Modelling serological data using reversible jump Markov chain Monte Carlo

David Hodgson, James Hay(?) and Adam Kucharski
Center of Mathematical Modelling of Infectious Diseases
London School of Hygiene and Tropical Medicine

January 26, 2024

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 hold 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 provide protection 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. This phase is characterized by a rapid increase in antibody levels as the immune system recognizes and mounts a defence against the antigen.
- *Peak Antibody Level*: The trajectories then rise to apeak antibody level, which is the highest concentration of antibodies reached during the immune response. This peak can vary depending on factors like the strength of the immune response and the nature of the antigen.
- *Decay Phase*: After the peak, there is a decline in antibody levels as antibodies have a finite lifespan 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, and therefore, decay phase trajectories often converge to a set point titre.[16]

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 are also useful for optimizing vaccination strategies and can help 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 is an immune function or biomarker that correlates with and or may be biologically responsible for protection against infection or disease. Correlates of protection have been established for influenza, Hepatitis A and B, Measles, Polio, Rabies, Yellow fever and more.[19] These established correlates of protection help researchers understand the specific immune responses needed to prevent or control an infectious disease. This knowledge is pivotal in designing and optimizing vaccines to induce the required immune response effectively.[20]

Determining the immunological profiles of individuals who are exposed to an infection but manage to abort it is crucial for determining a universal correlate of protection. Usually, these individuals are identified through challenge studies, where individuals' immune profiles are analysed, and they are then challenged with live viruses.[21] 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, these studies are expensive, difficult to run, have limited generalisability, and are only possible for pathogens with mild pathogenesis.[22]

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.[23] Therefore, serological studies can be augmented with either immunological profiling to determine other immunological biomarkers that indicate abortive infection[24] or include intensive contact tracing with the serological studies to determine exposure rates between individuals.[25] Augmented serological studies increases their complexity and cost and, therefore, are not feasible in many

settings. Consequently, there is a motivation for the methodological development of statistical and mathematical inference techniques, which augment the inference of existing serological studies to elucidate latent epidemiological characteristics of a pathogen without the need for more complex study designs.[26]

1.4 Overview of modelling framework

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 seroconverted throughout the study, ii) subsequent antibody kinetics of these infection individuals, iii) which individuals are 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 infection and exposure status of individuals and the population are recoverable through the interdependencies of mechanisms i)–iv) within a Bayesian probabilistic framework.

2 Simulated serological data from `serosim`[27]

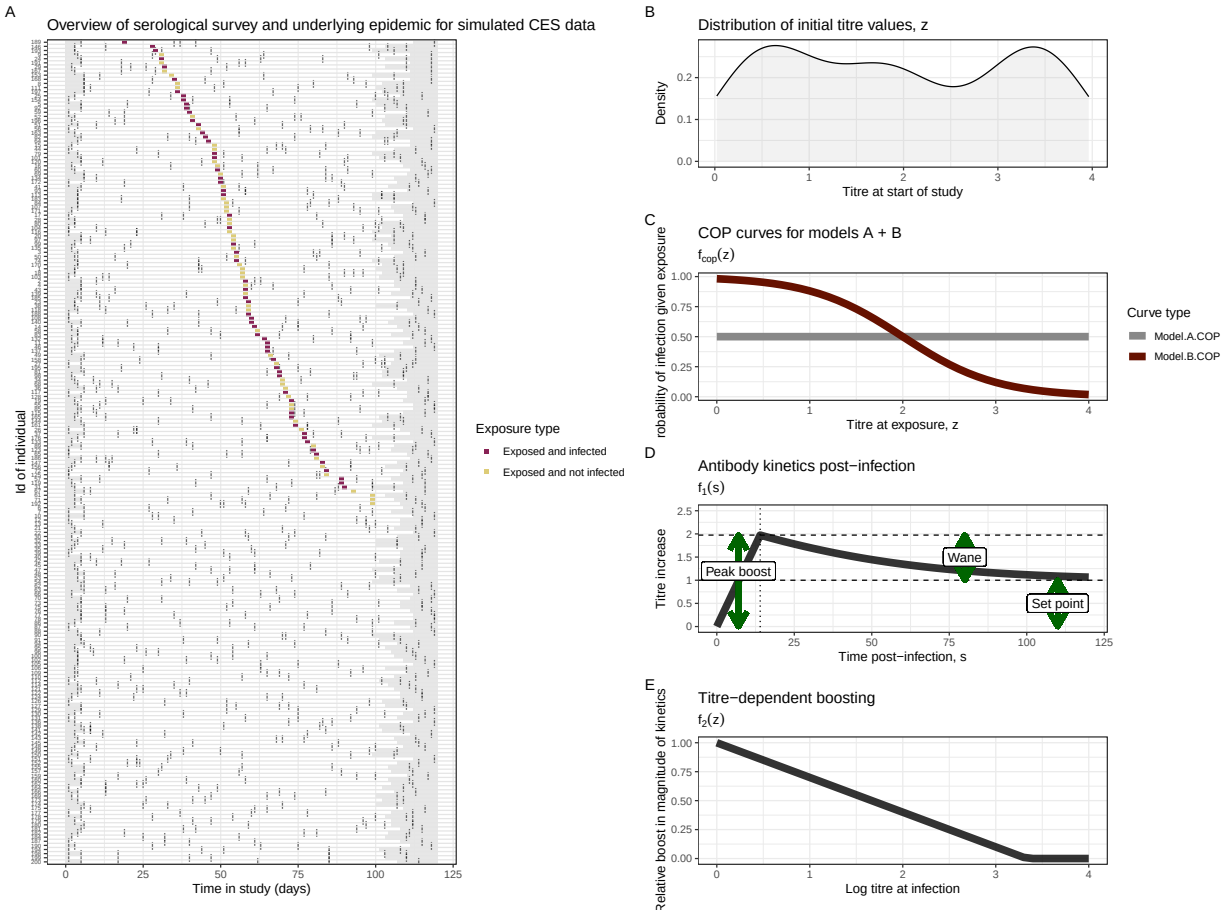


Figure 2.1: Schematics showing the simulated data structure from `serosim`[27]

We simulate serological data using the `serosim`[27] R package to demonstrate our modelling framework's capabilities at simulation recovery. We simulate continuous epidemic serosurveillence (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-B**.

We define the correlate of protection (COP) as the probability of infection given a titre value at exposure. 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:

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

where $\beta_0 = -4$ and $\beta_1 = 2$ in the simulated data and x is the titre value at exposure. This represents a pathogen for which higher antibody titres are associated with higher levels of protection from infection. Note in these data, we assume antibody trajectories remain constant until the timing of infection, such $f_{cop}(X_{j,t}, \beta_0, \beta_1) = f_{cop}(Z_j^0, \beta_0, \beta_1)$ where Z_j^0 is the antibody titre of an individual at their first bleed at the start of the study. A figure showing the two COP models is given in **Figure 2.1C**.

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**.

$$f_{ab}^1(s, a, b, c) = \begin{cases} \ln(\exp(a) + \exp(b))/14, & \text{if } s \leq 14 \\ \ln(\exp(a)\exp(-(b/10)(t-14)) + \exp(c)), & \text{if } s > 14 \end{cases} \tag{2.2}$$

where $a = 1.5$, $b = 2$, and $c = 1$ are values in the simulated data (**Figure 2.1D**) and s is the number of days post infection. We also assume that the magnitude of these dynamics depends on pre-existing titre values, with higher pre-existing values seeing attenuated dynamics relative to lower pre-existing tire values. The titre-dependent boosting is assumed to follow a linear decay truncated at 0, that is, given a titre value z, $f_{ab}^2(Z_j^0, \alpha) = \max(1 - \alpha z, 0)$. where $\alpha = 0.3$ in the simulated data (**Figure 2.1E**). Therefore, the model estimated titre value at time t for individual j, given time s since infection with a titre value at infection of Z_j^0, is given by;

$$f_{ab}(s, Z_j^0, a, b, c, \alpha) = f_{ab}^1(s, a, b, c) f_{ab}^2(Z_j^0, \alpha) \tag{2.3}$$

t

To model heterogeneity in individual-level antibody kinetics, we simulate a, b, c, and α from normal distributions where the mean μ is their simulated values and the standard deviation given by $\mu\sigma^*$. We simulate three levels of uncertainty for $\sigma^* = \{0.1, 0.3, 0.5\}$ 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.

We note that the simulated data for the correlate of protection does not exactly match the functional form chosen when inferred from the infected population. We plot the realised COP simulated 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 ??**.

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^τ	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
α	Parameter describing the titre-dependent boosting of the antibody kinetics	0.3
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 2.3**
f_{cop}	Logistic correlate of protection	**Equation 2.1**

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

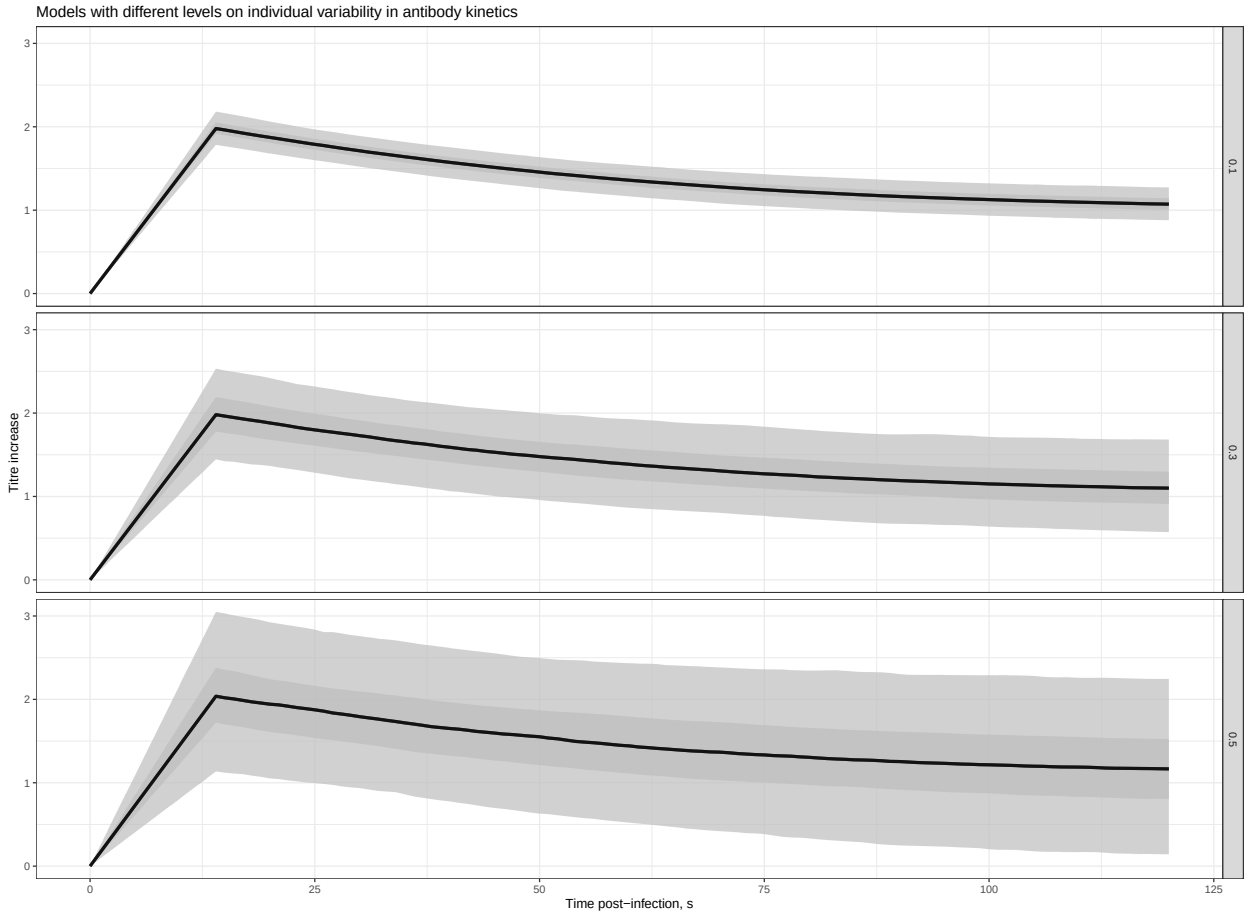


Figure 2.2: Schematics showing three levels of individual-level uncertainty and the impact on the variability of antibody kinetics.

3 Inference with known exposure status

Before we look at the reversible-jump mcmc algorithm (RJMCMC),[28] we will 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 $\mathbf{E}^\tau$) 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 and infection state $\mathbf{I}$ work without needing to describe the more complex inference using RJMCMC.

3.1 Mathematical representation of framework

Let the binary vector $\mathbf{E} = \{E_1, E_2, \ldots, E_M\}$, 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 $\mathbf{E}^\tau = \{E_1^\tau, E_2^\tau, \ldots, E_{n_{\mathbf{E}}}^\tau\}$ and $\mathbf{I} = \{I_1, I_2 \ldots, I_{n_{\mathbf{E}}}\}$ 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 $I_j = 0$ is not infected, and $I_j = 1$ is infected. Let $Z_{j,t} \in \mathbf{Z}$ 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 assume that the model predicted antibody titre at time t in the study ($X_{j.t}$) can be derived given the infection status I_j and timing of exposure E_j^τ. If a person is not infected, their starting titre value (Z_i^0) remains unchanged from the

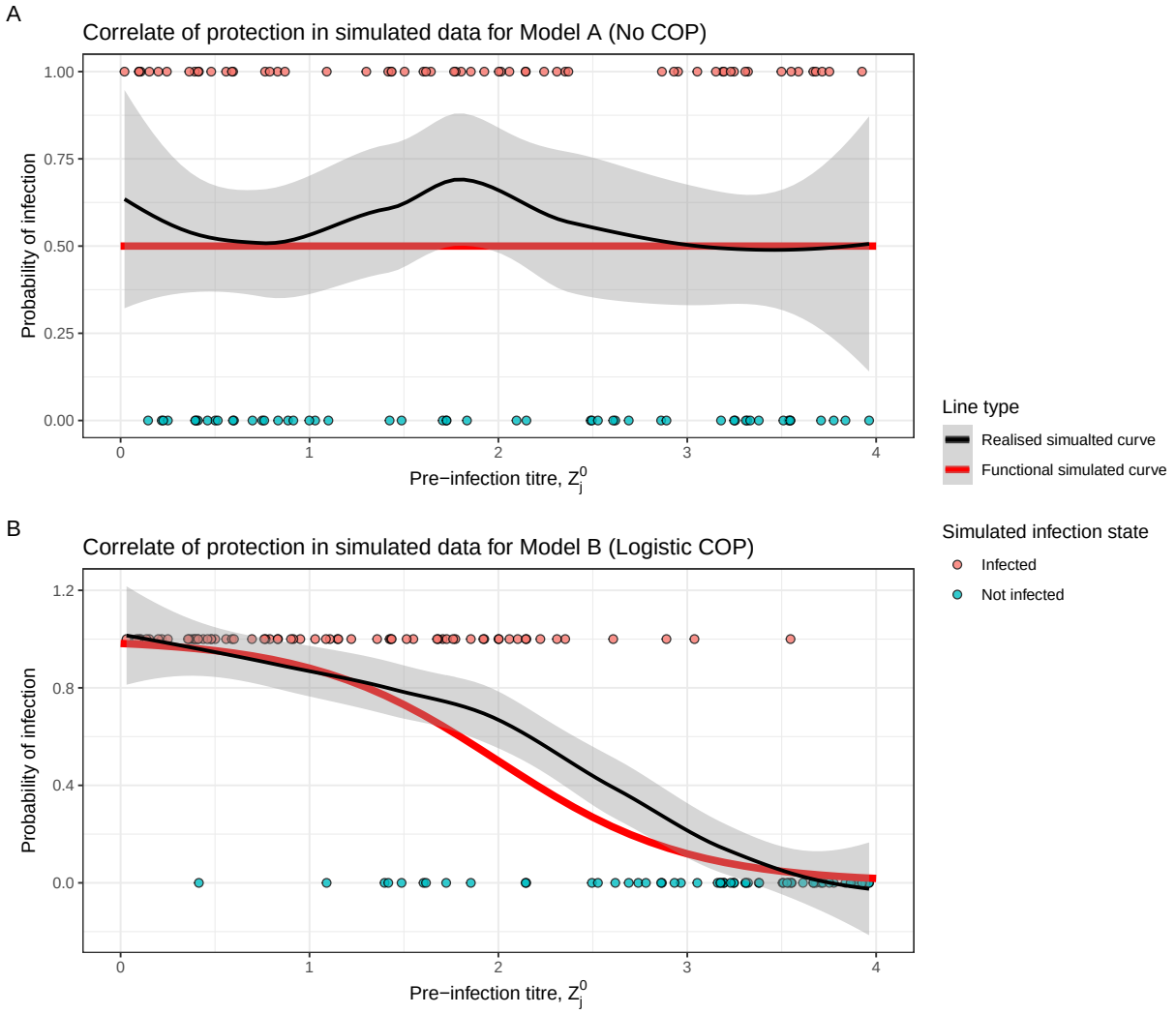


Figure 2.3: Schematics showing the difference between the functional form chosen to simulate the COP and the recovered COP from the exposed individuals.

start of the study. If the person is infected, their titre remains unchanged until the point of infection, at which point they follow the dynamics highlighted in **Equation 2.3**. The deterministic function for calculating $X_{j,t}$ value is given by

$$X_{j,t} = F_{ab}(I_j, E_j^\tau, \theta_{ab}, Z_j^0) = \begin{cases} Z_j^0 + f_{ab}(t - E_j^\tau, \theta_{ab}, Z_j^0), & \text{If } I_j = 1,\ E_j = 1, \text{ and } t > E_j^\tau \\ Z_j^0, & \text{Otherwise} \end{cases} \tag{3.1}$$

Where $\theta_{ab} = \{a, b, c, \alpha\}$. 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 $X_{j,t}$, 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.

$$P_{cop}(I_j \mid Z_j^0, \theta_{cop}) = f_{cop}(Z_j^0, \theta_{cop})^{I_j}(1 - f_{cop}(Z_j^0, \theta_{cop}))^{1-I_j} \tag{3.2}$$

$\theta_{cop} = \{\beta_0, \beta_1\}$. Finally, we define an observational model to capture variability between hosts and measurement error. Given $X_{j.t}$ and the serological antibody data at the same time point is given by $Z_{j,t}$,

we assume the measurement error follows a normal distribution with a PDF given by Equation 3.3.

$$P_{obs}(Z_{j,t} \mid X_{j,t}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left(\frac{(Z_{j,t}-X_{j,t})^2}{2\sigma^2}\right)} \tag{3.3}$$

Let $\theta = \{a, b, c, \alpha, \beta_0, \beta_1, \sigma\}$ 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 ($E_i = 1$) or not ($E_i = 0$).

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:

$$L_{E_j=0}(Z_j|\theta) = \prod_{t \in T_j} P_{obs}(Z_{j,t}|Z_j^0, \sigma) \tag{3.4}$$

as $X_{j.t} = Z_j^0$ for all t.

3.2.2 Likelihood for an exposed individual $E_j = 1$

In this case, the infectious status is determined by the correlate of the protection likelihood (P_{cop}) and the antibody kinetics. 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:

$$L_{E_j=1}(Z_j|I_j, \theta) = \prod_{t \in T_j} P_{obs}(Z_{j,t}|X_{j,t}, \sigma) P_{cop}(I_j \mid Z_j^0, \theta_{cop}) \tag{3.5}$$

where $X_{j,t} = F_{ab}(I_j, E_j^\tau, \theta_{ab}, Z_j^0)$.

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

$$L(\mathbf{Z}|\mathbf{I}, \theta) = \prod_{j \in \mathbf{E_0}} L_{E_j=0}(Z_j|\theta) \prod_{j \in \mathbf{E_1}} L_{E_j=1}(Z_j|I_j, \theta) \tag{3.6}$$

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 ($\mathcal{S}$)
a	$\mathcal{N}(1.5, 0.5)$	$[0.5, 4]$
b	$\mathcal{N}(0.3, 0.05)$	$[0, 1]$
c	$\mathcal{U}(0, 4)$	$[0, 4]$
α	$\mathcal{U}(0, 1)$	$[0, 1]$
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 $n_{\mathbf{I}}$ given the number of exposed individuals $n_{\mathbf{E}}$ to be a Beta Binomial distribution: $\pi(\mathbf{I}) = \text{BetaBinomial}(n_{\mathbf{I}}|n_{\mathbf{E}}, 1, 1)$. Choosing this prior prevents any implicit priors that might rise from products of Bernoulli trials[29] as $\text{BetaBinomial}(n_{\mathbf{I}}|n_{\mathbf{E}}1, 1) = 1/n_{\mathbf{E}}$ for all $0 \leq n_{\mathbf{I}} \leq n_{\mathbf{E}}$.

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

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

3.3 Metropolis-Hasting algorithm

3.3.1 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 distribttion $P(x)$. 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 $\{x_1, x_2, \ldots, x_n\}$, such that as $n \to \infty$ we can guarantee that $x_i \sim P(x)$. To do this, we uniquely define the Markov chain by it's transition probabilities from x to x', $F(x', x)$, that must satisfy the detailed balance condition:

$$F(x' \mid x)P(x) = F(x \mid x')P(x') \tag{3.8}$$

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 $\alpha(x, x')$:

$$F(x' \mid x) = q(x'|x)\alpha(x, x') \tag{3.9}$$

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

$$\alpha(x, x') = \min\left(1, \frac{P(x')}{P(x)} \cdot \frac{Q(x \mid x')}{Q(x' \mid x)}\right) \tag{3.10}$$

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

Algorithm 1 Generic Metropolis-Hastings Algorithm

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

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

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

3.3.2 MH for serological inference with known exposure

In our model, we wish to sample from the posterior density function given by **Equation 3.7**, which infers θ, and infectious statuses $I_j \in \mathbf{I}$, for $1 \leq j \leq M$ individuals. For the proposal distribution, we define independent proposal distribution for θ and $\mathbf{I}$, such that $Q(\cdot|\theta, \mathcal{I}) = q_\theta(\cdot|\theta)q_I(\cdot|\mathbf{I})$. 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 $\theta \sim q_\theta(\cdot|\theta^{(i)}, \psi^{(i)}_{adapt})$. This proposal is a multivariate normal distribution with an adaptive covariance matrix, which is defined by the set of parameters $\psi^{(i)}_{adapt}$, which are updated at each time step.[30, 31] (See Appendix). For $\mathbf{I}$, we propose a new infection state $\mathbf{I}'$ by selecting an exposed individual j, which has infection status $I^{(i)}_j$ at step i of the current Markov chain, we sample a proposal value for their infection status I'_j by the proposal distribution for $I'_j \sim q_I(\cdot|I^{(i)}_j) =$ Bernoulli(0.5). Therefore the proposal for $q_I(\mathbf{I}'|\mathbf{I}) = 1/n_{\mathbf{E}}0.5$ for all j. Both of these proposals $q_\theta\left(\theta'|\theta^{(i)}, \psi^{(i)}_{adapt}\right)$, $q_I(\mathbf{I}'|\mathbf{I})$ are both symmetric and thus cancel out the acceptance ratio (**Equation 3.10**). Further, the prior distribution $\pi(\mathbf{I}) = 1/n_{\mathbf{E}}$ for all $0 \leq n_{\mathbf{I}} \leq n_{\mathbf{E}}$, and thus also cancels out in the acceptance ratio, therefore we need only calculate: $P(\theta, \mathbf{I}|\mathbf{Z}) \propto \mathcal{L}(\mathbf{Z}|\mathbf{I}, \theta)\pi(\theta)$

Consequently, we construct a new algorithm for inference of the known exposure model (**Algorithm 2**).

Algorithm 2 Metropolis-Hastings Algorithm for antibody kinetics and infection inference

Initialize the chain with an initial state $\theta^{(0)}$ from the priors $\pi(\cdot)$ and $I_j^{(0)} \sim$ Bernoulli(0.5) for all $1 \leq j \leq M$ individuals to intialise $\mathbf{I}^{(0)}$, and initialise $\psi_{adapt}^{(0)}$.
for $i = 1$ to N **do**
  Generate a candidate state $\theta' \sim q_\theta\left(\theta^{(i)}, \psi_{adapt}^{(0)}\right)$
  Generate a candidate individual $j \in \mathbf{E_1}$, then a candidate state $I_j' \sim$ Bernoulli(0.5) to propose $\mathbf{I}'$
  Compute the acceptance ratio:

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

  Sample $u \sim \mathcal{U}(0, 1)$
  if $u \leq \alpha$ **then**
    Accept the candidate state: $\theta^{(i+1)} \leftarrow \theta'$ and $\mathbf{I}^{(i+1)} \leftarrow \mathbf{I}'$
  else
    Reject the candidate state: $\theta^{(i+1)} \leftarrow \theta^{(i)}$ and $\mathbf{I}^{(i+1)} \leftarrow \mathbf{I}^{(i)}$
  end if
  Update $\psi_{adapt}^{(i+1)} \leftarrow \psi_{adapt}^{(i)}$
end for

3.4 Implementation

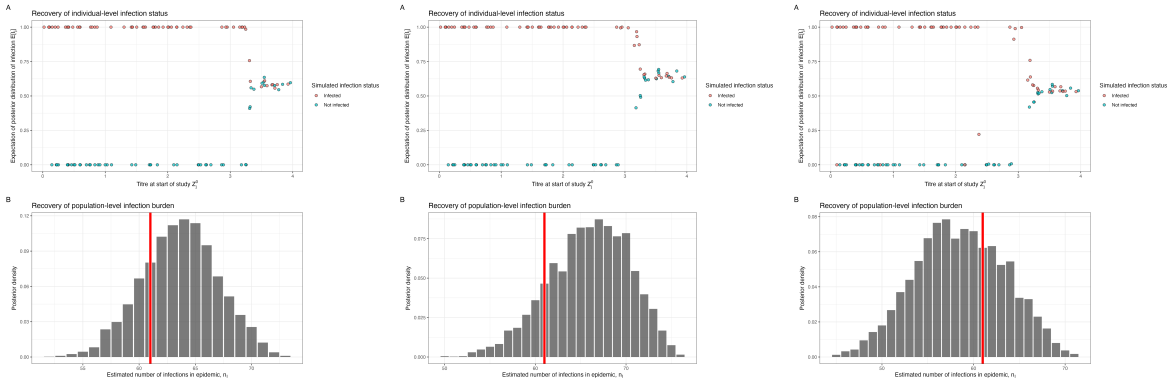
Algorithm 2 is coded manually in R and cpp through Rcpp. We run the algorithm for four chains, each with 200,000 steps, with 100,000 burn-ins steps. The initial values for θ and $\mathbf{I}$ are sampled from their prior distributions. We initialise the adaptive covariance by running with an identity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

3.5 Simulation recovery

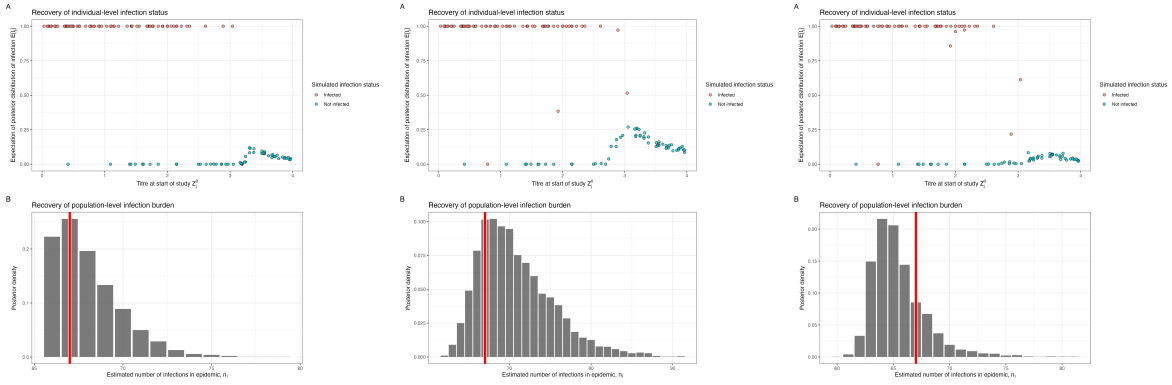
After running **Algorithm 2**, we plot the posterior samples, $\hat{\theta}$ and $\hat{\mathbf{I}}$ and compare with the simulated parameters.

3.5.1 Infection recovery

We assess the ability of the algorithm to recover the infection status of each individual in the study. If the set posterior samples of the infection status for individual j is given by $\hat{I}_j$, then we plot the expectation $\mathbb{E}[\hat{I}_j]$ so we can assess the ability of the algorithm to recover the individual-level simulated infection status (**Figure 4.3a**). Given no COP model A, we find when the pre-infection titre is less than 3.3 that all six models considered can recover the infection status of almost all individuals. When the pre-infection titre is greater than 3.3, the attenuation of boosting for infected individuals causes no meaningful change in the antibody kinetics ($f_{ab}^2(Z, \alpha) = 0$ when $Z > 3.3$). Thus, these individuals' infections are difficult to infer serologically as their titre dynamics are equivalent and independent of their infection status. In our COP model B, we find that including the correlate of protection influences the infection status. As the inferred correlate (**Figure 3.2**) has a low probability of infection at higher titres, this causes the $\mathbb{E}[\hat{I}_j]$ to be more likely to be 0 at higher titre values. Thus, the infection statuses for nearly all individuals are recoverable for COP model B.



(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error

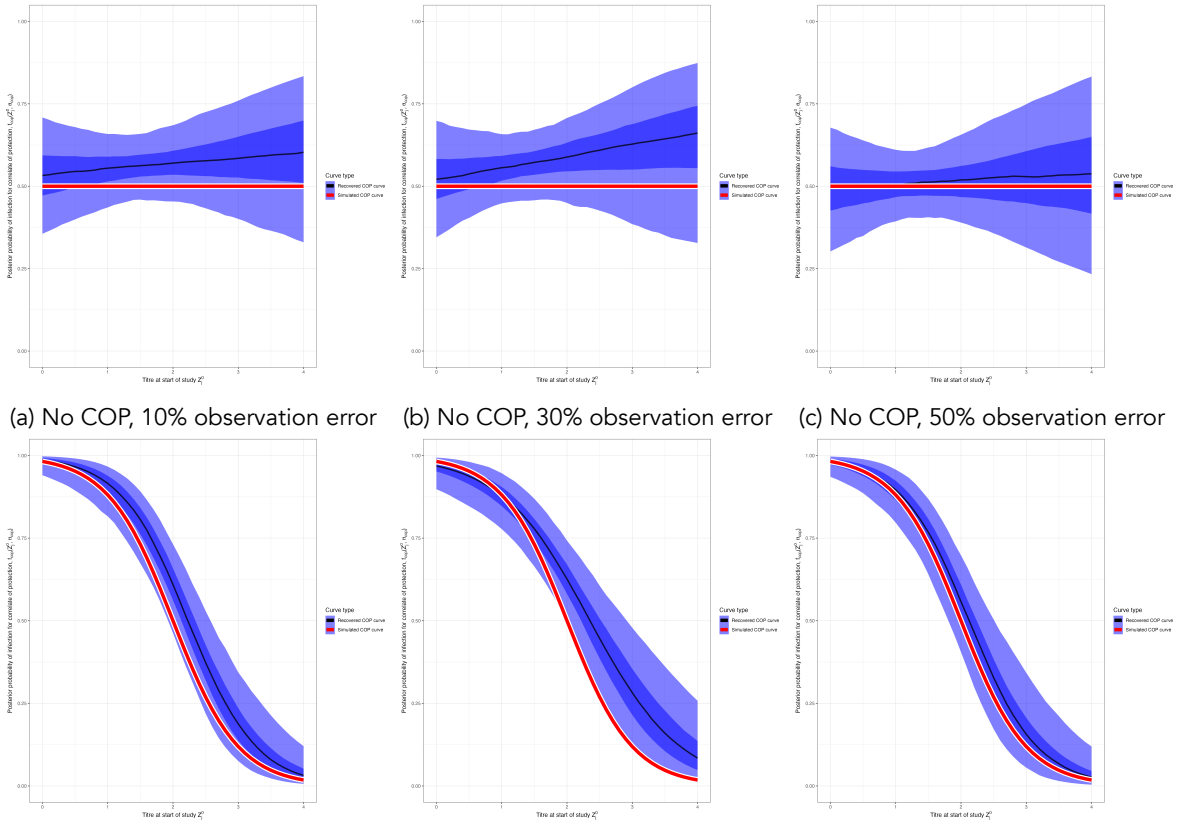


(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

Figure 3.1: Simulation recovery of the individual infection status, $\mathbb{E}[\hat{I}_j]$, 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 2** 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 mostly recovered, with the simulated line within a 95% confidence interval of the posterior sample (**Figure 3.2**). 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. To understand the difference between the simulated functional form in **Figure 3.2** in red, and the posterior samples, we have plotted the inferred COP from the simulated infection states in **Figure 2.3**. Here, it is clear that although we have a pre-defined function for the correlate of protection, when we simulate the data the COP curve is not perfectly recovered. Thus, the posterior distribution of our inference method can, at best, recover the inferred COP curve.



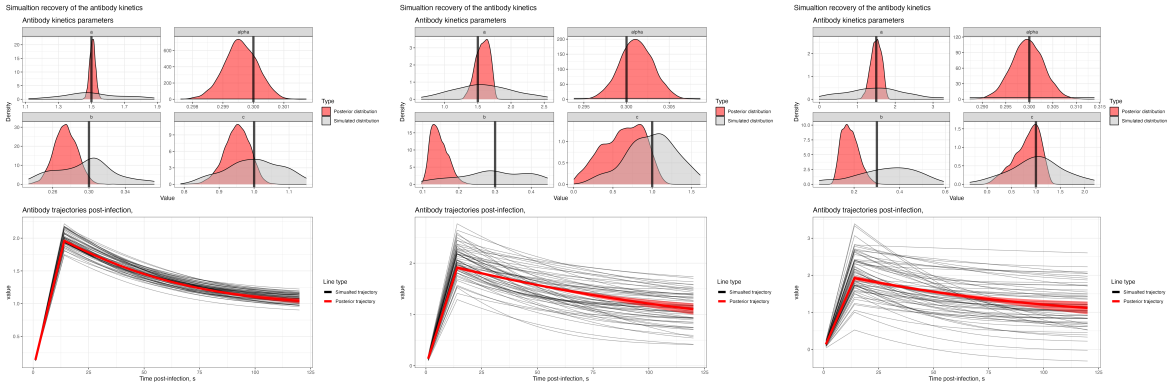
(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error

(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

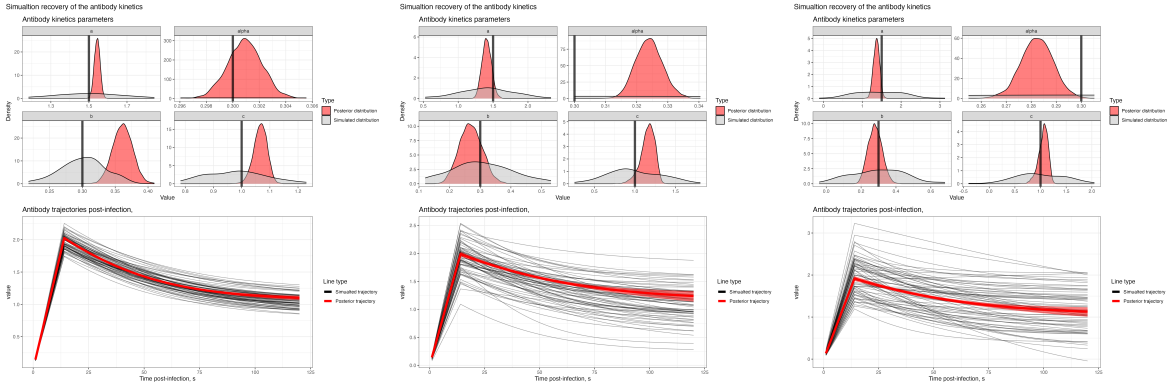
Figure 3.2: Simulation recovery of the COP function, with posterior samples plot $f_{cop}(x, \hat{\theta}_{cop})$. We have two different COP models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

3.5.3 Antibody kinetics

Algorithm 2 also succesfully recovers the simualted antibody kinetics. Let us plot $f^1_{ab}(s, \hat{a}, \hat{b}, \hat{c})$, the posterior predictive distribution for the antibody kinetic boosting, given posterior distributions $\hat{a}, \hat{b}$, and $\hat{c}$. At all three levels of kinetic uncertainty, the antibody kinetics are recovered, though increasing uncertainty weakens the accuracy of the recovered curves compared to the simulated. (**Figure 3.3**).



(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error



(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

Figure 3.3: Simulation recovery of the antibody kinetics function with posterior samples plot $f^1_{ab}(s, \hat{a}, \hat{b}, \hat{c})$. 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 2** 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. In this section, we will expand on this algorithm for the case when exposure status is unknown throughout the serosurvey.

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 dependencies:

$$L_{E_j=1}(Z_j|I_j, E_j^\tau, \theta) = \prod_{t \in T_j} P_{obs}(Z_{j,t}|X_{j,t}, \sigma) P_{cop}(I_j \mid E_j^\tau, \theta) \tag{4.1}$$

where $X_{j,t} = P_{ab}(I_j, E_j^\tau, \theta_{ab}, Z_i^0)$.

4.1.1 Likelihood and priors

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

$$\mathcal{L}(\mathbf{Z}|\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I}) = \prod_{j \in \mathbf{E_0}} L_{E_j=0}(Z_j|\theta) \prod_{j \in \mathbf{E_1}} L_{E_j=1}(Z_j|I_j, E_j^\tau, \theta) \tag{4.2}$$

We now must additionally define priors for $\pi(\mathbf{E})$ and $\pi(\mathbf{E}^\tau)$. Similar to $\pi(\mathbf{I})$ 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 $\pi(\mathbf{I})$ are as described in **Section 3**.

Consequently, we sample from the posterior distribution

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

If we use a **Algorithm 2** 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, as those who are exposed, have parameters I_j and E_j^τ to infer, whereas an individual who is not exposed has neither. Therefore, regardless of the proposal distribution we choose for inferring $\mathbf{E}$, we cannot use the existing algorithm highlighted in **Algorithm 2** as the detailed balance condition now fails.

4.2 The Reversible-Jump MCMC

The Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) algorithm[28] 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.

4.2.1 Mathematical overview

Let $\{k \in \mathcal{K}\}$ denote a collection models and θ_k be the parameter space of model k. A full Bayesian model for inferring k and θ_k can be written:

$$p(k)p(\theta_k|k)p(Z|k,\theta_k)$$

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 $p(Z|k,\theta_k)$ is the likelihood for the observed data for model k. We wish to build a Markov chain Monte Carlo algorithm to sample from the stationary distribution:

$$P(\theta_k, k|Z) \propto p(k)p(\theta_k|k)p(Z|\theta_k,k) \tag{4.4}$$

However, as the dimensions of vector θ_k change as we switch between models with different dimensions, there is no way obvious way to define Q and α such that the detailed balance condition (**Equation 3.8**) 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 is not converging to a single posterior distribution.

The RJ-MCMC proposes a solution to this issue [28]. 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 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

$$Q(x'|x) = Q\left((k', \theta_{k'})|(k, \theta_k)\right) = q_X(\theta_{k'}|\theta_k, k', k) \cdot q_k(k'|k) \tag{4.5}$$

where $q_k(k'|k)$ is the probability of selecting model k' from model k and q_X the probabiltiy 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, $\mathbf{u} = (u_1, \ldots, u_s) \sim q_1(\mathbf{u})$ and one of length s', $\mathbf{u'} = (u'_1, \ldots, u'_s) \sim q_2(\mathbf{u'})$ such that $d_{k'} + s' = d_k + s$. We then define a bijection, T

$$(\theta_{k'}, \mathbf{u'}) = T(\theta_k, \mathbf{u}) \tag{4.6}$$

to ensure the reversibility of the proposal distribution.

For the detailed balance condition to hold, Green[28] shows a prospoal distribution given by

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

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

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

$$\alpha\left(x, x'\right) = \min\left(1, \frac{P(x)q_k(k|k')q_2(\mathbf{u'})}{P(x')q_k(k'|k)q_1(\mathbf{u})} \cdot \left|\frac{\partial(\theta_{k'}, \mathbf{u'})}{\partial(\theta_k, \mathbf{u})}\right|\right) \tag{4.7}$$

ensures the stationary distribution chain samples:

$$P(\theta_k, k|Z) \propto p(k)p(\theta_k|k)p(Z|\theta_k,k) \tag{4.8}$$

A general form of the RJ-MCMC then follows **Algorithm 3**.

Algorithm 3 Reversible-Jump MCMC Algorithm

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

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

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

4.3 Application of RJ-MCMC to serological data

Algorithm 3 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 $\mathbf{E} = \{E_0, E_1, \ldots, E_{n_\mathbf{E}}\}$ 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 move (adding a new exposure), death move (remove an existing exposure), and parameter updating (exposure state remains the same)[32].

4.3.1 Birth move

For a given exposure vector $\mathbf{E}$, a birth move 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 move is selected at model k, $q_{birth}(k'|k)$ and the probability of choosing individual j uniformly from the non-exposed individuals:

$$q_k(k'|k) = q_{birth}(k'|k) \cdot \frac{1}{n_{\mathbf{E_0}}} \tag{4.9}$$

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

$$\prod_{t \in T} P_{obs}(Z_{j,t}|Z_j^0, \sigma) \rightarrow \prod_{t \in T} P_{obs}(Z_{j,t}|X_{j,t}, \sigma) P_{cop}(I_j \mid Z_j^0, \theta) P_t(E_j^t) \tag{4.10}$$

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^τ and their infection status I_j. In the notation of the previous section, it is convenient to define $\mathbf{u} = (E_j^\tau, I_j)$. 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^τ, we sample from the probability density function for the timing $P_t(\cdot)$, and for the infection status, we sample from $P_{cop}(\cdot|Z_j^0, \theta_{cop})$ (see **Equation 3.2**).

This sampling procedure results in a proposed sample which is in the proposed parameter space $\Theta_{k'}$: $(\theta_k, E_j^\tau, I_j) = \theta_{k'} \in \Theta_{k'}$. Consequently, we can choose the identify function for the required bijection T in **Equation 4.6**, 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 I_j and E_j^τ:

$$q_1(\mathbf{u}) = q_1(E_j^\tau, I_j) = P_{cop}(I_j|E_j^\tau, \theta_{cop})P_t(E_j^\tau) \tag{4.11}$$

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

$$q_k(k|k') = q_{birth}(k|k') \cdot \frac{1}{1 + n_{\mathbf{E}}} \tag{4.12}$$

In this move, we are removing the parameters E_j^τ and I_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 (**Equation 4.7**) for a birth move, where our current state is $k = \{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I}\} \to k' = \{\theta, \mathbf{E}', \mathbf{E}^{\tau\prime}, \mathbf{I}'\}$ is updated according to a uniformed sampled non-exposed individual j:

$$\alpha(k, k') = \min\left(\frac{P(\theta, \mathbf{E}', \mathbf{E}^{\tau\prime}, \mathbf{I}', |\mathbf{Z})q_{birth}(k|k')n_{\mathbf{E_0}}}{P(\theta, \mathbf{E}, \mathbf{E}^{\tau\prime}, \mathbf{I}, |\mathbf{Z})P_{cop}(I_j'|E_j', \theta_{cop})P_t(E_j^{\tau\prime})q_{birth}(k'|k)(n_{\mathbf{E}} + 1)}\right) \tag{4.13}$$

4.3.2 Death move

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

$$q_k(k'|k) = q_{death}(k'|k) \cdot \frac{1}{n_{\mathbf{E}}} \tag{4.14}$$

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

$$q_k(k|k') = q_{death}(k|k') \cdot \frac{1}{1 + n_{\mathbf{E_0}}} \tag{4.15}$$

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:

$$\prod_{t \in T} P_{obs}(Z_{j,t}|X_{j,t}, \sigma)P_{cop}(I_j \mid Z_j^0, \theta)P_t(E_j^\tau) \to \prod_{t \in T} P_{obs}(Z_{j,t}|Z_j^0, \sigma) \tag{4.16}$$

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 I_j. Therefore, by using the same argument as in the 'birth move' section, defining $q_1(\mathbf{u}) = 1$ and $q_2(\mathbf{u}') = F_t(E_j^\tau)F_{cop}(I_j|E_j^\tau, \theta_{cop})$, we can take the identity bijection as sample a value in state $\theta' \in \Theta'$ directly. The acceptance ratio (**Equation 4.7**) for a

death move, where our current state is $k = \{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I}\} \rightarrow k' = \{\theta, \mathbf{E}', \mathbf{E}^{\tau\prime}, \mathbf{I}'\}$ is updated according to a uniformed sampled exposed individual j:

$$\alpha(k, k') = \min\left(\frac{P(\theta, \mathbf{E}', \mathbf{E}^{\tau\prime}, \mathbf{I}, |\mathbf{Z})P_{cop}(I_j|E_j, \theta_{cop})P_t(E_j^\tau)q_{death}(k|k')n_{\mathbf{E}}}{P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I}, |\mathbf{Z})q_{death}(k'|k)(n_{\mathbf{E_0}} + 1)}\right) \tag{4.17}$$

4.3.3 Parameter updating

In this case, $\mathbf{E_i} = \mathbf{E}'$, that is the exposure vector remains unchanged, with probability $q_{par}(k|k)$. Therefore, the detailed balanced conditions are met without dimensional adjustment. In this case, we can use sample values for θ and $\mathcal{I}$ and use the acceptance ratio highlighted in Algorithm 3.

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 4.13** and **Equation 4.17**. If k is such that $n_{\mathbf{E}} = 0$, then select 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 4.13** and **Equation 4.17**) for all values of k' given k.

An algorithm describing the Birth-Death RJ-MCMC algorithm for this data is given in **Algorithm 4**.

Algorithm 4 Birth-Death Reversible Jump MCMC Algorithm

Chose a model k and initialize the chain with an initial states $\theta_k^{(0)}$, $\mathbf{E^{(0)}}$, $\mathbf{E}^{\tau,(\mathbf{0})}$ and $\mathbf{I^{(0)}}$. 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 $n_{\mathbf{E}} = M$, $p_{death} = 0.67, p_{par} = 0.33, p_{birth} = 0$.
for $i = 1$ to N **do**
  $u_1 \sim \mathcal{U}(0,1)$
  if $u_1 \leq p_{birth}$ **then**
    Birth move. Select $j' \in \mathbf{E_0^{(i)}}$, set $E_{j'} = 1$, sample $E_{j'}^{\tau} \sim P_t(\cdot)$, $I_{j'} \sim P_{cop}(\cdot|Z_{j'}^0, \theta_{cop})$ and update $\{\theta^{(i)}, \mathbf{E^{(i)}}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I^{(i)}}\} \rightarrow \{\theta', \mathbf{E'}, \mathbf{E}^{\tau\prime}, \mathbf{I'}\}$. Then calculate the acceptance probability

$$\alpha(k^{(i)}, k') = \min\left(\frac{P(\theta', \mathbf{E'}, \mathbf{E}^{\tau\prime}, \mathbf{I'}, |\mathbf{Z}) n_{\mathbf{E_0}}}{P(\theta, \mathbf{E^{(i)}}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I^{(i)}}, |\mathbf{Z}) P_{cop}(I_{j'}|E_{j'}, \theta_{cop}) P_t(E_{j'}^{\tau})(n_{\mathbf{E}}+1)}\right)$$

  else if $u_1 \leq (p_{birth} + p_{death})$ **then**
    Death move. Select $j' \in \mathbf{E_1^{(i)}}$, set $E_{j'} = 0$ and update $\{\theta^{(i)}, \mathbf{E^{(i)}}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I^{(i)}}\} \rightarrow \{\theta', \mathbf{E'}, \mathbf{E}^{\tau\prime}, \mathbf{I'}\}$. Then calculate the acceptance probability

$$\alpha(k^{(i)}, k') = \min\left(\frac{P(\theta', \mathbf{E'}, \mathbf{E}^{\tau\prime}, \mathbf{I}|\mathbf{Z}) P_{cop}(I_{j'}|E_{j'}, \theta_{cop}) P_t(E_{j'}^{\tau}) n_{\mathbf{E}}}{P(\theta, \mathbf{E^{(i)}}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I^{(i)}}, |\mathbf{Z})(n_{\mathbf{E_0}}+1)}\right)$$

  else
    Sample a candidate state $\theta' \sim q_\theta(\theta^{(i)}, \psi_{adapt}^{(i)})$
    Sample $j' \in \mathbf{E_1^{(i)}}$, and then a candidate state $I_j' \sim$ Bernoulli(0.5)
    Compute the acceptance ratio:

$$\alpha(k^{(i)}, k') = \min\left(1, \frac{P(\theta', \mathbf{E'}, \mathbf{E}^{\tau\prime}, \mathbf{I'}|\mathbf{Z})}{P(\theta^{(i)}, \mathbf{E^{(i)}}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I}^{(i)}|\mathbf{Z})}\right)$$

    Update $\psi_{adapt}^{(i+1)} \leftarrow \psi_{adapt}^{(i)}$
  end if
  Sample $u \sim \mathcal{U}(0,1)$
  if $u \leq \alpha$ **then**
    Accept the candidate state: Let $\{\theta^{(i+1)}, \mathbf{E^{(i+1)}}, \mathbf{E}^{\tau,(\mathbf{i+1})}, \mathbf{I^{(i+1)}}\} \leftarrow \{\theta', \mathbf{E'}, \mathbf{E}^{\tau\prime}, \mathbf{I'}\}$
  else
    Reject the candidate state: Let $\{\theta^{(i+1)}, \mathbf{E^{(i+1)}}, \mathbf{E}^{\tau,(\mathbf{i+1})}, \mathbf{I^{(i+1)}}\} \leftarrow \{\theta^{(i)}, \mathbf{E^{(i)}}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I^{(i)}}\}$
  end if
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. $\pi(\mathbf{I})$ cancels out in the parameter updating acceptance ratio (as described in **Algorithm 2**. 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 $\pi(\mathbf{I}) = 1/n_{\mathbf{E_1}} \neq 1/n_{\mathbf{E_1'}} = \pi(\mathbf{I'})$ 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: $P_t(E_j^{\tau}) = 1/120$.

4.3.4 Within Gibbs sampling of exposure times

Algorihtm 4 allows for efficient sampling of the θ, $\mathbf{I}$, and $\mathbf{E}$. However, the timing of the exposures, $\mathbf{E}^{\tau}$ is not efficiently explored as it can only be changed at a birth or death move. Therefore, it is desirable that values of $\mathbf{E}^{\tau}$ can be explored for fixed values of $\mathbf{I}$, and $\mathbf{E}$. To do this, we modify **Algorihtm 4** to allow the possibility of exploration of the $\mathbf{E}^{\tau}$ timings for a given model k whilst $\mathbf{I}$, and $\mathbf{E}$ remain fixed. To implement

this, after the candidate state has been accepted or rejected in **Algorithm 4**, we then resample a proportion of the $\mathbf{E}^{tau}$, and for each individual, we sample a new time $E_{j'}^{\tau}$ 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:

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

where α is the metropolis hasting ratio for $E_{j'}^{\tau}$ vs $E_j^{\tau,(i)}$.
Therefore the final RJMCMC algorithm (**Algorithm 5**) which effectively samples values from θ, $\mathbf{E}$, $\mathbf{E}_{\tau}$, $\mathbf{I}$.

Algorithm 5 Efficient Birth-Death Reversible Jump MCMC Algorithm

Chose a model k and initialise the chain with an initial states $\theta_k^{(0)}$, $\mathbf{E}^{(\mathbf{0})}$, $\mathbf{E}^{\tau,(\mathbf{0})}$ and $\mathbf{I}^{(\mathbf{0})}$.
for $i = 1$ to N **do**
  Update $\{\theta^{(i+1)}, \mathbf{E}^{(\mathbf{i+1})}, \mathbf{E}^{\tau,(\mathbf{i+1})}, \mathbf{I}^{(\mathbf{i+1})}\}$ according to **Algroithm 4**.
  for $k = 1$ to N_k **do**
    Select $j' \in N_{E=1}$ and resample from the proposal $E_{j'}^{\tau} \sim \mathcal{N}\left(E_j^{\tau,(i)}, \sigma_j^{(i)}\right)$ and update $\mathbf{E}^{\mathbf{t},(\mathbf{i})} \rightarrow \mathbf{E}^{\mathbf{t},*}$
    Compute the acceptance ratio:

$$\alpha(k^{(i)}, k') = \min\left(1, \frac{P(\theta^{(i)}, \mathbf{E}^{(\mathbf{i})}, \mathbf{E}^{\tau\prime}, \mathbf{I}^{(\mathbf{i})}|\mathbf{Z})}{P(\theta^{(i)}, \mathbf{E}^{(\mathbf{i})}, \mathbf{E}^{\tau,(\mathbf{i})}, \mathbf{I}^{(\mathbf{i})}|\mathbf{Z})}\right)$$

    Sample $u \sim \mathcal{U}(0,1)$
    if $u \leq \alpha$ **then**
      Accept the candidate state: $\mathbf{E}^{\tau,(\mathbf{i+1})} \leftarrow \mathbf{E}^{\tau\prime}$
    else
      Reject the candidate state:$\mathbf{E}^{\tau,(\mathbf{i+1})} \leftarrow \mathbf{E}^{\tau,(\mathbf{i})}$
    end if
    Update $\log\left(\sigma_{j'}^{(i_k+1)}\right) \leftarrow \log\left(\sigma_{j'}^{(i_k)}\right) + (1+i_k)^{-0.5}(\alpha - 0.44)$
    $i_{k+1} \leftarrow i_k + 1$
  end for
end for

4.3.5 Summary

Through **Algorithm 5** we define an efficient sampling procedure for the state space $k = \{\theta, \mathbf{E}, \mathbf{E}^{\tau}, \mathbf{I}\}$ 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.

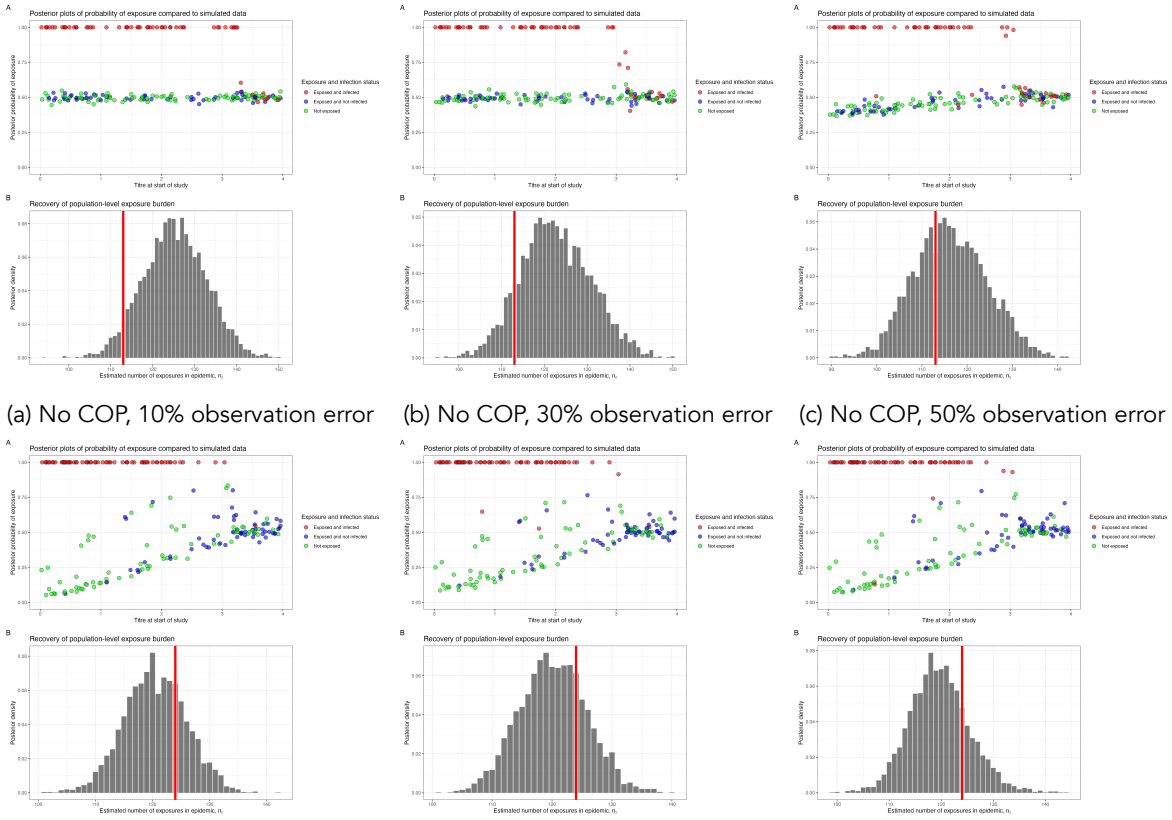
4.4 Simulation recovery

After running **Algorithm 5**, we plot the posterior samples, $\hat{\theta}$, $\hat{\mathbf{I}}$, $\hat{\mathbf{E}}$, and $\hat{\mathbf{E}^{\tau}}$ and compare with the simulated parameters.

4.4.1 Exposure state recovery

Algorithm 5 can predict and recover the population-level exposure rates, but struggles on the individual level (**Figure 4.2**). For those exposed and infected, the exposure rate is consistently recovered, except when the pre-infection titre is greater than 3.3, in which the boosting is completely attenuated. At these high titres, infection, exposed and not infected, and non-exposure all have the same antibody kinetics (i.e.

unchanged titre throughout the study), and thus, the model cannot differentiate between these exposure-infection states on an individual level. The expectation of the posterior values $\mathbb{E}[\hat{E}] = 0.5$, implying no preference between a positive or negative state. The model also cannot identify which individuals are exposed and not infected and which are not exposed for all titre values. This is unsurprising, as both these individuals have the same antibody kinetics regardless of pre-infection titre (i.e. unchanged throughout the study). In the case of a correlation of protection with a logistic function, there is some inference on the posterior probability of exposure $\hat{E}$ given pre-infection titre. At low titre values, it is unlikely an individual is exposed and not infected, as the probability of infection given exposure is high. Therefore, the model infers these individuals are unlikely to have been exposed $\mathbb{E}[\hat{E}] < 0.5$. This COP influence causes the increasing value of the $\mathbb{E}[\hat{E}]$ for each individual as the titre increases in **Figure 4.2**. Finally, the number of exposed individuals is approximately recovered for all six models.



(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error

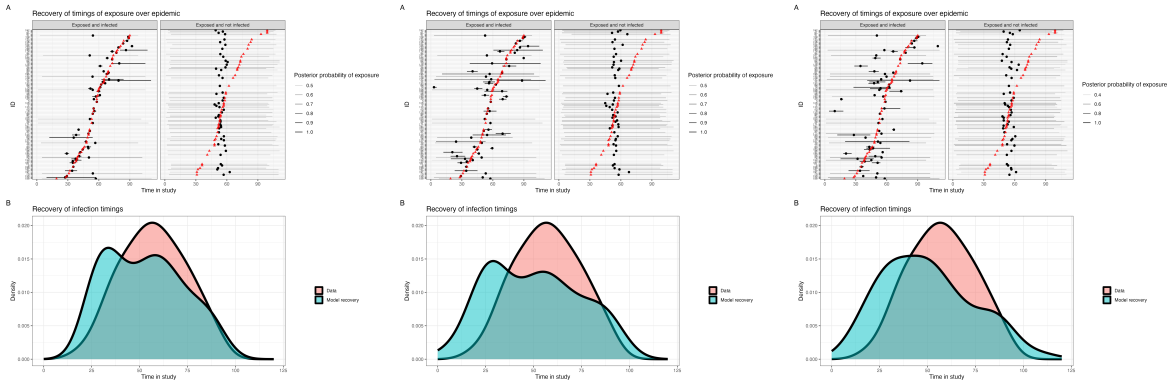
(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

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%)

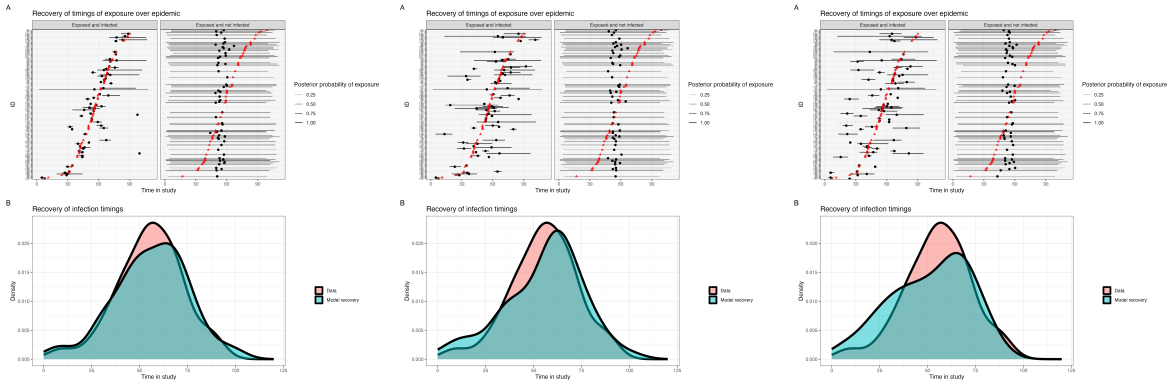
4.4.2 Exposure times recovery

Algorithm 5 can recover the exposure times under specific conditions. The posterior of the exposure times for individual j is given by $\hat{E}_j^\tau$ 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 those who are exposed and not infected, the model cannot infer the exposure time as there is an inference method for this in the likelihood (i.e. exposure time is determined by antibody kinetics, and antibody kinetics remain unchanged for those exposed and not infected). Therefore the inferred exposure time for this individual is the same as the prior entered into the model. For all six models, the epidemic curve is

reasonably recovered (**Figure 4.2**).



(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error

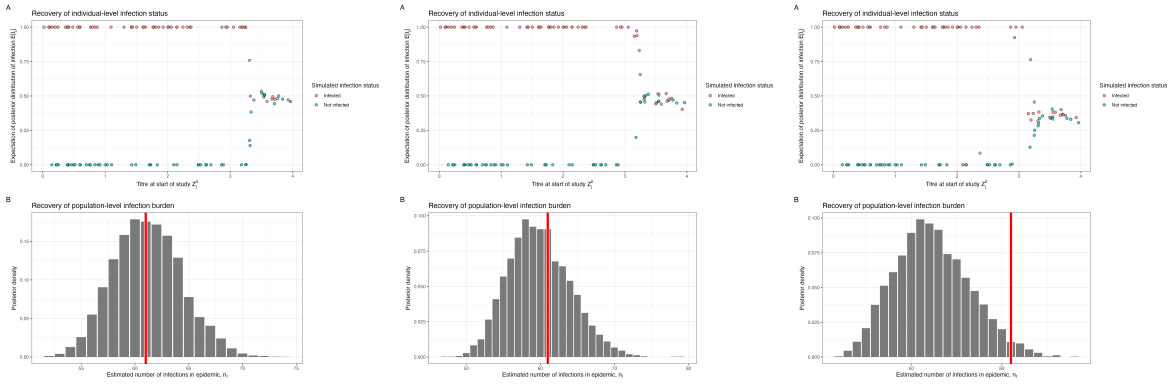


(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

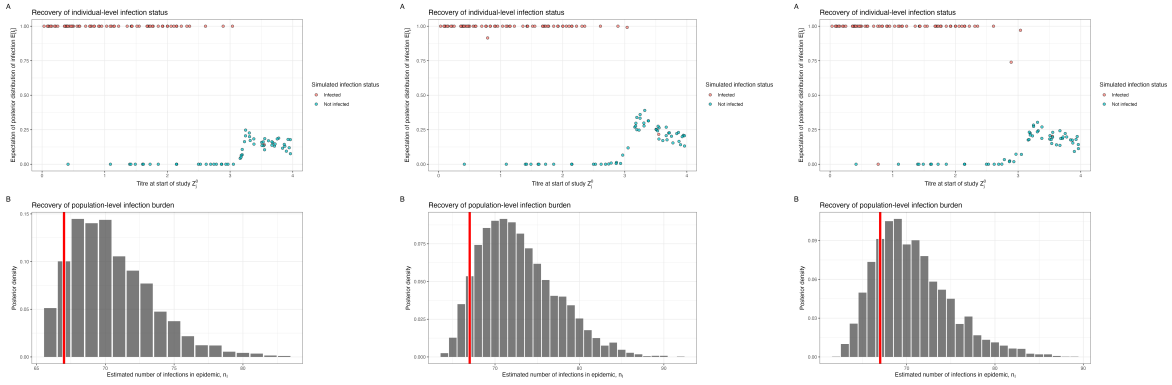
Figure 4.2: Simulation recovery of exposure and infection timtings $\hat{E^{\tau}}$ and epidemic curve for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%)

4.4.3 Infection state recovery

We also recover the infection status of each individual from the simulated data. If the set posterior samples of the infection status for individual j is given by $\hat{I}_j$, then we plot the expectation $\mathbb{E}(\hat{I}_j)$ so we can assess the ability of the algorithm to recover the individual-level simulated infection status (**Figure 4.3**). As before, we find when the pre-infection titre < 3.3 log titre value that all six models considered can recover the infection status of almost all individuals. When the pre-infection titre is greater than 3.3, the attenuation of boosting for infected individuals causes no meaningful change in the antibody kinetics ($f^2_{ab}(Z, \alpha) = 0$ when $Z > 3.3$). Thus, these individuals' infections are difficult to infer serologically as their titre dynamics are equivalent to independent of their infection status. In our COP model B, we find that including the correlation of protection influences the infection status. As the inferred correlate has a low probability of infection at higher titres, this causes the $\hat{I}_j$ to be more likely to be 0 at higher titre values.



(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error

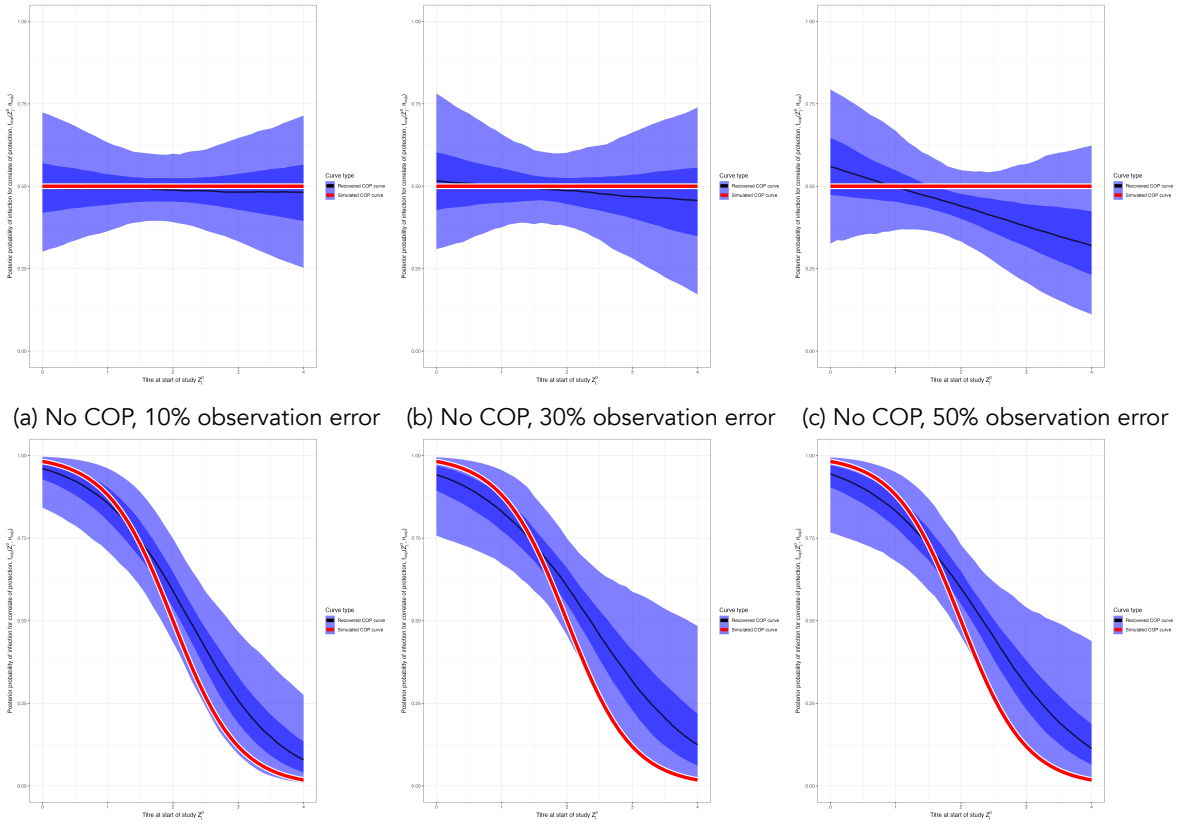


(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

Figure 4.3: Simulation recovery of the individual infection status, $\hat{I}_j$, for two COP models (top: No COP, bottom: logistic COP) and three different levels antibody kinetics variability (10%, 30%, 50%)

4.4.4 Correlate of protection

Algorithm 5 performs well at]ecovering 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 4.4**). For Model B, we find the logistic shape of the COP is recovered in the posterior samples.



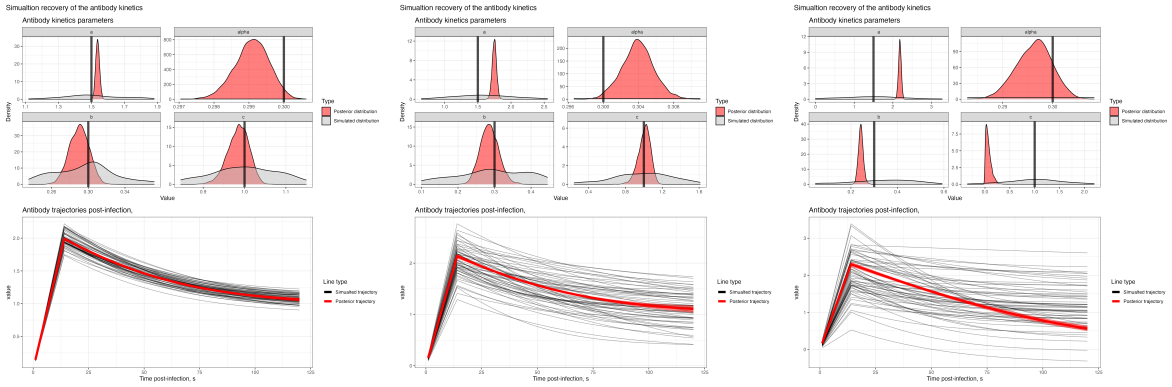
(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error

(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

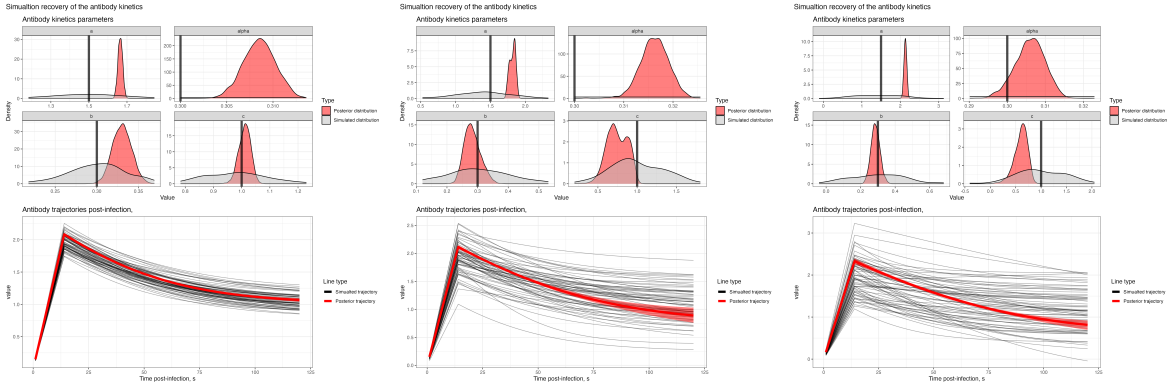
Figure 4.4: Simulation recovery of the COP function, posterior samples plot $f_{cop}(x, \hat{\theta}_{cop})$. We have two different COp models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

4.4.5 Antibody kinetics

Algorithm 5 also succesfully recovers the simualted antibody kinetics. Let us plot $f^1_{ab}(s, \hat{a}, \hat{b}, \hat{c})$, the posterior predictive distribution for the antibody kinetic boosting, given posterior distributions for $\hat{a}, \hat{b}$, and $\hat{c}$. At all three levels of kinetic uncertainty, the antibody kinetics are recovered, though increasing uncertainty weakens the accuracy of the recovered curves compared to the simulated. (**Figure 4.5**).



(a) No COP, 10% observation error (b) No COP, 30% observation error (c) No COP, 50% observation error



(d) COP, 10% observation error (e) COP, 30% observation error (f) COP, 50% observation error

Figure 4.5: Simulation recovery of the antibody kinetics function with posterior samples plot $f^1_{ab}(s, \hat{a}, \hat{b}, \hat{c})$. We have two different COP models (top: No COP, bottom: logistic COP) and three different levels of antibody kinetics variability (10%, 30%, 50%).

5 Looking forward

We have shown the capacity of **Algorithm 5** to recover the state variables $\{\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I}\}$ from simulated serological data.[27] We have shown that the correlation of protection and antibody kinetics functions are robustly recovered for all six simulated datasets. The model cannot infer individual-level exposure status for those not infected. However, this is unsurprising as the antibody kinetics between these groups in the framework here are equivalent. Despite this, the proportion of the population who is exposed and infected is well recovered across all datasets. At high levels of variability in the antibody kinetics, the ability of the model to infer the exposure time on the individual level weakens, and the epidemic curves start to differ from the simulated curve.

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 studies 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 printing 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 to improve inference. 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 life history of infections to be inferred given immunological titre landscapes, similar to the techniques in serosolver[29]. 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. Combining this mcmc algorithm will, however, be mathematically complex.

This document has provided details of the theoretical underpinning and implementation of an RJMCMC algorithm, which can infer important epidemiology and immunological information from individual-level serological data. It can recover the exposure status, exposure times, and infection status, given broad structural forms for antibody kinetics and the correlation of protection. 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 areas of health science which require the exploration of multidimensional model spaces.

Appendices

A Notation

Symbol *State variables*	**Description**
$Z_{j,t} \in \mathbf{Z}$	Data on the antibody titre at time t for individual j.
Z_j^0	Initial titre (titre at first value of t for individual j).
$X_{j,t} \in \mathbf{X}$	Model estimated antibody titre at time t for individual j.
$\mathbf{E} = \{E_1, \ldots, E_j, \ldots, E_M\}$	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.
$\mathbf{E_1} = \{j_1, \ldots, j_{j^*} \ldots, j_{n_{\mathbf{E}}}\}$	Vector of individuals j^* who are exposed. A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{E_0} = \{j_1, \ldots, j_{j^*} \ldots, j_{M-n_{\mathbf{E}}}\}$	Vector of individuals j^* who are not exposed. A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{E}^\tau = \{E^\tau_{j_1}, \ldots, E^\tau_{j_{j^*}}, \ldots, E^\tau_{j_{n_{\mathbf{E}}}}\}$	Vector of exposure times for each individual j. A superscript (i) specifies a specific value in the Markov chain.
$\mathbf{I} = \{I_1, \ldots, I_{j^*}, \ldots, I_{n_{\mathbf{E}}}\}$	Vector of infection statuses (binary vector) for each individual j. A superscript (i) specifies a specific value in the Markov chain.
$n_{\mathbf{I}}$	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
$\theta_{ab} = \{a, b, c, \alpha\}$	Fitted parameters for the antibody kinetics model
$\theta = \{\theta_{cop}, \theta_{ab}, \sigma\}$ *Functions*	All fitted parameters in the model. A superscript (i) specifies a specific value in the Markov chain.
$P(\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I}\|\mathbf{Z})$	Posterior distribution function given inputs θ, **E**, $\mathbf{E}^\tau$, **I** and data **Z**
$\mathcal{L}(\mathbf{Z}\|\theta, \mathbf{E}, \mathbf{E}^\tau, \mathbf{I})$	Likelihood function given inputs θ, **E**, $\mathbf{E}^\tau$, **I** and data **Z**.
$\mathcal{L}_{E_j=0}(Z_j, \|I_j, E_j^\tau, E_j, \theta)$	Likelihood function for individual j who is not exposed
$\mathcal{L}_{E_j=1}(Z_j, \|I_j, E_j^\tau, E_j, \theta)$	Likelihood function for individual j who is exposed
$P_t(E_j^\tau \| E_j)$	Likelihood of an exposure at time E_j^τ given individual j is exposed
$X_{j,t} = F_{ab}(t, I_i, E_j^\tau, \theta_{ab}, Z_j^0)$	Deterministic function for the estimated antibody titre at time t for individual j and starting titre values from the data, Z_j^0 for an exposure at time E_j^τ and infection status I_j.
$f^1_{ab}(s, a, b, c)$	The function which determines the antibody titres at time s after E_j^τ
$f^2_{ab}(Z_j^0, \alpha)$	The function which scales the trajectory given pre-titre Z_j^0.
$P_{cop}(I_j \| Z_j^0, \theta_{cop})$	Likelihood for the correlate of protection for an individual j with an exposure at time E_j^τ, and estimated titre value X_{j,E_j^τ} and infection status I_j.
$f_{cop}(Z_j^0, \beta_0, \beta_1)$	Function describing the correlate of protection for infection at time t. (logistic).
$P_{obs}(Z_{j,t} \| X_{j,t}, \sigma)$	Likelihood of the observation model for the data $Z_{j,t}$ given model-estimated titre values $X_{j,t}$ for individual j at time t.
$\pi(\theta) = \pi(a)\pi(b)\ldots\pi(\sigma)\pi(\mathbf{I})\pi(\mathbf{E})$	Prior distributions for all fitted parameters in the model.

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

Symbol	Description
N	Length of chain in metropolois hasting algorihtm
$q_\theta\left(\cdot\|\theta^{(i)}\right)$	Proposal distribution for the values of θ at given state i.
$q_I\left(\cdot\|I_j^{(i)}\right)$	Proposal distribution for the values of I_j for individual j at given state i (See XX).
$q_k\left(\cdot\|k\right)$	Proposal distribution for a new model given model is at $\mathcal{M}_k$ (See XX).
$q_\tau\left(\cdot\|E_j^{\tau,(i)}\right)$	Proposal distribution for a new time of exposure for individual j for a given state i.(See XX).

Table A.2: Symbols used in the mcmc algorithms

B Adaptive Proposal Distribution

I use an adaptive proposal distribution $q_\theta(\theta)$ to sample the parameter space θ. The adaptive metropolis hasting algorithm provides systematic method for modifying the shape of the proposal distribution based on the accepted steps of the current markov chain, allowing for 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.

Problematically, the proposal distribution using the updated covariance matrix, $\Sigma^{(i)}$, is no longer memoryless, and therefore 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 is given by

i	$i \leq T_{init}$	$i > T_{init}$		
$q(\cdot\|\theta^{(i)})$	$\mathcal{N}(\theta^{(i)}, \exp(\lambda^{(i)})I_s; \mathcal{S})$	$\mathcal{N}(\theta^{(i)}, \Sigma^{(i)}; \mathcal{S})$ $\mathcal{N}(\theta^{(i)}, \exp(\lambda^{(T_{init})})I_s; \mathcal{S})$	with probability β, with probability $1-\beta$	(B.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:

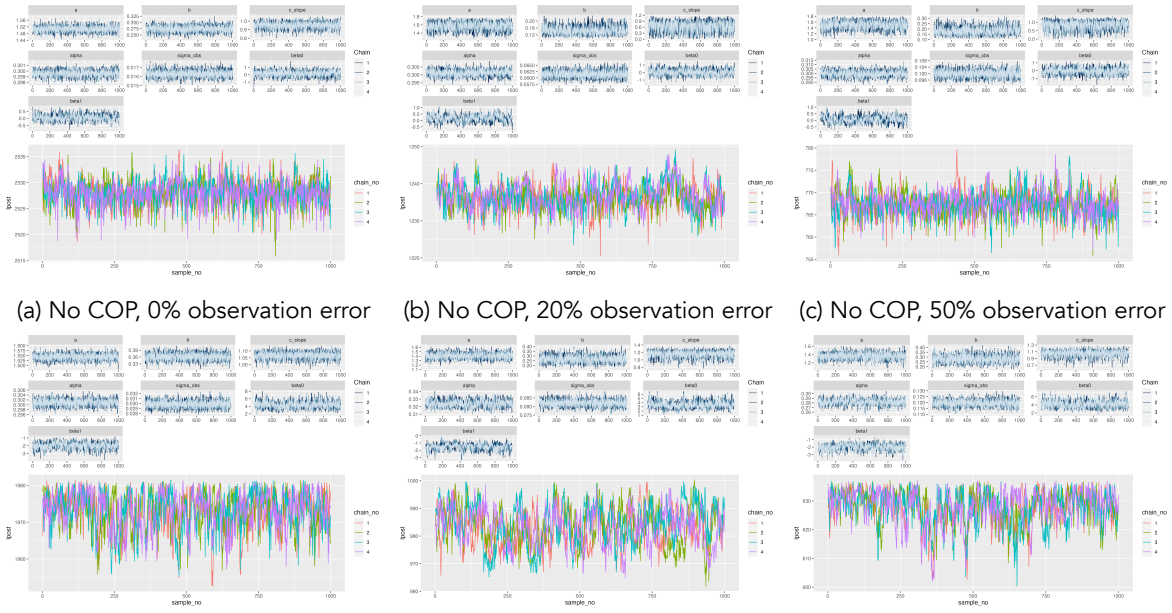
$$\begin{aligned}
\lambda^{(i+1)} &= \lambda^{(i)} + \gamma_1(i)(a(\theta^{(i)}, \theta') - 0.234) \\
M^{(i+1)} &= M^{(i)} + \gamma_2(i)(a(\theta^{(i)}, \theta') - 0.234) \\
\mu^{(i+1)} &= \mu^{(i)} + \gamma_3(i)(\mu^{(i)} - \theta^{(i)}) \\
\Gamma^{(i+1)} &= \Gamma^{(i)} + \gamma_4(i)[(\theta^{(i)}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

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

C Trace plots

C.1 Known expsosure

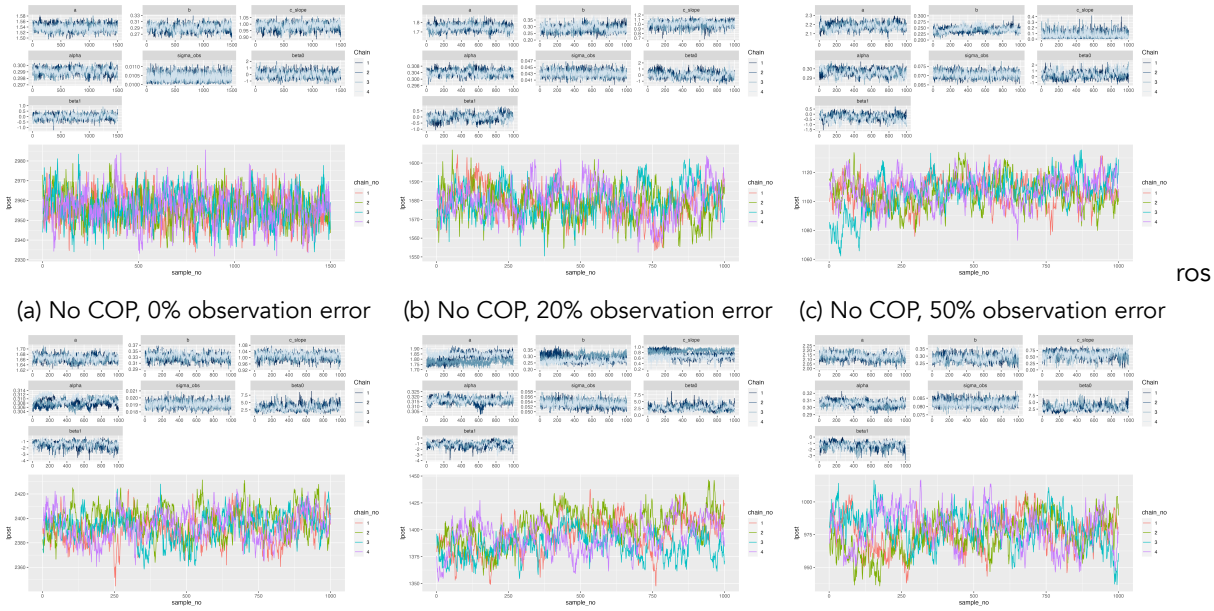


(a) No COP, 0% observation error (b) No COP, 20% observation error (c) No COP, 50% observation error

(d) COP, 0% observation error (e) COP, 20% observation error (f) COP, 50% observation error

Figure C.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 (0, 20%, 50%)

C.2 Inferred expsosure



(a) No COP, 0% observation error (b) No COP, 20% observation error (c) No COP, 50% observation error

ros

(d) COP, 0% observation error (e) COP, 20% observation error (f) COP, 50% observation error

Figure C.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 (0, 20%, 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 Drebot. 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, Dixion 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 Kudesia, 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, Binzhi 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 Eijkeren, 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] Stanley A Plotkin. Recent updates on correlates of vaccine-induced protection. *Front. Immunol.*, 13:1081107, 2022.

[20] 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.

[21] 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.

[22] Amrita Sekhar and Gagandeep Kang. Human challenge trials in vaccine development. *Semin. Immunol.*, 50:101429, August 2020.

[23] 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.

[24] 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.

[25] 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.

[26] Joshua T Schiffer. Correlates of protection via modeling. *Nat Comput Sci*, 2(3):140–141, March 2022.

[27] 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.

[28] Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, December 1995.

[29] 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.

[30] 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.

[31] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive MCMC. *Stat. Comput.*, 18:343–373, 2008.

[32] Faming Liang, Chuanhai Liu, and Raymond Carroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley, August 2010.