

1 Looking forward

We have shown the capacity of **Algorithm ??** to recover the state variables $\{\theta, \mathbf{E}, \mathbf{E}^T, \mathbf{I}\}$ from simulated serological data.[?] 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 whole 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 COP curves.

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[?]. 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 RJMCMC 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 very 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 where intense surveillance is challenging. We also hope this document sheds light on a mathematically complex but powerful inferring tool and encourages others to implement similar algorithms in other areas of health science which require the exploration of multidimensional model spaces.

References