

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**Modelling the role of immunity, climate and behaviour
in viral outbreak dynamics and control**

Emilie Finch

Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy

of the
University of London

July 2024

Department of Infectious Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Medical Research Council

Grant number: MR/No13638/1

Research group affiliation: Centre for Mathematical Modelling of Infectious Diseases

Declaration

Statement of own work

I, Emilie Finch, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, this has been indicated in the thesis.

Emilie Finch

July 2024

Acronyms

AOR	Adjusted odds ratio
CFR	Case fatality ratio
CI	Confidence interval
CrI	Credible interval
COVAX	COVID-19 Vaccines Global Access
COVID-19	Coronavirus disease
CRPS	Continuous ranked probability score
CRPSS	Continuous ranked probability skill score
DENV	Dengue virus
DIC	Deviance information criterion
ELISA	Enzyme-linked immunosorbent assay
ENSO	El Niño Southern Oscillation
EWS	Early warning system
FOI	Force of infection
GAM	Generalised additive model
GAMM	Generalised additive mixed model
GISAID	Global Initiative on Sharing All Influenza Data
GLM	Generalised linear model
GLMM	Generalised linear mixed model
HMC	Hamiltonian Monte Carlo
ICU	Intensive care unit
IFR	Infection fatality ratio
IgA	Immunoglobulin class A
IgG	Immunoglobulin class G
IgM	Immunoglobulin class M
INLA	Integrated nested Laplace approximation
IVM	Integrated vector management
MCMC	Markov chain Monte Carlo

NAAT Nucleic Acid Amplification Test
NEA National Environment Agency of Singapore
NOAA National Oceanic and Atmospheric Administration
NS1 Nonstructural protein 1
PRNT Plaque reduction neutralization test
RNA Ribonucleic acid
ROC Receiver operating characteristic
RT PCR Reverse transcription polymerase chain reaction
SAGE Scientific Advisory Group for Emergencies
SARS Severe acute respiratory syndrome
SARS-CoV-2 Severe acute respiratory syndrome coronavirus 2
SEIR Susceptible Exposed Infected Recovered
SIR Susceptible Infected Recovered
SINAVE Sistema Nacional de Vigilancia Epidemiológica
SPI Standardised Precipitation Index
SSTA Sea surface temperature anomaly
TSCV Time series cross-validation
VOC Variant of Concern
WAIC Widely applicable information criterion
WHO World Health Organization
WIC Weighted interval score

Abstract

Understanding the drivers of viral outbreaks is crucial for assessing future epidemic potential and options for outbreak control. However, disentangling the role of intrinsic drivers, such as population immunity, from extrinsic drivers, such as climatic variation or human behaviour, is challenging, particularly in complex settings with heterogenous immune landscapes. In this thesis, I integrate serological and climate data streams within modelling frameworks to disentangle the impact of immune, climatic and behavioural drivers of acute viral outbreaks, and consider the implications for disease control. I focus on SARS-CoV-2 and dengue virus as case studies, two pressing health threats with complex epidemic dynamics and the ability to cause large outbreaks with the potential to overwhelm health systems.

Through analysis of a seroepidemiological workplace cohort in the United States from April 2020 – February 2021, I found that primary infection with SARS-CoV-2 provided protection against reinfection for most individuals. I then used a multi-strain, age-stratified compartmental model to estimate the impact of immunity, population behaviour and vaccination on SARS-CoV-2 transmission in the Dominican Republic from 2020 – 2022. By jointly fitting to serological and surveillance data, I found that epidemic dynamics were largely driven by the accumulation of post-infection immunity but that, despite this, vaccination was essential in enabling a return to pre-pandemic behaviour without incurring considerable additional morbidity and mortality.

Next, I used a Bayesian hierarchical mixed model to quantify the effects of immunity and climate on dengue dynamics in Singapore from 2000 – 2023, with a view to improving early warning for outbreaks. I found non-linear and delayed impacts of climatic variation on dengue risk, with increased risk at intermediate temperature and rainfall levels, during El Niño events, and in the period following a switch in dominant serotype. I then adapted this model into an early warning framework, generating dengue forecasts at 2-8 week lead times. Accounting for serotype dynamics as a proxy for immunity, as well as climatic variation, improved the predictive power of the forecasting model, and particularly the prediction of outbreaks. Finally, I used a Bayesian space-

time hierarchical model, fit to surveillance data from 2013 – 2023 in the 155 municipalities of the Dominican Republic, to estimate the role of climatic and epidemic drivers in spatiotemporal dengue dynamics. I leveraged serological data to estimate a proxy for the build-up of immunity in a dengue season and then accounted for this, as well as autocorrelation in case counts from the force of infection, within the modelling framework. I found evidence of increased dengue risk at higher maximum temperatures and humidity, as well as in drought or El Niño conditions. I found that El Niño and drought indicators are influential predictors of temporal dengue dynamics, while lagged cases, weighted to represent the force of infection, predict both spatial and temporal patterns.

Overall, in this thesis, I consider immune drivers alongside important extrinsic drivers within modelling frameworks to better characterize acute viral transmission dynamics in complex epidemiological settings. I found that, by including multiple data streams within modelling frameworks, I was able to distinguish between epidemiological hypotheses underlying disease transmission and improve prediction of outbreak risk to inform epidemic response.

Acknowledgments

First and foremost, I would like to take this opportunity to thank my supervisors Prof. Adam Kucharski and Prof. Rachel Lowe, for all their guidance during this PhD. I'm deeply grateful for your insight and advice, which have made me a better scientist, and for your encouragement and support in developing my career, even in the midst of a global pandemic.

I would like to thank Dr. Eric Nilles and Prof. Colleen Lau, as well as their respective teams working on the Project Meridian serosurvey. Thank you for inviting me to help pilot the survey in the Dominican Republic - my time there was an invaluable experience, and I'm grateful to have been a part of such an inspirational collaboration. I would like to extend my gratitude to Cecilia Then Paulino and Dr. Ronald Skewes-Ramm for their generosity in hosting me at the Dirección de Epidemiología during my visits. This work wouldn't have been possible without the field teams in the Dominican Republic and all those who participated in the serosurvey, les agradezco. I would also like to thank Dr. Lee-Ching Ng and Dr. Shuzhen Sim in the National Environment Agency of Singapore, for sharing your expertise with me. Thank you to the MRC-LID for funding this work, as well as the conferences, research visits and placements that have so enriched the past few years.

Thank you to Dr. Anton Camacho for supervising me during my placement in Epicentre, MSF, from whom I learnt a great deal about how to make modelling have real world impact, as well as how to order lunch in French. Thanks also to Jonathon Mellor and the Infectious Disease Modelling group at the UK Health Security Agency - I've loved working with such an impactful and collaborative team.

I'm grateful to have been part of the Centre for Mathematical Modelling of Infectious Diseases, and to have met so many supportive and inspirational colleagues. Thank you to everyone who makes it such a brilliant place to work, from seminars and conferences to overly competitive Christmas quizzes and in-depth discussions at the pub. Thanks especially to all those who have

given their time to discuss my research over the past few years, and to Dr. Ellie Rees and Dr. Alexis Robert for doing so much to welcome me to LSHTM in a very strange time.

I feel lucky to have made incredible friends during my PhD, who have helped me through the challenges and made coming into the office such a joy. To David, Billy, Greg, Ciara M, Ciara J, Marleen and Kirsty - thank you for the pints, the tea breaks, the notes on the wall, the dancing, the swimming, for all the advice, and for being wonderful scientists and wonderful people.

Thank you to Immy, whatever souls are made of etc., and to Jorgen for your singularly joyful outlook on life. Thank you to Hannah for all the great times in the flats we've shared.

To my school friends, who are always there, and who I am so grateful to have met. To Leo, Alex and Will - science is lucky to have you, and so am I. To Amy and Freya, for your unwavering confidence in me and for so many years of love and friendship. I wouldn't choose to be in lockdown with anyone else.

To my family in Bath, for the wonderful chaos and the loving support. To my Mum, for her belief in me. To Nat, for everything. And to my Dad, whose passion, wit and kindness inspire me always, and without whom I couldn't have done any of this.

This thesis is dedicated to my Nana, Brenda Vicary-Finch, who did not get to see me complete it but is, I think, the reason I thought it possible to try.

Contents

Introduction.....	1
1.1 Background.....	1
1.2 SARS-CoV-2	1
1.2.1 Epidemiology and transmission.....	1
1.2.2 Clinical manifestation.....	2
1.2.3 Diagnosis and serology.....	3
1.2.4 Control measures	4
1.2.5 SARS-CoV-2 in the Dominican Republic.....	5
1.3 Dengue virus.....	7
1.3.1 Epidemiology and transmission.....	7
1.3.2 Clinical manifestation.....	7
1.3.3 Diagnosis and serology.....	8
1.3.4 Control measures	8
1.3.5 Climate and dengue	10
1.3.6 Dengue in Singapore and the Dominican Republic.....	11
1.4 Aim and objectives of thesis.....	12
1.5 Thesis structure.....	13
References.....	15
Methods.....	28
2.1 Study sites.....	28
2.1.1 Seroepidemiological cohort in the United States.....	28
2.1.2 Seroprevalence survey in the Dominican Republic.....	28
2.1.3 Dengue surveillance in Singapore	30
2.2 Modelling methods	30
2.2.1 Mechanistic modelling.....	31
2.2.3 Statistical modelling	34
2.2.4 Applications of modelling to COVID-19 and dengue virus	36
2.2.5 Model fitting	38
References.....	41
Investigating immunity to SARS-CoV-2 in a community seroepidemiological cohort in the United States.....	47
3.1 Abstract.....	51
3.2 Introduction.....	51

3.3 Results.....	52
3.3.1 SARS-CoV-2 infection and reinfection.....	53
3.4 Discussion.....	57
3.5 Methods	58
3.5.1 Seroepidemiological cohort description	58
3.5.2 Ethics Statement	59
3.5.3 Statistical Analysis.....	59
References.....	65
Effects of mobility, immunity and vaccination on SARS-CoV-2 transmission in the Dominican Republic: a modelling study	68
4.1 Abstract.....	72
4.2 Introduction.....	73
4.3 Methods	75
4.3.1 Data.....	75
4.3.2 Transmission dynamic model.....	78
4.3.3 Vaccination parameters	79
4.3.4 Mobility	80
4.3.5 Model fitting	81
4.3.6 Counterfactual analysis.....	81
4.4 Results.....	82
4.4.1 COVID-19 transmission dynamics between 2020 – 2022	82
4.4.2 Impact of vaccination campaign	86
4.4.3 Trade-off between vaccination and population mobility.....	88
4.5 Discussion.....	90
References.....	95
Disentangling the role of climate and serotype competition to forecast dengue outbreaks in Singapore	99
5.1 Abstract.....	103
5.2 Introduction.....	104
5.3 Results.....	107
5.3.1 Reported dengue cases in Singapore over the past two decades	107
5.3.2 Climate and serotype dynamics shape dengue risk with non-linear and delayed effects	109
5.3.3 Accounting for serotype and climate dynamics improves probabilistic predictions of dengue case incidence and outbreak detection.....	112
5.3.4 Dengue forecasting for early warning with 2-8 weeks lead time	115
5.4 Discussion.....	119
5.5 Methods	123

5.5.1 Data.....	123
5.5.2 Model framework	124
5.5.3 Model selection.....	125
5.5.4 Model evaluation	126
5.5.5 Adapting the model for early warning.....	128
References.....	130
Understanding the role of climate and epidemic drivers in spatiotemporal dengue outbreak dynamics in the Dominican Republic.....	135
6.1 Abstract.....	139
6.2 Introduction.....	140
6.3 Methods	142
6.3.1 Study setting	142
6.3.2 Surveillance data.....	142
6.3.3 Population data	142
6.3.4 Climate data	143
6.3.5 Seroprevalence study 2021	144
6.3.6 Serocatalytic model	144
6.3.7 Bayesian hierarchical model framework	145
6.4 Results.....	152
6.4.1 Dengue epidemiology in the Dominican Republic between 2013 and 2023.....	152
6.4.2 Estimating the annual force of infection.....	154
6.4.3 Quantifying the impact of climate variation on dengue risk	156
6.4.4 Out-of-sample dengue predictions.....	159
6.5 Discussion.....	162
References.....	167
Discussion	173
7.1 Summary of findings	173
7.1 Strengths of this work.....	177
7.1.1 Integrating multiple data streams.....	177
7.1.2 Representative data	179
7.1.3 Reproducibility	180
7.2 Limitations	181
7.2.1 Data.....	181
7.2.2 Model assumptions	183
7.2.3 Generalisability.....	185
7.3 Implications and future work.....	185
7.4 Concluding remarks.....	187
References.....	188

Ethical Approval	193
Supplementary material for Chapter 4.....	205
Supplementary material for Chapter 5.....	228
Supplementary material for Chapter 6.....	241

List of Figures

Figure 1.1 Schematic of viral and antibody kinetics for SARS-CoV-2.....	4
Figure 3.1 PCR and serological tests in seroepidemiological cohort	53
Figure 3.2 PCR positivity, seropositivity and reinfection risk.....	55
Figure 3.3 Unadjusted and adjusted odds of reinfection.....	57
Figure 3.4 Propensity to seek a PCR test between the seronegative and seropositive groups	61
Figure 4.1 Map of the study setting and time-series of COVID-19 cases.	75
Figure 4.2 Google mobility data, SARS-CoV-2 sequence data, and estimated vaccination coverage by age.....	77
Figure 4.3 Model schematic showing a two-variant SEIR model structure with a vaccination compartment (V).....	78
Figure 4.4 Comparison of model fit to observed data in the Dominican Republic from February 2020 to December 2021	83
Figure 4.5 Immune status, contact rates and reproduction number estimates from January 2020 to December 2021	85
Figure 4.6 Impact of vaccination campaign.....	87
Figure 4.7 Modelled total additional hospital admissions, ICU admissions and deaths in the 6 months following the vaccination campaign launch (16th February 2021 – 16th August 2021). 90	
Figure 5.1 Dengue cases, climate variability and serotype dominance in Singapore.....	108
Figure 5.2 Effects of climatic variability and switches in dominant serotype on dengue incidence in Singapore	111
Figure 5.3 Comparing time series cross-validated predictions of candidate models.....	114
Figure 5.4 Comparing outbreak detection of candidate models.	115
Figure 5.5 Dengue forecasts for early warning at 4 and 8 week forecast horizons.	117
Figure 5.6 Predictive performance over different forecast horizons.	118
Figure 6.1: Dengue epidemiology in the Dominican Republic	153
Figure 6.2: DENV seropositivity by age and mean age of dengue cases	155
Figure 6.3: Effects of climate on dengue risk in the Dominican Republic.....	158

Figure 6.4: Time-series cross validated posterior predictions of dengue cases from 2015 – 2023
..... 160

Figure 6.5: Influence of climatic and epidemic drivers on dengue predictions..... 162

List of Tables

Table 3.1: Estimated total hospital admissions, ICU admissions and deaths under different counterfactual vaccination scenarios.	88
--	----

Chapter 1

Introduction

1.1 Background

The transmission dynamics of acute viral infections are complex and challenging to predict as they are driven by an interplay between intrinsic properties of pathogen biology and the development of host immunity, and extrinsic factors such as climate or changes in population behaviour [1,2]. Nonlinearity is often a key feature of viral outbreak dynamics; for instance, where the build-up of immunity in a population reduces later transmission, or where contact patterns driving transmission change in response to the outbreak itself.

Towards the end of the 20th century, epidemiological data provided to the WHO demonstrated that the burden of disease attributable to infectious diseases was decreasing, due to improved medical treatment, vaccination and sanitation [3]. However, large outbreaks of SARS, pandemic influenza, dengue, Ebola, SARS-CoV-2 and mpox have demonstrated the ongoing threat to human health posed by emerging viruses [3,4]. In recent decades, mechanistic and statistical models have been used to investigate the dynamics of acute viral outbreaks, and to inform public health decision making around epidemic response [5–11]. Modelling can be used to support multiple aspects of public health decision making. First, modelling can enable the estimation of key epidemiological parameters. Second, scenario modelling can be used to understand the potential impact of different public health interventions, or, similarly, retrospective counterfactual analysis can assess the impact of past interventions, informing pandemic preparedness strategies. Third, models can be used to generate short-term forecasts (one or two weeks ahead) or longer-term seasonal forecasts (several months ahead). These can aid outbreak response by improving situational awareness or informing early warning systems linked to anticipatory response activities.

Given the complexity of viral outbreaks, accurately characterising dynamics requires the disentangling of multiple processes driving transmission. As such, modelling frameworks are increasingly incorporating multiple data streams to reflect the multifaceted nature of viral disease systems, capitalising on novel data streams available in the ‘big data’ era [12,13]. This introduces challenges in how to harmonize different data sources, which may be at different spatiotemporal resolutions and have complex underlying mechanistic relationships, and how to compare a model to these data to inform parameterisation and fitting [12]. However, considering multiple data streams simultaneously within a single modelling framework can allow us to distinguish between different epidemiological hypotheses for viral outbreak risk, and avoid incorrectly attributing variation in risk to a single factor (say a climatic variable, or the introduction of a new serotype) when it is actually the result of multiple important drivers working in combination [12,14,15].

This thesis focuses on the challenge of disentangling key drivers of viral transmission, particularly focusing on the role of immunity, climate and behaviour, by integrating multiple data streams into models that can inform public health decision making. In this thesis I investigate SARS-CoV-2 and dengue virus as case studies; two pressing challenges to global health due to their complex epidemic dynamics and their ability to cause large outbreaks with the potential to overwhelm health systems. I leverage data streams from the USA, the Dominican Republic and Singapore and explore the role of immunity and extrinsic drivers such as climate and behaviour in complex epidemiological settings.

1.2 SARS-CoV-2

1.2.1 Epidemiology and transmission

SARS-CoV-2, a novel coronavirus, was first identified in January 2020 in Wuhan, China, following the identification of a cluster of pneumonia cases of unknown aetiology in 2019 [16–18]. Since its emergence, SARS-CoV-2 has spread to every continent and led to substantial morbidity and mortality, particularly in settings where health-care systems were overwhelmed. The WHO declared SARS-CoV-2 a public health emergency of international concern on 30th

January 2020. This remained in place until 5th May 2023, by which point over 750 million cases had been reported and nearly 7 million deaths [19–21].

SARS-CoV-2 is a respiratory virus transmitted primarily through direct contact or via respiratory droplets or aerosols [22–24]. Early analysis estimated a reproduction number (that is, the average number of secondary infections generated by an infected individual) of around 2-3 [25–27]. SARS-CoV-2 transmission is overdispersed, with around 10% of individuals thought to cause 80% of secondary transmission [28].

In September 2020, a new variant (Alpha or B.1.1.7) was first sequenced in the United Kingdom. Alpha was designated a variant of concern (VOC) by the World Health Organization in February 2021 and estimated to be 43-90% more transmissible than the original lineage as well as displaying increased disease severity [29,30]. The Gamma variant (or P.1) was identified in Brazil in early 2021 and was estimated to be 2.4-fold more transmissible and exhibited immune evasion, whereby past infection was less protective against Gamma infection compared with previous variants [31]. The Delta variant (B.1.617.2) became the dominant strain globally in mid-2021 and was again estimated to be more transmissible than prior variants [32]. The Omicron variant (B.1.1.529) was reported in late 2021 and subsequently spread to global dominance, displaying high levels of immune evasion to both vaccine and post-infection immunity [33–35].

1.2.2 Clinical manifestation

Infection with SARS-CoV-2 is most often asymptomatic or causes mild symptoms including fever, dry cough, shortness of breath or loss of taste or smell. In some individuals infection can lead to severe disease requiring hospitalisation or ICU admission, which can progress to critical disease and death. Susceptibility to infection and subsequent clinical severity are both age-dependent, with clinical symptoms estimated to manifest in around 20% of 10-19 year olds and around 70% of over 70 year olds [15,36,37]. The infection-fatality ratio (risk of death per infection) increases with age, following a log-linear relationship [38]. As such, early estimated infection-fatality ratios differ substantially between countries, depending on their age structure; for instance, overall IFR was estimated to be around 1.41% in England compared with 0.24% in Kenya [38,39].

1.2.3 Diagnosis and serology

Viral testing for COVID-19 diagnosis can be performed using RT-PCR (reverse transcriptase polymerase chain reaction), which amplifies and quantifies viral RNA from a nasopharyngeal sample and is considered the gold standard for confirming current infection. Viral testing can also be done using rapid antigen testing, which detects SARS-CoV-2 antigen and has lower sensitivity than PCR testing [40]. Studies have estimated that over 95% of symptomatic COVID-19 cases develop detectable immunoglobulin G (IgG), immunoglobulin M (IgM) and immunoglobulin A (IgA) antibodies against SARS-CoV-2, with most individuals developing antibodies 1-3 weeks after symptom onset. Similarly to other viral infections, IgM and IgA antibodies wane faster than IgG antibodies [41–43]. Serological assays measuring antibody levels can be used to determine if someone has experienced past infection or vaccination. Assays can target antibodies against specific SARS-CoV-2 antigens such as the spike (anti-S) or nucleocapsid (anti-N) glycoproteins. While post-infection immunity to COVID-19 results in ‘broad’ immunity to SARS-CoV-2 antigens, several key COVID-19 vaccines were developed using the S protein (such as Pfizer/BioNTech, Oxford/AstraZeneca or Moderna) and therefore different assays can give information on an individual’s immune history. For instance, an individual with anti-S antibodies but not anti-N antibodies is likely to have been vaccinated but not previously infected, while an individual with both anti-S antibodies and anti-N antibodies is likely to have been previously infected, and may or may not have been vaccinated [44–46]. It should be noted this is not the case for inactivated virus vaccines, such as Sinovac-Coronavac, which result in both anti-S and anti-N responses [47]. In this context, seroepidemiological studies measuring antibody responses in a sample of the population have been used throughout the pandemic to understand SARS-CoV-2 transmission and immunity [48].

A schematic of viral and antibody kinetics relevant for PCR and serological testing is shown in Figure 1.1. Studies have estimated a median incubation period (time between exposure and symptom onset) of around 5 days, ranging from around 2 to 14 days [49,50]. Initial estimates of the generation interval (the time interval between an infector and infectee being infected) were around 5 days [49]. However, evidence suggests that the incubation period and generation interval vary for different variants [52–54].

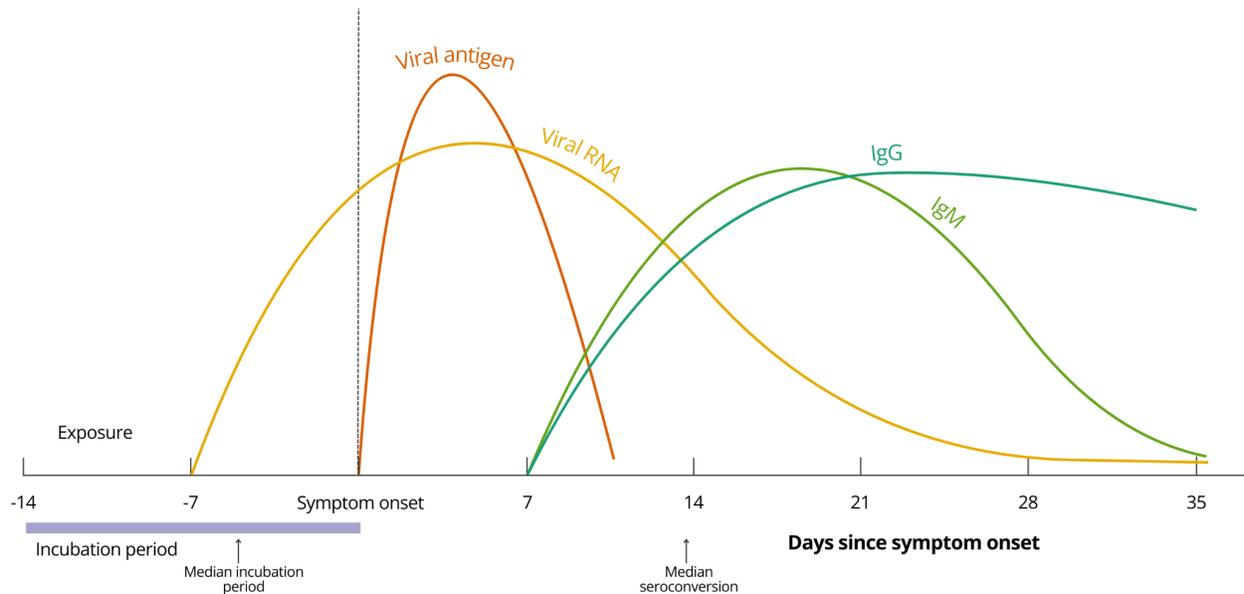


Figure 1.1 Schematic of viral and antibody kinetics for SARS-CoV-2

Lines showing levels of viral RNA (yellow line), viral antigen (orange line), immunoglobulin M, IgM (green line) and immunoglobulin G, IgG (turquoise line) following exposure and symptom onset [55]. Median incubation period of ~ 5 days is indicated with an arrow, while the maximum incubation period of ~ 14 days is shown with a light purple bar [50]. Median seroconversion of ~13 days is shown with an arrow [43].

1.2.4 Control measures

Before the development of vaccines or antivirals, the global response to COVID-19 was focused on non-pharmaceutical interventions [56]. Throughout the pandemic, countries differed in their approach to response, with some aiming to completely control transmission (with ‘zero covid’ policies) and others aiming to protect the most vulnerable and mitigate the impact of the pandemic, for instance by ensuring that hospital capacity was not overwhelmed. As with many other infectious disease outbreaks, contact tracing and case isolation are key interventions aiming to break chains of transmission. Other non-pharmaceutical interventions rely on physical distancing measures aiming to keep a distance of 1-2m between individuals to reduce transmission. Physical

distancing measures range from cancellation of large public events, closure of schools and workplaces and stay-at-home recommendations or restrictions, including curfews or lockdowns [57]. Many countries also mandated the use of face coverings when in public places. Finally, travel restrictions (and post-travel quarantines) were imposed by many countries during the pandemic based on the emergence of variants, incidence or domestic COVID-19 strategy.

By late 2020, the first COVID-19 vaccine was approved for use by the WHO and several countries, and the first vaccination programme was started in the United Kingdom [58]. Several vaccines such as Pfizer BNT 162b2 vaccine, AstraZeneca ChAdOx1 and Moderna mRNA-1273 were developed around the same time, and were found to be safe and effective against symptomatic disease and more severe outcomes [44–46]. By 5th May 2023, over 5.5 million people are thought to have received a COVID-19 vaccination [21]. However, despite international initiatives for equitable vaccine sharing, vaccine availability and rollout was highly inequitable globally, with low and middle income countries (LMICs) often receiving delayed or limited vaccine doses, or vaccines with lower efficacy profiles [59,60].

1.2.5 SARS-CoV-2 in the Dominican Republic

The Dominican Republic is a small-island developing state in the Caribbean, on the eastern side of Hispaniola, which it shares with Haiti. It is classified as an upper-middle income country by the World Bank, with a population of 10.8 million [61]. The Dominican Republic reported their first case of COVID-19 on 1st March 2020. On 17th March a state of emergency was declared, implementing border closures, and closures of schools and commercial businesses. Subsequently, a mandatory curfew (“toque de queda”) was introduced on 27th March. The Dominican Republic, like many countries in Latin America and the Caribbean, experienced a high burden of SARS-CoV-2 in 2020, with hospital capacity nearly reached or exceeded at several points [62,63]. Non-pharmaceutical measures began to be relaxed in July 2021 and were mostly lifted by October 2021 with the reopening of schools and relaxation of curfew measures.

A community-based seroprevalence survey was undertaken in the Dominican Republic between May – June 2020, targeting emerging hotspots of transmission. Researchers used a

chromatographic test targeting the S1 domain of the spike protein with reported IgG sensitivity and specificity of 56% and 100% respectively. The study found overall seropositivity for anti-SARS-CoV-2 IgG of 5.5%, varying from 18.6% (Duarte) to 1.8% (San Jose de Ocoa) in surveyed regions [64]. While these findings were based on a sampling approach that targeted known hotspots, and are not representative of the general population, they suggest there was high heterogeneity in seroprevalence in May-June 2020 and indicate that substantial immunity may have accrued in some provinces by this point.

Subsequently, a nationally representative seroprevalence study was conducted in the Dominican Republic between June and October 2021 [65]. This study estimated that 85.0% (95% CI 82.1 – 88.0) of the ≥ 5 years population had been immunologically exposed and 77.5% (95% CI 71.3 – 83) had been previously infected. Data from this survey are used in several chapters of this thesis, and further details of survey methodology and findings are available in Section 2.1.2.

Vaccination against SARS-CoV-2 began in February 2021, beginning with health-care professionals and then following a three-phase age-based approach first targeting individuals over 60 years of age and subsequently expanding to individuals over 50 on 3rd May 2021 and all adults over 18 on 10th May 2021 [66]. The Dominican Republic was the first country in the Americas to approve vaccination with a third dose and booster vaccination for highly vulnerable individuals began in July 2021 [67]. Most doses administered for the primary vaccination series were Sinovac-CoronaVac inactivated COVID-19 vaccine. This is an inactivated whole-virion vaccine with lower reported efficacy compared with Pfizer-BioNTech, Moderna or Oxford-Astrazeneca vaccines. A phase III trial in Brazil found that two doses resulted in a 50.4% (95% CI: 35.3 – 62.0) efficacy against symptomatic SARS-CoV-2 infection and 100% efficacy against hospitalisation [68]. A real-world effectiveness study in Chile estimated vaccine effectiveness of 65.9 % against symptomatic disease (95% CI: 65.2 – 66.6) with efficacies over 80% against hospitalisation or death outcomes [69]. Contrastingly, phase III trials of the Oxford-Astrazeneca in the United States, Chile and Peru found vaccine efficacy of 74.0% (95% CI: 65.3% – 80.5%) with 100% efficacy against severe or critical COVID-19 [70]. A real-world effectiveness study of the Pfizer-BioNTech vaccine in Israel found 97.0% vaccine effectiveness against symptomatic disease (95% CI: 96.7 – 97.2%) and 97.2% effectiveness against hospitalisation (95% CI: 97.1 – 97.8%) [71].

1.3 Dengue virus

1.3.1 Epidemiology and transmission

Dengue virus is a rapidly expanding arbovirus (arthropod-borne virus) causing an estimated 390 million infections a year, with 96 million of these manifesting as disease [72]. Dengue is primarily transmitted by *Aedes aegypti* and *Aedes albopictus* in urban and peri-urban areas [73]. Both are day-time biting mosquitoes that typically lay eggs in domestic water containers. The global spread of *Aedes* vectors (and subsequent dengue transmission) is thought to be driven by international travel and trade, particularly the used tire and plant trade, as well as rapid population growth, urbanisation and climate change [74,75]. Around half of the global population are estimated to be at risk of dengue transmission, with highest risk in the Asia-Pacific, followed by Latin America and the Caribbean, and reported dengue cases have doubled every decade in the past 30 years [72,76]. These increases in geographical range and burden are projected to continue as global heating increases climate suitability for *Aedes* vectors and dengue transmission [77]. Dengue can exhibit both endemic and epidemic transmission patterns. Within endemic settings, transmission typically varies seasonally and can display large interannual variation, causing occasional large outbreaks with the potential to overwhelm health systems [78,76].

1.3.2 Clinical manifestation

While most dengue infections are asymptomatic or cause mild illness, they can also cause acute flu-like symptoms (dengue) or potentially fatal severe disease (severe dengue). Dengue symptoms appear around 4–10 days after infection and typically include fever, as well as headaches, muscle and joint pain and potentially nausea or vomiting. Severe dengue usually develops after the febrile phase and is characterised by plasma leakage. Symptoms of severe dengue include severe abdominal pain, repeated vomiting, bleeding gums and fatigue [79]. Risk factors for dengue infection include; age (where older adults and children have higher risk), sex (where female sex is associated with severe dengue) previous dengue infection, ethnicity, comorbidities, and use of personal protection against mosquitos [80,81]. Risk is also influenced by environmental factors

including; hygiene and water infrastructure in the home, and living in an urban or peri-urban dwelling [82].

1.3.3 Diagnosis and serology

There are four immunologically distinct serotypes of dengue virus (DENV-1 to DENV-4). Primary infection results in a temporary increase in IgM, which peaks about two weeks after onset of symptoms and wanes over 2-3 months [83]. Evidence suggests levels of IgG increase more slowly and persist for life, conferring life-long serotype-specific immunity and short lived cross-immunity. A secondary infection with a heterologous serotype results in reduced peak IgM levels but a more rapid increase in IgG which is then detectable at high levels. IgM/IgG ratios can thus be used to distinguish between primary and secondary infection [73]. Secondary infection is also associated with an increased risk of severe disease, possibly through antibody-dependent enhancement [83].

Due to the broad range of dengue symptoms, clinical diagnosis alone is often unreliable and laboratory testing is required for case confirmation. Laboratory diagnosis for dengue virus can be performed using a nucleic acid amplification test (NAAT) such as RT-PCR or an NS1 antigen test (nonstructural protein 1) in the acute phase of infection up to ~ 7 days post symptom onset, depending on the assay used [83–85]. IgM testing can also be used to detect infection, and is the most widely used test in laboratory surveillance [78]. IgG testing can be used to test for past infection; although IgG ELISA assays lack specificity between dengue virus and other flaviviruses such as Zika [83].

1.3.4 Control measures

To date, two dengue vaccines have been licensed; Dengvaxia (CYD-TDV) a live, recombinant tetravalent vaccine with a yellow fever backbone, and Qdenga (TAK-003), a live-attenuated tetravalent vaccine with a DENV-2 backbone. CYD-TDV is licensed on a 3-dose schedule with 6 month intervals between doses while TAK-003 is licensed on a 2-dose series, three months apart. However, due to evidence of decreased vaccine efficacy among seronegative individuals as well

as increased risk of severe dengue in seronegative vaccinees, CYD-TDV is only recommended for use following pre-vaccination screening [86–89]. TAK-003 is currently only recommended for use in high intensity areas, due to a lack of data on the efficacy-risk profile for DENV-3 and DENV-4 in seronegative individuals [89].

As such, the primary control measure to reduce dengue burden is currently *Aedes* vector control. Integrated vector management (IVM) is recommended by the WHO to optimise the use of resources for vector control [90]. Control methods can be environmental, biological or chemical. Environmental control methods involve improving housing or water and sanitation infrastructure to prevent the creation of mosquito-breeding habitats and reduce *Aedes* proliferation [91]. Biological and chemical control aims to reduce the population of adult or larval *Aedes* mosquitos; for instance, through the introduction of biological larvicides or spraying homes with adulticides. Other approaches focus on reducing human-vector contact using mosquito nets (which may be impregnated with insecticide) or personal repellents [92].

Novel vector control strategies have been developed using *Wolbachia*, a maternally inherited obligate intracellular bacteria, to infect *Aedes* mosquitos. *Wolbachia* inhibits the ability of dengue, and other pathogens, from infecting the mosquito [93]. There are two primary *Wolbachia*-mediated vector control strategies: introgression, where *Wolbachia*-infected mosquitoes are introgressed into wild mosquito populations; and the incompatible insect technique, where male *Wolbachia*-infected mosquitoes are released and cannot produce viable eggs with wildtype females due to cytoplasmic incompatibility [94,95].

Climate-informed early warning systems (EWS) offer the potential for advanced warning of disease outbreaks, with longer lead times for epidemic response than decision-support tools based on epidemiological surveillance [96]. However, despite much academic research into EWS for dengue and other vector-borne diseases, few countries have fully operationalised systems able to inform epidemic response [97–99].

1.3.5 Climate and dengue

In endemic settings there is often strong seasonality in dengue incidence, accompanied by interannual variation. The influences of climate on dengue virus transmission are well recognised, if complex, affecting various stages of the pathogen and vector life cycle including viral replication, mosquito breeding, mosquito survival and mosquito biting patterns. Temperature is an important predictor of transmission, shaping the timing, length and geographical extent of dengue seasons [77,100]. The optimal temperature for transmission is thought to be around 29°C, with thermal limits between 17.8°C - 34.5°C [101]. While typically wet and humid conditions lead to the creation of mosquito breeding habitats, the effects of rainfall are nuanced, and can depend on local context such as the level of urbanisation, hygiene and water infrastructure, and population behaviour during drought or flooding [102,103]. For instance, excess rainfall can lead to flushing events, where mosquito breeding habitats are washed away, resulting in decreased transmission risk [104,105]. On the other hand, drought can increase the risk of transmission at long lead times, in cases where households increase the number of water storage containers around the home. Experimental evidence suggests that increased humidity increases vector survival and egg production [106,107]. The effect of humidity on dengue transmission is less well studied, with some studies finding evidence of a statistical association between humidity and dengue cases [108].

Dengue transmission is also affected by the El Niño Southern Oscillation (ENSO). This is an interannual climatic phenomena affecting oceanic and atmospheric temperatures around the Pacific Ocean. El Niño events are defined as periods of above average Pacific sea surface temperatures, while La Niña events are periods with below average temperatures. ENSO accounts for a large proportion of interannual variation in climate, and often impacts the occurrence of extreme weather, such as temperature and precipitation extremes, drought or flooding [109]. El Niño events are associated with an average global temperature increase of 0.5°C; however, the effects on precipitation vary geographically. For instance, El Niño events have been associated with increased risk of drought in Southeast Asia, Southern Africa and Australia but with increased rainfall along the western coast of South America [110]. In Singapore, El Niño events typically lead to hotter and drier climatic conditions. Similarly, in the Dominican Republic, El Niño events

are associated with warmer temperatures and risk of droughts, while La Niña events are associated with an increased intensity of the Atlantic hurricane season [111–114]. The 2014-2016 El Niño event was one of the strongest events on record, according to the National Oceanic and Atmospheric Administration (NOAA) with evidence of increased disease incidence across the globe [115]. ENSO indicators may be particularly useful for dengue forecasting models, as they predict interannual climatic variability, and could therefore be helpful to predict high incidence seasons as opposed to seasonal patterns [110]. The extent to which a weather variable (such as temperature or precipitation) is sensitive to ENSO varies across space and time, and can vary even between instances of strong El Niño events, making disentangling the causal impact of ENSO and downstream weather effects on dengue transmission challenging [109].

1.3.6 Dengue in Singapore and the Dominican Republic

Singapore is an equatorial city-state in Southeast Asia with a tropical climate, experiencing hot and humid conditions year-round and monsoon seasons from December – March and June – September [116]. Dengue is hyperendemic in Singapore, with all four serotypes in circulation, displaying cyclical outbreaks with increasing frequency and magnitude [117,118]. Dengue outbreaks typically follow switches in the dominant circulating serotype which, until 2020, switched between DENV-1 and DENV-2 [117,119]. This pattern has been disrupted with an increasing contribution from DENV-3 between 2019 – 2020, resulting in a DENV-3 outbreak in 2022 [120]. Unlike many countries in southeast Asia and Latin America, dengue transmission in Singapore is thought to have increased during the COVID-19 pandemic, as social distancing measures resulted in more of the population working from home, which is associated with greater dengue risk [121,122].

Over the past 50 years Singapore has implemented strict dengue prevention measures, centered around vector control and public education. This has led to sharp decreases in *Aedes* indices such as the *Aedes* House index, as well as estimated dengue seroprevalence in the population [118]. Despite this, reported dengue cases have been increasing since the 1990s, alongside more frequent outbreaks. This could be the result of low levels of population immunity to dengue increasing

Singapore's vulnerability to outbreaks; particularly given that increasing travel and trade are likely to have resulted in more frequent introductions of new viral genotypes [118,100]. Improved case ascertainment is also likely to have played a role, particularly following the introduction of NS1 rapid tests in clinical laboratories in 2008 [123]. Alongside traditional vector control, Singapore has also launched Project *Wolbachia*, a *Wolbachia-Aedes* suppression strategy, which began targeted releases in May 2020 and aims to cover 35% of Singapore's households in 2024 [95,124].

The Dominican Republic also experiences endemic dengue transmission, with multiple serotypes circulating. The Dominican Republic has a tropical climate but is geographically diverse, with multiple climate zones [125]. In much of the country the rainy season falls between May and October, while along the northern coast it typically occurs between November and January. The Dominican Republic is vulnerable to hurricanes and cyclones due to its position on the Atlantic hurricane belt, which typically occur between August and October. In recent decades, outbreaks have been increasing in size, with the largest outbreak to date in 2019 reporting over 20,000 cases. There is limited serotype surveillance in the Dominican Republic, with low numbers of samples sequenced each year to detect which serotypes are circulating. Between 2005 and 2022 DENV-2 was the most common serotype detected, followed by DENV-1. DENV-3 and DENV-4 are reported more rarely and can coincide with large outbreaks; for instance DENV-3 was reported only in 2015 and 2019, both large outbreak years [126].

1.4 Aim and objectives of thesis

The aim of this thesis is to integrate multiple data streams within statistical and mechanistic modelling frameworks to investigate the role of immunity, climate and behaviour in driving viral outbreak dynamics, considering the implications for disease control.

The aim will be met through the following objectives:

1. Estimate the risk of reinfection following SARS-CoV-2 infection
2. Understand the dynamics of SARS-CoV-2 in the Dominican Republic and quantify the role of mobility, immunity and vaccination on transmission and control

3. Develop a forecasting model for dengue in Singapore incorporating climate and serotype drivers of transmission
4. Analyse the role of climate and epidemic drivers in spatiotemporal dengue outbreak dynamics in the Dominican Republic

1.5 Thesis structure

This thesis is structured in a research paper style with four chapters written as journal articles that have either been published or are in the process of publication. This first introductory chapter provides background on SARS-CoV-2 and dengue virus, while the second methods chapter describes modelling approaches used to understand viral outbreak dynamics. This is followed by four results chapters, described below, and a discussion of the findings.

Chapter Three: Investigating immunity to SARS-CoV-2 in a community seroepidemiological cohort in the United States

This chapter was published in *PLOS Biology* in 2022 [127]. It contains an analysis of longitudinal PCR and serological testing data from a seroepidemiological workplace cohort in the United States, estimating reinfection risk for SARS-CoV-2. This chapter addresses objective one.

Chapter Four: Modelling the impact of population mobility, post-infection immunity and vaccination on SARS-CoV-2 transmission in the Dominican Republic

This chapter has been published in *The Lancet Regional Health - Americas* in 2024 [128] and addresses objective two. It presents a mechanistic modelling analysis of SARS-CoV-2 transmission in the Dominican Republic from 2020 – 2022, quantifying the role of immune and behavioural drivers of SARS-CoV-2 transmission. The study also details a counterfactual analysis assessing the impact of the vaccination campaign and considering alternative vaccine counterfactuals around timing and efficacy of vaccine products.

Chapter Five: Disentangling the role of climate and serotype competition to forecast dengue outbreaks in Singapore

This chapter has been submitted for publication and addresses objective 3. It contains an analysis of the role of climate and serotype competition in shaping dengue risk in Singapore from 2000 – 2023. The paper then presents a statistical forecasting model for early warning and evaluates predictive performance for 2 –8 week ahead forecasts.

Chapter Six: Understanding the role of climate and epidemic drivers in spatiotemporal dengue outbreak dynamics in the Dominican Republic

This chapter addresses objective 4. It describes a spatiotemporal analysis of dengue in the Dominican Republic using surveillance data from 2013 – 2023. It details the results from Bayesian hierarchical modelling to explore the role of climatic drivers (including temperature, rainfall, humidity and ENSO) as well as epidemic drivers in dengue outbreak dynamics. The study evaluates the influence of climatic and epidemic drivers in predicting temporal and spatial patterns of disease risk.

References

1. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, Angelis DD, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science* [Internet]. 2015 Mar 13 [cited 2022 Feb 1]; Available from: <https://www.science.org/doi/abs/10.1126/science.aaa4339>
2. Moran KR, Fairchild G, Generous N, Hickmann K, Osthus D, Priedhorsky R, et al. Epidemic Forecasting is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast. *J Infect Dis*. 2016 Dec 1;214(Suppl 4):S404–8.
3. Institute of Health Metrics and Evaluation (IHME). Global Burden of Disease - Deaths and DALYs [Internet]. Available from: <https://www.healthdata.org/research-analysis/gbd-data>
4. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. 2022 Apr;20(4):193–205.
5. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science*. 2003 Jun 20;300(5627):1961–6.
6. Baguelin M, Hoek AJV, Jit M, Flasche S, White PJ, Edmunds WJ. Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. *Vaccine*. 2010 Mar 11;28(12):2370–84.
7. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*. 2018 Mar;22:56–61.
8. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*. 2020 Jul 1;5(7):e375–85.
9. Keeling MJ, Dyson L, Tildesley MJ, Hill EM, Moore S. Comparison of the 2021 COVID-19 roadmap projections against public health data in England. *Nat Commun*. 2022 Aug 22;13(1):4924.
10. Endo A, Murayama H, Abbott S, Ratnayake R, Pearson CAB, Edmunds WJ, et al. Heavy-tailed sexual contact networks and monkeypox epidemiology in the global outbreak, 2022. *Science*. 2022 Oct 7;378(6615):90–4.

11. Overton CE, Abbott S, Christie R, Cumming F, Day J, Jones O, et al. Nowcasting the 2022 mpox outbreak in England. *PLOS Computational Biology*. 2023 Sep 18;19(9):e1011463.
12. De Angelis D, Presanis AM, Birrell PJ, Tomba GS, House T. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*. 2015 Mar 1;10:83–7.
13. McCrea R, King R, Graham L, Börger L. Realising the promise of large data and complex models. *Methods in Ecology and Evolution*. 2023;14(1):4–11.
14. Kucharski AJ, Kama M, Watson CH, Aubry M, Funk S, Henderson AD, et al. Using paired serology and surveillance data to quantify dengue transmission and control during a large outbreak in Fiji. *eLife*. 7:e34848.
15. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*. 2020 Aug;26(8):1205–11.
16. Novel Coronavirus (2019-nCoV) situation report-1 [Internet]. WHO; [cited 2021 Jul 16]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4
17. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*. 2020 Feb 20;382(8):727–33.
18. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020 Mar;579(7798):270–3.
19. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) [Internet]. [cited 2024 Jun 9]. Available from: [https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))
20. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic [Internet]. [cited 2024 Jun 9]. Available from: [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic)

21. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus Pandemic (COVID-19). Our World in Data [Internet]. 2020 Mar 5 [cited 2024 Jun 10]; Available from: <https://ourworldindata.org/covid-cases>
22. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*. 2020 Feb 15;395(10223):514–23.
23. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020 Feb 15;395(10223):497–506.
24. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*. 2020 Mar 26;382(13):1199–207.
25. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*. 2020 May 1;20(5):553–8.
26. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, et al. Report 3 - Transmissibility of 2019-nCoV [Internet]. Imperial College London; [cited 2024 Jun 9]. Available from: <https://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/disease-areas/covid-19/report-3-transmissibility-of-covid-19/>
27. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill*. 2020 Jan;25(4):2000058.
28. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*. 2020;5:67.
29. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* [Internet]. 2021 Apr 9 [cited 2021 Jul 16];372(6538). Available from: <https://science.sciencemag.org/content/372/6538/eabg3055>

30. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*. 2021 May;593(7858):270–4.
31. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021 May 21;372(6544):815–21.
32. Kucharski AJ, Eggo RM, Davies NG, Funk S, CMMID COVID-19 Working Group. Dynamics of B.1.617.2 in the UK from importations, traveller-linked and non-traveller-linked transmission [Internet]. 2021 May [cited 2021 Jul 16]. Available from: <https://cmmid.github.io/topics/covid19/import-analysis.html>
33. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *N Engl J Med*. 2022 Apr 21;386(16):1532–46.
34. Willett BJ, Grove J, MacLean OA, Wilkie C, De Lorenzo G, Furnon W, et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat Microbiol*. 2022 Aug;7(8):1161–79.
35. Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*. 2022 Feb;602(7898):657–63.
36. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*. 2020 Aug 1;20(8):911–9.
37. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Medicine*. 2020 Sep 22;17(9):e1003346.
38. Brazeau NF, Verity R, Jenks S, Fu H, Whittaker C, Winskill P, et al. Estimating the COVID-19 infection fatality ratio accounting for seroreversion using statistical modelling. *Commun Med*. 2022 May 19;2(1):1–13.
39. O’Driscoll M, Dos Santos GR, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*. 2020 Nov 2;1–6.

40. Antigen-detection in the diagnosis of SARS-CoV-2 infection [Internet]. [cited 2024 Jun 10]. Available from: <https://www.who.int/publications-detail-redirect/antigen-detection-in-the-diagnosis-of-sars-cov-2infection-using-rapid-immunoassays>
41. Seow J, Graham C, Merrick B, Acors S, Pickering S, Steel KJA, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology*. 2020 Dec;5(12):1598–607.
42. Gudbjartsson DF, Norddahl GL, Melsted P, Gunnarsdottir K, Holm H, Eythorsson E, et al. Humoral Immune Response to SARS-CoV-2 in Iceland. *New England Journal of Medicine*. 2020 Oct 29;383(18):1724–34.
43. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature Medicine*. 2020 Jun;26(6):845–8.
44. Baden Lindsey R., El Sahly Hana M., Essink Brandon, Kotloff Karen, Frey Sharon, Novak Rick, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine*. 2021 Feb 3;384(5):403–16.
45. Polack Fernando P., Thomas Stephen J., Kitchin Nicholas, Absalon Judith, Gurtman Alejandra, Lockhart Stephen, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*. 2020 Dec 30;383(27):2603–15.
46. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet*. 2021 Jan 9;397(10269):99–111.
47. Tanriover MD, Doğanay HL, Akova M, Güner HR, Azap A, Akhan S, et al. Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): interim results of a double-blind, randomised, placebo-controlled, phase 3 trial in Turkey. *The Lancet*. 2021 Jul 17;398(10296):213–22.
48. Clapham H, Hay J, Routledge I, Takahashi S, Choisy M, Cummings D, et al. Seroepidemiologic Study Designs for Determining SARS-COV-2 Transmission and Immunity. *Emerg Infect Dis*. 2020 Sep;26(9):1978–86.
49. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S mok, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus

Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *J Clin Med*. 2020 Feb 17;9(2):538.

50. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020 Mar 10;M20-0504.

51. Lau YC, Tsang TK, Kennedy-Shaffer L, Kahn R, Lau EHY, Chen D, et al. Joint Estimation of Generation Time and Incubation Period for Coronavirus Disease 2019. *The Journal of Infectious Diseases*. 2021 Nov 15;224(10):1664–71.

52. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, et al. Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis. *The Lancet Infectious Diseases*. 2022 May 1;22(5):603–10.

53. Wang Y, Chen R, Hu F, Lan Y, Yang Z, Zhan C, et al. Transmission, viral kinetics and clinical characteristics of the emergent SARS-CoV-2 Delta VOC in Guangzhou, China. *EClinicalMedicine*. 2021 Oct 1;40:101129.

54. Park SW, Sun K, Abbott S, Sender R, Bar-on YM, Weitz JS, et al. Inferring the differences in incubation-period and generation-interval distributions of the Delta and Omicron variants of SARS-CoV-2. *Proceedings of the National Academy of Sciences*. 2023 May 30;120(22):e2221887120.

55. Xiang F, Wang X, He X, Peng Z, Yang B, Zhang J, et al. Antibody Detection and Dynamic Characteristics in Patients With Coronavirus Disease 2019. *Clinical Infectious Diseases*. 2020 Nov 5;71(8):1930–4.

56. Liu Y, Morgenstern C, Kelly J, Lowe R, CMMID COVID-19 Working Group, Jit M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Med*. 2021 Feb 5;19(1):40.

57. ECDC. Guidelines for the implementation of non-pharmaceutical interventions against COVID-19 [Internet]. 2020 Sep [cited 2024 Jun 10]. Available from: <https://www.ecdc.europa.eu/en/publications-data/covid-19-guidelines-non-pharmaceutical-interventions>

58. Covid-19 vaccine: First person receives Pfizer jab in UK. *BBC News* [Internet]. 2020 Dec 8 [cited 2024 Jun 10]; Available from: <https://www.bbc.com/news/uk-55227325>

59. Gozzi N, Chinazzi M, Dean NE, Longini Jr IM, Halloran ME, Perra N, et al. Estimating the impact of COVID-19 vaccine inequities: a modeling study. *Nat Commun.* 2023 Jun 6;14(1):3272.
60. Duan Y, Shi J, Wang Z, Zhou S, Jin Y, Zheng ZJ. Disparities in COVID-19 Vaccination among Low-, Middle-, and High-Income Countries: The Mediating Role of Vaccination Policy. *Vaccines (Basel).* 2021 Aug 14;9(8):905.
61. World Bank. Dominican Republic | Data [Internet]. [cited 2021 Jul 29]. Available from: <https://data.worldbank.org/country/dominican-republic>
62. Kincaid J. Dominican Republic hospitals hit capacity for COVID-19 patients [Internet]. *Miami Herald.* [cited 2021 Jul 30]. Available from: <https://www.miamiherald.com/news/nation-world/world/americas/article244728677.html>
63. Diario L. Sistema hospitalario a punto del colapso por el Covid-19 [Internet]. *listindiario.com.* 2020 [cited 2024 Jun 10]. Available from: <https://listindiario.com/la-republica/2020/06/20/622736/sistema-hospitalario-a-punto-del-colapso-por-el-covid-19.html>
64. Paulino-Ramirez R, Báez AA, Degaudenzi AV, Tapia L. Seroprevalence of Specific Antibodies against SARS-CoV-2 from Hotspot Communities in the Dominican Republic. *The American Journal of Tropical Medicine and Hygiene.* 2020 Oct 21;103(6):2343–6.
65. Nilles EJ, Paulino CT, de St. Aubin M, Restrepo AC, Mayfield H, Dumas D, et al. SARS-CoV-2 seroprevalence, cumulative infections, and immunity to symptomatic infection – A multistage national household survey and modelling study, Dominican Republic, June–October 2021. *The Lancet Regional Health - Americas.* 2022 Dec 1;16:100390.
66. Pérez-Then E, Miric M, Qian HZ, Chen YQ, Wang Y, Vallejo V, et al. Population-Level Effectiveness of an Inactivated Whole-Virion COVID-19 Vaccine: A Test Negative Case-Control Study in the Dominican Republic. *Open Forum Infectious Diseases.* 2023 Mar 1;10(3):ofad075.
67. Pérez-Then E, Lucas C, Monteiro VS, Miric M, Brache V, Cochon L, et al. Neutralizing antibodies against the SARS-CoV-2 Delta and Omicron variants following heterologous CoronaVac plus BNT162b2 booster vaccination. *Nat Med.* 2022 Mar;28(3):481–5.
68. Palacios R, Batista AP, Albuquerque CSN, Patiño EG, Santos J do P, Tilli Reis Pessoa Conde M, et al. Efficacy and Safety of a COVID-19 Inactivated Vaccine in Healthcare Professionals in Brazil: The PROFISCOV Study [Internet]. Rochester, NY: Social Science

Research Network; 2021 Apr [cited 2022 Jan 13]. Report No.: ID 3822780. Available from: <https://papers.ssrn.com/abstract=3822780>

69. Jara A, Undurraga EA, González C, Paredes F, Fontecilla T, Jara G, et al. Effectiveness of an Inactivated SARS-CoV-2 Vaccine in Chile. *New England Journal of Medicine*. 2021 Sep 2;385(10):875–84.
70. Falsey AR, Sobieszczyk ME, Hirsch I, Sproule S, Robb ML, Corey L, et al. Phase 3 Safety and Efficacy of AZD1222 (ChAdOx1 nCoV-19) Covid-19 Vaccine. *New England Journal of Medicine*. 2021 Dec 15;385(25):2348–60.
71. Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *The Lancet*. 2021 May 15;397(10287):1819–29.
72. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013 Apr;496(7446):504–7.
73. Simmons CP, Farrar JJ, van Vinh Chau N, Wills B. Dengue [Internet]. <http://dx.doi.org/10.1056/NEJMra1110265>. Massachusetts Medical Society; 2012 [cited 2021 Jul 15]. Available from: <https://www.nejm.org/doi/10.1056/NEJMra1110265>
74. Lowe R, Ryan SJ, Mahon R, Meerbeeck CJV, Trotman AR, Boodram LLG, et al. Building resilience to mosquito-borne diseases in the Caribbean. *PLOS Biology*. 2020 Nov 24;18(11):e3000791.
75. Kraemer MUG, Reiner RC, Brady OJ, Messina JP, Gilbert M, Pigott DM, et al. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nat Microbiol*. 2019 May;4(5):854–63.
76. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases*. 2016 Jun 1;16(6):712–23.
77. Colón-González FJ, Sewe MO, Tompkins AM, Sjödin H, Casallas A, Rocklöv J, et al. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study. *The Lancet Planetary Health*. 2021 Jul 1;5(7):e404–14.

78. Guzman MG, Harris E. Dengue. *The Lancet*. 2015 Jan 31;385(9966):453–65.
79. Dengue and severe dengue [Internet]. World Health Organisation; [cited 2021 Jul 15]. Available from: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>
80. Paraná VC, Feitosa CA, da Silva GCS, Gois LL, Santos LA. Risk factors associated with severe dengue in Latin America: A systematic review and meta-analysis. *Tropical Medicine & International Health*. 2024;29(3):173–91.
81. Sangkaew S, Ming D, Boonyasiri A, Honeyford K, Kalayanarooj S, Yacoub S, et al. Risk predictors of progression to severe disease during the febrile phase of dengue: a systematic review and meta-analysis. *The Lancet Infectious Diseases*. 2021 Jul 1;21(7):1014–26.
82. Gibb R, Colón-González FJ, Lan PT, Huong PT, Nam VS, Duoc VT, et al. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nat Commun*. 2023 Dec 11;14(1):8179.
83. Organization WH. Dengue guidelines for diagnosis, treatment, prevention and control : new edition [Internet]. World Health Organization; 2009 [cited 2024 Jun 16]. Available from: <https://iris.who.int/handle/10665/44188>
84. Santiago GA, Vergne E, Quiles Y, Cosme J, Vazquez J, Medina JF, et al. Analytical and clinical performance of the CDC real time RT-PCR assay for detection and typing of dengue virus. *PLoS Negl Trop Dis*. 2013;7(7):e2311.
85. Lima M da RQ, Nogueira RMR, Schatzmayr HG, de Filippis AMB, Limonta D, dos Santos FB. A New Approach to Dengue Fatal Cases Diagnosis: NS1 Antigen Capture in Tissues. *PLoS Negl Trop Dis*. 2011 May 3;5(5):e1147.
86. Wilder-Smith A, Hombach J, Ferguson N, Selgelid M, O'Brien K, Vannice K, et al. Deliberations of the Strategic Advisory Group of Experts on Immunization on the use of CYD-TDV dengue vaccine. *The Lancet Infectious Diseases*. 2019 Jan 1;19(1):e31–8.
87. Sridhar S, Luedtke A, Langevin E, Zhu M, Bonaparte M, Machabert T, et al. Effect of Dengue Serostatus on Dengue Vaccine Safety and Efficacy. *N Engl J Med*. 2018 Jul 26;379(4):327–40.
88. Flasche S, Jit M, Rodríguez-Barraquer I, Coudeville L, Recker M, Koelle K, et al. The Long-Term Safety, Public Health Impact, and Cost-Effectiveness of Routine Vaccination with a Recombinant, Live-Attenuated Dengue Vaccine (Dengvaxia): A Model Comparison Study. *PLoS Med*. 2016 Nov;13(11):e1002181.

89. World Health Organization. WHO position paper on dengue vaccines, May 2024 [Internet]. [cited 2024 May 27]. Available from: <https://www.who.int/publications-detail-redirect/who-wer-9918-203-224>
90. WHO. Global vector control response 2017–2030 [Internet]. [cited 2024 Jul 9]. Available from: <https://www.who.int/publications/i/item/9789241512978>
91. Buhler C, Winkler V, Runge-Ranzinger S, Boyce R, Horstick O. Environmental methods for dengue vector control – A systematic review and meta-analysis. *PLOS Neglected Tropical Diseases*. 2019 Jul 11;13(7):e0007420.
92. Handbook for Integrated Vector Management in the Americas - PAHO/WHO | Pan American Health Organization [Internet]. 2019 [cited 2024 Jun 17]. Available from: <https://www.paho.org/en/documents/handbook-integrated-vector-management-americas>
93. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, et al. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*. *Cell*. 2009 Dec 24;139(7):1268–78.
94. Utarini Adi, Indriani Citra, Ahmad Riris A., Tantowijoyo Warsito, Arguni Eggi, Ansari M. Ridwan, et al. Efficacy of *Wolbachia*-Infected Mosquito Deployments for the Control of Dengue. *New England Journal of Medicine*. 2021 Jun 9;384(23):2177–86.
95. Lim JT, Bansal S, Chong CS, Dickens B, Ng Y, Deng L, et al. Efficacy of *Wolbachia*-mediated sterility to reduce the incidence of dengue: a synthetic control study in Singapore. *The Lancet Microbe* [Internet]. 2024 Feb 8 [cited 2024 Mar 5];0(0). Available from: [https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247\(23\)00397-X/fulltext](https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(23)00397-X/fulltext)
96. Finch E, Batista ML, Alcayna T, Lee SA, Fletcher IK, Lowe R. Early warning systems for vector-borne diseases: engagement, methods and implementation. In: *Planetary health approaches to understand and control vector-borne diseases* [Internet]. Wageningen Academic; 2023 [cited 2024 Jan 22]. p. 347–86. Available from: <https://brill.com/edcollchap-ooa/book/9789004688650/BP000023.xml>
97. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, et al. Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environ Health Perspect*. 2016 Sep;124(9):1369–75.

98. Codeco C, Coelho F, Cruz O, Oliveira S, Castro T, Bastos L. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. *Revue d'Épidémiologie et de Santé Publique*. 2018 Jul 1;66:S386.
99. Colón-González FJ, Bastos LS, Hofmann B, Hopkin A, Harpham Q, Crocker T, et al. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Medicine*. 2021 Mar 4;18(3):e1003542.
100. van Panhuis WG, Choisy M, Xiong X, Chok NS, Akarasewi P, Iamsirithaworn S, et al. Region-wide synchrony and traveling waves of dengue across eight countries in Southeast Asia. *Proceedings of the National Academy of Sciences*. 2015 Oct 20;112(42):13069–74.
101. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. *Ecology Letters*. 2019;22(10):1690–708.
102. Lowe R, Gasparrini A, Meerbeeck CJV, Lippi CA, Mahon R, Trotman AR, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLOS Medicine*. 2018 Jul 17;15(7):e1002613.
103. Lowe R, Lee SA, O'Reilly KM, Brady OJ, Bastos L, Carrasco-Escobar G, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*. 2021 Apr 1;5(4):e209–19.
104. Benedum CM, Seidahmed OME, Eltahir EAB, Markuzon N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. *PLOS Neglected Tropical Diseases*. 2018 Dec 6;12(12):e0006935.
105. Koenraadt CJM, Harrington LC. Flushing effect of rain on container-inhabiting mosquitoes *Aedes aegypti* and *Culex pipiens* (Diptera: Culicidae). *J Med Entomol*. 2008 Jan;45(1):28–35.
106. Schmidt CA, Comeau G, Monaghan AJ, Williamson DJ, Ernst KC. Effects of desiccation stress on adult female longevity in *Aedes aegypti* and *Ae. albopictus* (Diptera: Culicidae): results of a systematic review and pooled survival analysis. *Parasites & Vectors*. 2018 Apr 25;11(1):267.
107. Lega J, Brown HE, Barrera R. *Aedes aegypti* (Diptera: Culicidae) Abundance Model Improved With Relative Humidity and Precipitation-Driven Egg Hatching. *J Med Entomol*. 2017 Sep 1;54(5):1375–84.
108. Xu HY, Fu X, Lee LKH, Ma S, Goh KT, Wong J, et al. Statistical Modeling Reveals the Effect of Absolute Humidity on Dengue in Singapore. *PLOS Neglected Tropical Diseases*. 2014 May 1;8(5):e2805.

109. McGregor GR, Ebi K. El Niño Southern Oscillation (ENSO) and Health: An Overview for Climate and Health Researchers. *Atmosphere*. 2018 Jul;9(7):282.
110. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. El Niño and health. *The Lancet*. 2003 Nov 1;362(9394):1481–9.
111. Meteorological Service Singapore. El Niño/La Niña Status [Internet]. Available from: <http://www.weather.gov.sg/climate-el-la/>
112. Earnest A, Tan SB, Wilder-Smith A. Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections. *Epidemiol Infect*. 2012 Jul;140(7):1244–51.
113. Giannini A, Kushnir Y, Cane MA. Interannual Variability of Caribbean Rainfall, ENSO, and the Atlantic Ocean. *Journal of Climate*. 2000 Jan 15;13(2):297–311.
114. Taylor MA, Stephenson TS, Owino A, Chen AA, Campbell JD. Tropical gradient influences on Caribbean rainfall. *Journal of Geophysical Research: Atmospheres* [Internet]. 2011 [cited 2024 May 28];116(D21). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015580>
115. Anyamba A, Chretien JP, Britch SC, Soebiyanto RP, Small JL, Jepsen R, et al. Global Disease Outbreaks Associated with the 2015–2016 El Niño Event. *Sci Rep*. 2019 Feb 13;9(1):1930.
116. NEA. Guidebook on Climate of Singapore [Internet]. 2009 [cited 2024 Jan 9]. Available from: <https://www.nea.gov.sg/docs/default-source/resource/guidebook-on-climate-of-singapore.pdf>
117. Rajarethinam J, Ang LW, Ong J, Ycasas J, Hapuarachchi HC, Yap G, et al. Dengue in Singapore from 2004 to 2016: Cyclical Epidemic Patterns Dominated by Serotypes 1 and 2. *Am J Trop Med Hyg*. 2018 Jul;99(1):204–10.
118. Ho SH, Lim JT, Ong J, Hapuarachchi HC, Sim S, Ng LC. Singapore’s 5 decades of dengue prevention and control—Implications for global dengue control. *PLOS Neglected Tropical Diseases*. 2023 Jun 22;17(6):e0011400.
119. Lee KS, Lai YL, Lo S, Barkham T, Aw P, Ooi PL, et al. Dengue Virus Surveillance for Early Warning, Singapore - Volume 16, Number 5—May 2010 - *Emerging Infectious Diseases journal - CDC*. [cited 2022 Oct 22]; Available from: https://wwwnc.cdc.gov/eid/article/16/5/09-1006_article

120. National Environment Agency (NEA). Weekly Infectious Diseases Bulletin [Internet]. [cited 2024 Jan 15]. Available from: <https://www.moh.gov.sg/resources-statistics/infectious-disease-statistics/2023/weekly-infectious-diseases-bulletin>
121. Chen Y, Li N, Lourenço J, Wang L, Cazelles B, Dong L, et al. Measuring the effects of COVID-19-related disruption on dengue transmission in southeast Asia and Latin America: a statistical modelling study. *The Lancet Infectious Diseases*. 2022 May 1;22(5):657–67.
122. Lim JT, Chew LZ, Choo ELW, Dickens BSL, Ong J, Aik J, et al. Increased Dengue Transmissions in Singapore Attributable to SARS-CoV-2 Social Distancing Measures. *The Journal of Infectious Diseases*. 2021 Feb 1;223(3):399–402.
123. Pok KY, Lai YL, Sng J, Ng LC. Evaluation of Nonstructural 1 Antigen Assays for the Diagnosis and Surveillance of Dengue in Singapore. *Vector Borne Zoonotic Dis*. 2010 Dec;10(10):1009–16.
124. Wolbachia-Aedes Mosquito Suppression Strategy [Internet]. [cited 2024 Jun 18]. Available from: <https://www.nea.gov.sg>
125. Izzo M, Roszkopf CM, Aucelli PPC, Maratea A, Méndez R, Pérez C, et al. A New Climatic Map of the Dominican Republic Based on the Thornthwaite Classification. *Physical Geography* [Internet]. 2010 Sep 1 [cited 2024 May 27]; Available from: <https://www.tandfonline.com/doi/abs/10.2747/0272-3646.31.5.455>
126. PAHO/WHO. PAHO/WHO Data - Dengue fever cases [Internet]. Pan American Health Organization / World Health Organization. 2015 [cited 2024 May 28]. Available from: <https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/252-dengue-pais-ano-en.html>
127. Finch E, Lowe R, Fischinger S, Aubin M de S, Siddiqui SM, Dayal D, et al. SARS-CoV-2 antibodies protect against reinfection for at least 6 months in a multicentre seroepidemiological workplace cohort. *PLOS Biology*. 2022 Feb 10;20(2):e3001531.
128. Finch E, Nilles EJ, Paulino CT, Skewes-Ramm R, Lau CL, Lowe R, et al. Effects of mobility, immunity and vaccination on SARS-CoV-2 transmission in the Dominican Republic: a modelling study. *The Lancet Regional Health – Americas* [Internet]. 2024 Sep 1 [cited 2024 Nov 9];37. Available from: [https://www.thelancet.com/journals/lanam/article/PIIS2667-193X\(24\)00187-X/fulltext](https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(24)00187-X/fulltext)

Chapter 2

Methods

2.1 Study sites

2.1.1 Seroepidemiological cohort in the United States

Chapter 3 makes use of data from a seroepidemiological cohort of SpaceX employees in the United States. Employees from seven sites in California, Florida, Texas and Washington state were invited to partake by email, with no exclusion criteria. Serological testing was performed with the Ragon/MGH ELISA assay, with 82.4% sensitivity and 99.6% specificity [1,2]. Serological and PCR testing data were available for 1,800 individuals, with serological samples taken during four rounds of testing between April and September 2021, and PCR testing data available until January 2021.

2.1.2 Seroprevalence survey in the Dominican Republic

Chapters 4 and 6 of this thesis make use of seroprevalence data from a nationally representative seroprevalence survey conducted in the Dominican Republic from June – October 2021 [3]. Nilles *et al* followed a three-stage sampling design, selecting 134 clusters from 12,565 communities in the Dominican Republic. In the first stage, clusters were allocated to each of the 32 provinces weighted by population while also maximising spatial distribution and balancing urban and rural environments. In the second stage, clusters within each province were selected using a grid method to maximise their spatial dispersion. Finally, 23 households were chosen for each cluster, again

using grid methodology on satellite images. Two provinces (San Pedro de Macorís and Espaillat) were oversampled, as a longitudinal acute febrile illness hospital surveillance study was taking place in these provinces. Any household member ≥ 5 years old was invited to participate. Written consent was obtained for all participants; for children < 18 this was obtained from a legal guardian. 6,683 study participants were enrolled from 3832 households in 134 clusters, with an enrolment rate of 84.4% of eligible individuals present at the time of the visit.

Samples were tested with Roche Elecsys SARS-CoV-2 electrochemiluminescence immunoassays for antibodies against SARS-CoV-2 spike protein (anti-S) and nucleocapsid protein (anti-NC). Large non-manufacture sponsored studies demonstrated specificity and sensitivity of 99.8% (95% CI: 99.3–100) and 98.2% (95% CI: 96.5–99.2) for the anti-S assay and 99.6% (95% CI: 98.9–100) and 90.8% (95% CI: 81.3–95.7) for the anti-NC assay [4,2]. A subset of 200 samples from Espaillat and San Pedro de Macoris were tested for dengue IgG positivity using an in-house DENV1-4 ELISA assay.

In Chapter 4, I make use of publicly available COVID-19 surveillance data streams including daily reported deaths from the Dominican Republic’s COVID-19 dashboard, as well as daily hospital and intensive care unit (ICU) bed occupancy from daily COVID-19 bulletins by the Ministerio de Salud Pública y Asistencia Social, available from 19th September 2020 onwards [5]. In Chapter 6, I investigate dengue surveillance data provided by the Ministerio de Salud Pública y Asistencia Social from the Sistema Nacional de Vigilancia Epidemiológica (SINAVE), aggregated to weekly reported case counts, including clinically and laboratory confirmed cases. I also obtained weekly, municipality-level mean, minimum and maximum weekly temperature ($^{\circ}\text{C}$), precipitation (mm), relative humidity (%), absolute humidity (g/m^3) and specific humidity (g/kg) from the ERA5-Land reanalysis dataset from coauthors in the Barcelona Supercomputing Centre (BSC). Niño 3.4 sea surface temperature anomaly (SSTA) data is publicly available from the National Oceanic and Atmospheric Administration, NOAA [6].

2.1.3 Dengue surveillance in Singapore

In Chapter 3, I investigate over 20 years of dengue surveillance data in Singapore, from 1st January 2000 – 31st December 2022, provided by the National Environment Agency (NEA) and the Ministry of Health. We used weekly laboratory-confirmed cases, with laboratory confirmation through antigen detection of nonstructural protein 1 (NS1) or the detection of viral RNA by PCR. From 2006, a virus surveillance programme was set up by the NEA’s Environmental Health institute, where a subset of dengue samples are serotyped using RT-PCR each week [7]. I also made use of yearly population size estimates, obtained from the Department of Statistics, weekly minimum, mean and maximum temperature (°C), and relative (%) and absolute (g/m³) humidity. Daily precipitation (mm) values were also provided.

2.2 Modelling methods

This thesis draws on both statistical and mechanistic modelling approaches, with elements drawn from machine learning literature, such as cross-validation methodology. Statistical models, also known as phenomenological models, make inferences from a sample regarding the entire population, under a set of assumptions about the probability distribution giving rise to the observed data [8]. Contrastingly, mechanistic approaches, also referred to as dynamic or mathematical modelling, make inferences by constructing a simplified representation of a process using mathematical equations. In epidemic modelling, mechanistic models represent an epidemic, or how an infectious disease spreads through a population [9]. While typical statistical approaches aim to understand the probability of observing some epidemic data Y , given the underlying process X , or $P(Y|X = x)$, mechanistic approaches often aim to understand the joint probability of the underlying epidemic state and the probability of observing data, or $P(Y|X)P(X) = P(X, Y)$ [10]. The choice of modelling approach depends on the study aims, the type and extent of data available and the model outputs needed. While no model is able to perfectly represent the complexities of infectious disease transmission, models should balance accuracy (the ability to reproduce observed epidemic dynamics), transparency (being able to understand how different parameters interact to affect modelled dynamics) and flexibility to suit their purpose, and be

parameterisable from available data [11]. Another important concept in model choice is that of parsimony, that a good model should be as simple as possible, while capturing the features of a biological system that are needed to answer the question at hand. Modern model building techniques often incorporate different aspects of statistical and mechanistic approaches.

While these modelling approaches have different strengths and weaknesses, they can both be used as predictive tools, or to further understanding of underlying epidemiological processes. Within statistical modelling, a distinction is often made between models that aim to explain, or test causal explanations (for instance, asking what climatic conditions increase transmission), and those that aim to predict (for instance, asking how many cases we are likely to see in two weeks time). Whether a model is explanatory or predictive can be used to guide decisions in model formulation, selection and evaluation [12]. However, there are also benefits to integrating explanatory and predictive modelling approaches, synthesizing causal relationships to make improved predictions [13]. For instance, quantifying the predictive power of hypothesis-driven explanatory models can test the generalisability of inferred causal relationships. In cases where predictive power is limited, this could motivate future work on new hypotheses or reveal intrinsic limits to the predictability of a system [13,14]. Similarly predictive models, such as forecasting models, based on inferred causal relationships may be more generalisable and provide more insight when considering a system under long-term changes, such as the climate crisis. Forecasting models with a clear explanatory framework are also often more interpretable than their ‘black-box’ alternatives, which can be helpful when communicating outcomes to policymakers.

2.2.1 Mechanistic modelling

Compartmental models

Mechanistic models, or transmission dynamic models, typically represent host disease states, in human or animal hosts, and track changes in these states over time [11,9]. A common transmission dynamic model framework is the compartmental or *SIR* model which, in its simplest form, divides the population into compartments of ‘Susceptible’, ‘Infected’ and ‘Recovered’ (or ‘Immune’) states [15]. These models can be deterministic (where given the same initial conditions and the

same parameters we will always observe the same trajectory) or stochastic (which approximate the element of chance in transmission). Furthermore, models can be constructed in continuous time using differential equations, or in discrete time using difference equations. In the simplest, deterministic case in continuous time, the SIR model is given by the following ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) \tag{2.1}$$

$$\frac{dI(t)}{dt} = \beta(t)S(t)I(t) - \gamma I(t) \tag{2.2}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{2.3}$$

Here $\beta(t)$ represents the effective contact rate (the rate at which two individuals come into effective contact per unit time). The force of infection, λ , is the per-capita rate at which susceptible individuals become infected. Assuming homogenous mixing, where contact between any two individuals occurs with equal probability, this is defined as $\beta(t)I(t)$. γ represents the recovery rate and $\frac{1}{\gamma}$ represents the average infectious period. In this simple scenario, $\frac{\beta(t)}{\gamma}$ defines the basic reproduction number, typically given by R_0 . This is defined as the average number of secondary cases arising from an average primary case in a completely susceptible population.

This framework can be extended, depending on the complexity of the system studied. A common extension is to add an E compartment for diseases with a latent period (where individuals are exposed but not yet infectious) resulting in an $SEIR$ model. Similarly, for infectious diseases that are not fully immunising, reinfection can be added where individuals can move from the R compartment, back to the S compartment.

For diseases where contact patterns, and transmission, is age-dependent, each compartment can be stratified by age. This is particularly useful when modelling transmission of respiratory diseases, such as influenza or COVID-19 [16,17]. In this case, the force of infection for age group i is given by:

$$\lambda_i(t) = \sum_{j=1}^{j=J} c_{i,j} \cdot I_j(t) \quad 2.4$$

Here i represents the age group of the susceptible individual and j represents the age group of the infected individual and c represents contact rates between age groups, scaled for reciprocity of contacts between age groups [18,19]. In this case, we can calculate R_0 using the method outlined by Diekmann et al, whereby R_0 is the dominant eigenvalue of the next generation matrix [20]. The next generation matrix K is a matrix of numbers of newly infected individuals in each age category in consecutive generations and is given by:

$$K_{i,j} = \frac{\beta_{i,j} D S_i}{N_i} \quad 2.5$$

Where $\beta_{i,j}$ is the transmission rate between age group i and age j , D is the duration of infectiousness, S_i is the number of susceptible individuals in age class i and N_i is the population size in age class i . These frameworks can then incorporate data on age-dependent contact rates, such as those estimated through the POLYMOD or the CoMix survey, as well as age-dependent susceptibility or infectiousness [21,22].

Catalytic models

Another type of compartmental model commonly used in epidemiology is the catalytic model. Originally formulated by Muench in the 1930s, this assumes individuals are infected at a constant rate (given by the force of infection λ), which can then be estimated from age-stratified seroprevalence data [23]. For a simple catalytic model, assuming a constant force of infection independent of time or age and that individuals remain seropositive after seroconversion, the probability of being infected by age a is given by:

$$z(a) = 1 - e^{-\lambda a} \quad 2.6$$

This can be extended in several ways. In cases where measured antibodies wane over time a reverse catalytic model can be used to jointly estimate the force of infection, λ , and antibody waning, ρ [24]. Here:

$$\frac{dz(a)}{da} = \lambda(1 - z(a)) - \rho z(a) \quad 2.7$$

$$\frac{dz(a)}{da} = \lambda - (\lambda + \rho)z(a) \quad 2.8$$

By integrating the formula above we find that the probability of being infected by age a follows:

$$z(a) = \frac{\lambda}{\lambda + \rho}(1 - e^{-a(\lambda + \rho)}) \quad 2.9$$

There are several limitations to mechanistic models. Firstly, they rely on assumptions around the transmission process of the pathogen in question, and as such are less suitable for instances where the causal pathways underlying transmission are poorly characterised. Additionally, mechanistic models can be challenging to fit, particularly in cases where many individuals or compartments are being modelled, as they become increasingly time consuming to simulate and the number of fitted parameters increases.

2.2.3 Statistical modelling

Regression models are a class of statistical models that quantify the associations between covariates and an outcome which, in epidemiological use cases, is usually a disease outcome. Generalized linear models (GLMs) are an extension of linear regression, which estimate a linear relationship between explanatory covariates and the outcome. GLMs allow for non-linear associations by relating a linear model to a response variable through a link function. Logistic regression is one form of GLM, which models the probability of a binary outcome (for instance, whether an individual tests positive or negative for a particular infection). As probability p of testing positive is bounded between 0 and 1, it is transformed via a logit link function to the log

odds scale, where $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$. Here, the outcome, for instance test positivity, y is binomially distributed with probability of testing positive p . The equation for this is given by:

$$y \sim \text{Binomial}(1, p) \quad 2.10$$

$$\text{logit}(p) = \alpha + \sum \beta_k X_k \quad 2.11$$

Here, α is the intercept, X represents k explanatory variables and β represents their coefficients. Similarly, disease count data (for instance, weekly surveillance data of reported cases or hospitalisations) can be modelled assuming a Poisson distribution. One limitation of Poisson regression is that it assumes variance is equal to the mean; however, overdispersion, where variance is greater than the mean, is common in epidemiological data. In these cases, a quasi-Poisson, which assumes variance is a linear function of the mean, or negative binomial regression, which assumes variance is a quadratic function of the mean and estimates a dispersion parameter, can be used. In this thesis I have assumed surveillance count data follows a negative binomial distribution as this accounts for the overdispersion seen in epidemiological data, in a flexible manner. Here:

$$y_t \sim \text{NegBin}(\mu_t, \kappa) \quad 2.12$$

$$\log(\mu_t) = \alpha + \sum \beta_k X_{k,t} \quad 2.13$$

Generalised additive models extend this framework by allowing the linear predictor to be defined by smooth functions of some or all predictor variables and are able to capture non-linear predictor-outcome relationships [25].

Hierarchical or mixed-effects models (such generalised linear mixed models, GLMMs or generalised additive mixed models, GAMMs) allow predictors to be organised into groups where the relationship between the predictor and response can vary across groups [25–27]. A typical example of this would be to account for region-specific effects in a population model, or individual-level effects if a study involves repeat measurements from the same individuals. Random effects can be used to model varying intercepts between groups, or varying slopes, where

estimated regression coefficients can vary by group [26]. For time-series based regression models, random effects can be used to incorporate seasonality or interannual effects. Similarly, spatial or spatiotemporal models can take into account the underlying spatial structure of data, based on a range of assumptions about the dynamics across space [28]. Statistical methods are often limited by the extent and quality of available data, which determine the statistical power. This can make them less useful when considering policy questions around emerging or re-emerging diseases with little historical data.

2.2.4 Applications of modelling to COVID-19 and dengue virus

COVID-19

During the COVID-19 pandemic, a wide variety of modelling techniques were applied to understand key epidemiological parameters and inform public health decision making through scenario projections and forecasts [10]. Examples of these diverse approaches include the use of branching process models to estimate overdispersion in SARS-CoV-2 transmission, network modelling to estimate the efficacy of local control strategies, renewal equations to estimate R_t or individual-level models to model the impact of testing and quarantine strategies [29–33]. Compartmental models were used extensively, particularly when fitting to data at a national level to model options for epidemic response at a national level. These frameworks often adapted model structures originally developed for other respiratory viruses, particularly pandemic influenza. As part of this thesis, I adapt *covidm*, a transmission dynamic model originally designed by Davies and colleagues to quantify the impact of NPIs in the UK [34–36]. This was used in conjunction with modelling from other UK universities to generate evidence for the Scientific Advisory Group for Emergencies (SAGE) to advise the government on epidemic response.

Catalytic modelling is more typically applied to endemic diseases, where age-stratified seroprevalence reflects transmission intensity over time, as opposed to exposure during a large outbreak. However, Rees and colleagues used catalytic modelling to estimate the duration of immunity following infection for seasonal coronaviruses, thereby providing insight into potential dynamics for SARS-CoV-2 early on in the pandemic [37]. Statistical modelling was also used to

address a wide range of questions during the pandemic from evaluating efficacy in trials for pharmaceutical interventions, estimating key epidemiological parameters such as the incubation period from delayed and censored data, understanding the association between COVID-19 incidence with climatic variables, and short-term forecasting of hospital admissions [38–40].

Dengue

Mechanistic models for dengue are often derived from the Ross-Macdonald model of mosquito-borne pathogen transmission [41–43]. This describes R_0 for vector-borne diseases as a function of; mosquito-to-human ratio, human biting rate, probability of human infection, probability of vector infection, the extrinsic incubation period (or the time taken for the pathogen to develop inside the host) and the vector survival rate. Agent-based models have been used to model *Aedes* and dengue spread through a population, and to assess the potential impact of interventions such as vector control or vaccination, often at a city or local level [44]. Compartmental models of dengue transmission typically include compartments for both host and vector populations, as well as potentially modelling multiple serotypes, and have been used to understand theoretical behaviour of dengue transmission, investigate vector-control strategies or vaccine impact, often over larger spatial scales [45–51]. Within the field of dengue forecasting, time-series based regression approaches are the most regularly employed approach, with fewer examples of mechanistic and machine learning models [52–58].

In 2015, an open forecasting challenge, the Dengue Forecasting Project, was launched. This compared predictions from 16 forecasting teams for two dengue endemic locations, Iquitos in Peru and San Juan in Puerto Rico [58]. This study found that while forecasts performed well mid-season to provide situational awareness, forecast skill was generally lowest early in a season and for seasons with large outbreaks, both situations where accurate forecasts would have high operational utility. Additionally, mechanistic models had lower average forecast skill than statistical approaches that did not explicitly include biological processes underlying transmission (91).

2.2.5 Model fitting

Fitting models to data can be performed using Bayesian or frequentist inference. Frequentist inference considers probabilities in terms of the frequency of an event in a very large sample. Contrastingly, Bayesian inference incorporates a prior belief in the probability of an event, which is then updated with data. Until recently, fitting complex Bayesian spatiotemporal statistical models and mechanistic models to data was challenging due to the long computation times required for statistical inference of model parameters. However, techniques such as Markov chain Monte Carlo (MCMC) and alternative methods have allowed for more efficient estimation of model parameters, increasing computational tractability of complex models. Bayesian inference is based on Bayes' theorem, whereby the posterior distribution is proportional to the likelihood of the data y given the parameters θ , multiplied by the prior distribution $\pi(\theta)$.

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \quad 2.14$$

Until recently, Bayesian inference could be challenging, or impossible to perform, as the normalizing constant $\pi(y)$ required integrating the likelihood with respect to the prior, and was often intractable. The key innovation of MCMC algorithms is to instead consider the ratio of probabilities of proposal distributions (such that the normalizing constant cancels out) and to sample repeatedly, building a Markov chain with an equilibrium distribution equal to the posterior. Metropolis-Hastings MCMC algorithms follow the following steps to sample from the posterior distribution. First, a starting point is selected such that $\theta^{t-1} = \theta_0$. Then a new sample is proposed from a proposal or jumping distribution $g(\theta'|\theta^{t-1})$:

$$\theta' = \theta^{t-1} + \epsilon | \epsilon \sim Q \quad 2.15$$

The new proposal is then accepted with probability a , where

$$a = \min\left(1, \frac{p(\theta'|y)/g(\theta'|\theta^{t-1})}{p(\theta^{t-1}|y)/g(\theta^{t-1}|\theta')}\right) \quad 2.16$$

If the proposal is rejected, the current proposal is used again and these steps are repeated until convergence is achieved. Typically, a multivariate normal distribution is used as a proposal (or jumping) distribution, and a common issue with traditional MH-MCMC is specifying the covariance matrix, which determines the scale and orientation of the proposed jump in parameter space.

Hamiltonian Monte Carlo (HMC) is an MCMC method which uses derivatives of the sampled density function (similar to the concept of ‘momentum’ in physics) to move through the target distribution [26,59]. This allows for more efficient exploration of the posterior distribution than the random walk behaviour exhibited in Metropolis-Hasting MCMC algorithms, and is implemented in the programming language Stan [60]. For MCMC based algorithms it is important to assess convergence. This can be done by assessing chain mixing through trace plots as well as through the Gelman-Rubin (R-hat) statistic, which compares variation between chains to variation within chains and the effective sample size (ESS), giving the number of independent samples. Larger values of ESS are better while R-hat < 1.1 is typically indicative of convergence [26].

Similarly, Differential Evolution Markov Chain (DE-MC) offers a similar improvement on traditional MH-MCMC [61]. Here, chains are run in parallel and can learn from each other and the jump for one chain is chosen using the remaining chains. First, two of the other chains are chosen at random and then the difference of the vectors of these chains is taken and multiplied by a factor. This difference vector then determines the scale and orientation of the proposal for the original chain. This allows for the parameter space to be explored more efficiently, particularly in cases of collinearity between parameters.

Another method for Bayesian inference is the integrated nested Laplace approximation, INLA [62]. Instead of estimating the joint posterior distribution $\pi(\theta|y)$ this focuses on approximating univariate posterior distributions $\pi(\theta_i|y)$ and requires models to be expressed as latent Gaussian Markov random fields. This is a computationally efficient method of Bayesian inference, particularly suited to large spatiotemporal models.

Throughout this thesis information criteria are calculated to inform Bayesian model selection. In particular, we use the widely applicable information criteria, WAIC, and the deviance information criteria, DIC [63,64]. These criteria aim to approximate out-of-sample prediction error, correcting for biases introduced through using in-sample model fit. The WAIC computes the log pointwise predictive density, $lppd$, adjusted by a penalty term proportional to the variance in the posterior predictions. This is given by:

$$WAIC(y, \theta) = -2(lppd - \sum_i var_{\theta} \log(p(y_i|\theta))) \quad 2.17$$

where y is the data and θ is the posterior distribution.

The DIC is defined by:

$$DIC = -2\log(p(y|\hat{\theta})) + 2p_{DIC} \quad 2.19$$

where $2p_{DIC}$ is the effective number of parameters [26].

References

1. Roy V, Fischinger S, Atyeo C, Slein M, Loos C, Balazs A, et al. SARS-CoV-2-specific ELISA development. *Journal of Immunological Methods*. 2020 Sep 1;484–485:112832.
2. Nilles EJ, Karlson EW, Norman M, Gilboa T, Fischinger S, Atyeo C, et al. Evaluation of Three Commercial and Two Non-Commercial Immunoassays for the Detection of Prior Infection to SARS-CoV-2. *J Appl Lab Med*. 2021 Jul 1;6(6):1561–70.
3. Nilles EJ, Paulino CT, de St. Aubin M, Restrepo AC, Mayfield H, Dumas D, et al. SARS-CoV-2 seroprevalence, cumulative infections, and immunity to symptomatic infection – A multistage national household survey and modelling study, Dominican Republic, June–October 2021. *The Lancet Regional Health - Americas*. 2022 Dec 1;16:100390.
4. Ainsworth M, Andersson M, Auckland K, Baillie JK, Barnes E, Beer S, et al. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. *The Lancet Infectious Diseases* [Internet]. 2020 Sep 23 [cited 2020 Sep 30];0(0). Available from: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30634-4/abstract](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30634-4/abstract)
5. Ministerio de Salud Pública, de la República Dominicana. Boletines sobre el COVID-19 [Internet]. [cited 2022 Nov 18]. Available from: https://www.msp.gob.do/web/?page_id=6948
6. National Oceanic and Atmospheric administration (NOAA). Climate Prediction Center - Monitoring & Data Index [Internet]. [cited 2024 Jan 15]. Available from: <https://www.cpc.ncep.noaa.gov/data/>
7. Lee KS, Lai YL, Lo S, Barkham T, Aw P, Ooi PL, et al. Dengue Virus Surveillance for Early Warning, Singapore - Volume 16, Number 5—May 2010 - *Emerging Infectious Diseases journal* - CDC. [cited 2022 Oct 22]; Available from: https://wwwnc.cdc.gov/eid/article/16/5/09-1006_article
8. Cox DR. *Principles of Statistical Inference* [Internet]. Cambridge: Cambridge University Press; 2006 [cited 2024 Jun 12]. Available from: <https://www.cambridge.org/core/books/principles-of-statistical-inference/BCD3734047D403DF5352EA58F41D3181>
9. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, Angelis DD, Dye C, et al. Modeling

- infectious disease dynamics in the complex landscape of global health. *Science* [Internet]. 2015 Mar 13 [cited 2022 Feb 1]; Available from: <https://www.science.org/doi/abs/10.1126/science.aaa4339>
10. Cori A, Kucharski AJ. Inference of epidemic dynamics in the COVID-19 era and beyond [Internet]. OSF; 2024 [cited 2024 Jun 14]. Available from: <https://osf.io/mg497>
 11. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals* [Internet]. Princeton University Press; 2008 [cited 2024 Jun 12]. Available from: <https://www.jstor.org/stable/j.ctvcvcm4gk0>
 12. Shmueli G. To Explain or to Predict? *Statistical science*. 2010;25(3):289–310.
 13. Hofman JM, Watts DJ, Athey S, Garip F, Griffiths TL, Kleinberg J, et al. Integrating explanation and prediction in computational social science. *Nature*. 2021 Jul;595(7866):181–8.
 14. Gibb R, Colón-González FJ, Lan PT, Huong PT, Nam VS, Duoc VT, et al. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nat Commun*. 2023 Dec 11;14(1):8179.
 15. Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*. 1997 Jan;115(772):700–21.
 16. Eames KTD, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. *PLOS Computational Biology*. 2012 Mar 8;8(3):e1002425.
 17. Gimma A, Munday JD, Wong KLM, Coletti P, van Zandvoort K, Prem K, et al. Changes in social contacts in England during the COVID-19 pandemic between March 2020 and March 2021 as measured by the CoMix survey: A repeated cross-sectional study. *PLoS Med*. 2022 Mar;19(3):e1003907.
 18. Anderson RM, May RM. Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *J Hyg (Lond)*. 1985 Jun;94(3):365–436.
 19. Wallinga J, Teunis P, Kretzschmar M. Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *American Journal of Epidemiology*. 2006 Nov 15;164(10):936–44.

20. Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface*. 2010 Jun 6;7(47):873–85.
21. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine*. 2008 Mar 25;5(3):e74.
22. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*. 2017 Sep 12;13(9):e1005697.
23. Muench H. Derivation of Rates from Summation Data by the Catalytic Curve. *Journal of the American Statistical Association*. 1934;29(185):25–38.
24. Muench H. *Catalytic Models in Epidemiology* [Internet]. Harvard University Press; 2013 [cited 2024 Jun 24]. Available from: <https://www.degruyter.com/document/doi/10.4159/harvard.9780674428928/html>
25. Wood SN. *Generalized Additive Models: An Introduction with R*, Second Edition. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2017. 496 p.
26. Rubin AG, John B, Carlin, Hal S, Stern, David B, Dunson, Aki Vehtari, Donald B. *Bayesian Data Analysis*. 3rd ed. New York: Chapman and Hall/CRC; 2015. 675 p.
27. Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*. 2019 May 27;7:e6876.
28. Lee SA, Jarvis CI, Edmunds WJ, Economou T, Lowe R. Spatial connectivity in mosquito-borne disease models: a systematic review of methods and assumptions. *Journal of The Royal Society Interface*. 2021 May 26;18(178):20210096.
29. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*. 2020;5:67.
30. Firth JA, Hellewell J, Klepac P, Kissler S, Kucharski AJ, Spurgin LG. Using a real-world network to model localized COVID-19 control strategies. *Nat Med*. 2020 Oct;26(10):1616–22.
31. Pung R, Firth JA, Spurgin LG, Lee VJ, Kucharski AJ. Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events. *Nat Commun*. 2022 Apr 12;13(1):1956.
32. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the

- time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res.* 2020 Jun 1;5:112.
33. Quilty BJ, Clifford S, Hellewell J, Russell TW, Kucharski AJ, Flasche S, et al. Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study. *The Lancet Public Health.* 2021 Mar 1;6(3):e175–83.
 34. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health.* 2020 Jul 1;5(7):e375–85.
 35. Davies NG, Barnard RC, Jarvis CI, Russell TW, Semple MG, Jit M, et al. Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. *The Lancet Infectious Diseases.* 2021 Apr 1;21(4):482–92.
 36. Barnard RC, Davies NG, Jit M, Edmunds WJ. Modelling the medium-term dynamics of SARS-CoV-2 transmission in England in the Omicron era. *Nat Commun.* 2022 Aug 19;13(1):4879.
 37. Rees EM, Waterlow NR, Lowe R, Kucharski AJ. Estimating the duration of seropositivity of human seasonal coronaviruses using seroprevalence studies. *Wellcome Open Res.* 2021 Dec 21;6:138.
 38. Overton CE, Stage HB, Ahmad S, Curran-Sebastian J, Dark P, Das R, et al. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. *Infectious Disease Modelling.* 2020 Jan 1;5:409–41.
 39. Nottmeyer L, Armstrong B, Lowe R, Abbott S, Meakin S, O'Reilly KM, et al. The association of COVID-19 incidence with temperature, humidity, and UV radiation – A global multi-city analysis. *Science of The Total Environment.* 2023 Jan 1;854:158636.
 40. Meakin S, Abbott S, Bosse N, Munday J, Gruson H, Hellewell J, et al. Comparative assessment of methods for short-term forecasts of COVID-19 hospital admissions in England at the local level. *BMC Medicine.* 2022 Feb 21;20(1):86.
 41. Report on the prevention of malaria in Mauritius / by Ronald Ross. [Internet]. Wellcome Collection. [cited 2024 Jun 14]. Available from: <https://wellcomecollection.org/works/rjj6m9qm>
 42. Ross SR. The Prevention of malaria. J. Murray; 1910. 774 p.
 43. Macdonald G. The analysis of equilibrium in malaria. *Trop Dis Bull.* 1952 Sep;49(9):813–

- 29.
44. Perkins TA, Jr RCR, España G, Bosch QA ten, Verma A, Liebman KA, et al. An agent-based model of dengue virus transmission shows how uncertainty about breakthrough infections influences vaccination impact projections. *PLOS Computational Biology*. 2019 Mar 20;15(3):e1006710.
 45. Esteva L, Vargas C. Analysis of a dengue disease transmission model. *Math Biosci*. 1998 Jun 15;150(2):131–51.
 46. Newton EA, Reiter P. A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ULV) insecticide applications on dengue epidemics. *Am J Trop Med Hyg*. 1992 Dec;47(6):709–20.
 47. Ferguson N, Anderson R, Gupta S. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc Natl Acad Sci U S A*. 1999 Jan 19;96(2):790–4.
 48. Adams B, Holmes EC, Zhang C, Mammen MP, Nimmannitya S, Kalayanarooj S, et al. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Proc Natl Acad Sci U S A*. 2006 Sep 19;103(38):14234–9.
 49. Wearing HJ, Rohani P. Ecological and immunological determinants of dengue epidemics. *Proc Natl Acad Sci U S A*. 2006 Aug 1;103(31):11802–7.
 50. Ferguson NM, Rodríguez-Barraquer I, Dorigatti I, Mier-y-Teran-Romero L, Laydon DJ, Cummings DAT. Benefits and risks of the Sanofi-Pasteur dengue vaccine: Modeling optimal deployment. *Science*. 2016 Sep 2;353(6303):1033–6.
 51. Group WVDVM. Assessing the Potential of a Candidate Dengue Vaccine with Mathematical Modeling. *PLOS Neglected Tropical Diseases*. 2012 Mar 27;6(3):e1450.
 52. Colón-González FJ, Bastos LS, Hofmann B, Hopkin A, Harpham Q, Crocker T, et al. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Medicine*. 2021 Mar 4;18(3):e1003542.
 53. Lowe R, Coelho CA, Barcellos C, Carvalho MS, Catão RDC, Coelho GE, et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. *Elife*. 2016 Feb 24;5:e11285.
 54. Lowe R, Stewart-Ibarra AM, Petrova D, García-Díez M, Borbor-Cordova MJ, Mejía R, et al. Climate services for health: predicting the evolution of the 2016 dengue season in Machala,

- Ecuador. *The Lancet Planetary Health*. 2017 Jul 1;1(4):e142–51.
55. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Medicine*. 2018 Aug 6;16(1):129.
 56. Mincham G, Baldock KL, Rozilawati H, Williams CR. Development of a mechanistic dengue simulation model for Guangzhou. *Epidemiology & Infection*. 2019 Jan;147:e125.
 57. Henderson AD. *Mathematical Modelling of Arbovirus Outbreak Dynamics in Fiji and the Wider Pacific* [Internet] [doctoral]. London School of Hygiene & Tropical Medicine; 2020 [cited 2024 Jun 18]. Available from: <https://researchonline.lshtm.ac.uk/id/eprint/4660713/>
 58. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *PNAS*. 2019 Nov 26;116(48):24268–74.
 59. Betancourt MJ, Girolami M. *Hamiltonian Monte Carlo for Hierarchical Models* [Internet]. arXiv; 2013 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/1312.0906>
 60. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual* [Internet]. Stan. 2024 [cited 2024 Jun 13]. Available from: <https://mc-stan.org/users/documentation/>
 61. Braak CJFT. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Stat Comput*. 2006 Sep 1;16(3):239–49.
 62. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009;71(2):319–92.
 63. Watanabe S. A widely applicable Bayesian information criterion. *J Mach Learn Res*. 2013 Mar 1;14(1):867–97.
 64. Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002 Oct 1;64(4):583–639.

Chapter 3

Investigating immunity to SARS-CoV-2 in a community seroepidemiological cohort in the United States

The extent and duration of immunity following acute viral infection is an important determinant of potential epidemic dynamics, as well as optimal disease control strategies. Longitudinal serological studies can provide insights into antibody dynamics over time which, when coupled with routine testing for current infection, can be used to estimate reinfection risk. Analysis for this chapter began in late 2020, when there was limited understanding of the effect of SARS-CoV-2 antibodies on reinfection risk or the implications of post-infection immunity on transmission in the population. The question of whether SARS-CoV-2 infection would provide long-lasting immunity in the majority of individuals (similar to SARS-CoV-1) or immunity which waned rapidly (similar to human seasonal coronaviruses) had important implications for public health strategy, such as the intensity and frequency of non-pharmaceutical interventions, as well as the potential for vaccine development. In this study, I aimed to use seroepidemiological cohort data from April 2020 to February 2021 in the United States to estimate SARS-CoV-2 reinfection risk. I also discuss difficulties in estimating reinfection risk in real-time in the early phase of an outbreak, and address these by adjusting for individual-level variation in infection risk and considering the impact of population-level epidemic dynamics in the study period.

This paper was published in PLOS Biology in February 2022. Supplementary material is included in Appendix B.



London School of Hygiene & Tropical Medicine
 Keppel Street, London WC1E 7HT
 T: +44 (0)20 7299 4646
 F: +44 (0)20 7299 4656
 www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1802674	Title	Miss
First Name(s)	Emilie		
Surname/Family Name	Finch		
Thesis Title	Modelling the role of immunity, climate and behaviour in viral outbreak dynamics and control		
Primary Supervisor	Adam Kucharski		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	PLOS Biology		
When was the work published?	10th February 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

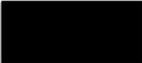
SECTION C – Prepared for publication, but not yet published

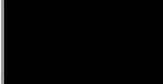
Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>Author list: Emilie Finch, Rachel Lowe, Stephanie Fischinger, Michael de St Aubin, Sameed M. Siddiqui, Diana Dayal, Michael A. Loesche, Justin Rhee, Samuel Beger, Yiyuan Hu, Matthew J. Gluck, Benjamin Mormann, Mohammad A. Hasdianda, Elon R. Musk, Galit Alter, Anil S. Menon, Eric J. Nilles, Adam J. Kucharski, on behalf of the CMMID COVID-19 working group and the SpaceX COVID-19 Cohort Collaborative</p> <p>I worked on: conceptualisation, data curation, formal analysis, methodology, visualization, writing original draft and review and editing of submitted manuscript.</p> <p>Copyright: © 2022 Finch et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.</p>
---	---

SECTION E

Student Signature	
Date	08/07/2024

Supervisor Signature	
Date	12/07/2024

3.1 Abstract

Identifying the potential for SARS-CoV-2 reinfection is crucial for understanding possible long-term epidemic dynamics. We analysed longitudinal PCR and serological testing data from a prospective cohort of 4411 US employees in four states between April 2020 and February 2021. We conducted a multivariable logistic regression investigating the association between baseline serological status and subsequent PCR test result in order to calculate an odds ratio for reinfection. We estimated an odds ratio for reinfection ranging from 0.14 (95% CI: 0.019 - 0.63) to 0.28 (95% CI: 0.05-1.1), implying that the presence of SARS-CoV-2 antibodies at baseline is associated with around 72-86% reduced odds of a subsequent PCR positive test based on our point estimates. This suggests that primary infection with SARS-CoV-2 provides protection against reinfection in the majority of individuals, at least over a sixth month time period. We also highlight two major sources of bias and uncertainty to be considered when estimating the relative risk of reinfection, confounders and the choice of baseline time point, and show how to account for both in reinfection analysis.

3.2 Introduction

The rapid global spread of COVID-19 throughout 2020 occurred as a result of the introduction of a highly transmissible virus, SARS-CoV-2, into populations with little pre-existing immunity [1]. Identifying the extent and duration of protective immunity afforded by natural infection is therefore of crucial importance for understanding possible long-term epidemic dynamics of SARS-CoV-2 [2].

Studies have estimated that over 95% of symptomatic COVID-19 cases develop antibodies against SARS-CoV-2, with most individuals developing antibodies within three weeks of symptom onset [3,4]. Several serological studies have also characterised individual-level immune dynamics, with some finding evidence for antibody waning and others for sustained antibody responses over several months [5–10]. Antibody kinetics are thought to vary between individuals and are possibly associated with severity of illness, where asymptomatic or mildly symptomatic individuals may

develop lower levels of antibodies that wane more rapidly [3,7,11]. While neutralising antibodies are thought to be associated with protection from reinfection, there are still limited studies on the impact of post-infection seropositivity on future reinfection risk [12]. Confirmed cases of reinfection with SARS-CoV-2 have been reported since August 2020 [13]. However, existing large studies examining the relative risk of reinfection in antibody positive individuals have typically involved specific cohorts who may not be representative of the wider community, such as closed communities or healthcare worker cohorts [14–17]. To evaluate the relative risk of SARS-CoV-2 infection and reinfection over time, we analysed PCR and serological testing data from a prospective cohort of SpaceX employees in the USA between April 2020 and February 2021 [18,19].

3.3 Results

Of 4411 individuals enrolled, 309 individuals tested seropositive during the study period (Figure 1). This resulted in an overall adjusted percentage ever-seropositive of 8.2% (95% CI: 7.3-9.1%) by the end of August 2020, after the final round of serological testing (Figure 2B). Here, imperfect test sensitivity and specificity were adjusted for using the Rogan-Gladen correction [20]. We defined a possible reinfection as a new positive PCR test more than 30 days after initial seropositive result. This identified 14 possible reinfections with a median time of 66.5 days between initial seropositive test and PCR positive test (Figure 2C).

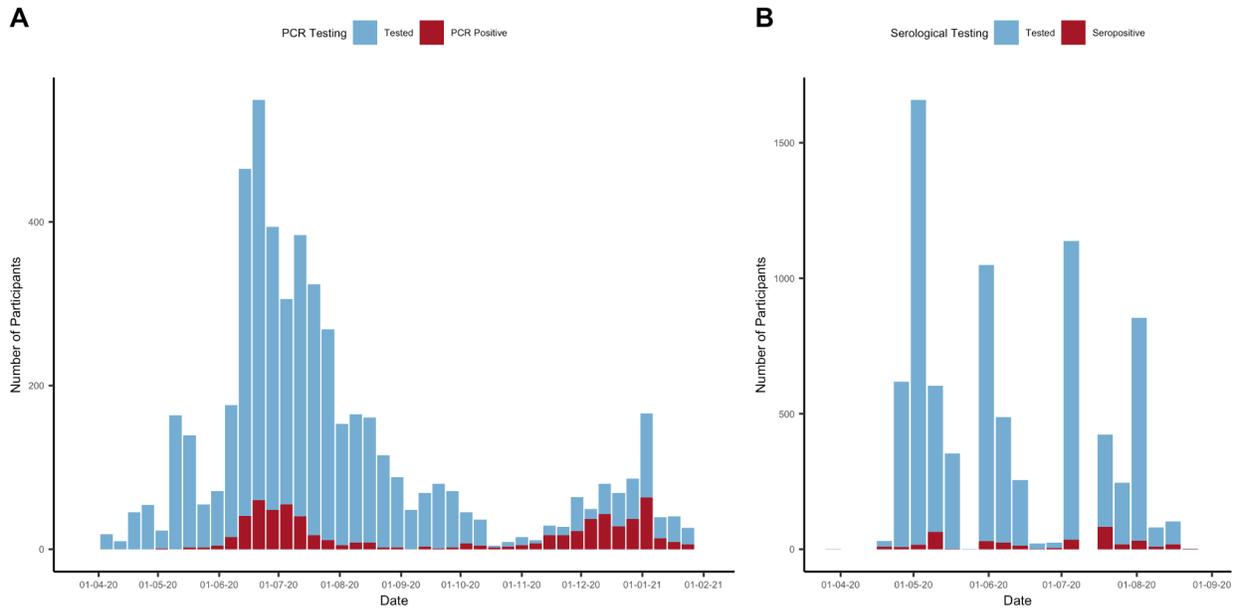


Figure 3.1 PCR and serological tests in seroepidemiological cohort

A) Number of PCR tests and PCR positive tests in the cohort between 5th April 2020 and 31st January 2021 from 3296 participants. B) Number of serological tests and seropositive tests between 29th March 2020 and 23rd August 2020 from 4411 participants.

3.3.1 SARS-CoV-2 infection and reinfection

We estimated the odds ratio for SARS-CoV-2 reinfection using multivariable logistic regression, to adjust for any background individual-level variation in the risk of infection (see Methods). This required us to choose a cut-off week in order to define baseline seroprevalence and the subsequent observation period for PCR testing. To examine how our estimate for the odds ratio for reinfection varied depending on the cut-off week chosen, we repeated the analysis using every possible cut-off week.

We considered that the most robust estimation of the odds ratio for reinfection would occur mid-epidemic when using cut-off weeks in between two ‘waves’ of the epidemic seen in the study cohort. We validated this methodological assumption by conducting a simulation study (see Supplementary Materials).

We defined a mid-epidemic period in between two epidemic waves where PCR positivity in the study cohort was below the WHO specified threshold of 5%, which occurred between 26th July 2020 and 27th September 2020 (see Figure 2A). During these cut-off weeks, estimates of the odds ratio for reinfection (Figure 2D) ranged from 0.14 (95% CI: 0.019 - 0.63) to 0.28 (95% CI: 0.05-1.1). Our point estimates suggest that the presence of SARS-CoV-2 antibodies confers around 72-86% protection against reinfection with SARS-CoV-2, at least over a six-month period. As a sensitivity analysis, we conducted the same analysis but excluding records where individuals had recorded a specific trigger reason for testing such as symptom onset or potential exposure (and so reflecting individuals tested at random). Considering the weeks between 26th July 2020 and 27th September 2020, we found estimates of the odds ratio for reinfection ranged from 0.18 (95% CI: 0.024-0.80) to 0.36 (95% CI: 0.06-1.5).

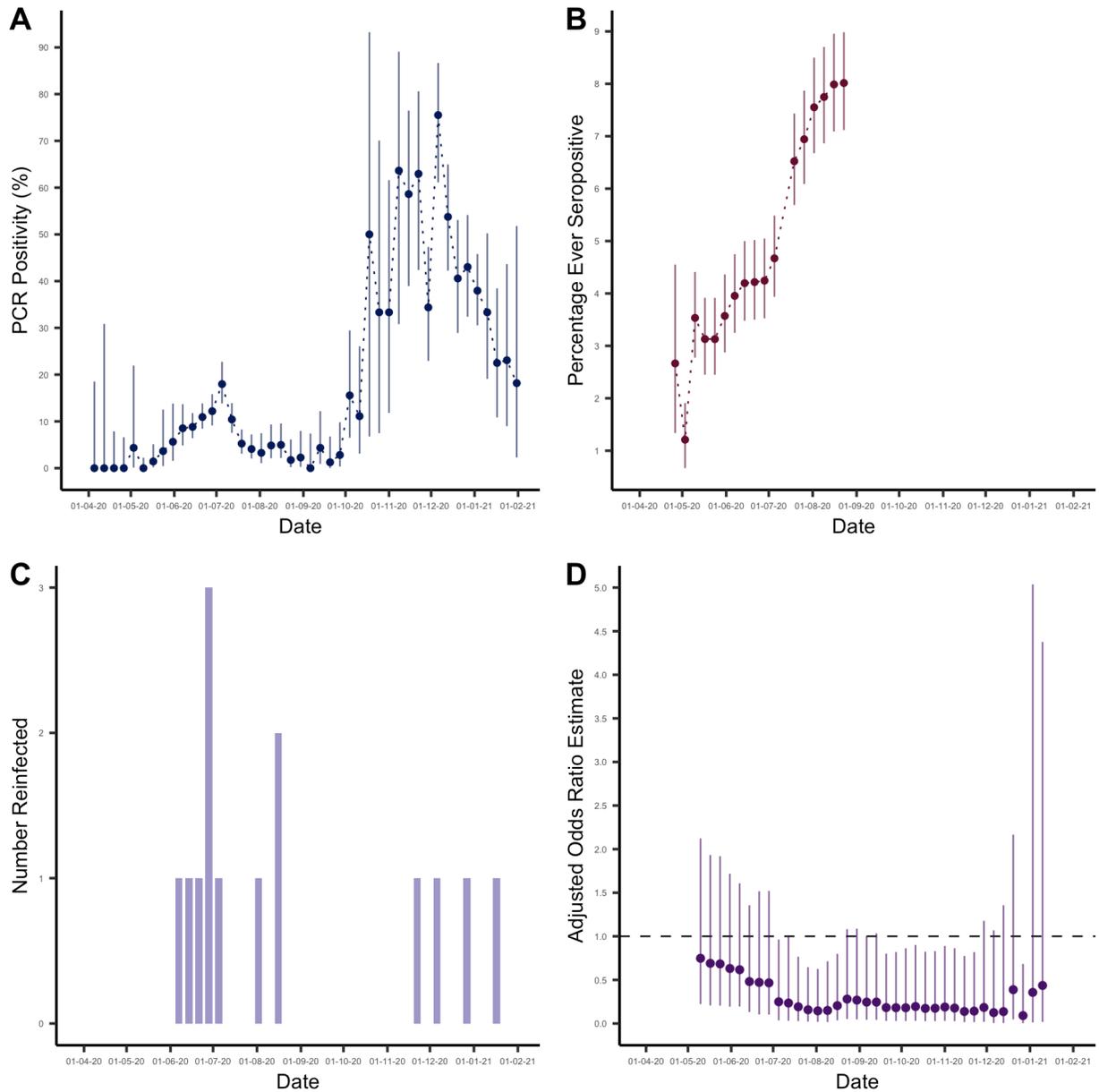


Figure 3.2 PCR positivity, seropositivity and reinfection risk

A) PCR positivity (%) in the cohort between 5th April 2020 and 31st January 2021. B) Percentage ever seropositive in the cohort (number ever seropositive/ cumulative number enrolled) between 29th March 2020 and 23rd August 2020. Note that the percentage ever positive decreases initially as participants continue to be enrolled in the study. C) Number of possible reinfections in cohort over time (defined as a new positive PCR test more than 30 days after initial seropositive result). D) Odds ratio estimates comparing odds of reinfection in the seropositive group with odds of primary infection in the seronegative group, estimated using logistic regression and adjusted for potential confounders. The estimates are presented with

their associated 95% confidence intervals and with the cut-off week used to define baseline seroprevalence on the x-axis.

In the adjusted analyses, odds ratio estimates for reinfection converged to similar values for cut-offs spanning a period after the first peak of infection in early July. By this point, sufficient numbers of participants had been both recruited and tested seropositive (see Figure 2B) that we had enough data to distinguish infection dynamics in seropositive and seronegative groups. Adjusted odds ratio estimates for reinfection then lost precision when using late cut-off weeks from mid-December onwards due to increasingly small numbers of participants experiencing PCR infection after the cut-off point, consistent with our simulation study (see Supplementary Materials).

Unadjusted odds ratio estimates tended to overestimate the odds ratio for reinfection compared with primary infection, particularly when using early cut-off weeks (Figure 3). Notably, with early cut-off weeks the unadjusted analysis estimated a higher odds of reinfection compared to primary infection, albeit with wide confidence intervals. This is the result of a subset of individuals who are at higher risk of initial seroconversion (who would be included in analyses at earlier time thresholds) and also at higher risk of later reinfection, giving a biased estimate of the association between antibodies and subsequent infection when using earlier cut-off weeks.

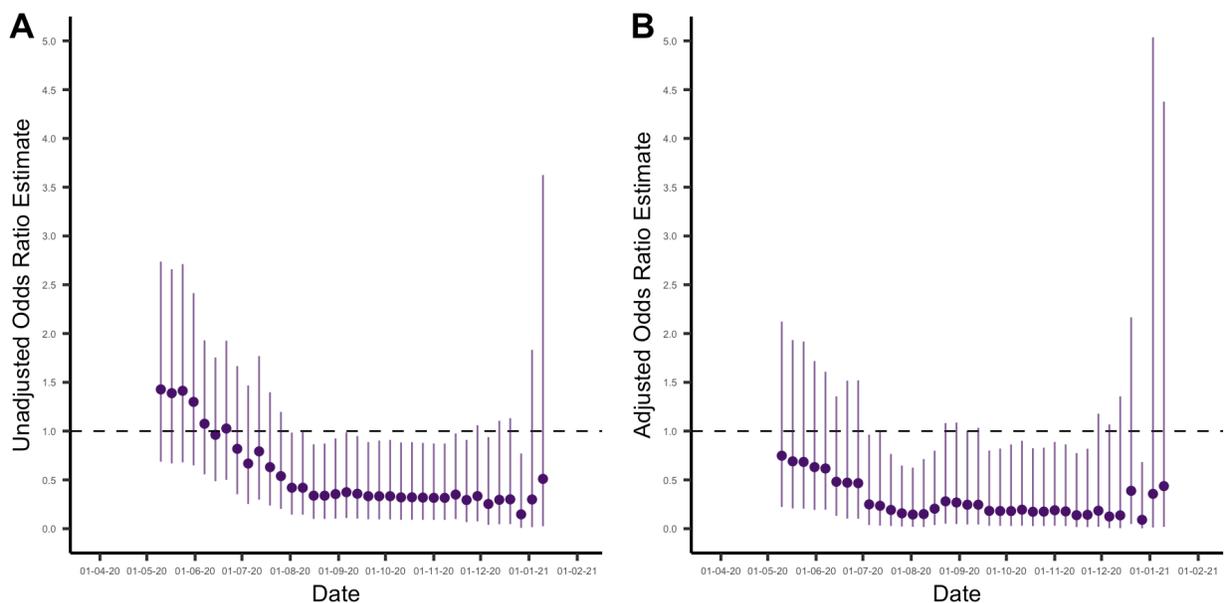


Figure 3.3 Unadjusted and adjusted odds of reinfection

A) Unadjusted odds ratio estimates comparing odds of reinfection in the seropositive group with odds of primary infection in the seronegative group. The estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline seroprevalence on the x-axis. B) Odds ratio estimates comparing odds of reinfection in the seropositive group with odds of primary infection in the seronegative group, estimated using logistic regression and adjusted for potential confounders. The estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline seroprevalence on the x-axis.

3.4 Discussion

We identified 14 possible reinfections out of 309 seropositive individuals in the prospective seroepidemiological cohort between April 2020 and February 2021, estimating an odds ratio for reinfection ranging from 0.14 (95% CI: 0.019 - 0.63) to 0.28 (95% CI: 0.05-1.1). This provides evidence that primary infection with SARS-CoV-2 results in protection against reinfection in the majority of individuals, at least over a sixth month time period. Our findings are broadly consistent with estimates of 0.17 (95% CI 0.13-0.24) odds ratio (14) and 0.11 (0.03-0.44) incidence rate ratio (15) for healthcare workers, 0.18 (0.11-0.28) incidence rate ratio for military recruits (16) and 0.195 (95% CI 0.155–0.246) incidence rate ratio from a Danish population-level study [21].

Our analysis addressed two key sources of bias and uncertainty in estimating the relative risk of reinfection. First, confounders may inflate estimates; if a specific subset of the cohort is at higher risk of infection (e.g. due to underlying health conditions or increased risk of exposure), these participants will be more likely to be both initially seropositive and to have a subsequent reinfection. Second, the time period considered could increase uncertainty; defining the baseline seroprevalence at an early time point means few will be seropositive, whereas defining it at a later point means there is less time to observe possible reinfections. We accounted for these two factors by adjusting for key confounders to calculate an adjusted odds ratio for reinfection. We then investigated how changing the cut-off date to define baseline seroprevalence impacted the accuracy of the adjusted odds ratio calculated. We assumed that for a two-wave epidemic scenario, a cut-off week in the period in between the two waves of infection risk would result in the most

robust estimates of the odds ratio for reinfection, which we validated using a simulation study (see Supplementary Materials). This suggests that the robustness of estimates of the relative risk of reinfection will be sensitive to the study period chosen, relative to population-level epidemic dynamics.

There are several limitations to the underlying data that should be considered when interpreting these findings. This prospective cohort was recruited opportunistically from employees at one US company and is unlikely to be representative of the general population. However, as we did not identify any workplace outbreaks, infections in this cohort are likely to be more reflective of community transmission than in health-care worker cohorts or other specialised populations. Additionally, we only considered possible reinfections (as opposed to probable or true reinfections). As possible reinfections did not meet a stringent case definition, such as confirmation through genomic sequencing, they may include cases of prolonged viral shedding following an initial infection. This would result in an overestimation of the odds ratio for reinfection and so our analysis reflects the minimum possible effect of antibodies on future SARS-CoV-2 infection risk according to the mechanism investigated here. Finally, the date of infection among seropositive participants is unknown, limiting inference on exact duration of protection.

As well as quantifying the relative risk of reinfection over a six-month period among a prospectively followed workplace population, our study highlights the importance of accounting for both individual-level heterogeneity in infection risk and population-level variation in epidemic dynamics when assessing the potential for reinfections.

3.5 Methods

3.5.1 Seroepidemiological cohort description

We used data from a seroepidemiological study of US employees at SpaceX, as described previously [19]. In brief, this study involved employees from seven work locations in California, Florida, Texas and Washington State, with ages ranging from 18-71. 4411 employees volunteered

to participate in the study and were enrolled from ~8400 total employees. All employees were invited to participate by email and there were no exclusion criteria. Study participants were offered SARS-CoV-2 IgG receptor-binding domain (RBD) antibody testing with an in-house ELISA assay with 82.4% sensitivity and 99.6% specificity [22]. Serological samples were taken during four rounds of testing between April - September 2020. A questionnaire including demographic, symptom and exposure information was conducted at enrolment, and with each round of serological testing. Individuals continued to be enrolled throughout the study period, and around half of the total participants (48%) were tested at more than one time point. Participants occupied a range of job positions within SpaceX including office-based and factory-based jobs. Additionally, symptomatic and asymptomatic PCR testing were widely available for employees using the Infinity BiologiX (IBX) TaqPath rRT-PCR assay, with data available from April 2020 - January 2021. Employees could request a test for any reason and testing was also specifically performed for symptomatic individuals, individuals with potential exposure and mission critical employees. Both serology and PCR testing data were available for 1800 individuals.

3.5.2 Ethics Statement

The study protocol was approved by the Western Institutional Review Board. The use of de-identified data and biological samples was approved by the Mass General Brigham Healthcare Institutional Review Board. Secondary data analysis was approved by the LSHTM Observational Research Ethics Committee (ref 22466). All participants provided written informed consent.

3.5.3 Statistical Analysis

To estimate the odds ratio for SARS-CoV-2 reinfection, we conducted multivariable logistic regression analysis investigating the association between baseline serological status and subsequent PCR test result, given a test was sought.

The choice of cut-off week used to define participants' baseline seroprevalence and the subsequent observation period for PCR testing has important implications in the estimation of the odds ratio

for reinfection. For instance, a cut-off week early in the study period will result in few seropositive individuals, while a cut-off week later in the study period leaves less time to observe subsequent PCR testing and detect possible reinfections, impacting the accuracy of estimates. To assess how the choice of cut-off week affected estimates of the odds ratio for reinfection, we repeated the multivariable logistic regression for every possible cut-off week. We assumed the most robust estimation of the relative risk of reinfection would occur in the between the two ‘waves’ of infection risk seen in the study cohort. We validated this assumption by conducting a simulation analysis using a known underlying probability distribution of infection and reinfection (see Supplementary Materials).

Potential confounding variables included; age, sex, race, ethnicity, BMI, state, work location, job category, household size, history of chronic disease, history of smoking and test frequency. We used a backwards selection procedure to select which variables to adjust for in our analyses, minimising root mean square error (RMSE) at each step [23]. Age and sex were considered ‘forced’ variables which we decided to control for a priori and were adjusted for in all analyses [24,25]. We conducted variable selection separately for each cut-off week and the variable sets adjusted for in each regression analysis are listed in Supplementary Materials S1. For most cut-off weeks (specifically those between 19th May 2020 and 22nd November 2020) all potential confounders were adjusted for, while early weeks (between 26th April 2020 and 3rd May 2020) and late weeks (between 20th December 2020 and 17th January 2021) adjusted for a subset of potential confounders.

As a sensitivity analysis, we performed the same analysis but excluding records where individuals had recorded a specific reason for test such as onset of symptoms or potential exposure to a COVID-19 case. As such, this sensitivity analysis included only individuals tested at random.

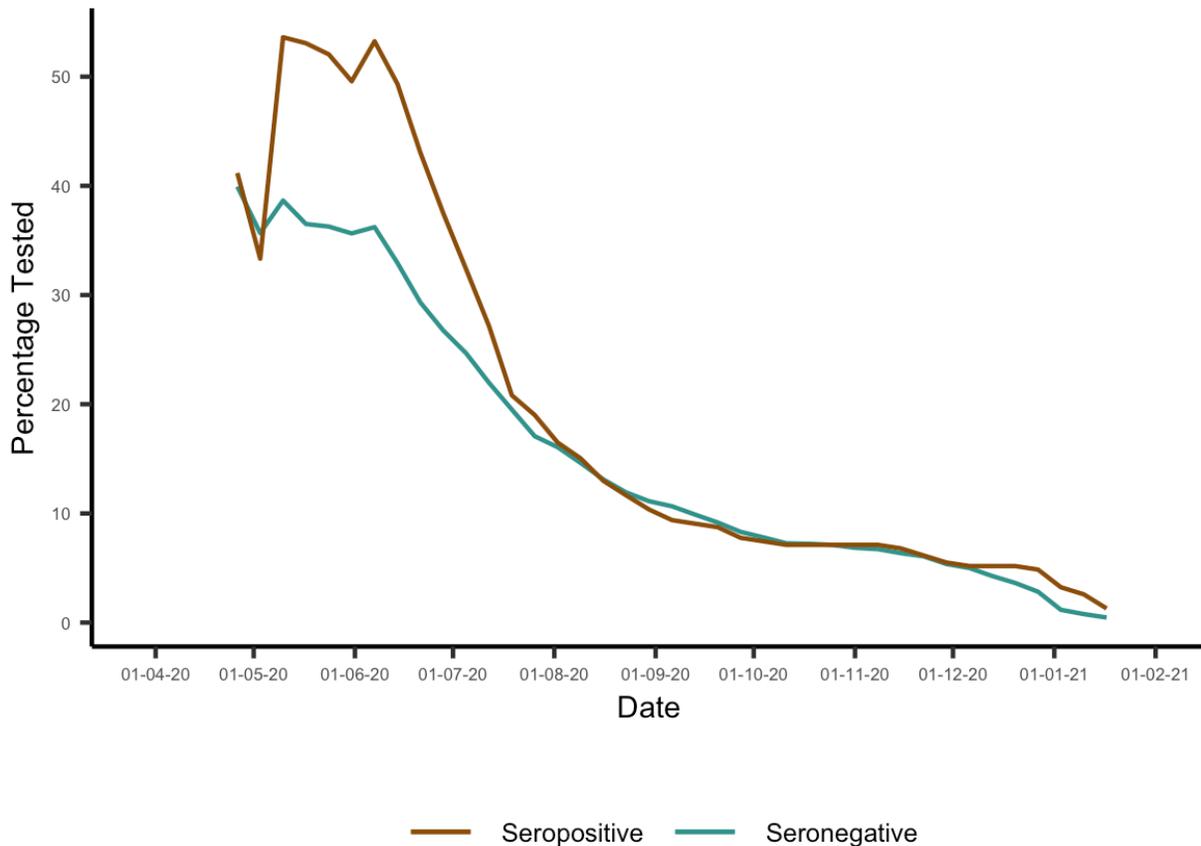


Figure 3.4 Propensity to seek a PCR test between the seronegative and seropositive groups

Propensity to seek a PCR test shown for each cut-off week considered in the main analysis. This was calculated as the percentage of those enrolled by the cut-off week shown on the x-axis who received at least one PCR test in the subsequent observation period.

We investigated the propensity to be tested among seronegative and seropositive individuals for each cut-off week by examining the percentage of those enrolled in the study by each cut-off week who had at least one test in the subsequent observation period and found that for cut-off weeks from mid-July onwards they were broadly similar between the two groups (Figure 5). However, the average distribution of test frequency differed between the seropositive and seronegative groups, with higher frequency of testing more common in the seronegative group. To account for this, we included PCR test frequency as a potential confounder in our analysis, defined as the number of PCR tests each individual took during the observation period (1-2, 3-5 or 6+). Protection

against infection with SARS-CoV-2 conferred by the presence of antibodies was estimated from the adjusted odds ratio (AOR) such that

$$Protection_{AB} = 1 - AOR_{Reinfection}.$$

Analysis was conducted in R version 4.0.3. Code to reproduce the figures and simulation analysis presented here can be found at <https://github.com/EmilieFinch/covid-reinfection>.

Acknowledgments

We thank SpaceX employees Lindsay Chapman, Jordan Steinhart, Suzanne Siebert, and Kyle Meade for their valuable support, in addition to the dedication and commitment of the many SpaceX employees that volunteered to participate in this study.

CMMID COVID-19 working group information

The following authors were part of the Centre for Mathematical Modelling of Infectious Disease COVID-19 Working Group. Each contributed in processing, cleaning and interpretation of data, interpreted findings, contributed to the manuscript, and approved the work for publication: Jiayao Lei, Sebastian Funk, Fiona Yueqian Sun, Amy Gimma, Emily S Nightingale, Graham Medley, Sam Abbott, Fabienne Krauer, Nicholas G. Davies, Mark Jit, Akira Endo, Oliver Brady, Anna M Foss, Yung-Wai Desmond Chan, Thibaut Jombart, Kevin van Zandvoort, Rosalind M Eggo, Yang Liu, Gwenan M Knight, Carl A B Pearson, Kaja Abbas, Katherine E. Atkins, Samuel Clifford, Mihaly Koltai, Yalda Jafari, Damien C Tully, Christopher I Jarvis, Kathleen O'Reilly, Nikos I Bosse, Kiesha Prem, Billy J Quilty, Simon R Procter, Rosanna C Barnard, William Waites, Ciara McCarthy, James D Munday, David Hodgson, W John Edmunds, Alicia Rosello, C Julian Villabona-Arenas, Hamish P Gibbs, Stefan Flasche, Timothy W Russell, Sophie R Meakin, Joel Hellewell, Naomi R Waterlow, Matthew Quaife, Frank G Sandmann.

Funding statements for the CMMID COVID-19 working group are as follows. KvZ: KvZ is supported by the UK Foreign, Commonwealth and Development Office (FCDO)/Wellcome Trust Epidemic Preparedness Coronavirus research programme (ref. 221303/Z/20/Z), and Elrha's Research for Health in Humanitarian Crises (R2HC) Programme, which aims to improve health outcomes by strengthening the evidence base for public health interventions in humanitarian crises. The R2HC programme is funded by the UK Government (FCDO), the Wellcome Trust, and the UK National Institute for Health Research

(NIHR). SC: Wellcome Trust (grant: 208812/Z/17/Z). FYS: NIHR EPIC grant (16/137/109). SFunk: Wellcome Trust (grant: 210758/Z/18/Z), NIHR (NIHR200908). GFM: NTD Modelling Consortium by the Bill and Melinda Gates Foundation (OPP1184344). YJ: LSHTM, DHSC/UKRI COVID-19 Rapid Response Initiative. SRM: Wellcome Trust (grant: 210758/Z/18/Z). WJE: European Commission (EpiPose 101003688), NIHR (NIHR200908). MQ: European Research Council Starting Grant (Action Number #757699); Bill and Melinda Gates Foundation (INV-001754). NRW: Medical Research Council (grant number MR/N013638/1). RME: HDR UK (grant: MR/S003975/1), MRC (grant: MC_PC 19065), NIHR (grant: NIHR200908). NGD: UKRI Research England; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK MRC (MC_PC_19065). JYL: Bill & Melinda Gates Foundation (INV-003174). MK: Foreign, Commonwealth and Development Office / Wellcome Trust. FK: Innovation Fund of the Joint Federal Committee (Grant number 01VSF18015), Wellcome Trust (UNS110424). DCT: No funding declared. JDM: Wellcome Trust (grant: 210758/Z/18/Z). AS: No funding declared. AMF: No funding declared. KP: Gates (INV-003174), European Commission (101003688). SFlasche: Wellcome Trust (grant: 208812/Z/17/Z). SA: Wellcome Trust (grant: 210758/Z/18/Z). BJQ: This research was partly funded by the National Institute for Health Research (NIHR) (16/137/109 & 16/136/46) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care. BJQ is supported in part by a grant from the Bill and Melinda Gates Foundation (OPP1139859). TJ: RCUK/ESRC (grant: ES/P010873/1); UK PH RST; NIHR HPRU Modelling & Health Economics (NIHR200908). AR: NIHR (grant: PR-OD- 1017-20002). GMK: UK Medical Research Council (grant: MR/P014658/1). MJ: Gates (INV- 003174, INV-016832), NIHR (16/137/109, NIHR200929, NIHR200908), European Commission (EpiPose 101003688). YL: Gates (INV-003174), NIHR (16/137/109), European Commission (101003688). JW: NIHR Health Protection Research Unit and NIHR HTA. JH: Wellcome Trust (grant: 210758/Z/18/Z). KO'R: Bill and Melinda Gates Foundation (OPP1191821). YWDC: No funding declared. TWR: Wellcome Trust (grant: 206250/Z/17/Z). CIJ: Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1). SRP: Bill and Melinda Gates Foundation (INV-016832). AE: The Nakajima Foundation. ESN: Gates (OPP1183986). NIB: Health Protection Research Unit (grant code NIHR200908). CJVA: European Research Council Starting Grant (Action number 757688). FGS: NIHR Health Protection Research Unit in Modelling & Health Economics, and in Immunisation. AG: European Commission (EpiPose 101003688). KA: Bill & Melinda Gates Foundation (OPP1157270, INV-016832). WW: MRC (grant MR/V027956/1). KEA: European Research Council Starting Grant (Action number 757688). RCB: European Commission (EpiPose 101003688). PK: This research was partly funded by the Royal Society under award RP\EA\180004, European Commission (101003688), Bill & Melinda Gates Foundation

(INV-003174). HPG: This research was produced by CSIGN which is part of the EDCTP2 programme supported by the European Union (grant number RIA2020EF-2983-CSIGN). The views and opinions of authors expressed herein do not necessarily state or reflect those of EDCTP. This research is funded by the Department of Health and Social Care using UK Aid funding and is managed by the NIHR. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health and Social Care (PR-OD-1017-20001). CABP: CABP is supported by the Bill & Melinda Gates Foundation (OPP1184344) and the UK Foreign, Commonwealth and Development Office (FCDO)/Wellcome Trust Epidemic Preparedness Coronavirus research programme (ref. 221303/Z/20/Z). OJB: Wellcome Trust (grant: 206471/Z/17/Z). WJE: European Commission (EpiPose 101003688), NIHR (NIHR200908, MRC: MC_PC_19065). CIJ: Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1). NGD: UKRI Research England; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK MRC (MC_PC_19065).

References

1. Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science*. 2020 Jul 17;369(6501):315–9.
2. Sabino EC, Buss LF, Carvalho MPS, Prete CA, Crispim MAE, Fraiji NA, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *The Lancet*. 2021 Feb 6;397(10273):452–5.
3. Seow J, Graham C, Merrick B, Acors S, Pickering S, Steel KJA, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nat Microbiol*. 2020 Dec;5(12):1598–607.
4. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med*. 2020 Jun;26(6):845–8.
5. Gudbjartsson DF, Norddahl GL, Melsted P, Gunnarsdottir K, Holm H, Eythorsson E, et al. Humoral Immune Response to SARS-CoV-2 in Iceland. *N Engl J Med*. 2020 Oct 29;383(18):1724–34.
6. Ibarondo FJ, Fulcher JA, Goodman-Meza D, Elliott J, Hofmann C, Hausner MA, et al. Rapid Decay of Anti-SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *N Engl J Med* [Internet]. 2020 Jul 21 [cited 2021 Apr 28]; Available from: <https://www.nejm.org/doi/10.1056/NEJMc2025179>
7. Chia WN, Zhu F, Ong SWX, Young BE, Fong SW, Bert NL, et al. Dynamics of SARS-CoV-2 neutralising antibody responses and duration of immunity: a longitudinal study. *Lancet Microbe* [Internet]. 2021 Mar 23 [cited 2021 Apr 28];0(0). Available from: [https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247\(21\)00025-2/abstract](https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(21)00025-2/abstract)
8. Duysburgh E, Mortgat L, Barbezange C, Dierick K, Fischer N, Heyndrickx L, et al. Persistence of IgG response to SARS-CoV-2. *Lancet Infect Dis*. 2021 Feb 1;21(2):163–4.
9. Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* [Internet]. 2021 Feb 5 [cited 2021 Apr 28];371(6529). Available from: <https://science.sciencemag.org/content/371/6529/eabf4063>
10. Ward H, Cooke G, Atchison C, Whitaker M, Elliott J, Moshe M, et al. Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. *medRxiv*. 2020

Oct 27;2020.10.26.20219725.

11. Long QX, Tang XJ, Shi QL, Li Q, Deng HJ, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med*. 2020 Aug;26(8):1200–4.
12. Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat Med*. 2021 Jul;27(7):1205–11.
13. To KKW, Hung IFN, Ip JD, Chu AWH, Chan WM, Tam AR, et al. Coronavirus Disease 2019 (COVID-19) Re-infection by a Phylogenetically Distinct Severe Acute Respiratory Syndrome Coronavirus 2 Strain Confirmed by Whole Genome Sequencing. *Clin Infect Dis* [Internet]. 2020 Aug 25 [cited 2021 Apr 28];(ciaa1275). Available from: <https://doi.org/10.1093/cid/ciaa1275>
14. Hall VJ, Foulkes S, Charlett A, Atti A, Monk EJ, Simmons R, et al. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). *The Lancet* [Internet]. 2021 Apr 9 [cited 2021 Apr 14];0(0). Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(21\)00675-9/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00675-9/abstract)
15. Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, et al. Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. *N Engl J Med*. 2021 Feb 11;384(6):533–40.
16. Letizia AG, Ge Y, Vangeti S, Goforth C, Weir DL, Kuzmina NA, et al. SARS-CoV-2 seropositivity and subsequent infection risk in healthy young adults: a prospective cohort study. *medRxiv*. 2021 Jan 29;2021.01.26.21250535.
17. Addetia A, Crawford KHD, Dingens A, Zhu H, Roychoudhury P, Huang ML, et al. Neutralizing Antibodies Correlate with Protection from SARS-CoV-2 in Humans during a Fishery Vessel Outbreak with a High Attack Rate. *J Clin Microbiol* [Internet]. 2020 Oct 21 [cited 2021 Mar 3];58(11). Available from: <https://jcm.asm.org/content/58/11/e02107-20>
18. Bartsch YC, Fischinger S, Siddiqui SM, Chen Z, Yu J, Gebre M, et al. Discrete SARS-CoV-2 antibody titers track with functional humoral stability. *Nat Commun*. 2021 Dec;12(1):1018.
19. Nilles EJ, Siddiqui SM, Fischinger S, Bartsch YC, de St. Aubin M, Zhou G, et al. Epidemiological and Immunological Features of Obesity and SARS-CoV-2. *Viruses*. 2021 Nov;13(11):2235.

20. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol.* 1978 Jan 1;107(1):71–6.
21. Hansen CH, Michlmayr D, Gubbels SM, Mølbak K, Ethelberg S. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study. *The Lancet* [Internet]. 2021 Mar 17 [cited 2021 Mar 24];0(0). Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(21\)00575-4/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00575-4/abstract)
22. Roy V, Fischinger S, Atyeo C, Slein M, Loos C, Balazs A, et al. SARS-CoV-2-specific ELISA development. *J Immunol Methods.* 2020 Sep 1;484–485:112832.
23. Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. *Int J Epidemiol.* 2016 Apr;45(2):565–75.
24. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med.* 2020 Aug;26(8):1205–11.
25. Pijls BG, Jolani S, Atherley A, Derckx RT, Dijkstra JIR, Franssen GHL, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open.* 2021 Jan 1;11(1):e044640.

Chapter 4

Effects of mobility, immunity and vaccination on SARS-CoV-2 transmission in the Dominican Republic: a modelling study

SARS-CoV-2 transmission is determined by a complex and non-linear interplay of post-infection immunity from prior transmission, changes in population behaviour (which can vary seasonally, as a result of changes in non-pharmaceutical interventions or in response to epidemic trends), the introduction of new variants and vaccination campaigns. Disentangling the relative role of these drivers on dynamics requires the integration of multiple data sources within modelling frameworks. Serological data, estimating past exposure in a population, is particularly valuable to distinguish between hypotheses behind observed epidemiological patterns. During the pandemic, mechanistic models were used routinely to provide decision-support to policy makers. However, there has been a lack of in-depth modelling for low- and middle-income countries. In particular, SARS-CoV-2 dynamics in countries that were unable to fully suppress transmission through non-pharmaceutical interventions prior to vaccine roll-out are not well understood. The Dominican Republic saw high incidence of SARS-CoV-2, with around three-quarters of the population previously infected by the autumn of 2021, and launched a vaccination campaign in February 2021, primarily using Sinovac-CoronaVac.

In this study, I jointly fit a multi-strain, age-stratified compartmental model to serological and surveillance data to investigate the transmission dynamics of SARS-CoV-2 in the Dominican Republic between 2020 and 2022. I also incorporated vaccination data and mobility data as model inputs. Once I had reconstructed the underlying epidemic dynamics, I was able to explore

counterfactual scenarios by changing model inputs, first assessing the impact of the vaccination campaign. I then explored the effect of using alternative vaccine products with higher efficacy vaccine, as well as the effect of delaying the campaign to wait for a higher efficacy product, retrospectively examining key policy decisions. Finally, I investigated the trade-off between levels of vaccination coverage and population mobility on COVID-19 burden, estimating the ‘return-to-normality’ afforded by the vaccination campaign.

This chapter was published in the *Lancet Regional Health – Americas* in September 2024. Supplementary Material for this paper is included as Appendix C.



London School of Hygiene & Tropical Medicine
 Keppel Street, London WC1E 7HT
 T: +44 (0)20 7299 4646
 F: +44 (0)20 7299 4656
 www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed **for each** research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1802674	Title	Miss
First Name(s)	Emilie		
Surname/Family Name	Finch		
Thesis Title	Modelling the role of immunity, climate and behaviour in viral outbreak dynamics and control		
Primary Supervisor	Adam Kucharski		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	The Lancet Regional Health - Americas
Please list the paper's authors in the intended authorship order:	Emilie Finch, Eric J Nilles, Cecilia Then Paulino, Ronald Skewes-Ramm, Colleen Lau, Rachel Lowe1, Adam J Kucharski
Stage of publication	In press

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I worked on: conceptualisation, methodology, project administration, investigation, data analysis, figures, writing original draft and review and editing of submitted manuscript.
--	--

SECTION E

Student Signature	
Date	08/07/2024

Supervisor Signature	
Date	12/07/2024

4.1 Abstract

Background

COVID-19 dynamics are driven by a complex interplay of factors including population behaviour, new variants, vaccination and immunity from prior infections. We quantify drivers of SARS-CoV-2 transmission in the Dominican Republic, an upper-middle income country of 10.8 million people, and assess the impact of the vaccination campaign implemented in February 2021 in saving lives and averting hospitalisations.

Methods

We fit an age-structured, multi-variant transmission dynamic model to reported deaths, hospital bed occupancy, and seroprevalence data until December 2021, and simulate epidemic trajectories under different counterfactual scenarios.

Findings

We estimate that vaccination averted 7210 hospital admissions (95% CrI: 6830 –7600), 2180 ICU admissions (95% CrI: 2080 –2280) and 766 deaths (95% CrI: 694 – 859) in the first 6 months of the campaign. If no vaccination had occurred, we estimate that an additional decrease of 10-20% in population mobility would have been required to maintain equivalent death and hospitalisation outcomes. We also found that early vaccination with Sinovac-CoronaVac was preferable to delayed vaccination using a product with higher efficacy.

Interpretation

SARS-CoV-2 transmission dynamics in the Dominican Republic were driven by a substantial accumulation of immunity during the first two years of the pandemic but, despite this, vaccination was essential in enabling a return to pre-pandemic mobility levels without considerable additional morbidity and mortality.

4.2 Introduction

During 2020-22, many countries experienced a significant burden of COVID-19 and imposed non-pharmaceutical interventions (NPIs) aiming to control SARS-CoV-2 transmission. Despite this, countries experienced markedly different epidemic dynamics, mediated by a complex interplay of factors including population behaviour, government interventions, the introduction of new variants, the roll-out of vaccination campaigns and prior levels of transmission. Serological surveys have proven crucial to understand global and national landscapes of population immunity, as well as indicating the extent of prior exposure to SARS-CoV-2 [1]. Some countries pursued an elimination strategy throughout the pre-Omicron era, with stringent public health interventions resulting in low levels of seroprevalence towards the end of 2021. For example, in Hong Kong, serosurveillance studies found <1% of sera tested was positive for anti-N SARS-CoV-2 IgG prior to March 2021 [2]. Other countries, such as the UK and many EU countries, saw epidemic waves linked to the strengthening and relaxing of public health interventions, alongside the emergence of more transmissible variants. In England, approximately 20% of the population were estimated to have been infected with SARS-CoV-2 between April 2020 and July 2021 [3]. In contrast, SARS-CoV-2 transmission was largely unmitigated in some settings such as in Manaus, Brazil, where 76% of the population are thought to have been infected by October 2020 [4].

Latin America and the Caribbean was a global hotspot for SARS-CoV-2 transmission during 2020-21, prior to the emergence of the Omicron variant, which caused large epidemics globally [5,6]. The Dominican Republic is an upper-middle income country with a population of 10.8 million, which shares the island of Hispaniola in the Caribbean with Haiti. They reported their first case of COVID-19 on the 1st of March 2020, which was followed by the imposition of strict public health measures including the closure of schools and workplaces, the cancellation of public events and the imposition of curfews. These began to be relaxed in July 2021 and were mostly lifted with the reopening of schools and relaxation of curfew measures in October 2021. In the first two years of the pandemic, the Dominican Republic experienced four waves of transmission: the first peaked in August 2020, with cases increasing again from November 2020 before a second peak in January 2021. A third wave of transmission took place over the summer of 2021, following the introduction of more transmissible variants, including Mu, with cases rising sharply to peak in July 2021.

Following the introduction of the Delta variant, a fourth wave took place in October and November 2021 [7]. A nationally representative serological survey involving 6683 individuals from 3832 households took place between June and October 2021 [8]. Results from the serological survey estimated that 76.6% (95% CI 70.1 – 82.5) of the population had been previously infected by the study midpoint. This vastly exceeded earlier estimates constructed from coarse reported data for the Dominican Republic, and the wider region of Latin America and the Caribbean [9].

The government of the Dominican Republic launched a national COVID-19 vaccination programme on 16th February 2021 initially focusing on health-care professionals and then following a three-phase age-based approach [10]. The campaign first targeted individuals over 60 years of age, then expanding to individuals over 50 on 3rd May 2021 and all adults over 18 on 10th May 2021 (Supplementary Table 1). Booster vaccination for highly vulnerable individuals began in July 2021, with the Dominican Republic being the first country in the Americas to approve vaccination with a third dose [11]. During the study period, approximately 90% of vaccine doses administered were Sinovac-CoronaVac (an inactivated viral vaccine), with Oxford/AstraZeneca vaccine (ChAdOx1-S, an adenovirus vector vaccine) and Pfizer/BioNTech (BNT162b2, mRNA vaccine) also administered [8]. By the current study's endpoint (15th December 2021), 62% of the population had received at least one dose of a COVID-19 vaccine and 50% had received two doses.

Mathematical models have been used throughout the pandemic to provide decision-support to policy makers through estimation of key epidemiological parameters, forecasts of future incidence, projections of epidemic trajectories under different scenarios, and quantification of the impact of non-pharmaceutical interventions. However, despite regular and in-depth modelling decision-support for high-income countries, there has been a lack of equivalent modelling analysis to understand transmission and control in low- and middle-income countries [12–18]. To address this gap, we used an age-structured transmission dynamic model to quantify the drivers of epidemic dynamics in the Dominican Republic during the first two years of the pandemic, and to assess the impact of the vaccination campaign on COVID-19 hospitalisations and deaths.

4.3 Methods

4.3.1 Data

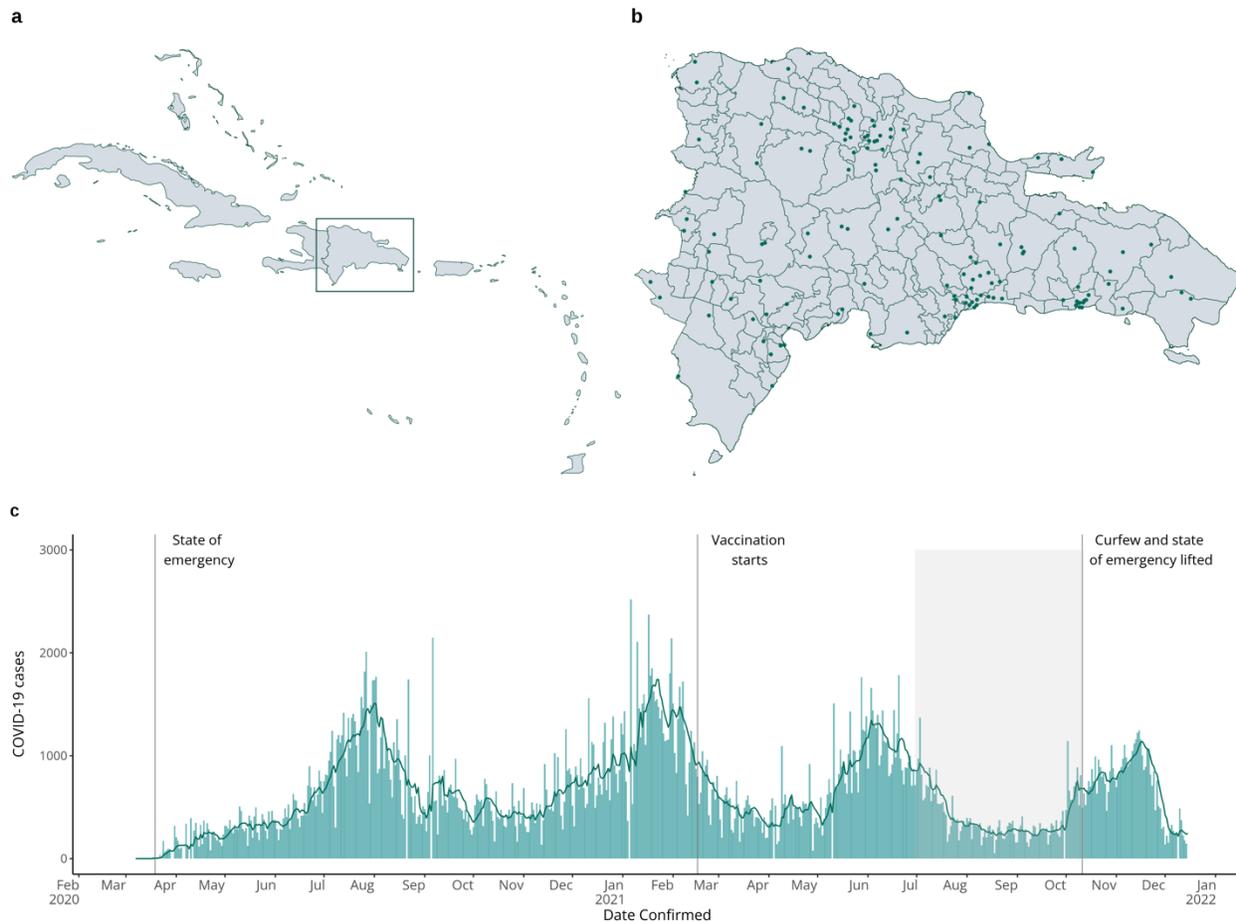


Figure 4.1 Map of the study setting and time-series of COVID-19 cases.

Figure showing a map of the Caribbean with the location of the Dominican Republic shown in a box (a), a map of the Dominican Republic showing clusters sampled in the 2021 serosurvey (b), and daily COVID-19 cases in the Dominican Republic (bars) and the 7-day moving average (line) from March 2020 – January 2022 (c). The shaded grey area indicates the timing of the serological survey.

For this analysis we incorporated multiple data streams. Aggregated daily reported deaths were collected from the Dominican Republic’s COVID-19 Dashboard. Aggregated daily hospital and ICU bed occupancy were scraped from daily COVID-19 bulletins published by the Ministerio de Salud Pública y Asistencia Social, available from 19th September 2020 onward [19]. We also used

serological data from a nationally representative multistage SARS-CoV-2 seroprevalence survey undertaken between June - October 2021. This survey employed a multistage sampling method which assigned clusters to provinces, taking into account population, urban-rural divide, and the spatial dispersal of clusters. Individuals aged ≥ 5 years old were eligible to enroll in the study and 6683 individuals were surveyed in total. SARS-CoV-2 antibodies were measured using the Roche Elecsys SARS-CoV-2 electrochemiluminescence immunoassays. Large non-manufacture-sponsored studies demonstrated specificities and sensitivities of 99.8% (CI 99.3–100) and 98.2% (CI 96.5–99.2) for the Elecsys anti-S assay and 99.6% (CI 98.9–100) and 90.8% (CI 81.3–95.7) for the anti-NC assay [20,21]. Seroprevalence estimates were adjusted for study design, national demographics and assay characteristics. Further details of survey methodology and findings are available elsewhere [8].

To obtain estimates of age stratified vaccination rates, we used data on the daily number of second vaccine doses distributed in the population and assumed that daily doses were evenly distributed between eligible age groups as per the government’s vaccination program [22], Supplementary Table 1). If an eligible age group became fully vaccinated during the vaccination allocation, remaining vaccine doses were distributed between the remaining eligible age groups or, if all were fully vaccinated, between adults > 20 , mimicking the vaccination of younger health care workers or those with chronic health conditions. Estimated vaccination coverage by age group over time is shown in Figure 2.

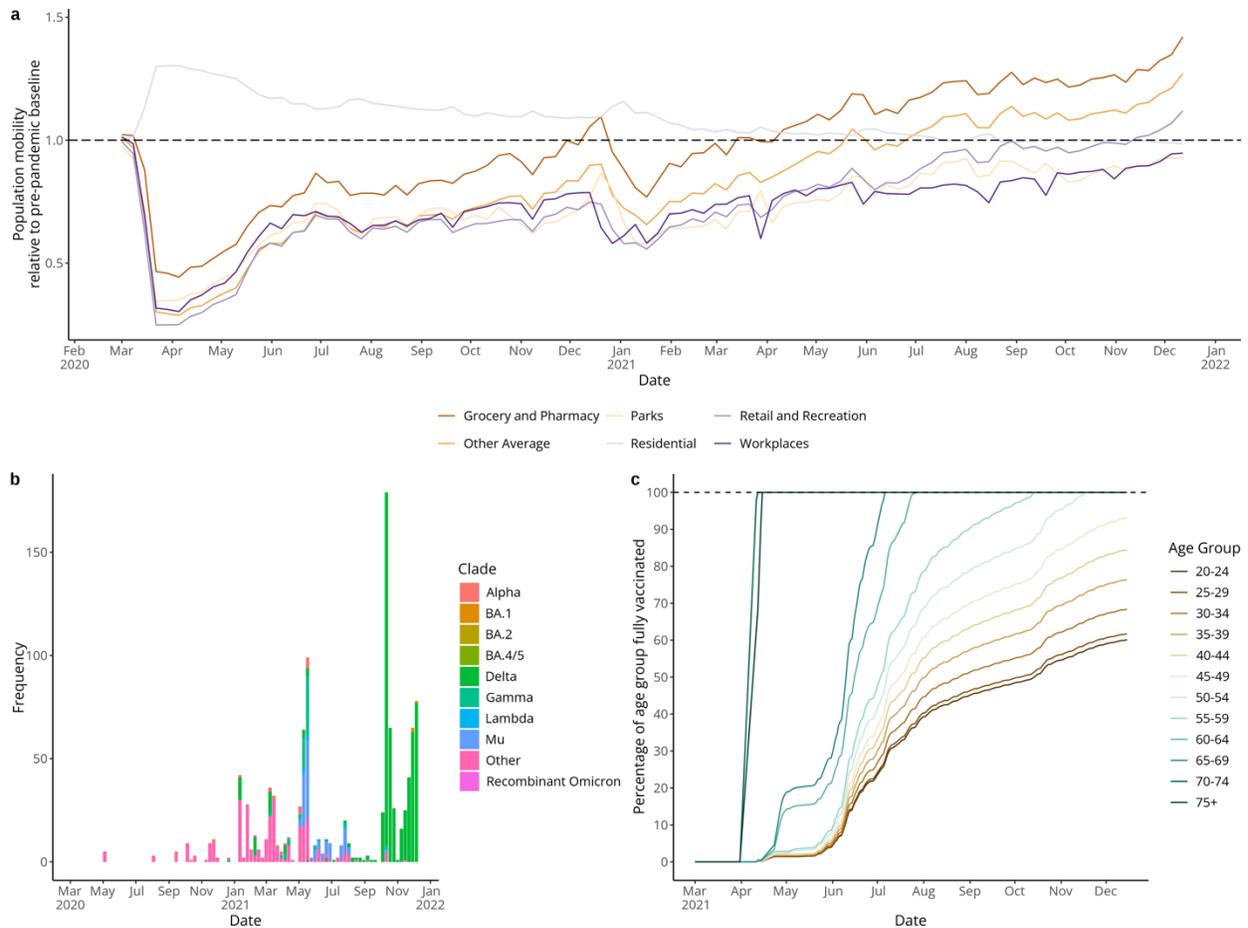


Figure 4.2 Google mobility data, SARS-CoV-2 sequence data, and estimated vaccination coverage by age. Panel a shows Google mobility data showing the proportional change in population mobility in different locations relative to a pre-pandemic baseline, panel b shows the frequency of SARS-CoV-2 sequences from the Dominican Republic on GISAID by variant and panel c shows estimated vaccination coverage by age assuming vaccine doses were evenly distributed between eligible age groups.

4.3.2 Transmission dynamic model

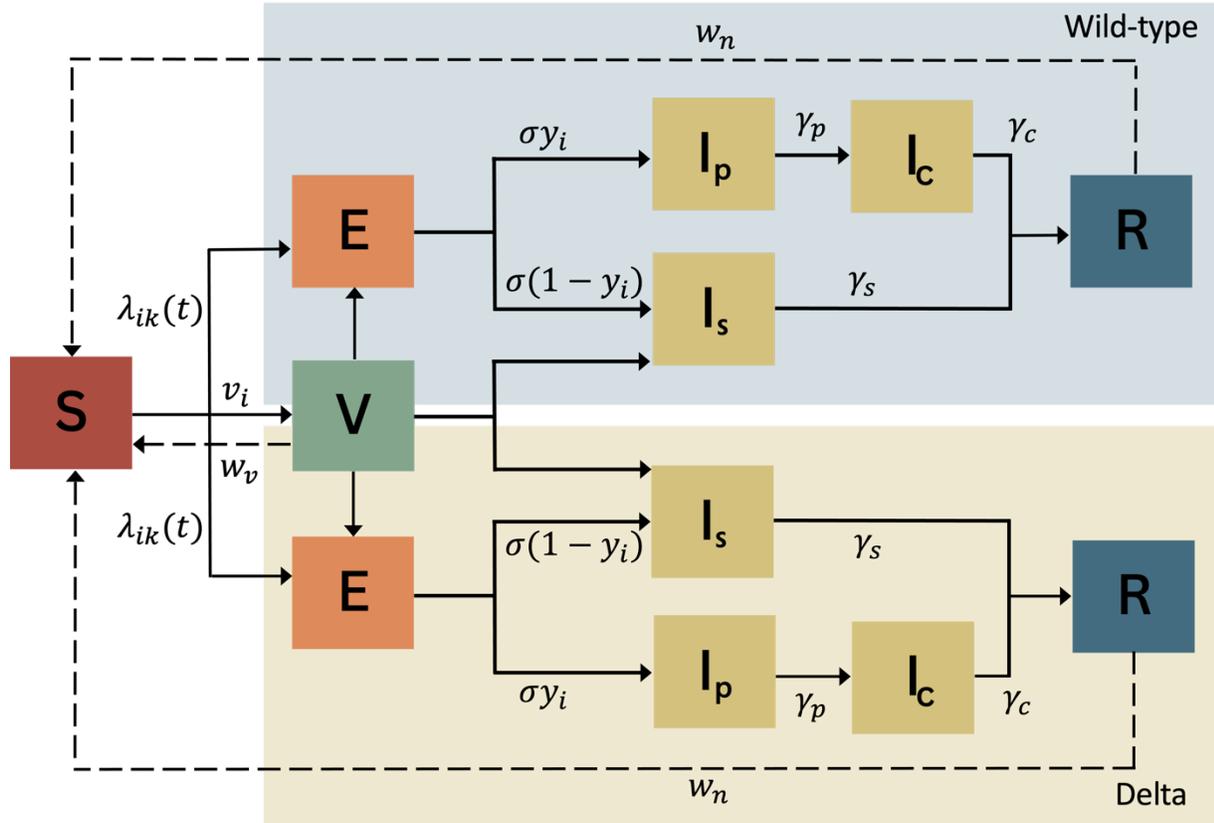


Figure 4.3 Model schematic showing a two-variant SEIR model structure with a vaccination compartment (V)

Model compartments $S_i, E_i, I_{pi}, I_{c,i}, I_{si}$ represent the number of individuals who are susceptible, exposed, infected (pre-clinical), infected (clinical) and infected (sub-clinical or asymptomatic) in age group i and either the wild-type or Delta variant respectively. Parameters are defined as follows: λ_{ik} is the force of infection for age group i and variant k , v_i is the number of daily vaccinations for age group i ; σ is $1/dE$ (where dE is the duration in the exposed compartment or the latent period); γ_p is $1/dP$ (where dP is the duration of preclinical infectiousness); γ_s is $1/dS$ (where dS is the duration of clinical infectiousness); γ_c is $1/dC$ (where dC is the duration of sub-clinical infectiousness). y_i is the probability of clinical symptoms given infection for age group i . w_n is the rate of waning of post-infection immunity and w_v is the rate of waning for post-vaccination immunity. A latent compartment L between compartments V and I_s for vaccinated individuals who have been exposed to infection but protected from disease has been omitted for simplicity but is included in Supplementary Materials Table 2. Further description of model equations and parameter values is given in the Supplementary Materials Tables 2-6.

We used covidM, an age-stratified, deterministic, compartmental model originally developed to model the effects of NPIs on SARS-CoV-2 transmission in the UK, and described fully elsewhere [23–25]. In brief, covidM is structured into 5-year age groups, with individuals moving from a susceptible state (S) to an exposed state (E) and then either to a pre-clinical and clinical infected state (I_p followed by I_c) or a sub-clinical infected state (I_s) and finally to a recovered state (R). The model explicitly considers two variants of SARS-CoV-2; wild-type and B.1.617.2 (Delta), while the introduction of other variants in the Dominican Republic in the intervening period (which include Mu, Gamma and Alpha) is captured by fitting a gradual increase in transmissibility in the first half of 2021, following a logistic function (Supplementary Materials, Table 2). We fit the model to daily reported deaths, daily hospital and ICU bed occupancy and a cross-sectional seroprevalence estimate from June - October 2021. We fit the model using data until 15th December 2021, when cases began to increase due to the Omicron variant, and only simulate epidemic trajectories until this point.

Hospitalisation, ICU admission and death are modelled as observation processes according to age-specific infection-severe ratios, infection-critical ratios, infection-fatality ratios and length of stay based on estimates from the literature. These are adjusted on the log odds scale by several fitted parameters and delays from infection to hospitalisation, ICU admission and death are estimated during the model fitting process. We assume that the observed number of deaths, hospital bed occupancy and ICU bed occupancy are distributed according to a negative binomial distribution, with the overdispersion parameter estimated during the model fitting process. We used a skew-normal likelihood for seroprevalence with the same mean and 95% confidence interval as reported for the data evaluated for the period of the serosurvey.

We consider a central waning assumption corresponding to 15% loss of post-infection protection after 1 year. Full details on model equations, fixed and fitted model parameters can be found in Supplementary Tables 2-6.

4.3.3 Vaccination parameters

We also incorporated information on SARS-CoV-2 vaccination in the Dominican Republic, using data collated by Our World in Data [22]. Fully vaccinated individuals moved to a vaccinated model compartment (V) from which, subject to vaccine waning parameters, they can move to the exposed state (E) or through a latent compartment (L) directly to a sub-clinical infection (I_s). We assume vaccinated individuals have a lower probability of clinical or sub-clinical infection but that, once infected, they have the same infectiousness as non-vaccinated individuals. We model vaccine efficacy against infection (VE_{inf}) and vaccine efficacy against severe disease given infection ($VE_{sd|inf}$) as described in Figure 3 and Supplementary Table 5. We used vaccine efficacy against severe disease rather than symptomatic disease as we fit directly to hospitalisations, ICU admissions and deaths. However, we conducted sensitivity analysis around the vaccination parameters used, as well as waning assumptions (Supplementary Tables 7-9).

As 90% of primary course vaccinations given in the Dominican Republic were CoronaVac (Sinovac COVID-19 vaccine) we used vaccination efficacy parameters based on the literature available for this vaccine product. We model differing vaccine efficacy by strain but assume the same vaccine efficacy across age groups [26,27]. Note that we do not directly model vaccine efficacy against hospitalisation or death.

4.3.4 Mobility

In this analysis, we used separate age-stratified social contact matrices for contacts in the home, at work, in school, or in other locations (for instance leisure or transport), simulated from contact surveys and demographic data [28]. To estimate changes in contact rates during the pandemic, we used population mobility data captured by Google's COVID-19 Community Mobility Reports as a proxy for changes in population behaviour. We mapped changes in mobility to changes in contact rates using the relationship between mobility and contact survey data found in the UK [23,29]. School contact rates were set to zero during school closures and school holiday periods. The representativeness of Google mobility data is dependent on the proportion of the population with smartphones using Google products. As this differs between the UK and the Dominican Republic, we infer a weighting between UK-adjusted contact rates and pre-pandemic baseline contact rates

in the Dominican Republic, fitting a separate weighting parameter for each year of the simulation period [28]. We found the UK-adjusted value was given more weight in the first year of the pandemic than the second, suggesting that the relationship between measured population mobility and contact rates changed during the pandemic.

4.3.5 Model fitting

We performed Bayesian inference using Markov chain Monte Carlo to estimate model parameters. We used the Differential Evolution Markov Chain Monte Carlo (DE-MCMC) algorithm which combines a genetic algorithm (Differential Evolution) with MCMC [30]. Here, multiple Markov chains are run in parallel and learn from one another to determine the scale and orientation of the proposal distribution, which allows for more efficient exploration of a complex parameter space than traditional Metropolis-Hastings algorithms, particularly when considering correlated parameters. Model convergence was assessed using trace plots of MCMC chains and the Rhat statistic [31].

4.3.6 Counterfactual analysis

We conducted counterfactual scenario analysis using the fitted model to simulate epidemic trajectories under different scenarios. We considered five key scenarios with changes applied from the beginning of the vaccination campaign (15th February 2021) until the end of the analysis period:

1. No vaccination
2. Vaccination using a vaccine with a Pfizer/BioNTech efficacy profile
3. Delayed vaccination using a vaccine with a Pfizer/BioNTech efficacy profile
4. Vaccination using a vaccine with an Oxford/AstraZeneca efficacy profile
5. Delayed vaccination using a vaccine with an Oxford/AstraZeneca efficacy profile

In Scenario 1 we assume no vaccine doses are distributed. In Scenarios 2 and 3 we assume the same timing of vaccine distribution, but with a Pfizer/BioNTech or Oxford/AstraZeneca efficacy

profile. For Scenarios 3 and 5 we consider a vaccination programme using a Pfizer/BioNTech or Oxford/AstraZeneca efficacy profile with a delay of two months. Here vaccination would begin 15th April 2021, aligning with initial deliveries of Oxford/AstraZeneca vaccine doses through COVAX [32].

For each scenario we ran 500 simulations, drawing parameters from the posterior distribution of each fitted parameter, and calculated the difference in hospital admissions, ICU admissions and deaths from the original model fit to estimate the impact of the scenario considered.

We also conducted a counterfactual analysis to examine the trade-off between differing levels of vaccination coverage and changes in population mobility on deaths, hospital admissions and ICU admissions during the time period between 16th February - 16th August 2021. We considered 11 vaccination coverage scenarios with coverage by 16th August 2021 ranging from 0 to 100% in 10% increments, and 9 population mobility scenarios changing ‘Work’ and ‘Other’ mobility by an extra -40% to +40% compared with the Google mobility data during the simulation period. Vaccination allocation for each scenario was performed in a similar way as described above, where daily vaccinations are multiplied by a factor equal to $\frac{Coverage_{simulation}}{Coverage_{actual}}$. Vaccination doses are then distributed between eligible age groups and, once these are fully vaccinated, between age groups > 20 and then finally between age groups < 20 . We then used the fitted model to simulate epidemic trajectories for 99 scenarios, considering all combinations of vaccination coverage and population mobility change, and estimated the impact of the scenario considered as above.

4.4 Results

4.4.1 COVID-19 transmission dynamics between 2020 – 2022

Our modelling analysis suggests that, after an initial decline in transmission following a sharp reduction in social interactions, COVID-19 dynamics were driven by substantial accumulation of immunity throughout 2020-2022 (Figure 4), as well as the spread of novel variants such as Mu in mid-2021 and Delta in late 2021. By jointly fitting to reported deaths, hospital bed occupancy,

ICU bed occupancy and seroprevalence data, the model reproduced the overall observed epidemic dynamics. Estimated deaths did not track closely with observed deaths during the first wave from March - September 2020, and deaths and hospitalisations were slightly underestimated during the second wave in January 2021.

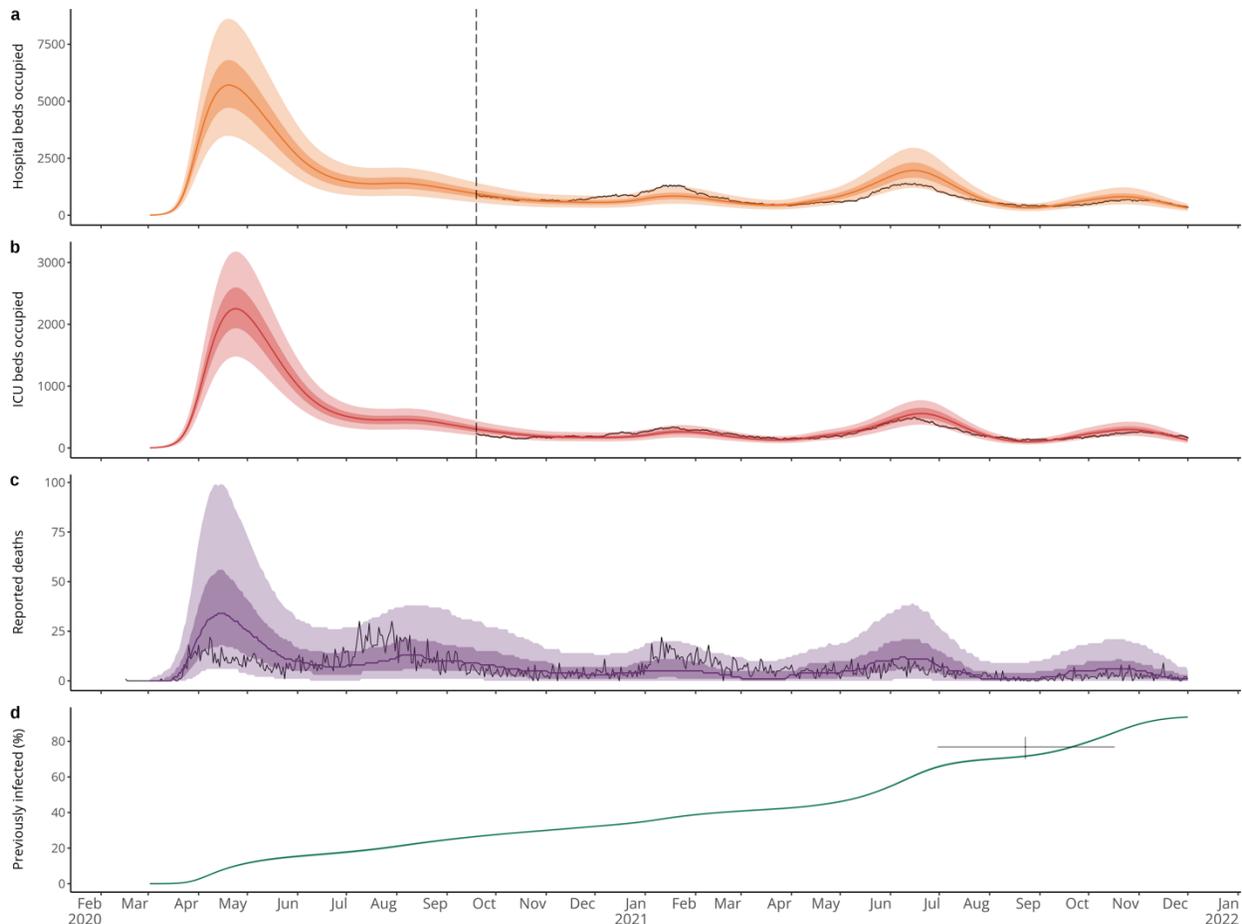


Figure 4.4 Comparison of model fit to observed data in the Dominican Republic from February 2020 to December 2021

Panels show comparison between modelled and observed hospital bed occupancy (a), ICU bed occupancy (b), reported deaths (c) and proportion of the population previously infected (d). Black lines show observed data, with vertical dashed lines in panels (a) and (b) indicating the point at which hospitalisation data became available. For panel (d) the black cross shows the duration of the serosurvey (horizontal line) and the 95% confidence interval around the central estimate (vertical line). Modelled hospital bed occupancy, ICU bed occupancy, deaths, and proportion previously infected, are shown in orange, red, purple and green

respectively, with associated 50% and 95% credible intervals in surrounding ribbons. Note that uncertainty in the observation process is included in modelled outputs for surveillance data streams (a, b and c) but not for the proportion previously infected (d).

Reconstructing the underlying epidemic dynamics, we found that changes in the effective reproduction number, R_t , reflected changes in contact rates derived from Google mobility data during 2020, but were less strongly associated with contact rates during 2021 (Figure 5). We estimated that 33.4% (95% CrI: 33.3-33.5) of the population had been infected by the end of 2020 (Figure 5), ranging from around 45% in those aged 20-39 to around 20% in those aged under 19 and above 70 (Figure S.3). This accumulation of population immunity contributed to a decline in transmission, with R_t remaining around 1 despite gradual increases in contact rates from May 2020. By the end of 2021, we estimated that 82.1% (95% CrI: 81.9 - 82.4%) of the population had been infected, ranging from above 90% in those aged 20-39 to around 55% in those over 70 (Figure S.3). Again, high levels of post-infection immunity resulted in reduced levels of transmission despite contact rates approaching pre-pandemic baseline levels, except during the emergence of more transmissible variants in May and September 2021.

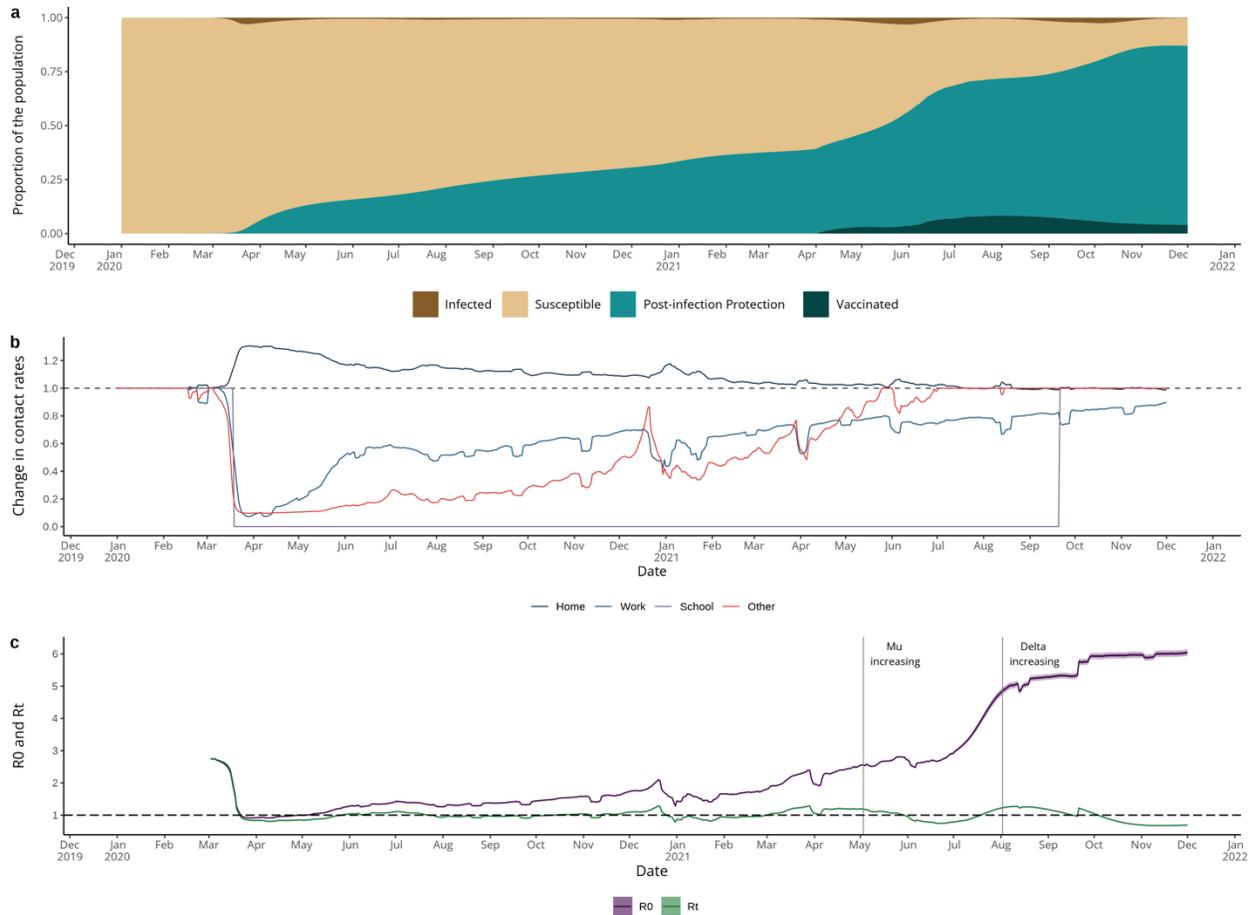


Figure 4.5 Immune status, contact rates and reproduction number estimates from January 2020 to December 2021

Panel a shows the modelled distribution of immune states in the Dominican Republic over time showing the proportion of the population that are: currently infected (brown), susceptible (beige), protected post-infection (blue) and protected post-vaccination (dark blue). Note that the vaccinated area (dark blue) does not include individuals that were vaccinated post-infection and so does not correspond with observed vaccination coverage. Panel b shows inferred contact rates in home, work, school and other settings relative to a pre-pandemic baseline. Panel c shows estimated R_0 and R_t . R_0 , the basic reproduction number, is defined as the average number of secondary cases generated by a primary cases in a susceptible population, while R_t , is its time-varying equivalent. Here, changes in R_0 reflect changing contact rates and transmissibility while changes in R_t incorporate both changing contacts and transmissibility as well as the build-up of immunity over time. The vertical lines on Panel c show the time at which Mu sequences start to increase, according to GISAID, the global data science initiative. Note that within the model, increases in transmissibility due to Mu and Delta are allowed to begin earlier in time (from February and May 2021 respectively) to reflect likely delays in their detection due to low levels of sequencing [33].

4.4.2 Impact of vaccination campaign

To estimate the impact of vaccination, we used our calibrated model to simulate counterfactual epidemic trajectories under different scenarios. First, we considered a ‘no vaccination scenario’, estimating deaths, hospital and ICU admissions in the absence of any vaccination during 2021 (Figure 6). By comparing these counterfactual outcomes to the original model estimates we were able to estimate the burden averted by the vaccination campaign in the 6 and 10 months following its launch in February 2021 (Table 2). It should be noted that this analysis assumes all other factors remain constant and, in particular, we assume the same changes in contact rates and variant transmissibility as in the original model fitting displayed in Figure 4. This is a simplifying assumption, as in reality it is possible countries would respond to rising COVID-19 cases with impositions of further restrictions or would see accompanying changes in population behaviour.

We estimated that the vaccination campaign averted 7210 hospital admissions (95% CrI: 6830 – 7600) hospital admissions, 2180 ICU admissions (95% CrI: 2080 – 2280) and 766 deaths (95% CrI: (694—859) in the 6 months following its launch. This is equivalent to averting 27.0% (95% CrI: 25.6 - 28.5) of hospital admissions, 33.2% (95% CrI: 31.8 - 34.8) of ICU admissions and 36.2% (95% CrI: 32.8 - 40.6) of reported deaths considering the median values expected under a ‘no vaccination’ scenario. Notably, we estimated hospital bed and ICU bed capacity would have been exceeded under a ‘no vaccination’ scenario, given population behaviour and variant introductions observed in 2021 [19].

Given the challenges and inequities of vaccine availability in real-time, with some products available at scale before others, we evaluated the impact of using alternative vaccine products with or without a delay in the vaccination programme on deaths, hospital and ICU admissions.

We estimated that while vaccination with a more efficacious product would have reduced hospitalisations, ICU admissions and deaths, delaying the vaccination campaign to vaccinate with a more efficacious product would have resulted in a higher overall burden in subsequent waves in 2021 (Figure 7). This is due both to the speed of vaccination rollout, with 50% of the population

receiving a two-dose primary series by the end of 2021, as well as the introduction of variants of concern or interest (particularly the Mu variant) in the summer of 2021 [22].

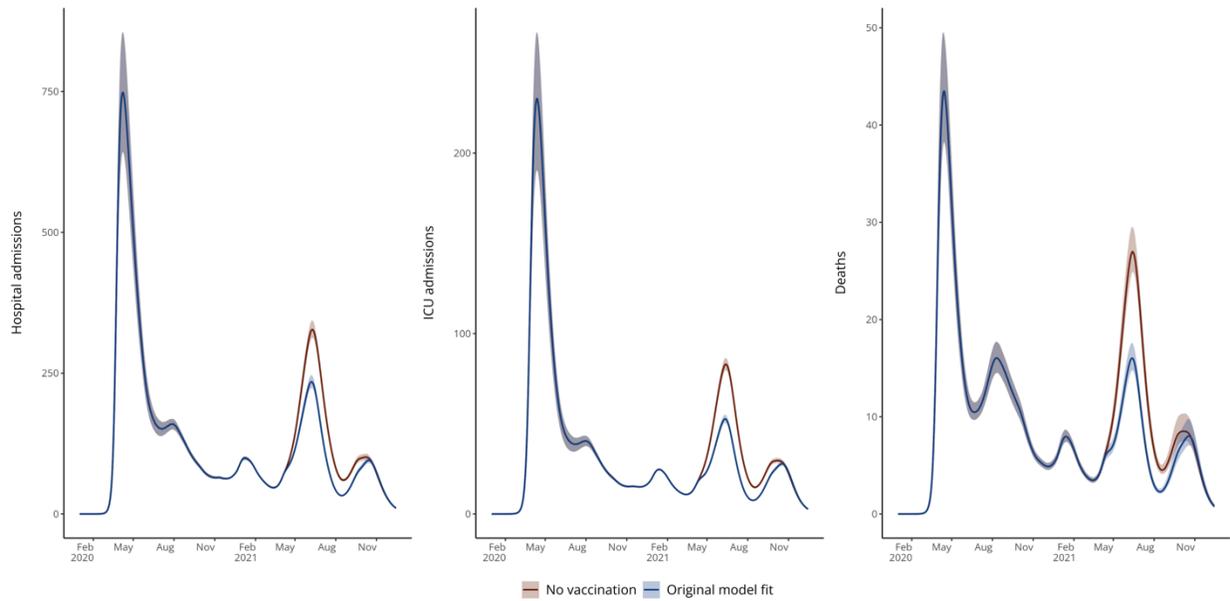


Figure 4.6 Impact of vaccination campaign

Figure showing modelled deaths (a), hospital admissions (b) and ICU admissions (c) from the original model fit (blue) and from a ‘no vaccination’ counterfactual (red). Lines show the median value from 500 simulations. To facilitate comparison between scenarios, modelled deaths do not include uncertainty generated through the observation process.

Scenario	Additional hospital admissions in next 6 months	Additional hospital admissions in next 10 months	Additional ICU admissions in next 6 months	Additional ICU admissions in next 10 months	Additional deaths in next 6 months	Additional deaths in next 10 months
No vaccination	7210 (6830 – 7600)	8330 (7850 – 8830)	2180 (2080 – 2280)	2500 (2380 – 2640)	766 (694 – 859)	863 (773 – 978)
Pfizer efficacy or equivalent	-565 (-597 – -535)	-2070 (-2320 – -1840)	-149 (-157 – -143)	-599 (-660 – -541)	-48 (-54.2 – -43.5)	-179 (-239 – -129)
Pfizer efficacy or equivalent and delay	4840 (4540 – 5140)	2070 (1550 – 2580)	1380 (1310 – 1450)	533 (415 – 658)	475 (412 – 548)	228 (101 – 344)
AZ efficacy or equivalent	-88.8 (-107 – -71.5)	-1410 (-1600 – -1240)	-11.9 (-15.8 – -7.99)	-408 (-452 – -365)	-0.447 (-4.21 – 2.53)	-115 (-161 – -78.7)
AZ efficacy or equivalent and delay	4970 (4670 – 5280)	2470 (1970 – 2960)	1420 (1360 – 1500)	654 (540 – 776)	490 (429 – 564)	267 (146 – 379)

Table 3.1: Estimated total hospital admissions, ICU admissions and deaths under different counterfactual vaccination scenarios.

Median values and 95% credible intervals are shown from 500 simulations. Estimates are split into the cumulative burden estimated during the 6 months following the beginning of the vaccination campaign (until 16th August 2021, aligning with the Mu wave) and the 10 months following the beginning of the vaccination campaign (until 15th December 2021, aligning with the Mu and Delta waves).

4.4.3 Trade-off between vaccination and population mobility

Finally, we investigated the trade-off between levels of vaccination coverage and population mobility on hospitalisations, ICU admissions and deaths. Here, we explore different counterfactual combinations of vaccination coverage and population mobility change to understand how much relaxation of social distancing measures vaccination could ‘buy’ in the later stages of a pandemic.

We found that, overall, changes in population mobility resulted in greater variation in disease burden than vaccination, where an increase in population mobility resulted in more additional deaths, hospitalisation and ICU admissions than a corresponding decrease in vaccination (Figure 8). For instance, our analysis suggests an increase of 20% in population mobility would result in around 7,200 extra hospital admissions, 1,700 extra ICU admissions and 570 extra deaths, while a reduction of 20% in vaccination coverage would result in around 2,400 extra hospital admissions, 690 extra ICU admissions and 230 extra deaths (Figure 8).

We also compared combinations of population mobility reduction and vaccination coverage that resulted in the same outcomes (looking at lines of equivalence in Figure 8). We found that in the absence of vaccination, an additional 10-20% reduction in population mobility would have been required to obtain the same hospitalisation and death outcomes seen in this period, quantifying the ‘return-to-normality’ associated with the first 6 months of vaccination in this setting.

Additionally, if population mobility had remained as measured in this period but perfect vaccination coverage had been achieved, an estimated an additional 4530 (95% CrI: 4250 – 4840), 976 ICU admissions (95% CrI: 923 – 1035) and 288 deaths (95% CrI: 252 – 331) would have been averted. This additional burden averted (going from the observed coverage of 43% to 100%) is lower than the burden averted that we estimated in the earlier ‘no vaccination’ scenario (going from 0% coverage to the observed 43% coverage, Table 1). This illustrates the importance of an age-targeted approach in reducing morbidity and mortality. Finally, a simulated change in mobility of +30% would have returned contacts in February 2021 back to (or slightly above) pre-pandemic baseline levels. This suggests that a return to baseline mobility at the beginning of the vaccination campaign would have resulted in an extra 8,210 hospital admissions (95% CrI: 6530 – 9840), 2070 ICU admissions (95% CrI: 1760 – 2370) and 678 deaths (95% CrI: 380 – 976).

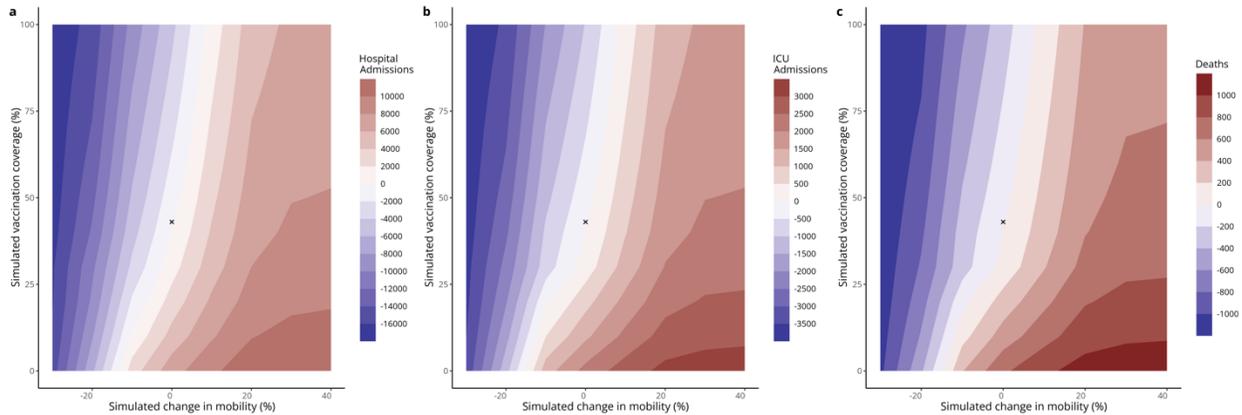


Figure 4.7 Modelled total additional hospital admissions, ICU admissions and deaths in the 6 months following the vaccination campaign launch (16th February 2021 – 16th August 2021)

These contour plots show the number of additional deaths (a), hospital admissions (b) and ICU admissions (c) compared to the original model fit under different levels of simulated vaccination coverage reached by 16th August 2021 (y-axis) and changes in population mobility during the 6 month period (x-axis). The actual vaccination coverage observed on 16th August 2021 is shown by a cross (vaccination coverage = 43% and mobility change = 0).

4.5 Discussion

We used an age-structured transmission dynamic model to quantify the drivers of SARS-CoV-2 transmission in the Dominican Republic and investigated the impact of the vaccination campaign and other counterfactual vaccination scenarios. We found that despite substantial prior accumulation of post-infection immunity, the vaccination campaign had an important impact on disease burden in 2021 and was essential in enabling a return to pre-pandemic mobility levels without incurring substantial additional burden. We estimate the campaign averted 7210 hospital admissions (95% CrI: 6830 – 7600), 2180 ICU admissions (95% CrI: 2080 – 2280) and 766 deaths (95% CrI: (694 – 859), in the first 6 months of the campaign. In addition to this, we found that earlier vaccination with Sinovac-Coronavac was preferable to delayed vaccination with a higher efficacy product, resulting in a lower burden over the subsequent waves of 2021. While these findings are contingent on local epidemiological factors in the Dominican Republic (particularly the speed of vaccine roll-out and the introduction of the Mu variant in the summer of 2021), they

illustrate the importance of timely and equitable access to vaccines during pandemics and large-scale outbreaks.

From 2020-2021 the Dominican Republic experienced four distinct waves of SARS-CoV-2 transmission. We found that, after the initial emergence of SARS-CoV-2 in March 2020, the first wave was largely controlled by the imposition of NPIs and the associated sharp drop in contact rates. The subsequent build-up of immunity in the following months maintained an estimated R_t of around 1 until the end of the 2020, despite the gradual increase of social contact rates from their trough. This contrasts with other settings with well-characterised transmission dynamics, such as the United Kingdom, where SARS-CoV-2 dynamics were largely driven by the imposition of NPIs before the widespread rollout of vaccination (19). The second wave from November - February 2021 was driven by a spike in contacts in December 2021, while the third wave during the summer of 2021 is best explained by gradually increasing contact rates back to pre-pandemic levels alongside the emergence of more transmissible variants, including the Mu variant. Finally, the fourth wave between September and December 2021 was driven by the Delta variant, alongside the reopening of schools and the relaxation of NPIs including curfews.

We estimated that the 43% two-dose coverage achieved by the vaccination campaign by mid-August 2021 would have offset a 10-20% increase in mobility – a proxy for social interactions – in this period, quantifying the ‘return-to-normality’ enabled by the vaccination campaign. Indeed, from July 2021 the Dominican Republic began to reopen the economy, culminating in the removal of curfew measures in October 2021 with population mobility almost returning to pre-pandemic baseline levels. This contrasts with other settings where higher levels of vaccination coverage were required to lift measures, such as the United Kingdom, which relaxed many measures in the summer of 2021 with a two-dose vaccine coverage of around 60% and population mobility still well below baseline [22,34]. Many other countries were unable to lift measures before intense Omicron transmission generated substantial population immunity or until very high vaccination coverage was achieved [35]. The trade-off between vaccination and population mobility on disease burden is likely to differ depending on the setting and epidemiological context. For instance, countries with lower levels of post-infection immunity would likely see greater changes in burden associated with changes in vaccination coverage. This balance will also be affected by the

emergence of new variants which may be more transmissible or exhibit immune evasive properties.

There are several limitations to this analysis. During the first wave (March - September 2020), the model struggled to reproduce the observed pattern in reported deaths. This may reflect limits in testing infrastructure and COVID-19 death reporting during this period, as observed in many countries globally [36]. Hospitalisation and ICU data were only publicly reported from September 2020, and deaths were probably under-reported early in the pandemic. While we partially accounted for changes in COVID-19 death reporting by allowing the infection-fatality ratio to vary over time, there remains substantial uncertainty in the modelled size and timing of the first wave.

Our modelling framework only incorporates protection from full primary vaccination with two doses and does not incorporate protection from a single dose. This would result in an underestimation of the impact of vaccination. However, as estimated protection of a single dose of Coronavac is low (particularly against the Delta strain), we do not expect this to have an important impact on our results [27]. Additionally, we assume fully vaccinated individuals are immediately afforded protection according to vaccine efficacy estimates used. Similarly, we also do not take into account protection from a single dose of Oxford-Astrazeneca and Pfizer-BioNTech vaccination in our counterfactual analysis and assume that fully vaccinated individuals are immediately afforded protection once vaccinated. The combined effect of these two assumptions may be to overestimate the vaccine efficacy of the alternative vaccine products. However, when we conducted sensitivity analysis considering lower vaccine efficacy estimates we found the direction and magnitude of our results did not change, providing evidence that the conclusions from our assessment of potential vaccination strategies are robust (Supplementary Table 8). We do not consider the impact of booster vaccination, which had begun in the Dominican Republic by late 2021, and we do not consider any additional benefit afforded by vaccination for individuals with post-infection immunity. Due to limited information on the introduction and epidemiological characteristics of variants introduced in early 2021, we parameterised the model for wild-type and Delta variants and modelled the effect of Mu and other VOC/VOIs in mid 2021 through a fitted sinusoidal increase in transmissibility over this period. We therefore do not capture the effect of immune evasion of Mu (or other variants such as Gamma) on the epidemic dynamics. Assuming

increased transmissibility of Mu or other variants, rather than immune evasion, would result in an underestimation of the impact of the vaccination campaign, as fewer individuals would remain susceptible during vaccine roll-out and therefore able to benefit from post-vaccination rather than post-infection protection.

Despite this, our analysis untangles the complex interactions between population behaviour, the introduction of variants and changes in population immunity in the Dominican Republic, enabling us to estimate the impact of vaccination in a complex immunological landscape and consider other counterfactual scenarios. We quantify the impact of these epidemic drivers in a setting with high seroprevalence during vaccination rollout, providing alternative insights to much comparable modelling in high-income countries. Our conclusions are therefore likely to be relevant to many other countries that were unable to suppress transmission through NPIs prior to vaccination roll-out. We also highlight the importance of having multiple data streams available to accurately characterise transmission dynamics during an epidemic. A particular strength of this study is the representativeness of the serological data used to parameterise the model and estimate population infection history, which was generated through one of the few national SARS-CoV-2 serological studies conducted in Latin America and the Caribbean during the pandemic using a rigorous multistage study design [8]. Similarly, the availability of hospital and ICU occupancy data is crucial for understanding how the relationship between infection, severe outcomes and death modulates during the epidemic due to improved treatment, the introduction of VOC/VOIs, and vaccination. Understanding these dynamics in real-time is essential to avoid potential problems such as reopening the economy too late when the population has high levels of immunity or delaying the re-imposition of NPIs when new variants emerge or contact rates increase unexpectedly. Ensuring that reliable data streams can be set-up quickly across both high income and low- and middle income countries should be a priority for future pandemic planning and preparedness.

Data sharing

All code and data used for this analysis are available at: <https://github.com/EmilieFinch/DR-covid19>

Contributors

Conceptualization: EF, AJK, RL, EJM, CL. Funding acquisition: AJK, RL, EJM, CL. Methodology: EF, AJK, RL. Project administration: EF. Investigation: EF. Supervision: AJK, RL. Data analysis: EF. Figures: EF. Writing - original draft: EF, AJK, RL. Writing - review and editing: all authors. EF and AJK accessed and verified the data.

Declaration of interests

EF was supported by the Medical Research Council (MR/N013638/1); AJK was supported by Wellcome Trust (206250/Z/17/Z); RL was supported by a Royal Society Dorothy Hodgkin Fellowship; EJM was supported by the US CDC (U01GH002238), CL was supported by an Australian National Health and Medical Research Council Investigator Grant (APP1158469). AJK, EF, CL have received salaries, consultancy fees, or travel paid through the US CDC award (U01GH002238). CTP and RS-R are employees of the Ministry of Health and Social Assistance, Dominican Republic, that was subcontracted with funds from the US CDC award.

Acknowledgments

We would like to thank the participants of the serosurvey in the Dominican Republic and the field teams that collected the data as well as the Dominican Republic Ministry of Health and Social Assistance for their support for this work. We would also like to thank Nicholas G. Davies, Rosana C. Barnard and Lloyd A.C. Chapman for helpful discussions in the completion of this work.

References

1. Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat Med*. 2021 Jul;27(7):1205–11.
2. Chen LL, Abdullah SMU, Chan WM, Chan BPC, Ip JD, Chu AWH, et al. Contribution of low population immunity to the severe Omicron BA.2 outbreak in Hong Kong. *Nat Commun*. 2022 Jun 24;13(1):3618.
3. Coronavirus (COVID-19) Infection Survey technical article: Cumulative incidence of the number of people who have tested positive for COVID-19, UK - Office for National Statistics [Internet]. [cited 2023 May 26]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticlecumulativeincidenceofthenumberofpeoplewhohavetestedpositiveforcovid19uk/22april2022>
4. Buss LF, Prete CA, Abraham CMM, Mendrone A, Salomon T, Almeida-Neto C de, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*. 2021 Jan 15;371(6526):288–92.
5. COVID-19 in Latin America—emergency and opportunity. *The Lancet*. 2021 Jul 10;398(10295):93.
6. Caribe CE para AL y el. Dos años de pandemia de COVID-19 en América Latina y el Caribe: reflexiones para avanzar hacia sistemas de salud y de protección social universales, integrales, sostenibles y resilientes [Internet]. CEPAL; 2022 [cited 2023 Aug 14]. Available from: <https://www.cepal.org/es/publicaciones/47914-anos-pandemia-covid-19-america-latina-caribe-reflexiones-avanzar-sistemas-salud>
7. Paulino-Ramírez R, López P, Mueses S, Cuevas P, Jabier M, Rivera-Amill V. Genomic Surveillance of SARS-CoV-2 Variants in the Dominican Republic and Emergence of a Local Lineage. *Int J Environ Res Public Health*. 2023 Jan;20(8):5503.
8. Nilles EJ, Paulino CT, de St. Aubin M, Restrepo AC, Mayfield H, Dumas D, et al. SARS-CoV-2 seroprevalence, cumulative infections, and immunity to symptomatic infection – A multistage national household survey and modelling study, Dominican Republic, June–October 2021. *Lancet Reg Health - Am*. 2022 Dec 1;16:100390.

9. Barber RM, Sorensen RJD, Pigott DM, Bisignano C, Carter A, Amlag JO, et al. Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis. *The Lancet*. 2022 Jun 25;399(10344):2351–80.
10. Pérez-Then E, Miric M, Qian HZ, Chen YQ, Wang Y, Vallejo V, et al. Population-Level Effectiveness of an Inactivated Whole-Virion COVID-19 Vaccine: A Test Negative Case-Control Study in the Dominican Republic. *Open Forum Infect Dis*. 2023 Mar 1;10(3):ofad075.
11. Pérez-Then E, Lucas C, Monteiro VS, Miric M, Brache V, Cochon L, et al. Neutralizing antibodies against the SARS-CoV-2 Delta and Omicron variants following heterologous CoronaVac plus BNT162b2 booster vaccination. *Nat Med*. 2022 Mar;28(3):481–5.
12. Caldwell JM, de Lara-Tuprio E, Teng TR, Estuar MRJE, Sarmiento RFR, Abayawardana M, et al. Understanding COVID-19 dynamics and the effects of interventions in the Philippines: A mathematical modelling study. *Lancet Reg Health West Pac*. 2021 Sep;14:100211.
13. Adhikari K, Gautam R, Pokharel A, Dhimal M, Uprety KN, Vaidya NK. Insight into Delta variant dominated second wave of COVID-19 in Nepal. *Epidemics*. 2022 Dec;41:100642.
14. Yang W, Shaman J. COVID-19 pandemic dynamics in India, the SARS-CoV-2 Delta variant and implications for vaccination. *J R Soc Interface*. 2022 Jun 6;19(191):20210900.
15. Pearson CAB, van Zandvoort K, Jarvis C I, Davies N, Checchi F, CMMID COVID-19 Working Group, et al. Projections of COVID-19 epidemics in LMIC countries [Internet]. 2020 Apr [cited 2021 Jun 21]. Available from: <https://cmmid.github.io/topics/covid19/LMIC-projection-reports.html>
16. Pearson CAB, Bozzani F, Procter SR, Davies NG, Huda M, Jensen HT, et al. COVID-19 vaccination in Sindh Province, Pakistan: A modelling study of health impact and cost-effectiveness. *PLoS Med*. 2021 Oct;18(10):e1003815.
17. Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *Lancet Infect Dis*. 2022 Sep;22(9):1293–302.
18. Gozzi N, Chinazzi M, Dean NE, Longini Jr IM, Halloran ME, Perra N, et al. Estimating the impact of COVID-19 vaccine inequities: a modeling study. *Nat Commun*. 2023 Jun 6;14(1):3272.
19. Ministerio de Salud Pública, de la República Dominicana. Boletines sobre el COVID-19

- [Internet]. [cited 2022 Nov 18]. Available from: https://www.msp.gob.do/web/?page_id=6948
20. Ainsworth M, Andersson M, Auckland K, Baillie JK, Barnes E, Beer S, et al. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. *Lancet Infect Dis* [Internet]. 2020 Sep 23 [cited 2020 Sep 30];0(0). Available from: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30634-4/abstract](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30634-4/abstract)
 21. Nilles EJ, Karlson EW, Norman M, Gilboa T, Fischinger S, Atyeo C, et al. Evaluation of Three Commercial and Two Non-Commercial Immunoassays for the Detection of Prior Infection to SARS-CoV-2. *J Appl Lab Med*. 2021 Jul 1;6(6):1561–70.
 22. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B, et al. Coronavirus Pandemic (COVID-19) - Dominican Republic Country Profile. *Our World Data* [Internet]. 2020 Mar 5 [cited 2021 Jul 30]; Available from: <https://ourworldindata.org/coronavirus/country/dominican-republic>
 23. Davies NG, Barnard RC, Jarvis CI, Russell TW, Semple MG, Jit M, et al. Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. *Lancet Infect Dis*. 2021 Apr 1;21(4):482–92.
 24. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health*. 2020 Jul 1;5(7):e375–85.
 25. Barnard RC, Davies NG, Jit M, Edmunds WJ. Modelling the medium-term dynamics of SARS-CoV-2 transmission in England in the Omicron era. *Nat Commun*. 2022 Aug 19;13(1):4879.
 26. Wu D, Zhang Y, Tang L, Wang F, Ye Y, Ma C, et al. Effectiveness of Inactivated COVID-19 Vaccines Against Symptomatic, Pneumonia, and Severe Disease Caused by the Delta Variant: Real World Study and Evidence - China, 2021. *China CDC Wkly*. 2022 Jan 28;4(4):57–65.
 27. Imai N, Hogan AB, Williams L, Cori A, Mangal TD, Winskill P, et al. Interpreting estimates of coronavirus disease 2019 (COVID-19) vaccine efficacy and effectiveness to inform simulation studies of vaccine impact: a systematic review [Internet]. *Wellcome Open Research*; 2021 [cited 2022 Jan 13]. Available from: <https://wellcomeopenresearch.org/articles/6-185>

28. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Comput Biol*. 2017 Sep 12;13(9):e1005697.
29. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Auzenbergs M, O'Reilly K, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med*. 2020 May 7;18(1):124.
30. Braak CJFT. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Stat Comput*. 2006 Sep 1;16(3):239–49.
31. Rubin AG John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, Donald B. Bayesian Data Analysis. 3rd ed. New York: Chapman and Hall/CRC; 2015. 675 p.
32. COVAX roll-out - Dominican Republic | Gavi, the Vaccine Alliance [Internet]. [cited 2023 May 25]. Available from: <https://www.gavi.org/covax-vaccine-roll-out/dominican-republic>
33. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*. 2017 Mar 30;22(13):30494.
34. Google. COVID-19 Mobility Reports [Internet]. Available from: <https://www.google.com/covid19/mobility/>
35. Poon RWS, Chan BPC, Chan WM, Fong CHY, Zhang X, Lu L, et al. SARS-CoV-2 IgG seropositivity after the severe Omicron wave of COVID-19 in Hong Kong. *Emerg Microbes Infect*. 2022 Dec 31;11(1):2116–9.
36. Russell TW, Golding N, Hellewell J, Abbott S, Wright L, Pearson CAB, et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med*. 2020 Oct 22;18(1):332.

Chapter 5

Disentangling the role of climate and serotype competition to forecast dengue outbreaks in Singapore

Dengue transmission in hyperendemic settings occurs in a landscape of complex population immunity. While SARS-CoV-2 population immunity (discussed in the previous chapter) is complicated by vaccination, waning immunity and the introduction of novel variants, dengue dynamics (the focus of this chapter) are the result of complex interactions between serotypes. As infection with one serotype results in lifelong immunity to the infecting serotype and short-term cross-immunity, changes in the distribution of serotype impact population vulnerability to dengue outbreaks. Climatic variation also impacts transmission, shaping dengue seasonality as well as the timing and magnitude of dengue outbreaks. Despite this, to date few climate-informed early warning forecasting frameworks account for the effects of immunity, potentially limiting the ability to forecast large outbreaks with operationally useful lead time.

In this study, I aimed to address this research gap by incorporating climate and serotype data into a Bayesian hierarchical forecasting model for Singapore, using data from 2000-2023. I used weekly climate, case and serotype surveillance data, leveraging daily rainfall data to construct more informative indicators for dengue transmission, and tested climatic variables with different aggregations and lags. I quantified the impact of temperature, rainfall and El Niño sea surface temperature anomalies on dengue incidence, as well as the time since a switch in dominant

serotype, disentangling the impacts of climate and immunity. I then adapted the model for use in early-warning, producing probabilistic forecasts of dengue cases with 2 to 8 week lead times. I compared forecast skill of a model including both climate and serotype information with a climate-only and serotype-only model, as well as a seasonal baseline, including only weekly random effects. I assessed predictive accuracy using proper scoring rules as well as outbreak detection metrics to evaluate the model's potential public health utility in an early warning system.

The supplementary material for this chapter is included as Appendix D.



London School of Hygiene & Tropical Medicine
 Keppel Street, London WC1E 7HT
 T: +44 (0)20 7299 4646
 F: +44 (0)20 7299 4656
 www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1802674	Title	Miss
First Name(s)	Emilie		
Surname/Family Name	Finch		
Thesis Title	Modelling the role of immunity, climate and behaviour in viral outbreak dynamics and control		
Primary Supervisor	Adam Kucharski		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Nature Microbiology
Please list the paper's authors in the intended authorship order:	Emilie Finch, Adam Kucharski, Shuzhen Sim, Lee-Ching Ng, Rachel Lowe
Stage of publication	Not yet submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I worked on: conceptualisation, methodology, data curation, investigation, writing original draft and review and editing of submitted manuscript.
--	---

SECTION E

Student Signature	
Date	08/07/2024

Supervisor Signature	
Date	12/07/2024

5.1 Abstract

Dengue poses a rapidly increasing threat to global health, with Southeast Asia as one of the worst affected regions. Climate-informed early warning systems can help to mitigate the impact of outbreaks; however, prediction of large outbreaks with sufficient lead time to guide interventions remains a challenge. In this work, we quantify the role of climatic variation and serotype competition in shaping dengue risk in Singapore using over 20 years of weekly case data. We integrated these findings into an early warning system framework able to predict dengue outbreaks up to 2 months ahead. While climatic variation improved the predictive power of the model by 54% compared to a seasonal baseline, including additional serotype information increased predictive performance to 60%, helping to explain interannual variation. By incorporating serotype competition as a proxy for population immunity, this work advances the field of dengue prediction and demonstrates the value of long-term virus surveillance

5.2 Introduction

Climate-informed forecasting models can be used to give advance warning of infectious disease outbreaks and mitigate their impact. Early warning systems integrating climate information are becoming an increasingly important tool for epidemic response as the climate crisis leads to more frequent climatic extremes and shifts the dynamics of climate-sensitive infectious diseases, such as dengue [1,2].

Dengue is an emerging vector-borne disease, transmitted by *Aedes aegypti* and *Aedes albopictus* in urban and peri-urban areas [3]. Global reported dengue incidence has increased by 30-fold over the past 50 years, alongside increases in the geographical range of transmission, and approximately half of the global population is thought to be at risk of dengue transmission [4]. There are four antigenically distinct serotypes of dengue virus (DENV1-4): infection with one serotype results in life-long immunity to the infecting serotype, and limited cross-immunity to others [5,6]. Dengue epidemic dynamics are complex and challenging to predict, with large outbreaks driven by multiple factors including climate variation, competition between the four dengue serotypes and traditional and novel vector-control efforts.

Singapore, an equatorial city-state in Southeast Asia, experiences hyperendemic dengue transmission with all four serotypes in circulation and cyclical replacements in the dominant serotype [7]. Singapore experiences warm and humid temperatures year round, with suitable conditions for mosquito breeding and dengue transmission. Rainfall is affected by two monsoon seasons, with the Northeast monsoon occurring from December until early March and the Southwest monsoon occurring from June to September [8]. The peak dengue season usually occurs between June and October, typically following the warmest and most humid months of the year. Since the 1960s, Singapore has implemented stringent dengue prevention measures focused on vector control and public education. This has led to a reduction in the *Aedes* House Index (AHI, a measure of the percentage of houses positive for *Aedes* breeding) from 48% in 1966 to around 1% in the 1990s [9]. Periodic seroprevalence surveys have demonstrated a concurrent decrease in seroprevalence in almost all age groups, with an estimated decrease in force of infection (FOI) from around 0.1 per year in the 1960s to 0.01 per year from the 1990s onwards [10,11]. This is

also reflected in surveillance data showing an increase in the average age of reported cases from children to young adults [7,9,12]. Despite this, reported cases have increased in recent years. Possible explanations for this include improved case detection and reporting, particularly after 2008 where a campaign was launched encouraging the use of NS1 rapid tests in laboratories or higher population vulnerability to dengue outbreaks (for example, after the importation of a new viral genotype) due to lower immunity [9,11].

Singapore experiences cyclical dengue outbreaks which have been increasing in frequency and magnitude [7,9]. These have followed switches in dominant circulating serotypes, with the exception of the 2019 outbreak. Between 2006 and 2020, virus surveillance showed the dominant serotype switching between DENV1 and DENV2 [7,13]. Accordingly, population immunity to DENV1 and DENV2 is believed to be higher than against DENV3 and DENV4. For example, a study of healthy blood donors in 2009 amongst 16 to 60 year olds found seropositivity of 35.8% anti-DENV1 and 36.4% anti-DENV2 compared with 15.4% anti-DENV3 and 7.7% anti-DENV4 [9,10]. However, the cyclical pattern of DENV1 and 2 dominance has been disrupted in recent years. A large, predominantly DENV2 outbreak in 2019-2020 saw increasing contribution from DENV3, which then gained predominance in the next outbreak year in 2022 [14]. Dengue outbreaks have also been associated with El Niño events, the warm phase of the El Niño Southern Oscillation (ENSO), involving warmer than normal oceanic and atmospheric temperatures in the Pacific, which typically lead to hotter and drier climatic conditions in Singapore [15–17].

Climate influences dengue transmission through effects on the vector (*Aedes* mosquitos) and the dengue virus itself. Temperature affects mosquito survival, development and reproduction, as well as the viral extrinsic incubation period, with an optimal temperature for transmission of around 29°C and thermal limits between 17.8°C – 34.5°C [18]. Temperature has been found to shape the timing, length and geographical extent of dengue seasons [2,19]. Contrastingly, the impact of rainfall on dengue transmission is more nuanced. While increasingly wet and humid conditions can lead to the creation of mosquito breeding sites, the effect of rainfall on transmission is dependent on human water storage behaviour, and the availability of water and sanitation infrastructure, which can lead to non-linear and delayed impacts of rainfall on dengue transmission [20–22]. In particular, excessive rainfall can lead to flushing effects, where mosquito breeding

habitats are washed away entirely. This has been documented in Singapore, where dengue outbreak risk was found to decrease following flushing events [23]. Previous research in Singapore has demonstrated the utility of temperature, precipitation and absolute humidity in predicting dengue incidence [24–26]. Currently, Singapore uses a machine learning approach to generate operational forecasts of dengue incidence for outbreak alerts and decision-support [27]. This methodology is based on LASSO (least absolute shrinkage and selection operator) regression analysis, which applies a penalty term to improve predictive ability through regularization and variable selection.

Forecasting models can capitalise on inherent lags between climatic variation and dengue transmission to predict outbreak risk at operationally useful lead times [28]. In Singapore, an analysis of vector control found that local authorities needed an average of 2 months to mitigate the impact of a dengue outbreak, suggesting that early warning forecasts with several months lead time would be optimal [29]. Forecasts at shorter horizons may also be helpful to inform situational awareness. To date, dengue forecasting models have struggled to predict interannual variability in dengue seasons and shown worse predictive performance for high incidence seasons, which have the greatest public health impact [30]. Additionally, forecast skill is typically lower earlier in the season when aiming to forecast several months ahead, which is when forecasts have the most potential operational value. While immunity is theoretically recognised as an important driver of interannual dengue dynamics, to our knowledge no current dengue forecasting models directly account for serotype dynamics or changes in immunity, which are likely to be particularly important in hyperendemic regions such as Southeast Asia. To address this, we incorporate climate and serotype dynamics within a Bayesian hierarchical modelling framework to forecast dengue incidence. We quantify the effect of climatic variables, the Niño 3.4 index and switches in dominant serotype on dengue incidence, and evaluated forecasts of dengue incidence with a 2-8 week forecast horizon.

5.3 Results

5.3.1 Reported dengue cases in Singapore over the past two decades

Between 1 January 2000 and 31 December 2022, 234,358 cases of dengue were reported in Singapore. Most cases were reported between June and September, typically following warm and humid climatic conditions (Figure 1, panels D-F). Cyclical outbreaks, which historically were thought to occur every 5-6 years, have become more frequent and of larger magnitude (Figure 1). For instance, while the 2004 dengue outbreak led to 9,459 cases overall and peaked at 332 reported weekly cases, the 2022 outbreak resulted in 32,259 reported cases and peaked at 1,568 reported weekly cases. We defined an outbreak week using a seasonal moving 75th percentile threshold. For a given month and year, we defined an outbreak threshold of the 75th percentile of weekly cases in that month using all years up to (but not including) the given year. This identifies the same outbreak periods as the endemic channel threshold used within the NEA but has several added benefits. Firstly, a percentile threshold is simpler to calculate and adjusts on a rolling basis rather than year-on-year. Additionally, by incorporating the seasonal patterns underlying dengue transmission, this definition allows outbreak periods to be defined earlier than with a fixed, non-seasonal threshold (Supplementary information, Figure 1).

We then defined an outbreak year as a year containing more than 12 outbreak weeks. This identified outbreaks in; 2004, 2005, 2007, 2013, 2014, 2015, 2016, 2019, 2020 and 2022. In some years, dengue outbreaks coincided with switches in dominant serotype; such as the switch from DENV-2 to DENV-1 in 2013 or from DENV-2 to DENV-3 in 2022 (Figure 1). Additionally, outbreaks sometimes coincided with El Niño events, defined when sea surface temperature (SSTs) in the Niño 3.4 region exceed 0.5°C for 5 consecutive months.

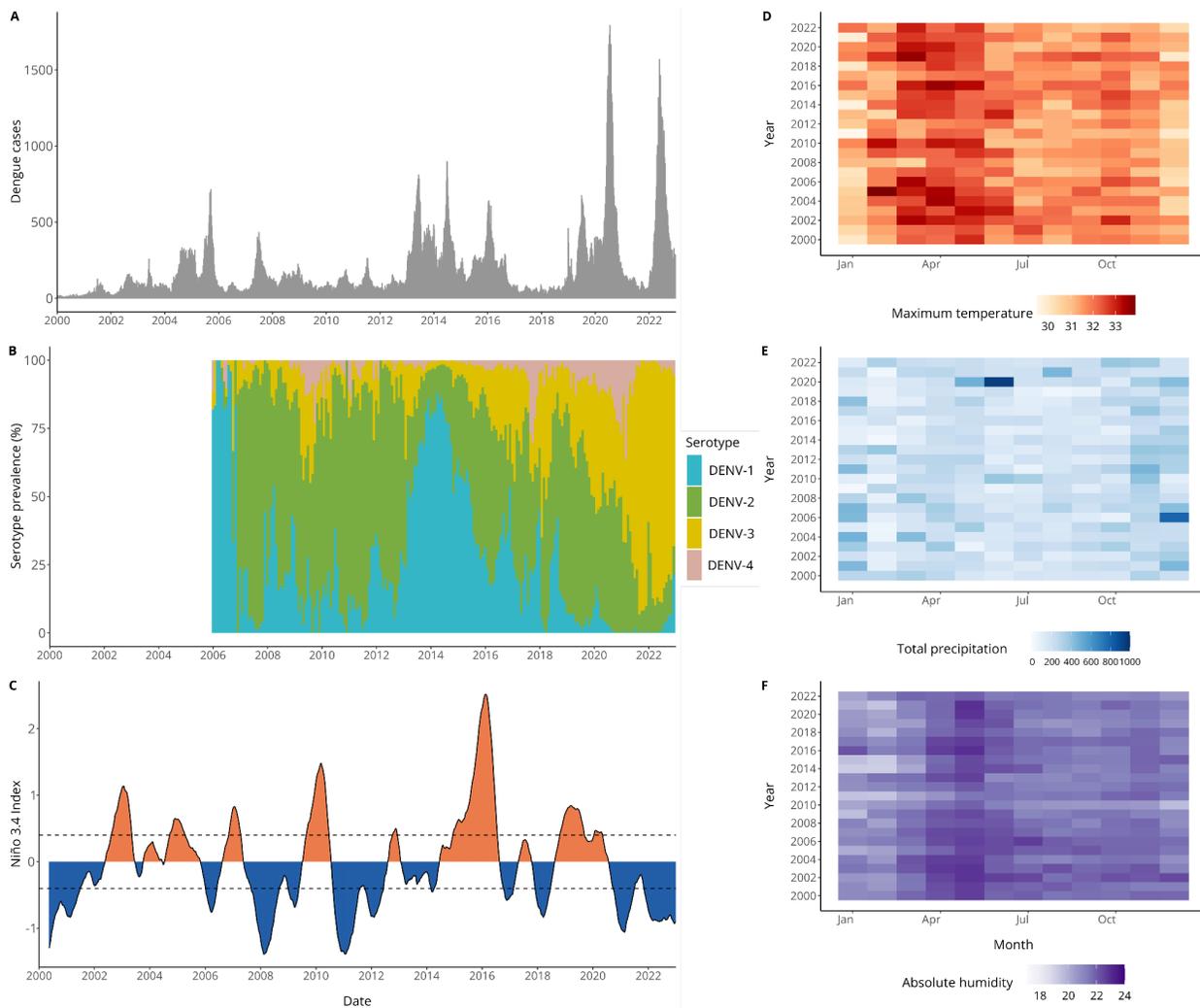


Figure 5.1 Dengue cases, climate variability and serotype dominance in Singapore.

Figure showing epidemiological and climatic data for Singapore from 2000 – 2023. A) Bars show weekly reported dengue cases from 1 January 2000 – 31 December 2022. B) Stacked bars show the proportion of each serotype detected through virus surveillance. To calculate this, we aggregate the number of serotyped cases for DENV1-4 at a monthly level and divide each by the total number of cases serotyped in that month. C) Line graph showing weekly Niño 3.4 sea surface temperature anomalies, dashed lines indicate $+0.5^{\circ}\text{C}$ and -0.5°C which are often used as thresholds to define El Niño and La Niña events. Blue shading indicates negative values of the Niño 3.4 index (indicating La Niña conditions), while orange shading indicates positive values, associated with El Niño conditions. Heatmaps for (D) monthly mean maximum temperature ($^{\circ}\text{C}$), (E) total precipitation (mm), and (F) mean absolute humidity (g/m^3).

5.3.2 Climate and serotype dynamics shape dengue risk with non-linear and delayed effects

To quantify the effect of climate, ENSO, and changes in serotype on dengue risk, we fit a Bayesian hierarchical model to weekly case data. We used a negative binomial likelihood and incorporated weekly random effects to capture seasonality and yearly random effects to account for unexplained interannual variation in dengue risk, for instance, due to control measures (Methods). We then compared a baseline model (including only weekly and yearly random effects) with models containing climatic and serotype covariates to see whether their inclusion improved model adequacy statistics, and reduced unexplained interannual variation in the model. We tested temperature (non-linear and linear), precipitation (non-linear and linear), humidity (linear) and ENSO (non-linear and linear) variables considering lags from 0:20 weeks, as well as serotype variables (non-linear and linear). We selected a final model including: 12-week rolling average maximum temperature in °C; 12-week total days without rain; 12-week rolling average Niño 3.4 SSTA with a 4 week lag; and a time-varying covariate measuring the number of weeks since a switch in dominant serotype (Methods). We found a non-linear relationship between maximum temperature and dengue incidence risk, with increased risk around 32 °C and decreased risk at particularly low or high maximum temperatures (Figure 2, panel A). Similarly, we found increased risk of dengue at intermediately wet conditions, with around 30 days without rain in the previous 3 months, and decreased risk in dry conditions, with more than 45 days without rain in the previous 3 months (Figure 2, panel B). We found decreased dengue risk with negative Niño 3.4 SSTA values and non-linearly increasing dengue risk with values of Niño 3.4 SSTA upwards of around 1.4 (Figure 2, panel C). Finally, we found a non-linear relationship between the time since a switch in dominant serotype and dengue risk, with increased risk in the first two years following a switch, decreased risk between 2-6 years following a switch, and subsequent increased risk at 6+ years following a switch (Figure 2, panel D).

We then compared the yearly random effects estimated for our final *climate and serotype* model, a *climate only* model including climatic covariates and random effects, a *serotype only* model including the serotype covariate and random effects, and a baseline model including weekly and yearly random effects (Figure 2, panel E). As yearly random effects account for unexplained

interannual variation in dengue incidence, we would expect these values to be closer to 0 when other model covariates are able to explain this variation. We found no difference in estimated yearly random effects between the four models before 2006, which we would expect as no serotype information is available before this date. From 2007 onwards, overall, models including serotype information tended to have yearly random effects closer to 0. By calculating the percentage reduction in mean absolute yearly random effects between covariate models and the baseline model, we can quantify how much model covariates explain interannual variation in dengue incidence (Methods). While a *climate only* model reduced unexplained interannual variation in dengue incidence by 4.1%, including additional serotype information (in the *climate and serotype* model) resulted in a 26.8% reduction. Contrastingly, a *serotype only* model reduced unexplained interannual variation by 19.4%.

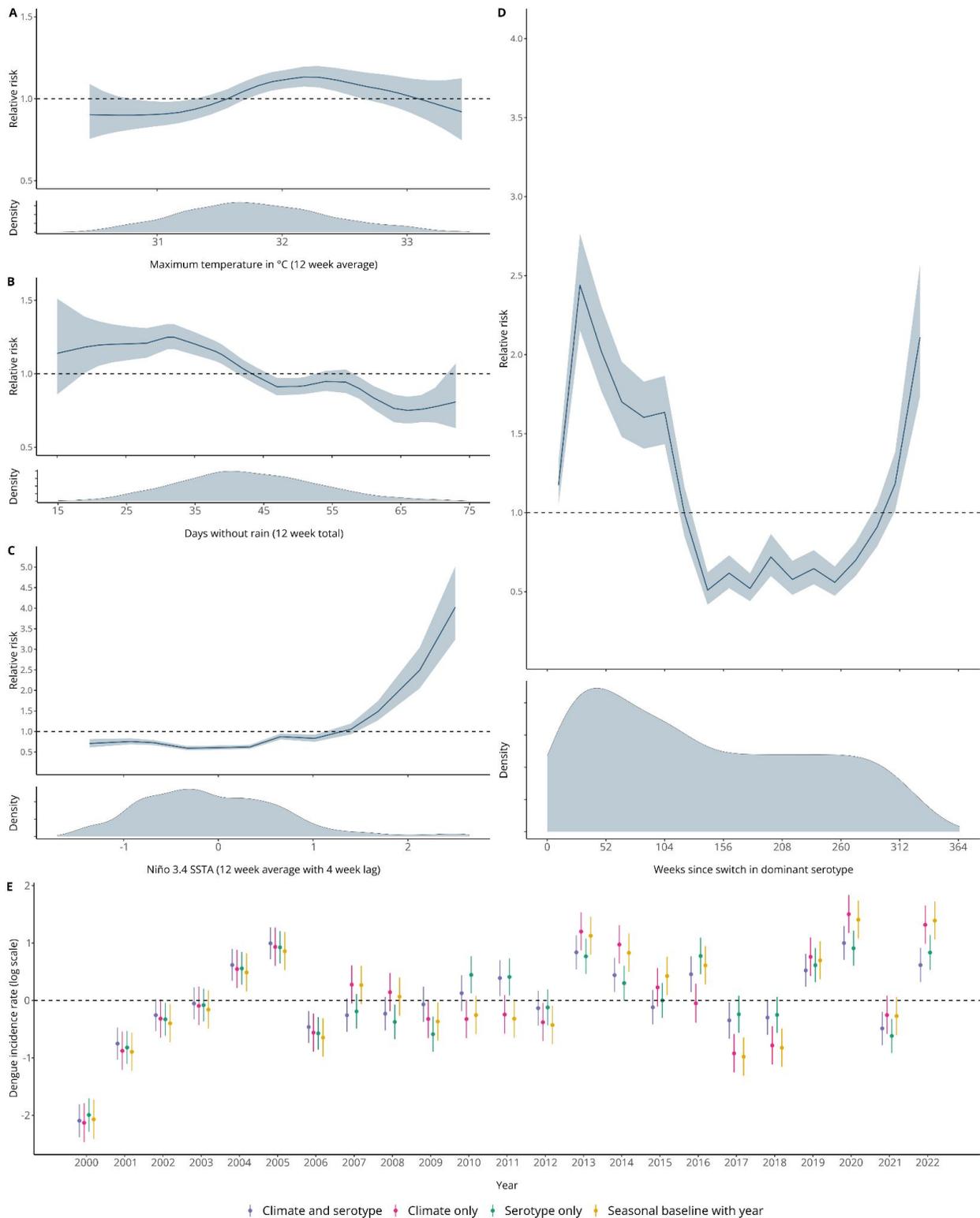


Figure 5.2 Effects of climatic variability and switches in dominant serotype on dengue incidence in Singapore

Panels A-D show posterior marginal effects and density plots for covariates in the final selected model. These include maximum temperature in °C (12 week running average), days without rain (12 week total), Niño 3.4 SSTA (12 week average with a 4 week lag) and weeks since switch in dominant serotype. These are shown on the relative risk scale displaying the median value and associated 95% credible interval and can be interpreted as the effect of the covariate on dengue incidence rate with all other parameters held constant. Panel E compares the estimated yearly random effect, $\gamma_{a[t]}$, for a *seasonal baseline with year* model including only weekly and yearly random effects $\gamma_{a[t]} + \delta_{w[t]}$ (in yellow), the final selected *climate and serotype* model including all climate and serotype covariates and random effects (in purple), a *climate only model* with random effects (in pink), and a *serotype only* model with random effects (in green). The estimated yearly random effect from the baseline model indicates whether dengue incidence was higher or lower for that year than the overall mean incidence. We would expect the estimated yearly random effects for covariate models to be closer to 0 (indicated with a dashed line) when covariates are able to account for interannual variability in dengue incidence.

5.3.3 Accounting for serotype and climate dynamics improves probabilistic predictions of dengue case incidence and outbreak detection

We used a time series cross-validation approach to produce probabilistic dengue predictions and calculate the probability of exceeding a predefined outbreak threshold in a given week [31]. This is an appropriate design to assess model utility for forecasting, as we preserve the underlying time-order of the data. We first generated probabilistic predictions using our candidate models with no lead time (using information up to a target date to predict dengue cases on that date) to compare out-of-sample predictive ability between 2009 and 2022 (Figure 3). The first 8 years of data were used exclusively for training (Methods). We compared our final selected *climate and serotype* model with a *climate only* model and a *serotype only* model containing only climatic and serotype covariates respectively (Supplementary information, Table 3). We compared these to a *seasonal baseline* model which only included weekly random effects. This is equivalent to a climatological baseline model which uses the average seasonal pattern in dengue incidence to predict cases in a given target week. We assessed forecast skill in predicting weekly dengue cases using the continuous ranked probability score (CRPS), where smaller values indicate better performance. We also calculated the continuous ranked probability skill score (CRPSS) which is defined as the percentage improvement in CRPS compared to a baseline model. We also assessed the predictive

ability of the candidate models for outbreak detection using the Brier score and conducted receiver operating characteristic (ROC) analysis. Here, we compared model hit rate (proportion of outbreak weeks correctly identified) with false alarm rate (proportion of weeks without an outbreak where an outbreak was predicted to occur). We calculated the area under the curve (AUC) to measure model skill in classifying outbreak and non-outbreak weeks (Methods).

The *climate and serotype* model was able to reproduce dengue epidemic dynamics in Singapore between 2009 and 2022 and had a lower CRPS than all other models, indicating better performance. In particular, the *climate and serotype* model is better able to predict the decrease in early 2016 than a *serotype-only* model (which predicts a late peak around July 2016). Similarly, the *climate and serotype* model outperforms the *climate-only* model in predicting the decrease in cases following peaks in mid-2013 and mid-2014, as well as better predicting peak cases in July 2019. All covariate models underpredicted the peak in 2020 but were able to accurately predict peak timing. Similarly in 2022, covariate models underpredicted the peak and predicted later peak timing than what was seen. The *climate and serotype* model showed a 60% relative improvement over the *seasonal baseline* model according to the CRPSS. The *climate only* and *serotype only* models also performed well, with a 54% and 49% relative improvement over the *seasonal baseline* model respectively (Supplementary information, Table 3). The *climate and serotype* model also outperformed other models in outbreak detection, with a lower Brier score (indicating better performance). This can be seen in Figure 4, where the *climate and serotype* model is better able to assign high probability of an outbreak to actual outbreak weeks and lower probability of an outbreak to non-outbreak weeks. The *climate and serotype* model also had the highest area-under-the-curve (AUC) (98%, 95% CI: 97.7 - 99.0%) and corresponding lowest false alarm rate (2.1%) and the highest hit rate (92%) of the candidate models, with an optimal model outbreak alert threshold of 71% (Supplementary information, Table 4). The *climate only* and *serotype only* models also performed well in outbreak detection, as can be seen from the overlapping ROC curves in Figure 4.

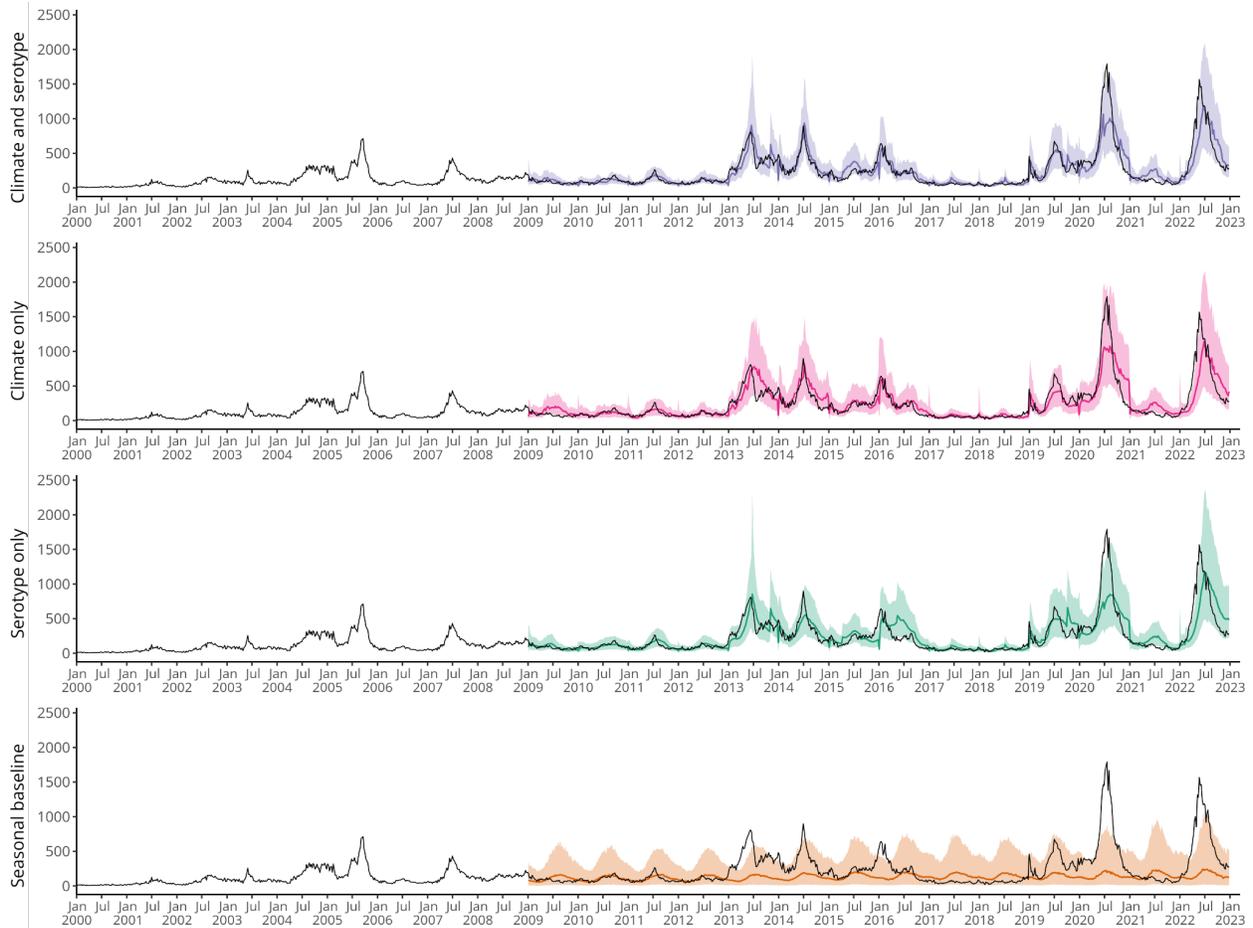


Figure 5.3 Comparing time series cross-validated predictions of candidate models.

Figure showing time series cross-validated posterior predictions of dengue cases for each model from 2009 – 2022. We used an expanding window cross-validation methodology, where the model is trained on data up to but not including the target week and then posterior predictions are generated for the target week. Coloured lines show the median posterior prediction of weekly dengue cases, shaded areas show the 95% credible interval and the dark grey lines show the data. From top to bottom the figure shows: predictions for the final selected ‘*Climate and serotype*’ model with weekly and yearly random effects $\gamma_{a[t]} + \delta_{w[t]}$ in purple; predictions for a ‘*Climate only*’ model with weekly and yearly random effects in pink; predictions for a ‘*Serotype only*’ model with weekly and yearly random effects in green; and predictions from a ‘*Seasonal baseline*’ model with only weekly random effects $\delta_{w[t]}$ in orange.

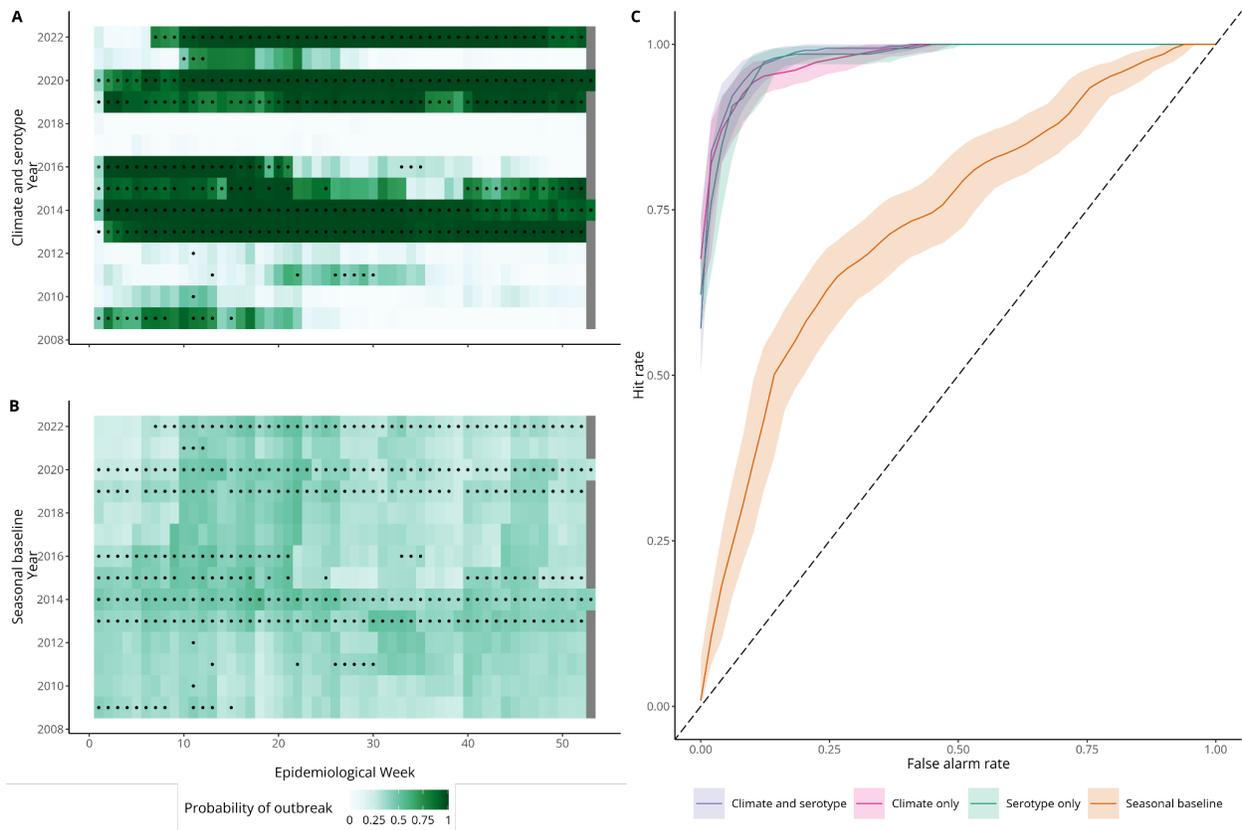


Figure 5.4 Comparing outbreak detection of candidate models.

Panels A and B show tile plots of model posterior predictions of exceeding the outbreak threshold for each week between 2009 – 2022. White indicates $P_{outbreak} = 0$, while dark green indicates $P_{outbreak} = 1$. Circles indicate observed outbreak weeks defined using a seasonal moving 75th percentile threshold. For a given month and year, we defined an outbreak threshold using the 75th percentile of weekly cases in that month using all years up to, but not including, the given year. Panel E shows an ROC curve, plotting hit rate against false alarm rate for different model outbreak alert thresholds. Hit rate (or sensitivity) is defined as the proportion of outbreak weeks that were correctly predicted. False alarm rate (1 - specificity) is defined as the proportion of weeks without an outbreak where an outbreak was predicted to occur. The shaded area shows the 95% confidence interval around the ROC curve.

5.3.4 Dengue forecasting for early warning with 2-8 weeks lead time

Having identified a model able to predict dengue cases and outbreak weeks with no lead time, we then adapted our framework for use in an early-warning context, producing forecasts with 2 to 8

weeks lead time. For each forecast horizon, we used the best approximation of the covariates in the final models available at the lead time considered (Methods, Supplementary information, Figure 3). We then produced probabilistic predictions of dengue using our candidate models to compare predictive ability with 2-8 week forecast horizons. Forecasts for 4 weeks ahead and 8 weeks ahead are shown in Figure 5, with forecasts for all horizons in Supplementary information, Figure 2. As expected, predictive ability declined with forecast horizon, with better performance at shorter lead times and increased uncertainty around model predictions at longer lead times (Figure 6). For instance, at an 8 week forecast horizon, the *climate and serotype* model struggled to predict peaks in late 2015-2016, first over and then under-predicting. However, all three covariate models considered were able to capture broad epidemic dynamics from 2009-2022 even at longer lead times and offered considerable improvement in performance compared with the *seasonal baseline*. For instance, at an 8 week forecast horizon the *climate and serotype* model showed 32% relative skill improvement over the *seasonal baseline*. Additionally, the *climate and serotype* model continued to accurately detect outbreak weeks at all forecast horizons, with an AUC of 94% (95% CI: 92.7 - 95.7%) at an 8 week forecast horizon (Supplementary information, Table 4).

We compared forecast skill metrics for each model from 0-8 weeks ahead and found the *climate and serotype* model had a lower CRPS than other models (indicating better performance) until a 6 week forecast horizon, when its performance was equivalent to the *climate only* model. However, when considering outbreak detection metrics (Brier score and AUC) the *climate and serotype* model outperformed all other models at all forecast horizons (Figure 6).

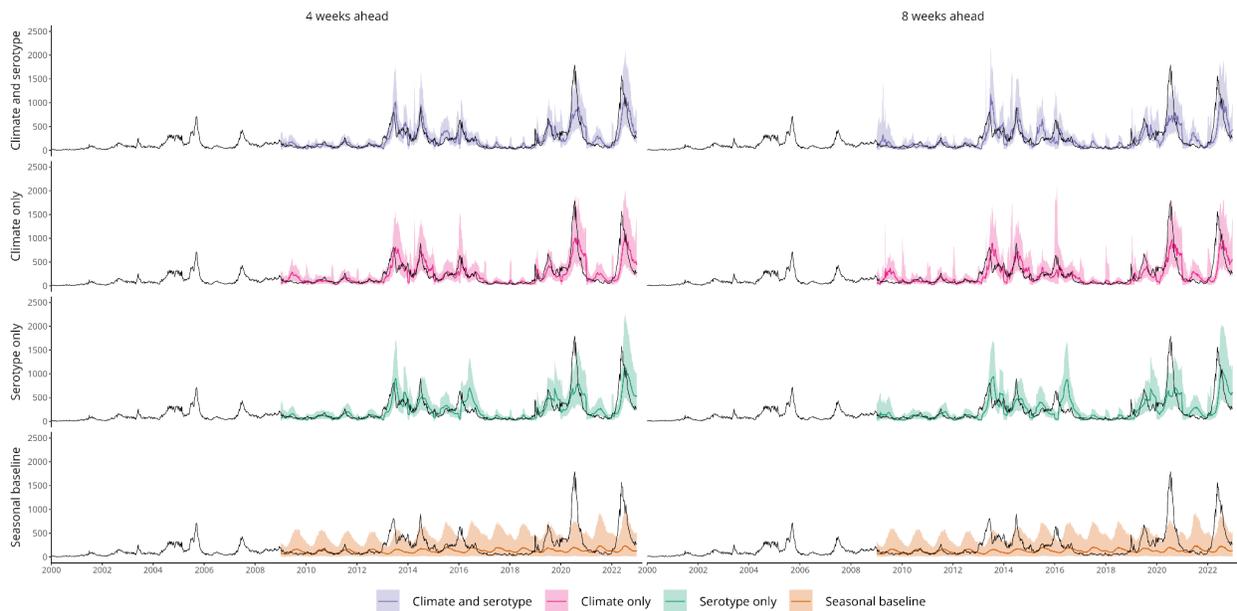


Figure 5.5 Dengue forecasts for early warning at 4 and 8 week forecast horizons.

Figure showing time series cross-validated posterior predictions of dengue cases for each model from 2009 – 2022 at 2 - 8 week forecast horizons. We used an expanding window time series cross-validation methodology, where posterior predictions for each week are generated from a model fit to data up to and including the target week. Coloured lines show the median posterior prediction of weekly dengue cases, shaded areas show the 95% credible interval and the dark grey line shows the data. From top to bottom the figure shows: predictions for the final selected *climate and serotype* model with weekly and yearly random effects $\gamma_{a[t]} + \delta_{w[t]}$ in purple; predictions for a *climate only* model with weekly and yearly random effects in pink; predictions for a *serotype only* model with weekly and yearly random effects in green; and predictions from a *seasonal baseline* model with only weekly random effects $\delta_{w[t]}$ in orange. From left to right each column shows forecasts at 4 and 8 weeks ahead respectively.

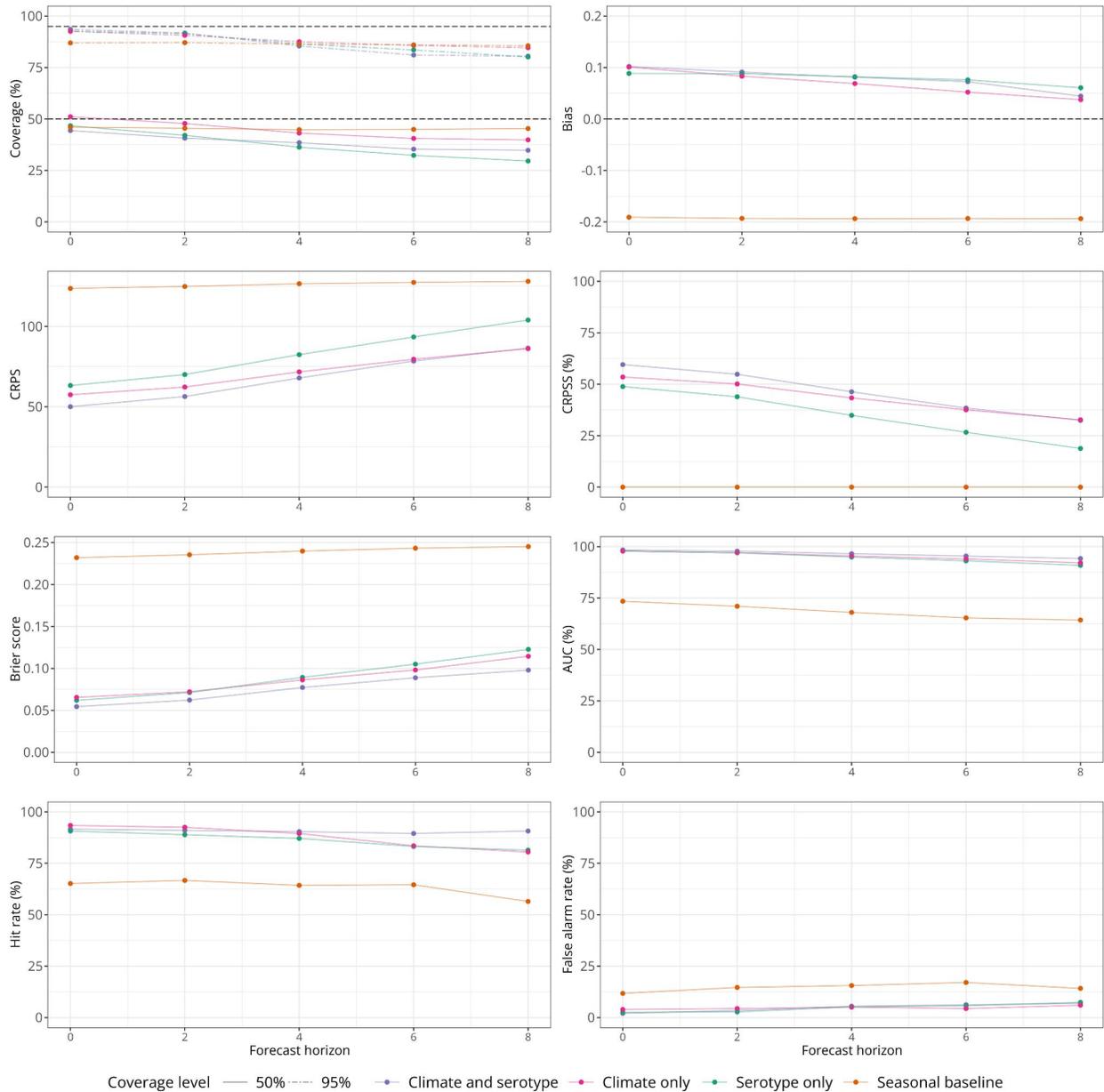


Figure 5.6 Predictive performance over different forecast horizons.

Figure showing forecast metrics for each model across all forecast horizons from 2009 – 2022. From top left to bottom right these show: interval coverage %, bias; CRPS (continuous ranked probability score), CRPSS (continuous ranked probability skill score, %), Brier score, AUC (area under the curve, %), hit rate (%) and false alarm rate (%). Interval coverage shows the percentage of observations falling inside a given prediction interval. A perfectly calibrated forecast would have coverage equal to the nominal prediction interval; that is, 95% coverage equal to 95% and 50% coverage equal to 50%, indicated by dashed horizontal lines. Bias measures the relative tendency of the model to over- or under-predict, and is bounded

between -1 and 1, with 0 indicating unbiased forecasts. The CRPS can take values between 0 and infinity, with smaller values indicating better performance. The CRPSS indicates the relative improvement of each covariate model over the *seasonal baseline* model and can take values from 0%, indicating that the model performs the same as the baseline, and 100%, indicating perfect forecasting skill. The Brier score can take values from 0 – 1, with smaller values indicating better performance. The AUC can take values from 0-100% with 100% indicating perfect classification. Hit rate and false alarm rate also take values from 0 – 100% with higher and lower values indicating better performance respectively.

5.4 Discussion

Climate and serotype dynamics have important impacts on dengue transmission and outbreak risk; however, to date few statistical forecasting models account for both drivers within the same framework. This could lead to misattribution of dengue risk caused by changes in population immunity to climatic or other non-climatic factors, and limit the ability to forecast large outbreaks in advance. Additionally, mechanistic approaches aiming to better capture the biological processes underlying transmission have been found to perform less well than statistical approaches to forecast dengue [30]. We analyse 20 years of data from Singapore to understand the relative impact of climatic and serotype dynamics on dengue risk and to produce probabilistic forecasts from 0 - 8 weeks ahead. Our approach integrates a proxy for changes in population immunity and the epidemic potential of the dominant virus in circulation, within a statistical framework, aiming to generate accurate probabilistic predictions of dengue incidence through improved inference.

The impacts of climatic variables on dengue risk in Singapore were complex, with non-linear and delayed effects. We found increased risk of dengue at a maximum temperature of 32 °C and at intermediately wet conditions, with decreasing risk in very hot and very dry conditions. This suggests that dengue seasonality in Singapore may change as climate change leads to increasing temperatures, with fewer cases in the middle of the year. We also found non-linearly increasing dengue risk with increasing Niño 3.4 SSTAs, reflecting increased risk during El Niño conditions. We found a non-linear relationship between the time since a switch in dominant serotype and dengue transmission. In the first two years following a switch, when population immunity to the new serotype is low, we found increased dengue risk, followed by decreased dengue risk in the

subsequent 4 years as immunity to the dominant serotype increases in the population. We then found evidence of increased risk at 6+ years following a serotype switch, which likely reflects the accumulation of susceptibles in the population as well as the risk associated with the growth of a non-dominant serotype before it reaches dominance, for instance growing DENV3 prevalence in 2019. Our results are in line with previous research on climate-dengue relationships in Singapore, finding that increases in temperature and precipitation increase dengue risk, as well as El Niño conditions [16,29]. However, we also found evidence of non-linearity in the temperature-dengue relationship, with decreased risk at high maximum temperatures. Xu and colleagues found that absolute humidity was a better predictor for dengue incidence than other climatic variables, due to the stability of its relationship with dengue incidence during different subperiods of serotype circulation [25]. In this work, we accounted jointly for the impact of serotype circulation and local weather indicators on dengue risk, and therefore were able to estimate the marginal effects of each climatic indicator without confounding from concurrent serotype dynamics. As a result of this we found stronger evidence for a role of temperature and rainfall in transmission, with little additional benefit of humidity information.

El Niño is thought to affect dengue transmission through changes in local weather conditions. We included covariates to capture both El Niño and local weather conditions within the same model structure, following the logic that the full effect of ENSO on transmission is unlikely to be fully captured by the weather covariates included in the model. For instance, ENSO may affect humidity, which was not explicitly included in the model, or temperature and precipitation metrics other than those included in the model, potentially affecting dengue season length or timing, or the spatial spread of dengue in Singapore.

We adapted our model into a proposed early-warning framework, generating accurate forecasts at operationally useful lead times [29]. We used a rigorous time series cross-validation methodology to realistically evaluate model performance for early warning. This retrospective statistical validation is a necessary step in the construction of a forecasting model for early-warning. When a model is implemented, forecasts are used for decision-making which directly impacts dengue transmission and complicates the evaluation of forecast accuracy. This analysis was designed in collaboration with stakeholders in the National Environment Agency of Singapore to address key

questions around local dengue transmission in Singapore and offers a more interpretable forecasting framework than current LASSO based dengue forecasts [27]. This framework could also be used to generate probabilistic forecasts under different scenarios (e.g. comparing forecasts with or without a switch in dominant serotype or under different possible Niño 3.4 SSTAs) to understand potential outbreak risks.

Accounting for serotype dynamics within our forecasting framework helped to explain interannual variability in dengue transmission, improved dengue case forecasts at shorter lead times (up to 4 weeks ahead) and increased outbreak detection accuracy at all lead times. By comparing our full *climate and serotype* model with models only containing *climate* or *serotype* information, we were able to identify time periods where particular covariates were improving predictions. For instance, including serotype information helped to predict decreases following peaks in 2013 and 2014, as well as the peak in 2019. Contrastingly, including climate information helped to predict peak intensity and timing in 2016, during an El Niño event. It should also be noted that while our *climate and serotype* and *climate only* models perform equally well when forecasting dengue cases at longer lead times (6-8 weeks ahead), this is largely due to the flexible random effects incorporated in the modelling framework. When generating probabilistic forecasts, the covariate models estimate yearly random effects based on data available for that year up until the forecast target week. We conducted sensitivity analysis around this, running these models without a yearly random effect. In this case, the *climate only* model performs worse than the *climate and serotype* model at all forecast horizons (Supplementary information, Figures 4-5). While this demonstrates that including serotype information is helping to predict interannual variability, it also highlights the power of a flexible random effects structure to account for unmeasured variation and improve seasonal forecasts, particularly for settings without regular virus surveillance.

Despite this, there are several limitations to this study. Our *climate and serotype* model struggled to predict peak weekly cases in 2020 and 2022. This is likely due to the increase in dengue transmission in Singapore seen under SARS-CoV-2 social distancing measures, which is not accounted for in our model framework [32]. We included serotype dynamics using the time since a switch in dominant serotype as a proxy for population immunity, but it would arguably be preferable to include measures of seroprevalence or estimates of the proportion susceptible

directly. We were also unable to include serotype-specific dynamics due to the low number of switches in dominant serotype, even within a 17 year period of virus surveillance. This could be important to include as certain serotypes are associated with greater clinical severity and secondary infections are thought to be associated with increased severity as a result of antibody dependent enhancement (27, 28). Similarly, we only consider switches in antigenic serotype and don't consider changes in genotype prevalence, which are also hypothesised to increase outbreak risk [35]. We did not incorporate vector density into the model, which mediates the relationship between climatic conditions and dengue transmission. We did not have access to data on vector control efforts, therefore any effects of interventions would be accounted for via yearly random effects. However, it should be noted that vector control is stringently maintained within Singapore and vector density has been low for several decades. We also do not account for the early phase of pilots of Wolbachia *Aedes* suppression control strategy in Singapore with targeted releases from May 2020. This has demonstrated effectiveness in reducing mosquito populations and dengue cases, and will likely impact climate-dengue relationships and serotype dynamics in the future [36]. Finally, we conducted model evaluation with forecast skill metrics and ROC analysis. In an ideal implementation scenario, forecast outputs such as outbreak warning alerts would be tied to specific control interventions or public health actions [28]. This would allow for the cost of a false alarm or missed event to be calculated and enable a cost-effectiveness analysis of the early warning system as a whole.

Our analysis disentangles the role of climate and serotype dynamics in driving dengue outbreaks over a 23-year period in Singapore, capitalising on a rich dataset of epidemiological, weather station and virus surveillance data. We translate these findings into an early warning framework able to forecast dengue cases and generate outbreak alert predictions with 0-8 weeks lead time. In this study, we integrate explanatory and predictive modelling approaches with a view that understanding the key causal relationships underlying transmission allows for the construction of forecasting models that are more generalisable [37]. This can also result in greater interpretability, allowing for clearer communication of forecasts to policymakers and a more intuitive understanding of how transmission may vary under future large-scale changes, such as climate change. Climate-driven early warning systems will become increasingly important adaptation measures as climate change alters the geographical range of dengue transmission and leads to

greater climatic extremes. We demonstrate the additional value of viral surveillance in improving forecast accuracy, and particularly in addressing the challenge of predicting dengue outbreak years. Future extensions of this work could include comparing this approach to other forecasting methods, including the current forecasting model in Singapore as well as alternative statistical and mechanistic approaches more broadly.

5.5 Methods

5.5.1 Data

In Singapore, dengue case reporting by clinicians and laboratories is mandatory. Weekly laboratory-confirmed cases from 1 January 2000 – 31 December 2022 were provided by the Ministry of Health, Singapore [14]. Laboratory confirmation is performed through antigen detection of nonstructural protein 1 (NS1) or detection of viral RNA by polymerase chain reaction (PCR) in the first five days of illness, or serological detection of immunoglobulin M (IgM) after five days of illness [7]. Since 2006, the National Environment Agency’s (NEA) Environmental Health Institute has serotyped a subset of dengue samples using RT-PCR as part of a virus surveillance programme [13]. In our dataset, for years where serotype information is available, ~27% of reported dengue cases are serotyped. Weekly DENV1-4 frequencies and total number of serotyped samples were provided by the Ministry of Health [14]. We calculated a smoothed proportion for DENV1-4 for each week with a GAM multinomial logistic regression using *mgcv* 1.8.36 [38]. To smooth the serotype data, we only used data up to and including the week of interest to enable resulting models to be useful in a forward-looking early warning framework. We then use these smoothed DENV proportions to classify the dominant serotype for each week. We defined a switch event as occurring where the current dominant serotype is different from the dominant serotype in the previous week, identifying 4 switch events in our dataset. We then defined the time since a switch in dominant serotype as the number of weeks since the most recent switch event.

Midyear population size estimates were obtained from the Singapore Department of Statistics, which included both local and foreigner populations.

Weekly maximum, minimum and mean temperature ($^{\circ}\text{C}$), absolute (g/m^3) and relative humidity (%), and precipitation (mm) were provided by the NEA. Daily precipitation (mm) was also provided and used to calculate: the number of days without rain per week (calculated as $\sum P_d = 0$ where P_d represents rainfall on a given day); number of days with heavy rain per week (calculated as $\sum P_d \geq 40$); number of days with moderate to heavy rain per week (calculated as $\sum P_d \geq 20$); and number of days with consecutive rainfall (i.e. a count of consecutive days where $P_d \geq 1$). Thresholds were chosen based on exploratory analysis of daily rainfall data and broadly align with 90th and 97.5th percentiles. Weekly Niño 3.4 SSTA was obtained from the National Oceanic and Atmospheric Administration (NOAA) [39]. This index measures the El Niño Southern Oscillation (ENSO); interannual fluctuations in the oceanic and atmospheric temperature around the Pacific Ocean. The index is commonly used to define El Niño events (unusually warm) and La Niña events (unusually cool). Sea surface temperature anomalies are calculated by subtracting the observed sea surface temperatures from a historical mean, calculated for the period 1981 - 2010. El Niño and La Niña events are typically defined by a sea surface temperature anomaly of $\pm 0.5^{\circ}\text{C}$ for over 6 months.

5.5.2 Model framework

We used a Bayesian hierarchical mixed-effects model to produce probabilistic predictions of weekly dengue incidence. Inference was performed using integrated nested Laplace approximation in INLA 23.04.24 [40]. Weekly dengue counts (y_t) were assumed to follow a negative binomial distribution to account for overdispersion in the data, with a mean μ_t and overdispersion parameter κ .

$$y_t \sim \text{NegBin}(\mu_t, \kappa)$$

$$\log(\mu_t) = \log(P_{a[t]}) + \alpha + \sum \beta_k X_{k,t} + \gamma_{a[t]} + \delta_{w[t]}$$

Here $\log(\mu_t)$ is the linear predictor, where $\log(P_{a[t]})$ is a population offset with population per 100,000 by year a and α is the model intercept. $\sum \beta_k X_{k,t}$ are k climate and/or serotype covariates, $\gamma_{a[t]}$ is a yearly random effect and $\delta_{w[t]}$ is a weekly random effect. We use the weekly random effect to account for seasonality and seasonal autocorrelation, while the yearly random effect accounts for any unexplained interannual variation in the data, for example as a result of vector control efforts or COVID-19 restrictions [32]. Details on model prior specifications and hyperparameters are available in the Supplementary Materials.

5.5.3 Model selection

We calculated Pearson's rank correlation index to assess correlation between variables using *corrplot* 0.92 and considered $r \geq 0.5$ as indicative of high correlation. We also calculated the variance inflation factor to assess multicollinearity. This measures how easily a model covariate is predicted from a linear regression of other covariates and we considered $VIF \geq 5$ as evidence of high collinearity. Based on this, we excluded relative humidity from further analysis due to high correlation with temperature variables and rainfall.

We tested all remaining variables with a 0, 4, 8, 12 and 16 week lag. For temperature, humidity and Niño variables we tested 1, 4, 8 and 12 week running averages, while for precipitation variables we tested 1, 4, 8 and 12 week running totals. Finally, we also tested non-linear formulations of temperature, precipitation and Niño 3.4 variables. We explored the best combinations of different classes of covariate (temperature, precipitation, humidity and Niño), conducting model selection in a forward stepwise manner, comparing models of increasing complexity (Supplementary information, Table 2). Covariate models were compared to a baseline model including only weekly and yearly random effects ($\gamma_{a[t]} + \delta_{w[t]}$). Overall 505 model formulations of climatic and serotype covariates were tested. We used various model adequacy criteria including: the widely applicable information criteria (WAIC) and deviance information criteria (DIC). WAIC and DIC are metrics which aim to maximise model fit while also penalising model complexity, with lower scores indicating a more parsimonious model.

Once we had selected the best performing climate model, we tested the inclusion of serotype variables including: dengue serotype proportions (individually and in combination, numeric variables), dengue serotype growth rates (individually and in combination, numeric variables), yearly or weekly dominant serotype (factor variable with four levels), serotype switch event (binary variable), time since serotype switch (numeric).

We quantified how much model covariates were able to explain interannual variation in dengue incidence by comparing the mean absolute value of the yearly random effects $\gamma_{a[t]}$ between covariate models and the baseline model. We calculated the proportion of interannual variation explained by the model as:

$$I = \frac{G_{model}}{G_{baseline}}$$

where $G = \frac{\sum |\gamma_{a[t]}|}{A}$ and A is the total number of years in the dataset.

5.5.4 Model evaluation

We evaluated the performance of four candidate models: a *climate and serotype* model; a *climate only* model; a *serotype only* model; a *seasonal baseline* model. We conducted model evaluation using time series cross-validation methodology to produce posterior predictions, and comparing observed and predicted outcomes to evaluate predictive performance. This is an appropriate cross-validation design when conducting statistical validation of forecasting models, as it preserves the time order of the underlying data, i.e. only data prior to an observation occurring is used to generate the prediction. To do this, we refit the model for each week in the dataset from 2009-2023 using an expanding window approach [31]. Data for the first eight years was used solely for training. Then, for each target week t in the dataset, we trained the model on data until week $t - 1$ and then simulated a posterior predictive distribution for dengue cases in week t , using climatic data up until time t . Serotype covariates are constructed only using data until $t - 1$ for each time point t as serotype frequencies are linked to dengue case counts. The posterior predictive distribution was simulated using 1000 samples from the posterior distribution of model parameters and hyperparameters. Note that all our final models (except the *seasonal baseline*) included a yearly

random effect, $\gamma_{a[t]}$, which is estimated using only using data available until week t for the year of the target week.

We also calculated posterior predictive probabilities of exceeding the outbreak threshold. This both evaluates the model’s ability to distinguish between outbreak and non-outbreak periods, and provides an operationally useful model output for decision makers. For instance, it may be easier to tie specific response measures or public health decisions to a probability of exceeding an operationally meaningful threshold than to the full probabilistic forecast. Here, we calculated outbreak thresholds for each week based on the seasonal moving 75th percentile of cases (Supplementary information, Figure 1). For each week t , we then calculated the proportion of posterior predictive samples that were greater than the threshold value.

Forecasts were scored using the *scoringutils 1.2.1* package [41]. We aimed to evaluate both how successful models were at predicting dengue case incidence and detecting outbreak weeks. Dengue case forecasts were scored using the continuous ranked probability score (CRPS); this is a proper scoring rule which can be considered a generalisation of mean absolute error that takes into account the entire predictive distribution [42]. This is calculated such that:

$$CRPS(F, y) = \int_{-\infty}^{\infty} (F(x) - 1(x \geq y))^2 dx$$

where y is the observed value and F is the cumulative density function (CDF) of the predictive distribution. Smaller CRPS values indicate a better forecast, and the metric penalises both under- and over-prediction. Sharper forecasts (where predictions are concentrated in a narrower range) will also score better. We then calculated the continuous ranked probability skill score (CRPSS), which is calculated as $1 - \frac{CRPS_{model}}{CRPS_{baseline}}$ and measures the percentage improvement of the considered model over a baseline model. A value of 1 indicates that the model has perfect skill compared to the baseline, 0 indicates the model is equivalent to the baseline and a negative value indicates that the model is worse than the baseline.

We also assessed model calibration by calculating interval coverage at the 50% and 95% levels. Interval coverage measures the proportion of observed values falling in a given prediction interval range. For a given prediction interval, a perfectly calibrated model would have interval coverage

equal to the nominal prediction interval (that is, 95% of observations falling within the 95% prediction interval). We also calculated bias B , measuring a model's tendency to over or under-predict. This was calculated for a data point y_t such that:

$$B(P, y_t) = 1 - (P(y_t) + P(y_t - 1))$$

Where $P(y_t)$ is the predicted probability mass for all outcomes smaller or equal to y_t .

To score the models' ability to forecast future outbreaks we calculated the Brier score comparing the posterior predictive probability of exceeding the outbreak threshold with observed outbreak weeks. The Brier score is a proper scoring rule for binary outcomes where smaller values indicate better forecasts [43]. This is defined as:

$$Brier\ score = \frac{1}{N} \sum_{n=1}^N (f_n - y_n)^2$$

where f_n are the predicted probabilities that corresponding events y_n will be equal to one. Finally, we used receiver operating characteristic (ROC) analysis to determine the optimum threshold for issuing outbreak alerts, balancing hit rate with false alarm rate, using the *pROC package 1.18.4* [44,45]. Here, hit rate (or sensitivity) is defined as the proportion of events (outbreak weeks) that were correctly predicted. False alarm rate (or 1 - specificity) is defined as the proportion of weeks without an outbreak where an outbreak was predicted to occur. We generated ROC curves for each model, which show hit rate against false alarm rate at different outbreak alert thresholds and calculated the area under the ROC curve (AUC). The AUC is a measure of model performance in classifying outbreak and non-outbreak weeks, with higher values indicating a better classification [46]. We selected outbreak alert thresholds by choosing the point closest to the top left of the ROC plot (representing perfect sensitivity or specificity).

5.5.5 Adapting the model for early warning

To adapt our model framework for use in an early warning scenario, we produced and evaluated dengue case forecasts at 2, 4, 6 and 8 week ahead forecast horizons. To do this, we used the best

approximation of each covariate used in the final models available at the forecast horizon being considered. For instance, as in our final models we use a 12 week running average of maximum temperature ($^{\circ}\text{C}$), to produce forecasts, at a 4 week ahead time horizon we instead use an 8 week running average of maximum temperature ($^{\circ}\text{C}$), with a four week lag. As the rainfall covariate we include is a 12 week total of days without rain, to approximate this for a 4 week ahead time horizon we use an 8 week total of days without rain, scaled up by a factor of 1.5. Full details of the variables used for prediction at each time horizon are available in Supplementary information, Table 5.

When conducting model evaluation for week t and forecast horizon h , we used the same expanding window time series cross-validation approach described earlier. We trained the model using the final selected covariates on data available until week $t - h$. Unlike the model evaluation design described earlier, we then predicted dengue incidence in week t using lagged covariates available at time $t - h$. For example, considering only the temperature covariate for simplicity, when predicting dengue cases with a 4 week ahead time horizon, we fit the model up until week $t - 4$ using a 12 week running average temperature to estimate model parameters. Then, using these estimated model parameters, we predict dengue cases at week t by inputting 8 week running average temperature with a 4 week lag (alongside other lagged covariates). A schematic showing the cross-validation design for different forecast horizons is shown in Supplementary Figure 3. This approach allows us to preserve the key relationships between climate and serotype covariates, and dengue cases that we estimate in full model fitting and then use the best climate data available at different lead times to generate forecasts for early warning.

Code availability

All code and data used for this analysis are available at: <https://github.com/EmilieFinch/dengue-singapore>

References

1. Romanello M, Napoli C di, Green C, Kennard H, Lampard P, Scamman D, et al. The 2023 report of the Lancet Countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms. *The Lancet*. 2023 Dec 16;402(10419):2346–94.
2. Colón-González FJ, Sewe MO, Tompkins AM, Sjödin H, Casallas A, Rocklöv J, et al. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study. *The Lancet Planetary Health*. 2021 Jul 1;5(7):e404–14.
3. Simmons CP, Farrar JJ, Nguyen van VC, Wills B. Dengue. *N Engl J Med*. 2012 Apr 12;366(15):1423–32.
4. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013 Apr;496(7446):504–7.
5. Sabin AB. Research on Dengue during World War II. *The American Journal of Tropical Medicine and Hygiene*. 1952 Jan 1;1(1):30–50.
6. Halstead SB. Dengue. *The Lancet*. 2007 Nov 10;370(9599):1644–52.
7. Rajarethinam J, Ang LW, Ong J, Ycasas J, Hapuarachchi HC, Yap G, et al. Dengue in Singapore from 2004 to 2016: Cyclical Epidemic Patterns Dominated by Serotypes 1 and 2. *Am J Trop Med Hyg*. 2018 Jul;99(1):204–10.
8. NEA. Guidebook on Climate of Singapore [Internet]. 2009 [cited 2024 Jan 9]. Available from: <https://www.nea.gov.sg/docs/default-source/resource/guidebook-on-climate-of-singapore.pdf>
9. Ho SH, Lim JT, Ong J, Hapuarachchi HC, Sim S, Ng LC. Singapore’s 5 decades of dengue prevention and control—Implications for global dengue control. *PLOS Neglected Tropical Diseases*. 2023 Jun 22;17(6):e0011400.
10. Low SL, Lam S, Wong WY, Teo D, Ng LC, Tan LK. Dengue Seroprevalence of Healthy Adults in Singapore: Serosurvey among Blood Donors, 2009. *Am J Trop Med Hyg*. 2015 Jul 8;93(1):40–5.
11. Tan LK, Low SL, Sun H, Shi Y, Liu L, Lam S, et al. Force of Infection and True Infection Rate of Dengue in Singapore: Implications for Dengue Control and Management. *Am J*

- Epidemiol. 2019 Aug;188(8):1529–38.
12. Yap G, Li C, Mutalib A, Lai YL, Ng LC. High Rates of Inapparent Dengue in Older Adults in Singapore. *Am J Trop Med Hyg.* 2013 Jun 5;88(6):1065–9.
 13. Lee KS, Lai YL, Lo S, Barkham T, Aw P, Ooi PL, et al. Dengue Virus Surveillance for Early Warning, Singapore - Volume 16, Number 5—May 2010 - *Emerging Infectious Diseases journal* - CDC. [cited 2022 Oct 22]; Available from: https://wwwnc.cdc.gov/eid/article/16/5/09-1006_article
 14. National Environment Agency (NEA). Weekly Infectious Diseases Bulletin [Internet]. [cited 2024 Jan 15]. Available from: <https://www.moh.gov.sg/resources-statistics/infectious-disease-statistics/2023/weekly-infectious-diseases-bulletin>
 15. Meteorological Service Singapore. El Niño/La Niña Status [Internet]. Available from: <http://www.weather.gov.sg/climate-el-la/>
 16. Earnest A, Tan SB, Wilder-Smith A. Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections. *Epidemiol Infect.* 2012 Jul;140(7):1244–51.
 17. Johansson MA, Cummings DAT, Glass GE. Multiyear climate variability and dengue--El Niño southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. *PLoS Med.* 2009 Nov;6(11):e1000168.
 18. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. *Ecology Letters.* 2019;22(10):1690–708.
 19. van Panhuis WG, Choisy M, Xiong X, Chok NS, Akarasewi P, Iamsirithaworn S, et al. Region-wide synchrony and traveling waves of dengue across eight countries in Southeast Asia. *Proceedings of the National Academy of Sciences.* 2015 Oct 20;112(42):13069–74.
 20. Lowe R, Gasparini A, Meerbeeck CJV, Lippi CA, Mahon R, Trotman AR, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLOS Medicine.* 2018 Jul 17;15(7):e1002613.
 21. Lowe R, Lee SA, O'Reilly KM, Brady OJ, Bastos L, Carrasco-Escobar G, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary Health.* 2021 Apr 1;5(4):e209–19.
 22. Gibb R, Colón-González FJ, Lan PT, Huong PT, Nam VS, Duoc VT, et al. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in

- Vietnam. *Nat Commun.* 2023 Dec 11;14(1):8179.
23. Benedum CM, Seidahmed OME, Eltahir EAB, Markuzon N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. *PLOS Neglected Tropical Diseases.* 2018 Dec 6;12(12):e0006935.
 24. Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J. Forecast of Dengue Incidence Using Temperature and Rainfall. *PLOS Neglected Tropical Diseases.* 2012 Nov 29;6(11):e1908.
 25. Xu HY, Fu X, Lee LKH, Ma S, Goh KT, Wong J, et al. Statistical Modeling Reveals the Effect of Absolute Humidity on Dengue in Singapore. *PLOS Neglected Tropical Diseases.* 2014 May 1;8(5):e2805.
 26. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Medicine.* 2018 Aug 6;16(1):129.
 27. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, et al. Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environ Health Perspect.* 2016 Sep;124(9):1369–75.
 28. Finch E, Lotto Batista M, Alcayna T, Lee SA, Fletcher IK, Lowe R. Early Warning Systems for vector-borne diseases: engagement, methods and implementation. In: *Planetary Health. in press. (Ecology and Control of Vector-borne Diseases).*
 29. Hii YL, Rocklöv J, Wall S, Ng LC, Tang CS, Ng N. Optimal Lead Time for Dengue Forecast. *PLOS Neglected Tropical Diseases.* 2012 Oct 18;6(10):e1848.
 30. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *PNAS.* 2019 Nov 26;116(48):24268–74.
 31. Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice (3rd Edition)* [Internet]. OTexts; 2021 [cited 2022 Jan 28]. 380 p. Available from: OTexts.com/fpp3
 32. Lim JT, Chew LZ, Choo ELW, Dickens BSL, Ong J, Aik J, et al. Increased Dengue Transmissions in Singapore Attributable to SARS-CoV-2 Social Distancing Measures. *The Journal of Infectious Diseases.* 2021 Feb 1;223(3):399–402.
 33. Soo KM, Khalid B, Ching SM, Chee HY. Meta-Analysis of Dengue Severity during Infection by Different Dengue Virus Serotypes in Primary and Secondary Infections. *PLOS ONE.* 2016 May 23;11(5):e0154760.

34. Katzelnick LC, Gresh L, Halloran ME, Mercado JC, Kuan G, Gordon A, et al. Antibody-dependent enhancement of severe dengue disease in humans. *Science*. 2017 Nov 17;358(6365):929–32.
35. O'Connor O, Ou TP, Aubry F, Dabo S, Russet S, Girault D, et al. Potential role of vector-mediated natural selection in dengue virus genotype/lineage replacements in two epidemiologically contrasted settings. *Emerg Microbes Infect*. 2021 Dec;10(1):1346–57.
36. Lim JT, Bansal S, Chong CS, Dickens B, Ng Y, Deng L, et al. Efficacy of Wolbachia-mediated sterility to reduce the incidence of dengue: a synthetic control study in Singapore. *The Lancet Microbe* [Internet]. 2024 Feb 8 [cited 2024 Mar 5];0(0). Available from: [https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247\(23\)00397-X/fulltext](https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(23)00397-X/fulltext)
37. Hofman JM, Watts DJ, Athey S, Garip F, Griffiths TL, Kleinberg J, et al. Integrating explanation and prediction in computational social science. *Nature*. 2021 Jul;595(7866):181–8.
38. Wood SN. *Generalized Additive Models: An Introduction with R*, Second Edition. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2017. 496 p.
39. National Oceanic and Atmospheric administration (NOAA). Climate Prediction Center - Monitoring & Data Index [Internet]. [cited 2024 Jan 15]. Available from: <https://www.cpc.ncep.noaa.gov/data/>
40. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009;71(2):319–92.
41. Bosse NI, Gruson H, Cori A, van Leeuwen E, Funk S, Abbott S. Evaluating Forecasts with scoringutils in R [Internet]. arXiv; 2022 [cited 2024 Jan 17]. Available from: <http://arxiv.org/abs/2205.07090>
42. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* [Internet]. 2012 Jan 1 [cited 2021 Aug 2]; Available from: <https://www.tandfonline.com/doi/abs/10.1198/016214506000001437>
43. Brier GW. Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*. 1950 Jan 1;78:1.
44. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd Edition [Internet]. John Wiley & Sons, Ltd; 2011 [cited 2024 Jan 17]. Available from:

<https://www.wiley.com/en-us/Forecast+Verification%3A+A+Practitioner%27s+Guide+in+Atmospheric+Science%2C+2nd+Edition-p-9781119960003>

45. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011 Mar 17;12(1):77.
46. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006 Jun 1;27(8):861–74.

Chapter 6

Understanding the role of climate and epidemic drivers in spatiotemporal dengue outbreak dynamics in the Dominican Republic

Following on from Chapter 5, where I explored using serotype surveillance data as a proxy for population immunity to improve dengue forecasting in Singapore, this chapter focuses on leveraging cross-sectional serological data to improve dengue prediction in the Dominican Republic. Serocatalytic models can be used to estimate average annual force of infection from age-stratified seroprevalence data. In this study, I applied a serocatalytic model to samples from a serological survey undertaken in 2021 in the Dominican Republic to estimate average annual attack rate. This allowed me to estimate the proportion of dengue infections reported and a proxy for population immunity built up in a dengue season.

I then constructed a spatiotemporal Bayesian hierarchical model for 155 municipalities in the Dominican Republic, using weekly surveillance data and gridded climate data from 2013 - 2023. By incorporating a proxy for the build-up of immunity and lagged cases, weighted by the serial interval, I was able to better quantify the effect of climate on dengue transmissibility. I explored temperature, precipitation, humidity, El Niño and drought indicators, investigating different aggregations and lags. I then conducted model evaluation using different cross-validation schema to assess the model's predictive ability, and to quantify the influence of included model covariates

on temporal and spatial predictive ability. This analysis forms an important first step in the development of a climate-informed early warning system for dengue in the Dominican Republic.

The supplementary material for this chapter is included as Appendix E.



London School of Hygiene & Tropical Medicine
 Keppel Street, London WC1E 7HT
 T: +44 (0)20 7299 4646
 F: +44 (0)20 7299 4656
 www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1802674	Title	Miss
First Name(s)	Emilie		
Surname/Family Name	Finch		
Thesis Title	Modelling the role of immunity, climate and behaviour in viral outbreak dynamics and control		
Primary Supervisor	Adam Kucharski		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Lancet Planetary Health
Please list the paper's authors in the intended authorship order:	Emilie Finch, Daniela Luhrsen, Eric Nilles, Colleen Lau, Ronald Skewes-Ramm, Cecilia Then Paulino, Adam Kucharski, Rachel Lowe
Stage of publication	Not yet submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I worked on: conceptualisation, methodology, data curation, investigation, writing original draft and review and editing of submitted manuscript.
--	---

SECTION E

Student Signature	
Date	08/07/2024

Supervisor Signature	
Date	12/07/2024

6.1 Abstract

Background

Emerging vector-borne diseases, such as dengue, are an important health threat in the Caribbean, a region particularly vulnerable to climate change. Dengue is hyperendemic in the Dominican Republic, causing large, irregular outbreaks; however, the drivers of these are not well understood. Here, we quantify the effects of climatic and epidemic drivers of spatiotemporal dengue risk in the Dominican Republic from 2013 to 2023.

Methods

We developed a Bayesian space-time hierarchical model to quantify the effect of climatic variation on dengue transmission in the 155 municipalities of the Dominican Republic, using 10 years of surveillance data. By leveraging serological survey data from 2021, we account for epidemic drivers in our modelling framework, incorporating the build-up of immunity during a dengue season, as well as autocorrelation in cases.

Findings

We found evidence for increased risk of dengue at higher maximum temperatures and relative humidity, as well as in drought or El Niño conditions. While including climatic variation improved the predictive power of the model by 8.1% over a baseline including seasonal and year-specific spatial random effects, incorporating epidemic factors increased this to 12.7%. We found El Niño and drought indicators are influential predictors of temporal dengue dynamics, while weighted lagged cases, representing the force of infection, predict both spatial and temporal patterns.

Interpretation

We found seasonal and interannual climatic variation shape dengue risk in the Caribbean, with non-linear and delayed impacts. Additionally, incorporating proxies for seasonal immunity and the force of infection improved the ability of the model to predict unseen dengue observations. Early warning systems are an increasingly important climate change adaptation tool. In this study, we identify climate-dengue relationships that are able to predict space-time variations in dengue risk, in a first step towards an early warning system.

6.2 Introduction

Dengue is a rapidly expanding mosquito-borne disease, transmitted primarily by *Aedes aegypti* and *Aedes albopictus*. Climatic changes, alongside increasing travel, trade and urbanisation, have shifted the dynamics and geographic distribution of dengue with reported cases doubling every decade in the past 30 years, and around half of the global population estimated to be at risk [1–3]. These trends are projected to continue, with increasing suitability for dengue transmission driven by future global heating [4,5].

While two dengue vaccines have been licensed, they are currently only recommended for use after pre-vaccination screening or in high transmission settings due to potential risks in seronegative individuals [6]. As such, many countries remain focused on vector control to limit transmission and disease burden. Climate-informed forecasting models can be used to mitigate the impact of dengue outbreaks by giving advanced warning of when outbreaks are likely to occur and helping policy makers to decide when and where to target vector control efforts, as well as informing public communication strategies [7]. Early warning systems integrating forecasting models with appropriate response mechanisms are becoming an increasingly important tool for disease control as climate change leads to more frequent climatic extremes and shifts the dynamics and distribution of dengue, as well as other climate-sensitive infectious diseases.

Climate-informed forecasting models capitalise on inherent lags between climatic variation and dengue transmission to predict periods with increased risk. Climate can influence transmission through effects on the vector life-cycle, the dengue virus itself, or both. Temperature affects both mosquito survival and reproduction, as well as the viral extrinsic incubation period, and temperatures of 29°C are thought to be optimal for transmission [8,9]. While increased rainfall can create vector breeding sites, excessive rainfall can cause flushing events, washing larval habitats away entirely [10,11]. The effects of rainfall also depend on population water storage behaviour; for instance, in cases where droughts lead to an increase in water storage containers around the home that can act as larval habitats [12,13]. The role of humidity in dengue transmission is less well-studied, although there is a clear mechanistic link between water vapour content in the air and vector desiccation stress [14]. Experimental evidence suggests increased humidity

increases vector survival and egg production, while previous studies have found statistical associations between humidity and dengue transmission [15–17]. Finally, dengue risk can be affected by the El Niño Southern Oscillation (ENSO); interannual fluctuations in oceanic and atmospheric temperature around the Pacific Ocean. This can result in El Niño events (periods of above average Pacific sea surface temperatures) or La Niña events (periods of below average Pacific sea surface temperatures). This is an important driver of interannual climatic variability in the Caribbean, with El Niño events associated with warmer temperatures and drought while La Niña events are associated with increased intensity of the Atlantic hurricane season [18,19]. ENSO indicators can be particularly valuable in forecasting frameworks as they offer the potential to predict interannual variation in dengue risk, and periods when large outbreaks may be more likely [20,21].

Emerging vector-borne diseases are a key health threat facing Caribbean islands, with small island developing states particularly vulnerable to health impacts of climate change [22,23]. The Dominican Republic experiences endemic dengue transmission, with large outbreaks that have been increasing in frequency and magnitude over past decades and the highest case-fatality ratio in the Caribbean [24]. Previous work investigating correlation between climate indicators and dengue cases in the Dominican Republic between 2015-2019 found that relative humidity was most frequently correlated with dengue cases in the nine provinces studied [25]. Contrastingly, another study of the 2019 outbreak in eight provinces found that temperature and rainfall were better predictors of dengue cases than relative humidity at lags of 2-5 weeks [25,26].

In this work, we use a Bayesian hierarchical modelling framework to quantify the effect of climate on spatiotemporal dengue risk in the Dominican Republic for 155 municipalities between 2013 and 2023, accounting for the build-up of immunity in a dengue season as well as autocorrelation in weekly cases. This is an important first step in constructing a forecasting framework able to provide early-warning of dengue risk and inform targeted public health action.

6.3 Methods

6.3.1 Study setting

The Dominican Republic is an upper-middle income country in the Caribbean, sharing the island of Hispaniola with Haiti. It is classified as a small island developing state by the United Nations [27]. The country is divided into 32 provinces and 155 municipalities and has an estimated population of around 10.8 million [28]. While the Dominican Republic has a tropical climate, it is a geographically diverse country, with the highest peak in the Caribbean of 3,101 m and several climate zones [29].

Most of the country experiences a rainy season between May and October, although this occurs later in the year on the northern coast, typically between November and January (Supplementary Figure 2). The Dominican Republic is situated on the Atlantic hurricane belt, and experiences hurricanes and cyclones between August and October. The frequency of hurricanes and extreme weather events is affected by ENSO, with more frequent hurricanes and cyclones during La Niña conditions, which also bring cooler and wetter weather [30].

6.3.2 Surveillance data

Weekly dengue cases between 31st March 2013 and 26th March 2023 were provided by the Dirección de Epidemiología from the Sistema Nacional de Vigilancia Epidemiológica (SINAVE). Data was aggregated by municipality and week of reporting, using epidemiological weeks running from Sunday - Saturday.

6.3.3 Population data

We used municipality-level population estimates available from 2015-2020 and province-level population estimates available between 2000-2030, both from the Oficina Nacional de Estadística [28]. To obtain municipality-level estimates for other years in the dataset we extrapolated the

nearest value in time, scaled by province-level trends in population. For instance, to obtain municipality-level population estimates in 2021, we used estimates from 2020, scaled by the relative increase or decrease seen at the province level from 2020 to 2021.

6.3.4 Climate data

In this study, we obtained mean, minimum and maximum weekly temperature ($^{\circ}\text{C}$) and precipitation (mm) from the ERA5-Land reanalysis dataset [31]. We calculated relative humidity, absolute humidity and specific humidity from mean temperature, surface pressure and dew point temperature extracted from the ERA5-Land dataset. We then spatially aggregated climate data by municipality (using a geographic mean over administrative boundaries) and by epidemiological week running from Sunday - Saturday.

We also used the standard precipitation index (SPI), an indicator recommended by the World Meteorological Organization to characterise meteorological drought on a range of timescales [32,33]. This quantifies the distance in standard deviations between observed precipitation and long-term mean precipitation, modelled by a gamma distribution. This was calculated at a weekly level for each municipality using the *SPEI* package [34]. Negative SPI values indicate drier conditions than would be expected while positive values indicate abnormally wet conditions, a value of >2 indicates extremely wet conditions, while <-2 indicates extremely dry conditions. SPI can be calculated at different timescales, reflecting the impact of drought on different water resources. At shorter time-scales, the index reflects soil moisture conditions while at longer timescales the index reflects groundwater, streamflow and reservoir storage conditions [33]. We calculated the SPI at 1, 3, 6 and 12 month timescales.

We obtained the Niño 3.4 sea surface temperature anomaly (SSTA) index from the National Oceanic and Atmospheric Administration, NOAA [35]. This measures sea surface temperature anomalies in the Niño 3.4 region (an area across the Pacific from the dateline to the South American coast) and is one of the indices commonly used to define El Niño (warm) and La Niña (cool) events. Anomalies are calculated by subtracting observed sea surface temperature from the

historical mean (1981-2010). El Niño and La Niña events are typically defined when the 5-month running average Niño 3.4 SSTs exceed +/- 0.5°C for a period of 6 months or more [20].

6.3.5 Seroprevalence study 2021

We used age-stratified seroprevalence data from a subset of samples collected during a cross-sectional serological survey in the Dominican Republic between June and October 2021 [36]. The full serological survey enrolled 6683 participants from 3832 households. In brief, this survey employed a multistage sampling method to select 134 clusters out of 12,565 communities. First, clusters were assigned to provinces, taking into account population, urban-rural divide, and the spatial dispersal of clusters. Then, clusters were selected within each province using grid methodology to maximise the spatial distribution of clusters and, finally, households were selected using grid methods applied to satellite images. Individuals were eligible for the study if they were a household member ≥ 5 years old.

We used data from a subset of 200 samples from Espaillat and San Pedro de Macorís provinces, tested for DENV IgG positivity using an in-house DENV1-4 ELISA assay.

6.3.6 Serocatalytic model

Serocatalytic models can be used to estimate the force of infection from age-stratified seroprevalence data [37,38]. The force of infection λ is defined as the rate at which susceptible individuals become infected. In a simple catalytic model, we assume a constant force of infection independent of time or age, and that individuals remain IgG positive after seroconversion. In this case, the probability of being infected by age a is given by

$$z(a) = 1 - e^{-\lambda a}$$

For a reverse catalytic model we continue to assume a time-constant force of infection but allow IgG positivity to wane over time with waning rate ρ [37]. Here:

$$\frac{dz(a)}{da} = \lambda - (\lambda + \rho)z(a)$$

By integrating the formula above we find that the probability of being infected by age a follows:

$$z(a) = \frac{\lambda}{\lambda + \rho}(1 - e^{-a(\lambda + \rho)})$$

As we assume infections follow a Poisson process, and therefore that the time until next infection follows an exponential distribution, the probability that a susceptible individual becomes infected in a year (or the annual attack rate) follows the cumulative distribution function of the exponential distribution such that:

$$\text{Attack rate} = 1 - e^{-\lambda}$$

We fit models using the RSero package, with parameter estimation implemented in Stan using Hamiltonian MCMC with a No-U-Turn sampler (NUTS) algorithm [39]. We assessed convergence using trace plots (Supplementary Figure 4) and the Gelman-Rubin \hat{R} statistic, ensuring $\hat{R} < 1.1$ and the effective sample size is >200 . We selected the best fitting model using the widely applicable information criterion, WAIC, and deviance information criterion, DIC [40,41].

6.3.7 Bayesian hierarchical model framework

Statistical approaches for infectious disease forecasting often use generalised linear modelling frameworks where disease counts y are assumed to follow a Poisson or Negative Binomial distribution, with a matrix of space-time varying covariates explaining variation in the outcome. For climate-sensitive infectious diseases, such as dengue, these often use climatic, environmental and socio-economic factors to explain variation in disease incidence. However, there are several features of the dynamics of infectious disease outbreaks that are not fully captured by these

frameworks [42]. In particular, the build-up of immunity (and depletion of the population susceptible to infection) and the inherent autocorrelation in the epidemic process (whereby cases at time t are dependent on past cases via the force of infection). Previous studies have accounted for the build-up of immunity using cumulative counts of cases in the current or previous seasons, or by using known proxies of population immunity [43]. Alternatively, generalised linear mixed modelling frameworks can use random effects to account for unexplained spatiotemporal variation caused by changes in population immunity, as well as other factors [12,44]. Similarly, autocorrelation between current and previous infected individuals has been accounted for by including cases at time $t - 1$ as a predictor, as in ARIMA models, or weighted cumulative previous cases for the duration of the serial interval [45,46].

It is also possible to derive a regression model framework from a semi-mechanistic renewal equation. Here we use the method by Cori and colleagues that links the instantaneous reproduction number with current and past incidence data [47]. Following an approach taken by Camacho and colleagues [42,43], we consider that dengue cases $y_{i,t}$ for week t and municipality i can be modelled using a renewal process such that:

$$y_{i,t} \sim Pois(R_{i,t} \sum_{s=1}^t y_{i,t-s} \omega_s)$$

Here, $R_{i,t}$ is the instantaneous reproduction number at week t and municipality i and ω is a discretised probability distribution of the serial interval by week s . $R_{i,t}$ can be expressed as a product of the basic instantaneous reproduction number $R_{0,i,t}$ and the proportion of the population susceptible individuals ($S_{i,t}$) in the population ($N_{i,t}$). When discretising $R_{i,t}$, we assume that $R_{i,t}$ is fixed at the value at the beginning of the time-step, and so is determined by the proportion of susceptible individuals $S_{i,t-1}$ in the population $N_{i,t}$. Similarly, we assume that the force of infection is determined by cases up to $t-1$, and that there is no probability mass on a serial interval of 0.

$$R_{i,t} = R_{0,i,t} \frac{S_{i,t-1}}{N_{i,a[t]}} \text{ and } S_{i,t-1} = N_{i,a[t]} - \sum_{s=t_0}^{t-1} y_{i,s}$$

We make the simplifying assumption that immunity lasts for the duration of the dengue season, with t_0 set to the beginning of the respective dengue season at time t . We consider dengue seasons running from epiweek 14 (around the beginning of April) until epiweek 13 of the subsequent year, to align with the seasonality of dengue in the Dominican Republic.

By substituting this expression for $R_{i,t}$ into the original renewal equation, and taking the logarithm of the expectation we find:

$$\log(E[y_{i,t}]) = \log(R_{0,i,t}) + \log\left(\sum_{s=1}^t y_{i,t-s}\omega_s\right) + \log\left(1 - \frac{\sum_{s=t_0}^{t-1} y_{i,s}}{N_{i,a[t]}}\right)$$

Following previous studies [43], this allows us to consider $\log(E[I_{i,t}])$ to be composed of three components: a transmissibility term $\log(R_{0,i,t})$, a term representing the force of infection

$$\log\left(\sum_{s=0}^t y_{i,t-s}\omega_s\right) \quad \text{and a term representing decreasing susceptibility from the accumulation of immunity } \log\left(1 - \frac{\sum_{s=0}^{t-1} y_{t,s}}{N_{i,t}}\right).$$

We use a discretised serial interval for dengue estimated by Siraj and colleagues, weighting municipality-level lagged cases with a 1-5 week lag with weights 0, 0.2, 0.425, 0.25, 0.125 [48]. Reported dengue cases $y_{i,t}$ are likely to be substantially underreported (as many cases are asymptomatic or present with mild flu-like symptoms). As such, cumulative reported cases divided by the total population is likely to underestimate the accumulation of immunity. To correct for this, we estimate an underreporting factor j . We first estimate the number of infections per year by comparing the estimated average annual attack rate from our previous catalytic model with the estimated population susceptible in 2021 according to the 2021 seroprevalence study. We then compare this average number of yearly infections with the average number of yearly reported cases in our dataset, estimating that $\sim 7.5\%$ of infections are reported. We then scale up reported cases such that the depletion of susceptibility is given by:

$$\log\left(1 - \frac{\sum_{s=t_0}^{t-1} y_{i,s} \times j}{N_{i,a[t]}}\right)$$

Finally, we model the effect of climate on the transmissibility term by:

$$\log(R_{0,i,t}) = \alpha + \beta_i X_{i,t,k} + \delta_{w[t]} + u_{i,a[t]} + v_{i,a[t]}$$

Where α is the model intercept, $\beta_i X_{i,t,k}$ is a vector of k climate covariates $\delta_{w[t]}$ is a weekly random effect accounting for seasonality and $u_{i,a[t]} + v_{i,a[t]}$ are year-specific spatially structured and unstructured random effects.

We fit this model in a Bayesian hierarchical model, implemented in R-INLA. We use a negative binomial parameterisation to account for overdispersion in the data.

Model priors and hyperparameters

The weekly random effect $\delta_{w[t]}$ was assigned a random walk 2 prior distribution (second order difference prior distribution) where we assume that second order increments follow a Gaussian distribution with zero mean and precision τ . This was defined to be cyclic (so the first week of the year is dependent on the last two weeks of the previous year).

We also included year-specific spatial random effects using the *bym2* model [49]. This includes a structured component u_i assuming a conditional autoregressive model (CAR) following a neighbourhood matrix, as well as an unstructured component v_i of normal random effects. Here, the structured component takes into account that municipalities closer together are likely to experience similar dengue transmission while the unstructured component accounts for other spatial heterogeneities, for instance due to spatial variation in vector control. The *bym2* model is defined as:

$$\eta = \frac{1}{\sqrt{\tau}}(\sqrt{1 - \phi}v_i + \sqrt{\phi}u_i)$$

with a precision parameter τ and a mixing parameter ϕ , which controls the contribution to the variance of v and u . We used the default hyperparameters implemented in INLA with penalised complexity (PC) priors.

Non-linear covariates were modelled by setting a random walk 2 prior on the coefficients of the covariates. We used PC priors for precision τ , with hyperparameters $\sigma_0 = 0.5$ and $\alpha = 0.01$ for all non-linear covariates and the weekly random effect. The PC prior is defined on the standard deviation $\sigma = \tau^{-1/2}$ such that $P(\sigma > \sigma_0) = \alpha$. these penalize departure from $\sigma = 0$. PC priors follow the principle of parsimony, favouring a base model (where $\sigma = 0$) unless evidence is provided against it [50].

Model selection

In our analysis we grouped climatic variables into: temperature variables (including mean, minimum and maximum temperature); rainfall variables (including precipitation and SPI indicators); humidity variables (including absolute, relative and specific humidity), and the Niño 3.4 index. First, we assessed correlation between classes of climatic variables. We calculated Pearson's rank correlation index using *corrplot 0.92* and considering $r \geq 0.5$ as indicative of high correlation. We then calculated the variance inflation factor to assess multicollinearity with a threshold of $VIF \geq 5$ for evidence of high collinearity. We found that absolute humidity and specific humidity were highly correlated with temperature variables. Based on this, and their mechanistic relationship with temperature, we excluded these from further analysis.

We tested temperature, precipitation and humidity variables at 0, 2, 4, 6, 8, 12, 14 and 16 week lags, based on previous lagged climate-dengue relationships identified in the literature and exploratory analysis [9, 12]. For SPI indicators and Niño 3.4 we considered lags up until 20 weeks, in 2-weekly increments, as these capture longer-term climatic processes. We also explored different aggregations of climate variables, to obtain the optimal signal from available weekly data. For temperature, humidity and Niño 3.4 variables we considered 1, 4, 6, 8, 10 and 12 week running averages, while for precipitation we considered 1, 4, 6, 8, 10 and 12 week running totals. We tested 264 lagged climate indicators in total, with all climate indicators tested as non-linear covariates. We performed model selection using stepwise forward selection, comparing models of increasing complexity. Covariate models included susceptibility and force of infection terms as

detailed above, and were compared to a baseline model including only weekly and year-specific spatial random effects, $\delta_{w[t]} + u_{s,y[t]} + v_{s,y[t]}$.

We conducted model selection based on the widely applicable information criteria (WAIC) and deviance information criteria (DIC). These are metrics aiming to maximise model fit while penalising model complexity, with lower scores indicating a more parsimonious model [40,41]. We also considered the R^2 statistic in model selection, which provides a measure of goodness-of-fit and can be interpreted as the ability of the model to account for variation in dengue cases. At each step of model selection, the best performing variable was carried forwards and tested against all variables in remaining climate classes. Overall, 635 models were tested; model adequacy criteria for the best performing model at each step of the model selection process are shown in Supplementary Table 1.

Model evaluation

We evaluated model performance using two evaluation designs, allowing us to compare posterior predictions with observed outcomes to evaluate predictive performance. We conducted a time-series cross-validation using an expanding window approach [46,51]. This is an appropriate cross-validation design when evaluating the potential utility of models for forecasting, as it preserves the time order of the underlying data. Data from the first two dengue seasons was used solely as training data, then for each week from the start of the 2015/16 dengue season until the end of the 2022/2023 dengue season, we refit the model to generate posterior predictions for each municipality and week. For each target week t in the dataset, we trained the model on data until week $t-1$ and then simulated a posterior predictive distribution for dengue cases in week t .

We then conducted a k-fold block cross-validation with a spatial design [52]. Municipalities were randomly assigned to 5 folds. For each fold k , the model was fit to data for the remaining four folds and posterior predictions were generated for the entire time series of municipalities in fold k . To account for variation in fold-allocation we repeated this 10 times, aggregating samples from each repetition to generate posterior distributions at the municipality and province level. The cross-validation schemes used for both experimental designs are illustrated in Supplementary Figure 7.

We estimated the posterior predictive distribution using 1000 samples from the posterior distribution of model parameters and hyperparameters.

To investigate the influence of each model covariate in explaining dengue dynamics in the Dominican Republic, we repeated the cross-validation excluding each model covariate, following an approach by Gibb and colleagues [53]. We then compared the out-of-sample continuous ranked probability score (CRPS) to that of the full model. The continuous ranked probability score is a proper scoring rule which can be considered a generalisation of mean absolute error that takes into account the entire predictive distribution [54]. Smaller CRPS values indicate better prediction.

We scored model predictions over time (Supplementary Figure 8). Here we scored quantile predictions using the weighted interval score (WIS), an approximation to the CRPS for quantile forecasts, for computational tractability. The WIS is also a proper scoring rule, where smaller values indicate better performance, that converges to the CRPS as more quantiles are reported [55]. We also calculated model bias over time; this measures the tendency of the model to over or under predict. If the prediction is smaller than the observed value, the bias is calculated as the maximum percentile rank where the prediction is smaller than the observed value. Conversely, if the prediction is larger than the observed value the bias is the maximum percentile rank where the prediction is larger than the observed value. We also calculated the interval coverage to assess probabilistic calibration. We calculated coverage as the proportion p of observations that fall within the 50% and 95% prediction intervals. For a perfectly calibrated model coverage would equal the nominal prediction interval (i.e. 50% coverage for the 50% prediction interval and 95% coverage for the 95% prediction interval). Scoring was performed using the *scoringutils 1.2.1* package [56].

6.4 Results

6.4.1 Dengue epidemiology in the Dominican Republic between 2013 and 2023

Between 31st March 2013 (epiweek 14) and 26th March 2023 (epiweek 13) 85,903 dengue cases were reported in the Dominican Republic. Almost half of these were reported in the provinces containing the two major cities and their immediate surroundings (Santo Domingo, Distrito Nacional and Santiago). Contrastingly, dengue incidence rates were highest in the smaller provinces of San Jose de Ocoa and Hermanas Mirabal. Dengue transmission varies seasonally, with peak incidence between June and November, coinciding with the rainy season and high average temperatures (Supplementary Figures 1 and 2). The Dominican Republic has experienced large outbreaks of dengue, for example, in 2013, 2015, 2019 and 2022 (Figure 1, panel b and c). In 2019, the largest outbreak to date was reported with over 20,000 cases.

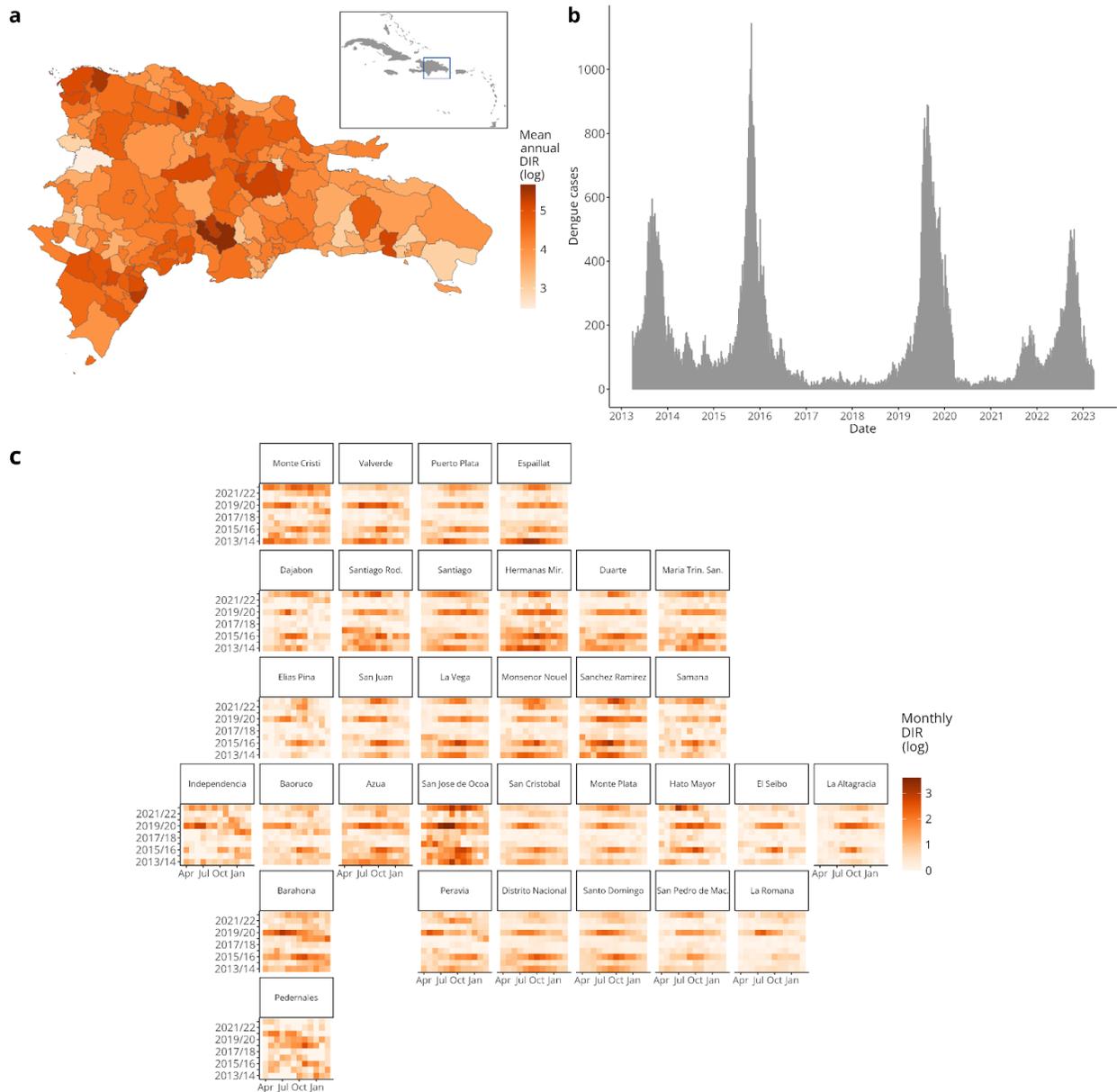


Figure 6.1: Dengue epidemiology in the Dominican Republic

(a) Map of the Dominican Republic showing mean annual dengue incidence rate (DIR) from 2013 - 2023 at the municipality level. The map insert shows a map of the Caribbean with the location of the Dominican Republic outlined in a rectangle. (b) Weekly reported dengue cases (grey bars) between 31st March 2013 and 26th March 2023. (c) Mean monthly dengue incidence per 100,000 population by province, with months running from April - March on the x-axis and dengue season on the y-axis. Provinces are arranged to mirror their geographic location in the Dominican Republic. Dengue incidence rates are log+1 transformed for visualisation.

6.4.2 Estimating the annual force of infection

We estimated the average annual force of infection by fitting serocatalytic models to age-stratified seroprevalence data. We considered a catalytic and reverse catalytic model structure (with the latter allowing for waning of seropositivity). As both fit similarly well to the data according to WAIC and DIC values, we used a simple catalytic model, following the principle of parsimony. We estimated an average annual force of infection of 0.129 (95% CrI: 0.103 - 0.162), corresponding to an average annual attack rate of 12.1% (95% CrI: 9.75% - 15.0%).

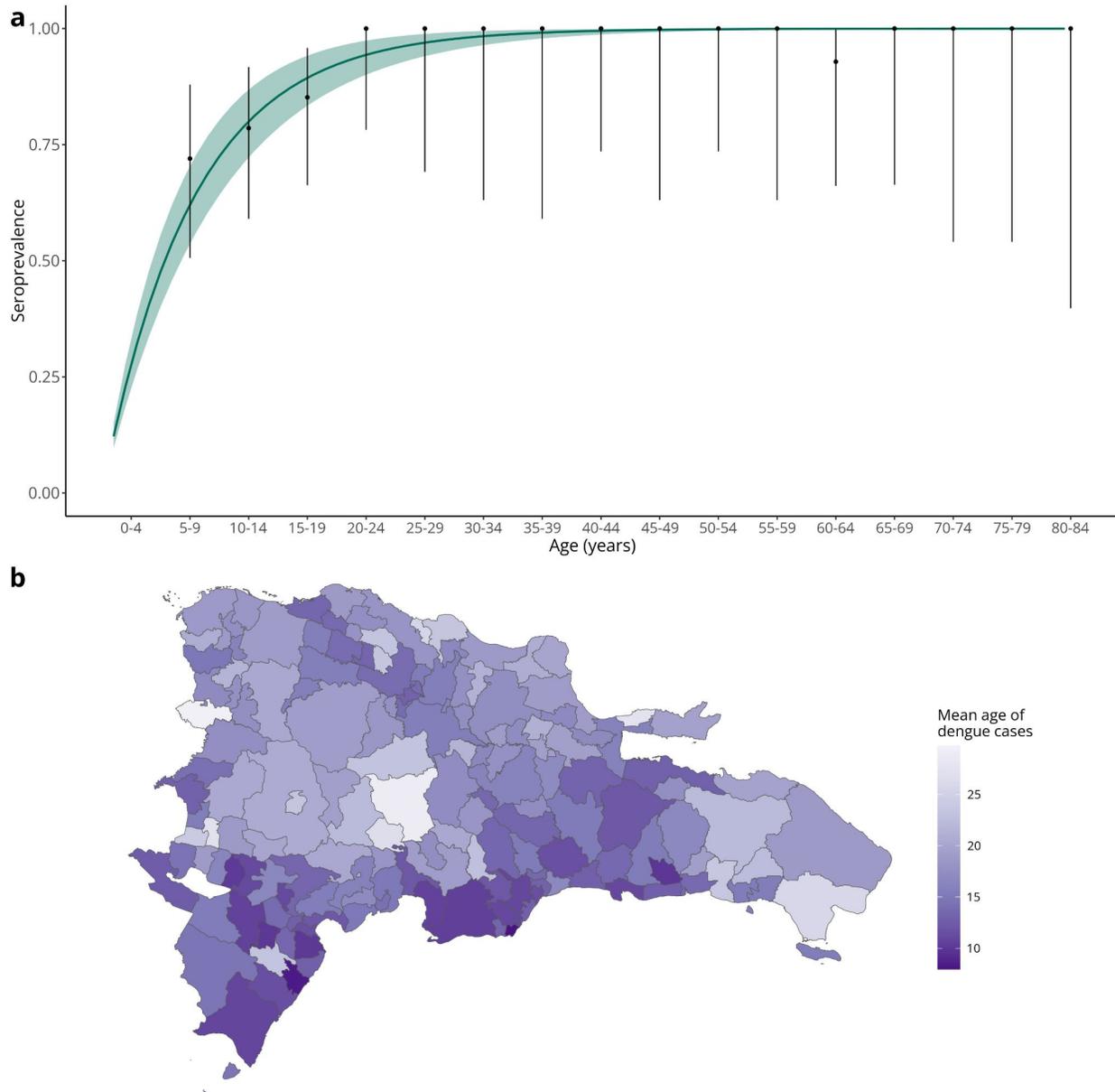


Figure 6.2: DENV seropositivity by age and mean age of dengue cases

Estimated seroprevalence by age from the serocatalytic model is shown in panel a (green line) with associated 95% model credible intervals (light green shaded area), alongside the proportion of individuals seropositive to a DENV1-4 ELISA assay by age (black points and line bars). The mean age of reported dengue cases reported to the Dominican Republic’s surveillance system (SINAVE) by municipality is shown in panel b, with earlier mean ages indicated by darker purple.

Once we obtained an estimated attack rate, we then estimated the proportion of infections reported by first approximating the number of infections a year, multiplying the attack rate by the proportion

susceptible. As ~9% of the population were IgG seronegative in our seroprevalence sample we considered the proportion susceptible to be 9% of the 2021 population (10.53 million) resulting in an estimated ~ 11,500 infections per year. We then divided the average reported yearly cases in our dataset (8,590) by the estimated infections a year to obtain an estimated reporting rate of 7.47%, assuming a constant reporting rate over time. For each dengue season, we estimated the population susceptible, as outlined in the Methods.

6.4.3 Quantifying the impact of climate variation on dengue risk

We fit a Bayesian spatiotemporal model to weekly surveillance data, using a negative binomial likelihood. We incorporated temporally correlated weekly random effects to account for seasonality and seasonal autocorrelation (*seasonal random effects*) and year-specific municipality level spatially structured and unstructured random effects to account for unexplained spatiotemporal variation in the data (*spatiotemporal random effects*).

We investigated whether the inclusion of temperature, rainfall, humidity and Niño 3.4 covariates improved model adequacy criteria when compared with a baseline model only including random effects (Supplementary Table 1). We tested climatic covariates with different aggregations, with non-linear effects specified using a second-order random walk (Methods). Given the inherent lag in the relationship between climate and dengue risk, we tested including climatic variables with lags from 0 weeks up until 16 weeks, or, for the Niño 3.4 and standard precipitation indices, until 20 weeks. Each model tested also included the municipality-level estimated proportion susceptible over each dengue season and a running average of cases with a 2-5 week lag, weighted by the serial interval distribution [48] (Methods). Following model selection, we found the best performing model included: Niño 3.4 sea surface temperature anomaly (SSTA), using a 12 week average with a 6 week lag; maximum temperature (°C), using an 8 week average with a 4 week lag; relative humidity (%) using a 10 week average with a 2 week lag; and the standard precipitation index (SPI) on a 12 month time scale (Supplementary Table 1). The full model also included lagged cases weighted by a discretised serial interval distribution (representing the force of infection) and an estimated proportion susceptible as described in the Methods.

To understand the relative role of climate in shaping dengue risk we estimated the marginal effects of each variable, these were exponentiated so the effects can be interpreted on the multiplicative response scale (Figure 3, panels a-d). We found that positive values of the lagged Niño 3.4 SSTA (indicative of El Niño conditions) were associated with increased dengue risk (Figure 3, panel a). The effect of maximum temperature on dengue risk was non-linear; with peak risk upwards of 28°C and risk plateauing as temperatures exceed 30°C (Figure 3, panel b). Similarly, we found increased dengue relative risk in humid conditions, with greatest risk upwards of 80% relative humidity (Figure 3, panel c). Drought conditions, indicated by negative values of the SPI, were associated with increased dengue risk, while abnormally wet conditions (indicated by positive values) were mostly protective (Figure 3, panel d).

We explored how model covariates were able to explain spatiotemporal variation in dengue incidence by investigating estimated *spatiotemporal random effects*. The *spatiotemporal random effects* indicate whether dengue incidence is higher or lower than expected for a given municipality and year than the overall mean incidence: therefore, if model covariates are able to account for spatiotemporal variation in dengue incidence we would expect the estimated *spatiotemporal random effects* to be closer to 0. We calculate the average municipality-level difference between the absolute value for *spatiotemporal random effects* in the final model and the baseline model, such that negative values indicate municipalities where included covariates are able account for spatiotemporal variation in dengue cases (Figure 4, panels e-f).

We found that, during outbreak years (2013, 2015, 2019 and 2022), included model covariates reduced unexplained variation in dengue incidence in most municipalities, and particularly in those in the East and South of the Dominican Republic (Figure 2, panel e). During non-outbreak years, model covariates again reduced unexplained variation in most municipalities, with effects concentrated in the West and south-East (Figure 2, panel f). We compared the contribution of climatic covariates and epidemic covariates in reducing unexplained spatiotemporal variation in the model in Supplementary Figure 6.

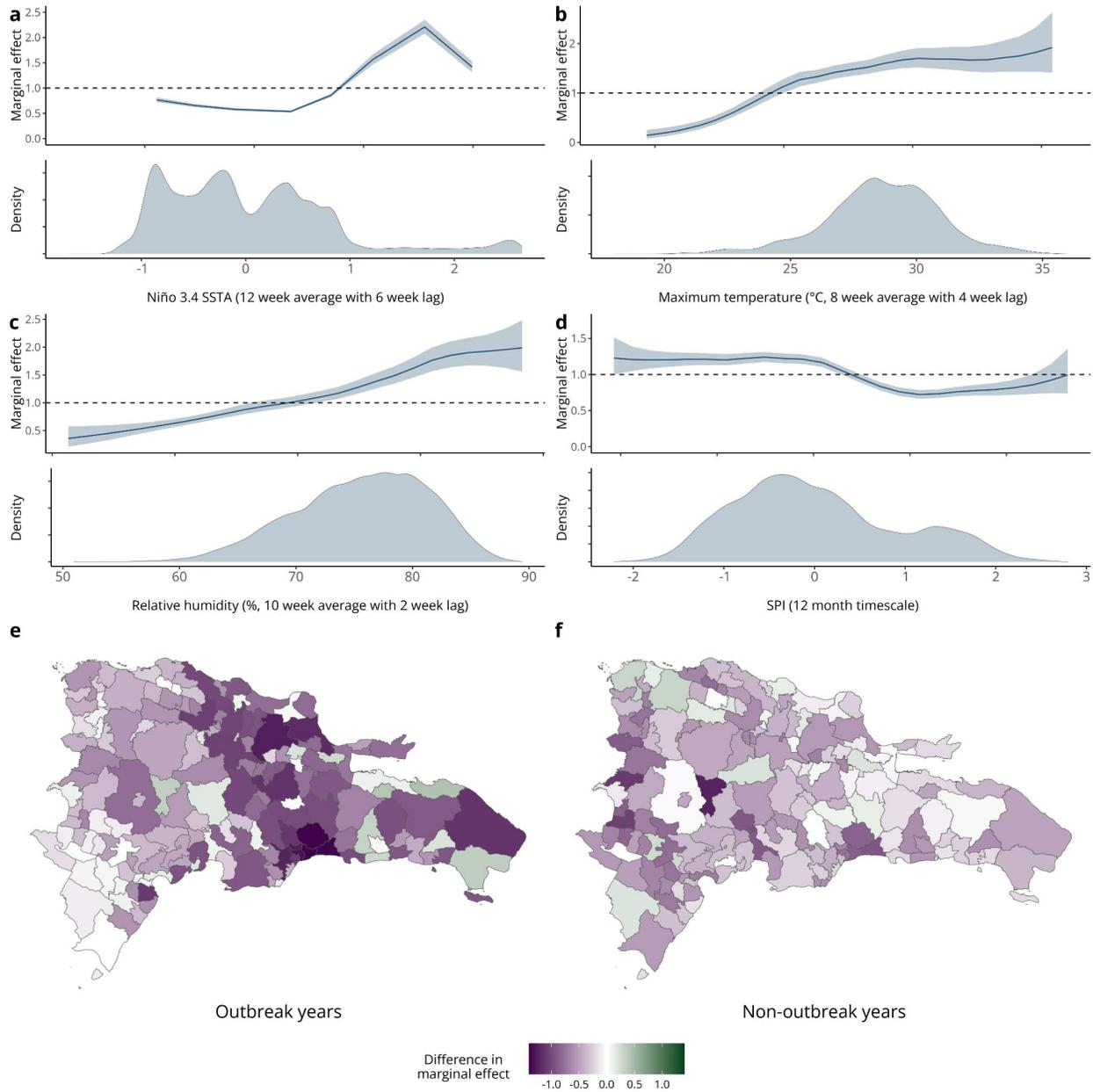


Figure 6.3: Effects of climate on dengue risk in the Dominican Republic

Posterior marginal effects and density plots for climatic covariates in the full model are shown in panels a-d. These include: Niño 3.4 sea surface temperature anomaly (SSTA) using a 12 week average with a 6 week lag, mean temperature in °C (8 week average with 4 week lag), relative humidity (% , 8 week average with a 4 week lag) and SPI (12 month timescale). These are shown on the outcome scale displaying the median value and associated 95% credible interval, a value of 1 suggests no association between the covariate and the outcome, while positive and negative values indicate the covariate is associated with increased and decreased risk respectively. The impact of including model covariates on unexplained spatiotemporal

variation is shown in panels d and e. These show the absolute difference in spatiotemporal random effects in the full covariate model compared with a baseline model including only *seasonal* and *spatiotemporal random effects*. Panels d and e show the difference in absolute average marginal spatial random effects during outbreak years and non-outbreak years respectively. This is calculated as:

$|u_{i,a[t]} + v_{i,a[t]}|_{Full} - |u_{i,a[t]} + v_{i,a[t]}|_{Baseline}$. Municipalities where the magnitude of the spatial random effects was smaller (i.e. closer to zero) after the inclusion of model covariates are shown in purple. The more negative the difference (and the darker the purple) the greater the contribution of climatic and epidemic model covariates in explaining variation in dengue risk. Municipalities where covariates increased the marginal random effects are shown in green and those with no change are shown in white.

6.4.4 Out-of-sample dengue predictions

We generated out-of-sample probabilistic dengue predictions using a time series cross-validation methodology (TSCV) and assessed model performance. Probabilistic predictions were generated at the municipality level and samples were aggregated to the province level for visualisation (Figure 4). The model was able to capture spatial patterns in dengue cases, ranging from high intensity of dengue cases in Distrito Nacional, Santiago and Santo Domingo to very low numbers of cases in provinces such as Baoruco and Pedernales. The model was also able to capture broad temporal outbreak dynamics. However, while the peak cases in the 2015 and 2022 outbreaks were successfully captured in most provinces, the peak of the 2019 outbreak was under-predicted in nearly all provinces (Supplementary Figure 8).

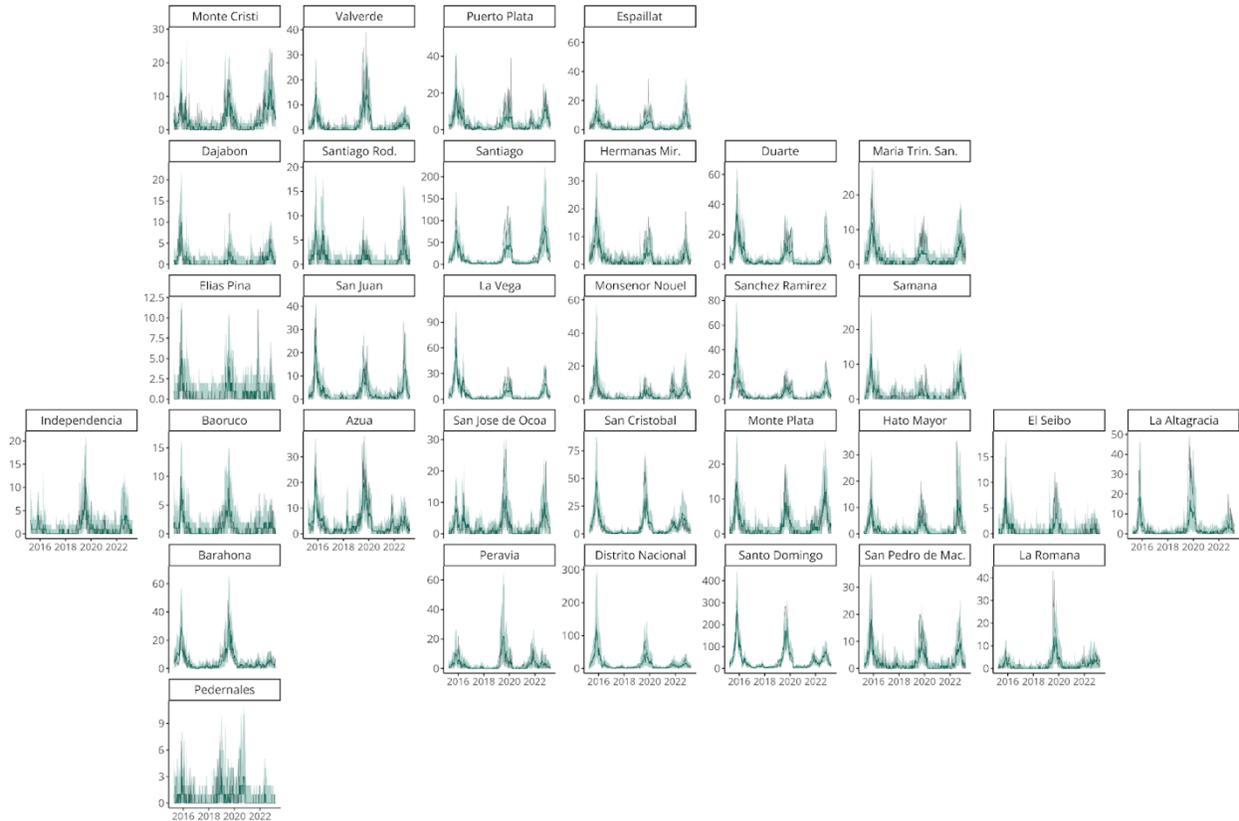


Figure 6.4: Time-series cross validated posterior predictions of dengue cases from 2015 – 2023

Figure showing time series cross-validated posterior predictions of weekly dengue cases for the full model from 2015 – 2023. We used an expanding window cross-validation methodology, where the model is trained on data up to, but not including, the target week and then posterior predictions are generated for the target week (Supplementary Figure 7). The first two years of data were used exclusively for training, and predictions are therefore shown for the following 8 dengue seasons from 2015 – 2023. Green lines represent median posterior predictions of weekly dengue cases, the shaded green area shows the associated 95% credible interval, and the grey lines show the data. The model was fit at the municipality level and posterior samples for each time step were then aggregated to the province level. Provinces are arranged according to their approximate geographic position in the Dominican Republic.

We also generated out-of-sample dengue predictions using a k-fold cross-validation design, shown in Supplementary Figure 9.

We assessed the influence of each covariate in predicting temporal and spatial dynamics by repeating the cross-validation experiments, excluding one model covariate from the full model at a time (Figure 5). We used the time series cross-validation design to understand a covariate’s

influence in predicting temporal dynamics, and a k-fold spatial holdout design to understand a covariate's influence in predicting spatial dynamics. Under a k-fold spatial holdout design, predictions of dengue cases are inferred from neighbouring municipalities through spatial random effects. The change in predictive ability when a covariate is excluded therefore reflects the extent to which that covariate explained differences in the expected similarities between neighbouring municipalities. Contrastingly with a TSCV design, predictions are inferred from *seasonal random effects* capturing seasonality in transmission, as well as the estimated year-specific *spatiotemporal random effect*. As such, changes in the predictive ability of the model when the covariate is excluded reflect how the covariate explains differences from expected seasonal and interannual variation up to the prediction week. The influence of a covariate on dengue predictions was measured as the percentage change in the continuous ranked probability score (CRPS) when the covariate was excluded (Figure 5).

For both spatial and temporal holdout designs the full model showed improved predictive accuracy over a baseline. Weighted lagged cases and the Niño 3.4 indicator were particularly influential covariates in predicting temporal dynamics, followed by a smaller role for SPI-12 and relative humidity. The influence of maximum temperature and susceptibility was small but positive. We found that including climatic covariates improved the predictive power of the model by 8.1%, over a baseline, while incorporating climatic and epidemic factors increased this to 12.7% (Supplementary Table 2). Weighted lagged cases, representing municipality-level force of infection, were the most influential covariate in predicting spatial dynamics. Climatic covariates and susceptibility appear less influential in driving spatial dynamics, with very small median increases in CRPS, and 95% credible intervals crossing zero (Supplementary Table 3).

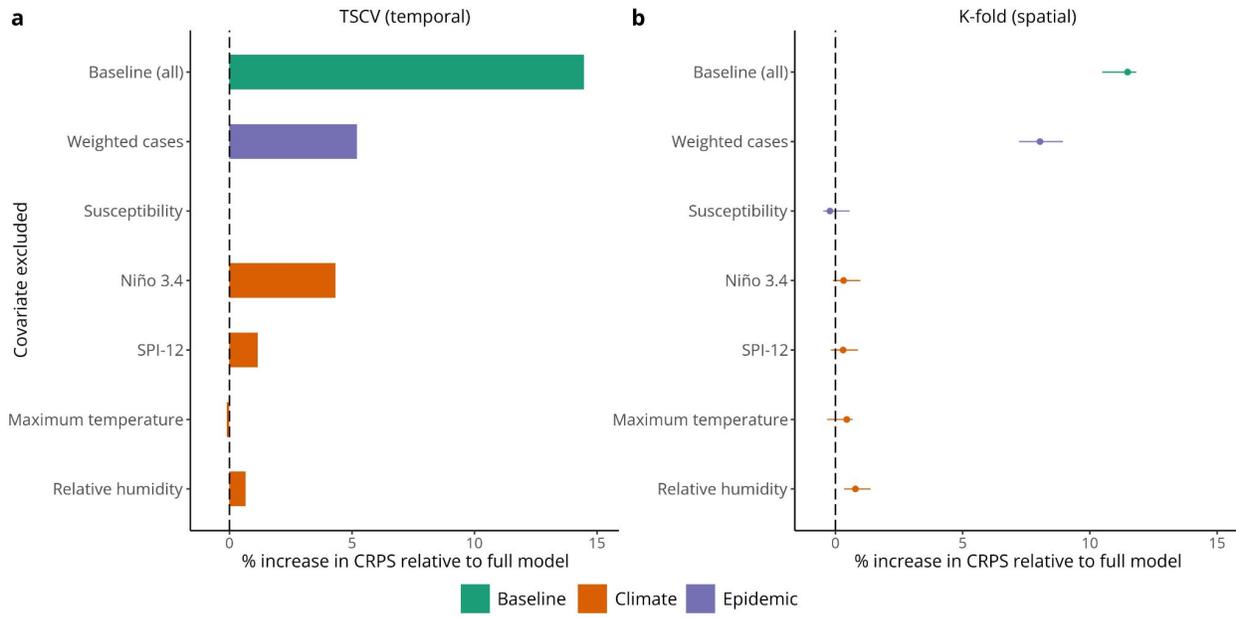


Figure 6.5: Influence of climatic and epidemic drivers on dengue predictions

Influence of model covariates on out-of-sample continuous ranked probability score (CRPS). This is estimated by calculating the percentage increase in CRPS score when each model covariate is excluded (note that smaller CRPS scores indicate better predictions). We evaluated the influence of model covariates using time-series cross-validation (panel a), which represents the importance of the covariate in explaining temporal dynamics, and 5-fold spatial cross-validation, representing the importance of the covariate in driving spatial patterns of dengue risk (panel b). To account for random variation in fold allocation we repeated the k-fold evaluation 10 times, and report a median % increase in CRPS (point) and the associated 95% credible interval (lines). The colours reflect the type of covariate considered with climate covariates shown in orange, epidemic covariates shown in purple and the baseline model (where all covariates are excluded) shown in green.

6.5 Discussion

To date, dengue early warning systems have typically been based on statistical models, which have been found to outperform mechanistic methods [57]. However, statistical regression methods often do not account for inherent epidemic drivers of infectious disease transmission, absent from other typical use-cases such as environmental epidemiology or economics [42]. In particular, statistical regression models typically do not account for changes in population immunity, influencing the

underlying population at risk of disease, which may limit their ability to predict the timing of an outbreak peak or, conversely, when a large outbreak may be more likely to occur.

We analyse 10 years of climate and dengue data in the Dominican Republic to quantify key drivers of dengue transmission; a first step in the development of an early warning system for dengue. We develop a spatiotemporal Bayesian framework, incorporating both climatic and epidemic drivers of transmission, which is able to generate probabilistic predictions of dengue cases at the municipality level with improved out-of-sample predictive performance over a baseline model. Our climate-driven statistical modelling framework was informed by catalytic modelling of age-stratified seroprevalence data. We estimated an average annual attack rate of 12.1% (95% CrI: 9.75% - 15.0%), suggesting that around 7.5% of dengue infections are reported and allowing us to approximate the proportion of the population remaining susceptible during each dengue season. We included both this and an average of lagged cases, weighted by the serial interval distribution. These terms represent the accumulation of immunity and the force of infection during each dengue season, capturing epidemic drivers and allowing us to quantify the impact of climate on dengue transmissibility (as measured by the reproduction number, R_0) directly. The best performing model also included the following climatic covariates: Niño 3.4 sea surface temperature anomaly (SSTA), using a 12 week average with a 6 week lag; maximum temperature ($^{\circ}\text{C}$), using an 8 week average with a 4 week lag; relative humidity (%) using a 10 week average with a 2 week lag; and the standard precipitation index (SPI) on a 12 month time scale (Supplementary Table 1).

We found non-linear and delayed effects of climate on dengue risk, as has been seen in other settings in the Caribbean, as well as Latin America more widely [12,13]. We found increased dengue risk during El Niño conditions, as well as with increasing temperatures and relative humidity. We also found evidence of increased dengue risk during drought conditions with reduced risk during periods of heavy rainfall. While this is in line with other studies, the impact of drought and extreme rainfall on dengue risk is known to be mediated by local water and sanitation infrastructure, as well as population water storage behaviour, and is likely to differ depending on the setting studied [58,53].

Our results are in line with a previous study on dengue in several provinces of the Dominican Republic, which found that lagged relative humidity and temperature were correlated with dengue cases, and that precipitation (measured in mm) was not a strong predictor [25]. However, unlike Robert and colleagues we found a positive relationship between relative humidity and dengue cases, which is likely a result of accounting for both temperature and relative humidity together in a multivariable model [25]. Our work also builds on previous research identifying useful predictors for dengue outbreaks in the Dominican Republic, which found that probable cases could be used to predict outbreaks of hospitalised cases but that meteorological data obtained for the study did not capture enough geographic variation to improve predictions [59]. Here, we were able to understand spatiotemporal impacts of climate on disease transmission beyond the areas represented by weather station data by leveraging fine-scale gridded climate reanalysis data [31].

We then conducted rigorous model evaluation to test the ability of our inferred climatic and epidemic drivers to predict unseen observations [53,60]. Using a time-series cross-validation approach to assess the potential utility of a model framework for forecasting, we found that the model was able to capture broad spatiotemporal outbreak dynamics in the Dominican Republic. However, the model underpredicted the 2019 peak in cases in almost all provinces. This is likely due to our simplification of dengue immune dynamics in the Dominican Republic, and particularly our inability to incorporate the impact of switches in dominant serotype on population susceptibility. While there is limited historical serotype surveillance in the Dominican Republic, DENV3 was reported in 2019 for the first time since 2015, suggesting there may have been increased population susceptibility for circulating serotypes or genotypes at this time [24].

We found the most influential drivers of temporal dengue risk were weighted lagged cases, representing the force of infection, as well as the Niño 3.4 index and SPI-12, representing interannual climatic drivers. Contrastingly, we found that only weighted lagged cases were the most influential driver of spatial risk, with less evidence for an influence of included climatic covariates on spatially structured dynamics (that is, how different transmission is from neighbouring municipalities than what would be expected). Despite this, we did find that included epidemic and climatic model covariates reduced the unexplained spatiotemporal variation in dengue cases, suggesting they are able to account for non-spatially structured variation in

transmission. The inclusion of socioeconomic variables could improve the spatial predictive ability of the model but would possibly have less impact in a forecasting application as these factors are unlikely to vary significantly over time. While we lacked the data needed to parameterise a fully mechanistic model in this setting, we compared our mechanistically informed statistical framework with a more typical *climate-only* model. We found that while including climatic variation improved the predictive power of the model by 8.1% over a baseline including seasonal and year-specific spatial random effects, incorporating epidemic drivers increased this to 12.7% (Supplementary Table 3).

There are several limitations to this work. We also did not have access to data on vector control efforts, which may have changed in intensity, particularly following the 2015-16 Zika outbreak, or on vector density, which both mediate the relationship between climate and dengue risk. Additionally, we only had access to seroprevalence data from two provinces in the Dominican Republic and for individuals ≥ 5 years old. Due to high seropositivity in this sample, we were unable to investigate temporal trends in the force of infection, which would have allowed a more precise estimation of reporting rate and population susceptibility over time. In this study, we use DENV1-4 IgG ELISA testing data, which can be affected by cross-reactivity with other flaviviruses. This would inflate our estimates of DENV seropositivity. To address this, we compared ELISA assay data with DENV1 and DENV2 neutralisation test data using fluorescent reported virus and live imaging, finding similar patterns of age-stratified seropositivity (Supplementary Figure 10). We assumed a constant force of infection (reflecting endemic dengue transmission in the country), estimating the average FOI over circulating serotypes. As we lacked the serotype surveillance data necessary to understand serotype dynamics over time, we also calculated susceptibility for each dengue season, making the simplifying assumption that population immunity was seasonal. We therefore did not incorporate long-term homotypic and short-term heterotypic immunity of dengue, which likely limited our ability to predict large outbreak peaks such as in 2019 [61,62]. Future extensions of this work could include evaluating dengue forecasts generated by this modelling framework at appropriate lead times for public health response, either through the use of identified lagged climatic covariates or climate forecasts. Additionally, further work could consider alternative parametric assumptions for the discretised serial interval distribution, which may improve predictive ability [45].

In this work, we leverage serological data to account for important aspects of the dengue transmission process in our modelling framework, enabling us to quantify the impact of climate on transmissibility. We used rigorous cross-validation methodology to estimate out-of-sample predictive error to test whether the climate-dengue relationships identified enable prediction of unseen observations; an important first step in the development of a dengue early warning system. Early warning systems able to accurately forecast dengue dynamics will be increasingly important in the Dominican Republic, as climate change impacts future outbreak risk.

References

1. Simmons CP, Farrar JJ, van Vinh Chau N, Wills B. Dengue [Internet]. <http://dx.doi.org/10.1056/NEJMra1110265>. Massachusetts Medical Society; 2012 [cited 2021 Jul 15]. Available from: <https://www.nejm.org/doi/10.1056/NEJMra1110265>
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013 Apr;496(7446):504–7.
3. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infect Dis*. 2016 Jun 1;16(6):712–23.
4. Romanello M, Napoli C di, Green C, Kennard H, Lampard P, Scamman D, et al. The 2023 report of the Lancet Countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms. *The Lancet*. 2023 Dec 16;402(10419):2346–94.
5. Colón-González FJ, Sewe MO, Tompkins AM, Sjödin H, Casallas A, Rocklöv J, et al. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study. *Lancet Planet Health*. 2021 Jul 1;5(7):e404–14.
6. World Health Organization. WHO position paper on dengue vaccines, May 2024 [Internet]. [cited 2024 May 27]. Available from: <https://www.who.int/publications-detail-redirect/who-wer-9918-203-224>
7. Finch E, Batista ML, Alcayna T, Lee SA, Fletcher IK, Lowe R. Early warning systems for vector-borne diseases: engagement, methods and implementation. In: *Planetary health approaches to understand and control vector-borne diseases* [Internet]. Wageningen Academic; 2023 [cited 2024 Jan 22]. p. 347–86. Available from: <https://brill.com/edcollchap-oo/book/9789004688650/BP000023.xml>
8. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA, et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLoS Negl Trop Dis*. 2017 Apr 27;11(4):e0005568.
9. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. *Ecol Lett*. 2019;22(10):1690–708.

10. Benedum CM, Seidahmed OME, Eltahir EAB, Markuzon N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. *PLoS Negl Trop Dis*. 2018 Dec 6;12(12):e0006935.
11. Koenraadt CJM, Harrington LC. Flushing effect of rain on container-inhabiting mosquitoes *Aedes aegypti* and *Culex pipiens* (Diptera: Culicidae). *J Med Entomol*. 2008 Jan;45(1):28–35.
12. Lowe R, Gasparrini A, Meerbeeck CJV, Lippi CA, Mahon R, Trotman AR, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLOS Med*. 2018 Jul 17;15(7):e1002613.
13. Lowe R, Lee SA, O'Reilly KM, Brady OJ, Bastos L, Carrasco-Escobar G, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *Lancet Planet Health*. 2021 Apr 1;5(4):e209–19.
14. Brown JJ, Pascual M, Wimberly MC, Johnson LR, Murdock CC. Humidity – The overlooked variable in the thermal biology of mosquito-borne disease. *Ecol Lett*. 2023;26(7):1029–49.
15. Schmidt CA, Comeau G, Monaghan AJ, Williamson DJ, Ernst KC. Effects of desiccation stress on adult female longevity in *Aedes aegypti* and *Ae. albopictus* (Diptera: Culicidae): results of a systematic review and pooled survival analysis. *Parasit Vectors*. 2018 Apr 25;11(1):267.
16. Lega J, Brown HE, Barrera R. *Aedes aegypti* (Diptera: Culicidae) Abundance Model Improved With Relative Humidity and Precipitation-Driven Egg Hatching. *J Med Entomol*. 2017 Sep 1;54(5):1375–84.
17. Xu HY, Fu X, Lee LKH, Ma S, Goh KT, Wong J, et al. Statistical Modeling Reveals the Effect of Absolute Humidity on Dengue in Singapore. *PLoS Negl Trop Dis*. 2014 May 1;8(5):e2805.
18. Giannini A, Kushnir Y, Cane MA. Interannual Variability of Caribbean Rainfall, ENSO, and the Atlantic Ocean. *J Clim*. 2000 Jan 15;13(2):297–311.
19. Taylor MA, Stephenson TS, Owino A, Chen AA, Campbell JD. Tropical gradient influences on Caribbean rainfall. *J Geophys Res Atmospheres* [Internet]. 2011 [cited 2024 May 28];116(D21). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015580>
20. McGregor GR, Ebi K. El Niño Southern Oscillation (ENSO) and Health: An Overview for

- Climate and Health Researchers. *Atmosphere*. 2018 Jul;9(7):282.
21. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. El Niño and health. *The Lancet*. 2003 Nov 1;362(9394):1481–9.
 22. CARPHA. State of Public Health in the Caribbean 2017-18 [Internet]. [cited 2024 May 28]. Available from: <https://carpha.org/What-We-Do/Health-Information/State-of-Public-Health>
 23. Climate change and health in small island developing states: A WHO special initiative [Internet]. Geneva: World Health Organisation; [cited 2021 Jul 29]. Available from: <https://www.who.int/publications-detail-redirect/climate-change-and-health-in-small-island-developing-states>
 24. PAHO/WHO. PAHO/WHO Data - Dengue fever cases [Internet]. Pan American Health Organization / World Health Organization. 2015 [cited 2024 May 28]. Available from: <https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/252-dengue-pais-ano-en.html>
 25. Robert MA, Rodrigues HS, Herrera D, de Mata Donado Campos J, Morilla F, Del Águila Mejía J, et al. Spatiotemporal and meteorological relationships in dengue transmission in the Dominican Republic, 2015–2019. *Trop Med Health*. 2023 Jun 2;51(1):32.
 26. Freitas A, Rodrigues HS, Martins N, Iutis A, Robert MA, Herrera D, et al. Multiplicative Mixed-Effects Modelling of Dengue Incidence: An Analysis of the 2019 Outbreak in the Dominican Republic. *Axioms*. 2023 Feb;12(2):150.
 27. United Nations. List of Small Island Developing States - Sustainable Development Knowledge Platform [Internet]. [cited 2021 Jul 29]. Available from: <https://sustainabledevelopment.un.org/topics/sids/list>
 28. Estadística (ONE) ON de. Estimaciones y proyecciones demográficas [Internet]. Oficina Nacional de Estadística (ONE). [cited 2024 May 24]. Available from: <https://www.one.gob.do/datos-y-estadisticas/temas/estadisticas-demograficas/estimaciones-y-proyecciones-demograficas/>
 29. Izzo M, Roskopf CM, Aucelli PPC, Maratea A, Méndez R, Pérez C, et al. A New Climatic Map of the Dominican Republic Based on the Thornthwaite Classification. *Phys Geogr* [Internet]. 2010 Sep 1 [cited 2024 May 27]; Available from: <https://www.tandfonline.com/doi/abs/10.2747/0272-3646.31.5.455>
 30. World Bank. Climate Change Knowledge Portal [Internet]. [cited 2024 May 27]. Available

- from: <https://climateknowledgeportal.worldbank.org/>
31. Muñoz Sabater J. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2019;
 32. McKee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration to time scales. Eighth Conf Appl Climatol Am Meteorol Soc. 1993;
 33. Svoboda M, Hayes M, Wood DA, World Meteorological Organization (WMO). Standardized Precipitation Index User Guide [Internet]. [cited 2024 May 24]. Available from: <https://library.wmo.int/records/item/39629-standardized-precipitation-index-user-guide>
 34. Beguería S, Vicente-Serrano SM. SPEI: Calculation of the Standardized Precipitation-Evapotranspiration Index [Internet]. 2023 [cited 2024 May 24]. Available from: <https://cran.r-project.org/web/packages/SPEI/index.html>
 35. National Oceanic and Atmospheric administration (NOAA). Climate Prediction Center - Monitoring & Data Index [Internet]. [cited 2024 Jan 15]. Available from: <https://www.cpc.ncep.noaa.gov/data/>
 36. Nilles EJ, Paulino CT, de St. Aubin M, Restrepo AC, Mayfield H, Dumas D, et al. SARS-CoV-2 seroprevalence, cumulative infections, and immunity to symptomatic infection – A multistage national household survey and modelling study, Dominican Republic, June–October 2021. *Lancet Reg Health - Am.* 2022 Dec 1;16:100390.
 37. Muench H. Catalytic Models in Epidemiology [Internet]. Harvard University Press; 2013 [cited 2024 Jun 24]. Available from: <https://www.degruyter.com/document/doi/10.4159/harvard.9780674428928/html>
 38. Hens N, Aerts M, Faes C, Shkedy Z, Lejeune O, Van Damme P, et al. Seventy-five years of estimating the force of infection from current status data. *Epidemiol Infect.* 2010 Jun;138(6):802–12.
 39. Hoze N. Rsero-package: Rsero: Estimate the annual force of infection using... in nathoze/Rsero: Estimate the annual force of infection using serological data [Internet]. [cited 2024 May 23]. Available from: <https://rdr.io/github/nathoze/Rsero/man/Rsero-package.html>
 40. Watanabe S. A widely applicable Bayesian information criterion. *J Mach Learn Res.* 2013 Mar 1;14(1):867–97.
 41. Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD. Bayesian measures of model complexity

- and fit. *J R Stat Soc Ser B Stat Methodol.* 2002 Oct 1;64(4):583–639.
42. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. *Environ Res.* 2015 Oct 1;142:319–27.
 43. Camacho A, Bouhenia M, Alyusfi R, Alkohlani A, Naji MAM, de Radiguès X, et al. Cholera epidemic in Yemen, 2016–18: an analysis of surveillance data. *Lancet Glob Health.* 2018 Jun 1;6(6):e680–90.
 44. Lowe R, Stewart-Ibarra AM, Petrova D, García-Díez M, Borbor-Cordova MJ, Mejía R, et al. Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador. *Lancet Planet Health.* 2017 Jul 1;1(4):e142–51.
 45. Bracher J, Held L. Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *Int J Forecast.* 2022 Jul 1;38(3):1221–33.
 46. Colón-González FJ, Bastos LS, Hofmann B, Hopkin A, Harpham Q, Crocker T, et al. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Med.* 2021 Mar 4;18(3):e1003542.
 47. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am J Epidemiol.* 2013 Nov 1;178(9):1505–12.
 48. Siraj AS, Oidtmann RJ, Huber JH, Kraemer MUG, Brady OJ, Johansson MA, et al. Temperature modulates dengue virus epidemic growth rates through its effects on reproduction numbers and generation intervals. *PLoS Negl Trop Dis.* 2017 Jul 19;11(7):e0005797.
 49. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math.* 1991 Mar 1;43(1):1–20.
 50. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Stat Sci.* 2017 Feb;32(1):1–28.
 51. *Forecasting: Principles and Practice (2nd ed)* [Internet]. [cited 2024 May 31]. Available from: <https://otexts.com/fpp2/>
 52. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillerá-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography.* 2017;40(8):913–29.

53. Gibb R, Colón-González FJ, Lan PT, Huong PT, Nam VS, Duoc VT, et al. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nat Commun.* 2023 Dec 11;14(1):8179.
54. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. *J Am Stat Assoc* [Internet]. 2012 Jan 1 [cited 2021 Aug 2]; Available from: <https://www.tandfonline.com/doi/abs/10.1198/016214506000001437>
55. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. *PLOS Comput Biol.* 2021 Feb 12;17(2):e1008618.
56. Bosse NI, Gruson H, Cori A, van Leeuwen E, Funk S, Abbott S. Evaluating Forecasts with scoringutils in R [Internet]. arXiv; 2022 [cited 2024 Jan 17]. Available from: <http://arxiv.org/abs/2205.07090>
57. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proc Natl Acad Sci.* 2019 Nov 26;116(48):24268–74.
58. Li C, Liu Z, Li W, Lin Y, Hou L, Niu S, et al. Projecting future risk of dengue related to hydrometeorological conditions in mainland China under climate change scenarios: a modelling study. *Lancet Planet Health.* 2023 May 1;7(5):e397–406.
59. Bowman LR, Tejeda GS, Coelho GE, Sulaiman LH, Gill BS, McCall PJ, et al. Alarm Variables for Dengue Outbreaks: A Multi-Centre Study in Asia and Latin America. *PLOS ONE.* 2016 Jun 27;11(6):e0157971.
60. Hofman JM, Watts DJ, Athey S, Garip F, Griffiths TL, Kleinberg J, et al. Integrating explanation and prediction in computational social science. *Nature.* 2021 Jul;595(7866):181–8.
61. Sabin AB. Research on Dengue during World War II. *Am J Trop Med Hyg.* 1952 Jan 1;1(1):30–50.
62. Halstead SB. Dengue. *The Lancet.* 2007 Nov 10;370(9599):1644–52

Chapter 7

Discussion

In this thesis I have explored the ways in which intrinsic immune drivers and extrinsic drivers such as population behaviour or climatic variation shape the dynamics of viral outbreaks. Understanding the relative role of these drivers can have important implications for disease control, particularly when considering viruses that pose major public health risks. This research focused on two pressing viral threats; SARS-CoV-2 and dengue virus, exploring their transmission dynamics in complex high-incidence settings.

7.1 Summary of findings

At the outset of this PhD in October 2020, there was limited understanding of SARS-CoV-2 immune dynamics. Key questions included: the extent of protection against reinfection afforded by primary SARS-CoV-2 infection, and indeed whether phenomena such as antibody-dependent enhancement might be observed; the duration of protection from neutralising antibodies; and the extent of cross-immunity from previous infection with seasonal coronaviruses [1–3]. As a result of this, it was not possible to simulate potential long-term dynamics for SARS-CoV-2, without assuming dynamics similar to SARS or influenza [4]. By September 2020, preliminary evidence from a SARS-CoV-2 outbreak on a fishing vessel suggested neutralising antibodies were protective against reinfection, as the three individuals with prior neutralising antibodies were not infected [5]. However, the extent of this protection was unclear, as were the implications of this for widespread transmission in the general population.

In my first published paper (Chapter 3), I analysed longitudinal PCR and serological testing data from a large workplace cohort of over 4,400 employees in the United States. In this study, we

found that primary infection with SARS-CoV-2 provides protection against reinfection for most individuals, at least over a 6-month time period. Our analysis adjusted for confounders, which is particularly important when considering reinfection risk. If a specific subset of the cohort is at higher risk of infection then they will be more likely to be both initially seropositive and then also to have a subsequent reinfection, which would inflate estimates of reinfection risk. Similarly, we found that accurate estimation of reinfection risk was dependent on the cutoff date used to define baseline seroprevalence (and the subsequent window to observe possible reinfections). Notably, we found that unadjusted estimates using early cutoff weeks resulted in point estimates for an odds ratio of reinfection above one, demonstrating how early data can give misleading impressions of immune dynamics in evolving outbreaks with novel pathogens.

These findings then informed the parameterisation of a compartmental model of SARS-CoV-2 transmission in Chapter 4, which I fit to surveillance and serological data to understand drivers of transmission in the Dominican Republic. In this analysis, I found that COVID-19 dynamics were largely driven by substantial accumulation of immunity, estimating that around a third of the population were infected by the end of 2020 and over 80% by the end of 2021. The increasing levels of population immunity kept R_t around one despite increases in contact rates, except during periods where novel variants emerged. I retrospectively assessed the impact of the vaccination campaign, which primarily administered CoronaVac, finding that the speed and timing of the campaign had an important impact on disease burden, despite high levels of post-infection immunity in the population. I also quantified the potential effect of using vaccine products with higher efficacy (such as Pfizer/BioNTech or Oxford/AstraZeneca), aiming to understand the impact of inequities and delays in COVID-19 vaccine roll out in low- and middle-income countries. Here I found that, while using a higher efficacy product would have reduced disease burden over subsequent waves, delaying the vaccination program to wait for such a product to be delivered (for instance, through the COVAX mechanism) would have resulted in increased burden. A novel aspect of this work involved directly assessing the trade-off between changes in population mobility and vaccination on disease burden. This illustrated important nonlinearities in public health response measures; for instance, quantifying how initial increases in vaccine coverage had a bigger impact on disease burden than later ones due to the age-targeted rollout. We also quantified the return-to-normality afforded by the vaccination campaign, finding that in the

absence of vaccination, a 10-20% reduction of mobility would have been required to maintain equivalent death and hospitalisation outcomes.

Chapter 5 presented the results of an analysis exploring how climate variation and serotype competition drive dengue outbreak dynamics in Singapore. While both climate and changes in population immunity caused by shifts in dominant serotypes are well-recognised drivers of dengue outbreaks, few modelling frameworks account for both factors [6,7]. I used a Bayesian hierarchical modelling framework to quantify the impact of temperature, rainfall, El Niño Southern Oscillation (ENSO) and switches in the dominant DENV serotype on dengue incidence. I explored different climatic indicators at a range of biologically plausible lags to identify the best predictors of weekly dengue incidence, as well as different serotype indicators. In this analysis I found delayed and non-linear relationships between climatic variation and dengue incidence. I found evidence of a non-linear association between maximum temperature and dengue risk, with decreasing risk at very high temperatures. This has important implications for future dengue transmission in Singapore as climate change may shift seasonal transmission patterns, resulting in more transmission earlier or later in the season. I found that intermediately wet conditions were associated with increased dengue risk, with decreased risk in very dry conditions, as well as non-linearly increasing risk in El Niño conditions. Finally, I found a non-linear relationship between the time since a switch in the dominant serotype and dengue risk, with increased risk in the years immediately following a switch and increased risk again at 6+ years since a switch. This likely reflects the accumulation of susceptible individuals in the population, as well as risk associated with the growth of non-dominant serotypes.

I then adapted this model to forecast dengue incidence for early warning, generating accurate forecasts from 2-8 weeks ahead. I assessed the model's ability to predict dengue incidence and the probability of an outbreak using strictly proper scoring rules [8]. Additionally, I conducted receiver operator characteristic (ROC) analysis, calculating the model's hit rate (sensitivity) and false alarm rate (1-specificity) in detecting outbreaks and providing a more operationally focused assessment of model performance. We found that including serotype dynamics helped to explain interannual variability in dengue transmission and improved the quality of the dengue forecasts, particularly the ability to discriminate between outbreak and non-outbreak weeks. These findings demonstrate

the utility of incorporating serotype competition as a proxy for immune dynamics within forecasting frameworks, particularly to predict interannual variability in incidence.

In Chapter 6, I used a Bayesian spatiotemporal model to quantify the impact of climatic variation on dengue risk in the Dominican Republic. Again, I explored different climatic indicators at a range of lags. While I found similar non-linear relationships between Niño 3.4 SSTA, maximum temperature and dengue incidence as those in Singapore, in the Dominican Republic, relative humidity and drought indices were also strong predictors. I found evidence of increased dengue risk with increasing humidity, as well as during drought periods, with decreased risk following exceptionally wet periods. The relationship between temperature and dengue incidence found in Singapore was broadly similar to the non-linear thermal performance curves identified in laboratory experiments with decreasing dengue risk at very high temperatures (upwards of 32 °C) [45]. Contrastingly, in the Dominican Republic, I found increasing uncertainty around the effect estimates at higher temperatures. This could be because fewer locations and/or weeks are too hot for dengue transmission in the Dominican Republic, making this non-linearity harder to detect.

In this analysis, I also incorporated additional epidemic drivers into the modelling framework, leveraging dengue seroprevalence data. Specifically, I included weighted lagged cases (representing the force of infection) as well as an estimated proportion susceptible in the current dengue season, informed by serocatalytic modelling. I estimated an average annual attack rate of around 12%, suggesting approximately 7.5% of dengue infections are reported. Despite the approximate nature of the epidemic terms, their inclusion helped to account for additional unexplained interannual variation when compared with a climate-only model and improved the model's predictive ability. I then assessed the influence of each covariate in predicting temporal and spatial dynamics. The most influential drivers of temporal dengue risk were weighted lagged cases (representing the force of infection) as well as the Niño 3.4 index and SPI-12 (representing interannual climatic drivers). I then found only weighted lagged cases were the most influential driver of spatial risk, with less evidence for an influence of climatic drivers. Finally, I generated and scored out-of-sample probabilistic predictions of dengue cases to assess the predictive skill of the full epidemic- and climate-informed model. I found that the model was able to capture key

spatiotemporal outbreak dynamics and outperformed a baseline model in a step towards the development of an early warning system.

7.1 Strengths of this work

7.1.1 Integrating multiple data streams

One of the key strengths of this work is the use of modelling frameworks to integrate and jointly analyse different data streams, enabling a better understanding of the drivers of viral transmission. In Chapter 4, I jointly fit a transmission dynamic model to seroprevalence data, reported deaths, hospital occupancy and ICU occupancy over time in the Dominican Republic. This enabled me to parameterise a complex age-structured model, accounting for vaccination as well as changes in hospitalisation and death risk due to improved treatment or the introduction of new variants, and changes in the relationship between measured mobility data and true social contacts. Fitting to multiple data streams can inform different components of the transmission process and make use of uncorrelated error or bias when data sources are from different studies or surveillance systems [9]. In addition to this, I made use of other valuable datasets including: Google mobility reports which, when combined with a simulated contact matrix for the Dominican Republic, helped to inform changes in contact rates over time; vaccination data collected by Oxford's Our World in Data to inform vaccination rates in the population; and SARS-CoV-2 sequence data published on GISAID to inform the timing of introduction of new variants [10–13].

The ideal complexity of a transmission dynamic model is a trade-off between accuracy, the ability to reproduce observed epidemic dynamics, and transparency, being able to understand how different parameters interact to affect modelled dynamics [14]. In this case, a relatively complex model structure was required to characterize the non-linear effects of immunity, social contacts and viral evolution on transmission, and conduct plausible counterfactual analysis. Fitting to multiple data streams allowed me to distinguish between alternative hypotheses for disease transmission. For instance, from the data it is difficult to assess the true size of the first wave of SARS-CoV-2 transmission, as cases (and probably deaths) were likely substantially underreported

due to limited testing capacity. In light of this, without the understanding of population immunity gained through the serosurvey, the model would not have been able to distinguish between a scenario with a very large first wave generating high levels of immunity and reducing the effective reproduction number, R_t , or a scenario with low transmission with a lower initial estimated basic reproduction number, R_0 , or increased influence of reductions in social contacts due to non-pharmaceutical interventions. By jointly fitting to both data streams, and incorporating changes in contacts over time, I found that the scenario best supported by the data lay in the middle of these two extremes; where cases and deaths were underreported in the first wave and a combination of a steady build-up of immunity and reduction in social contacts drove R_t to around 1 for much of 2020.

Similarly, in Chapter 5, by combining dengue case data, climate data, an ENSO indicator and serotype surveillance data, I was able to both better quantify the impact of climatic variation on dengue incidence and to improve forecasting ability. While dengue forecasting models have shown good performance at predicting seasonal patterns of transmission, a longstanding challenge in the field has been how to predict large outbreaks in advance, with models generally also showing poor predictive ability early in the season [6]. Ideally, dengue forecasting models would be able to give early warning of large outbreaks, as well as accurately predict peak timing and intensity. This warning should be given with operationally useful lead time to allow for vector control and other public health interventions to be carried out [6,7,15]. Previous research in Singapore found that a lead time of three months was optimal for dengue control, with 1-3 months lead time also offering operational utility [16].

Immunity is well-recognised to be an important driver of dengue transmission, with much theoretical modelling work demonstrating the role of serotype competition in determining dynamics [17,18]. However, to date few dengue forecasting models incorporate serotype or immunity data. This is partly because the majority of dengue early warning forecasting models (particularly those employed operationally) are based on statistical regression frameworks, where it is more challenging to account for the nonlinearity introduced by the accumulation of immunity in the population [19]. Chapters 5 and 6 address this question, aiming to incorporate proxies of immunity into forecasting frameworks, using different approaches based on the data available.

Chapter 5 contributes to the field of dengue prediction by demonstrating that the inclusion of serotype data helps to account for unexplained interannual variation in dengue case data. By including an indicator for the time since a switch in serotype, I was able to improve predictions of dengue incidence at short lead times and improve outbreak detection at 2–8 weeks lead time, when compared with a climate-only model. The model developed also could also be used to generate forecasts under different scenarios, for instance, what would forecasts look like with typical climatology and a switch in dominant serotype or with an El Niño event? This work was developed in collaboration with the National Environmental Agency in Singapore and offers a more interpretable forecasting framework than the current LASSO-based forecasting model [20]. As part of this ongoing collaboration with the NEA, the Bayesian hierarchical model presented in Chapter 5 will be integrated into Singapore’s routine dengue forecasting, as part of an ensemble forecasting approach.

In Chapter 6, I combine gridded climate data, ENSO and drought indicators with dengue seroprevalence data from 2021, incorporating proxies for underlying immune drivers into a climate-informed forecasting model. Specifically, I used an estimate of the annual force of infection for dengue to calculate the average reporting ratio for dengue and approximate the proportion infected in each dengue season. As well as this, I incorporated lagged cases weighted by the serial interval distribution, to approximate the force of infection. While this was a relatively crude approximation, including epidemic elements representing the force of infection and changes in population immunity allowed me to estimate the impact of climate drivers on transmissibility, or R_0 , rather than on dengue cases, and improved spatiotemporal predictive ability.

7.1.2 Representative data

Another strength of this thesis is the variety and quality of the data used throughout. In particular, the serological data used in Chapters 4 and 6 is from a high quality nationally representative dataset [21]. The serological study used a multistage survey design, accounting for population, urban-rural divide, and spatial distribution of selected clusters. Adults and children ≥ 5 years old were eligible to partake in the study and over 6,600 individuals were surveyed. Nationally representative serosurveys are relatively unusual, with serological surveys often leveraging samples from blood

donors or other convenience samples, limiting the representativeness of the data [22]. This survey was performed between June and October 2021, providing insight into SARS-CoV-2 dynamics as vaccines began to be rolled out in the Dominican Republic. Serosurveys of this quality were rarely performed outside of high-income countries in the pandemic, and few nationally representative studies were conducted in Latin America and the Caribbean [23]. Serosurvey samples were tested for SARS-CoV-2 using the Roche Elecsys electrochemiluminescence immunoassay, which had high sensitivity and specificity, according to large non-manufactured sponsored surveys [24,25]. This SARS-CoV-2 seroprevalence data informed analysis in Chapter 4. Subsequently, a subset of samples were selected for testing with a DENV1-4 ELISA assay, which I was able to leverage to improve the predictive power of a climate-informed model in Chapter 6.

This thesis also makes use of Singapore’s dengue surveillance data, through a collaboration with the National Environmental Agency in Singapore. This is high quality surveillance data, with laboratory confirmation of dengue cases and a world-class dataset of DENV serotype surveillance from 2006 [26].

7.1.3 Reproducibility

Throughout my PhD, I have aimed to make my research as reproducible and transparent as possible. A reproducible data analysis is defined as one where the data and code needed to perform the analysis are made available to others for independent study and analysis [27]. This must, however, be balanced with ethical issues surrounding individual-level data, such as ensuring anonymity for study participants.

Chapter 3 relied on individual-level serological and PCR data, including sensitive data such as race and ethnicity, job categories, and history of chronic disease. As a result of this, weekly aggregated PCR and serological data were made available on GitHub, as well as code to reproduce the simulation analysis. I also included a file of odds ratio estimates such that all figures in the manuscript could be recreated. All data and code used in the analysis in Chapter 4 are available on GitHub. As well as making the analysis reproducible, this brings together several useful datasets for the Dominican Republic (including surveillance data, GISAID sequence data, Google mobility

report data and vaccination data), creating a repository for potential future analysis. Similarly, for Chapter 5, all data and code needed to reproduce the analysis are available on GitHub. Additionally, a more comprehensive code base was shared with the National Environment Agency, to enable analysts to run the model independently and incorporate it into routine dengue ensemble forecasts. This contained data wrangling steps needed to routinely run forecasts from raw data inputs. I also plan to publish code and data needed to reproduce the analysis presented in Chapter 6 as part of the publication process.

7.2 Limitations

7.2.1 Data

While the variety and quality of the data used throughout this thesis is a strength of the work presented, there are limitations, often inherent to the data used, that must be considered when interpreting study results.

Chapter 3 analysed data from an opportunistic workplace cohort which is unlikely to be representative of the national population. However, I did not identify any workplace outbreaks within the study period, and so infections are likely to represent community transmission unlike, say, in healthcare worker cohorts where transmission is often nosocomial [28]. Additionally, reinfection was not confirmed by genomic sequencing and as such may reflect instances of prolonged viral shedding rather than true reinfection. While these cases appear to be relatively rare, this would result in an overestimation of the odds ratio for reinfection, meaning that our estimate of reinfection risk is conservative, reflecting the minimum possible association between antibodies and future infection risk.

In Chapter 4, COVID-19 surveillance data streams were likely to have been poorer quality in the early phase of the pandemic. Hospital and ICU data only began to be reported in September 2020, and deaths may have been underreported in the first wave due to limited testing capacity. Indeed, test positivity was high in this period, above 15%, which is indicative of underreporting [11]. To

address this, I considered different weighting options for surveillance and serological data within the likelihood, giving the serological data more weight. As in many other low- and middle-income country settings, genomic sequencing was limited throughout the pandemic, which meant I was unable to precisely date the introduction and rise to dominance of novel variants from available data. As a result of this, I fit parameters describing these dynamics for Delta and conducted sensitivity analysis around different introduction periods for Mu. Similarly, while the real-time publication of mobility data during an outbreak, such as the Google mobility reports used here, is unprecedented, there are limitations in using this as a proxy for social contacts. Firstly, Google mobility reports are only informative of the mobility patterns among users of Google products. Secondly, the relationship of this data to social contacts affecting transmission may change during the pandemic, as non-pharmaceutical interventions such as physical distancing measures and face masks began to be employed. To take this into account I fit two parameters modifying the relationship between Google mobility data and contact rates within the model framework, one for each year of the pandemic.

In Chapter 6, I used reported dengue cases in the analysis, which have not necessarily been laboratory confirmed. Due to the nonspecific clinical presentation of most dengue cases, this could result in misdiagnosis of other acute febrile illness as dengue, particularly during the large chikungunya outbreak of 2014 and Zika outbreak in 2015/16. The impact of dengue misdiagnosis, particularly if it were to vary over time, is ameliorated by the use of year-specific spatial random effects. These random effects can account for interannual variation in dengue incidence, potentially due to increased misdiagnosis due to a concurrent outbreak of non-dengue acute febrile illness, as well as other factors such as changing vector control efforts. Additionally, outside of the large chikungunya and Zika outbreaks, dengue burden is much higher than other similarly presenting acute febrile illness, suggesting misdiagnosis is unlikely to substantially impact inference of spatiotemporal dynamics. Chapter 6 also makes use of serological data, tested using an DENV1-4 ELISA assay. Cross-reaction is a common problem faced in serological analysis of flaviruses, particularly when using ELISA assays, which have lower specificity than other assays [29–31]. Despite this, advancements have been made in distinguishing between flavivirus infections. Where longitudinal data is available, this can be achieved by looking at changes in titre values over time or, if only cross-sectional data is available, by jointly modelling dengue and Zika

seroprevalence [32–34]. In this analysis, we use cross-sectional DENV seroprevalence data from 2021, where cross-reaction from previous infection with other flaviviruses could inflate estimates of DENV seropositivity. However, I found support for high seroprevalence of DENV antibodies in the population from supplementary datasets, testing for DENV1 and DENV2 neutralising antibodies. These found similar patterns of age-stratified seropositivity as the broader, less specific DENV1-4 ELISA assay, suggesting high levels of endemic dengue transmission with nearly all adults having had prior infection.

In Chapters 5 and 6, I include different approximations for changing population immunity to dengue virus into forecasting frameworks. Predictions for both of these models would likely be improved by more accurate measurements of immunity over time, for instance, by incorporating modelled seroprevalence to the currently dominant serotype. However, until serological testing is routinely performed, using well-validated assays able to distinguish between prior infection with a particular DENV serotype, this work shows proxies for temporal immune dynamics drawn from serotype surveillance or cross-sectional serology can add valuable contributions to forecasting frameworks.

7.2.2 Model assumptions

While trade-offs between model accuracy, transparency and, in some cases, analytical tractability are intrinsic to infectious disease modelling, the work presented here is limited by modelling assumptions, which are important to consider when interpreting model findings. In Chapter 4, we lacked age-stratified vaccination data, and as such made several assumptions regarding this process. I only incorporated protection from full vaccination (after two doses in a full primary vaccination schedule) and did not incorporate the partial protection afforded by a single dose. However, as vaccine efficacy from Sinovac-CoronaVac from a single dose is low, I do not expect this to have a meaningful impact on our results [35,36]. As we lacked age-stratified data, I assumed that vaccine doses were divided equally between eligible age groups for vaccination, according to VacúnateRD policy [37]. If all eligible age groups were completely vaccinated, excess doses were distributed throughout the rest of the population at random, reflecting the vaccination of healthcare workers and other vulnerable individuals of all ages. Finally, I only incorporated vaccine efficacy

against severe disease and not against symptomatic disease or death. Substantial uncertainty in the efficacy of CoronaVac meant that I was unable to construct a full pipeline of vaccine efficacy against multiple end points within the modelling framework and as I fit directly to hospitalisation and ICU admissions (but not cases), vaccine efficacy against severe disease was the most relevant estimate to incorporate. As this assumption was important when considering counterfactual scenarios around vaccination impact, I conducted sensitivity analyses around vaccine efficacy and waning assumptions. Specifically, in a sensitivity analysis using lower vaccine efficacy (based on estimates in the literature for protection against symptomatic disease rather than severe disease), we found a lower estimated impact of vaccination. However, results from our sensitivity analyses were in the same order of magnitude and direction, showing the key conclusions from the assessment of potential vaccination strategies are robust.

In Chapters 5 and 6, I made assumptions around causal processes underlying dengue transmission, informing my choices of which climatic and non-climatic variables to test in model selection. In both chapters I decided to test Niño 3.4 sea surface temperature anomalies (SSTAs) measuring the El Niño Southern Oscillation, as well as climatic variables such as temperature, precipitation and humidity within the same model. In other similar studies quantifying the impact of climatic variation on disease these were kept separate, following the logic that the impact of ENSO on transmission is mediated through local weather conditions [38]. Here, I reasoned that, while this is likely to be true to an extent, ENSO is also likely to have other climatic impacts not captured through say, changes in weekly average temperature or cumulative rainfall. For instance, ENSO could result in changes to longer-term climatic variation, influencing annual temperatures determining geographical and seasonal limits of dengue transmission [39]. Similarly, ENSO could impact climatic variables not incorporated in the modelling framework, such as humidity in Chapter 5, or the likelihood of a hurricane or tropical storm in Chapter 6. Finally, in the spatiotemporal analysis presented in Chapter 6, I did not incorporate socioeconomic variables such as level of urbanisation or access to piped water, which are known to impact dengue transmission. Spatial socioeconomic differences are instead accounted for through the inclusion of year-specific spatial random effects. While this likely limits the spatial predictive power of the model, this is not likely to impact the model's suitability as an early-warning framework as these factors often do not change in the time frame (weeks to months) relevant to short-term forecasting.

7.2.3 Generalisability

Another limitation of this work is the extent to which findings are generalisable. In Chapter 3, while we utilize a US-based workplace cohort, we aim to estimate a key epidemiological parameter which is generalizable to other settings. The results of Chapter 3 are dependent on the wider epidemiological context, particularly the variants circulating in the US in the study period between April 2020 and January 2021. However, this period predates the rise to dominance of the first variant of concern (Alpha), and so reinfection risk likely reflects that associated with wild-type SARS-CoV-2. Analyses of key epidemiological parameters, such as reinfection risk, were repeated as new variants emerged to characterise their epidemiological properties. For instance, a study in South Africa demonstrated increased risk of reinfection with the emergence of Omicron, suggestive of immune evasion [40].

In Chapter 4, I sought to develop an in-depth modelling study of SARS-CoV-2 transmission dynamics in the Dominican Republic, able to offer similar insights to studies performed regularly in high-income countries, such as the UK. As such, my approach, data streams and methods were tailored to the setting in question. Despite this, the findings may have broader relevance in understanding transmission dynamics in countries where non-pharmaceutical interventions alone were unable to suppress transmission, those with relatively high levels of population immunity at vaccine roll out and those using a primarily Sinovac-CoronaVac vaccination campaign. For Chapters 5 and 6, I demonstrated the value of developing climate-informed models tailored to local climate-dengue relationships, leveraging the data streams available in each setting to incorporate proxies for immunity, and building on previous work in other settings [41–43]. However, these findings are likely to be relevant to other dengue endemic areas, particularly those experiencing hyperendemic circulation of multiple serotypes.

7.3 Implications and future work

This work has been developed with a view to informing and enhancing epidemic response. This framing has shaped the analyses performed; for instance, designing the counterfactual scenarios

developed in Chapter 4 or the model evaluation methodology used in Chapters 5 and 6. The analysis in Chapter 3, focused on estimating the relative risk of reinfection, a key epidemiological parameter needed to consider the potential impact of future non-pharmaceutical interventions, and to inform vaccine development. In Chapter 4, counterfactual analysis around COVID-19 vaccination strategies in the Dominican Republic aimed to add to lessons learnt from the pandemic to inform future pandemic preparedness, particularly considering optimal vaccination strategies and arguing for improved global vaccine equity.

In Chapter 5, I developed a functional forecasting model for early warning in Singapore, and through ongoing work with the National Environmental Agency I am supporting its implementation into routine dengue forecasting. Similarly in Chapter 6 I developed a prototype space-time dengue model, demonstrating good predictive ability; an important first step in developing an early warning framework. These two chapters demonstrate the value of proxies for population immunity in improving predictive power of forecasting models, and particularly in addressing the challenge of predicting large outbreak years. Further work is needed to explore methods to incorporate measures of seroprevalence directly into forecasting frameworks. Particularly, work to disentangle serotype-specific population immunity levels, through improved algorithms for analysis of assay data or joint analysis of serotype surveillance and serological data, would allow for a better characterisation of outbreak risk associated with the growth of a specific serotype. Here, research aimed at improving the predictive performance of mechanistic models in hyperendemic settings could be beneficial, as these can be parameterised to include multiple serotypes and more fully capture dengue immune dynamics. Potential avenues for improving mechanistic dengue forecasting models include integrating local climate-dengue relationships to inform transmissibility or fitting jointly to serological or serotype surveillance data as well as dengue case surveillance. This work would rely on the expansion of reliable serological and serotype surveillance data streams, which are currently limited in many low and middle-income countries experiencing a high burden of dengue. Future work is also needed to compare state of the art climate-informed statistical dengue forecasting frameworks with mechanistic or semi-mechanistic approaches. Ultimately, ensemble forecasting frameworks, leveraging the benefits of multiple models (as well as mitigating their potential pitfalls) are likely to offer greatest predictive performance [44,6].

7.4 Concluding remarks

In this thesis, I aimed to bring together diverse modelling approaches and data sources to disentangle the role of immunity from extrinsic drivers, such as climate and behaviour, in driving acute viral outbreaks. The research presented has advanced understanding of these interactions in settings with complex landscapes of population immunity; for instance, with multiple SARS-CoV-2 variants in circulation and a mix of post-infection and post-vaccination immunity, or with hyperendemic circulation of dengue serotypes. Disentangling these dynamics within modelling frameworks allows us to more accurately predict future or counterfactual transmission patterns, thereby better informing epidemic response.

References

1. Lee WS, Wheatley AK, Kent SJ, DeKosky BJ. Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nat Microbiol*. 2020 Oct;5(10):1185–91.
2. Tillett RL, Sevinsky JR, Hartley PD, Kerwin H, Crawford N, Gorzalski A, et al. Genomic evidence for reinfection with SARS-CoV-2: a case study. *The Lancet Infectious Diseases* [Internet]. 2020 Oct 12 [cited 2020 Oct 13];0(0). Available from: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30764-7/abstract](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30764-7/abstract)
3. Overbaugh J. Understanding protection from SARS-CoV-2 by studying reinfection. *Nat Med*. 2020 Nov;26(11):1680–1.
4. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. 2020 May 22;368(6493):860–8.
5. Addetia A, Crawford KHD, Dingens A, Zhu H, Roychoudhury P, Huang ML, et al. Neutralizing Antibodies Correlate with Protection from SARS-CoV-2 in Humans during a Fishery Vessel Outbreak with a High Attack Rate. *Journal of Clinical Microbiology* [Internet]. 2020 Oct 21 [cited 2021 Mar 3];58(11). Available from: <https://jcm.asm.org/content/58/11/e02107-20>
6. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *PNAS*. 2019 Nov 26;116(48):24268–74.
7. Finch E, Batista ML, Alcayna T, Lee SA, Fletcher IK, Lowe R. Early warning systems for vector-borne diseases: engagement, methods and implementation. In: *Planetary health approaches to understand and control vector-borne diseases* [Internet]. Wageningen Academic; 2023 [cited 2024 Jan 22]. p. 347–86. Available from: <https://brill.com/edcollchap-oo/book/9789004688650/BP000023.xml>
8. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* [Internet]. 2012 Jan 1 [cited 2021 Aug 2]; Available from: <https://www.tandfonline.com/doi/abs/10.1198/016214506000001437>
9. De Angelis D, Presanis AM, Birrell PJ, Tomba GS, House T. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*. 2015 Mar

- 1;10:83–7.
10. Google. COVID-19 Mobility Reports [Internet]. Available from: <https://www.google.com/covid19/mobility/>
 11. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B, et al. Coronavirus Pandemic (COVID-19) - Dominican Republic Country Profile. Our World in Data [Internet]. 2020 Mar 5 [cited 2021 Jul 30]; Available from: <https://ourworldindata.org/coronavirus/country/dominican-republic>
 12. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*. 2017 Mar 30;22(13):30494.
 13. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*. 2017 Sep 12;13(9):e1005697.
 14. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals* [Internet]. Princeton University Press; 2008 [cited 2024 Jun 12]. Available from: <https://www.jstor.org/stable/j.ctvc4gk0>
 15. Wellcome Trust. Landscape mapping of software tools for climate-sensitive infectious disease modelling [Internet]. 2022 [cited 2022 Feb 4]. Available from: <https://wellcome.org/reports/landscape-mapping-software-tools-climate-sensitive-infectious-disease-modelling>
 16. Hii YL, Rocklöv J, Wall S, Ng LC, Tang CS, Ng N. Optimal Lead Time for Dengue Forecast. *PLOS Neglected Tropical Diseases*. 2012 Oct 18;6(10):e1848.
 17. Ferguson N, Anderson R, Gupta S. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc Natl Acad Sci U S A*. 1999 Jan 19;96(2):790–4.
 18. Wearing HJ, Rohani P. Ecological and immunological determinants of dengue epidemics. *Proc Natl Acad Sci U S A*. 2006 Aug 1;103(31):11802–7.
 19. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. *Environmental Research*. 2015 Oct 1;142:319–27.
 20. Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, et al. Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environ Health Perspect*. 2016 Sep;124(9):1369–75.
 21. Nilles EJ, Paulino CT, de St. Aubin M, Restrepo AC, Mayfield H, Dumas D, et al. SARS-

- CoV-2 seroprevalence, cumulative infections, and immunity to symptomatic infection – A multistage national household survey and modelling study, Dominican Republic, June–October 2021. *The Lancet Regional Health - Americas*. 2022 Dec 1;16:100390.
22. Chen X, Chen Z, Azman AS, Deng X, Sun R, Zhao Z, et al. Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *The Lancet Global Health*. 2021 May 1;9(5):e598–609.
 23. Arora RK, Joseph A, Wyk JV, Rocco S, Atmaja A, May E, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *The Lancet Infectious Diseases*. 2021 Apr 1;21(4):e75–6.
 24. Ainsworth M, Andersson M, Auckland K, Baillie JK, Barnes E, Beer S, et al. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. *The Lancet Infectious Diseases* [Internet]. 2020 Sep 23 [cited 2020 Sep 30];0(0). Available from: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30634-4/abstract](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30634-4/abstract)
 25. Nilles EJ, Karlson EW, Norman M, Gilboa T, Fischinger S, Atyeo C, et al. Evaluation of Three Commercial and Two Non-Commercial Immunoassays for the Detection of Prior Infection to SARS-CoV-2. *J Appl Lab Med*. 2021 Jul 1;6(6):1561–70.
 26. Lee KS, Lai YL, Lo S, Barkham T, Aw P, Ooi PL, et al. Dengue Virus Surveillance for Early Warning, Singapore - Volume 16, Number 5—May 2010 - *Emerging Infectious Diseases journal - CDC*. [cited 2022 Oct 22]; Available from: https://wwwnc.cdc.gov/eid/article/16/5/09-1006_article
 27. Peng RD, Hicks SC. Reproducible Research: A Retrospective. *Annual Review of Public Health*. 2021 Apr 1;42(Volume 42, 2021):79–93.
 28. Hall VJ, Foulkes S, Charlett A, Atti A, Monk EJ, Simmons R, et al. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). *The Lancet* [Internet]. 2021 Apr 9 [cited 2021 Apr 14];0(0). Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(21\)00675-9/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00675-9/abstract)
 29. Andrade DV, Harris E. Recent advances in understanding the adaptive immune response to Zika virus and the effect of previous flavivirus exposure. *Virus Res*. 2018 Aug 2;254:27–33.
 30. Mansfield KL, Horton DL, Johnson N, Li L, Barrett ADT, Smith DJ, et al. Flavivirus-induced antibody cross-reactivity. *J Gen Virol*. 2011 Dec;92(Pt 12):2821–9.

31. Beck C, Desprès P, Paulous S, Vanhomwegen J, Lowenski S, Nowotny N, et al. A High-Performance Multiplex Immunoassay for Serodiagnosis of Flavivirus-Associated Neurological Diseases in Horses. *BioMed Research International*. 2015;2015(1):678084.
32. Kucharski A, Riley S. Reducing uncertainty about flavivirus infections. *The Lancet Infectious Diseases*. 2017 Jan 1;17(1):13–5.
33. Salje H, Cummings DAT, Rodriguez-Barrquer I, Katzelnick LC, Lessler J, Klungthong C, et al. Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. *Nature*. 2018 May;557(7707):719–23.
34. Kucharski AJ, Kama M, Watson CH, Aubry M, Funk S, Henderson AD, et al. Using paired serology and surveillance data to quantify dengue transmission and control during a large outbreak in Fiji. *eLife*. 7:e34848.
35. Imai N, Hogan AB, Williams L, Cori A, Mangal TD, Winskill P, et al. Interpreting estimates of coronavirus disease 2019 (COVID-19) vaccine efficacy and effectiveness to inform simulation studies of vaccine impact: a systematic review [Internet]. Wellcome Open Research; 2021 [cited 2022 Jan 13]. Available from: <https://wellcomeopenresearch.org/articles/6-185>
36. Wu D, Zhang Y, Tang L, Wang F, Ye Y, Ma C, et al. Effectiveness of Inactivated COVID-19 Vaccines Against Symptomatic, Pneumonia, and Severe Disease Caused by the Delta Variant: Real World Study and Evidence - China, 2021. *China CDC Wkly*. 2022 Jan 28;4(4):57–65.
37. VacunateRD [Internet]. [cited 2024 Jul 3]. Available from: <https://vacunate.gob.do/>
38. Lotto Batista M, Rees EM, Gómez A, López S, Castell S, Kucharski AJ, et al. Towards a leptospirosis early warning system in northeastern Argentina. *J R Soc Interface*. 20(202):20230069.
39. Gibb R, Colón-González FJ, Lan PT, Huong PT, Nam VS, Duoc VT, et al. Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. *Nat Commun*. 2023 Dec 11;14(1):8179.
40. Pulliam JRC, van Schalkwyk C, Govender N, von Gottberg A, Cohen C, Groome MJ, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*. 2022 Mar 15;376(6593):eabn4947.
41. Lowe R, Coelho CA, Barcellos C, Carvalho MS, Catão RDC, Coelho GE, et al. Evaluating

- probabilistic dengue risk forecasts from a prototype early warning system for Brazil. *Elife*. 2016 Feb 24;5:e11285.
42. Lowe R, Stewart-Ibarra AM, Petrova D, García-Díez M, Borbor-Cordova MJ, Mejía R, et al. Climate services for health: predicting the evolution of the 2016 dengue season in Machala, Ecuador. *The Lancet Planetary Health*. 2017 Jul 1;1(4):e142–51.
 43. Colón-González FJ, Bastos LS, Hofmann B, Hopkin A, Harpham Q, Crocker T, et al. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Medicine*. 2021 Mar 4;18(3):e1003542.
 44. Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *PNAS*. 2019 Feb 19;116(8):3146–54.
 45. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. *Ecology Letters*. 2019;22(10):1690–708.

Appendix A

Ethical Approval

Appendix A: Ethical Approval

London School of Hygiene & Tropical Medicine

Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk



Observational / Interventions Research Ethics Committee

Ms Emilie Finch
LSHTM

13 May 2021

Dear Ms Emilie Finch

Study Title: Combining novel data streams to forecast acute febrile illness in the Dominican Republic

LSHTM Ethics Ref: 26214

Thank you for your application for the above research project which has now been considered by the Observational Committee via Chair's Action.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved is as follows:

Document Type	File Name	Date	Version
Consent form	Project_Meridian_ENGLISH_Adult Consent Form	11/11/2020	1
Consent form	Project_Meridian_ENGLISH_Adolescent Assent Form	12/11/2020	1
Consent form	Project_Meridian_SPANISH_Adult Consent Form	18/04/2021	1
Consent form	Project_Meridian_SPANISH_Adolescent Assent Form	18/04/2021	1
Investigator CV	Emilie Finch CV	03/05/2021	1
Protocol / Proposal	Main_protocol	04/05/2021	1
Protocol / Proposal	Main_protocol_COVID_supplement	04/05/2021	1
Investigator CV	2020_Kucharski	04/05/2021	1
Other	Training Research_Ethics_online_training_certificate Kucharski	04/05/2021	1
Local Approval	Main_approval_Carta de aprobación Conabios Harvard	04/05/2021	1
Local Approval	Main_approval_ReviewLetter	04/05/2021	1
Local Approval	Main_approval_UNPHU Bioethics Approval	04/05/2021	1
Investigator CV	LoweR_CV_2021_long	10/05/2021	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using the End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: <http://leo.lshtm.ac.uk>.

Further information is available at: www.lshtm.ac.uk/ethics.

Yours sincerely,



Professor Jimmy Whitworth
Chair

ethics@lshtm.ac.uk
<http://www.lshtm.ac.uk/ethics/>

Improving health worldwide

Appendix B

Supplementary material for Chapter 3

S1 Text: Confounder adjustment for logistic regression analyses

Table A: Confounding variables adjusted for in logistic regression analysis for each cut-off week.

Variable selection was conducted separately for each cut-off week. The final variable set adjusted for was selected using backwards selection minimising RMSE at each step from a list of potential confounders selected a priori. These included: age, sex, race, ethnicity, BMI, state, work location, job category, household size, history of chronic disease, history of smoking and test frequency. Age and sex were considered ‘forced’ variables and were adjusted for in all analyses.

Date	Variables adjusted for
26/04/2020	Sex, age, BMI, history of chronic disease, household size, job category
03/05/2020	Sex, age, history of chronic disease, household size, job category, race, history of smoking
10/05/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
17/05/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
24/05/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
31/05/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
07/06/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
14/06/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
21/06/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
28/06/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
05/07/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
12/07/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
19/07/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
26/07/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
02/08/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
09/08/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
16/08/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
23/08/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking

Appendix B: Supplementary Material Chapter 3

30/08/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
06/09/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
13/09/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
20/09/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
27/09/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
04/10/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
11/10/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
18/10/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
25/10/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
01/11/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
08/11/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
15/11/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
22/11/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
29/11/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
06/12/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
13/12/2020	Sex, age, state, BMI, race, ethnicity, job category, test frequency, history of chronic disease, household size, smoking
20/12/2020	Sex, age, BMI, history of chronic disease, ethnicity, household size, job category, history of smoking, state, test frequency
27/12/2020	Sex, age, BMI, history of chronic disease, ethnicity, household size, smoking, state, test frequency
03/01/2021	Sex, age, ethnicity, household size, smoking
10/01/2021	Sex, age, BMI, state, test frequency
17/01/2021	Sex, age, BMI, race, state

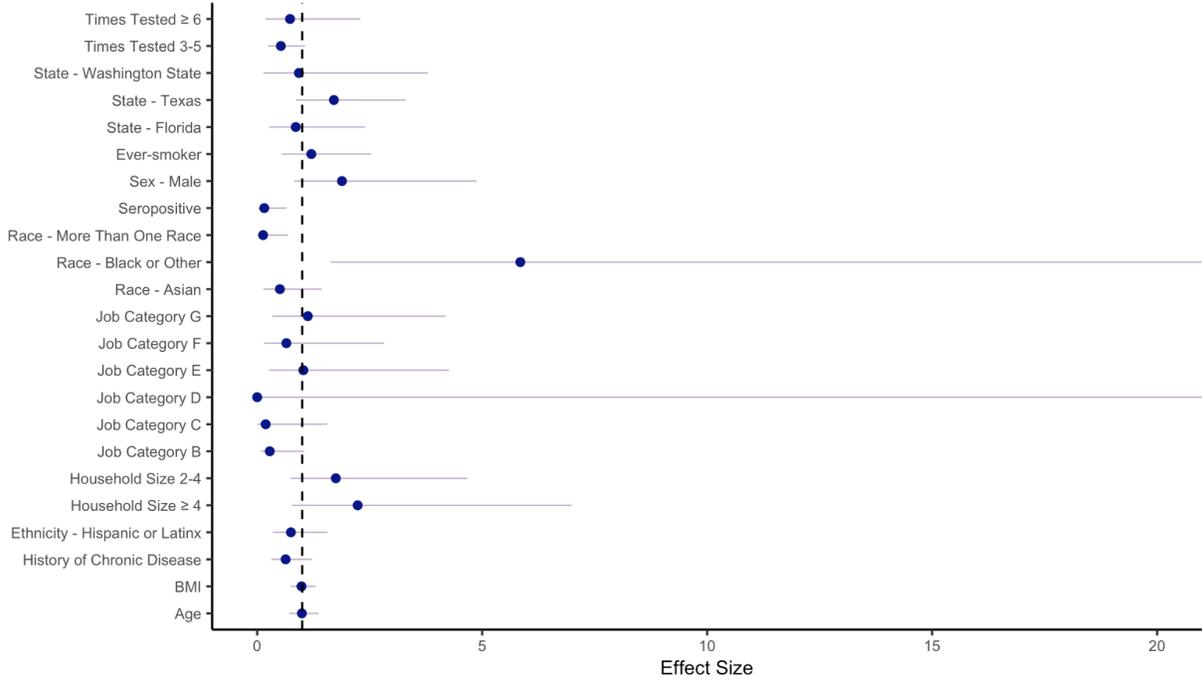


Figure A: Effect sizes of confounders adjusted for in main analysis for the cut-off week of 26/7/2021. Estimates are presented with their associated 95% confidence intervals. The upper bound of the 95% confidence interval for ‘Race - Black or Other’ of 23.7 is not shown. The estimate of ‘Job Category D’ is not well defined due to a low sample size, and the upper bound is not shown. Reference categories for categorical variables are: Times Tested 1-2; State - California; Never-smoker; Sex - Female; Seronegative; Race - White; Job Category A; Household Size 1; Ethnicity - Not Hispanic/Not Latinx; No History of Chronic Disease. Job categories have been anonymised for data protection.

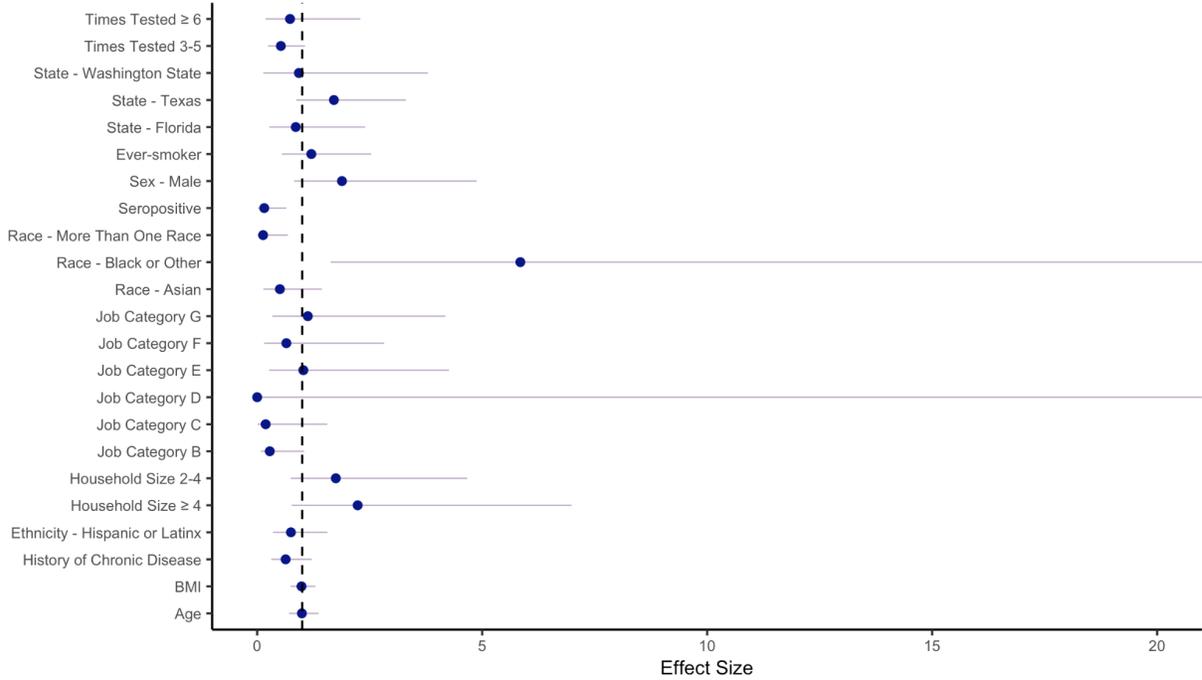


Figure B: Effect sizes of confounders adjusted for in main analysis for the cut-off week of 16/8/2020. Estimates are presented with their associated 95% confidence intervals. The upper bound of the 95% confidence interval for ‘Race - Black or Other’ of 51.4 is not shown. The estimate of ‘Job Category D’ is not well defined due to a low sample size, and the upper bound is not shown. Reference categories not shown for categorical variables are: Times Tested 1-2; State - California; Never-smoker; Sex - Female; Seronegative; Race - White; Job Category A; Household Size 1; Ethnicity - Not Hispanic/Not Latinx; No History of Chronic Disease. Job categories have been anonymised for data protection.

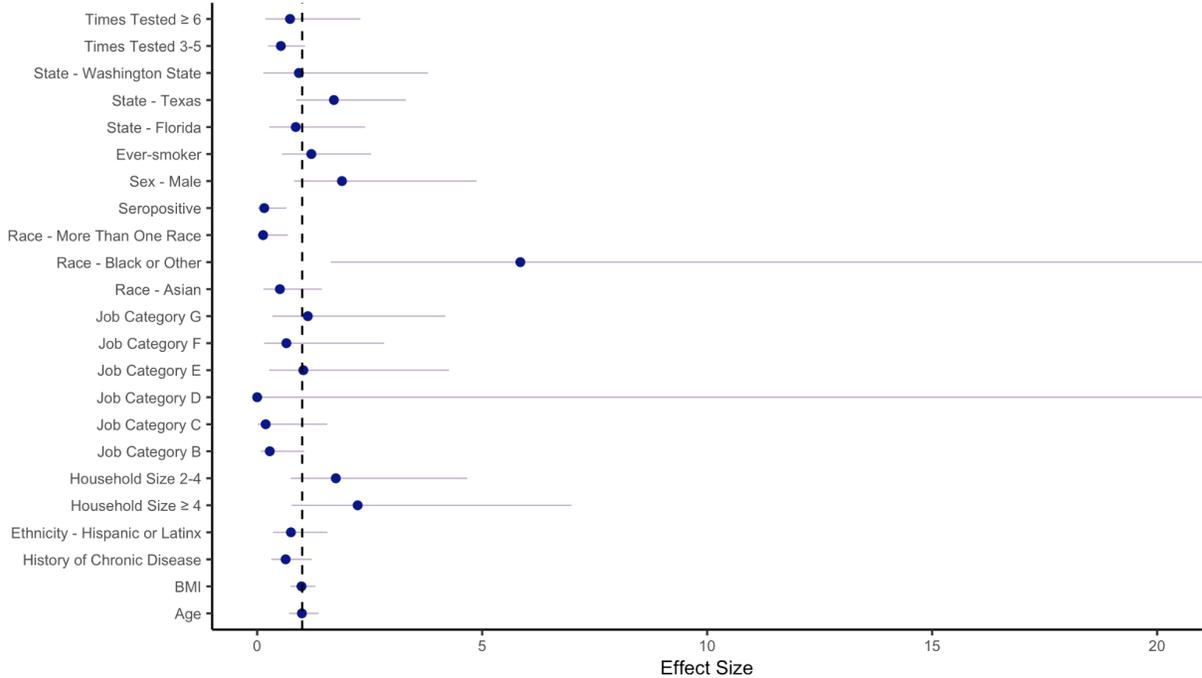


Figure C: Effect sizes of confounders adjusted for in main analysis for the cut-off week of 13/9/2020. Estimates are presented with their associated 95% confidence intervals. The upper bound of the 95% confidence interval for ‘Race - Black or Other’ of 93.7 is not shown. The estimate of ‘Job Category D’ is not well defined due to a low sample size, and the upper bound is not shown. Reference categories not shown for categorical variables are: Times Tested 1-2; State - California; Never-smoker; Sex - Female; Seronegative; Race - White; Job Category A; Household Size 1; Ethnicity - Not Hispanic/Not Latinx; No History of Chronic Disease. Job categories have been anonymised for data protection.

S2 Text: Description of simulation analysis investigating how the estimation of the relative risk of reinfection varies depending on population-level epidemic dynamics

We conducted a simulation analysis to validate our methodological assumption that the most robust estimation of the odds ratio for reinfection would occur in the mid-epidemic period, and to further investigate how the estimation of the odds ratios for reinfection varies depending on population-level epidemic dynamics.

We simulated a two-wave epidemic scenario using a cumulative probability distribution for infection derived from PCR testing data from the study cohort (see Figure 2A in the main text). This was scaled so the overall cumulative probability of infection reflected the overall percentage ever-seropositive in the cohort (8%). We considered a sample size of 2000 individuals over a period of 44 weeks. For each time point i , the number of seropositive and seronegative individuals up to and including that time point was simulated, according to the cumulative probability of infection at time point i (see Eq. 1-2). We then simulated how many of these seropositive and seronegative individuals would test PCR positive or negative after time point i , with a pre-set risk ratio for reinfection of 0.15 (see Eq. 3-6).

$$P(\text{Seropositive}) = \sum_{t=1}^i Pr(X = x_t) \tag{1}$$

$$P(\text{Seronegative}) = 1 - \sum_{t=1}^i Pr(X = x_t) \tag{2}$$

$$P(\text{PCR Positive} \mid \text{Seronegative}) = 1 - \Pi \left(1 - \sum_{t=i+1}^n Pr(X = x_t) \right) \tag{3}$$

$$P(\text{PCR Negative} \mid \text{Seronegative}) = 1 - \left(1 - \Pi \left(1 - \sum_{t=i+1}^n Pr(X = x_t) \right) \right) \tag{4}$$

$$P(\text{PCR Positive} \mid \text{Seropositive}) = 0.15 \times (1 - \prod (1 - \sum_{t=i+1}^n Pr(X = x_t))) \tag{5}$$

$$P(\text{PCR Negative} \mid \text{Seropositive}) = 1 - (0.15 \times (1 - \prod (1 - \sum_{t=i+1}^n Pr(X = x_t)))) \tag{6}$$

Finally, we re-estimated the risk ratio for reinfection for each cut-off time point and assessed the accuracy by comparing the estimated risk ratio using each cut-off week with the ‘true’ ratio of 0.15.

When considering this two-wave epidemic scenario, we found that the accuracy in the estimated risk ratio for reinfection was maximised in the middle of the simulation period (i.e. between the two ‘waves’ of infection risk). This supports our methodological assumption that the most robust estimation of the relative risk of reinfection would occur in between two ‘waves’ of infection in our study cohort, where we considered odds ratios given the need for an underlying regression analysis.

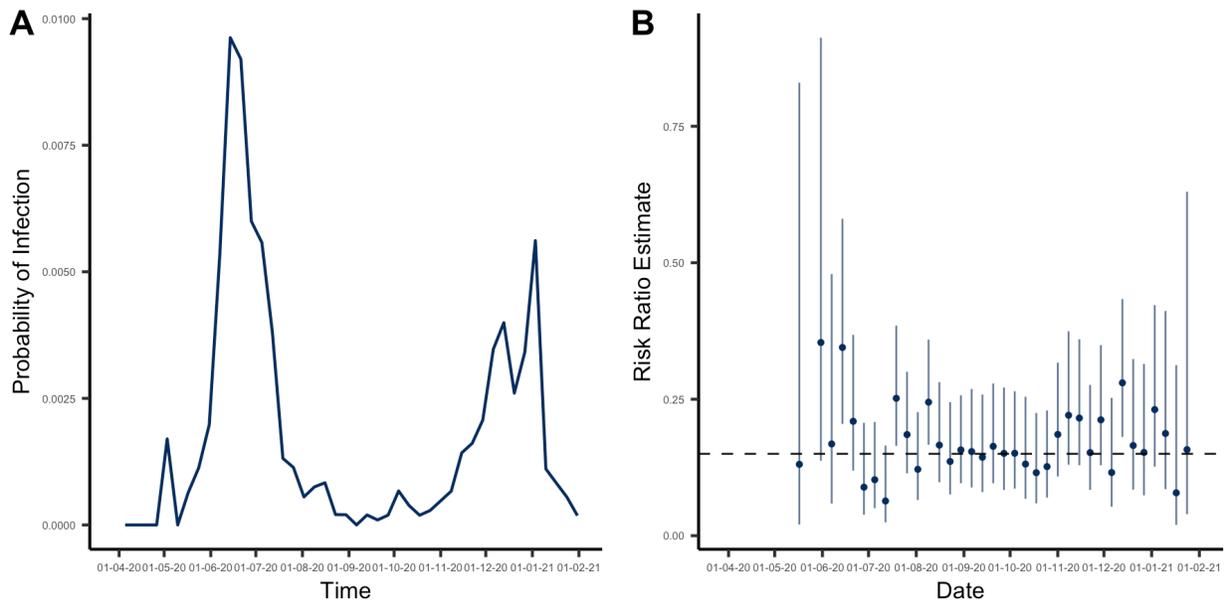


Figure A: A) Probability of infection used for simulation analysis, derived from PCR testing data in the study cohort. B) Risk ratio estimates comparing the risk of reinfection with the risk of primary infection. The

estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline infection status on the x-axis. The dashed line represents the pre-set reinfection risk ratio of 0.15.

Appendix C

Supplementary material for Chapter 4

Supplementary Materials for:

Estimating the impact of population mobility, post-infection immunity and vaccination on SARS-CoV-2 transmission in the Dominican Republic

Emilie Finch^{1*}, Eric J Nilles^{2,3}, Cecilia Then Paulino⁴, Ronald Skewes-Ramm⁴, Colleen Lau⁵, Rachel Lowe^{1,6,7}, Adam J Kucharski¹

¹Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom, ²Harvard Humanitarian Initiative, Cambridge, MA, USA, ³Brigham & Women's Hospital, Boston, MA, USA, ⁴Ministerio de Salud Pública y Asistencia Social, Santo Domingo, Dominican Republic, ⁵School of Public Health, The University of Queensland, Brisbane, Australia, ⁶Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain; ⁷Barcelona Supercomputing Center, Barcelona, Spain

Correspondence to: Emilie Finch, Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom
emilie.finch1@lshtm.ac.uk

Supplementary methods

VacúnateRD vaccination timeline

Supplementary Table 1: Eligibility of age groups for vaccination in the Dominican Republic

Phase	Date	Age group
Fase IC	25/02/2021	> 70
Fase ID	24/04/2021	> 60
Fase II	03/05/2021	> 50
Fase III	10/05/2021	> 18

Model parameters

Supplementary Table 2: Description of model compartments

Compartment	Description
S_i	Number of susceptible individuals in age group i .
V_i	Number of individuals in age group i protected from infection by full vaccination. Note that only susceptible individuals can enter the vaccinated compartment, and so this does not represent individuals who have both post-infection and post-vaccination immunity
L_{ik}	Number of individuals in latent compartment who have been exposed to infection but protected against disease by vaccination for variant k in age group i
E_{ik}	Number of exposed individuals (with a latent infection) with variant k in age group i
I_{sik}	Number of individuals with a sub-clinical (asymptomatic) infection with variant k in age group i
I_{pik}	Number of individuals with a pre-clinical infection with variant k in age group i
I_{cik}	Number of individuals with a clinical (symptomatic) infection with variant k in age group i
R_{ik}	Number of individuals recovered from variant k in age group i

Model equations

$$S_i(t+1) = S_i(t) \cdot (1 - \lambda_i(t)) - v_i(t+1) \cdot \frac{S_i(t)}{N_i(t)} + R_{ik}(t) \cdot w_n + V_i(t) \cdot w_v \quad (1)$$

$$V_i(t+1) = V_i(t) + v_i(t+1) \cdot \frac{S_i(t)}{N_i(t)} - V_i(t) \cdot \lambda_i(t) \cdot (1 - VE_{inf,k}) - V_i(t) \cdot w_v \quad (2)$$

$$L_{ik}(t+1) = L_{ik}(t) \cdot (1 - \sigma) + V_i(t) \cdot \lambda_i(t) \cdot (1 - VE_{inf,k}) \cdot VE_{sd|inf,k}$$

$$E_{ik}(t+1) = E_{ik}(t) \cdot (1 - \sigma) + V_i(t) \cdot \lambda_i(t) \cdot (1 - VE_{inf,k}) \cdot (1 - VE_{sd|inf,k}) \quad (3)$$

$$I_{pik}(t+1) = I_{pik}(t) \cdot (1 - \gamma_p) + E_{ik}(t) \cdot \sigma \cdot y_i \quad (4)$$

$$I_{cik}(t+1) = I_{cik}(t) \cdot (1 - \gamma_c) + I_{pik}(t) \cdot \gamma_p \quad (5)$$

$$I_{sik}(t+1) = I_{sik}(t) \cdot (1 - \gamma_s) + E_{ik}(t) \cdot \sigma \cdot (1 - y_i) + L_{ik}(t) \cdot \sigma \quad (6)$$

$$R_{ik}(t+1) = R_{ik}(t) \cdot (1 - w_n) + I_{sik}(t) \cdot \gamma_s + I_{cik}(t) \cdot \gamma_c \quad (7)$$

Model compartments S_i , V_i , E_{ik} , I_{pik} , I_{cik} , I_{sik} and R_{ik} are defined in Table 1. Parameters are defined as follows: λ_i is the force of infection for age group i and is the sum of variant specific $\lambda_{i,k}$ as defined below; v_i is the number of daily vaccinations for age group i ; σ is $1/dE$ (where dE is the duration in the exposed compartment or the latent period); γ_p is $1/dP$ (where dP is the duration of preclinical

infectiousness); γ_c is $1/dC$ (where dC is the duration of clinical infectiousness); γ_s is $1/dS$ (where dS is the duration of sub-clinical infectiousness). y_i is the probability of clinical symptoms given infection for age group i . $VE_{inf,k}$ is vaccine efficacy against infection with variant k and $VE_{sd|inf,k}$ is vaccine efficacy against disease given infection for variant k .

w_n is the rate of waning of post-infection immunity and w_v is the rate of waning for post-vaccination immunity. Further description of parameter values is given in Table 2.

The force of infection is given by:

$$\lambda_{ik}(t) = u_{ik} \cdot \sum_{j=1}^J C_{i,j,t} \cdot \frac{I_{pj}(t) + I_{cj}(t) + f \cdot I_{sj}(t)}{N_j} \quad (8)$$

Where u_{ik} is susceptibility for age group i and variant k . j also depicts age group with J of 16 and f represents the relative infectiousness of sub-clinical (or asymptomatic) infections, which is 50%.

In this study we model vaccine protection against two outcomes: infection, with effectiveness VE_{inf} ; severe disease given infection $VE_{sd|inf}$. Vaccine efficacy against severe disease given infection is calculated from overall efficacy against infection, VE_{inf} , and severe disease, VE_{sd} , as follows:

$$VE_{sd|inf} = \frac{VE_{sd} - VE_{inf}}{1 - VE_{inf}} \quad (9)$$

Hospitalisation, intensive care unit (ICU) admission and death are modelled as observation processes. Individuals enter observation processes when they mature from the I_p compartment based on associated delays and age-specific probabilities. We use infection-hospitalisation ratios, infection-severe ratios and infection-fatality ratios estimated in the literature and calculate age-specific probabilities of hospitalisation, ICU admission and death given symptomatic infection as follows:

$$P_{hosp,i} = \frac{IHR_i}{y_i} \quad (10)$$

$$P_{ICU,i} = \frac{ISR_i}{y_i} \quad (11)$$

$$P_{death,i} = \frac{IFR_i}{y_i} \quad (12)$$

These are then adjusted on the log odds scale by several fitted parameters to account for changes in the probability of severe outcomes due to improvements in treatment over time and the introduction of new variants, as described in Supplementary Table 2.

The delay between symptom onset and death is assumed to follow a gamma distribution where the mean is estimated during model fitting (and bounded between 5 and 30 days) and the shape parameter is 2.2. The probability density function is bounded at 60 days and given by:

$$Delay_{death} = Gamma(mean = death_{mean}, shape = 2.2) \quad (13)$$

The number of deaths occurring at time t is the sum of deaths occurring over all delays $t - d$ and is given by:

$$D(t) = \sum_{i=1}^{i=I} \sum_{d=1}^{d=D} I_p(t-d)_i \cdot Delay_{death}(d) \cdot P_{death,i,k} \quad (14)$$

Hospitalisation and ICU admissions are estimated using the same approach with shape parameters 0.71 and 1.91 respectively.

$$Delay_{hosp} = Gamma(mean = hosp_{mean}, shape = 0.71) \quad (15)$$

$$Hosp(t) = \sum_{i=1}^{i=I} \sum_{d=1}^{d=D} I_p(t-d)_i \cdot Delay_{hosp} \cdot P_{hosp,i,k} \quad (16)$$

$$Delay_{ICU} = Gamma(mean = icu_{mean}, shape = 1.91) \quad (17)$$

$$ICU(t) = \sum_{i=1}^{i=I} \sum_{d=1}^{d=D} I_p(t-d)_i \cdot Delay_{ICU} \cdot P_{ICU,i,k} \quad (18)$$

The full likelihood is the sum of the likelihoods listed below.

$$Y_{Hosp}(t) \sim NegBin(X_{Hosp}(t), \kappa_{Hosp}) \quad (19)$$

$$L_{Hosp} = \sum_{t=1}^{t=T} P_{NegBin}(Y_{Hosp}(t) | X_{Hosp}(t), \kappa_{Hosp}) \quad (20)$$

$$Y_{ICU}(t) \sim NegBin(X_{ICU}(t), \kappa_{ICU}) \quad (21)$$

$$L_{ICU} = \sum_{t=1}^{t=T} P_{NegBin}(Y_{ICU}(t) | X_{ICU}(t), \kappa_{ICU}) \quad (22)$$

$$Y_{Deaths}(t) \sim NegBin(X_{Deaths}(t), \kappa_{Deaths}) \quad (23)$$

$$L_{Deaths} = \sum_{t=1}^{t=T} P_{NegBin}(Y_{Deaths}(t) | X_{Deaths}(t), \kappa_{Deaths}) \quad (24)$$

$$Y_{Sero}(t) \sim SkewNorm(X_{Sero}(t), \xi_{Sero}, \omega_{Sero}, \alpha_{Sero}) \quad (25)$$

$$L_{Sero} = \sum_{t=1}^{t=T} P_{SkewNorm}(Y_{Sero}(t) | X_{Sero}(t), \xi_{Sero}, \omega_{Sero}, \alpha_{Sero}) \quad (26)$$

We use a logistic-shaped curve parameterized to be a smooth S-shaped function to parameterize parts of the model whereby:

$$asc(x, y_0, y_1, s_0, s_1) = y_0 + (y_1 - y_0) \frac{\text{logistic}(s_0 + x(s_1 - s_0)) - \text{logistic}(s_0)}{\text{logistic}(s_1) - \text{logistic}(s_0)} \quad (27)$$

where,

$$\text{logistic}(x) = \frac{e^x}{1 + e^x} \quad (28)$$

Here, the curve goes from y_0 at $x = 0$ to y_1 at $x = 1$, with an inflection point at $\frac{-s_0}{-s_0 + s_1}$ if $s_0 < 1$.

Appendix C: Supplementary Material Chapter 4

Supplementary Table 3: Fixed model parameters

Parameter	Description	Value	Reference
t_S	Start date of wild-type SARS-CoV-2 epidemic in days after 1 Jan 2020	20	Determines the date which seeding begins (1)
d_E	Latent period (E to Ip and E to Is in days)	$\sim\text{gamma}(\mu = 2.5, k = 2.5)$	2.5 so that the incubation period (latent period plus period of preclinical infectiousness is 5 days). Lauer et al, 2020 (2)
d_P	Duration of pre-clinical infectiousness	$\sim\text{gamma}(\mu = 2.5, k = 4)$	Assumed to be half the duration of total infectiousness in clinically-infected individuals (3)
d_C	Duration of clinical infectiousness	$\sim\text{gamma}(\mu = 2.5, k = 4)$	Infectious period set to 5 days to result in a serial interval of approximately 6 days (4 – 6)
d_S	Duration of subclinical infectiousness	$\sim\text{gamma}(\mu = 5.0, k = 4)$	Assumed to be the same duration as total infectious period for clinical cases, including preclinical transmission
y_i	Probability of clinical symptoms given infection for age group i	Estimated from case distribution across 6 countries	(7)
f	Relative infectiousness of subclinical (asymptomatic) cases	50%	Assumed (7,8)
C_{ij}	Number of age- j individuals contacted by an age- i -individual per day, prior to changes in mobility	Dominican Republic specific contact matrix	(9)
N_i	Number of age- i individuals	From demographic data	
Δt	Time step for discrete-time simulation	0.25 days	
$P_{\text{.death}}$	Infection-fatality ratio by age		Using estimates based on age-specific death data from 45 countries and 22 seroprevalence surveys from O'Driscoll et al (10)
$P_{\text{.hosp}}$	Infection-severe ratio by age (infections resulting in hospitalisation)		Using estimates based on on multi-country serology studies in Herrera-Esposito et al (11)
$P_{\text{.critical}}$	Infection-critical ratio (infections resulting in admission to ICU)		Using estimates based on multi-country serology studies in Herrera-Esposito et al (11)
extra_voc_takeoff	Dates during which transmissibility increases according to a logistic growth function	8th February - 5th April	Informed by GISAID sequence data (12)
los_{hosp}	Length of stay in hospital	$\sim\text{lognormal}(\mu_{log} = 11.08, sd_{log} = 1.20)$	Estimated from UK CO-CIN data (13)
los_{ICU}	Length of stay in ICU	$\sim\text{lognormal}(\mu_{log} = 13.33, sd_{log} = 1.25)$	Estimated from UK CO-CIN data (13)

Supplementary Table 4: Fitted parameters and prior distributions

Parameter	Description	Prior	Note
u	Basic susceptibility to infection	\sim normal(0.09, 0.02) ≥ 0.05 and ≤ 0.2	Determines the basic reproduction number, R_0
death_mean	Mean delay in days from start of infectious period to death	\sim normal(15, 2) ≥ 5 and ≤ 30	Delay is assumed to follow a gamma distribution with shape parameter 2.2. Prior and shape of distribution informed by analysis of CO-CIN data (13)
hosp_admission	Mean delay in days from start of infectious period to hospitalisation	\sim normal(8, 1) ≥ 4 and ≤ 20	Delay is assumed to follow a gamma distribution with shape parameter 0.71. Prior and shape of distribution informed by analysis of CO-CIN data
icu_admission	Mean delay in days from start of infectious period to ICU admission	\sim normal(12.5, 1) ≥ 8 and ≤ 14	Delay is assumed to follow a gamma distribution with shape parameter 1.91. Prior and shape of distribution informed by analysis of CO-CIN data (13)
cfr_rlo cfr_rlo2 cfr_rlo3	Relative log-odds of the case fatality rate (CFR) for COVID-19 for different time periods	\sim normal(0, 0.1) ≥ -2 and ≤ 2	Age-specific case fatality rates based on O'Drisoll et al (10). This was adjusted by cfr_flo, cfr_rlo2 and cfr_rlo3 over time
hosp_rlo hosp_rlo2	Log-odds of hospital admission given infection derived from Herrera-Eposito et al. relative to age-specific probabilities of hospital admission in first half of 2020	\sim normal(0, 0.1) ≥ -2 and ≤ 2	Age-specific probabilities of hospitalisation are based on Herrera-Eposito et al (11) then adjusted based on the icu_rlo and icu_rlo2 parameters. icu_rlo applies for the first half of 2020 while icu_rlo2 applies for the second half of 2020 into 2021.
icu_rlo icu_rlo2	Log-odds of intensive care unit (ICU) admission relative to age-specific probabilities of ICU admission in first half of 2020	\sim normal(0, 0.1) ≥ -2 and ≤ 2	Age-specific probabilities of ICU admission are based on Herrera-Eposito et al (11). These are then adjusted based on the icu_rlo and icu_rlo2 parameters. icu_rlo applies for the first half of 2020 while icu_rlo2 applies for the second half of 2020 into 2021.
disp_deaths disp_hosp_inc disp_hosp_prev disp_icu_prev	Negative binomial dispersion for deaths, hospital incidence (admissions), hospital prevalence (beds occupied), and ICU prevalence	\sim exponential(10)	We estimate the size parameter for negative binomial likelihood functions of deaths, hospital incidence, hospital prevalence and ICU prevalence, where size = $1/(\text{disp}^2)$
contact_adj_a contact_adj_b	Parameter determining the weighting given to comix-adjusted contacts vs baseline contacts from Google mobility data. contact_adj_a applies in 2020 and contact_adj_b applies in 2021	\sim beta(15, 1)	To account for differences in the relationship between Google mobility data and contact patterns between the UK and the Dominican Republic, we allow modelled contacts to be scaled towards pre-pandemic baseline values through fitted parameters contact_adj_a and contact_adj_b.
extra_voc_relu	Relative transmissibility of Mu (and other VOI in mid 2021)	\sim lognormal(0.4, 0.1)	Prior centred around estimated transmission advantage of variants of interest (VOIs) over wild-type (WT) of 1.5
v2_relu	Relative transmissibility of Delta variant	\sim lognormal(0.92, 0.1)	Prior centred around estimated transmission advantage of Delta over WT of 2.5. (14)
v2_when	Date of introduction of Delta variant in days after 1st Jan 2020	\sim uniform(486, 517)	On this date, ten random individuals contract B.1.617.2 (Delta). We use a uniform prior between 1st May 2021 - 1st June 2021 based on sequence data available for the Dominican Republic (see Main Text Figure 3)
v2_hosp_rlo v2_icu_rlo v2_cfr_rlo	Relative log-odds of hospitalisation, Intensive care unit (ICU) admission and death for Delta compared to pre-existing variants	\sim normal(0, 0.1) ≥ -4 and ≤ 4	Vague priors

Supplementary Table 5: Vaccine efficacy parameters

Parameter	Description	Value	Reference
$VE_{inf,WT}$	Overall efficacy against infection with ancestral variant	0.67	Assumed to be the same as efficacy against disease.
$VE_{sd,WT}$	Overall efficacy against severe disease with ancestral variant	0.92	Imai et al (Table 2) (15)
$VE_{inf,delta}$	Overall efficacy against infection with Delta variant	0.39	Assumed to be the same as efficacy against disease.
$VE_{sd,delta}$	Overall efficacy against severe disease with Delta variant	0.61	Wu et al 2022 (Table 3) (16)

Supplementary Table 6: Waning parameters

Parameter	Description	Default value (central waning)
w_n	Rate of waning from the recovered compartment to the susceptible compartment for all strains	$\log(0.85)/-365$, corresponding to exponential waning with a 15% loss of protection after 1 year
w_v	Rate of waning from the vaccinated compartment to the susceptible compartment	$\log(0.6)/-182.5$, corresponding to exponential waning with a 40% loss of protection after 6 months. Based on Cerqueira-Silva et al (17)

Appendix C: Supplementary Material Chapter 4

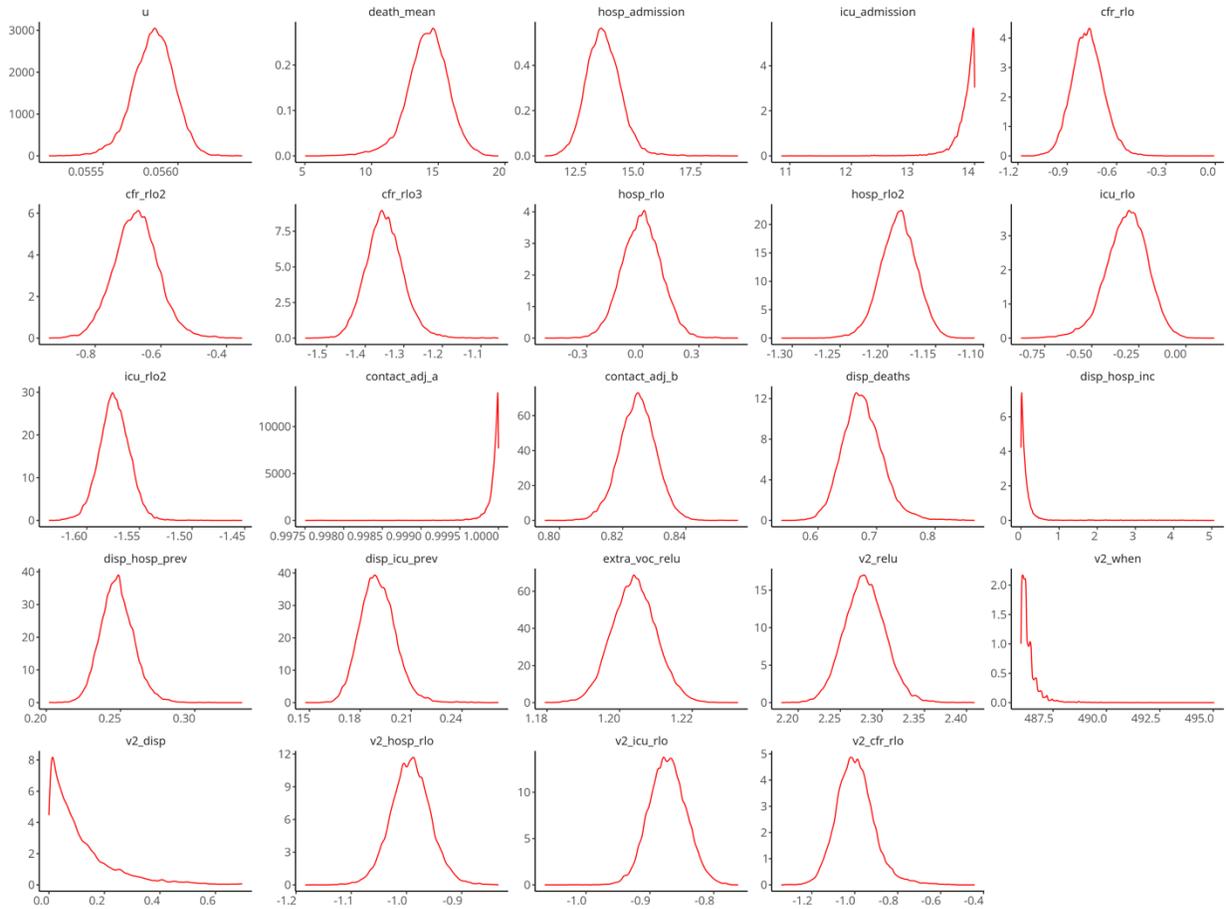
Supplementary Table 7: Vaccine efficacy parameters for two doses of Pfizer and AstraZeneca vaccines against wild-type (WT) and Delta variants

Description	Value
Pfizer overall efficacy against infection with WT	0.85
Pfizer overall efficacy against severe disease with WT	0.95
Pfizer overall efficacy against infection with Delta variant	0.8
Pfizer overall efficacy against severe disease with Delta variant	0.96
AZ overall efficacy against infection with WT	0.75
AZ overall efficacy against severe disease with WT	0.9
AZ overall efficacy against infection with Delta variant	0.63
AZ overall efficacy against severe disease with Delta variant	0.93

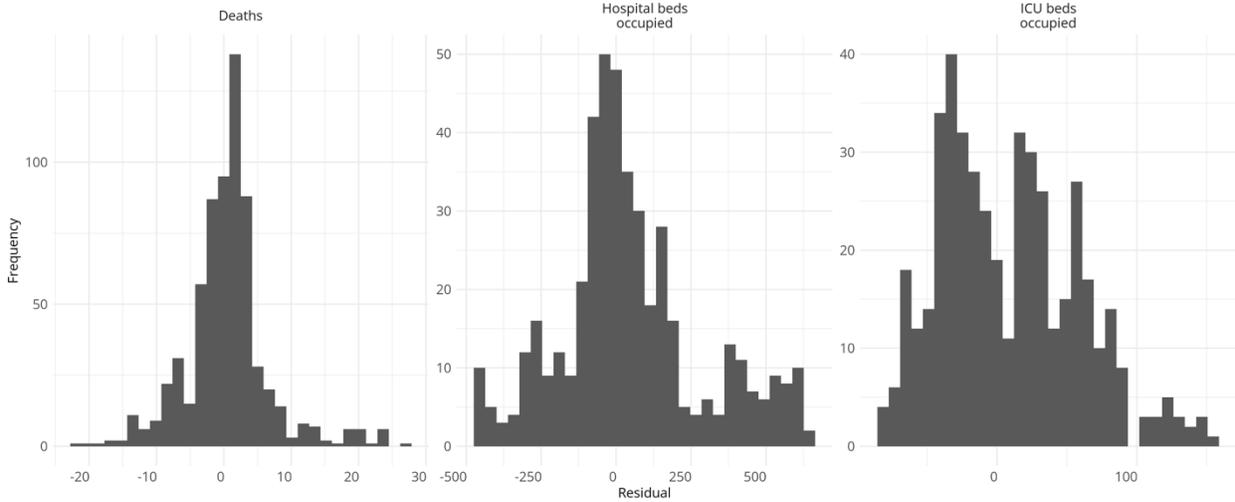
Table adapted from Supplementary Table 6 and 7 in Barnard et al 2022 (14)

Model validation

Supplementary Figure 1: Posterior estimates of fitted parameters



Supplementary Figure 2: Histogram of residuals



Sensitivity analyses

We conducted sensitivity analyses around key vaccine efficacy, waning and transmission parameters used in the main manuscript. In particular we consider:

1. A conservative estimate of vaccine efficacy against disease, using estimated vaccine efficacy against symptomatic disease rather than severe disease (VE_{SD})
2. High waning of post-infection and post-vaccination immunity (w_n and w_v)
3. Low waning of post-infection and post-vaccination immunity (w_n and w_v)
4. High relative infectiousness of subclinical (asymptomatic) individuals (f)

A full description of the parameters considered in the sensitivity analyses is given in Table 8, while estimated hospital admissions, ICU admissions and deaths under different vaccination assumptions and counterfactual scenarios are shown in Table 7.

Appendix C: Supplementary Material Chapter 4

Supplementary Table 8: Sensitivity analysis: estimated deaths, hospital admissions and intensive care unit (ICU) admissions under different waning and vaccine efficacy assumptions

Scenario	Additional hospital admissions in next 6 months	Additional hospital admissions in next 10 months	Additional ICU admissions in next 6 months	Additional ICU admissions in next 10 months	Additional deaths in next 6 months	Additional deaths in next 10 months
No vaccination (low VE against disease)	4870 (4590 – 5240)	5040 (4650 – 5510)	1440 (1360 – 1540)	1480 (1370 – 1610)	524 (466 – 645)	538 (457 – 681)
No vaccination (high waning)	6190 (5900 – 6480)	6700 (6340 – 7070)	1840 (1770 – 1920)	1970 (1870 – 2070)	646 (586 – 724)	685 (611 – 777)
No vaccination (low waning)	7640 (7270 – 8040)	9310 (8820 – 9840)	2320 (2210 – 2430)	2820 (2680 – 2970)	826 (749 – 914)	968 (872 – 1080)
No vaccination (high asymptomatic infectiousness)	6660 (6330 – 6980)	7780 (7370 – 8190)	2020 (1930 – 2110)	2340 (2230 – 2460)	719 (644 – 802)	814 (724 – 912)
Pfizer efficacy or equivalent (low VE against disease)	-1760 (-1900 – -1660)	-3360 (-3780 – -3010)	-518 (-554 – -491)	-997 (-1100 – -905)	-186 (-231 – -165)	-332 (-444 – -229)
Pfizer efficacy or equivalent (high waning)	-469 (-492 – -447)	-1220 (-1430 – -1010)	-124 (-130 – -119)	-342 (-401 – -284)	-41 (-46 – -37)	-111 (-165 – -60)
Pfizer efficacy or equivalent (low waning)	-616 (-649 – -585)	-2730 (-2980 – -2500)	-163 (-171 – -155)	-805 (-869 – -741)	-52 (-57 – -47)	-224 (-275 – -185)
Pfizer efficacy or equivalent (high asymptomatic infectiousness)	-555 (-582 – -528)	-2020 (-2230 – -1820)	-148 (-154 – -141)	-590 (-642 – -540)	-48 (-53 – -43)	-173 (-222 – -128)
Pfizer efficacy or equivalent and delay (low VE against disease)	2980 (2760 – 3300)	160 (-346 – 722)	820 (764 – 891)	-38 (-159 – 93)	287 (229 – 391)	24 (-92 – 223)
Pfizer efficacy or equivalent and delay (high waning)	4080 (3850 – 4310)	1850 (1410 – 2270)	1140 (1080 – 1200)	454 (348 – 559)	387 (336 – 449)	166 (58 – 275)
Pfizer efficacy or equivalent and delay (low waning)	5180 (4890 – 5480)	2140 (1640 – 2660)	1490 (1420 – 1570)	553 (426 – 683)	522 (460 – 592)	269 (156 – 376)
Pfizer efficacy or equivalent and delay (high asymptomatic infectiousness)	4430 (4180 – 4670)	1790 (1350 – 2210)	1270 (1210 – 1330)	445 (342 – 545)	440 (378 – 507)	206 (93 – 314)
AZ efficacy or equivalent (low VE against disease)	-995 (-1070 – -940)	-1890 (-2150 – -1670)	-294 (-314 – -278)	-566 (-632 – -507)	-105 (-130 – -93)	-189 (-259 – -123)
AZ efficacy or equivalent (high waning)	-50 (-62 – -38)	-756 (-910 – -610)	-4 (-7 – -1)	-211 (-254 – -170)	0 (-2 – 3)	-65 (-104 – -28)

Appendix C: Supplementary Material Chapter 4

AZ efficacy or equivalent (low waning)	-118 (-137 – -100)	-1940 (-2130 – -1770)	-19 (-23 – -14)	-575 (-623 – -527)	-1 (-5 – 2)	-151 (-190 – -121)
AZ efficacy or equivalent (high asymptomatic infectiousness)	-96 (-110 – -83)	-1370 (-1510 – -1230)	-15 (-18 – -12)	-399 (-435 – -366)	-2 (-5 – 1)	-110 (-145 – -79)
AZ efficacy or equivalent and delay (low VE against disease)	3200 (2980 – 3530)	1100 (656 – 1620)	893 (836 – 965)	250 (142 – 370)	314 (260 – 418)	117 (17 – 295)
AZ efficacy or equivalent and delay (high waning)	4200 (3960 – 4430)	2150 (1730 – 2550)	1180 (1120 – 1230)	547 (446 – 648)	402 (351 – 463)	198 (96 – 302)
AZ efficacy or equivalent and delay (low waning)	5310 (5020 – 5620)	2580 (2090 – 3080)	1540 (1460 – 1610)	687 (564 – 814)	538 (477 – 609)	309 (201 – 414)
AZ efficacy or equivalent and delay (high asymptomatic infectiousness)	4560 (4310 – 4800)	2190 (1770 – 2600)	1310 (1250 – 1370)	570 (471 – 666)	455 (393 – 523)	245 (137 – 349)

Supplementary Table 9: Sensitivity analysis: description of waning and vaccine efficacy assumptions

Scenario	Description	Value
Central waning (main manuscript results)	Rate of waning from the recovered compartment to the susceptible compartment for all strains	$\log(0.85)/-365$, corresponding to exponential waning with a 15% loss of protection after 1 year
	Rate of waning from the vaccinated compartment to the susceptible compartment	$\log(0.6)/-182.5$, corresponding to exponential waning with a 40% loss of protection after 6 months (17)
High waning	Rate of waning from the recovered compartment to the susceptible compartment for all strains	$\log(0.85)/-182.5$, corresponding to exponential waning with a 15% loss of protection after 6 months
	Rate of waning from the vaccinated compartment to the susceptible compartment	$\log(0.6)/-91.25$, corresponding to exponential waning with a 40% loss of protection after 3 months
Low waning	Rate of waning from the recovered compartment to the susceptible compartment for all strains	$\log(0.85)/-730$, corresponding to exponential waning with a 15% loss of protection after 2 years
	Rate of waning from the vaccinated compartment to the susceptible compartment	$\log(0.84)/-120$, corresponding to exponential waning with a 16% loss of protection after 4 months.
Low vaccine efficacy against disease	Sinovac-Coronavac	0.67 against disease for wild-type (WT) and 0.39 efficacy against disease for Delta (14)
	Pfizer-like efficacy	0.9 efficacy against disease for WT and 0.81 efficacy against disease for Delta (14)
	AZ-like efficacy	0.8 efficacy against disease for WT and 0.65 efficacy against disease for Delta (14)
High asymptomatic infectiousness	Higher relative infectiousness of asymptomatic cases	0.7 (assumed)

Supplementary Results

Main results broken down by age

Supplementary Figure 3: Modelled hospitalisations, deaths, and proportion previously infected by age group

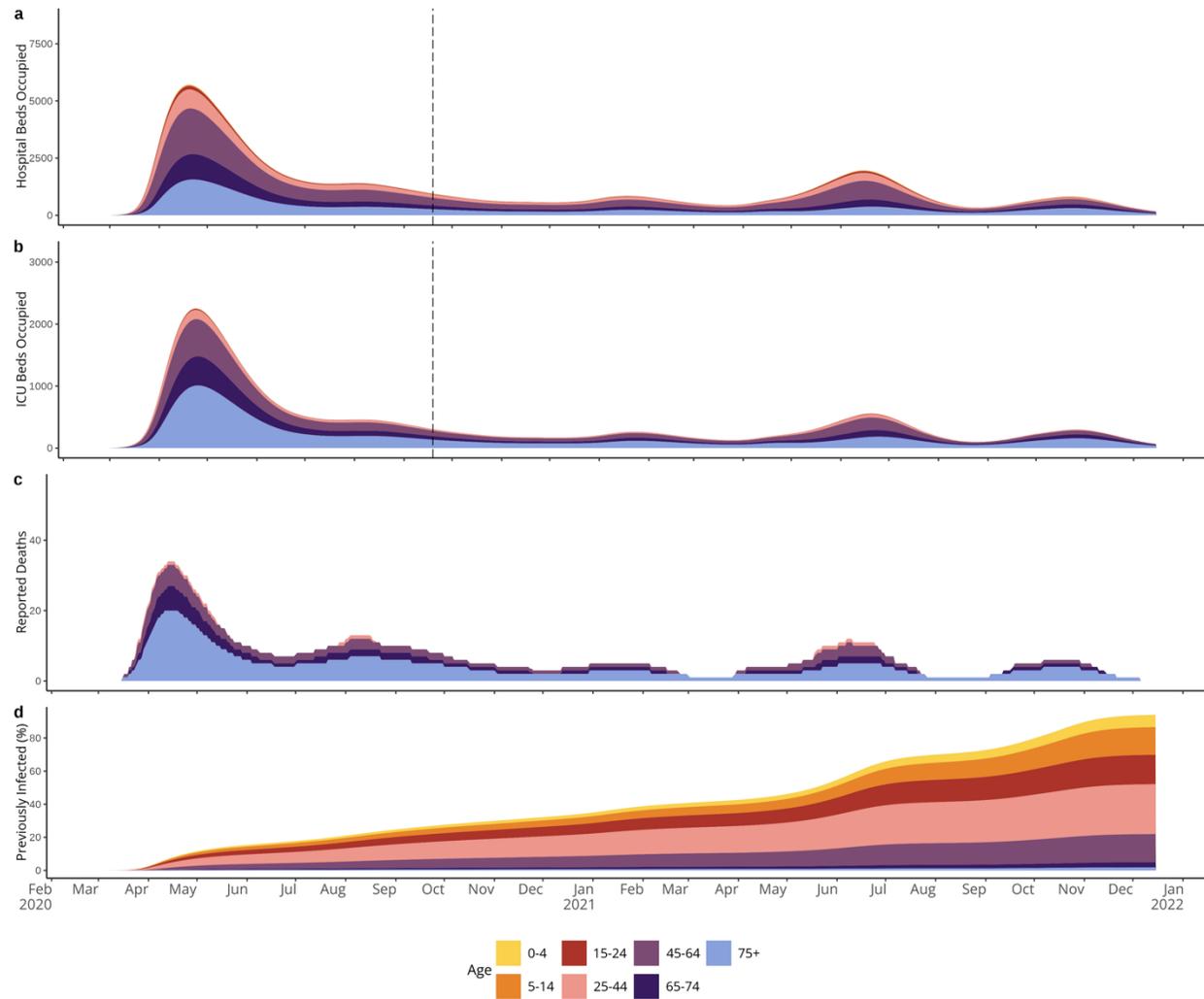


Figure shows median estimates of modelled hospital bed occupancy (a), ICU bed occupancy (b), reported deaths (c) and the proportion of the population previously infected (d). Shaded areas show the burden broken down by age groups. Vertical dashed lines in panels (a) and (b) indicate the date at which hospitalisation data became available.

Supplementary Figure 4: Modelled immune status of the population by age group

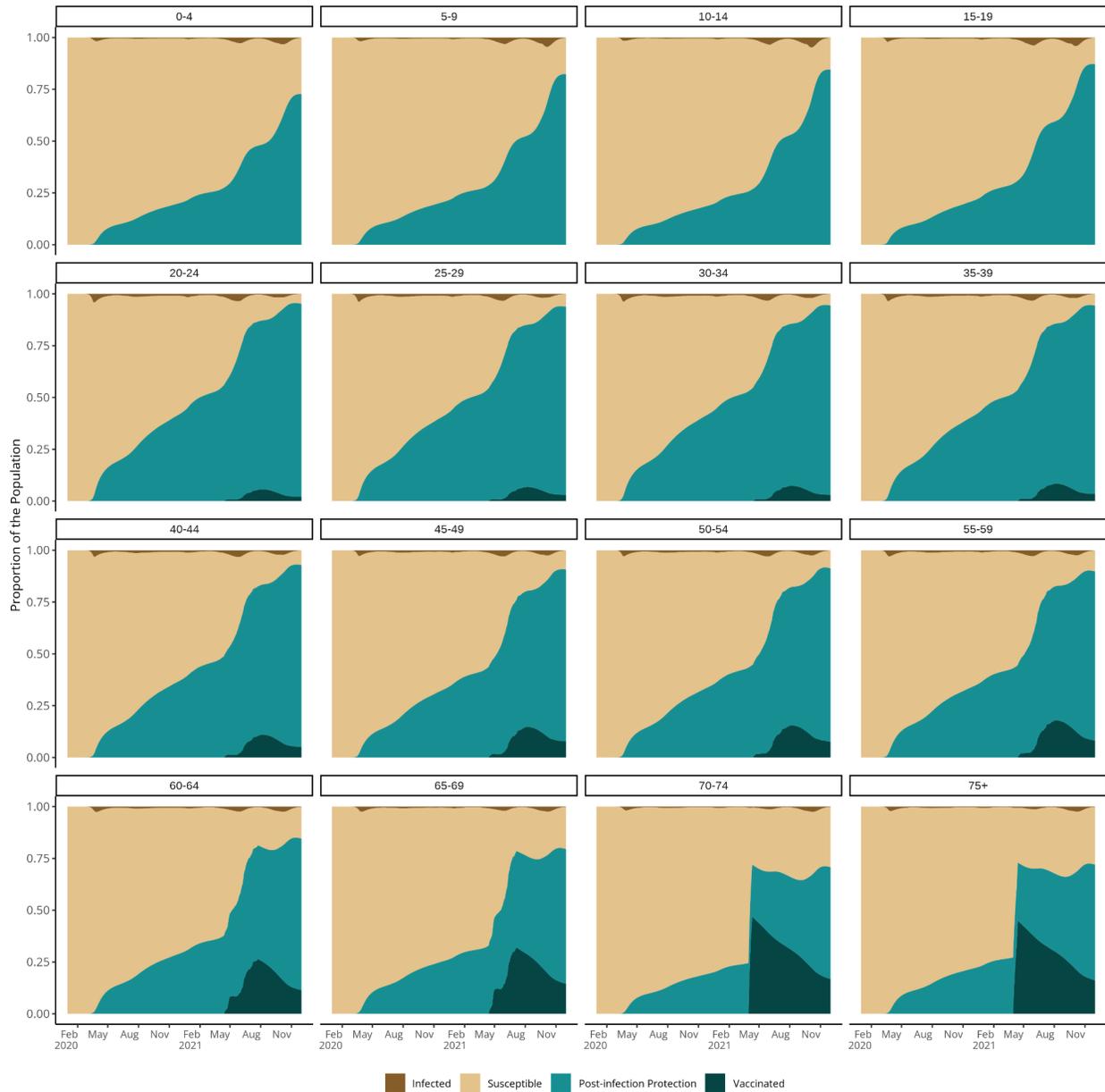


Figure showing the modelled distribution of immune states in the Dominican Republic over time, faceted by age group. The shaded areas show the proportion of the population that are; currently infected (brown), susceptible (beige), protected post-infection (blue) and protected post-vaccination (dark blue). Note that the vaccinated area (dark blue) does not include individuals that were vaccinated post-infection and so does not correspond with observed vaccination coverage.

Supplementary Figure 5: Impact of vaccination with alternative vaccine products with and without delay

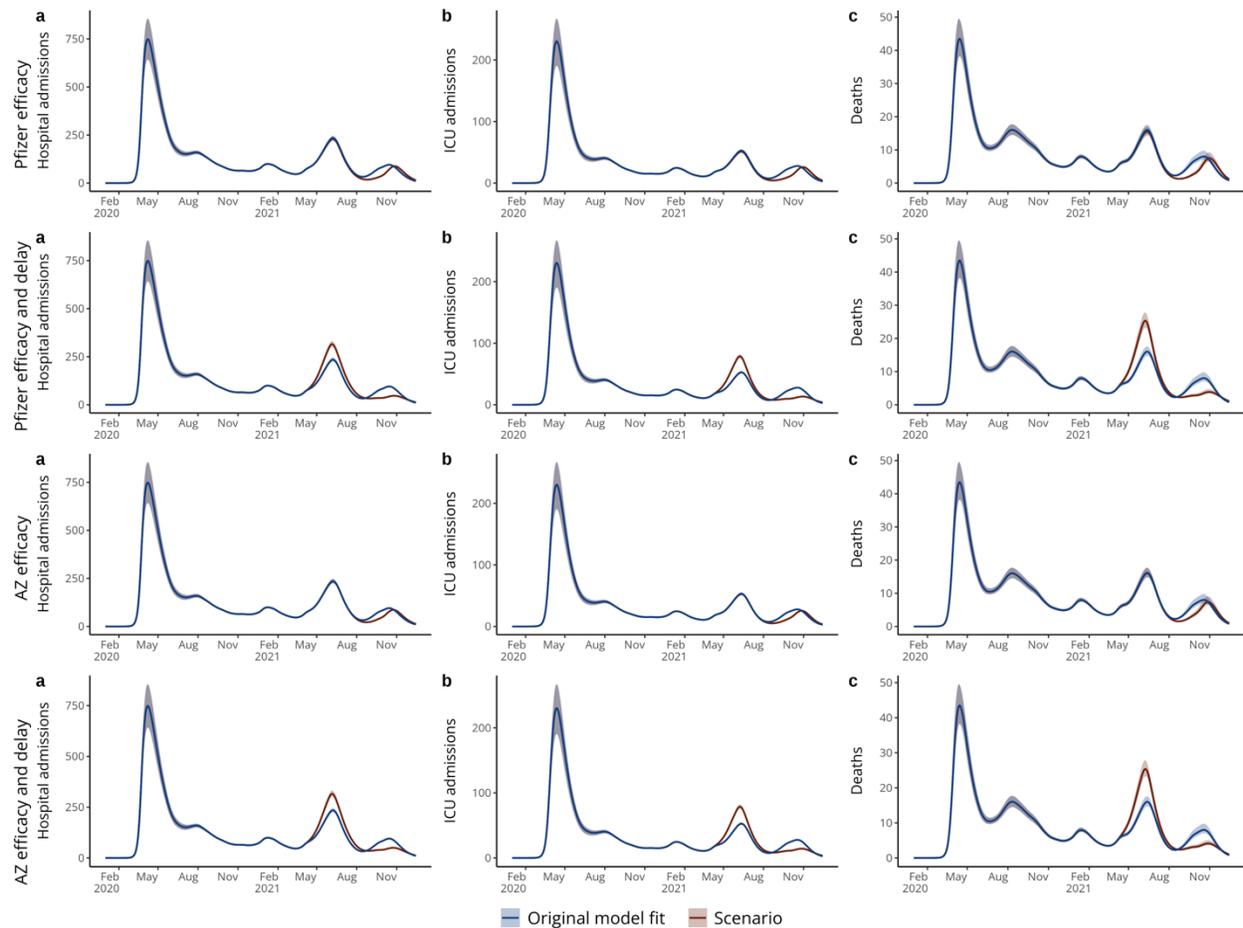


Figure showing modelled deaths (a), hospital admissions (b) and intensive care unit (ICU) admissions (c) from the original model fit (in blue) and from counterfactual scenarios (in red) with: vaccination with a Pfizer/BioNTech efficacy profile; vaccination with a Pfizer/BioNTech efficacy profile and a delay in vaccination of 2 months; vaccination with an Oxford/AstraZeneca efficacy profile; and vaccination with an Oxford/AstraZeneca efficacy profile with a delay of 2 months. Pfizer/BioNTech and Oxford/AstraZeneca efficacy profiles are labelled as Pfizer and AZ respectively. Lines show the median value from 500 simulations with associated 95% credible intervals in surrounding ribbons. To facilitate comparison between scenarios, modelled deaths do not include uncertainty generated through the observation process and are therefore higher than those shown in the model fit in Figure 4 of the main text.

References

1. Paulino-Ramirez R, Báez AA, Degaudenzi AV, Tapia L. Seroprevalence of Specific Antibodies against SARS-CoV-2 from Hotspot Communities in the Dominican Republic. *Am J Trop Med Hyg.* 2020 Oct 21;103(6):2343 – 6.
2. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med.* 2020 Mar 10;M20-0504.
3. Ferretti L, Ledda A, Wymant C, Zhao L, Ledda V, Abeler-Dörner L, et al. The timing of COVID-19 transmission [Internet]. medRxiv; 2020 [cited 2022 Nov 16]. p. 2020.09.04.20188516. Available from: <https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v2>
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus – Infected Pneumonia. *N Engl J Med.* 2020 Mar 26;382(13):1199 – 207.
5. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis.* 2020 Apr 1;93:284 – 6.
6. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis.* 2020 Aug 1;20(8):911 – 9.
7. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med.* 2020 Aug;26(8):1205 – 11.
8. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health.* 2020 Jul 1;5(7):e375 – 85.
9. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Comput Biol.* 2017 Sep 12;13(9):e1005697.
10. O’Driscoll M, Dos Santos GR, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature.* 2020 Nov 2;1 – 6.
11. Herrera-Esposito D, de los Campos G. Age-specific rate of severe and critical SARS-CoV-2 infections estimated with multi-country seroprevalence studies. *BMC Infect Dis.* 2022 Mar 29;22(1):311.
12. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance.* 2017 Mar 30;22(13):30494.
13. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ.* 2020 May 22;369:m1985.

14. Barnard RC, Davies NG, Jit M, Edmunds WJ. Modelling the medium-term dynamics of SARS-CoV-2 transmission in England in the Omicron era. *Nat Commun.* 2022 Aug 19;13(1):4879.
15. Imai N, Hogan AB, Williams L, Cori A, Mangal TD, Winskill P, et al. Interpreting estimates of coronavirus disease 2019 (COVID-19) vaccine efficacy and effectiveness to inform simulation studies of vaccine impact: a systematic review [Internet]. Wellcome Open Research; 2021 [cited 2022 Jan 13]. Available from: <https://wellcomeopenresearch.org/articles/6-185>
16. Wu D, Zhang Y, Tang L, Wang F, Ye Y, Ma C, et al. Effectiveness of Inactivated COVID-19 Vaccines Against Symptomatic, Pneumonia, and Severe Disease Caused by the Delta Variant: Real World Study and Evidence - China, 2021. *China CDC Wkly.* 2022 Jan 28;4(4):57 – 65.
17. Cerqueira-Silva T, Katikireddi SV, de Araujo Oliveira V, Flores-Ortiz R, Júnior JB, Paixão ES, et al. Vaccine effectiveness of heterologous CoronaVac plus BNT162b2 in Brazil. *Nat Med.* 2022 Apr;28(4):838 – 43.

Appendix D

Supplementary material for Chapter 5

Supplementary information for:

Climate variation and serotype competition drive dengue outbreak dynamics in
Singapore

Emilie Finch¹, Adam Kucharski¹, Shuzhen Sim², Lee-Ching Ng², Rachel Lowe^{1,3,4}

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

²National Environment Agency, Singapore

³Barcelona Supercomputing Center (BSC), Barcelona, Spain

⁴Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

Corresponding author: Emilie Finch (emilie.finch1@lshtm.ac.uk) and Rachel Lowe (rachel.lowe@lshtm.ac.uk)

Model priors and hyperparameter formulations

The weekly random effect $\delta_{w[t]}$ was assigned a random walk 2 prior distribution (second order difference prior distribution) where we assume that second order increments follow a Gaussian distribution with zero mean and precision τ . This was defined to be cyclic (so the first week of the year is dependent on the last two weeks of the previous year). We included independent and identically distributed (iid) random effects for each year $\gamma_{a[t]}$. Non-linear covariates were modelled by setting a random walk 2 prior on the coefficients of the covariates. We used penalized complexity priors (PC priors) for precision τ , with hyperparameters $\sigma_0 = 0.5$ and $\alpha = 0.01$ for all non-linear covariates and random effects in the model. The PC prior is defined on the standard deviation $\sigma = \tau^{-1/2}$ such that $P(\sigma > \sigma_0) = \alpha$. these penalize departure from $\sigma = 0$. PC priors follow the principle of parsimony, favouring a base model (where $\sigma = 0$) unless evidence is provided against it (1).

Table 1: Covariates tested in model selection

Variable name	Covariate class	Variable type
Minimum temperature °C	Temperature	Numeric
Mean temperature °C	Temperature	Numeric
Maximum temperature °C	Temperature	Numeric
Absolute humidity g/m ³	Humidity	Numeric
Relative humidity %	Humidity	Numeric
Total precipitation (mm)	Rainfall	Numeric
Number of days without rain	Rainfall	Numeric or categorical
Number of days with heavy rain	Rainfall	Numeric or categorical
Number of days with moderate or heavy rain	Rainfall	Numeric or categorical
Number of days with consecutive rainfall	Rainfall	Numeric or categorical
Niño 3.4 sea surface temperature anomalies (SSTA)	ENSO	Numeric
Serotype proportions	Serotype	Numeric
Serotype growth rates	Serotype	Numeric
Dominant serotype	Serotype	Factor
Switch in dominant serotype	Serotype	Binary
Time since switch in dominant serotype	Serotype	Numeric

Table 2: Details of model selection

We conducted forwards stepwise selection, grouping climatic indicators into classes of covariate including: temperature, rainfall, humidity and Niño 3.4. At each step of model selection, the best performing variable was carried forwards and tested against all variables in remaining climate classes. Note that at the third stage of model selection the 12 week total days without rain performed similarly well to 12 week total precipitation (mm); we selected the former due to the high influence of outlier precipitation values on the estimated effect size. Finally, we did not include absolute humidity as there was evidence for no effect on dengue incidence and only marginal improvements in model adequacy criteria.

Step	Variable	WAIC	DIC	Rsqr
1	Maximum temperature (12 week running average, non-linear, 0 lag)	12830.96	12829.96	0.067
2	Maximum temperature (12 week rolling average, 0 lag) + Niño 3.4 (12 week rolling average, non-linear, 1 month lag)	12786.84	12785.10	0.103
3	Maximum temperature (12 week rolling average, 0 lag) + Niño 3.4 (12 week rolling average, non-linear, 1 month lag) + 12 week total days without rain (0 lag)	12756.24	12754.75	0.139
4	Maximum temperature (12 week rolling average, non-linear, 0 lag) + Niño 3.4 (12 week rolling average, non-linear, 1 month lag) + 12 week total days without rain (non-linear, 0 lag) + absolute humidity (4 month lag)	12748.57	12746.83	0.146
Adding serotype	Maximum temperature (12 week rolling average, non-linear, 0 lag) + Niño 3.4 (12 week rolling average, non-linear, 1 month lag) + 12 week total days without rain (non-linear, 0 lag) + time since switch in dominant serotype (non-linear)	12498.13	12498.27	0.331

Table 3: Details of selected models

Full model formulae for full *climate and serotype* model and other models compared in the main text.

Final models	Formula
<i>Climate and serotype</i>	Time since switch in dominant serotype (non-linear) + maximum temperature °C (12 week average, non-linear) + days without rain (12 week total, non-linear) + Niño 3.4 SSTA (12 week average with a 4 week lag) + $\gamma_{a[t]}$ + $\delta_{w[t]}$
<i>Climate only</i>	Maximum temperature °C (12 week average, non-linear) + days without rain (12 week total, non-linear) + Niño 3.4 SSTA (12 week average with a 4 week lag) + $\gamma_{a[t]}$ + $\delta_{w[t]}$
<i>Serotype only</i>	Time since switch in dominant serotype (non-linear) + $\gamma_{a[t]}$ + $\delta_{w[t]}$
<i>Seasonal baseline</i>	$\delta_{w[t]}$

Table 4: Forecast metrics over different forecast horizons

Forecast skill metrics for each candidate model at forecast horizons from 0-8 weeks. Metrics include: CRPS (lower scores are better) CRPSS (higher scores are better), Brier score (lower scores are better), AUC (higher scores are better), false alarm rate (lower scores are better) and hit rate (higher scores are better). The trigger threshold which maximises the AUC for each model is also shown. The best score for each forecast metric and horizon is shown in bold.

Horizon	Model	CRPS	CRPSS	Brier score	AUC	False alarm (%)	Hit rate (%)	Trigger threshold (%)
0	<i>Climate and serotype</i>	50	59.5	0.0544	98.4 (95% CI: 97.69-99.03)	2.08	91.6	71.4
0	<i>Serotype only</i>	63.2	48.9	0.062	97.8 (95% CI: 97-98.65)	2.54	90.7	65.4
0	<i>Climate only</i>	57.5	53.5	0.0655	97.8 (95% CI: 96.95-98.6)	3.92	93.4	54.7
0	<i>Seasonal baseline</i>	124	0	0.232	73.4 (95% CI: 69.8-77.09)	11.8	65.2	37.8
2	<i>Climate and serotype</i>	56.4	54.9	0.0623	97.8 (95% CI: 96.96-98.58)	3.46	91	64
2	<i>Serotype only</i>	70	43.9	0.0713	97 (95% CI: 95.99-98.03)	2.77	88.9	65.6

Appendix D: Supplementary Material Chapter 5

2	<i>Climate only</i>	62.3	50.1	0.0722	97 (95% CI: 96.05-98.03)	4.38	92.5	49.6
2	<i>Seasonal baseline</i>	125	0	0.235	71 (95% CI: 67.22-74.79)	14.6	66.7	37
4	<i>Climate and serotype</i>	67.9	46.3	0.0773	96.5 (95% CI: 95.44-97.62)	5.54	90.4	49.2
4	<i>Serotype only</i>	82.4	34.9	0.0894	94.9 (95% CI: 93.45-96.4)	5.19	87.1	50.6
4	<i>Climate only</i>	71.7	43.4	0.0863	95.4 (95% CI: 94.11-96.78)	5.07	89.5	49.2
4	<i>Seasonal baseline</i>	127	0	0.24	68 (95% CI: 64.11-71.94)	15.6	64.3	37
6	<i>Climate and serotype</i>	86.5	32.4	0.0889	95.4 (95% CI: 94.1-96.74)	6.23	89.5	45.4
6	<i>Serotype only</i>	104	18.8	0.105	93.1 (95% CI: 91.31-94.86)	5.77	83.2	54.6
6	<i>Climate only</i>	86.2	32.7	0.0982	94 (95% CI: 92.37-95.57)	4.38	83.5	52.5
6	<i>Seasonal baseline</i>	128	0	0.243	65.3 (95% CI: 61.35-69.35)	17.1	64.6	36.2
8	<i>Climate and serotype</i>	86.5	32.4	0.098	94.2 (95% CI: 92.65-95.75)	6.92	90.7	41.7
8	<i>Serotype only</i>	104	18.8	0.123	90.9 (95% CI: 88.79-92.98)	7.38	81.4	51.5
8	<i>Climate only</i>	86.2	32.7	0.115	92.1 (95% CI: 90.24-93.92)	6	80.5	51.6
8	<i>Seasonal baseline</i>	128	0	0.245	64.3 (95% CI: 60.27-68.33)	14.2	56.5	37.8

CRPS: continuous ranked probability score; CRPSS: continuous ranked probability skill score; AUC: area under the curve; CI: confidence interval

Table 5: Variables used for prediction at each time horizon

Forecast horizon	Temperature	Precipitation	ENSO	Serotype
0	12 week average maximum temperature °C, 0 week lag	12 week total days with no rain, 0 week lag	12 week rolling average Niño 3.4 SSTA, 4 week lag	Time since serotype switch, 0 week lag
2	10 week average maximum temperature °C, 2 week lag	10 week total days with no rain, 2 week lag, scaled up by 1.2	12 week rolling average Niño 3.4 SSTA, 4 week lag	Time since serotype switch, 2 week lag + 2
4	8 week average maximum temperature °C, 4 week lag	8 week total days with no rain, 4 week lag, scaled up by 1.5	12 week rolling average Niño 3.4 SSTA, 4 week lag	Time since serotype switch, 4 week lag + 4
6	6 week average maximum temperature °C, 6 week lag	6 week total days with no rain, 6 week lag, scaled up by 2	10 week rolling average Niño 3.4 SSTA, 6 week lag	Time since serotype switch, 6 week lag + 6
8	4 week average maximum temperature °C, 8 week lag	4 week total days with no rain, 8 week lag, scaled up by 3	8 week rolling average Niño 3.4 SSTA, 8 week lag	Time since serotype switch, 8 week lag + 8

Figure 1: Dengue cases and outbreak threshold

Figure showing weekly reported dengue cases from 1 January 2000 – 31 December 2022 (grey bars) and seasonal moving 75th percentile outbreak threshold (green line) and the endemic channel threshold used operationally by the NEA (brown line). For a given month and year, we defined a seasonal moving 75th percentile outbreak threshold using the 75th percentile of weekly cases in that month using all years up to, but not including, the current year. The NEA uses an endemic channel threshold which is calculated as two standard deviations in excess of mean cases over the past 5 years, with outliers removed. Outliers are defined as any weekly cases greater than the threshold for that year.

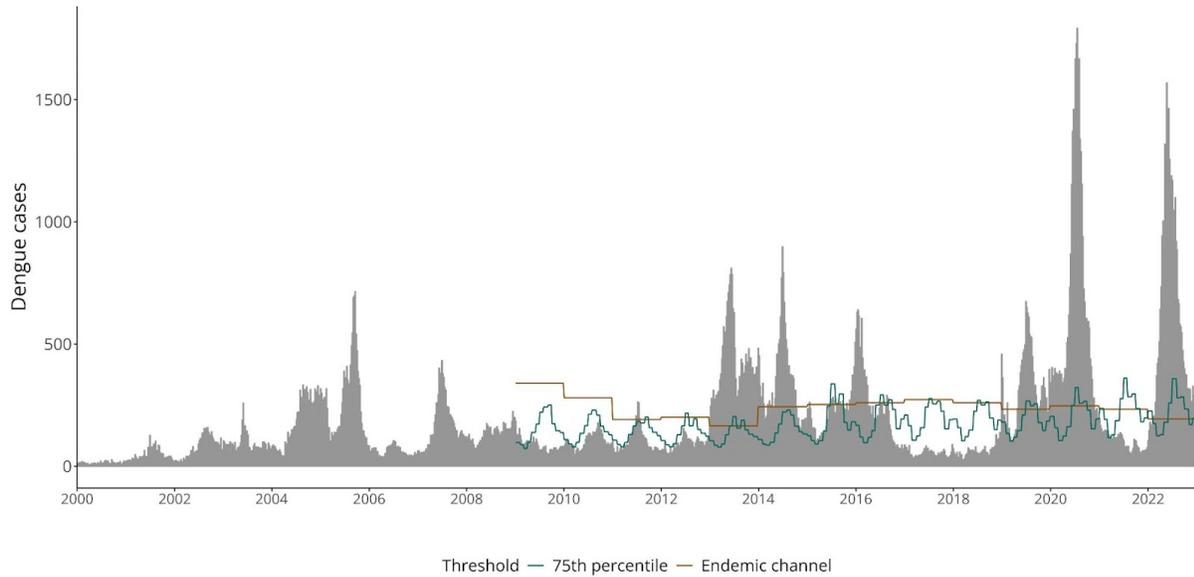


Figure 2: Dengue forecasts for early warning at 2-8 week forecast horizons

Figure showing time series of cross-validated posterior predictions of dengue cases for each model from 2009 – 2022 at 2 - 8 week forecast horizons. Columns for 4 weeks ahead and 8 weeks ahead are also shown in Figure 5 of the main text. Coloured lines show the median posterior prediction of weekly dengue cases, shaded areas show the 95% credible interval and the dark grey line shows the data. From top to bottom the figure shows: predictions for the final selected ‘*Climate and serotype*’ model weekly and yearly random effects $a[t]+w[t]$ in purple; predictions for a ‘*Climate only*’ model with random effects in pink; predictions for a ‘*Serotype only*’ model with random effects in green; and predictions from a ‘*Seasonal baseline*’ model with only weekly random effects $w[t]$ in orange. From left to right each column shows forecasts at 2, 4, 6 and 8 weeks ahead respectively.

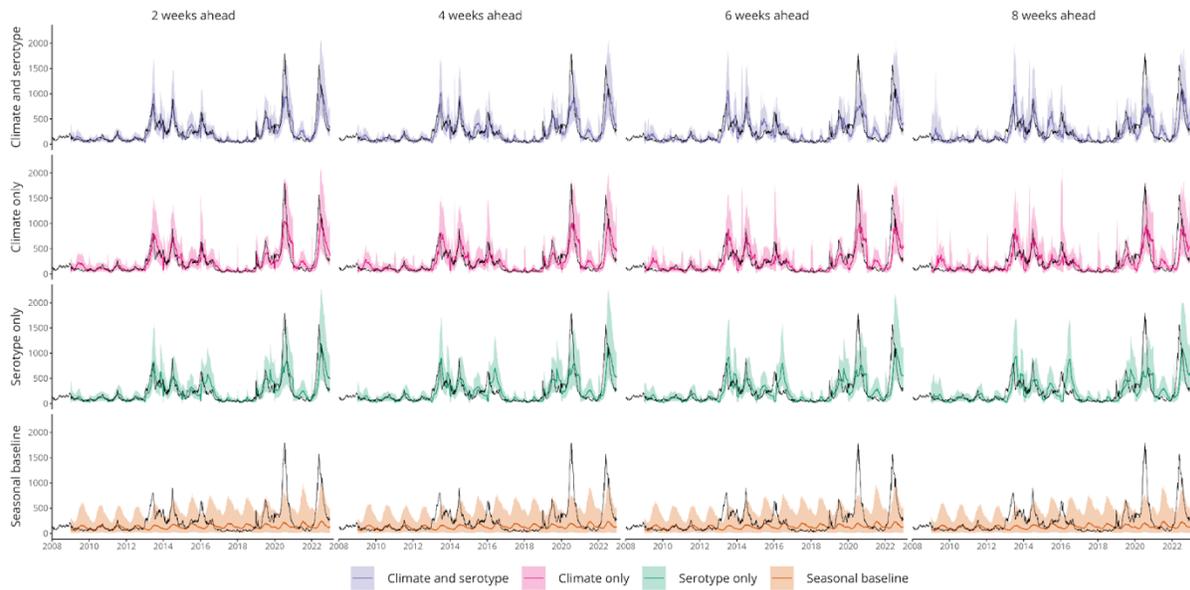


Figure 3: Schematic of time series cross validation design to generate forecasts

Schematic showing how h week ahead forecasts are generated using time series cross-validation. Blue blocks represent the training data, orange blocks time points to be forecasted and grey points represent data not included in the forecast generation. Columns show the training/testing design for 2-step ahead and 3-step ahead forecasts respectively, while rows show different example municipalities. We use an expanding window approach where, for each forecast horizon h , we train the model on data available until week $t - h$. We then predict dengue incidence in week t , using lagged covariates. When conducting model evaluation for week t and forecast horizon h , we used the same expanding window time series cross-validation approach (Methods). We trained the model using the final selected covariates on data available until week $t-h$. We then predicted dengue incidence in week t using the best approximation of each climatic covariate. For example, considering only the temperature covariate for simplicity, when predicting dengue cases with a 4 week ahead time horizon, we fit the model up until week $t-h$ using a 12 week running average temperature to estimate model parameters. Then, using these estimated model parameters, we predict dengue cases at week t by inputting 8 week running average temperature with a 4 week lag (alongside other lagged covariates). This allows us to preserve the key relationships between climate and serotype covariates, and dengue cases that we estimate in full model fitting and then use the best climate data available at different lead times to generate forecasts for early warning.

Note that this differs from the design used for model evaluation (shown in Figure 3 of the main text), where data until $t-1$ is used to train the model. Climatic data up to time t is then used to generate predictions for dengue incidence at time t . Serotype covariates are constructed using data until $t-1$ as serotype frequencies are dependent on case counts.

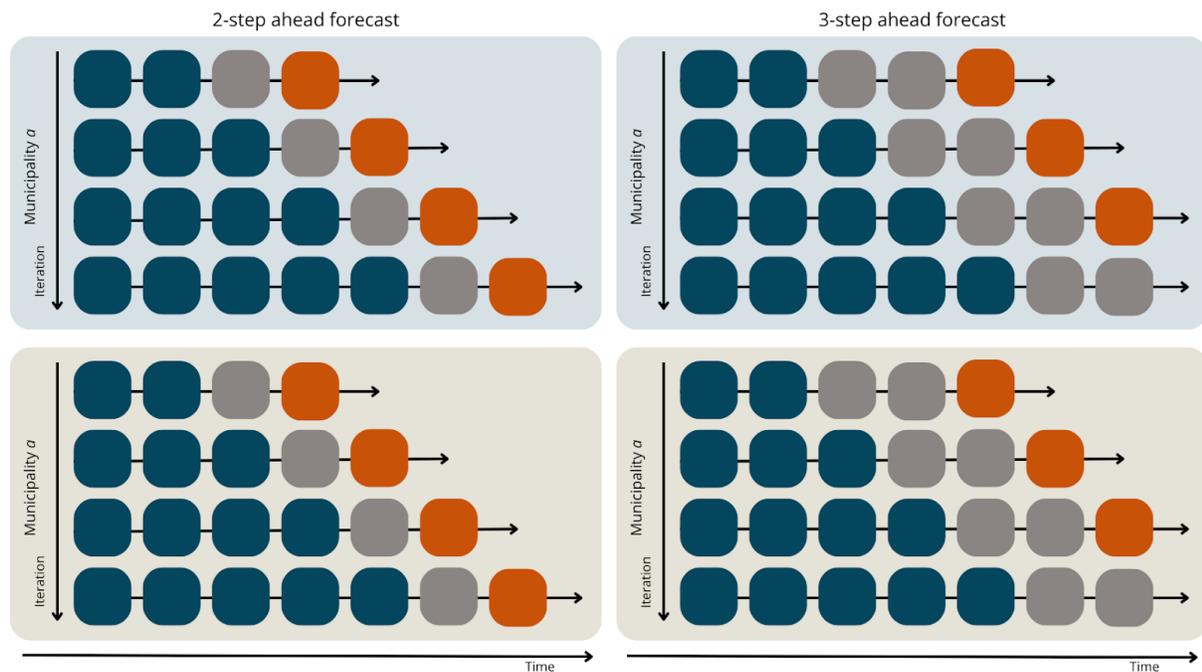


Figure 4: Dengue forecasts for early warning at 2-8 week forecast horizons showing covariate models without yearly random effects

Figure showing time series of cross-validated posterior predictions of dengue cases for each model from 2009 – 2022 at 2 - 8 week forecast horizons. Here, unlike Figure 5 in the main text, the *Climate and serotype*, *Climate-only* and *Serotype-only* models do not include a yearly random effect $a[t]$. Coloured lines show the median posterior prediction of weekly dengue cases, shaded areas show the 95% credible interval and the dark grey line shows the data. From top to bottom the figure shows: predictions for the final selected ‘*Climate and serotype*’ model with weekly random effects $w[t]$ in purple; predictions for a ‘*Climate only*’ model with weekly random effects in pink; predictions for a ‘*Serotype only*’ model with weekly random effects in green; and predictions from a ‘*Seasonal baseline*’ model with only weekly random effects in orange. From left to right each column shows forecasts at 2, 4, 6 and 8 weeks ahead respectively.

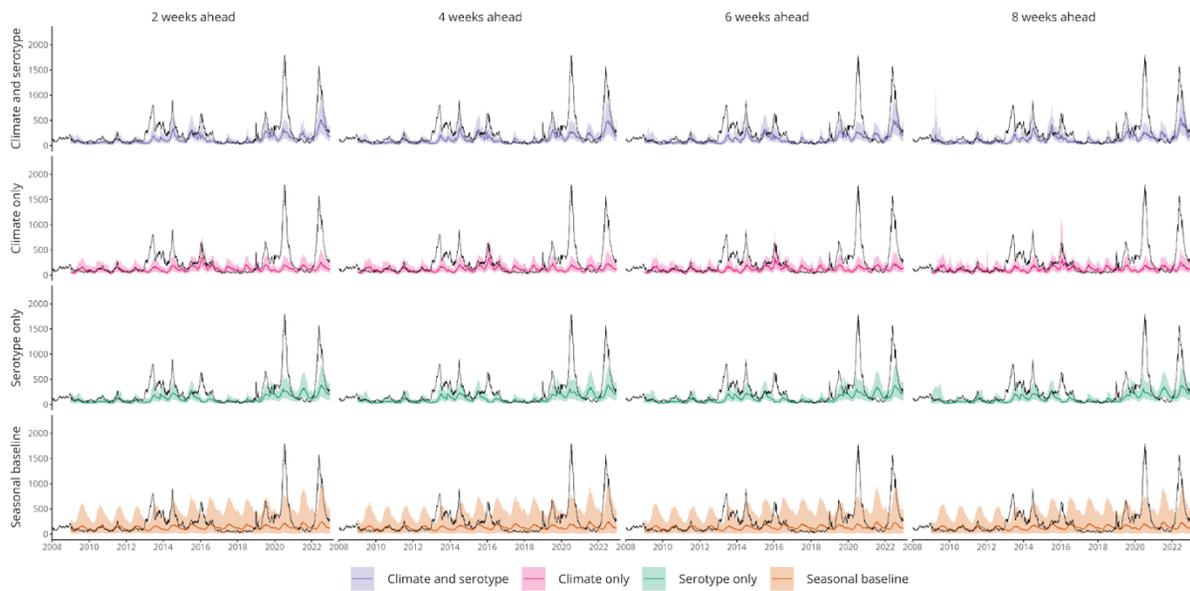
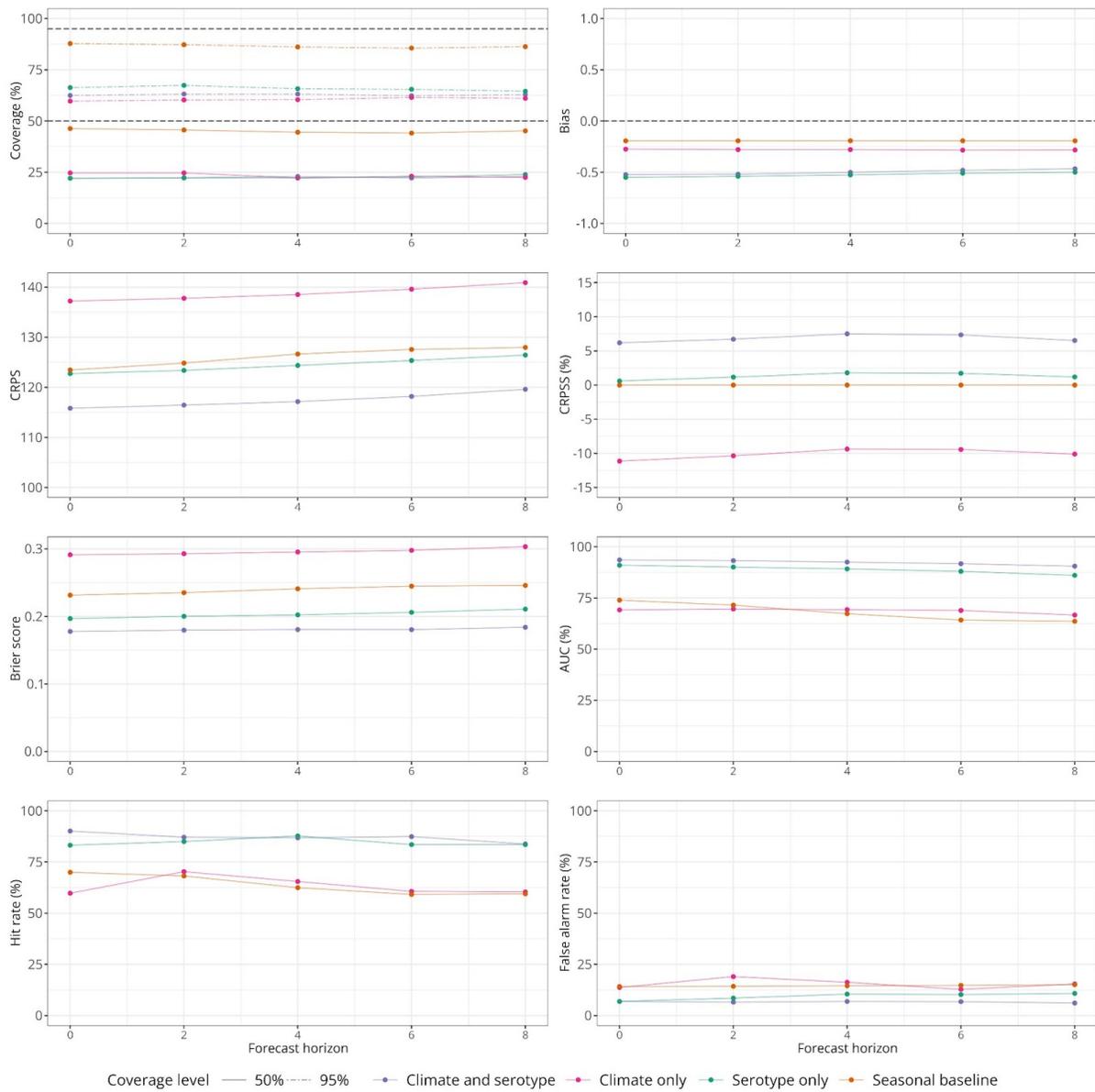


Figure 5: Predictive performance over different forecast horizons for covariate models without yearly random effects.

Figure showing forecast metrics for each model across all forecast horizons from 2009 – 2022. Unlike Figure 6 in the main text, here covariate models do not include a yearly random effect $a[t]$. Figure showing forecast metrics for each model across all forecast horizons from 2009 – 2022. From top left to bottom right these show: interval coverage %, bias; CRPS (continuous ranked probability score), CRPSS (continuous ranked probability skill score, %), Brier score, AUC (area under the curve, %), hit rate (%) and false alarm rate (%). Interval coverage shows the percentage of observations falling inside a given prediction interval. A perfectly calibrated forecast would have coverage equal to the nominal prediction interval; that is, 95% coverage equal to 95% and 50% coverage equal to 50%, indicated by dashed horizontal lines. Bias measures the relative tendency of the model to over- or under-predict, and is bounded between -1 and 1, with 0 indicating unbiased forecasts. The CRPS can take values between 0 and infinity, with smaller values indicating better performance. The CRPSS indicates the relative improvement of each covariate model over the *seasonal baseline* model and can take values from 0%, indicating that the model performs the same as the baseline, and 100%, indicating perfect forecasting skill. The Brier score can take values from 0 – 1, with smaller values indicating better performance. The AUC can take values from 0-100% with 100% indicating perfect classification. Hit rate and false alarm rate also take values from 0 – 100% with higher and lower values indicating better performance respectively.



References

1. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Stat Sci.* 2017 Feb;32(1):1–28.

Appendix E

Supplementary material for Chapter 6

Supplementary Materials for:

Understanding the role of climatic and epidemic drivers in spatiotemporal dengue outbreak dynamics in the Dominican Republic

Figure 1: Mean temperature in the Dominican Republic from 2013 - 2023

Tile plot showing average monthly mean temperature °C with months running from April - March on the x-axis and dengue season on the y-axis. Darker red tiles indicate higher average temperatures. Provinces are arranged to mirror their geographic location in the Dominican Republic.

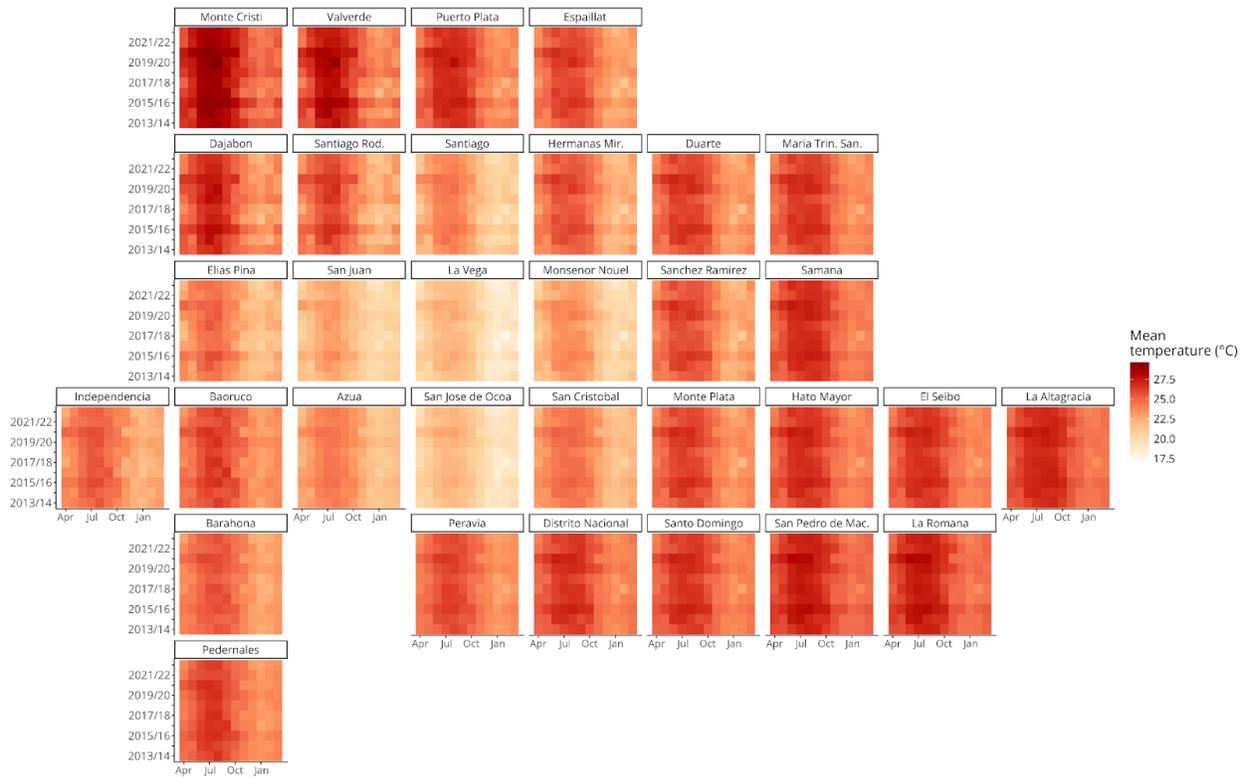


Figure 2: Cumulative precipitation in the Dominican Republic from 2013 - 2023

Tile plot showing average monthly cumulative precipitation in mm with months running from April - March on the x-axis and dengue season on the y-axis. Darker blue tiles indicate higher precipitation levels. Provinces are arranged to mirror their geographic location in the Dominican Republic.

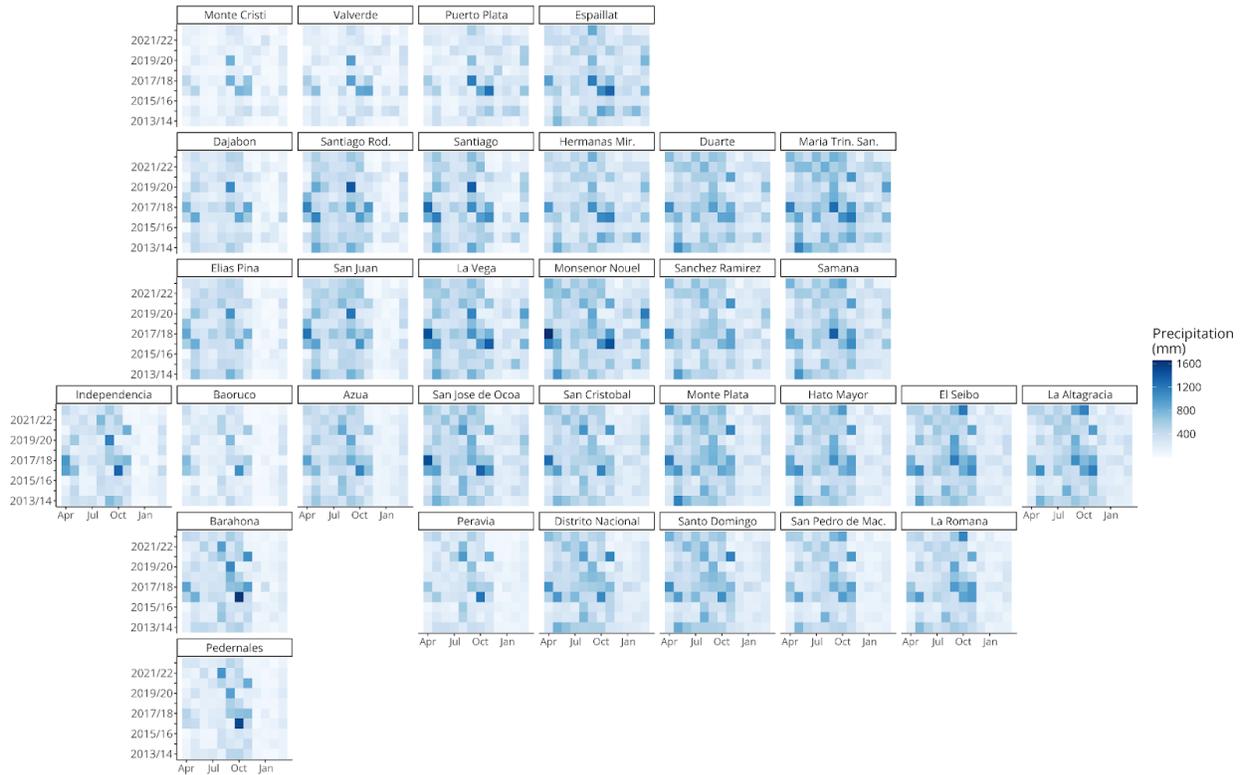


Figure 3: Average relative humidity (%) in the Dominican Republic 2013 - 2023.

Tile plot showing average monthly relative humidity (%) with months running from April - March on the x-axis and dengue season on the y-axis. Darker purple tiles indicate months with higher humidity. Provinces are arranged to mirror their geographic location in the Dominican Republic.

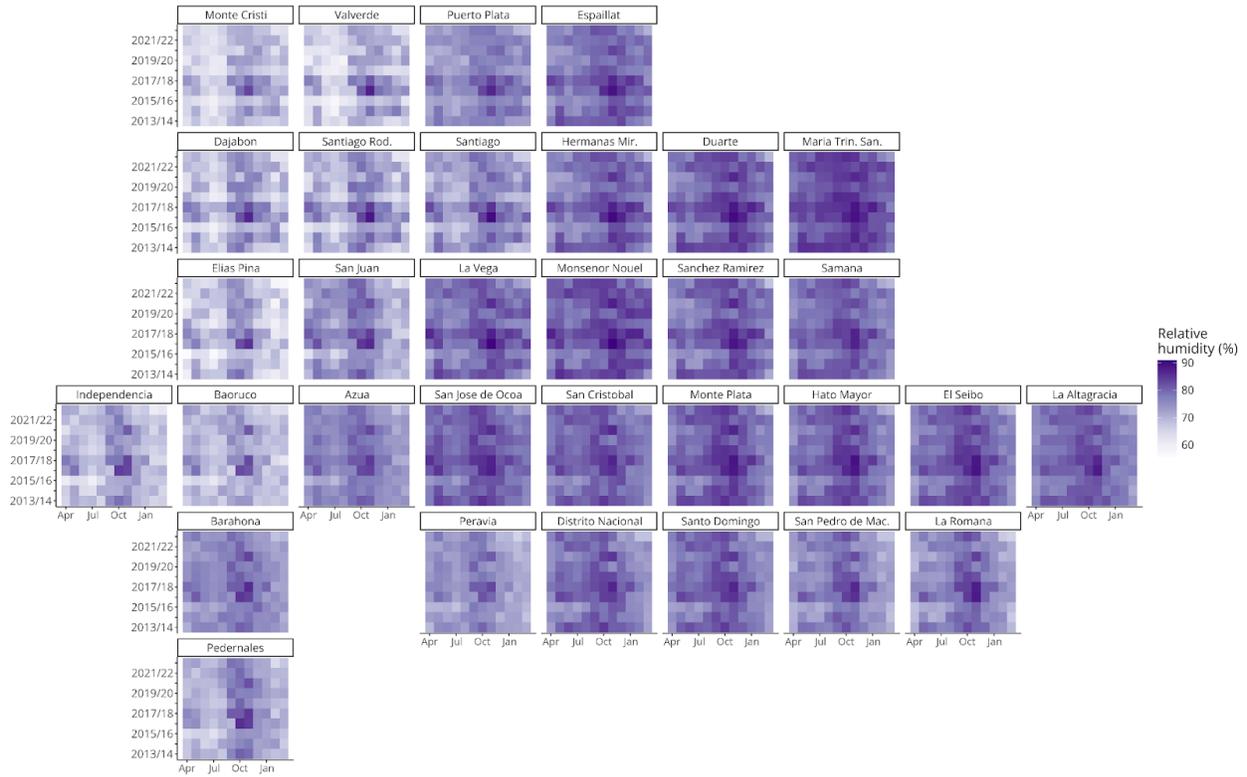


Figure 4: Catalytic model trace plot

Figure showing sampled values for the annual force of infection at each step of the Hamiltonian MCMC using a No-U-Turn sampler (NUTS) algorithm for four chains.

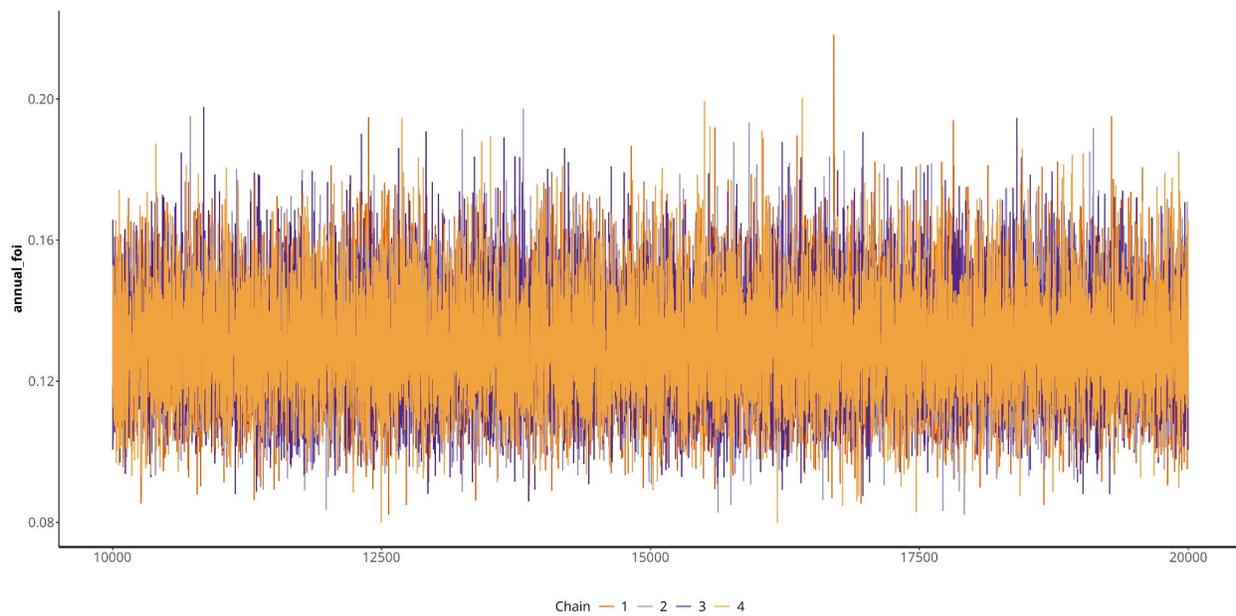


Table 1: Model adequacy statistics from model selection

We conducted forwards stepwise selection, grouping climatic indicators into classes of covariate including: temperature, rainfall, humidity and Niño 3.4. At each step of model selection, the best performing variable was carried forwards and tested against all variables in remaining climate classes. Model adequacy criteria considered include WAIC and DIC (where lower values are better) and R^2 (which can take values 0 to 1, where high values are better). All covariate models tested included weighted lagged cases (on a log scale) and estimated proportion susceptible (on a log scale). Note that while maximum temperature or SPI individually do not improve model adequacy criteria (in Steps 2a and 2b), including them together in Step 3 results in a lower WAIC and DIC. The final selected model is shown in Step 4.

Step	Variable	WAIC	DIC	Rsqr
1	Niño 3.4 (12 week running average, non-linear, 6 week lag) + log(weighted lagged cases) + log(susceptibility)	123513.13	123520.2	0.05
2a	Niño 3.4 (12 week running average non-linear, 6 week lag) + maximum temperature (8 week running average with 4 week lag, non-linear) + log(weighted lagged cases) + log(susceptibility)	123544.13	123549.65	0.05
2b	Niño 3.4 (12 week running average, non-linear, 6 week lag) + SPI (12 month time scale) + log(weighted lagged cases) + log(susceptibility)	123587.46	123589.14	0.049
3	Niño 3.4 (12 week running average, 6 week lag, non-linear) + SPI (12 month timescale) + maximum temperature (8 week running average with 4 week lag, non-linear) + log(weighted lagged cases) + log(susceptibility)	123384.68	123389.88	0.052
4	Niño 3.4 (12 week running average, 6 week lag, non-linear) + SPI (12 month timescale) + maximum temperature (8 week running average with 4 week lag, non-linear) + relative humidity (10 week running average, 2 week lag, non-linear) + log(weighted lagged cases) + log(susceptibility)	123268.5	123271.33	0.054

Figure 5: Comparing average year-specific spatiotemporal random effects for covariate and baseline models

Figure showing *spatiotemporal random effects* for each year, averaged over all municipalities by year. The value of the yearly average *spatiotemporal random effect* indicates whether, on average, dengue incidence is higher or lower than expected for a given year. These are shown on the additive, linear predictor scale where a value of 0 indicates no contribution of the *spatiotemporal random effect*. Therefore, where model covariates are able to account for interannual variation in dengue risk, we would expect these to be closer to zero. Average *spatiotemporal random effects* are shown for: a baseline model including only *seasonal* (weekly) and *spatiotemporal random effects* $\delta_{w[t]} + u_{s,y[t]} + v_{s,y[t]}$, in green; a model including only climate covariates, in orange, a model including only epidemic terms (that is, weighted lagged cases and susceptibility during each dengue season), in purple, and a full model including both climate and epidemic terms, in pink.

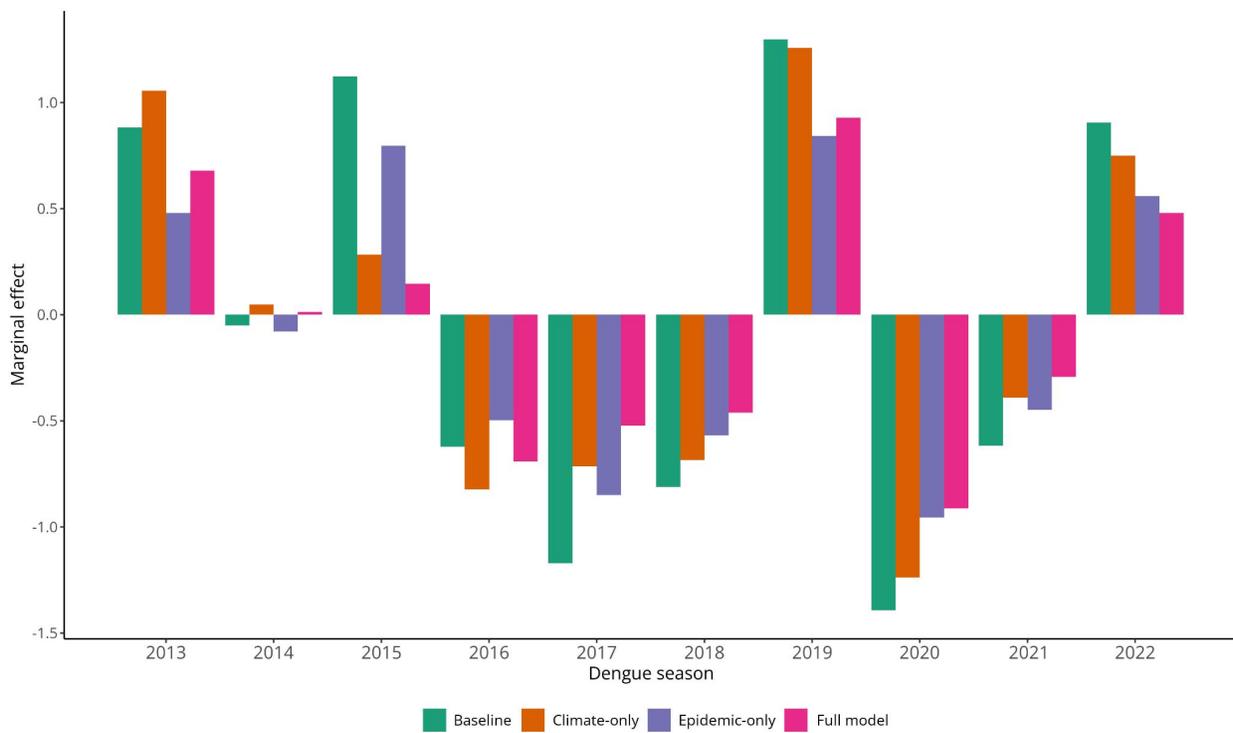


Figure 6: Change in municipality-specific spatiotemporal random effects between covariate and baseline models

Figure showing the absolute difference in spatiotemporal random effects in covariate models compared with a baseline model including only *seasonal* and *spatiotemporal random effects*, during outbreak and non-outbreak years respectively. This is calculated as:

$|u_{i,a[t]} + v_{i,a[t]}|_{Full} - |u_{i,a[t]} + v_{i,a[t]}|_{Baseline}$ Municipalities where the magnitude of the spatial random effects was smaller (i.e. closer to zero) after the inclusion of model covariates are shown in purple. The more negative the difference (and the darker the purple) the greater the contribution of climatic and epidemic model covariates in explaining variation in dengue risk. Municipalities where covariates increased the marginal random effects are shown in green and those with no change are shown in white. The first row shows the difference in *spatiotemporal random effects* between a model only including climatic covariates (panels a and b). The second row shows the difference in *spatiotemporal random effects* between a model only including epidemic components (weighted lagged cases and susceptibility) and a baseline model (panels c and d). Finally, the third row shows the difference in *spatiotemporal random effects* between the full covariate model (including both climatic and epidemic covariates) and the baseline model, as shown in Figure 3 in the main text.

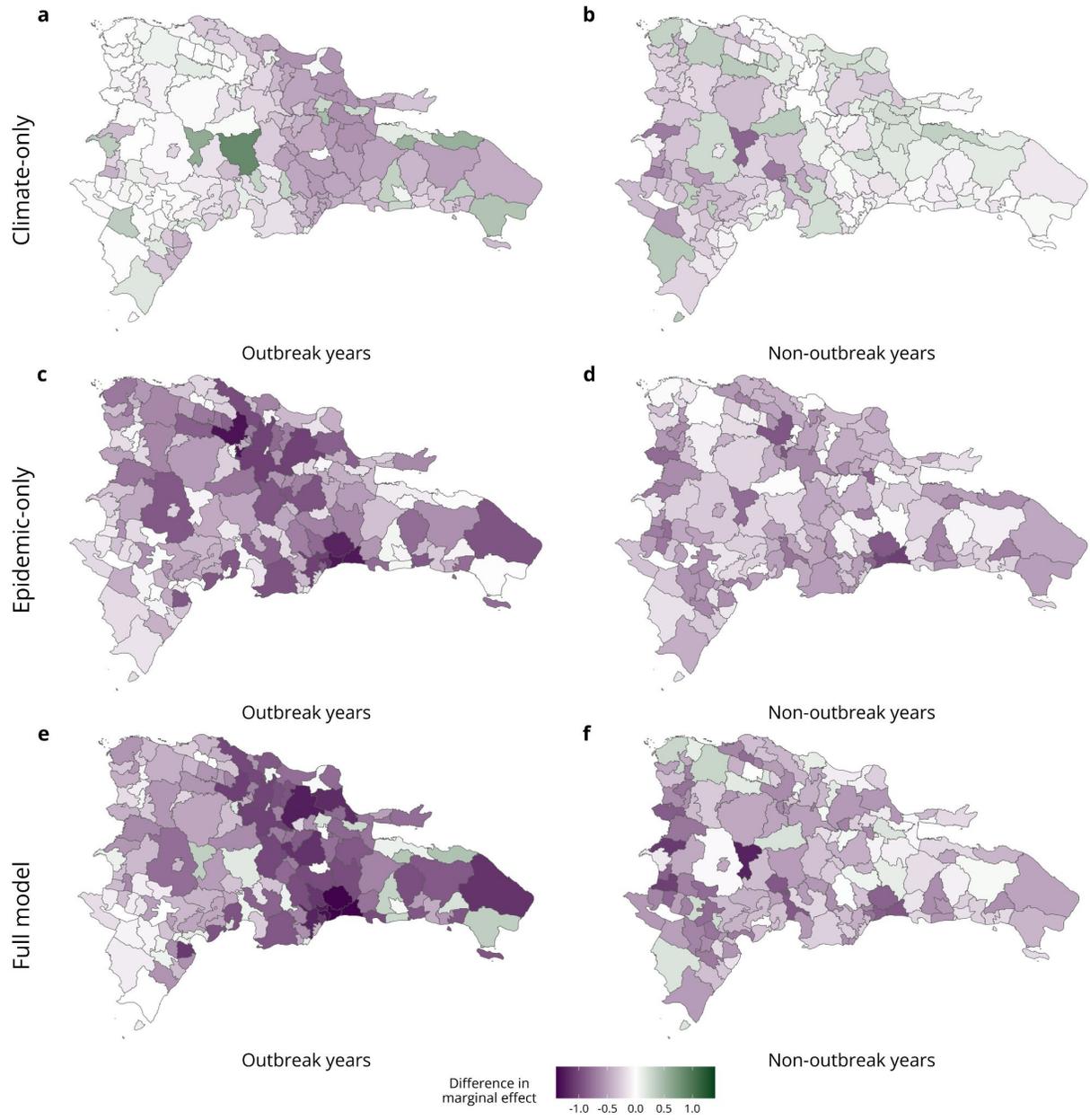


Figure 7: Model evaluation designs

Schematic showing model evaluation designs. The left panels represent time series cross validation with blue blocks representing training data and orange blocks representing testing data. Here for a given week t , the model is trained on data up until and including week $t-1$ with week t used to test the model. For each iteration, we refit the model with an additional week of data in the training set, until we have obtained posterior predictions of dengue incidence for each week and municipality. The right panels represent k-fold cross validation with a spatial holdout design. We conducted 5-fold cross validation, holding out entire municipalities at a time. Each municipality is randomly allocated to a fold from 1:5, the model is then refit 5 times. For each fold f all the municipalities in that fold will be used to test the model, while data from the remaining municipalities will be used for training. To account for random variation in assigning municipalities to folds, this was repeated 10 times with different fold allocations. Posterior predictions for each repetition were then aggregated to the municipality and week level as shown in Supplementary Figure 9.

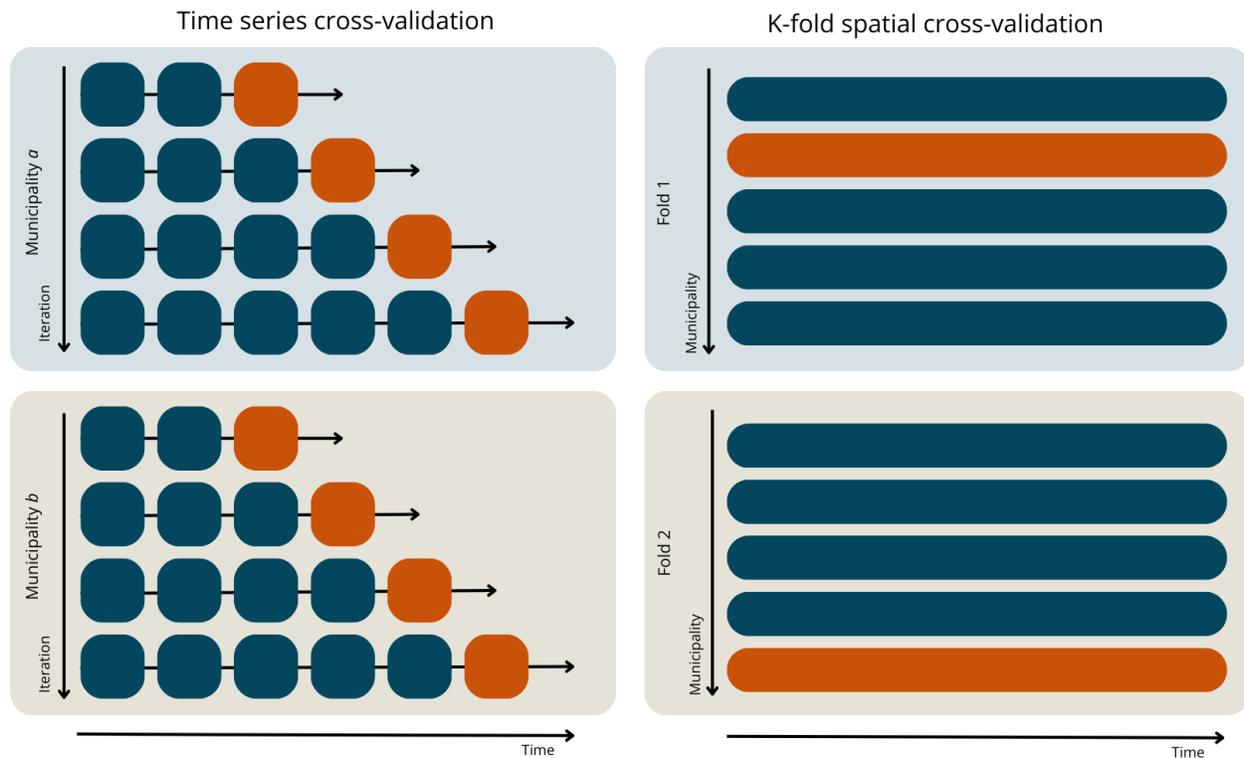


Figure 8: Comparing out-of-sample evaluation metrics over time for the full covariate model and a baseline model

Figure shows model evaluation metrics on out-of-sample predictions for a full model including climatic and epidemic covariates, in green, and a baseline model including only *seasonal* and *spatiotemporal random effects*, $\delta_w[t] + u_{s,y[t]} + v_{s,y[t]}$, in purple. Panel a shows the interval coverage at the 50% and 95% level (shown by solid lines and dashed lines, respectively). This metric assesses model calibration and is related to the concept of reliability. It is calculated as the proportion p of observations that fall within the nominal prediction interval. For a perfectly calibrated model, 50% of the observations would fall within the 50% prediction interval and 95% of the observations would fall within the 95% prediction interval. If p is higher than the prediction interval (k) then the model is underconfident, while if p is lower than the relevant prediction interval the model is overconfident. 95% and 50% are indicated with horizontal dashed lines. Panel b shows the model bias, values under 0 indicate that the model tends to underpredict, while values greater than 0 indicate overprediction. If the prediction is larger than the observed value the bias is calculated as the maximum percentile rank where the prediction is larger than the observed value, whereas if the prediction is smaller than the observed value it is the maximum percentile rank where the prediction is smaller than the observed value. Panel c shows the weighted interval score, this is an approximation to the continuous ranked probability score for quantile forecasts, where smaller values indicate better performance. It penalises predictions based on overprediction, underprediction and sharpness (width of the confidence interval).

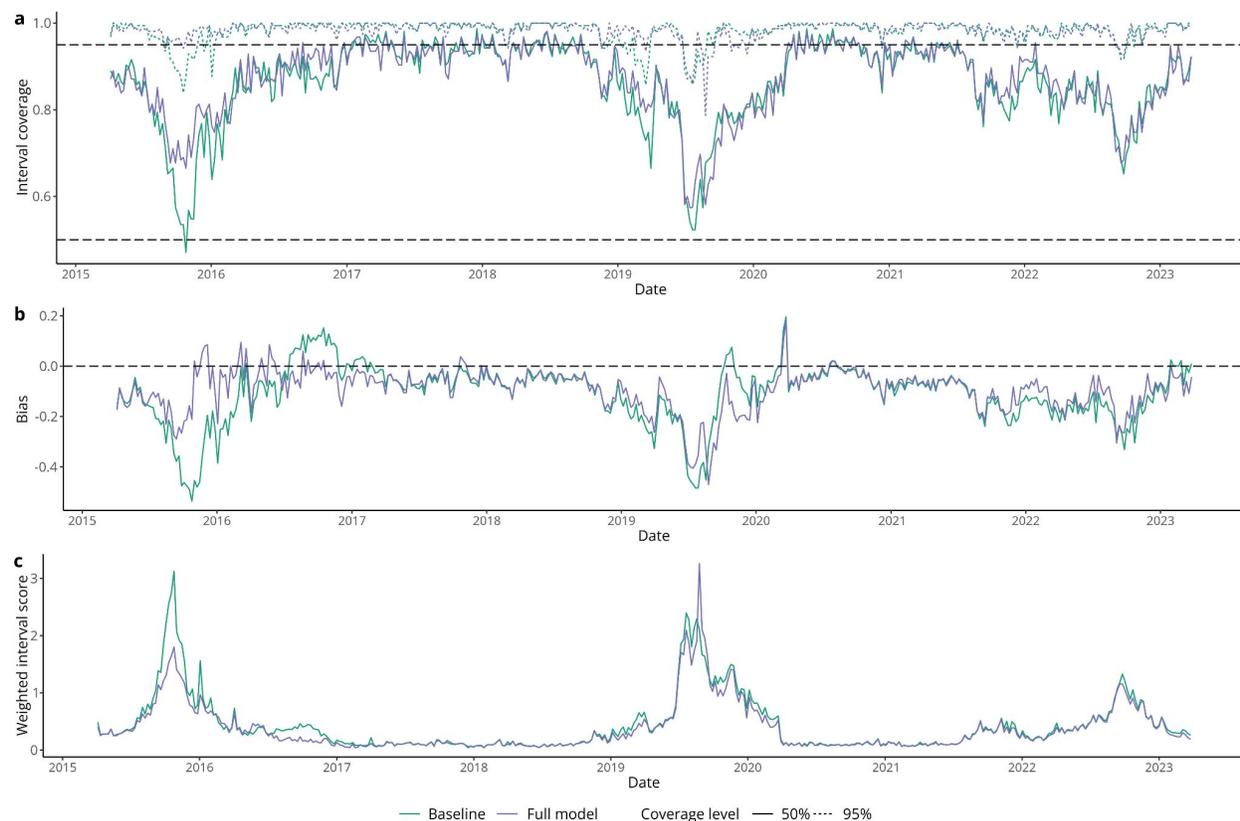


Figure 9: Posterior predictions of dengue cases from 5-fold spatial cross-validation

Figure showing out-of-sample posterior predictions of weekly dengue cases from the full model from 2015 – 2023. Posterior predictions are generated through 5-fold spatial cross validation, as described in Supplementary Figure 7. Green lines represent the median posterior prediction of weekly dengue cases, the shaded green area shows the 95% credible interval, and the grey lines show the data. The model was fit at the municipality level and posterior samples for each time step were then aggregated to the province level.



Table 2: Influence of covariates on dengue predictions (temporal)

Table comparing the continuous ranked probability score (CRPS) and continuous ranked probability skill score (CRPSS) for the full model to models with each covariate excluded. The CRPS is calculated by comparing out-of-sample posterior predictions of dengue cases from time-series cross validation with observed values. Lower CRPS values indicate better predictive ability. The CRPSS is defined as the percentage improvement in CRPS compared with a baseline model and can take values from 0% (indicating the model performs the same as the baseline) to 100% (indicating perfect forecasting skill).

Covariate excluded	CRPS	CRPSS (%)
<i>Full model (none)</i>	0.455	12.7
<i>Climate-only (weighted cases and susceptibility excluded)</i>	0.479	8.1
<i>Epidemic-only (climate covariates excluded)</i>	0.482	7.5
<i>Niño 3.4</i>	0.475	8.8
<i>SPI-12</i>	0.46	11.7
<i>Maximum temperature</i>	0.455	12.7
<i>Relative humidity</i>	0.458	12.1
<i>Susceptibility</i>	0.455	12.7
<i>Weighted cases</i>	0.479	8.1
<i>Baseline (all)</i>	0.521	0

Table 3: Influence of covariates on dengue predictions (spatial)

Table comparing the median continuous ranked probability score (CRPS) and continuous ranked probability skill score (CRPSS) for the full model to models with each covariate excluded. The CRPS is calculated by comparing out-of-sample posterior predictions of dengue cases from 5-fold spatial cross validation with observed values. Here, we show the median CRPS value over 10 repetitions of the 5-fold cross validation, with the associated 95% confidence interval. Lower CRPS values indicate better predictive ability. The CRPSS is defined as the percentage improvement in CRPS compared with a baseline model, and can take values from 0% (indicating the model performs the same as the baseline) to 100% (indicating perfect forecasting skill).

Model	CRPS	CRPSS (%)
<i>Full model (none)</i>	0.855 (95% CrI: 0.853 – 0.86)	10.3 (95% CrI: 9.49 – 10.57)
<i>Climate-only (weighted cases and susceptibility excluded)</i>	0.923 (95% CrI: 0.921 – 0.927)	3.084 (95% CrI: 2.744 – 3.385)
<i>Epidemic-only (climate covariates excluded)</i>	0.87 (95% CrI: 0.867 – 0.876)	3.084 (95% CrI: 2.744 – 3.385)
<i>Niño 3.4</i>	0.859 (95% CrI: 0.854 – 0.862)	9.9 (95% CrI: 9.57 – 10.41)
<i>SPI</i>	0.858 (95% CrI: 0.854 – 0.862)	9.8 (95% CrI: 9.49 – 10.54)
<i>Maximum temperature</i>	0.858 (95% CrI: 0.855 – 0.861)	9.91 (95% CrI: 9.53 – 10.26)
<i>Relative humidity</i>	0.862 (95% CrI: 0.86 – 0.87)	9.4 (95% CrI: 8.53 – 9.95)
<i>Susceptibility</i>	0.855 (95% CrI: 0.851 – 0.859)	10.2 (95% CrI: 9.73 – 10.73)
<i>Weighted cases</i>	0.924 (95% CrI: 0.922 – 0.931)	3.0 (95% CrI: 2.38 – 3.41)
<i>Baseline (all)</i>	0.953 (95% CrI: 0.95 – 0.955)	0

Figure 10: Comparing age-stratified seropositivity from DENV IgG ELISA and neutralisation tests

Figure comparing age-stratified seropositivity tested by DENV1-4 ELISA assay (panel a), DENV1 neutralisation test (panel b) and DENV2 neutralisation test (panel c). IgG ELISA testing was performed by a US CDC in-house assay, while neutralisation testing was performed using fluorescent reporter virus and live imaging of cells.

