Leveraging mathematical models and household studies: improving inference to gain insights into pathogen transmission

Drawing on the experience of the COVID-19 pandemic, this thesis aims to explore how mathematical modelling can be used effectively to analyse data collected in households to better understand the transmission of airborne pathogens.

A stochastic individual-based model of airborne pathogen transmission was developed and coupled with an inference model to estimate transmission parameters from household data. This inference framework, extension of existing models, notably Layan et al. 2020, allows to quantify infectiousness and susceptibility across age groups, symptoms, partial immunity and control measures applied in the household.

First, these methods were specifically applied to data from the PedCovid study (PI: Isabelle Sermet-Gaudelus), in which households were included through the recruitment of children who tested positive for SARS-CoV-2. A simulation study was conducted to assess the impact of this recruitment protocol on the ability of the classic inference framework to estimate parameters. It revealed a systematic bias in the estimates of age-related parameters. A new inference model correcting the age-related recruitment bias was formulated and validated. It was then applied to data from the PedCovid study to assess the impact of age on the transmission of the Alpha and Omicron variants of SARS-CoV-2. We show that with the Alpha variant, children under 6 years old were less infectious than adolescents and adults, which was not the case for children aged 6–11. With the Omicron variant, no significant difference in infectiousness or susceptibility was observed according to age.

Second, the simulation study was extended to explore the impact of different aspects of study design on the estimation of infectiousness and susceptibility across age groups and symptoms. Synthetic datasets were simulated to reproduce studies with different recruitment methods, sample sizes, inclusion criteria, and follow-up formats. The model was also generalised to different pathogens according to their incubation period, generation time, and proportion of asymptomatic cases.

The analysis first explored the impact of different protocols on the Alpha variant of SARS-CoV-2, before being generalised to potential future emerging pathogens with different characteristics. The results show that numerous recruitments (age-specific, symptoms-specific) can bias estimates when the classic model is used. They also show how the power of the inference framework depends on the sample size, but also on the inclusion criteria, and that depending on the pathogens studied, the spacing of visits impacts the quality of the estimates.

This thesis highlights the impact of the design of longitudinal household studies on their analysis using mathematical modelling. Taking this design into account in models will be crucial in order to improve the response to a future epidemic risk, whether from a known or an emerging pathogen.