infections and in the case of antibody kinetics, also infer correlates of protection

1 Overview of serological modelling

1.1 Introduction

Serological samples can be analysed to detect the presence of biomarkers made in response to an infection long after the infection has cleared.[1] Therefore, analysing serological samples allows researchers and healthcare professionals to deduce crucial information about the epidemiology of a pathogen at the individual and population level, which active virological surveillance systems may otherwise miss.

On the individual level, after measuring antibodies to a specific pathogen, infection is usually inferred using either i) an antibody threshold level (seropositive) or ii) a threshold fold-rise between a pair of samples (seroconverted).[2] Often, researchers are interested in understanding how seropositivity and seroconversion rates change according to controlled host factors, such as age, geography, living conditions, sexual behaviour, etc.[3, 4, 5] On the population level, serological samples which are representative of a population (e.g. cross-sectional samples) can be used to estimate the prevalence of infectious diseases (seroprevalence) and determine how seroprevalence changes over time according to host factors.[6, 7, 8] Estimates of seropositivity, seroconversion, and seroprevalence can help the understanding of the immune system's ability to combat various pathogens, aid in developing new targeted intervention programmes and provide insights into the transmission dynamics of infectious diseases. The methods used to analyse serological samples to inform epidemiology and public health policy of infectious diseases have been termed 'serodynamics' and have recently been reviewed.[9]

Serological samples play an increasingly important role in public health efforts to combat and control infectious diseases.[2, 10] However, inferring infection through seropositivity or seroconversion requires deriving an absolute or relative threshold value, and these are often determined by rule-of-thumb heuristics (e.g. for influenza: 4-fold-rise for conversion, titre of 1:40 HAI for seropositivity).[11] However, antibody responses vary greatly between individuals for many pathogens. Therefore, relying on these heuristics to determine infections in serology studies can lead to incorrect infection status being inferred, leading to biased estimates of prevalence.[6, 12] Consequently, a better understanding of the kinetics of antibody

trajectories post-vaccination and infection can help assess the accuracy of existing heuristics and better inform infection status.

1.2 Antibody kinetics

Modeling antibody kinetics involves using mathematical and statistical techniques to simulate the trajectories of antibodies in response to an infection or vaccination.[9] Typically, this involves using mathematical equations and statistical methods to describe the time-dependent changes in antibody levels within an individual or a population. This process is essential for understanding how antibody levels evolve and, therefore, potentially protect against infectious diseases. Various functional forms have been used to model the individual-level kinetics of antibody trajectories,[13, 14, 15, 16, 17, 18] typically it follows a three-stage process:

- *Initial Response:* The trajectories start by capturing the initial antibody response to a pathogen or vaccine. A rapid increase in antibody levels characterises this phase as the immune system recognises and mounts a defence against the antigen. The speed and relative contributions of memory versus de novo responses depend on whether the exposure is primary or secondary.
- *Peak Antibody Level:* The trajectories then rise to a peak antibody level, the highest antibody concentration reached during the immune response. This peak can vary depending on factors like the strength of the immune response and the route of and composition of antigen(s) in the exposure.
- *Decay Phase:* After the peak, there is a decline in antibody levels and the plasmablasts which secrete them. Therefore, antibodies have a finite half-life in the bloodstream, and their concentration gradually decreases as the pathogen is cleared or the vaccine antigen wanes. Over time, antibodies are continuously secreted from newly established long-lived plasma cells. Therefore, decay phase trajectories often converge to a set-point titre or a slow long-term waning rate.[16, 19, 20]

Modeling antibody kinetics provides several important benefits. First, by understanding the rate of antibody decline, models can estimate how long an individual's immunity is likely to last after infection or vaccination. Kinetics also help optimise vaccination strategies and identify the optimal timing and frequency of booster vaccinations to maintain protective antibody levels within a population. This is especially important for vaccine-preventable diseases with varying immunity levels, such as influenza and COVID-19.

1.3 Correlates of protection

A correlate of protection (COP), defined here, is an immune function or biomarker that predicts protection against infection by a pathogen of interest. Studies such as vaccine trials often estimate thresholds of protection from disease (i.e. >50% probability) for immune biomarkers, and these thresholds exist for influenza, Hepatitis A and B, Measles, Polio, Rabies, Yellow fever and more.[21] Vaccine trials which estimate thresholds of protection do not estimate the titre value at the time of exposure and, indeed, cannot estimate which individuals were even exposed to the pathogen throughout the trial. Therefore, though estimating thresholds of protection from disease is pivotal in designing and optimising vaccines to induce the required immune response effectively, they cannot be used to estimate the COP as defined. [22]

Determining the immunological profiles of individuals exposed to an infection but managed to abort it is crucial for determining a universal correlate of protection. Usually, these individuals are identified through challenge studies, where their immune profiles are measured and challenged with live viruses.[23] As all individuals in the study have been exposed to the virus, it is then possible to determine which biomarkers correlate with protection from infection and disease. However, it's unclear how applicable these estimates are for real-world exposures, as we don't know the typical infecting dose of a pathogen. The studies are also performed on fit and healthy individuals, which are not representative of the population vulnerable to severe infection by infectious disease. Finally, these studies focus on finding threshold values corresponding to 50% protection from infection and diseases and it is unclear what the protection dynamics are for individuals

outside this range. Further, these studies are also expensive, difficult to run, and are only possible for pathogens with mild pathogenesis.[24]

Using serological samples to establish a correlate of protection would solve the aforementioned problems with challenge studies as they are cheaper to conduct, more inclusive, and can be used for any circulating virus in a population. A natural biomarker for a correlate of protection from infection is the amount of neutralising antibodies in the serum, which measures a serum's ability to prevent viral particles from infecting vulnerable cells. Therefore, those with high levels of neutralising antibodies could abort an infection by neutralising viral particles in-host even if exposed to a virus. However, determining correlates of protection through serological studies and measuring neutralising capacity is challenging because those exposed but experiencing an abortive infection generally leave no measurable antibody imprint and cannot be identified through serological testing.[25] Therefore, serological studies can be augmented with either immunological profiling to determine other immunological biomarkers that indicate 'abortive' infection[26] or include intensive contact tracing with the serological studies to determine exposure rates between individuals.[27] Augmented serological studies increase their complexity and cost and, therefore, are not feasible in many settings. Using existing study designs, there is a need for new statistical and mathematical methods for inferring latent epidemiological parameters, such as past infections and correlates of protection.

1.4 Overview of modelling framework

Previous studies have used antibody kinetics to infer infections within serological datasets through "time-since-infection" modelling or Reversible-Jump MCMC methods.[28, 29, 30] These models infer functional forms of the antibody kinetics to estimate infection times and/or force of infection and then establish functional relationships between their estimated antibody titre at the time of infection and protection against disease. However, as these methods do not estimate exposure rates (i.e. those who are exposed but infections are aborted), they do not estimate a correlate of protection against infection.

In this document, we present a single modelling framework which takes individual-level serological sample data only and uses changes in antibody titres over time to determine i) which individuals have become infected throughout the study, ii) subsequent antibody kinetics of infected individuals, iii) the proportion of the population exposed throughout the study and iv) the correlate of protection preventing exposed individuals from becoming infected. Though the infection and exposure status are often unknown for most individuals in a serological study, we find that by using broad, biologically informed mechanistic forms for the antibody kinetics and correlation of protection, the latent infection and exposure status of individuals and the population are recoverable through the interdependencies of mechanisms i)–iv) within a Reversible-Jump MCMC framework.

2 Simulated serological data from serosim[31]

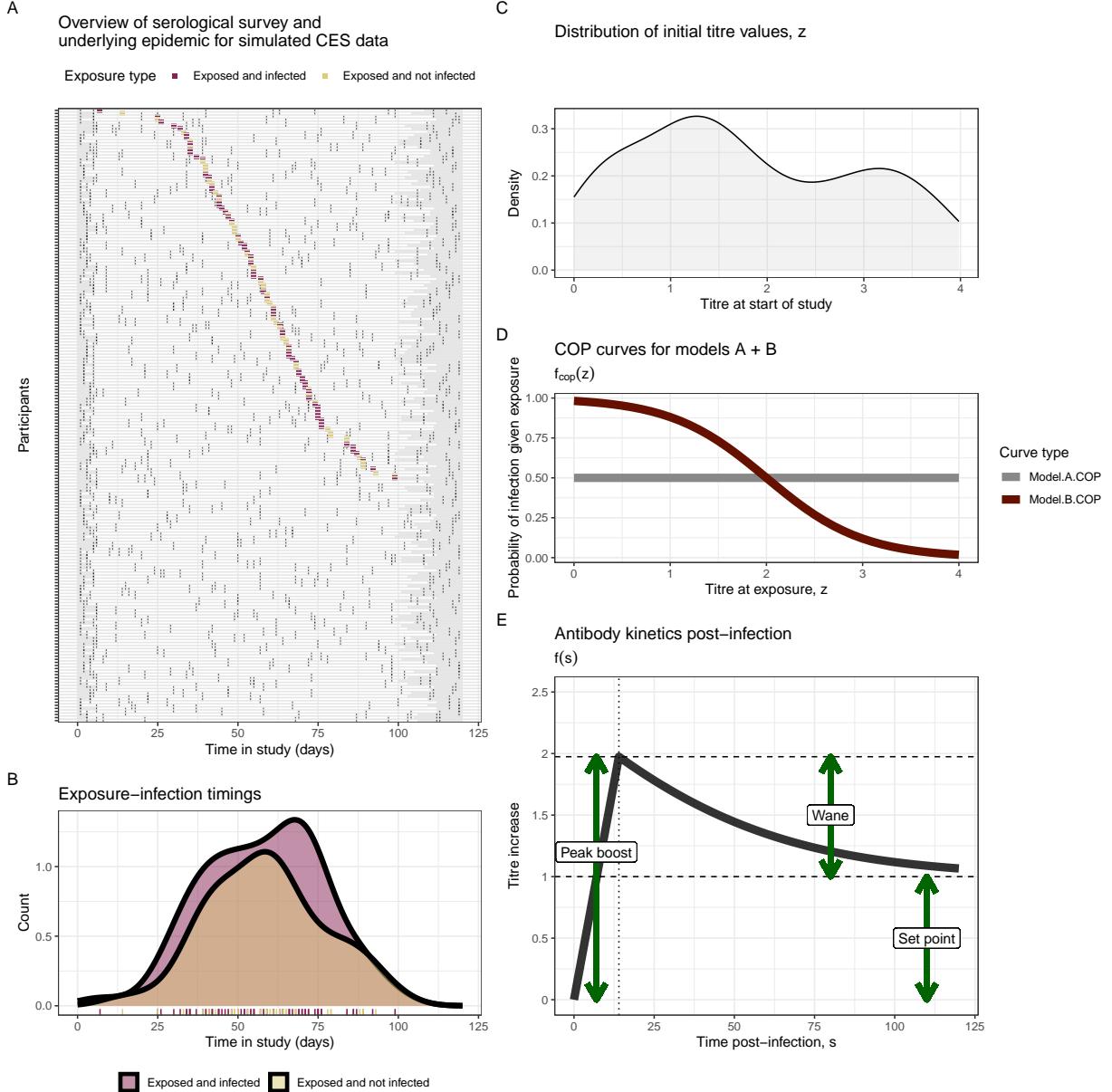


Figure 2.1: Schematics showing the simulated data structure from serosim[31]

We simulate a serological dataset using the `serosimR` package[31] to test the framework we later develop. We simulate continuous epidemic serosurveillance (CES) cohort data, which represents a serostudy in which individuals are followed over an epidemic wave and bled at multiple random time points throughout. The simulated data includes $M = 200$ individuals with serological samples taken within the first seven days of the study's starting and a sample within the last seven days of the study's ending. These individuals also had three samples taken randomly throughout the study (over the 120-day epidemic wave). Each individual has a 60% chance of exposure to the virus over the study timeframe and can have a maximum of one exposure. To model an even epidemic peak, we simulate the exposure time for each individual from a normal distribution, $\mathcal{N}(60, 20)$ days. A figure showing the timing of the bleeds, infections and exposures

and the starting titre values for each individual is given in **Figure 2.1A-C**.

Established correlates of protection have been defined as threshold values of immunological assays which convey protection against disease or infection with a probability of $\geq 50\%$. However, interpreting such thresholds is problematic as there is usually a high amount of variability between individuals, and it's unclear if this protection threshold can be considered completely protective. To account for this individual-level variability, we define a correlate of protection (COP) as the probability of infection given a titre value at exposure and assume that this curve is smooth. Data are simulated with two different correlates of protection; one is uniform at 50% for all titres at exposure (COP model A) and thus represents no titre-dependent COP. The second follows a logistic distribution (COP model B) of the form:

$$f_{cop}(x, \beta_0, \beta_1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))} \quad (2.1)$$

where $\beta_0 = -4$ and $\beta_1 = 2$ in the simulated data and x is the titre value at exposure. This is a common functional form for the correlate of protection and represents a pathogen for which higher antibody titres are associated with higher levels of protection from infection, such as for Influenza or SARS-CoV-2.[32] Note in these data, we assume antibody trajectories remain constant until the timing of infection, such $f_{cop}(x, \beta_0, \beta_1) = f_{cop}(Y_j^0, \beta_0, \beta_1)$ where Y_j^0 is the antibody titre of an individual, j , at their first bleed at the start of the study. A figure showing the two functional forms for the COP models is given in **Figure 2.1D**.

Following infection, the antibody kinetics are assumed to follow a linear rise to a peak at 14 days, followed by an exponential decay to a set-point value.[16] The formula for this biphasic trajectory is given by **Equation 2.2**.

$$f_{ab}(s, a, b, c) = \begin{cases} \ln(\exp(a) + \exp(b))/14, & \text{if } s \leq 14 \\ \ln(\exp(a) \exp(-(b/10)(t - 14)) + \exp(c)), & \text{if } s > 14 \end{cases} \quad (2.2)$$

where $a = 1.5$, $b = 2$, and $c = 1$ are values in the simulated data (**Figure 2.1E**) and s is the number of days post infection. If an individual is not infected over the period their titre remains unchanged and therefore equal to Y_j^0 throughout.

Models with different levels on individual variability in antibody kinetics

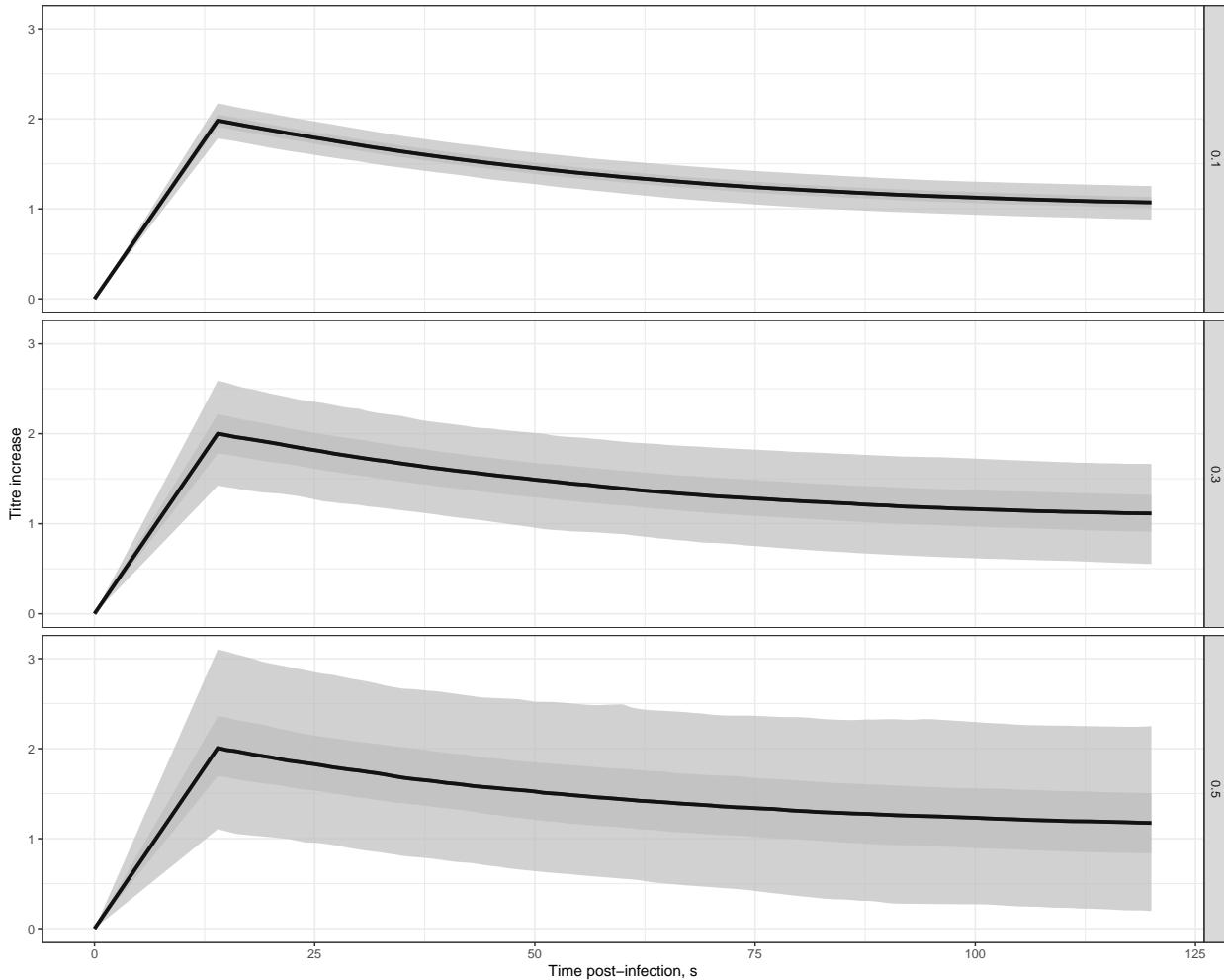


Figure 2.2: Schematics showing three levels of individual-level variation and the impact on the variability of antibody kinetics. The rows provide a summary of the trajectories when the antibody kinetics parameters are simulated with 10%, 30% and 50% individual-level variability.

To model heterogeneity in individual-level antibody kinetics, we simulate a, b, c from normal distributions where the mean μ is their simulated values and the standard deviation given by $\mu\sigma^*$. We simulate three levels of variation; $\sigma^* = \{0.1, 0.3, 0.5\}$ for both of the COP models, giving six simulated datasets. A schematic showing how these levels of uncertainty influence the variability of the antibody kinetics trajectories is shown in **Figure 2.2**. We will fit the to-be-described methods to these datasets and explore how the correlate of protection and level of variability in antibody kinetics impacts the framework's ability to recover simulated data.

Because simulations are stochastic, the relationship between titre in the simulated dataset does not exactly match the functional relationship. We plot the simulated COP curve and compare it to the functional COP curve in **Figure 2.3**. A summary of the parameter values and functions used to generate the simulated data can be found in **Table 2.1**.

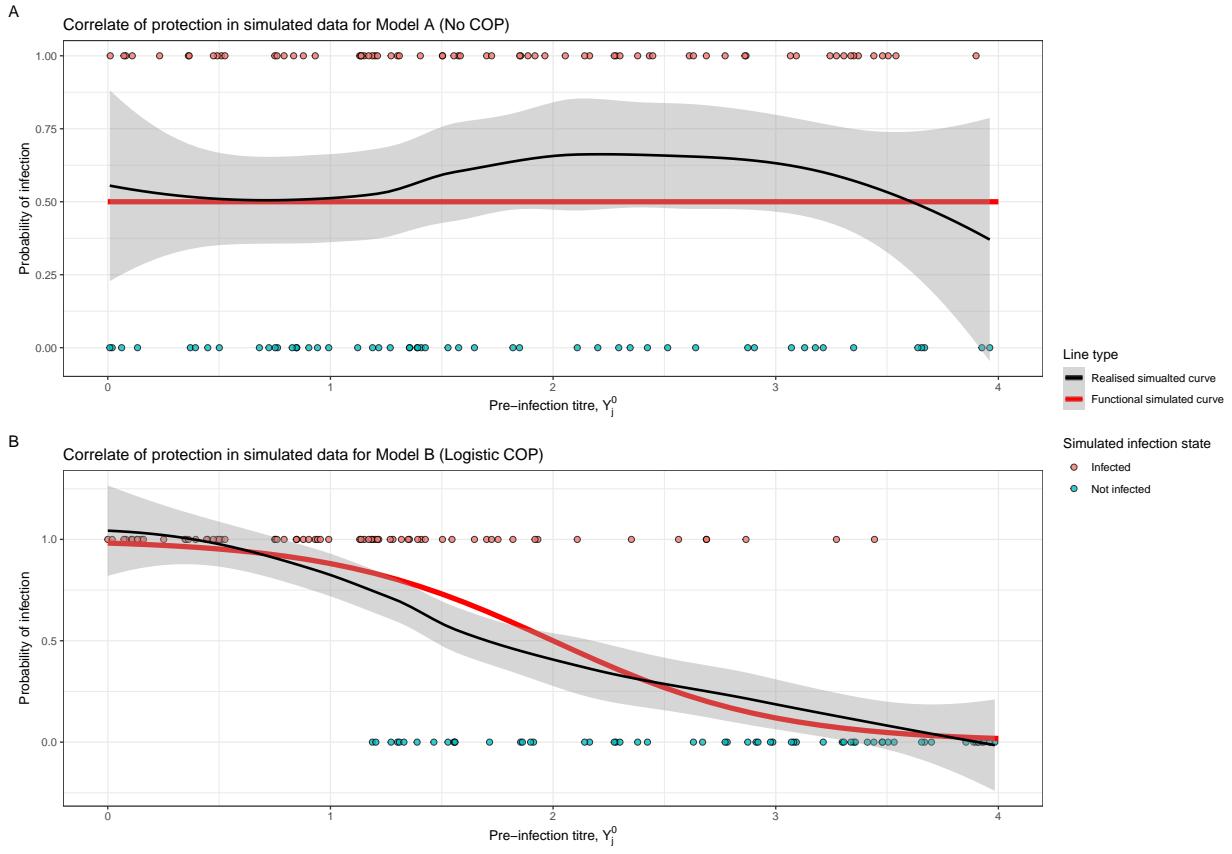


Figure 2.3: Schematics showing the difference between the functional form chosen to simulate the COP (red line) and the recovered COP from the exposed individuals (black line). The black line is a LOESS spline with 95% uncertainty defined by the gray area.

Symbol	Description	Value
M	Number of individuals in the sample, we use subscript j to refer to an individual.	200
T	Time over which the study is run.	120 days
$P(E_j)$	Probability of exposure for individual j over time period.	60%
E_j^T	Timing of exposure for exposed individuals.	$\mathcal{N}(60, 20)$
a	Parameter describing the peak boost of the post-infection antibody kinetics	1.5
b	Parameter describing the waning rate of the post-infection antibody kinetics	2
c	Parameter describing the set-point value of the post-infection antibody kinetics	1
β_0	Intercept parameter of the logistic function for the COP	-4
β_1	Slope parameter of the logistic function for the COP	2
f_{ab}	Antibody kinetics function	Equation 2.2
f_{cop}	Logistic correlate of protection	Equation 2.1

Table 2.1: Table of parameters associated with the simulated data and their values.

3 Inference with known exposure status

Using our simulated noisy serological data, this section will show how to infer several latent processes, including the population-level antibody kinetics, the correlate of protection, and the individual-level infectious state. We show simulation recovery in a simplified framework where we assume the exposure status of every individual (represented by vector \mathbf{E}) and exposure time (represented by vector \mathbf{E}^τ) is known. Though knowing this information is rarely feasible in practice, working through this example will help explain how the inference on the fitted parameters, θ (describing the COP and antibody kinetics), and infection state, \mathbf{Z} , work without needing to describe the more complex inference using RJ-MCMC.

3.1 Mathematical representation of framework

Let the binary vector $\mathbf{E} = \{E_1, E_2, \dots, E_M\}$, represent the exposure status of each individual j where $E_j = 0$ is not exposed and $E_j = 1$ is exposed and let $n_{\mathbf{E}} = \sum_{j=1}^M E_j$ be the number of exposed individuals. Then, let the vectors $\mathbf{E}^\tau = \{E_1^\tau, E_2^\tau, \dots, E_{n_{\mathbf{E}}}^\tau\}$ and $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_{n_{\mathbf{E}}}\}$ be the timing of the exposure and the infection state respectively for each individual which is exposed. The infection state is a binary vector where $Z_j = 0$ is not infected, and $Z_j = 1$ is infected. Let $Y_{j,t} \in \mathbf{Y}$ represent the dataset of titre values for individual j and at time t .

We define several functions to help us calculate the likelihood of our model. First, we define a biologically-informed function that models the latent antibody titre at time t for individual j , $A_{j,t}$, given their infection status Z_j and timing of exposure E_j^τ . If a person is not infected, their starting titre value (Y_j^0) remains unchanged. If the person is infected, their titre remains unchanged until the point of infection, after which they follow the kinetic function in **Equation 2.2**. Although our simulated data has individual-level trajectories that vary from a population mean according to our variability level (10%, 30%, 50%), here we estimate the population-level antibody kinetics function, averaging over the individual-level antibody kinetics. The deterministic function for calculating the latent antibody titre at time t for individual j , $A_{j,t}$ depends on the timing of exposure for the individual, so $A_{j,t}$ is given by

$$A_{j,t} = g_{ab}(Z_j, E_j^\tau, \theta_{ab}, Y_j^0) = \begin{cases} Y_j^0 + f_{ab}(t - E_j^\tau, \theta_{ab}, Y_j^0), & \text{If } Z_j = 1, E_j = 1, \text{ and } t > E_j^\tau \\ Y_j^0, & \text{Otherwise} \end{cases} \quad (3.1)$$

Where $\theta_{ab} = \{a, b, c\}$. Second, we define a likelihood function for the correlate of protection. For an individual j , with $E_j = 1$, the correlate of protection given exposure at time t with titre value $A_{j,t}$, is assumed to follow a Bernoulli distribution with the probability is given by **Equation 2.1**. The PDF of this likelihood is given by **Equation 3.2**.

$$P_{cop}(Z_j | Y_j^0, \theta_{cop}) = f_{cop}(Y_j^0, \theta_{cop})^{Z_j} (1 - f_{cop}(Y_j^0, \theta_{cop}))^{1-Z_j} \quad (3.2)$$

where $\theta_{cop} = \{\beta_0, \beta_1\}$. Finally, we define an observational model to capture variability between hosts and measurement error. Given $A_{j,t}$ and the serological antibody data at the same time point is given by $Y_{j,t}$, we assume the measurement error follows a normal distribution with a PDF given by **Equation 3.3**.

$$P_{obs}(Y_{j,t} | A_{j,t}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(Y_{j,t} - A_{j,t})^2}{2\sigma^2}\right)} \quad (3.3)$$

Let $\theta = \{a, b, c, \beta_0, \beta_1, \sigma\}$ be the set of continuous parameters which are to be fitted in the model.

3.2 Posterior distribution via Bayes rule

We have two different likelihoods depending on whether an individual is exposed ($E_j = 1$) or not ($E_j = 0$).

3.2.1 Likelihood for an non-exposed individual $E_j = 0$

In this case, the value of the timing of exposure and infection status is not applicable and thus not inferred. The likelihood for individual j with serological samples taken at times $t \in T_j$ is therefore equivalent to:

$$L_{E_j=0}(Y_j|\theta) = \prod_{t \in T_j} P_{obs}(Y_{j,t}|Y_j^0, \sigma) \quad (3.4)$$

as $A_{j,t} = Y_j^0$ for all t .

3.2.2 Likelihood for an exposed individual $E_j = 1$

In this case, the infection status is determined by the correlate of the protection likelihood (P_{cop}) and the antibody kinetics function. The likelihood for this individual with serological samples taken at times $t \in T_j$ and infection time E_j^τ is therefore equivalent to:

$$L_{E_j=1}(Y_j|Z_j, \theta) = \prod_{t \in T_j} P_{obs}(Y_{j,t}|A_{j,t}, \sigma) P_{cop}(Z_j | Y_j^0, \theta_{cop}) \quad (3.5)$$

where $A_{j,t} = g_{ab}(Z_j, E_j^\tau, \theta_{ab}, Y_j^0)$.

3.2.3 Total likelihood

If \mathbf{E}_0 and \mathbf{E}_1 are vectors representing the set of individuals who are not exposed and exposed, respectively. Then, the total likelihood can be written

$$L(\mathbf{Y}|\mathbf{Z}, \theta) = \prod_{j \in \mathbf{E}_0} L_{E_j=0}(Y_j|\theta) \prod_{j \in \mathbf{E}_1} L_{E_j=1}(Y_j|Z_j, \theta) \quad (3.6)$$

3.2.4 Prior distributions

We choose prior distributions for each parameter $\pi(\theta)$. **Table 3.1** summarises the chosen priors with their support.

Parameter	Prior (π)	Support (S)
a	$\mathcal{N}(1.5, 0.5)$	$[0.5, 4]$
b	$\mathcal{N}(0.3, 0.05)$	$[0, 1]$
c	$\mathcal{U}(0, 4)$	$[0, 4]$
β_0	$\mathcal{U}(-10, 10)$	$[-10, 10]$
β_1	$\mathcal{U}(-10, 10)$	$[-10, 10]$
σ	$\mathcal{U}(0.01, 1)$	$[0.01, 1]$

Table 3.1: Table with Headers: Parameter, Prior, and Support

We also choose the prior for the number of infections $n_{\mathbf{Z}}$ given the number of exposed individuals $n_{\mathbf{E}}$ to be a Beta Binomial distribution: $\pi(\mathbf{Z}) = \text{BetaBinomial}(n_{\mathbf{Z}}|n_{\mathbf{E}}, 1, 1)$. Choosing this prior prevents any implicit priors that might rise from products of Bernoulli trials[33] as $\text{BetaBinomial}(n_{\mathbf{Z}}|n_{\mathbf{E}}, 1, 1) = 1/n_{\mathbf{E}}$ for all $0 \leq n_{\mathbf{Z}} \leq n_{\mathbf{E}}$.

3.2.5 Posterior distributions

Bayes' rule stipulates that the product of the prior distribution and likelihood is proportional to the posterior distribution; we can use this rule to approximate the posterior for use in the metropolis algorithm. Specifically

$$P(\theta, \mathbf{Z}|\mathbf{Y}) \propto \mathcal{L}(\mathbf{Y}|\mathbf{Z}, \theta) \pi(\theta) \pi(\mathbf{Z}) \quad (3.7)$$

3.3 Metropolis-Hastings for serological inference with known exposure

An overview of the Metropolis-Hastings algorithm is in [Appendix Section A.1](#). In this section, we define the proposal distribution required to sample from [Equation 3.7](#), which infers θ , and infection statuses $Z_j \in \mathbf{Z}$, for $1 \leq j \leq M$ individuals. A schematic showing the relationship between the state variables and likelihood functions is given in [Figure 3.1](#).

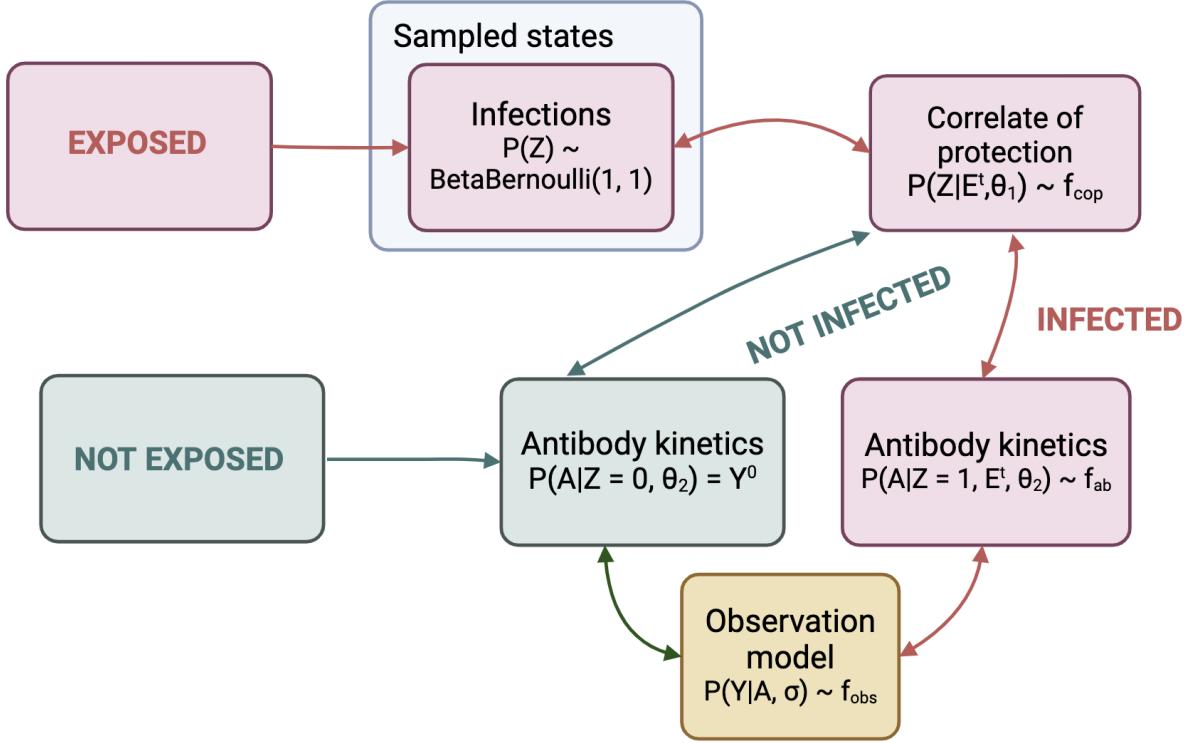


Figure 3.1: Schematic showing the relationship between the sampled states for Z and the likelihood functions.

We define independent proposal distribution for θ and \mathbf{Z} , such that $Q(\cdot|\theta, \mathbf{Z}) = q_\theta(\cdot|\theta)q_Z(\cdot|\mathbf{Z})$. At Markov chain step i , we have a value of the parameter space, $\theta^{(i)}$, and propose a new set of parameters θ' via the proposal distribution $\theta' \sim q_\theta(\cdot|\theta^{(i)}, \psi_{adapt}^{(i)})$. This proposal is a multivariate normal distribution with an adaptive covariance matrix, which is defined by the set of parameters $\psi_{adapt}^{(i)}$, which are updated at each time step.[34, 35] (see [Appendix C](#)). For \mathbf{Z} , we propose a new infection state \mathbf{Z}' by selecting an exposed individual j , which has infection status $Z_j^{(i)}$ at step i of the current Markov chain, and sample a proposal value for their infection status Z'_j by the proposal distribution for $Z'_j \sim q_Z(\cdot|Z_j^{(i)}) = \text{Bernoulli}(0.5)$. Therefore the proposal for $q_Z(\mathbf{Z}'|\mathbf{Z}) = 1/n_E 0.5$ for all j . Both of these proposals $q_\theta(\theta'|\theta^{(i)}, \psi_{adapt}^{(i)})$, $q_Z(\mathbf{Z}'|\mathbf{Z})$ are symmetric and thus cancel out the acceptance ratio ([Equation A.3](#)). Further, the prior distribution $\pi(\mathbf{Z}) = 1/n_E$ for all $0 \leq n_{\mathbf{Z}} \leq n_E$, and thus also cancels out in the acceptance ratio, therefore we need only calculate: $P(\theta, \mathbf{Z}|Y) \propto \mathcal{L}(Y|\mathbf{Z}, \theta)\pi(\theta)$. [Algorithm 3.1](#) is a Metropolis-Hastings algorithm which samples from this proposal.

Algorithm 3.1 Metropolis-Hastings Algorithm for antibody kinetics and infection inference

- 1: Initialize the chain with an initial state $\theta^{(0)}$ from the priors $\pi(\cdot)$ and $I_j^{(0)} \sim \text{Bernoulli}(0.5)$ for all $1 \leq j \leq M$ individuals to initialise $\mathbf{Z}^{(0)}$, and initialise $\psi_{adapt}^{(0)}$.
- 2: **for** $i = 1$ to N **do**
- 3: Generate a candidate state $\theta' \sim q_\theta(\theta^{(i)}, \psi_{adapt}^{(0)})$
- 4: Generate a candidate individual $j \in \mathbf{E}_1$, then a candidate state $Z_j \sim \text{Bernoulli}(0.5)$ to propose \mathbf{Z}'
- 5: Compute the acceptance ratio:

$$\alpha((\theta^{(i)}, \mathbf{Z}^{(i)}), (\theta', \mathbf{Z}')) = \min \left(1, \frac{P(\theta', \mathbf{Z}' | \mathbf{Y})}{P(\theta^{(i)}, \mathbf{Z}^{(i)} | \mathbf{Y})} \right)$$

- 6: Sample $u \sim \mathcal{U}(0, 1)$
- 7: **if** $u \leq \alpha$ **then**
- 8: Accept the candidate state: $\theta^{(i+1)} \leftarrow \theta'$ and $\mathbf{Z}^{(i+1)} \leftarrow \mathbf{Z}'$
- 9: **else**
- 10: Reject the candidate state: $\theta^{(i+1)} \leftarrow \theta^{(i)}$ and $\mathbf{Z}^{(i+1)} \leftarrow \mathbf{Z}^{(i)}$
- 11: **end if**
- 12: Update $\psi_{adapt}^{(i+1)} \leftarrow \psi_{adapt}^{(i)}$
- 13: **end for**

3.4 Implementation

Algorithm 3.1 is coded manually in R and C++ through Rcpp. We run the algorithm for four chains, each with 400,000 steps and 200,000 burn-in steps. The initial values for θ and \mathbf{Z} are sampled from their prior distributions. We initialise the adaptive covariance by running with an identity matrix with each parameter scale according to 1,000 steps, then sample from the adaptive scheme (see **Appendix C**). We thin the posterior samples by taking one in every 100, leaving 2,000 posterior samples at the end of the sampling. The trace plots for each of the six models fits are given in **Appendix Figure D.1**.

3.5 Simulation recovery

After running **Algorithm 3.1**, we plot the posterior samples, $\hat{\theta}$ and $\hat{\mathbf{Z}}$ and compare with the simulated parameters.

3.5.1 Infection recovery

We assess the algorithm's ability to recover each individual's infection status in the study. Suppose the set posterior samples of the infection status for individual j is given by \hat{Z}_j . In that case, we plot the expectation $\mathbb{E}[\hat{Z}_j]$ so we can assess the ability of the algorithm to recover the individual-level simulated infection status (**Figure 4.4a**). All models except one recovered the infection status for all individuals and correctly estimated the number of individuals infected across the study. The COP with 50% variability has one individual out of 200 misclassified, predicting this individual was not infected when it indeed was.

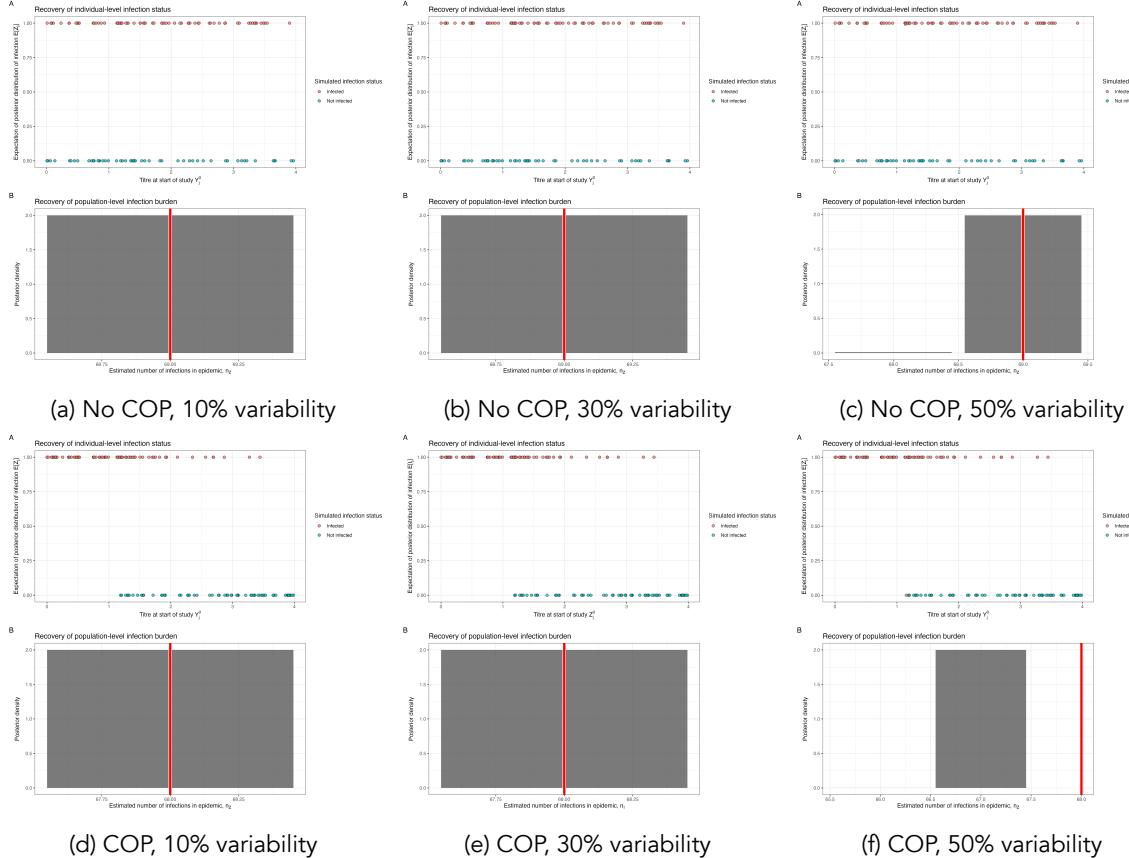


Figure 3.2: Simulation recovery of the individual infection status, $\mathbb{E}[\hat{Z}_j]$, for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%)

3.5.2 Correlate of protection

We next assess the ability of **Algorithm 3.1** to recover the correlate of protection function $f_{cop}(x, \hat{\theta}_{cop})$, where x is the titre value at infection and where $\hat{\theta}_{cop} = \{\hat{\beta}_0, \hat{\beta}_1\}$ are the posterior samples for β_0 and β_1 . We consider two COP models: COP model A, no correlate of protection, and COP model B, a logistic curve for COP. For Model A, we find that the COP curve is recovered, with the simulated line within a 95% confidence interval of the posterior sample (**Figure 3.3**). For Model B, we also find the logistic shape of the COP is recovered in the posterior samples. The variability in the antibody kinetics seemed to have a negligible effect on the recoverability of the COP curve.

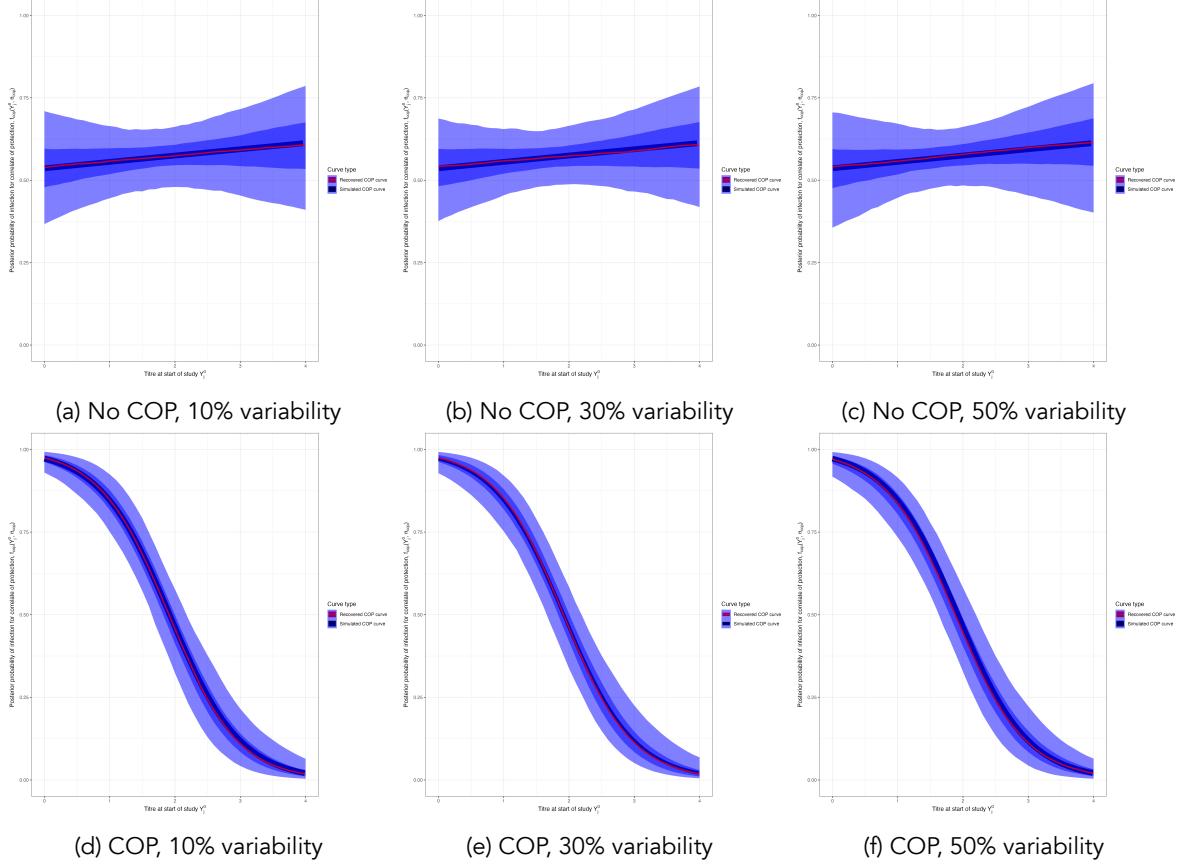


Figure 3.3: Simulation recovery of the COP function, with posterior samples plot $f_{cop}(x, \hat{\theta}_{cop})$. We have two different COP models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

3.5.3 Antibody kinetics

Algorithm 3.1 also recovers the simulated antibody kinetics. Let us plot $f_{ab}(s, \hat{a}, \hat{b}, \hat{c})$, the posterior predictive distribution for the antibody kinetic boosting, given posterior distributions \hat{a} , \hat{b} , and \hat{c} . For the no COP model A, the antibody kinetics are recovered, though increasing variability weakens the accuracy of the recovered curves compared to the simulated when we compare the posterior distributions for \hat{a} , \hat{b} , \hat{c} to the simulated value (**Figure 3.4**). For COP model B, there is a slight bias in the estimation for parameter c , being overestimated for variability levels 30% and 50%.

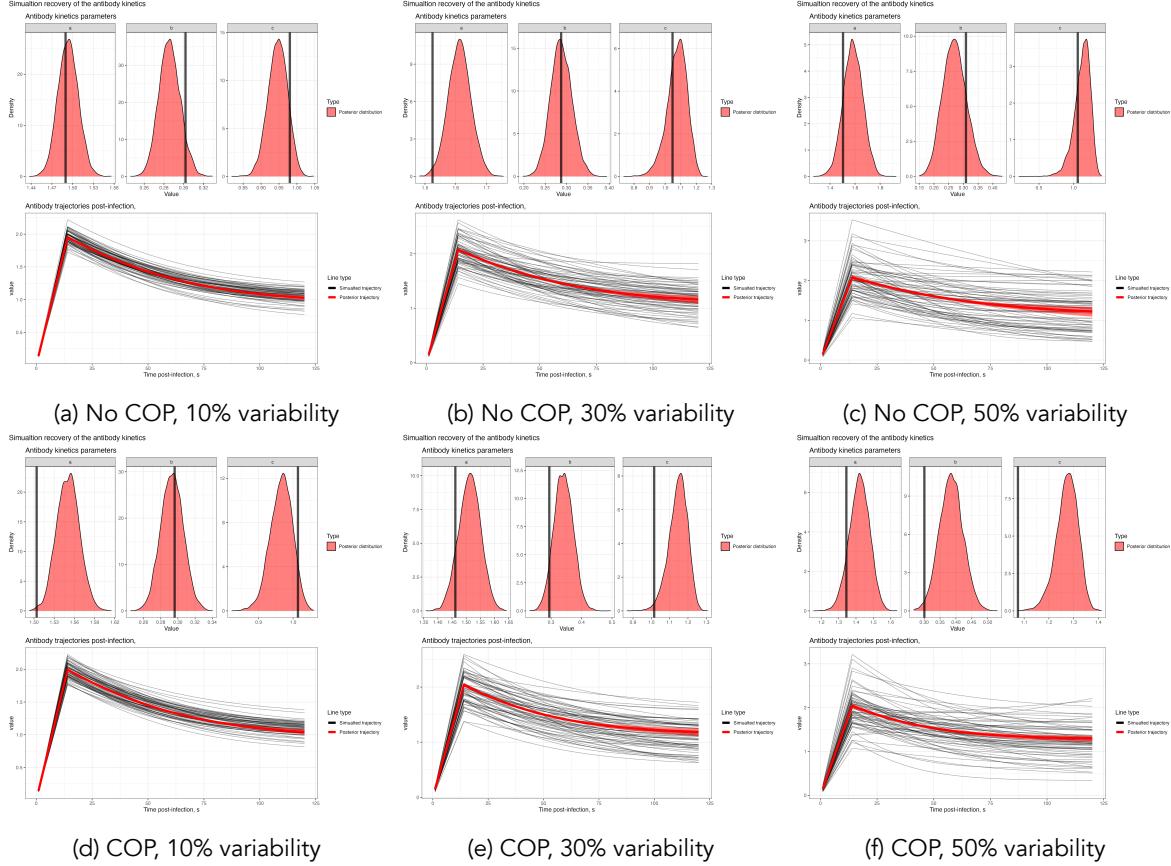


Figure 3.4: Simulation recovery of the antibody kinetics function with posterior samples plot $f_{ab}(s, \hat{a}, \hat{b}, \hat{c})$. We have two different COP models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

4 Inference with unknown exposure status

For the known exposure status of an individual, we have shown that **Algorithm 3.1** can recover the individual-level infection status, a population-level correlate of protection and the underlying antibody kinetics for two different correlate of protection assumptions and three different levels of individual-level kinetics variability. In practice, this algorithm is unlikely to be useful as the individual-level exposure state is unknown. This section will expand on this algorithm for the case when exposure status is unknown in a serological study.

4.1 Overview

In the case where the exposure status of each individual, j , is unknown, we must now infer their exposure state $E_j \in \{0, 1\}$ and the time of exposure given they are exposed $0 \leq E_j^\tau \leq 120$. In the case where $E_j = 0$, the likelihood is as derived in **Equation 3.4**. However, in the case where $E_j = 1$, the likelihood contains an additional dependency on E_j^τ :

$$L_{E_j=1}(Y_j|Z_j, E_j^\tau, \theta) = \prod_{t \in T_j} P_{obs}(Y_{j,t}|A_{j,t}, \sigma)P_{cop}(Z_j | E_j^\tau, \theta) \quad (4.1)$$

where $A_{j,t} = g_{ab}(Z_j, E_j^\tau, \theta_{ab}, Y_i^0)$.

4.1.1 Likelihood and priors

Let $\mathbf{E} = \{E_0, E_1, \dots, E_M\}$ be a vector describing the exposure status of each individual, let $\mathbf{E}^\tau = \{E_0^\tau, E_1^\tau, \dots, E_{n_\mathbf{E}}^\tau\}$ be a vector describing the exposure times for each exposed individual. The likelihood of this system is similar to before:

$$\mathcal{L}(\mathbf{Y}|\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}) = \prod_{j \in \mathbf{E}_0} L_{E_j=0}(Y_j|\theta) \prod_{j \in \mathbf{E}_1} L_{E_j=1}(Y_j|Z_j, E_j^\tau, \theta) \quad (4.2)$$

We now must additionally define priors for $\pi(\mathbf{E})$ and $\pi(\mathbf{E}^\tau)$. Similar to $\pi(\mathbf{Z})$ we define the prior distribution for $\pi(\mathbf{E})$ to be a Beta Binomial distribution: $\pi(\mathbf{E}) = \text{BetaBinomial}(n_\mathbf{E}|M, 1, 1)$, which is equal to $1/M$ for all $0 \leq n_\mathbf{E} \leq M$. For $\pi(\mathbf{E}^\tau)$, we assume that each element $E_j^\tau \in \mathbf{E}^\tau$ has a prior given by P_t such that $\pi(\mathbf{E}^\tau) = \prod_{j=1}^{n_\mathbf{E}} P_t(E_j^\tau)$. The priors for $\pi(\theta)$ and $\pi(\mathbf{Z})$ are as described in **Section 3**.

Consequently, we sample from the posterior distribution

$$P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}|\mathbf{Y}) \propto \mathcal{L}(\mathbf{Y}|\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z})\pi(\theta)\pi(\mathbf{E})\pi(\mathbf{E}^\tau)\pi(\mathbf{Z}) \quad (4.3)$$

If we use a **Algorithm 3.1** or any Metropolis Hasting algorithm to infer the exposure status and exposure time, we run into a problem. The number of parameters in the posterior distribution changes according to whether an individual is exposed. Those exposed have parameters Z_j and E_j^τ to infer, whereas an individual without exposure has neither. Therefore, regardless of the proposal distribution we choose for inferring \mathbf{E} , we cannot use the existing algorithm highlighted in **Algorithm 3.1** as the detailed balance condition now fails.

4.2 The Reversible-Jump MCMC

The reversible jump Markov chain Monte Carlo (RJMCMC) algorithm[36] is a Bayesian statistical method designed for model selection in situations where the number of model parameters can vary. It achieves this by introducing a stochastic mechanism that proposes moves between different models, including adding or removing parameters. The idea is to use a Metropolis-Hastings step to evaluate the acceptance probability of these proposed model changes, ensuring that the Markov chain explores the posterior distribution over

both model parameters and model structures. We give a derivation of RJ-MCMC and motivate its use in **Appendix A.2**.

4.3 Application of RJ-MCMC to serological data

Appendix Algorithm A.2 is a general framework for jumping from a model k with a parameter values $\theta_k \in \Theta_k$ and another model k' with a parameter values $\theta_{k'} \in \Theta_{k'}$. For our serological inference, our model k , represents different elements of the exposure state vector $\mathbf{E} = \{E_0, E_1, \dots, E_{n_E}\}$ be the number of exposed individuals and $n_{\mathbf{E}_0} = M - n_{\mathbf{E}}$ be the number of non-exposed individuals in model k . For a given exposure vector \mathbf{E}_k , we define three different possible ways to sample a new exposure state vector, \mathbf{E}' in our RJ-MCMC algorithm: a birth event (adding a new exposure), death event (remove an existing exposure), and a move event (exposure state resampled)[37, 38].

4.3.1 Birth event

For a given exposure vector \mathbf{E} , a birth event generates a new exposure vector \mathbf{E}' , by randomly selecting a non-exposed individual and changing their exposure status from $E_j = 0 \rightarrow E'_j = 1$ and appending it to \mathbf{E} . We can derive an expression for $q_k(k'|k)$ by separating into the probability that a birth event is selected at model k , $q_{birth}(k'|k)$ and the probability of choosing individual j uniformly from the non-exposed individuals:

$$q_k(k'|k) = q_{birth}(k'|k) \cdot \frac{1}{n_{\mathbf{E}_0}} \quad (4.4)$$

To understand how to evaluate $q_2(\mathbf{u}')$, $q_1(\mathbf{u})$ (see **Appendix A.2**), consider the change in likelihood and the addition prior term for an individual j who is chosen to be exposed:

$$\prod_{t \in T} P_{obs}(Y_{j,t}|Y_j^0, \sigma) \rightarrow \prod_{t \in T} P_{obs}(Y_{j,t}|A_{j,t}, \sigma) P_{cop}(Z_j | Y_j^0, \theta) P_t(E_j^t) \quad (4.5)$$

with the likelihood function staying the same for all other individuals. By changing the exposure state for individual j , the likelihood now depends on two parameters not in the previous likelihood: the timing of the exposure E_j^t and their infection status Z_j . In the notation of **Appendix A.2**, it is convenient to define $\mathbf{u} = (E_j^t, Z_j)$. Therefore, we must define a sampling procedure for \mathbf{u} and a probability density function $q_1(\mathbf{u})$ (Note as $d_{k'} > d_k$, we can assume \mathbf{u}' is empty). For E_j^t , we sample from the probability density function for the timing $P_t(\cdot)$, and for the infection status, we sample from $P_{cop}(\cdot | Y_j^0, \theta_{cop})$ (see **Equation 3.2**). This sampling procedure results in a proposed sample which is in the proposed parameter space $\Theta_{k'} : (\theta_k, E_j^t, Z_j) = \theta_{k'} \in \Theta_{k'}$. Consequently, we can choose the identify function for the required bijection T in **Equation A.6**, which means the Jacobian is equal to 1.

With this sampling proposal, we can then evaluate $q_1(\mathbf{u})$ through likelihood functions for each of Z_j and E_j^t :

$$q_1(\mathbf{u}) = q_1(E_j^t, Z_j) = P_{cop}(Z_j | E_j^t, \theta_{cop}) P_t(E_j^t) \quad (4.6)$$

Now let us consider the inverse, which is from the proposed model k' moving back to the model k . For this event, we randomly select one person from the proposed exposure state \mathbf{E}' and change their exposure state from $E_j = 1 \rightarrow E'_j = 0$. Similar to above, we derived an expression for $q_k(k|k')$ by separating into the probability that an inverse birth event is selected at model k , $q_{birth}(k|k')$ and the probability of choosing individual j uniformly from the exposed individuals of \mathbf{E}' :

$$q_k(k|k') = q_{birth}(k|k') \cdot \frac{1}{1 + n_{\mathbf{E}}} \quad (4.7)$$

In this event, we are removing the parameters E_j^τ and Z_j from $\theta_{k'}$. After this, we are left with $\theta_k \in \Theta_k$, so we do not need to sample new variables to generate samples from Θ_k . Thus \mathbf{u}' is empty and $q_2(\mathbf{u}') = 1$. The acceptance ratio for a birth move, where $k = \{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}\} \rightarrow k' = \{\theta, \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}'\}$ is updated according to a uniformly sampled non-exposed individual j :

$$\alpha(k, k') = \min \left(\frac{P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}, |\mathbf{Y}) q_{birth}(k|k') n_{\mathbf{E}_0}}{P(\theta, \mathbf{E}, \mathbf{E}^{\tau'}, \mathbf{Z}, |\mathbf{Y}) P_{cop}(Z'_j | Y_j^0, \theta_{cop}) P_t(E_j^\tau) q_{birth}(k'|k)(n_{\mathbf{E}} + 1)} \right) \quad (4.8)$$

4.3.2 Death event

For a given exposure vector \mathbf{E} , a death event generates a new exposure vector \mathbf{E}' , by randomly selecting an exposed individual and changing their exposure status from $E_j = 1 \rightarrow E'_j = 0$ therefore removing it from \mathbf{E} . We can derive an expression for $q_k(k'|k)$ by separating into the probability that a death event is selected at model k , $q_{death}(k'|k)$ and the probability of choosing individual j uniformly from the exposed individuals:

$$q_k(k'|k) = q_{death}(k'|k) \cdot \frac{1}{n_{\mathbf{E}}} \quad (4.9)$$

The reverse probability $q_k(k'|k)$ is equivalent to a birth event, that is, the probability of sampling a non-exposed person in \mathbf{E}' , or

$$q_k(k|k') = q_{death}(k|k') \cdot \frac{1}{1 + n_{\mathbf{E}_0}} \quad (4.10)$$

To understand how to evaluate $q_1(\mathbf{u}), q_2(\mathbf{u}')$, consider the change in likelihood for an individual j who is chosen to be exposed:

$$\prod_{t \in T} P_{obs}(Y_{j,t} | A_{j,t}, \sigma) P_{cop}(Z_j | Y_j^0, \theta) P_t(E_j^\tau) \rightarrow \prod_{t \in T} P_{obs}(Y_{j,t} | Y_j^0, \sigma) \quad (4.11)$$

with the likelihood function staying the same for all other individuals. By changing the exposure state for individual j , the likelihood now depends on two fewer parameters than were in the previous likelihood for k' : the timing of the exposure E_j^τ and their infection status Z_j . Therefore, by using the same argument as in the ‘birth event’ section, defining $q_1(\mathbf{u}) = 1$ and $q_2(\mathbf{u}') = P_t(E_j^\tau) P_{cop}(Z_j | E_j^\tau, \theta_{cop})$, we can take the identity bijection as sample a value in state $\theta' \in \Theta'$ directly. The acceptance ratio for a death event, where our current state is $k = \{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}\} \rightarrow k' = \{\theta, \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}'\}$ is updated according to a uniformly sampled exposed individual j :

$$\alpha(k, k') = \min \left(\frac{P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}, |\mathbf{Y}) P_{cop}(Z_j | E_j, \theta_{cop}) P_t(E_j^\tau) q_{death}(k|k') n_{\mathbf{E}}}{P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}, |\mathbf{Y}) q_{death}(k'|k)(n_{\mathbf{E}_0} + 1)} \right) \quad (4.12)$$

4.3.3 Move event

If we use only the birth and death event highlighted above, the timing of the exposures, \mathbf{E}^τ is not efficiently explored.[38] Therefore, it is desirable that values of \mathbf{E}^τ can be explored for fixed values of \mathbf{Z} , and \mathbf{E} . Therefore, we define a move event which resamples a proportion of the \mathbf{E}^τ , and for each individual, we sample a new time E_j^τ , from the proposal distribution

$$E_{j'}^\tau \sim q_t(E_j^\tau)$$

where we choose the proposal to be the symmetric $q_t(E_j^\tau) = \mathcal{N}(E_j^{t,(i)}, \sigma_j^{(i)})$. Where $\sigma_j^{(i)}$ is an adaptively updated standard deviation for the proposal for the normal, which updates according to the regime:

$$\log(\sigma_j^{(i+1)}) = \log(\sigma_j^{(i)}) + (1 + i)^{-0.5} * (\alpha - 0.44)$$

where α is the metropolis hasting ratio for $E_{j'}^\tau$ vs $E_j^{\tau,(i)}$.

With these three moves, we define the RJ-MCMC algorithm (**Algorithm 4.1**), which effectively samples values from $\{\theta, \mathbf{E}, \mathbf{E}_\tau, \mathbf{Z}\}$.

Note on q_{birth} , q_{death} , q_{par} : We can select values $q_{birth}(k'|k)$ and $q_{death}(k'|k)$ which simplify the expression in the acceptance ratios in **Equation 4.8** and **Equation 4.12**. If k is such that $n_{\mathbf{E}} = 0$, then we choose $q_{par} = 1/3$ and $q_{birth} = 2/3$. If $n_{\mathbf{E}} = M$, then we choose $q_{par} = 1/3$ and $q_{death} = 2/3$. Otherwise, we $q_{birth}(k'|k) = q_{death}(k'|k) = q_{par} = 1/3$ for all k' . Choosing these values means that the values of q_{birth} , q_{death} , and q_{par} cancel in all acceptance ratios (**Equation 4.8** and **Equation 4.12**) for all values of k' given k .

An algorithm describing the Birth-Death RJ-MCMC algorithm for this data is given in **Algorithm 4.1**.

Algorithm 4.1 Birth-Death Reversible Jump MCMC Algorithm

- 1: Choose a model k and initialize the chain with an initial states $\theta_k^{(0)}, \mathbf{E}^{(0)}, \mathbf{E}^{\tau,(0)}$ and $\mathbf{Z}^{(0)}$. If $0 < n_{\mathbf{E}} < M$, then $p_{birth} = p_{death} = p_{par} = 0.33$; if $n_{\mathbf{E}} = 0$, $p_{birth} = 0.67, p_{par} = 0.33, p_{death} = 0$; if $n_{\mathbf{E}} = M$, $p_{death} = 0.67, p_{par} = 0.33, p_{birth} = 0$.
 - 2: **for** $i = 1$ to N **do**
 - 3: Sample a candidate state $\theta' \sim q_\theta(\theta^{(i)}, \psi_{adapt}^{(i)})$
 - 4: $u_1 \sim \mathcal{U}(0, 1)$
 - 5: **if** $u_1 \leq p_{birth}$ **then**
 - 6: Birth move. Select $j' \in \mathbf{E}_0^{(i)}$, set $E_{j'} = 1$, sample $E_{j'}^\tau \sim P_t(\cdot)$, $Z_{j'} \sim P_{cop}(\cdot | Y_{j'}^0, \theta_{cop})$ and update $\{\theta^{(i)}, \mathbf{E}^{(i)}, \mathbf{E}^{\tau,(i)}, \mathbf{Z}^{(i)}\} \rightarrow \{\theta', \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}'\}$. Then, calculate the acceptance probability
 - 7:
$$\alpha(k^{(i)}, k') = \min \left(\frac{P(\theta', \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}', |\mathbf{Y}) n_{\mathbf{E}_0}}{P(\theta, \mathbf{E}^{(i)}, \mathbf{E}^{\tau,(i)}, \mathbf{Z}^{(i)}, |\mathbf{Y}) P_{cop}(Z_{j'} | E_{j'}, \theta_{cop}) P_t(E_{j'}^\tau)(n_{\mathbf{E}} + 1)} \right)$$
 - 8: **else if** $u_1 \leq (p_{birth} + p_{death})$ **then**
 - 9: Death move. Select $j' \in \mathbf{E}_1^{(i)}$, set $E_{j'} = 0$ and update $\{\theta^{(i)}, \mathbf{E}^{(i)}, \mathbf{E}^{\tau,(i)}, \mathbf{Z}^{(i)}\} \rightarrow \{\theta', \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}'\}$. Then calculate the acceptance probability
 - 10:
$$\alpha(k^{(i)}, k') = \min \left(\frac{P(\theta', \mathbf{E}', \mathbf{E}^{\tau'}, Z | \mathbf{Y}) P_{cop}(Z_{j'} | E_{j'}, \theta_{cop}) P_t(E_{j'}^\tau) n_{\mathbf{E}}}{P(\theta, \mathbf{E}^{(i)}, \mathbf{E}^{\tau,(i)}, \mathbf{Z}^{(i)}, |\mathbf{Y})(n_{\mathbf{E}_0} + 1)} \right)$$
 - 11: **else**
 - 12: Select $j' \in N_{E=1}$ and resample from the proposal $E_{j'}^\tau \sim \mathcal{N}(E_j^{\tau,(i)}, \sigma_j^{(i)})$ and update $\mathbf{E}^{\tau,(i)} \rightarrow \mathbf{E}^{\tau,*}$
 - 13: Compute the acceptance ratio:
 - 14:
$$\alpha(k^{(i)}, k') = \min \left(1, \frac{P(\theta', \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}' | \mathbf{Y})}{P(\theta^{(i)}, \mathbf{E}^{(i)}, \mathbf{E}^{\tau,(i)}, \mathbf{Z}^{(i)} | \mathbf{Y})} \right)$$
 - 15: **if** $u \leq \alpha$ **then**
 - 16: Accept the candidate state: Let $\{\theta^{(i+1)}, \mathbf{E}^{(i+1)}, \mathbf{E}^{\tau,(i+1)}, \mathbf{Z}^{(i+1)}\} \leftarrow \{\theta', \mathbf{E}', \mathbf{E}^{\tau'}, \mathbf{Z}'\}$
 - 17: **else**
 - 18: Reject the candidate state: Let $\{\theta^{(i+1)}, \mathbf{E}^{(i+1)}, \mathbf{E}^{\tau,(i+1)}, \mathbf{Z}^{(i+1)}\} \leftarrow \{\theta^{(i)}, \mathbf{E}^{(i)}, \mathbf{E}^{\tau,(i)}, \mathbf{Z}^{(i)}\}$
 - 19: **end if**
 - 20: **end for**
-

Note on prior distributions As M is fixed, then $\pi(\mathbf{E}) = 1/M$ for all $0 \leq n_{\mathbf{E}} \leq M$ and thus cancels out in the acceptance ratio for the birth, death and parameter update move. $\pi(\mathbf{Z})$ cancels out in the parameter updating acceptance ratio (as described in **Algorithm 3.1**). However, in the birth and death move, as $n_{\mathbf{E}}$ in the current state and $n_{\mathbf{E}'}$ is the proposed state have different values, then $\pi(\mathbf{Z}') = 1/n_{\mathbf{E}_1} \neq 1/n_{\mathbf{E}_1'} = \pi(\mathbf{Z}')$ no longer cancels in the ratio and must be included to ensure the detailed balance condition holds. We choose a non-informative for the timing of infection given exposure prior: $P_t(E_j^\tau) = 1/120$.

4.4 Implementation

Algorithm 4.1 is coded manually in R and C++ through Rcpp. We run the algorithm for four chains, each with 400,000 steps and 200,000 burn-in steps. The initial values for θ , \mathbf{Z} , \mathbf{E} , and \mathbf{E}^τ are sampled from their prior distributions. We initialise the adaptive covariance by running with an identity matrix with each parameter scale according to 1,000 steps, then sample from the adaptive scheme. We thin the posterior samples by taking one in every 100, leaving 2,000 posterior samples at the end of the sampling. The trace plots for each of the six model fits are given in **Appendix Figure D.2**.

Through **Algorithm 4.1** we define an efficient sampling procedure for the state space $k = \{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}\}$ within one single framework. This allows us to infer the exposure state, infection state, exposure and infection timings across the epidemic, the correlates of protection and the antibody kinetics function. We show the ability of this procedure to recover our simulated data displayed in the next section.

4.5 Simulation recovery

After running **Algorithm 4.1**, we plot the posterior samples, $\hat{\theta}$, $\hat{\mathbf{Z}}$, $\hat{\mathbf{E}}$, and $\hat{\mathbf{E}}^\tau$ and compare with the simulated parameters.

4.5.1 Exposure state recovery

Algorithm 4.1 can recover the population-level exposure rates across all six models (**Figure 4.1**). For Model A (no COP), we consistently overestimated the number of individuals exposed (estimating between 135 (95% CrI 120–151) exposures when 122 were simulated). However, the exposure state effectively acts as a nuisance parameter in the recovery of the COP as the probability someone is exposed is irrelevant once you condition on exposure, and we see that the COP function is still consistently recovered (see **Figure 4.5**)

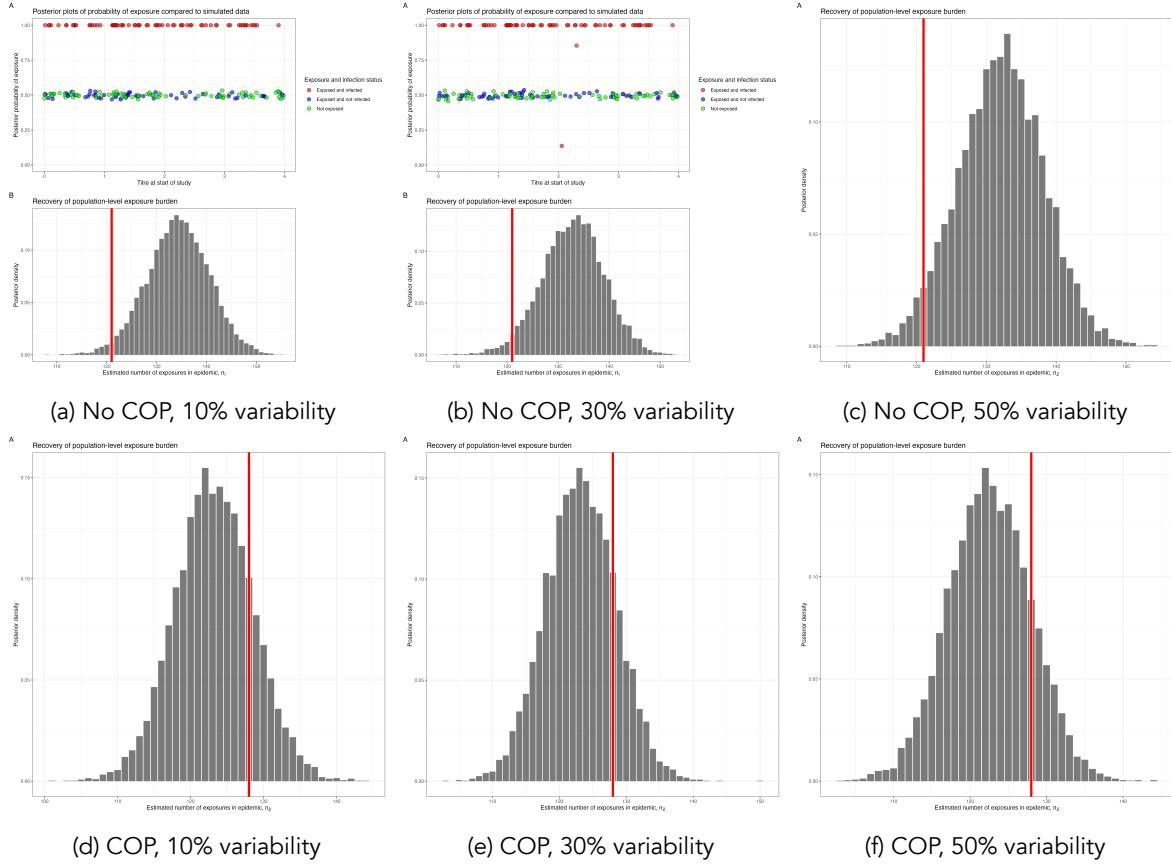


Figure 4.1: Simulation recovery of exposure status \hat{E} and epidemic curve for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%). Red lines in B figures show the true number of exposures from the simulations

4.5.2 Exposure times recovery

Algorithm 4.1 can recover the exposure times. The posterior of the exposure times for individual j is given by \hat{E}_j^τ and plotted with the simulated exposure time in **Figure 4.2** by their exposure status. For those infected, the model can reasonably recover the exposure times for each individual, though as the antibody kinetics variability increases, the exposure time becomes less recoverable. For example, for Model A (no COP), the proportion of individuals with infection times inferred within 14 days of their simulated infection day is 97%, 84%, and 73% for 10%, 30% and 50% variability, respectively (**Figure 4.3**). For all six models, the epidemic curve is reasonably recovered (**Figure 4.2**).

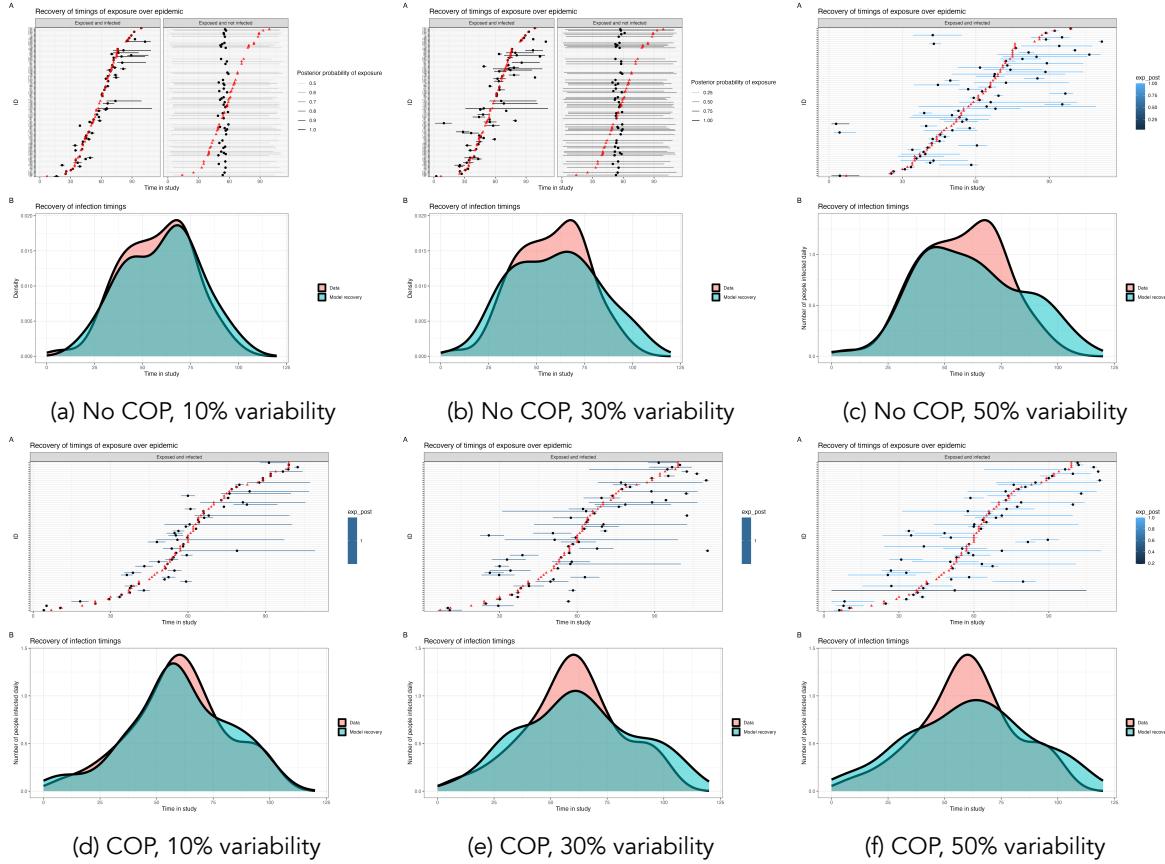


Figure 4.2: Simulation recovery of exposure and infection timings \hat{E}^τ and epidemic curve for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%)

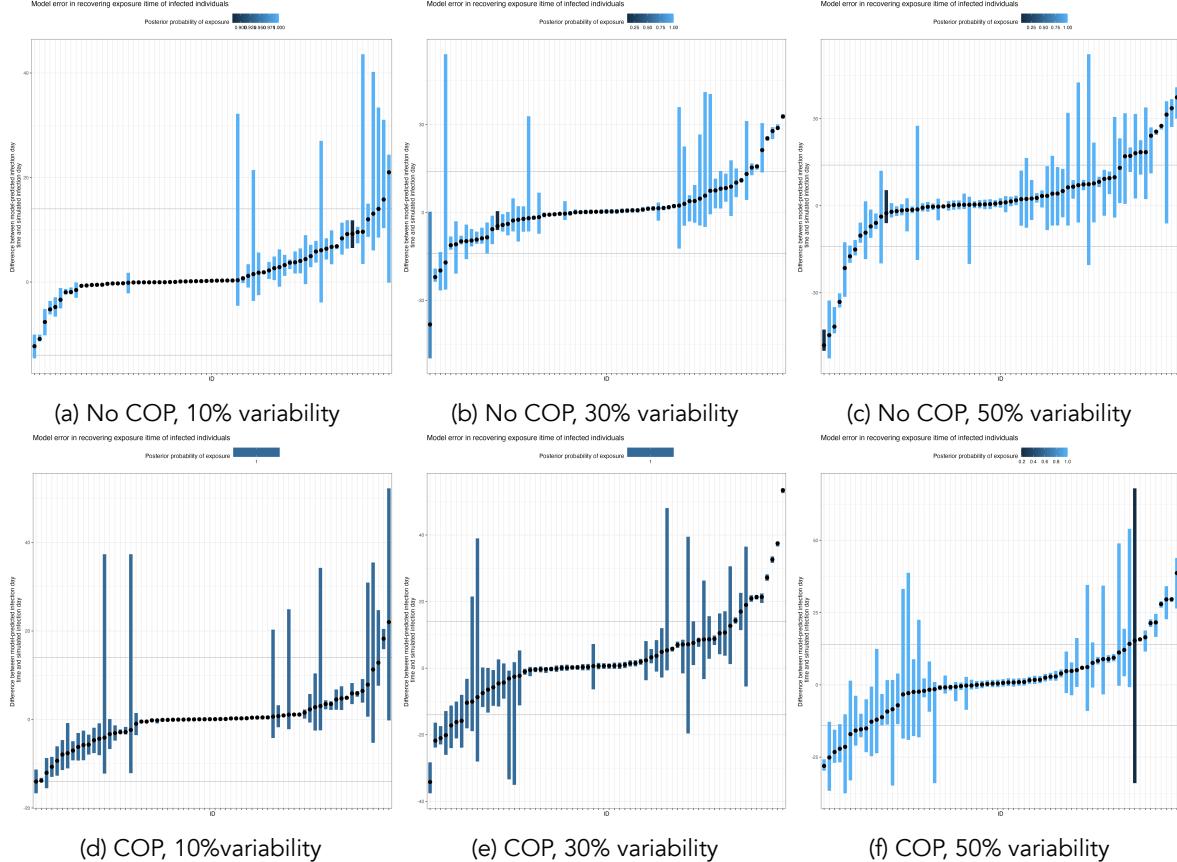


Figure 4.3: Error in estimating infection time two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%)

4.5.3 Infection state recovery

We also recover the infection status of each individual from the simulated data. Suppose the set posterior samples of the infection status for individual j is given by \hat{Z}_j . In that case, we plot the expectation $\mathbb{E}(\hat{Z}_j)$ so we can assess the ability of the algorithm to recover the individual-level simulated infection status (**Figure 4.4**). Nearly all infection states are accurately recovered for all six models.

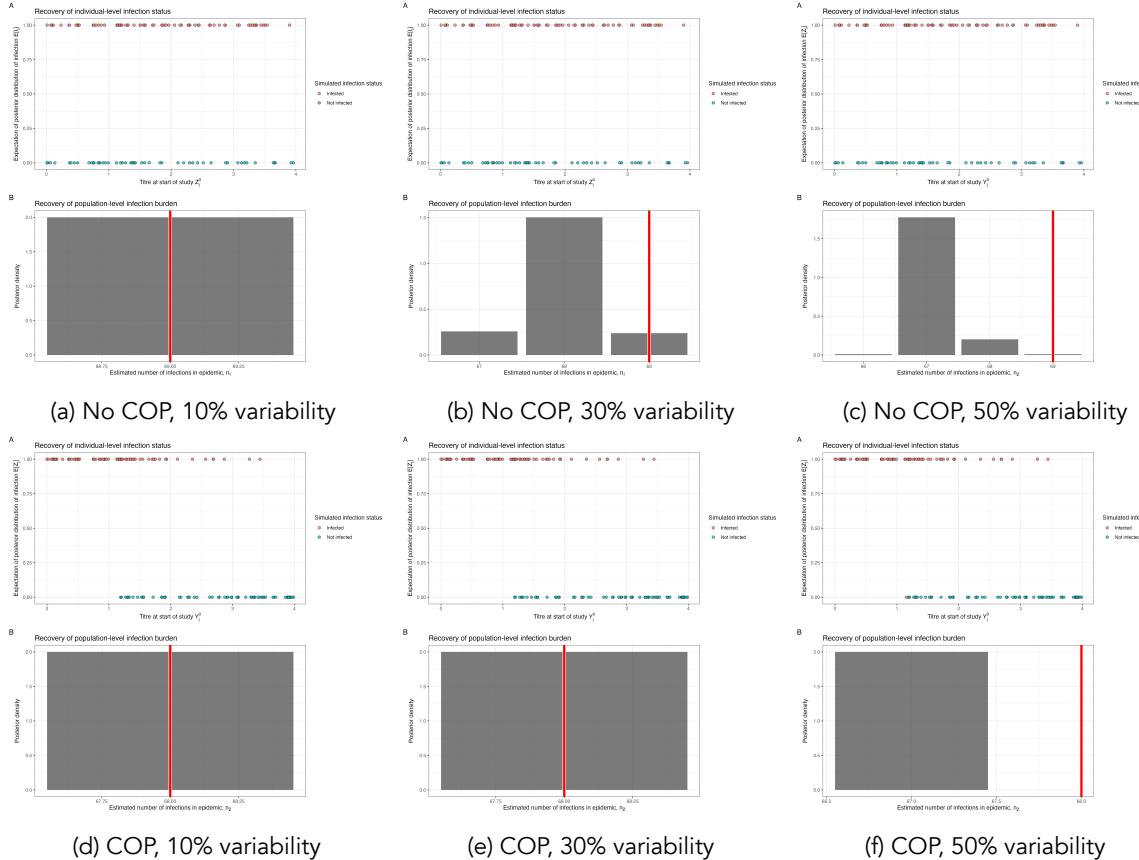


Figure 4.4: Simulation recovery of the individual infection status, \hat{Z}_j , for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%)

4.5.4 Correlate of protection

Algorithm 4.1 performs well at recovering the correlate of protection function $f_{cop}(x, \hat{\theta}_{cop})$, where x is the titre value at infection and where $\hat{\theta}_{cop} = \{\hat{\beta}_0, \hat{\beta}_1\}$ are the posterior samples for β_0 and β_1 . For Model A, we find that the COP curve is recovered, with the simulated line within a 50% confidence interval of the posterior sample (**Figure 4.5**). For Model B, we find the logistic shape of the COP is recovered in the posterior samples.

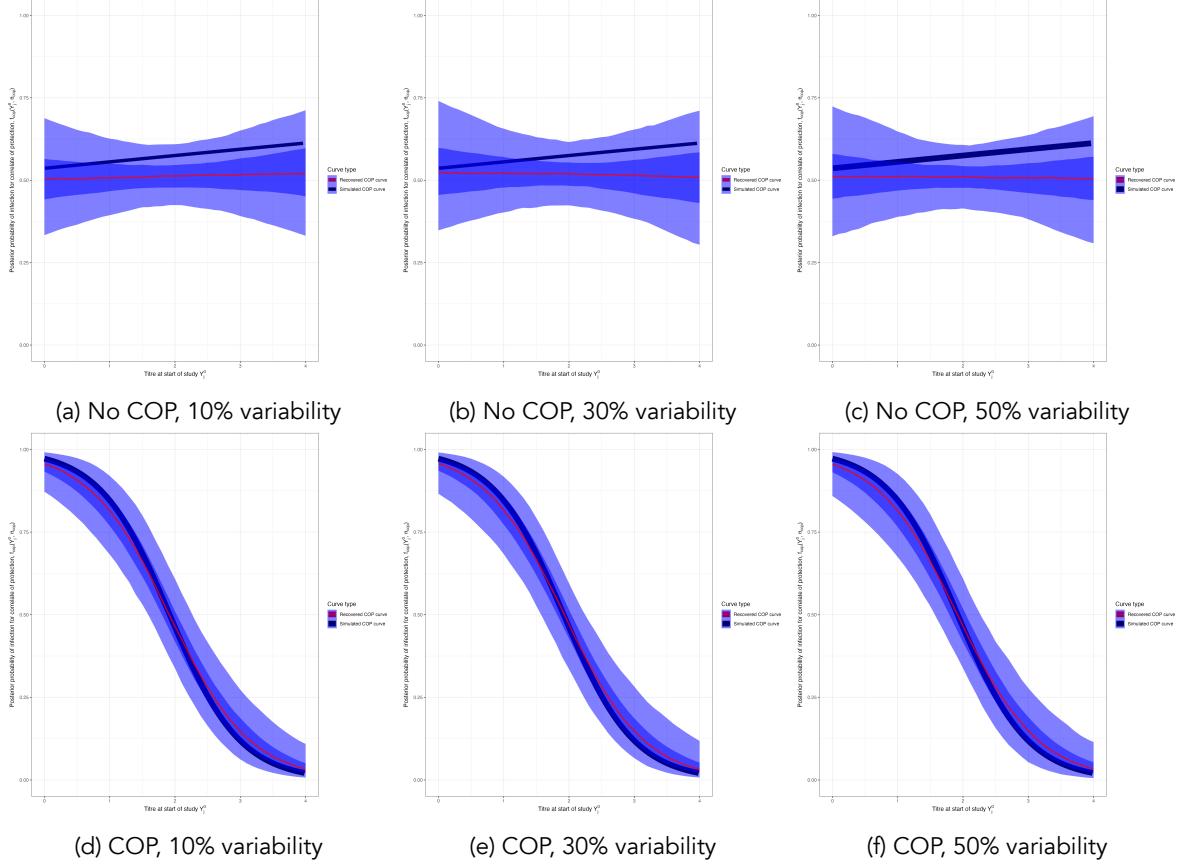


Figure 4.5: Simulation recovery of the COP function, posterior samples plot $f_{cop}(x, \hat{\theta}_{cop})$. We have two different COP models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

4.5.5 Antibody kinetics

Algorithm 4.1 also recovers the simulated antibody kinetics. Let us plot $f_{ab}(s, \hat{a}, \hat{b}, \hat{c})$, the posterior predictive distribution for the antibody kinetic boosting, given posterior distributions for \hat{a} , \hat{b} , and \hat{c} . For all six models, the some of the posterior distributions for \hat{a} , \hat{b} , and \hat{c} differ slightly from the simulated value. However, despite this, the recovered mean trajectory seems to well represent the individual-level kinetics from the simulated data (**Figure 4.6**).

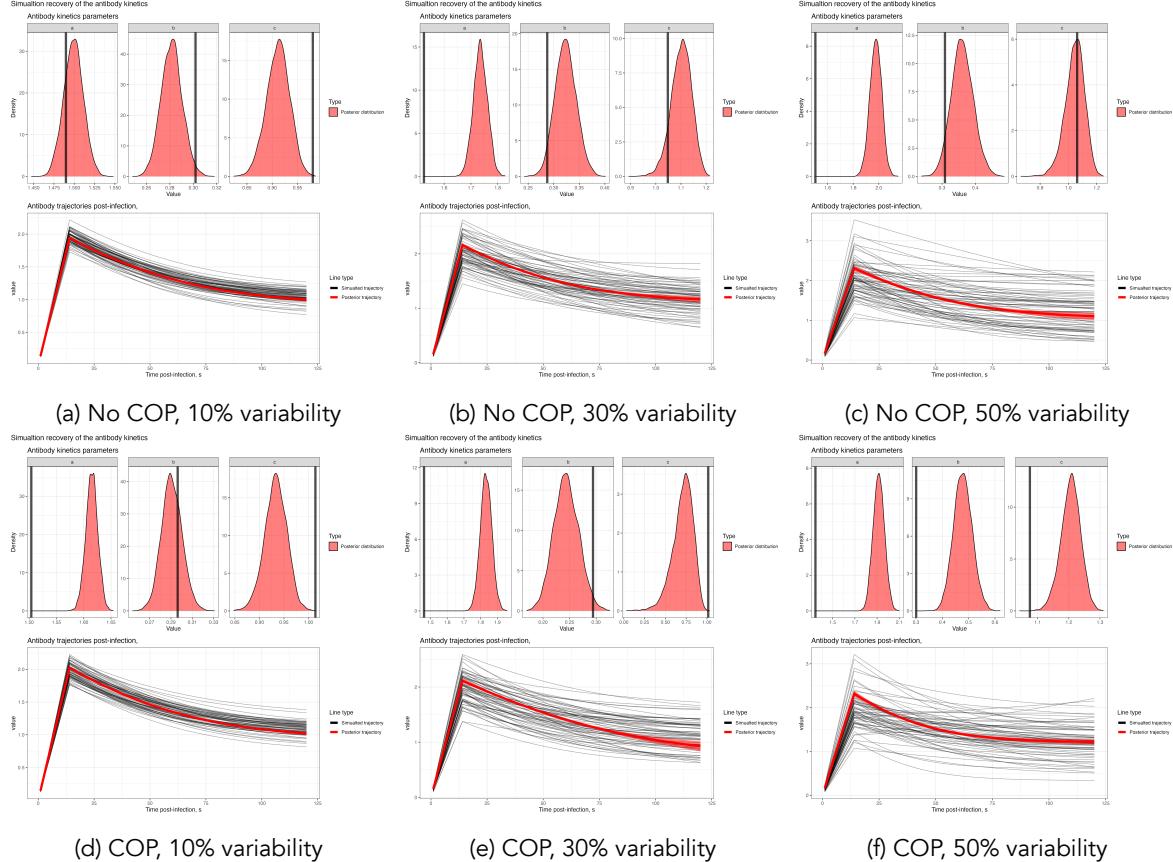


Figure 4.6: Simulation recovery of the antibody kinetics function with posterior samples plot $f_{ab}^1(s, \hat{a}, \hat{b}, \hat{c})$. We have two different COP models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

5 Looking forward

We have shown that **Algorithm 4.1** can recover the state variables $\{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}\}$, the correlate of protection and the antibody kinetics from simulated serological data.[31] The correlate of protection and infection states are well-recovered across all datasets simulated for contrasting COP curves. As the variability of the antibody kinetics increases, the ability of the algorithm to infer the exposure time on the individual level weakens, but, on the population level, the epidemic curve is still accurately recovered.

In the current framework, the user can choose the functional form of the antibody kinetics and COP, allowing for flexibility in the inference methods, which can be tailored to the pathogen being analysed. For example, if the timing of a vaccination programme is known (and stored in a vector, \mathbf{V}), then vaccination kinetics could be added into the antibody kinetics function and inferred within the framework. This is useful for prospective cohort studies which follow vaccinated cohorts and infer correlates of protection/infection by looking at breakthrough infections. Additionally, hierarchical effects can be added to either the antibody kinetics or the correlates of protection such that the impact of host factors on both of these immunological processes can be evaluated. This is crucial for determining the impact of host factors such as age and exposure history on resulting immunological characteristics over the season. Hierarchical models like this can be used to assess the legitimacy of immune imprinting or original antigenic sin by assessing how exposure history alters the COP.

Future extensions to this framework include i) adding the possibility of inferring multiple exposures for an individual over an epidemic period and ii) adding inference for multiple biomarkers and antigenically varied pathogens. Inferring multiple exposures per individual over a larger timeframe will allow the exploration of the impact of poorly understood longer-term immunological phenomena such as original antigenic sin, immune imprinting, etc. Adding inference for multiple biomarkers can help better infer infection status for antigenically varied pathogens such as influenza, SARS-CoV-2, etc. Combining these two new features permits the life history of infections to be inferred given immunological titre landscapes, similar to the methods such as the serosolver framework[33]. However, these extensions require a high number of parameters, greatly increasing the inferred state space and leading to sampling times which may be prohibitively long. These could be overcome by optimising existing analysis, e.g. i) coding the likelihood in C++, or ii) finding more optimal scalings in the RJ-MCMC to ensure optimal convergence. Alternatively, adding a population-level MCMC algorithm (such as parallel tempering) with a reversible jump could improve mixing considerably, allowing for more complex frameworks to be evaluated.

This document has provided details of the theoretical underpinning and implementation of an RJ-MCMC algorithm, which can infer important epidemiology and immunological information from individual-level serological data. Given broad structural forms for antibody kinetics and the correlation of protection, it can recover the exposure status, exposure times, and infection status. We hope this technique will be useful for inferring epidemiological information in a pathogen-agnostic setting, particularly pathogens for which intense surveillance is challenging. We also hope this document sheds light on a mathematically complex but powerful inference tool and encourages others to implement similar algorithms in other health science areas requiring the exploration of multidimensional model spaces.

Appendices

A Additional methods

A.1 Metropolis-Hastings algorithm

A.1.1 Overview

The Metropolis-Hastings (MH) algorithm is a widely used method for generating samples from a target probability distribution. It falls under the broader category of Markov Chain Monte Carlo (MCMC) methods and is particularly useful when direct sampling from the desired distribution is challenging or impossible such as the likelihood described above. A Markov chain-based approach that iteratively generates a sequence of samples, which eventually converge to the desired distribution.

Say we wish to sample from an intractable probability distribution $P(x)$. The idea of the MH is to define a Markov chain so that the stationary distribution of the Markov chain is $P(x)$. That is, the resulting Markov chain from MH generates a sequence of values, denoted $\{x_1, x_2, \dots, x_n\}$, such that as $n \rightarrow \infty$ we can guarantee that $x_i \sim P(x)$. To do this, we uniquely define the Markov chain by its transition probabilities from x to x' , $F(x'|x)$, that must satisfy the detailed balance condition:

$$F(x' | x)P(x) = F(x | x')P(x') \quad (\text{A.1})$$

This condition ensures that the i) probability density for the next step of the Markov chain is the same as the current density and that ii) this probability density is equal to the posterior. To construct a transition probability which satisfies this condition, we split P into a proposal distribution $q(x'|x)$ and an acceptance probability $\alpha(x, x')$:

$$F(x' | x) = q(x'|x)\alpha(x, x') \quad (\text{A.2})$$

A common choice for $\alpha(x, x')$, which satisfies the detailed balance condition, is the acceptance ratio given by

$$\alpha(x, x') = \min \left(1, \frac{P(x')}{P(x)} \cdot \frac{Q(x | x')}{Q(x' | x)} \right) \quad (\text{A.3})$$

With this, the user has a choice over the proposal distribution Q , which can be tailored to optimise the general algorithm given in **Algorithm A.1**.

Algorithm A.1 Generic Metropolis-Hastings Algorithm

- 1: Initialise the chain with an initial state $\theta^{(0)}$
- 2: **for** $i = 1$ to N **do**
- 3: Generate a candidate state θ' from the proposal distribution: $\theta' \sim Q(\cdot | \theta^{(i)})$
- 4: Compute the acceptance ratio:

$$\alpha(\theta^{(i)}, \theta') = \min \left(1, \frac{P(\theta')}{P(\theta^{(i)})} \cdot \frac{Q(\theta^{(i)} | \theta')}{Q(\theta' | \theta^{(i)})} \right)$$

- 5: Sample $u \sim \mathcal{U}(0, 1)$
 - 6: **if** $u \leq \alpha$ **then**
 - 7: Accept the candidate state: $\theta^{(i+1)} \leftarrow \theta'$
 - 8: **else**
 - 9: Reject the candidate state: $\theta^{(i+1)} \leftarrow \theta^{(i)}$
 - 10: **end if**
 - 11: **end for**
-

A.2 Reversible-Jump MCMC algorithm

Let $\{k \in \mathcal{K}\}$ denote a collection models and θ_k be the parameter space of model k . We wish to build a Markov chain Monte Carlo algorithm to sample from the stationary distribution:

$$P(\theta_k, k|Y) \propto p(k)p(\theta_k|k)p(Y|\theta_k, k) \quad (\text{A.4})$$

where $p(k)$ is the prior probability that model k is chosen, $p(\theta_k|k)$ is the prior distribution for parameters θ_k in model k , and $p(Y|k, \theta_k)$ is the likelihood for the observed data for model k . However, as the dimensions of vector θ_k change as we switch between models with different dimensions, there is no way straightforward way to define Q and α such that the detailed balance condition (**Equation A.1**) is met. That is, the posterior density for the proposal state cannot be the same as the current density as the dimensions have changed. Therefore, the sampler would not converge to a single posterior distribution.

The RJ-MCMC proposes a solution to this issue [36]. The idea is to augment both the current state and the proposed state with sampled parameters, define a bijection between these two augmented spaces, and then redefine the acceptance ratio, α , such that the detailed balance condition holds. Let $x = (k, \theta_k)$ denote the model number k and θ_k the parameters associated with model k ($\theta_k \in \mathbb{R}^{d_k}$) then define the proposed state as $x' = (k', \theta_{k'})$, with $\theta_{k'} \in \mathbb{R}^{d_{k'}}$. We write the proposal $Q(x'|x)$, the probability of moving to state x' from state x in the form

$$Q(x'|x) = Q((k', \theta_{k'})|(k, \theta_k)) = q_X(\theta_{k'}|\theta_k, k', k) \cdot q_k(k'|k) \quad (\text{A.5})$$

where $q_k(k'|k)$ is the probability of selecting model k' from model k and q_X the probability of sampling $\theta_{k'}$ given current parameters θ_k and with known k , and known proposed model k' . The challenge with q_X is that we must adjust for the change in dimensions of the parameter space of $\theta_{k'}$ compared to θ_k (i.e $d_k \neq d_{k'}$). To do this, we sample auxiliary variables to match the dimensions and define a bijection between the augmented spaces. Thus if $d_k \neq d_{k'}$, we generate a random variables of length s , $\mathbf{u} = (u_1, \dots, u_s) \sim q_1(\mathbf{u})$ and one of length s' , $\mathbf{u}' = (u'_1, \dots, u'_{s'}) \sim q_2(\mathbf{u}')$ such that $d_{k'} + s' = d_k + s$. We then define a bijection, T

$$(\theta_{k'}, \mathbf{u}') = T(\theta_k, \mathbf{u}) \quad (\text{A.6})$$

to ensure the reversibility of the proposal distribution.

For the detailed balance condition to hold, Green[36] shows that we can obtain the desired properties with a proposal distribution

$$Q(x|x') = q_k(k|k')q_X(\theta_{k'}|\theta_k, k, k') = q_k(k|k')q_2(\mathbf{u}') \left| \frac{\partial(\theta_{k'}, \mathbf{u}')}{\partial(\theta_k, \mathbf{u})} \right|$$

$$Q(x'|x) = q_k(k'|k)q_X(\theta_k|\theta_{k'}, k, k') = q_k(k'|k)q_1(\mathbf{u})$$

where $\left| \frac{\partial(\theta_{k'}, \mathbf{u}')}{\partial(\theta_k, \mathbf{u})} \right|$ is the jacobian of the transformation T . Then, choosing an acceptance ratio given

$$\alpha(x, x') = \min \left(1, \frac{P(x)q_k(k|k')q_2(\mathbf{u}')}{P(x')q_k(k'|k)q_1(\mathbf{u})} \cdot \left| \frac{\partial(\theta_{k'}, \mathbf{u}')}{\partial(\theta_k, \mathbf{u})} \right| \right) \quad (\text{A.7})$$

ensures the stationary distribution chain samples:

$$P(\theta_k, k|Y) \propto p(k)p(\theta_k|k)p(Y|\theta_k, k) \quad (\text{A.8})$$

A general form of the RJ-MCMC then follows **Algorithm A.2**.

Algorithm A.2 Reversible-Jump MCMC Algorithm

- 1: Choose a model k
- 2: Initialize the chain with an initial state $\theta_k^{(0)}$
- 3: **for** $i = 1$ to N **do**
- 4: Sample model $k' \sim q(\cdot|k^{(i)})$
- 5: Sample $\mathbf{u} \sim q_2(\mathbf{u})$
- 6: Set $(\theta_{k'}, \mathbf{u}') = T(\theta_k^{(i)}, \mathbf{u})$
- 7: Compute the acceptance ratio:

$$\alpha((k^{(i)}, \theta_k^{(i)}), (k', \theta_{k'})) = \min \left(1, \frac{P(k', \theta_{k'} | \mathbf{Y}) q(k^{(i)} | k') q_2(\mathbf{u}')}{P(k^{(i)}, \theta_k^{(i)} | \mathbf{Y}) q(k' | k^{(i)}) q_1(\mathbf{u})} \cdot \left| \frac{\partial(\theta_{k'}, \mathbf{u}')}{\partial(\theta_k^{(i)}, \mathbf{u})} \right| \right)$$

- 8: Generate a uniform random number u from the interval $[0, 1]$
 - 9: **if** $u \leq \alpha$ **then**
 - 10: Accept the candidate state: $k^{(i+1)} \leftarrow k'$ and $\theta^{(i+1)} \leftarrow \theta_{k'}$
 - 11: **else**
 - 12: Reject the candidate state: $k^{(i+1)} \leftarrow k^{(i)}$ and $\theta^{(i+1)} \leftarrow \theta^{(i)}$
 - 13: **end if**
 - 14: **end for**
-

B Notation

Symbol	Description
<i>State variables</i>	
$Y_{j,t} \in \mathbf{Y}$	Data on the antibody titre at time t for individual j .
Y_j^0	Initial titre (titre at first value of t for individual j).
$A_{j,t} \in \mathbf{A}$	Model estimated antibody titre at time t for individual j .
$\mathbf{E} = \{E_1, \dots, E_j, \dots, E_M\}$	Vector of exposure statuses (binary vector) for each individual j . A superscript (i) specifies a specific value in the Markov chain.
$n_{\mathbf{E}}$	Total number of individuals exposed. . A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{E}_1 = \{j_1, \dots, j_{j^*}, \dots, j_{n_{\mathbf{E}}}\}$	Vector of individuals j^* who are exposed. A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{E}_0 = \{j_1, \dots, j_{j^*}, \dots, j_{M-n_{\mathbf{E}}}\}$	Vector of individuals j^* who are not exposed. A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{E}^\tau = \{E_{j_1}^\tau, \dots, E_{j_{j^*}}^\tau, \dots, E_{j_{n_{\mathbf{E}}}}^\tau\}$	Vector of exposure times for each individual j . A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{Z} = \{Z_1, \dots, Z_{j^*}, \dots, Z_{n_{\mathbf{E}}}\}$	Vector of infection statuses (binary vector) for each individual j . A superscript (i) specifies a specific value in the Markov chain.
$n_{\mathbf{Z}}$	Total number of individuals infected. A superscript (i) specifies a specific value in the Markov chain.
$\theta_{cop} = \{\beta_0, \beta_1\}$	Fitted parameters for the correlate of protection model
$\theta_{ab} = \{a, b, c\}$	Fitted parameters for the antibody kinetics model
$\theta = \{\theta_{cop}, \theta_{ab}, \sigma\}$	All fitted parameters in the model. A superscript (i) specifies a specific value in the Markov chain.
<i>Functions</i>	
$P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z} \mathbf{Y})$	Posterior distribution function given inputs $\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}$ and data \mathbf{Y}
$\mathcal{L}(\mathbf{Y} \theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z})$	Likelihood function given inputs $\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{Z}$ and data \mathbf{Y} .
$\mathcal{L}_{E_j=0}(Y_j, Z_j, E_j^\tau, E_j, \theta)$	Likelihood function for individual j who is not exposed
$\mathcal{L}_{E_j=1}(Y_j, Z_j, E_j^\tau, E_j, \theta)$	Likelihood function for individual j who is exposed
$P_t(E_j^\tau)$	Likelihood of an exposure at time E_j^τ given individual j is exposed
$A_{j,t} = g_{ab}(t, Z_i, E_j^\tau, \theta_{ab}, Y_j^0)$	Deterministic function for the estimated antibody titre at time t for individual j and starting titre values from the data, Y_j^0 for an exposure at time E_j^τ and infection status Z_j .
$f_{ab}(s, a, b, c)$	The function which determines the antibody titres at time s after E_j^τ
$P_{cop}(I_j Y_j^0, \theta_{cop})$	Likelihood for the correlate of protection for an individual j with an exposure at time E_j^τ , and estimated titre value Y_j^0 and infection status Z_j .
$f_{cop}(Y_j^0, \beta_0, \beta_1)$	Function describing the correlate of protection for infection at time t . (logistic).
$P_{obs}(Y_{j,t} A_{j,t}, \sigma)$	Likelihood of the observation model for the data $Y_{j,t}$ given model-estimated titre values $A_{j,t}$ for individual j at time t .
$\pi(\theta) = \pi(a)\pi(b)\dots\pi(\sigma)\pi(\mathbf{Z})\pi(\mathbf{E})$	Prior distributions for all fitted parameters in the model.

Table B.1: Symbols used in calculating the posterior distribution

C Adaptive Proposal Distribution

We use an adaptive proposal distribution $q_\theta(\theta)$ to sample the parameter space θ . The adaptive metropolis hasting algorithm provides a systematic method for modifying the shape of the proposal distribution based on the accepted steps of the current Markov chain, allowing for a more efficient mixing of chains. That is the $q_\theta(\theta^{(i)}) = N(\theta^{(i)}, \Sigma^{(i)}(\theta^{(i)}))$ follows a Gaussian distribution. To provide a reasonable estimate for the covariance matrix $\Sigma^{(i)}$, the Markov chain runs for an initial number of steps (T_{init}) from a truncated

multivariate normal proposal distribution with a covariance matrix, I_s , whose entries are calculated using the upper and lower bounds of the support of the priors $[s_0^k, s_1^k] \in \mathcal{S}$, through $i_{k,k} = (s_1^k - s_0^k)/\zeta$ and $i_{i,j} = 0$ otherwise, where ζ is a scaling factor.

Problemsatically, the proposal distribution using the updated covariance matrix, $\Sigma^{(i)}$, is no longer memoryless, and therefore, the chain may no longer converge to the correct stationary distribution. To overcome this problem, the proposal distribution must also sample from a non-adaptive multivariate Gaussian distribution modified to ensure that changes to the covariance matrix diminish over time. Further, to improve chain mixing and to optimise convergence rates, I include adaptive scaling factors, $\lambda^{(i)}$ and $M^{(i)}$ for the initial non-adaptive and adaptive proposals, respectively, whose magnitude diminishes with the number of steps in the chain. The adaptive scaling factor for the non-adaptive proposal distributions stops once the model starts sampling from the adaptive proposal distributions. Overall, the combined non-adaptive and adaptive proposal distributions for the adaptive Metropolis-Hastings are given by

$$\frac{i}{q(\cdot|\theta^{(i)})} \begin{cases} i \leq T_{init} \\ \mathcal{N}(\theta^{(i)}, \exp(\lambda^{(i)})I_s; \mathcal{S}) \end{cases} \quad \begin{cases} i > T_{init} \\ \frac{\mathcal{N}(\theta^{(i)}, \Sigma^{(i)}; \mathcal{S})}{\mathcal{N}(\theta^{(i)}, \exp(\lambda^{(T_{init})})I_s; \mathcal{S})} \end{cases} \quad \begin{cases} \text{with probability } \beta, \\ \text{with probability } 1 - \beta \end{cases} \quad (C.1)$$

where $\Sigma^{(i)} = \exp(M^{(i)})\Gamma^{(i)}$ and $M^{(i)}$, $\lambda^{(i)}$ and $\Gamma^{(i)}$ are updated iteratively through the stochastic approximation algorithm:

$$\begin{aligned} \lambda^{(i+1)} &= \lambda^{(i)} + \gamma_1(i)(a(\theta^{(i)}, \theta') - 0.234) \\ M^{(i+1)} &= M^{(i)} + \gamma_2(i)(a(\theta^{(i)}, \theta') - 0.234) \\ \mu^{(i+1)} &= \mu^{(i)} + \gamma_3(i)(\mu^{(i)} - \theta^{(i)}) \\ \Gamma^{(i+1)} &= \Gamma^{(i)} + \gamma_4(i)[(\theta^{(i)} - \mu^{(i+1)})(\theta^{(i)} - \mu^{(i+1)})^T - \Gamma^{(i)}] \end{aligned}$$

where $\gamma_x(i)$ are gain factors. Note when $i > T_{init}$ up stop updaing $\lambda^{(i)}$.

In our implementation, we define $\psi_{adapt}^{(i)} = \{M^{(i)}, \mu^{(i)}, \Gamma^{(i)}, \lambda^{(i)}\}$, and choose values, $\beta = 0.05$, $\zeta = 100$, $\lambda^{(0)} = \log(0.1^2/|\theta_i|)$, $M^{(0)} = \log(2.382^2/|\theta^{(0)}|)$, $\mu^{(0)} = \pi_0$, $\Gamma^{(0)} = I_s$, and $\gamma_x(i) = (1+i)^{-0.5}$ for all x .

D Trace plots

D.1 Known exposure

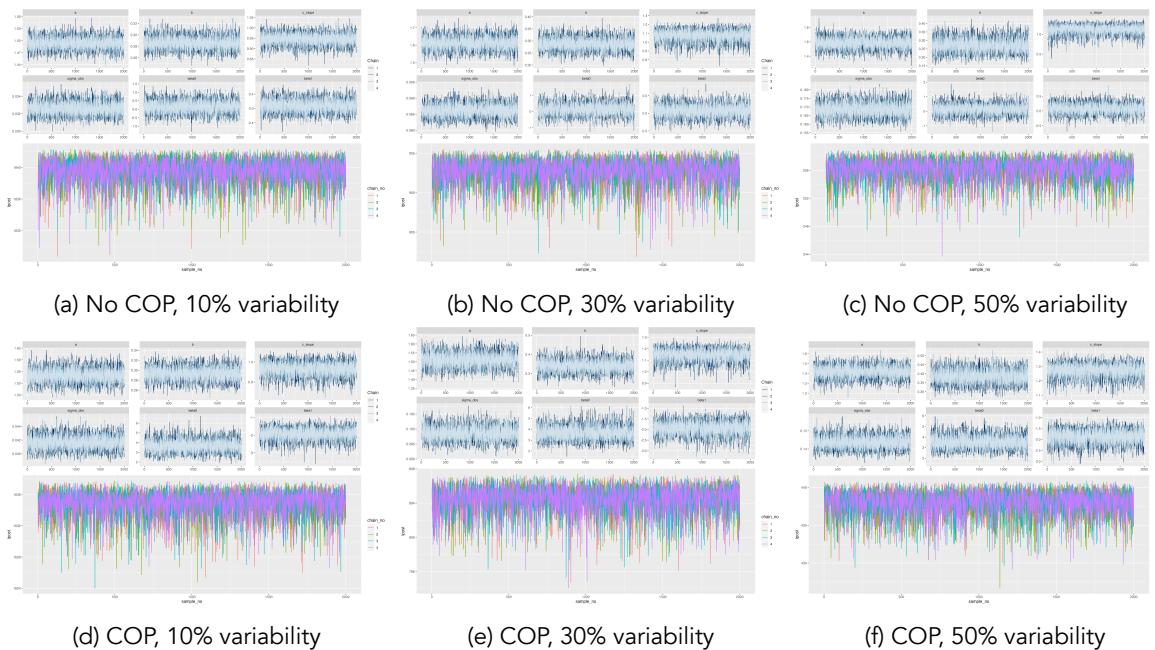


Figure D.1: Simulation recovery of infection status and epidemic curve for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%),

D.2 Inferred exposure

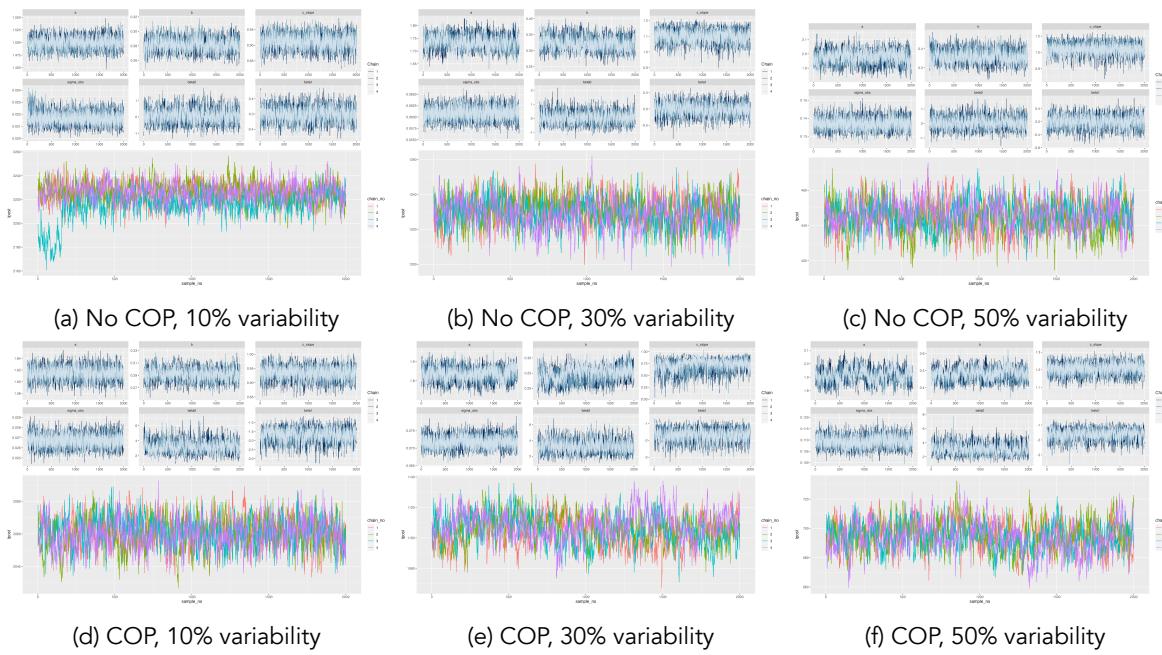


Figure D.2: Simulation recovery of infection status and epidemic curve for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%).

References

- [1] Felicity T Cutts and Matt Hanson. Seroepidemiology: an underused tool for designing and monitoring vaccination programmes in low- and middle-income countries. *Trop. Med. Int. Health*, 21(9):1086–1098, September 2016.
- [2] Andrea H Haselbeck, Justin Im, Kristi Prifti, Florian Marks, Marianne Holm, and Raphaël M Zellweger. Serology as a tool to assess infectious disease landscapes and guide public health policy. *Pathogens*, 11(7), June 2022.
- [3] Parnali Dhar-Chowdhury, Kishor Kumar Paul, C Emdad Haque, Shakhawat Hossain, L Robbin Lindsay, Antonia Dibernardo, W Abdullah Brooks, and Michael A Drobot. Dengue seroprevalence, seroconversion and risk factors in dhaka, bangladesh. *PLoS Negl. Trop. Dis.*, 11(3):e0005475, March 2017.
- [4] Tanyaporn Wansom, Sant Muangnoicharoen, Sorachai Nitayaphan, Suchai Kitsiripornchai, Trevor A Crowell, Leilani Francisco, Paileen Gilbert, Dixon Rwakasyaguri, Jittima Dhitavat, Qun Li, David King, Merlin L Robb, Kirsten Smith, Elizabeth A Heger, Siriwat Akapirat, Punnee Pitisuttithum, Robert J O’Connell, and Sandhya Vasan. Risk factors for HIV sero-conversion in a high incidence cohort of men who have sex with men and transgender women in bangkok, thailand. *EClinicalMedicine*, 38:101033, August 2021.
- [5] Dorothy H Crawford, Karen F Macsween, Craig D Higgins, Ranjit Thomas, Karen McAulay, Hilary Williams, Nadine Harrison, Stuart Reid, Margaret Conacher, Jill Douglas, and Anthony J Swerdlow. A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis. *Clin. Infect. Dis.*, 43(3):276–282, August 2006.
- [6] Yuyen Chan, Kimberly Fornace, Lindsey Wu, Benjamin F Arnold, Jeffrey W Priest, Diana L Martin, Michelle A Chang, Jackie Cook, Gillian Stresman, and Chris Drakeley. Determining seropositivity—a review of approaches to define population seroprevalence when using multiplex bead assays to assess burden of tropical diseases. *PLoS Negl. Trop. Dis.*, 15(6):e0009457, June 2021.
- [7] Oda E van den Berg, Kamelia R Stanoeva, Rens Zonneveld, Denise Hoek-van Deursen, Fiona R van der Klis, Jan van de Kassteele, Eelco Franz, Marieke Opsteegh, Ingrid H M Friesema, and Laetitia M Kortbeek. Seroprevalence of toxoplasma gondii and associated risk factors for infection in the netherlands: third cross-sectional national study. *Epidemiol. Infect.*, 151:e136, July 2023.
- [8] Hayley Colton, David Hodgson, Hailey Hornsby, Rebecca Brown, Joanne Mckenzie, Kirsty L Bradley, Cameron James, Benjamin B Lindsey, Sarah Birch, Louise Marsh, Steven Wood, Martin Bayley, Gary Dickson, David C James, Martin J Nicklin, Jon R Sayers, Domen Zafred, Sarah L Rowland-Jones, Goura Kudisia, Adam Kucharski, CMMID COVID-19 Working Group, Thomas C Darton, Thushan I de Silva, and Paul J Collini. Risk factors for SARS-CoV-2 seroprevalence following the first pandemic wave in UK healthcare workers in a large NHS foundation trust. *Wellcome Open Res*, 6:220, 2021.
- [9] James Hay, Isobel Routledge, and Saki Takahashi. Serodynamics: a review of methods for epidemiological inference using serological data. November 2023.
- [10] C Jessica E Metcalf, Jeremy Farrar, Felicity T Cutts, Nicole E Basta, Andrea L Graham, Justin Lessler, Neil M Ferguson, Donald S Burke, and Bryan T Grenfell. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet*, 388(10045):728–730, August 2016.
- [11] Cuiling Xu, Ling Liu, Binzhui Ren, Libo Dong, Shumei Zou, Weijuan Huang, Hejiang Wei, Yanhui Cheng, Jing Tang, Rongbao Gao, Lizhong Feng, Ruifu Zhang, Chaopu Yuan, Dayan Wang, and Jing Chen. Incidence of influenza virus infections confirmed by serology in children and adult in a suburb community, northern china, 2018-2019 influenza season. *Influenza Other Respi. Viruses*, 15(2):262–269, March 2021.

- [12] Simon Cauchemez, Peter Horby, Annette Fox, Le Quynh Mai, Le Thi Thanh, Pham Quang Thai, Le Nguyen Minh Hoa, Nguyen Tran Hien, and Neil M Ferguson. Influenza infection rates, measurement errors and the interpretation of paired serology. *PLoS Pathog.*, 8(12):e1003061, December 2012.
- [13] Irene Garcia-Fogeda, Hajar Besbassi, Ynke Larivière, Benson Ogunjimi, Steven Abrams, and Niel Hens. Within-host modeling to measure dynamics of antibody responses after natural infection or vaccination: A systematic review. *Vaccine*, 41(25):3701–3709, June 2023.
- [14] Sylvia Ranjeva, Rahul Subramanian, Vicky J Fang, Gabriel M Leung, Dennis K M Ip, Ranawaka A P M Perera, J S Malik Peiris, Benjamin J Cowling, and Sarah Cobey. Age-specific differences in the dynamics of protective immunity to influenza. *Nat. Commun.*, 10(1):1660, April 2019.
- [15] James A Hay, Karen Laurie, Michael White, and Steven Riley. Characterising antibody kinetics from multiple influenza infection and vaccination events in ferrets. *PLoS Comput. Biol.*, 15(8):e1007294, August 2019.
- [16] Komal Srivastava, Juan Manuel Carreño, Charles Gleason, Brian Monahan, Gagandeep Singh, Anass Abbad, Johnstone Tcheou, Ariel Raskin, Giulio Kleiner, Harm van Bakel, Emilia Mia Sordillo, PARIS Study Group, Florian Krammer, and Viviana Simon. Kinetics and durability of humoral responses to SARS-CoV-2 infection and vaccination. August 2023.
- [17] Xiahong Zhao, Yilin Ning, Mark I-Cheng Chen, and Alex R Cook. Individual and population trajectories of influenza antibody titers over multiple seasons in a tropical country. *Am. J. Epidemiol.*, 187(1):135–143, January 2018.
- [18] P F M Teunis, J C H van Eijkelen, W F de Graaf, A Bonačić Marinović, and M E E Kretzschmar. Linking the seroresponse to infection to within-host heterogeneity in antibody production. *Epidemics*, 16:33–39, September 2016.
- [19] Ian J Amanna and Mark K Slifka. Mechanisms that determine plasma cell lifespan and the duration of humoral immunity. *Immunol. Rev.*, 236(1):125–138, July 2010.
- [20] Mathieu Andraud, Olivier Lejeune, Jammbe Z Musoro, Benson Ogunjimi, Philippe Beutels, and Niel Hens. Living on three time scales: the dynamics of plasma cell and antibody populations illustrated for hepatitis a virus. *PLoS Comput. Biol.*, 8(3):e1002418, March 2012.
- [21] Stanley A Plotkin. Recent updates on correlates of vaccine-induced protection. *Front. Immunol.*, 13:1081107, 2022.
- [22] Frontiers in Immunology. Immune correlates of protection for emerging diseases – lessons from ebola and COVID-19. <https://www.frontiersin.org/research-topics/36500/immune-correlates-of-protection-for-emerging-diseases—lessons-from-ebola-and-covid-19>.
- [23] Meagan E Deming, Nelson L Michael, Merlin Robb, Myron S Cohen, and Kathleen M Neuzil. Accelerating development of SARS-CoV-2 vaccines - the role for controlled human infection models. *N. Engl. J. Med.*, 383(10):e63, September 2020.
- [24] Amrita Sekhar and Gagandeep Kang. Human challenge trials in vaccine development. *Semin. Immunol.*, 50:101429, August 2020.
- [25] Leo Swadling and Mala K Maini. Can T cells abort SARS-CoV-2 and other viral infections? *Int. J. Mol. Sci.*, 24(5), February 2023.
- [26] Leo Swadling, Mariana O Diniz, Nathalie M Schmidt, Oliver E Amin, Aneesh Chandran, Emily Shaw, Corinna Pade, Joseph M Gibbons, Nina Le Bert, Anthony T Tan, Anna Jeffery-Smith, Cedric C S Tan, Christine Y L Tham, Stephanie Kucykowicz, Gloryanne Aidoo-Micah, Joshua Rosenheim, Jessica Davies, Marina Johnson, Melanie P Jensen, George Joy, Laura E McCoy, Ana M Valdes, Benjamin M Chain, David Goldblatt, Daniel M Altmann, Rosemary J Boyton, Charlotte Manisty, Thomas A Treibel, James C Moon, COVIDsortium Investigators, Lucy van Dorp, Francois Balloux, Áine McKnight,

- Mahdad Noursadeghi, Antonio Bertoletti, and Mala K Maini. Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. *Nature*, 601(7891):110–117, January 2022.
- [27] Cheryl Cohen, Jackie Kleynhans, Anne von Gottberg, Meredith L McMorrow, Nicole Wolter, Jinal N Bhiman, Jocelyn Moyes, Mignon du Plessis, Maimuna Carrim, Amelia Buys, Neil A Martinson, Kathleen Kahn, Stephen Tollman, Limakatso Lebina, Floidy Wafawanaka, Jacques D du Toit, Francesc Xavier Gómez-Olivé, Fatimah S Dawood, Thulisa Mkhencele, Kaiyuan Sun, Cécile Viboud, Stefano Tempia, and PHIRST-C Group. SARS-CoV-2 incidence, transmission, and reinfection in a rural and an urban setting: results of the PHIRST-C cohort study, south africa, 2020-21. *Lancet Infect. Dis.*, 22(6):821–834, June 2022.
- [28] J Simonsen, K Mølbak, G Falkenhorst, K A Krogfelt, A Linneberg, and P F M Teunis. Estimation of incidences of infectious diseases based on antibody measurements. *Stat. Med.*, 28(14):1882–1895, June 2009.
- [29] Henrik Salje, Derek A T Cummings, Isabel Rodriguez-Barraquer, Leah C Katzelnick, Justin Lessler, Chonticha Klungthong, Butsaya Thaisomboonsuk, Ananda Nisalak, Alden Weg, Damon Ellison, Louis Macareo, In-Kyu Yoon, Richard Jarman, Stephen Thomas, Alan L Rothman, Timothy Endy, and Simon Cauchemez. Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. *Nature*, 557(7707):719–723, May 2018.
- [30] Tim K Tsang, Ranawaka A P M Perera, Vicky J Fang, Jessica Y Wong, Eunice Y Shiu, Hau Chi So, Dennis K M Ip, J S Malik Peiris, Gabriel M Leung, Benjamin J Cowling, and Simon Cauchemez. Reconstructing antibody dynamics to estimate the risk of influenza virus infection. *Nat. Commun.*, 13(1):1557, March 2022.
- [31] Arthur Menezes, Saki Takahashi, Isobel Routledge, C Jessica E Metcalf, Andrea L Graham, and James A Hay. serosim: An R package for simulating serological data arising from vaccination, epidemiological and antibody kinetics processes. *PLoS Comput. Biol.*, 19(8):e1011384, August 2023.
- [32] Andrew J Dunning, Jennifer Kensler, Laurent Coudeville, and Fabrice Bailleux. Some extensions in continuous models for immunological correlates of protection. *BMC Med. Res. Methodol.*, 15:107, December 2015.
- [33] James A Hay, Amanda Minter, Kylie E C Ainslie, Justin Lessler, Bingyi Yang, Derek A T Cummings, Adam J Kucharski, and Steven Riley. An open source tool to infer epidemiological and immunological dynamics from serological data: serosolver. *PLoS Comput. Biol.*, 16(5):e1007840, May 2020.
- [34] Gareth O Roberts, Jeffrey S Rosenthal, Gareth O R Oberts, and Jeffrey S R Osenthal. Examples of adaptive MCMC. *J. Comput. Graph. Stat.*, 18(2):349–367, 2012.
- [35] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive MCMC. *Stat. Comput.*, 18:343–373, 2008.
- [36] Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, December 1995.
- [37] Faming Liang, Chuanhai Liu, and Raymond Carroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley, August 2010.
- [38] Gavin J Gibson and Eric Renshaw. Estimating parameters in stochastic compartmental models using markov chain methods. *Math. Med. Biol.*, 15(1):19–40, March 1998.