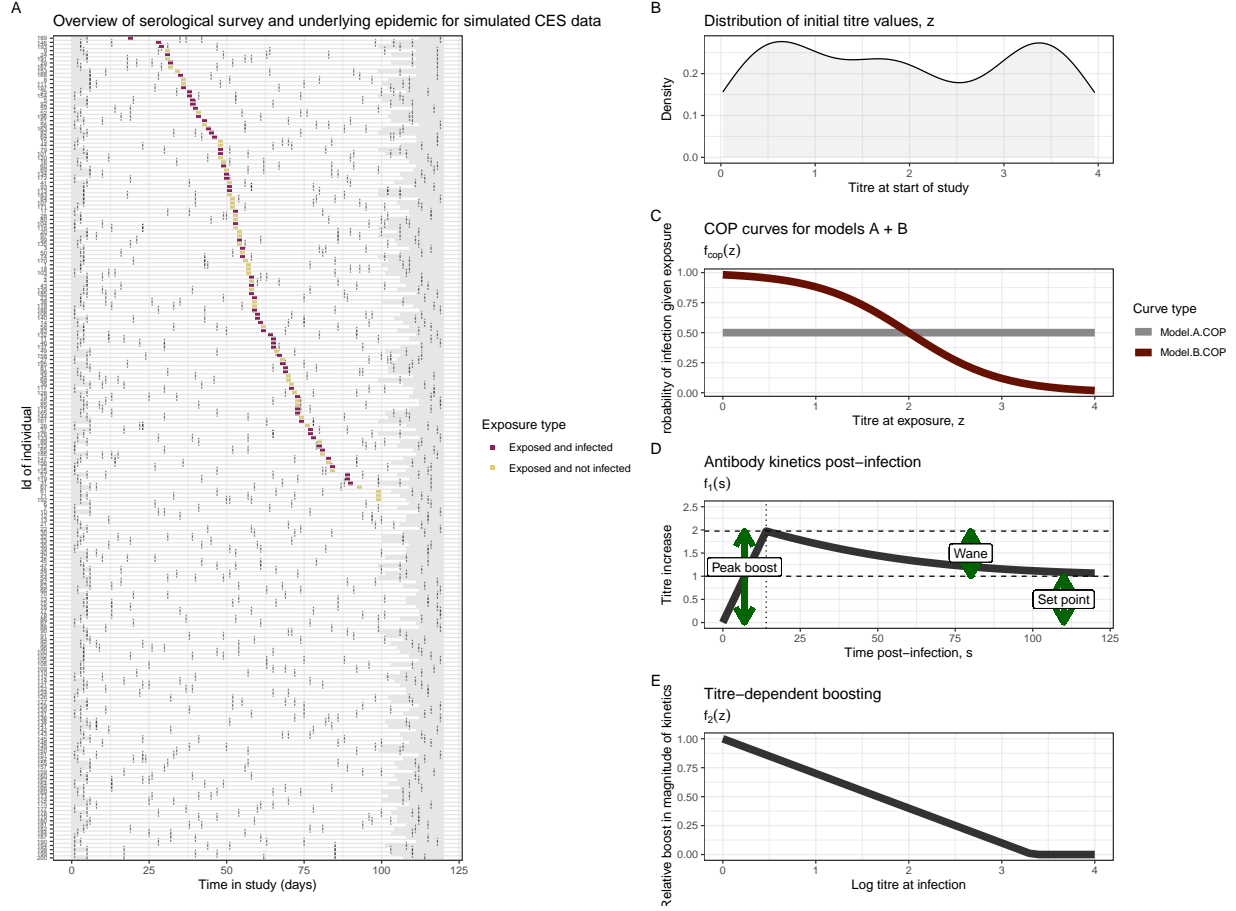# 1 Simulated serological data from `serosim`[1]

Figure 1: Schematics showing the simulated data structure from `serosim`[1]

We simulate serological data using the `serosim`[1] 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 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:

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

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 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.[2] The formula for this biphasic trajectory is given by **Equation 2**.

$$f_{ab}^1(s, a, b, c) = \begin{cases} \ln(\exp(a) + \exp(b))/14, & \text{if } s \le 14 \\ \ln(\exp(a)\exp(-(b/10)(t-14)) + \exp(c)), & \text{if } s > 14 \end{cases} \tag{2}$$

where $a = 1.5$, $b = 2$, and $c = 1$ are values in the simulated data (**Figure 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 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;

$$f_{ab}(s, Z_j^0, a, b, c, \alpha) = f_{ab}^1(s, a, b, c) f_{ab}^2(Z_j^0, \alpha) \tag{3}$$

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**. 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 3**. A summary of the parameter values and functions used to generate the simulated data can be found in **Table 1**.

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 3**
f_{cop}	Logistic correlate of protection	**Equation 1**

Table 1: Table of parameters associated with the simulated data and their values
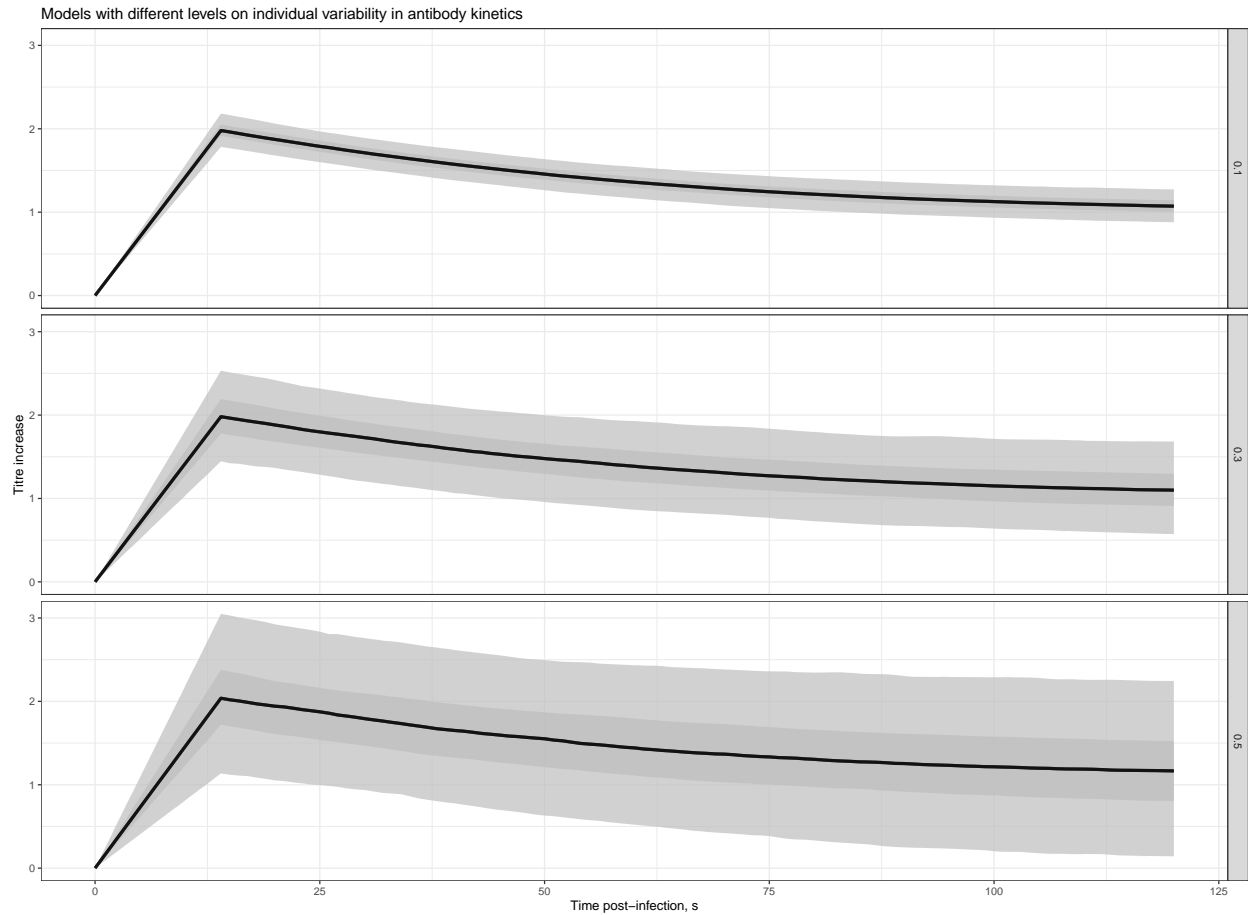
Figure 2: Schematics showing three levels of individual-level uncertainty and the impact on the variability of antibody kinetics.

References

[1] 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.

[2] 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.

A

Correlate of protection in simulated data for Model A (No COP)

B

Correlate of protection in simulated data for Model B (Logistic COP)
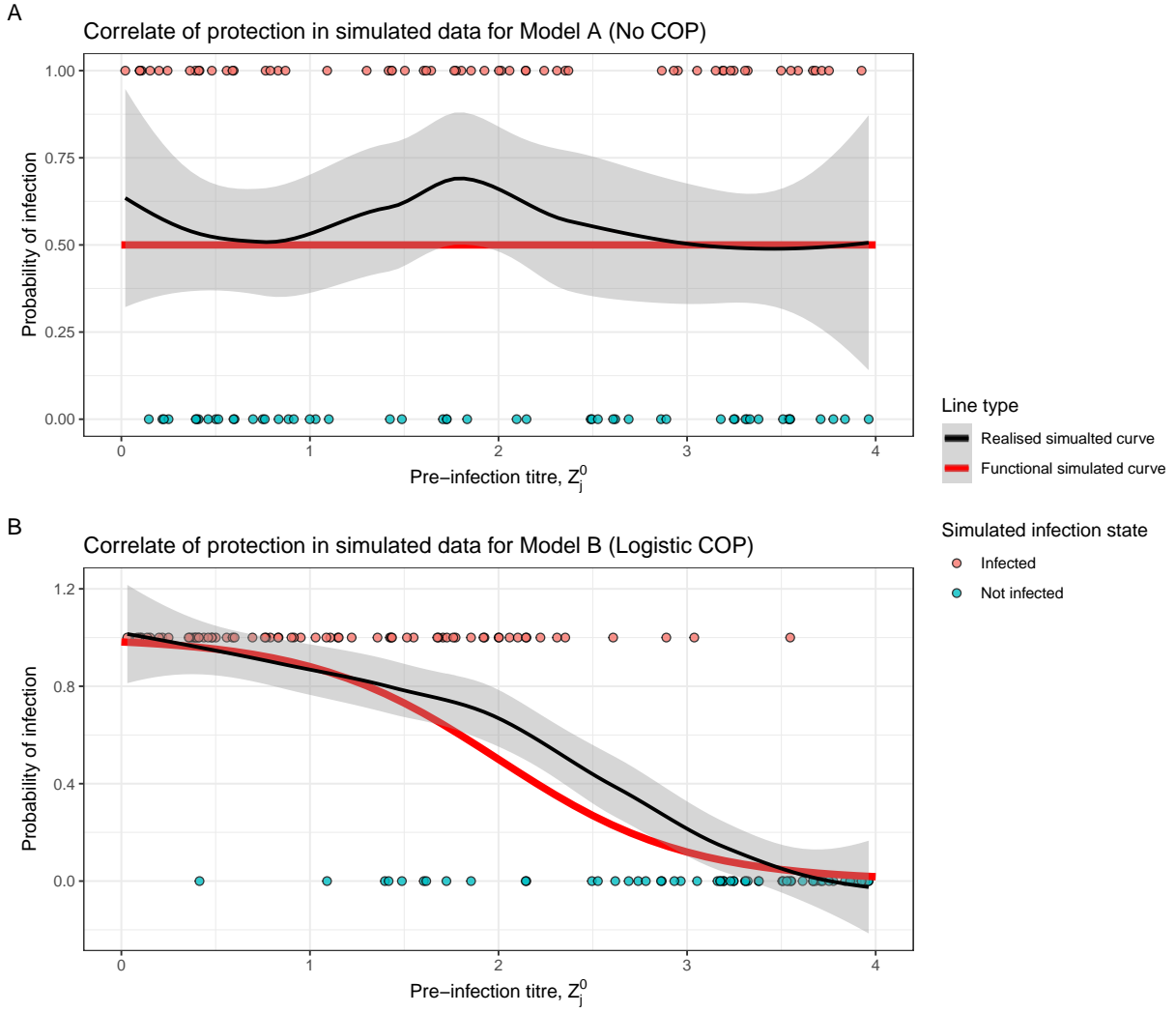

Figure 3: Schematics showing the difference between the functional form chosen to simulate the COP and the recovered COP from the exposed individuals.