COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore

Rachael Pung

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

June 2024

Department of Infectious Disease Epidemiology
Faculty of Epidemiology and Population Health
London School of Hygiene & Tropical Medicine

Declaration

Statement of Own Work

I, Rachael Pung, 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. I have read and understood the school's definition of plagiarism and cheating given in the Research Degree Handbook.

Rachael Pung,
June 2024

Abstract

The COVID-19 pandemic has prompted many countries to implement a mixture of traditional and novel outbreak control measures. This led to changes in human behaviour and the availability of innovative data sources. In this thesis, I integrated different outbreak surveillance data of COVID-19 cases in Singapore with statistical and mathematical modelling tools to understand the transmission dynamics of SARS-CoV-2 and its implications for outbreak control.

Outbreak control measures are often applied in combination but the effectiveness of each measure is seldom independently evaluated. In a retrospective analysis, I used granular epidemiological investigation data and showed that the effectiveness of contact tracing was dependent on the effectiveness of case finding. Furthermore, with the strict quarantine of incoming travellers, the number of imported cases in Singapore in the second half of 2020 was three times higher than that at the start of the outbreak but the effective reproduction number remained below 1. I also found that the outbreak metric on the proportion of cases with no known infectors among all notified cases is not always reflective of the proportion of missed infections among all infections.

In 2022, the relaxation of mainland China's 'zero-COVID' strategy led to a surge of cases within China. Using the same dataset as the previous study but focusing on the imported cases arriving from mainland China to Singapore, I analysed the outbreak trajectory in China in real time. I found that the outbreak in China peaked in mid-December 2022 and, together with no apparent risk from novel strains, helped policymakers in Singapore to decide against reactive border control measures.

With the emergence of new SARS-CoV-2 variants, there was an increase in observed cases within a short period which could be attributed to a decrease in the generation interval, often proxied by the serial interval. Thus, I also performed a real-time analysis but did not identify a large difference of more than one-day reduction in the Delta variant serial intervals as compared with the wild-type SARS-CoV-2. I further discussed how this finding could be attributed to the small sample size of less than 50 transmission pairs in each study period and could affect the power to detect changes, if any.

As a follow-up analysis, I developed a simulation framework to sample transmission pairs and studied the power to detect changes in the generation and serial intervals under varying pathogen biology, outbreak control measures, contact patterns and epidemic dynamics. For a decrease of 0–1.4 days in the incubation period of the Delta variant reported in the literature, I found that a one-day reduction in the serial interval of the Delta variant was unlikely. Overall, a sample size of at least 100 transmission

pairs would be required to provide 30–70% power to detect a one-day change in generation and serial intervals.

Scenario analysis using outbreak simulation models is also useful when planning for the resumption of large-scale events amidst potential threats of new variants that are more transmissible. Using high-resolution temporal contact networks on cruises, I estimated that mask-wearing interventions, in addition to baseline measures of case isolation and physical distancing, would further reduce the outbreak size by 50% after accounting for the periods of interaction in dining and sports settings when passengers are not wearing masks. Also, the risk of a large outbreak was reduced when regular testing of passengers prior to departure and halfway through the event was implemented without having to wear a mask.

Building on the temporal data collected from the cruises and from other studies, I explored the time-varying network properties in cruises, communities, high schools, hospitals and workplaces. The type of contacts that tend to be retained over consecutive timesteps varied across different settings. As the risk of transmission increases with longer contact duration, this implies that outbreak control measures have to be calibrated across each setting. Furthermore, as the terms 'superspreaders' and individuals driving 'superspreading events' are often used interchangeably in the literature, I classified individuals by ranking their connectivity over time. I found that less than 10% of the population in each network was consistently identified as being highly connected, and are potential 'superspreaders' if infectious. Instead, most of the population was highly connected for short periods and could drive 'superspreading events' if infectious.

Overall, I performed a retrospective analysis of the effectiveness of outbreak control measures, real-time analyses of the epidemiology of SARS-CoV-2, and predictive analyses of transmission dynamics in specific settings. Each study of this thesis helps to identify the strengths and weaknesses of the current surveillance system, and the work will help inform the future pandemic preparedness and response policies in Singapore and across the world.

Acknowledgements

This PhD has been a truly enriching journey and motivates me to improve my understanding of infectious disease modelling. My foremost and biggest thanks must go to my supervisors, Dr Adam Kucharski and Dr Vernon Lee. I would like to thank Adam for his consistent guidance — taking time off to fine-tune my research proposal even before I was enrolled as a PhD student, putting aside other work to co-edit on a time-sensitive outbreak analysis that was ready within 24 hours, sending references on mathematical theories which may not be related to my studies. I really appreciate these gestures and not forgetting your advice on career development and scientific writing. I also thank Vernon for his constant support at the Ministry of Health, Singapore (MOH). I was unsure about pursuing a PhD when I first started at MOH. However, you have provided me with many opportunities to perform research studies at the ministry and to showcase my work on international platforms. Your continuous support, including that very letter to HR and senior management (which I still have a copy of!), has provided me with a lot of encouragement and I was ready when the opportunity came. Despite being my director for the past 6 years, I have always viewed you as a mentor. You and Adam are truly inspirational figures to me.

Secondly, I would like to thank my academic committee members, Dr Timothy Russell, Dr Josh Firth and Dr Hannah Clapham. To Tim, thank you for setting aside time to run through very complex mathematical concepts and for being the funniest person in room 107. To Josh, thank you for all the Zoom calls to chat on network theories and for always offering your support to review my thesis. I will definitely pay you a visit in Oxford/Leeds when I am in the UK for my viva. To Hannah, thank you for the discussions on my very initial research study idea and also for the many chats, on PhD, work and life either when we meet in Singapore or in the UK.

Thirdly, I would like to thank MOH for funding this study and all the related publications. I will never forget how the team of HR staff expedited my application in order for me to successfully enrol into LSHTM during the January intake in 2021.

Finally, I would like to thank (ex-)colleagues at the Communicable Disease Division, MOH, my friends, and family. In no particular order, I would like to thank Joven Liew, Shelley Khoo, Cheong Jing, Han Xin Yuun, Tracy Koh, Sarah Hayes, Peh Xinyi, Valerie Koh, Constance Low, Carolyn Ho, Poh Cuiqin, Shanice Teo and Paul Lee for the many personal conversations on life, studies and career development. And to my family, thank you for all the love and patience throughout all these years.

Contents

List of Tables

List of Figures

Acronyms

ART Antigen Rapid Test

CFR Case Fatality Ratio

CI Confidence intervals

COVID-19 Coronavirus Disease – 2019

CrI Credible intervals

Ct Cycle threshold

GI Generation Interval

IgG Immunoglobulin class G

IFR Infection Fatality Ratio

MCMC Markov Chain Monte Carlo

NAAT Nucleic Acid Amplification Test

NPI Non-pharmaceutical Intervention

PCR Polymerase Chain Reaction

PHSM Public Health and Social Measure

RFID Radio Frequency Identification

RT-PCR Reverse Transcription-Polymerase Chain Reaction

SAR Secondary Attack Rate

SARS Severe Acute Respiratory Syndrome

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus 2

SI Serial Interval

VNT Virus Neutralisation Test

VOC Variant(s) of Concern

WHO World Health Organization

1   Introduction

1.1 Background

Infectious disease outbreaks are characterised by a mix of unobserved (e.g. time of infection, missed infections) and observed events (e.g. time of symptoms onset, notified cases), and data collection on the observed events are often incomplete. In combination, this limits our understanding of the epidemiology of the disease, transmission dynamics and severity of infection – critical questions posed at the start of an outbreak or whenever mutations of a pathogen are detected. In turn, the answers to these questions help determine what combination of measures are necessary for outbreak control, when we can relax these measures without sparking new outbreaks, and which sub-populations to prioritise for outbreak control, treatments, or vaccination. To formulate these outbreak control policies, outbreak data analysis is essential for evidence-based decision making.

Data streams during an outbreak are often varied and include, but are not limited to, data from disease notifications, epidemiological investigations, sentinel surveillance, and behavioural surveys. Reporting delay coupled with changing pathogen biology, human behaviour, outbreak control policies and population heterogeneity can create bias and, thus, influence our interpretation of these data. As such, mathematical and statistical tools are helpful for bias correction and identification of risk factors for disease transmission and severe outcomes. Specifically, mathematical models can be used to generate outbreaks, by incorporating the known factors and quantifying the unknown factors involved in the transmission process, and predict the case trajectory. On the other hand, statistical methods can be used to compare the modelled case trajectory with the observed data. Overall, when combined with the data, these tools can help estimate the epidemiology, transmission dynamics and severity of infection to determine the most effective outbreak control policies.

Using the COVID-19 pandemic as a case study, I integrated mathematical models and statistical methods with different datasets to investigate the transmission dynamics of SARS-CoV-2 in Singapore. This introduction provides a background to the COVID-19 pandemic and its epidemiology, the critical public health questions of interest and the types of mathematical and statistical tools available for outbreak analysis generally.

1.2 Epidemiology of COVID-19

COVID-19 is an acute respiratory disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 in humans was first reported as

1

cases of pneumonia of unknown aetiology with epidemiological links to the Huanan Seafood Wholesale Market in Wuhan, China, on 29 December 2019 [1]. However, by January 2020, there was an increase in cases with no direct exposure to the market, indicative of human-to-human transmission [1]. Early epidemiological investigations suggested that SARS-CoV-2 can be transmitted directly via (i) inhalation of respiratory droplets when in close contact with a COVID-19 case or (ii) exposure to aerosols generated during medical procedures or when exhaled respiratory particles of a COVID-19 case suspend in a poorly ventilated area, and indirectly via (iii) fomites [2].

1.2.1 Individual-level natural history of COVID-19

Among the unvaccinated individuals infected with wild-type SARS-CoV-2, about a third were asymptomatic, while the remaining could develop mild or severe disease [3,4]. Systematic reviews and meta-analysis estimated that the mean incubation period (i.e. duration from infection to symptoms onset) of wild-type SARS-CoV-2 cases (i.e. the background strain that contains no significant mutations) was 5–7 days [5,6]. The clinical manifestations in wild-type SARS-CoV-2 cases include fever, cough, fatigue, headache, myalgia, sore throat, coryza, dyspnoea, nausea, or diarrhoea [7,8]. Furthermore, unlike the previous SARS outbreak in 2003 and the H1N1 influenza pandemic in 2009, anosmia was reported in many COVID-19 cases [9].

The infectiousness profile can be proxied by the viral shedding profile, which includes the duration of viral shedding, peak viral load and time to peak. The viral shedding profile can be determined by measuring the changes in the reverse transcription-polymerase chain reaction (RT-PCR) cycle threshold (Ct) values over time. Viral shedding for wild-type SARS-CoV-2 precedes symptoms onset [10–12] and continues for 13.4 days (95%CI 10.9–15.8 days) post-symptoms onset [12]. As such, pre-symptomatic transmission can occur in symptomatic individuals, accounting for more than 40% of the transmissions in 19 out of 31 studies listed in a systematic review and meta-analysis study [3]. Peak viral load for wild-type SARS-CoV-2 occurred about 1–5 days post-symptoms onset [13,14].

Overall, the individual's incubation period, symptoms profile and viral shedding patterns can be influenced by individual-level factors such as age, comorbidities and immunity, and pathogen-related factors such as the type of SARS-CoV-2 variants. The elderly and those with medical comorbidities were at higher risk of developing severe COVID-19 [7,15]. Older age was also associated with earlier symptom onset and prolonged virus shedding [15]. Infection with the Alpha and Delta variants (labelled initially as B.1.1.7 and B.1.617.2 based on the Pango lineage classification [16,17]), which emerged in late 2020, was also characterised by a shorter incubation period, higher peak viral load and longer viral shedding among unvaccinated persons as compared to vaccinated individuals [6,18]. However, with increased population immunity, SARS-CoV-2 infection in vaccinated and/or previously infected individuals is now characterised by milder symptoms [19] and faster viral clearance [18].

Lastly, the interpretation of the above epidemiological characteristics was often based on observations in a population and epidemic dynamics can influence our understanding of the observed incubation period. During the exponential phase of an outbreak, most cases with longer incubation periods have yet to be observed. In real-time, most cases have recent symptoms onset and exposure, causing the observed incubation periods to be at the lower end of the intrinsic incubation period distribution [20,21].

1.2.2 Population-level epidemic growth dynamics

At the population-level, the growth of COVID-19 hospitalised cases and deaths over time is influenced by the timescale of disease transmission, the transmissibility of the virus and the severity of the cases. The timescale of disease transmission between an infector and an infectee is characterised by the generation interval (i.e. duration from infection in the infector to infection in the infectee). As the transmission process is generally unobserved, the generation interval is often estimated using the serial interval (i.e. duration from onset in infector to onset in infectee). With pre-symptomatic transmission, the serial interval for wild-type SARS-CoV-2 is 5.2 days (95%CI 4.9–5.5) and is shorter than the estimated incubation period [22]. As the generation and serial interval are a function of the infector and infectee's incubation period, factors influencing the observed incubation period would also affect the interpretation of these intervals. Furthermore, both intervals are a function of the time from symptoms onset in the infector to transmission. With early case isolation, most of the observed transmission would occur around the time of symptoms onset, thereby shortening the observed serial interval at the population-level [22].

Transmissibility can be characterised by the basic reproduction number, R_0, and the secondary attack rate (SAR). R_0 is defined as the typical number of secondary cases generated by an infectious individual in an otherwise susceptible population [23]. As the outbreak progresses, the effective reproduction number, R_t, measures the average number of secondary cases generated by an infectious individual over time. Exponential growth occurs when R_t exceeds 1 while disease extinction occurs when R_t falls below 1. In early 2020, the estimated R_0 of COVID-19 in China was 2.2 before the introduction of travel restrictions [1,24]. The early R_0 estimates for COVID-19 were higher than the H1N1 influenza pandemic in 2009, with an R_0 of 1.2 [25,26], and comparable with the SARS outbreak in 2003, with an R_0 of 2.2-2.8 [27,28]. R_t is a function of the contact rate between an infectious individual and other susceptible individuals, the duration of contact, and the probability of transmission between an infectious and susceptible pair of individuals. This probability is similar to the SAR, defined as the proportion of exposed individuals who acquired the infection. As such, SAR and R_t are linearly related. Household studies estimated an SAR of about 16% for the wild-type SARS-CoV-2 [29,30], and this estimate was about 1.5 times the household SAR for seasonal influenza [31].

Over the pandemic, pathogen biology and human behaviour changes have led to changes in the R_t and the SAR. Despite the continued implementation of strict outbreak control measures to limit social activities and hence the number of contacts, the R_t of the Delta variant outbreaks was 2.0–4.8 [32,33], and household SAR was 23–37% [34–36], higher than the wild-type SARS-CoV-2. Furthermore, the Omicron variant, which emerged in late 2021, had lower viral replication in the lungs, leading to lower pathogenicity compared to the wild-type SARS-CoV-2 [37,38], Delta and Gamma variants [37], but higher replication in the upper respiratory tract, resulting in increased viral loads. This enhanced the virus's transmissibility [39] with a household SAR of 35–50% [34]. Coupled with the resumption of social and economic activities, the R_t of the Omicron outbreak was reported to be 2–3 times higher than the Delta variant based on a review study [33]. However, vaccination lowered the transmissibility of the infector and susceptibility of the close contacts against Omicron infection by a combined effectiveness of 36% (i.e. SAR was expected to be 36% lower in fully vaccinated than fully unvaccinated households) [34].

The severity of the disease can be described by the frequency of clinical symptoms, complications of an infection and outcomes following infection (e.g. hospitalisation, deaths). Early estimates of the wild-type case fatality ratio, CFR, (i.e. the proportion of deaths among reported cases) were 1.0–1.7% [40–42], lower than the CFR of the 2003 SARS outbreak of 11% [43], but higher than the CFR of the 2009 H1N1 influenza pandemic of 0.008% [44]. After accounting for the underreporting of cases, the wild-type infection fatality ratio, IFR, (i.e. the proportion of deaths among all infections) was estimated to increase exponentially with age for the global population based on seroprevalence surveys in 2020 [45]. Globally, the elderly population aged 60 and above had an estimated IFR of 1% and above, while those aged 30 and below had an IFR of less than 0.06% [45]. A small number of European countries and parts of the American continent had high estimated age-standardised IFR exceeding 0.6% in July 2020 [45]. With the introduction of SARS-CoV-2 vaccines, based on a systematic review of randomised clinical trials in 2020–2022, the risk of death among individuals vaccinated against the wild-type SARS-CoV-2 was lowered by nearly 20 times [46].

1.3 Outbreak control measures

By the end of January 2020, the novel coronavirus SARS-CoV-2 had spread to 17 countries outside of mainland China, prompting the World Health Organization (WHO) to declare COVID-19 as a public health emergency of international concern and countries had implemented a series of different outbreak control measures [47]. Outbreak control measures can be classified as non-pharmaceutical interventions (NPIs) and pharmaceutical interventions that reduce the risk of transmission or burden on the healthcare system. NPIs, also known as public health and social measures (PHSMs), include border controls, case finding, contact tracing and other population-level measures to reduce the risk of transmission. Vaccines are a form of

pharmaceutical intervention for COVID-19 outbreak control but were not available at the start of the outbreak.

1.3.1 Border controls

Border control measures in the context of the COVID-19 pandemic were measures that limited the introduction of SARS-CoV-2 infections to a country or region. These included travel restrictions, which limited the outflow and inflow of travellers from the outbreak epicentre in Wuhan, China at the beginning of the COVID-19 pandemic [24,48] (Table 1.1). Outside of mainland China, countries and regions focused outbreak control resources at the borders to increase the detection of imported cases and reduce the introduction of SARS-CoV-2 into their community. The implemented measures included temperature screening, health declarations, targeted testing of symptomatic travellers and quarantine of travellers arriving from outbreak regions, and restriction of entry for those on non-essential travel. For European countries, border control measures were less stringent but helped delay the influx of imported cases [49]. This gave countries time to develop and validate SARS-CoV-2 diagnostic tests.

By late January 2020, with increasing signs of widespread transmission, China implemented an unprecedented lockdown of Wuhan and other cities in the Hubei province, China, to cease all travel into and out of the city and restrict social and economic activities within each of the cities. This was eventually expanded to other parts of the country, and other countries adopted similar lockdown measures to concentrate all outbreak control efforts on community control measures.

1.3.2 Case finding and contact tracing

In the community, case finding at different levels of the healthcare surveillance system involved testing suspect cases and early isolation of individuals infected with SARS-CoV-2 [50,51]. At the same time, contact tracing identified close contacts of a known case as they might be infected [50,51], and measures taken for these individuals may include quarantine and/or testing, thereby breaking the chain of onward transmission. These measures targeted individuals with known risk(s) of infection in the community. In general, the COVID-19 case finding strategy was dependent on factors such as the exposure history of the cases (i.e. potential transmission routes), the proportion of symptomatic individuals and their clinical features. This strategy ranged between targeted testing of close contacts and symptomatic suspect cases, and population-wide mass testing. Quarantine and isolation of measures would need to account for the duration of the incubation period and the duration of viral shedding respectively.

Early in the outbreak, studies showed the potential for pre-symptomatic SARS-CoV-2 transmission [3]. Thus, for contact tracing to be effective in identifying infected secondary cases before the start of their infectiousness, novel contact tracing methods, such as using digital contact tracing tools, were implemented. This minimised the chances of missed contacts arising from recall bias in case interviews

and reduced the time to inform close contacts of their exposure history and subsequent quarantine. As the number of unassociated clusters of SARS-CoV-2 infection increased, backward tracing, which involved retrospective reviews of the exposure histories of index cases (as opposed to the review of movement histories of cases in forward contact tracing), was sometimes conducted to establish epidemiological linkage of the clusters [52]. For transmission that is clustered around a primary case, backward tracing to identify this primary case would potentially reveal more transmission chains that have yet to be identified [53]. Backward tracing was implemented when countries aimed to suppress viral transmission to achieve low or no cases, and was more effective in outbreak control if the transmission was clustered.

1.3.3 Other non-pharmaceutical interventions

The use of other non-pharmaceutical interventions (NPIs) was progressively implemented [54,55] as knowledge on the modes of transmission and the proportion of pre-symptomatic transmission and, hence, the effectiveness of NPIs in preventing infection increased. Furthermore, population-wide implementation of large-scale NPIs had a substantial impact on the lifestyles of many and had to be communicated effectively to the public to achieve high adherence rates [56,57]. In Southeast and East Asia, and the Pacific, most of the population-wide NPIs were mandatory at the time of the outbreak [2,3]. However, with the introduction of new diagnostic tests and rapid roll-out of vaccination, NPIs were progressively relaxed when the healthcare capacities were not under pressure.

1.3.4 Pharmaceutical intervention

Vaccines were the key pharmaceutical intervention that reduced the risk of COVID-19 transmission and severe disease among SARS-CoV-2 naïve individuals and were rolled out to populations worldwide in December 2020. The development and rollout of the COVID-19 mRNA vaccines occurred at unprecedented speed. The efficacy of two doses of mRNA vaccines was about 91–95% in preventing symptomatic infection and 93–100% in preventing severe disease (i.e. hospitalisation, ICU, death) when infected with wild-type SARS-CoV-2 in randomised control trials conducted between July 2020 to March 2021 [58,59]. Based on systematic review and meta-analysis, the respective efficacies were lower at 70–82% and 86–95% in consideration of infection with pre-Omicron variants in randomised control trials [46]. Real-world effectiveness of two doses of mRNA vaccines in preventing symptoms upon infection with pre-Omicron variants was about 78–95% in those aged above 16 and 77–91% in those aged 60 and above [60]. Vaccine effectiveness against severe disease was about 85–99% [60].

As countries progressively reopened their borders for business and non-essential travels from the second half of 2020 onwards, COVID-19 vaccination certifications were required for travellers to enter the country or to be exempted from on-arrival tests or quarantine, similar to existing border control measures for Yellow Fever.

Furthermore, as part of the resumption of social and economic activities, proof of vaccination was a requirement for entry to certain events or venues such as nightclubs or concerts.

Outbreak Control Measures	Traditional	Novel
Border controls *Minimise disease introduction into the community*	• Quarantine or restrict movement of incoming travellers from countries with ongoing outbreak • Isolation and testing of symptomatic travellers	• Imposing total border closures or lockdowns in countries or regions
Case finding and contact tracing *Targeted at known or potential source(s) of infection in the community*	• Testing of suspect cases • Enhance surveillance in specific sub-populations or vulnerable groups • Activity mapping and case interviews to establish close contacts • Isolation of cases and quarantine of close contact	• Use of digital contract tracing tools to identify close contacts otherwise missed from case interviews • Backward tracing to identify source(s) of infection and potential sub-populations for targeted case finding
Other non-pharmaceutical interventions *Untargeted community or population-level preventive measures*	• Physical distancing • School and venue closures • General health advisory (e.g. hand washing, wearing of mask when sick)	• Large-scale population movement restrictions and corresponding work-from-home arrangements • Population-wide face mask usage; mandatory in some countries • Pre-event testing
Pharmaceutical interventions *Vaccines that minimises transmission or severe disease*	• Proof of vaccination prior to entry into countries	• Proof of vaccination prior to entry to events or venues • Accelerated development and roll-out of COVID-19 vaccines

Table 1.1 Summary of outbreak control measures used in various countries

1.4 Diagnostic testing

At the individual-level, testing and detection enabled early implementation of outbreak control measures such as case isolation and contact tracing, thereby reducing SARS-CoV-2 transmission. This also allowed for early administration of anti-virals, available in the later stages of the pandemic, and other therapeutics to help inhibit viral replication, and hence the risk of developing severe disease. Furthermore, as part of epidemiological investigations, testing helped to estimate the secondary attack rate, SAR. At the population-level, identifying cases over the COVID-19 pandemic helped establish the outbreak trajectory, sub-population at-risk and the burden of the disease. Overall, testing directly and indirectly aided in reducing the burden on healthcare by facilitating early treatment and reducing onward transmission.

The gold standard for diagnosing acute COVID-19 infection was through the detection of the SAR-CoV-2 RNA (i.e. genetic material of the virus) in nasopharyngeal swabs or bronchial aspirate using reverse transcription-polymerase chain reaction (RT-PCR) [61]. The turn-around time in Singapore was typically about 24–72 hours for this form of nucleic acid amplification test (NAAT). As the viral shedding patterns change over the course of the infection, the sensitivity of the test (i.e. probability of test to classify infected persons as positive) was about 80% at five days post-symptoms onset [62]. Furthermore, when used in the field (i.e. outside of a controlled environment), the test's sensitivity can be affected by other sample collection-related factors such as the sample collection site, sampling technique, and specimen storage.

By June 2020, rapid diagnostics tests for SARS-CoV-2 antigens were developed and widely distributed. These tests were mainly self-administered with a fast turn-around time, thus facilitating pre-event testing at large-scale events [63,64], routine testing of high-risk sub-populations or healthcare workers [65,66], or even regular testing in the population [67,68]. Unlike NAATs, there was no amplification of the target SARS-CoV-2 antigen (i.e. proteins on the virus's surface) for detection, making antigen tests less sensitive than RT-PCR but test results were ready within 30 minutes [61]. Higher viral load increases the test sensitivity [67,69] and, hence, assessing the test adequacy should account for the viral shedding patterns of the current circulating strains.

While understanding acute infections is essential for calibrating outbreak control measures, understanding past infections was helpful for backward tracing to identify potential source(s) of infection [52] or to retrospectively assess the underlying number of infections in the population over time to quantify the infection fatality ratio, IFR [70]. This was achieved through serological assays, which detect SARS-CoV-2 antibodies post-infection [61,70]. Similar to the RT-PCR and rapid antigen test, the sensitivity of the serological test was affected by the time of administering the test post-infection. A test sensitivity of less than 50% was reported when administered less than a week post-symptoms onset as the antibodies had yet to be developed [71]. The decline of

antibody titre levels, from a peak at 2–3 weeks post-symptoms onset to a stable plateau after 3 months, also led to a corresponding decline in test sensitivity [72–74].

1.5 Analysis of epidemic dynamics and implications for outbreak control

Understanding the drivers behind the COVID-19 outbreak trajectory was useful for prospective planning of outbreak control measures or retrospective evaluation of the effectiveness of these measures for future outbreak or pandemic preparedness planning. Different outbreak analyses required different methods, and the implications of the outcomes varied. Table 1.2 summarises the outbreak analyses performed during the COVID-19 pandemic. The listed types of studies aimed to understand the epidemiology of COVID-19 and the epidemic growth dynamics and do not include virological, clinical, immunological, behavioural and socio-economic analyses, which were important for the overall outbreak management but are not the focus of this thesis.

1.5.1 Statistical and mathematical models

The choice of statistical and mathematical models for outbreak analyses (Table 1.2) can differ based on several factors: (i) the type of data available for inference, (ii) the transmission process to model, and (iii) the uncertainty of the outcomes. If observed data was used for outbreak analyses, regardless of the model choice, the modelling framework could incorporate additional analyses for bias correction, data censoring or stratification by risk factors. In the following subsections, I described statistical and mathematical models and, in some instances, provided examples of their use cases during the COVID-19 pandemic.

1.5.1.1 Statistical: single distribution

At the early stages of the COVID-19 pandemic, epidemiological data pertaining to the individual was collected via case interviews. This data includes the symptoms of COVID-19 and key epidemiological delay distributions such as the incubation period, onset-to-report, onset-to-death. To interpret the data collected from a population of cases, parametric distributions (e.g. lognormal, gamma, Weibull) were commonly used to capture the long right tail distribution of the data [5,75,76]. This method does not involve modelling the transmission process, and parameters of the fitted distribution were often used as input parameters in mathematical models for population-level outbreak analyses.

Level of analysis	Type of analysis	Outcomes of interest	Implications to outbreak control strategies	Statistical (S) or Mathematical (M) Model (section 1.5.1)
Individual	Natural history of disease	• Disease states and clinical manifestation • Incubation period • Infectiousness period (proxied by viral shedding profile) • Proportion of symptomatic cases	• Case finding / Testing (targeted vs population) • Contact tracing • Quarantine • Isolation	• S: single distribution • S: multiple distributions • S: regression
Pairwise	Modes of transmission	• Proportion of contact or droplet borne, airborne, and fomite transmission	• Case finding / Testing (targeted vs population) • Contact tracing • Population-wide NPIs	• Direct observation/Informal inference
	Timescales of transmission	• Generation interval • Serial interval • Proportion of pre-symptomatic transmission	• Case finding • Contact tracing • Population-wide NPIs	• S: single distribution • S: multiple distributions
Population / Cluster	Disease introduction	• Incidence / prevalence among travellers	• Border control • Case finding • Quarantine / Testing (alternative to quarantine)	• S: single distribution • S: scaling • M: compartmental model

	Transmissibility	• Basic/Effective reproduction number • Secondary attack rate • Overdispersion in transmission	• Border control • Case finding • Contact tracing • Quarantine • Isolation • Population-wide NPIs • Vaccination	• S: regression • M: compartmental model • M: branching process model • M: renewal equation • M: network model
	Severity and burden of disease	• Risk / rates of hospitalisation or ICU admission • Case fatality ratio • Infection fatality ratio • Case ascertainment rate	• Case finding • Isolation (healthcare capacity planning) • Population-wide NPIs • Vaccination	• S: single distribution • S: regression • S: scaling • M: compartmental model • M: renewal equation • M: network model
	Immunity / Protection	• Duration of immunity post infection / vaccination • Proportion of susceptible population over time • Herd immunity threshold • Vaccine effectiveness against infection / severe disease • Correlates of protection	• Vaccination • Resource projection/plans	• S: scaling • S: regression • M: compartmental model
	Effectiveness of outbreak control measures	• Real-time / retrospective evaluation of measures	• Resource projection/plans	• S: regression • M: compartmental model • M: branching process model • M: renewal equation • M: network model

Table 1.2 Summary of outbreak analyses for the COVID-19 pandemic and the typical methods for analyses

The intrinsic incubation period distribution is the incubation period distribution we expect to observe for an average infected individual in the population [21] and implicitly assumes that the times of infection and symptoms onset are fully observed. During the exponential growth phase of the COVID-19 outbreak, the observed incubation period from infected individuals with similar onset times (i.e. backward incubation) was shorter than the intrinsic incubation period distribution (i.e. forward incubation) which was unaffected by exponential growth dynamics [20,21]. This was because most of the infected individuals with longer incubation had yet to be observed. As such, when we take a snapshot of the outbreak, a higher proportion of the cases with short (long) incubation periods were observed in the earlier (later) phase of the outbreak. To account for this bias in the backward incubation period and derive the forward incubation period, the following adjustment was applied [20]:

$$f(\tau) = \frac{exp(r\tau)b(\tau)}{\int_0^\infty exp(rx)b(x)dx} \tag{1.1}$$

where $f(\tau)$ and $b(\tau)$ are the forward and backward incubation period τ time since infection, r is the outbreak growth rate and x is the variable of integration. A similar form of bias correction was also applied to account for data censoring in COVID-19 deaths [77]. However, instead of using exponential distributions for bias correction, parametric or empirical distributions of the onset-to-death or hospitalisation-to-death were used to adjust for the delay in death observations. This delay is independent of exponential outbreak growth dynamics.

1.5.1.2 Statistical: multiple distribution

Through contact tracing, epidemiological data pertaining to pairs of infector and infectee was collected by contact tracers. Data collection was more extensive in the early stages of the COVID-19 pandemic in 2020. The pairwise transmission process was partially observed; the time of symptoms onset was often collected but information on the precise time of exposure was rarely available. With multiple transmission pairs, statistical likelihood inference methods were used to infer the parameters of the SARS-CoV-2 generation interval distribution. In one generation interval study [78], parametric forms (e.g. lognormal, gamma, Weibull) of multiple distributions of epidemiological properties that contributes to the likelihood of observing the collected data was first assumed. For each transmission pair with full data on the range of exposure times and symptoms onset, the likelihood of observing these data points were estimated from the product of the likelihood of the incubation period, f, and the assumed generation interval, ω, using the following simplified formulation:

$$L_{i,j} = \sum_{I,i=e_{min,i}}^{e_{max,i}} f\left(t_{S,i} - t_{I,i}\right) \sum_{I,j=e_{min,j}}^{e_{max,j}} f\left(t_{S,j} - t_{I,j}\right) \omega\left(t_{I,j} - t_{I,i} \mid \theta\right) \tag{1.2}$$

where t_S and t_I are the time of symptoms onset and infection of infector i or infectee j, e_{min} and e_{max} are the range of time of exposure for each case, θ is the unknown parameters of the generation interval distribution of interest. The overall likelihood was then estimated from the product of the likelihood of all pairs, and numerical methods such as Maximum Likelihood Estimation or Markov chain Monte Carlo (MCMC) were used to estimate θ. Data augmentation was also used to establish epidemiological links for cases with multiple potential exposures and the augmented data was jointly estimated with the generation and incubation period distribution [79]. Adjustment for the incubation period of the infector was sometimes performed when data was collected during the exponential growth phase [21,78].

In other COVID-19 generation interval studies [79–81], the serial interval distribution was modelled from the incubation period distribution of the infector i and infectee j, and the generation interval distribution, ω. The observed serial intervals were then used to compute the likelihood of the observation given the modelled parametric distribution of the serial intervals, and the parameters of ω that maximise the likelihood were inferred. Finally, in the process of inferring the times of exposure with this method, these studies also estimated the proportion of pre-symptomatic transmission [78,80].

1.5.1.3 Statistical: regression

Regression analysis on individual-level data was used to quantify the differences in the outcomes of interest due to characteristics of the COVID-19 case (e.g. age, symptoms, vaccination history). For studies on the viral shedding profile of individuals, linear regression analysis was commonly used to model the changes in the PCR cycle threshold (Ct) values of a case over time [62,82], while logistic regression was used to estimate the sensitivity of COVID-19 tests in correctly identifying cases at different post-infection time points [62] or for a given viral load [83,84]. For household analysis, logistic regression was used to estimate the secondary attack rates [85–87] and the proportion of asymptomatic cases when the study involved systematic testing of all household members regardless of symptoms [87]. Besides estimating the risk of infection, regression analysis was also used to estimate the risk or rates of developing severe COVID-19 infection and, hence, the risk of hospitalisation, admission to intensive care units (ICU) or death in the majority of the studies identified in systematic reviews and meta-analyses [88,89].

At the population-level, regression method was applied to quantify the impact of outbreak control measures on the time-varying reproduction number (estimated using other methods) [54]. Overall, this method does not model the transmission process but rather models the outcomes of transmission.

1.5.1.4 Statistical: scaling

At the population-level, the incidence or prevalence of underlying SARS-CoV-2 infection events in a population or among travellers was rarely observed. However, with information on the notified COVID-19 cases, an appropriate scale factor could be applied to estimate the true burden of infection. This scale factor could account for, but is not limited to, the underreporting of mildly symptomatic or asymptomatic cases [90–92], the probabilities of testing positive given test sensitivity over time of infection [90–93], and coverage of the sentinel sites in a population [90–92].

1.5.1.5 Mathematical: branching process model

Contact between pairs of infected and susceptible individuals may or may not result in transmission. Thus, branching process models were used to model or characterise the risk of transmission at the individual-level [94–98]. This is unlike the statistical inference methods in section 1.5.1.2 which performs inference on transmission events. The branching process modelling framework in COVID-19 studies often assumed a parametric form of the incubation period and the infectiousness profile, and was used to estimate the effectiveness of outbreak control measures [95,96,98] or the effective reproduction number and extent of overdispersion in COVID-19 transmission [97]. In non-COVID-19 transmission studies (e.g. influenza, MERS-CoV), branching process models were used to estimate the risk of transmission and susceptibility by age, contact type or modes of transmission, and the heterogeneity of the reproduction number [99–101].

In some outbreak simulation studies, the overall number of new infections and time of infections were modelled based on the reproduction number and the generation interval distributions [94–96]. These simulation studies were predominantly used for overall resource planning [95,96]. In other studies where the focus was to understand the epidemiological factors for transmission [99,101], transmission was modelled via a Poisson contact process and the probability of an infection p occurring at time t given that the infection did not occur in previous time steps is:

$$p(t) = e^{-\Lambda} \left(1 - e^{-\lambda(t)\Delta t}\right) \qquad (1.3)$$

where $\lambda(t)$ is the force of infection at time t and Λ is the cumulative force of infection since the start of infectiousness to time $t-1$. The first coefficient on the right-hand side of the equation is the probability that the infection did not occur in the previous time steps, and the second coefficient is the probability of infection occurring at time t. The force of infection between a contact can be modelled to incorporate individual-level data on the different intensities of contact or viral shedding profiles. For models that further assume continuous contact over time, the waiting times to the next event (i.e. the delay in generating new infection) is exponentially distributed.

Branching process models were used in household or cluster-level studies to fit with epidemiological data and evaluate the likelihood of infection between contact pairs. The overall likelihood was estimated from the product of the likelihood of all pairs, and unknown model parameters were inferred using numerical methods, similar to the statistical inference method in section 1.5.1.2.

1.5.1.6 Mathematical: renewal equation

The renewal equation models transmission over continuous time and was used by studies to estimate the population-level transmissibility or severity of SARS-CoV-2 infection [55,78]. The transmission process and hence the modelled number of infected individuals I over the calendar time t was modelled in a simplified formulation as follow:

$$I(t) \ = \ \int_0^\infty \beta(\tau)\, I(t - \tau) d\tau \qquad (1.4)$$

where β is the mean rate at which an individual infects others τ time since infection (i.e. infectiousness). $\beta(\tau)$ can be expressed as a product of the reproduction number, R, and generation interval, $\omega(\tau)$, which characterises the probability distribution of acquiring an infection from an individual infected τ days ago. Thus, the newly infected individuals at time t, would have acquired infection from previously infected individuals with varying levels of infectiousness τ days since infection. Infections can be modelled either through a deterministic or stochastic process. The latter provides an understanding of the uncertainty of the transmission dynamics but requires more computational time and memory. These are trade-offs to consider when modelling transmission in large populations and over long periods.

Using the above model framework, population-level studies inferred the effective reproduction number, R_t, a measure of transmissibility, by assuming a parametric distribution for ω [55,78]. This distribution can also be modelled using non-parametric distributions. In other words, the waiting time in a renewal equation is not necessarily exponentially distributed (as assumed in most branching process models). By incorporating data on the infection fatality ratio, the renewal equation model was used to simulate the expected notified cases or deaths over time [55]. Model fitting was done by evaluating the negative binomial likelihood of observing the COVID-19 cases or deaths given the modelled expectations (e.g. $I(t)$). A negative binomial distribution was used to account for the overdispersion of the surveillance data. With sufficient data on the individual-level attributes of transmission pairs, renewal equations were used to estimate the heterogeneity of β and hence R_t based on age, contact type or modes of transmission [78], similar to the branching process model in section 1.5.1.5.

As the renewal equation in most COVID-19 studies did not account for the population size, this model does not capture the effects of population immunity over time.

Immunity acquired through infection or vaccination would lower the average individual's risk of infection against the current circulating pathogen and R_t will decrease over time. This limitation can be overcome by fitting the model to different periods of the outbreak to allow for changes in R_t in each period or to model a time-varying R_t using a Gaussian process [55,102].

1.5.1.7 Mathematical: compartment models

Compartmental models were also commonly used to estimate the population-level transmissibility or severity of SARS-CoV-2 infection [103–105] or to study the herd immunity thresholds (i.e. proportion of the population that have acquired immunity to avoid large outbreaks) [106]. At the beginning of the COVID-19 pandemic, they were used to model the outbreak trajectory in the epicentre, Wuhan, China, by using data on exported COVID-19 cases and the travel volume to respective countries [105]. Unlike the renewal equation, COVID-19 studies involving compartmental models accounted for the overall population size. The models separated the population into sub-populations described by the COVID-19 disease states of susceptible (S), exposed but not infectious (E), infectious (I) and recovered (R). Heterogeneity was modelled by incorporating different sub-groups of individuals based on age, and random mixing was assumed in each of the (sub-)compartments.

As compartmental models generally take the form of ordinary differential equations, the overall waiting time in each compartment was assumed to be exponentially distributed (including infection times). Prior to the depletion of susceptible individuals, the early dynamics of a stochastic SIR model are similar to a branching process model with exponentially distributed waiting times [107]. Similar observations apply between a deterministic SIR model and a renewal equation [108].

As COVID-19 compartmental models studied the outbreak dynamics in populations consisting of thousands to millions of individuals, transmission was often modelled deterministically to reduce computational run time and memory for model fitting. Modelled outcomes of infections were then compared to the observed incidence to estimate parameters related to transmissibility or severity.

1.5.1.8 Mathematical: network models

Network models illustrate a complex population structure by explicitly modelling the contacts between individuals and tracking each individual's disease status, similar to compartmental models (e.g. S, E, I or R for the life course of COVID-19). In COVID-19 network models, contacts between individuals were modelled to be static (i.e. unchanging with time) [109–111] or dynamic (i.e. time-varying) [112]. Network models with contacts formed randomly in a population approximate a compartmental model that simulates population-level transmission [113]. In some studies, contacts were modelled based on population census. This involves the allocation of contacts and the associated location of the contacts to each individual. Numerical methods were

employed to swap contacts between individuals until the number of contacts and distance of travelling (i.e. location of contact) matched the observed population census [111,112]. In other studies, contacts were formed based on real-world recorded or reported contacts [109,110].

In most COVID-19 network models, transmission between a pair of individuals was modelled via a Poisson contact process [110], equivalent to a branching process model, or based on default transmission probabilities in a given location [111,112]. Scale factors were used to calibrate the basic reproduction number in these models to be representative of SARS-CoV-2. Overall, these models were commonly used to simulate disease transmission stochastically to investigate the impact of varying social structures in the population on the distribution of outbreak sizes and evaluate potential interventions [110,112,114]. In some instances, it was used to estimate the generation interval in household settings [109].

1.6 Modelling SARS-CoV-2 and gaps in the literature

Outbreak modelling over the course of the COVID-19 pandemic can be classified into three broad categories: (i) evaluating the effectiveness of past outbreak control measures, (ii) real-time analysis of the epidemiological characteristics of SARS-CoV-2 infection and nowcasting (i.e. short-term predictions) of the outbreak trajectory, (iii) examining the impact of potential outbreak control measures to be implemented.

1.6.1 Evaluating the effectiveness of outbreak control measures

For studies looking at the effectiveness of past population-level outbreak control measures, these studies often modelled the changes in the effective reproduction number, R_t, over different periods with varying outbreak control policies [54,55]. However, in the absence of granular individual-level data, there were limited studies concurrently evaluating the effectiveness of respective control measures such as border controls, case finding, and contact tracing. Furthermore, studies of R_0 and R_t were primarily based on notified cases, and few accounted for the cases' travel history [55,115,116]. The dynamics between missed infectors versus notified cases and local versus imported cases could be different. Specifically, for the same contact patterns, missed infectors are expected to generate more infections than notified cases who are isolated upon notification and, hence, unable to generate infections despite being infectious. For imported cases, their movement histories and contact patterns were potentially different from local cases. Factoring in these heterogeneities is essential for correctly interpreting the changes in R_t and, hence, the estimated effectiveness of the outbreak control measures when performing retrospective evaluations.

1.6.2 Real-time analysis of SARS-CoV-2 outbreaks

Real-time nowcasting of the overseas outbreak trajectory helped countries determine the risk of disease introduction. In January 2020, studies adjusted the number of

exported cases from Wuhan, China by the travel volume to respective cities and estimated the outbreak trajectory at the outbreak epicentre before corroborating with the reported cases in Wuhan, China [105,117]. When pre-departure and on-arrival testing policies in most countries were imposed from the end of 2020 to mid-2022, one study made additional adjustments to the observed imported case incidence to account for testing practices when attempting to nowcast the outbreak trajectory at the ports of departure [93]. However, by the end of 2022, many countries had withdrawn their surveillance measures and transitioned away from acute outbreak response. Furthermore, testing capacities in countries with ongoing outbreaks, such as China in Oct 2022, were limited [118]. This resulted in a paucity of information and suggested a need for robust real-time estimation of the short-term outbreak dynamics in countries with ongoing outbreaks along with the risk it poses to other countries.

Real-time analysis of the epidemiological characteristics of SARS-CoV-2 was performed during the COVID-19 pandemic, especially when viral mutation occurred. With the onset of new SARS-CoV-2 variants, it was important to evaluate the changes to the generation or serial intervals arising from the changes in the pathogen's characteristics, which would influence the speed of the outbreak. In turn, this would affect the speed required to expand outbreak control measures [119]. Most studies on the generation or serial interval reported these characteristics for the current circulating SARS-CoV-2 variant [21,22,32,78,79,120]. Studies that compared the difference in these characteristics between different variants were mainly retrospective analyses [20,94,109]. Across all studies, some stratified for different contact types [22,109], varying time from onset-to-isolation [22], or adjusted for exponential growth dynamics [20,21,78], and sample sizes ranged from 40 to over 1000. However, these factors were not consistently adjusted for when comparing the generation or serial intervals between different variants, thus making it challenging to compare results across different outbreak periods and settings. During the Delta variant outbreak in Singapore in April 2021, a rapid increase in hospitalised COVID-19 cases was observed. To calibrate the outbreak control measures, there was a need to understand the drivers of the outbreak: a shortened generation interval, an increase in the R_t or both.

To our knowledge, studies that compared the differences in generation and serial intervals over different outbreak periods did not perform further statistical inference to quantify the power to detect these differences for a given sample size [32,120]. Thus, having a framework to simulate changes to the pathogen characteristics while accounting for variations in external factors would allow us to evaluate the overall impact on the generation or serial intervals for a given sample size and ensure that a future study design is well-powered to detect changes in the timescales of infection.

1.6.3 Examining the impact of potential outbreak control measures

Examining the impact of potential outbreak control measures to be implemented was often performed when countries planned to resume social and economic activities and when new diagnostic tests and vaccinations were available. Most evaluations were either performed at the population-level or were specific to large-scale pilot events. For the latter, most events do not last for more than a day [63,64]. Furthermore, contact patterns at these events and the effect of interventions such as mask-wearing were not measured. On the other hand, contact studies in the pre-COVID era were conducted in schools, workplaces, and hospitals, and few were in large-scale events with high economic throughput, such as conferences [121–125]. The introduction of digital contact tracing devices now presents new opportunities to better understand contact patterns at large-scale events and how they affect disease transmission over longer timescales. Modelling disease transmission in these settings with combinations of outbreak control measures would also allow us to estimate the overall impact on the risk of transmission.

The high-resolution contact data collected during the COVID-19 pandemic and from previous studies also presents an opportunity to study the time-varying characteristics of contact networks and their implication to outbreak control measures. In particular, the overdispersion of the reproduction number was reported for COVID-19 and other diseases, but the ease of identifying individuals who account for a large number of transmission (e.g. 80% of the infected offspring) has yet to be investigated in real-world contact networks that change over time. Furthermore, temporal contact network properties are rarely normalised based on the population size, making comparisons between different contact settings difficult. Further analysis in this field would help public health officials to better calibrate outbreak control measures – targeted at individuals or sub-populations or untargeted mass intervention.

1.7 Aims

This PhD aims to understand the drivers of SARS-CoV-2 transmission patterns in Singapore and the effectiveness of outbreak control measures. Given the comprehensive genomic surveillance and epidemiological investigations of COVID-19 cases and contacts, Singapore was used as a case study. Nevertheless, disease-specific findings and derived epidemiological insights are expected to apply to other countries or settings.

1.8 Objectives

The objectives of the research presented in this PhD thesis are:

1a.   Quantify the relative role of border restrictions, case finding and contact tracing in controlling SARS-CoV-2 and estimate the number of missed COVID-19 infections.

1b.   Infer the outbreak dynamics in a country with an ongoing outbreak using data collected from travellers arriving from this country of interest.

2a.   Estimate the serial intervals observed in SARS-CoV-2 Delta variant cases and compare with those observed in the wild-type SARS-CoV-2 cases.

2b.   Evaluate the power to detect changes in the generation and serial intervals for a given sample size under varying pathogen characteristics, outbreak control measures and contact patterns.

3a.   Use high-resolution cruise ship contact networks to provide insights into the risk of SARS-CoV-2 transmission on cruises and identify optimal outbreak control strategies.

3b.   Understand how the structural feature of temporal contact networks affects the reliability of identifying potential superspreaders, the key drivers of transmission in different settings, and the impact on outbreak control resource planning.

1.9 References

1.   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;382: 1199–1207. doi:10.1056/NEJMoa2001316

2.   Karia R, Gupta I, Khandait H, Yadav A, Yadav A. COVID-19 and its Modes of Transmission. SN Compr Clin Med. 2020; 1–4. doi:10.1007/s42399-020-00498-4

3.   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 Med. 2020;17: e1003346. doi:10.1371/journal.pmed.1003346

4.   Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadas SM, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. Proceedings of the National Academy of Sciences. 2021;118: e2109229118. doi:10.1073/pnas.2109229118

5.   McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open. 2020;10: e039652. doi:10.1136/bmjopen-2020-039652

6.   Wu Y, Kang L, Guo Z, Liu J, Liu M, Liang W. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. JAMA Network Open. 2022;5: e2228008. doi:10.1001/jamanetworkopen.2022.28008

7.   Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine. 2020;382: 1708–1720. doi:10.1056/NEJMoa2002032

8.   World Health Organization. WHO COVID-19 Case definition. 22 Jul 2022 [cited 28 Oct 2023]. Available: https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-Surveillance_Case_Definition-2022.1

9.   Canas LS, Molteni E, Deng J, Sudre CH, Murray B, Kerfoot E, et al. Profiling post-COVID-19 condition across different variants of SARS-CoV-2: a prospective longitudinal study in unvaccinated wild-type, unvaccinated alpha-variant, and vaccinated delta-variant populations. The Lancet Digital Health. 2023;5: e421–e434. doi:10.1016/S2589-7500(23)00056-0

10.   Wei WE. Presymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. MMWR Morb Mortal Wkly Rep. 2020;69. doi:10.15585/mmwr.mm6914e1

11. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26: 672–675. doi:10.1038/s41591-020-0869-5

12. Byrne AW, McEvoy D, Collins AB, Hunt K, Casey M, Barber A, et al. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. BMJ Open. 2020;10: e039856. doi:10.1136/bmjopen-2020-039856

13. Ke R, Zitzmann C, Ho DD, Ribeiro RM, Perelson AS. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. Proceedings of the National Academy of Sciences. 2021;118: e2111477118. doi:10.1073/pnas.2111477118

14. Hakki S, Zhou J, Jonnerby J, Singanayagam A, Barnett JL, Madon KJ, et al. Onset and window of SARS-CoV-2 infectiousness and temporal correlation with symptom onset: a prospective, longitudinal, community cohort study. The Lancet Respiratory Medicine. 2022;10: 1061–1073. doi:10.1016/S2213-2600(22)00226-0

15. Lin Y, Wu P, Tsang TK, Wong JY, Lau EHY, Yang B, et al. Viral kinetics of SARS-CoV-2 following onset of COVID-19 in symptomatic patients infected with the ancestral strain and omicron BA.2 in Hong Kong: a retrospective observational study. The Lancet Microbe. 2023;4: e722–e731. doi:10.1016/S2666-5247(23)00146-5

16. World Health Organization. Tracking SARS-CoV-2 variants. [cited 10 Jul 2022]. Available: https://www.who.int/activities/tracking-SARS-CoV-2-variants

17. CDC. SARS-CoV-2 Variant Classifications and Definitions. In: Centers for Disease Control and Prevention [Internet]. 11 Feb 2020 [cited 28 Oct 2023]. Available: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html

18. Kissler SM, Fauver JR, Mack C, Tai CG, Breban MI, Watkins AE, et al. Viral Dynamics of SARS-CoV-2 Variants in Vaccinated and Unvaccinated Persons. New England Journal of Medicine. 2021;385: 2489–2491. doi:10.1056/NEJMc2102507

19. Bar-On YM, Goldberg Y, Mandel M, Bodenheimer O, Freedman L, Kalkstein N, et al. Protection of BNT162b2 Vaccine Booster against Covid-19 in Israel. New England Journal of Medicine. 2021;385: 1393–1400. doi:10.1056/NEJMoa2114255

20. 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;120: e2221887120. doi:10.1073/pnas.2221887120

21.   Park SW, Sun K, Champredon D, Li M, Bolker BM, Earn DJD, et al. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. Proceedings of the National Academy of Sciences. 2021;118: e2011548118. doi:10.1073/pnas.2011548118

22.   Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109. doi:10.1126/science.abc9004

23.   Diekmann O, Heesterbeek J a. P, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. Journal of The Royal Society Interface. 2009;7: 873–885. doi:10.1098/rsif.2009.0386

24.   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;20: 553–558. doi:10.1016/S1473-3099(20)30144-4

25.   Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. Science. 2009;324: 1557–1561. doi:10.1126/science.1176062

26.   Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. The Lancet Infectious Diseases. 2020;20: e238–e244. doi:10.1016/S1473-3099(20)30484-9

27.   Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. Science. 2003;300: 1966–1970. doi:10.1126/science.1086616

28.   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;300: 1961–1966. doi:10.1126/science.1086478

29.   Jing Q-L, Liu M-J, Zhang Z-B, Fang L-Q, Yuan J, Zhang A-R, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020;20: 1141–1150. doi:10.1016/S1473-3099(20)30471-0

30.   Madewell ZJ, Yang Y, Longini IM Jr, Halloran ME, Dean NE. Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. JAMA Network Open. 2020;3: e2031756. doi:10.1001/jamanetworkopen.2020.31756

31.   Cowling BJ, Chan KH, Fang VJ, Lau LLH, So HC, Fung ROP, et al. Comparative Epidemiology of Pandemic and Seasonal Influenza A in Households. New England Journal of Medicine. 2010;362: 2175–2184. doi:10.1056/NEJMoa0911530

32.   Zhang M, Xiao J, Deng A, Zhang Y, Zhuang Y, Hu T, et al. Transmission Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2 —

Guangdong Province, China, May–June 2021. CCDCW. 2021;3: 584–586. doi:10.46234/ccdcw2021.148

33. Liu Y, Rocklöv J. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. Journal of Travel Medicine. 2021;28: taab124. doi:10.1093/jtm/taab124

34. Madewell ZJ, Yang Y, Longini IM Jr, Halloran ME, Dean NE. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis. JAMA Network Open. 2022;5: e229317. doi:10.1001/jamanetworkopen.2022.9317

35. Ng OT, Koh V, Chiew CJ, Marimuthu K, Thevasagayam NM, Mak TM, et al. Impact of Delta Variant and Vaccination on SARS-CoV-2 Secondary Attack Rate Among Household Close Contacts. The Lancet Regional Health – Western Pacific. 2021;17. doi:10.1016/j.lanwpc.2021.100299

36. Singanayagam A, Hakki S, Dunning J, Madon KJ, Crone MA, Koycheva A, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. The Lancet Infectious Diseases. 2022;22: 183–195. doi:10.1016/S1473-3099(21)00648-4

37. Armando F, Beythien G, Kaiser FK, Allnoch L, Heydemann L, Rosiak M, et al. SARS-CoV-2 Omicron variant causes mild pathology in the upper and lower respiratory tract of hamsters. Nat Commun. 2022;13: 3519. doi:10.1038/s41467-022-31200-y

38. Halfmann PJ, Iida S, Iwatsuki-Horimoto K, Maemura T, Kiso M, Scheaffer SM, et al. SARS-CoV-2 Omicron virus causes attenuated disease in mice and hamsters. Nature. 2022;603: 687–692. doi:10.1038/s41586-022-04441-6

39. Hui KPY, Ho JCW, Cheung M, Ng K, Ching RHH, Lai K, et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. Nature. 2022;603: 715–720. doi:10.1038/s41586-022-04479-6

40. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. International Journal of Infectious Diseases. 2020;101: 138–148. doi:10.1016/j.ijid.2020.09.1464

41. 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 Medicine. 2020;18: 332. doi:10.1186/s12916-020-01790-9

42. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 2020;20: 669–677. doi:10.1016/S1473-3099(20)30243-7

43. Chan-Yeung M, Xu R-H. SARS: epidemiology. Respirology. 2003;8: S9–S14. doi:10.1046/j.1440-1843.2003.00518.x

44. Riley S, Kwok KO, Wu KM, Ning DY, Cowling BJ, Wu JT, et al. Epidemiological Characteristics of 2009 (H1N1) Pandemic Influenza Based on Paired Sera from a Longitudinal Community Cohort Study. PLOS Medicine. 2011;8: e1000442. doi:10.1371/journal.pmed.1000442

45. COVID-19 Forecasting Team. Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. The Lancet. 2022;399: 1469–1488. doi:10.1016/S0140-6736(21)02867-1

46. Yang Z-R, Jiang Y-W, Li F-X, Liu D, Lin T-F, Zhao Z-Y, et al. Efficacy of SARS-CoV-2 vaccines and the dose–response relationship with three major antibodies: a systematic review and meta-analysis of randomised controlled trials. The Lancet Microbe. 2023;4: e236–e246. doi:10.1016/S2666-5247(22)00390-1

47. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). 30 Jan 2020 [cited 19 Sep 2021]. Available: 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)

48. World Tourism Organization. COVID-19 related travel restrictions. a global review for tourism. Second report. 28 Apr 2020 [cited 25 Mar 2021]. Available: https://www.unwto.org/news/covid-19-travel-restrictions

49. Russell TW, Wu JT, Clifford S, Edmunds WJ, Kucharski AJ, Jit M. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. The Lancet Public Health. 2021;6: e12–e20. doi:10.1016/S2468-2667(20)30263-2

50. Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. The Lancet Infectious Diseases. 2020;20: 1151–1160. doi:10.1016/S1473-3099(20)30457-6

51. Ng Y, Li Z, Chua YX, Chaw WL, Zheng Z, Er B, et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore — January 2–February 29, 2020. MMWR Morb Mortal Wkly Rep. 2020;69. doi:10.15585/mmwr.mm6911e1

52. Yong SEF, Anderson DE, Wei WE, Pang J, Chia WN, Tan CW, et al. Connecting clusters of COVID-19: an epidemiological and serological investigation. The Lancet Infectious Diseases. 2020;20: 809–815. doi:10.1016/S1473-3099(20)30273-5

53.   Endo A, Leclerc QJ, Knight GM, Medley GF, Atkins KE, Funk S, et al. Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks. Wellcome Open Res. 2021;5: 239. doi:10.12688/wellcomeopenres.16344.3

54.   Liu Y, Morgenstern C, Kelly J, Lowe R, Munday J, Villabona-Arenas CJ, et al. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. BMC Medicine. 2021;19: 40. doi:10.1186/s12916-020-01872-8

55.   Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584: 257–261. doi:10.1038/s41586-020-2405-7

56.   Williams SN, Armitage CJ, Tampe T, Dienes K. Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: a UK-based focus group study. BMJ Open. 2020;10: e039334. doi:10.1136/bmjopen-2020-039334

57.   Ministry of Health, Singapore. MOH pandemic readiness and response plan for influenza and other acute respiratory diseases (revised April 2014). 2014 [cited 26 Sep 2021]. Available: https://www.moh.gov.sg/docs/librariesprovider5/diseases-updates/interim-pandemic-plan-public-ver-_april-2014.pdf

58.   El Sahly HM, Baden LR, Essink B, Doblecki-Lewis S, Martin JM, Anderson EJ, et al. Efficacy of the mRNA-1273 SARS-CoV-2 Vaccine at Completion of Blinded Phase. New England Journal of Medicine. 2021;385: 1774–1785. doi:10.1056/NEJMoa2113017

59.   Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine. 2020;383: 2603–2615. doi:10.1056/NEJMoa2034577

60.   Zheng C, Shao W, Chen X, Zhang B, Wang G, Zhang W. Real-world effectiveness of COVID-19 vaccines: a literature review and meta-analysis. International Journal of Infectious Diseases. 2022;114: 252–260. doi:10.1016/j.ijid.2021.11.009

61.   World Health Organization. Diagnostic testing for SARS-CoV-2. 11 Sep 2020 [cited 19 Sep 2021]. Available: https://www.who.int/publications-detail-redirect/diagnostic-testing-for-sars-cov-2

62.   Hellewell J, Russell TW, Matthews R, Severn A, Adam S, Enfield L, et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. BMC Medicine. 2021;19: 106. doi:10.1186/s12916-021-01982-x

63.   Government of the United Kingdom. Information on the Events Research Programme. In: GOV.UK [Internet]. 20 Aug 2021 [cited 19 Sep 2021].

Available: https://www.gov.uk/government/publications/information-on-the-events-research-programme/information-on-the-events-research-programme

64. Ministry of Health, Singapore. Safeguarding lives and livelihoods. 10 Nov 2020 [cited 29 Oct 2021]. Available: https://www.moh.gov.sg/news-highlights/details/safeguarding-lives-and-livelihoods

65. European Centre for Disease Prevention and Control. Options for the use of rapid antigen tests for COVID-19 in the EU/EEA and the UK. In: TECHNICAL REPORT [Internet]. 19 Nov 2020 [cited 19 Sep 2021]. Available: https://www.ecdc.europa.eu/sites/default/files/documents/Options-use-of-rapid-antigen-tests-for-COVID-19_0.pdf

66. Centres for Disease Control and Prevention. Interim Guidance for Antigen Testing for SARS-CoV-2. In: Centers for Disease Control and Prevention [Internet]. 11 Feb 2020 [cited 19 Sep 2021]. Available: https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html

67. Public Health England Porton Down, University of Oxford. Preliminary report from the Joint PHE Porton Down & University of Oxford SARS-CoV-2 test development and validation cell: Rapid evaluation of Lateral Flow Viral Antigen detection devices (LFDs) for mass community testing. 8 Nov 2020 [cited 19 Sep 2021]. Available: https://www.ox.ac.uk/sites/files/oxford/media_wysiwyg/UK%20evaluation_PHE%20Porton%20Down%20%20University%20of%20Oxford_final.pdf

68. Ministry of Health, Singapore. Updates on local situation, testing and vaccination efforts in transition towards COVID resilience. 27 Aug 2021 [cited 19 Sep 2021]. Available: https://www.moh.gov.sg/news-highlights/details/updates-on-local-situation-testing-and-vaccination-efforts-in-transition-towards-covid-resilience

69. Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database of Systematic Reviews. 2021 [cited 12 Jun 2021]. doi:10.1002/14651858.CD013705.pub2

70. Smerczak E. SARS-CoV-2 Antibody Testing: Where Are We Now? Laboratory Medicine. 2022;53: e19–e29. doi:10.1093/labmed/lmab061

71. Borremans B, Gamble A, Prager K, Helman SK, McClain AM, Cox C, et al. Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset. Malagón T, Davenport MP, Mina MJ, editors. eLife. 2020;9: e60122. doi:10.7554/eLife.60122

72. Gallais F, Gantner P, Bruel T, Velay A, Planas D, Wendling M-J, et al. Evolution of antibody responses up to 13 months after SARS-CoV-2 infection and risk of reinfection. EBioMedicine. 2021;71. doi:10.1016/j.ebiom.2021.103561

73. Post N, Eddy D, Huntley C, Schalkwyk MCI van, Shrotri M, Leeman D, et al. Antibody response to SARS-CoV-2 infection in humans: A systematic review. PLOS ONE. 2020;15: e0244126. doi:10.1371/journal.pone.0244126

74. Srivastava K, Carreño JM, Gleason C, Monahan B, Singh G, Abbad A, et al. Kinetics and durability of humoral responses to SARS-CoV-2 infection and vaccination. medRxiv; 2023. p. 2023.08.26.23294679. doi:10.1101/2023.08.26.23294679

75. 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;172: 577–582. doi:10.7326/M20-0504

76. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. PLOS ONE. 2020;15: e0234765. doi:10.1371/journal.pone.0234765

77. Russell TW, Hellewell J, Jarvis CI, Zandvoort K van, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. Eurosurveillance. 2020;25: 2000256. doi:10.2807/1560-7917.ES.2020.25.12.2000256

78. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368: eabb6936. doi:10.1126/science.abb6936

79. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Eurosurveillance. 2020;25: 2000257. doi:10.2807/1560-7917.ES.2020.25.17.2000257

80. Sender R, Bar-On YM, Park SW, Noor E, Dushoff J, Milo R. The unmitigated profile of COVID-19 infectiousness. 2021 Nov p. 2021.11.17.21266051. doi:10.1101/2021.11.17.21266051

81. Chen D, Lau Y-C, Xu X-K, Wang L, Du Z, Tsang TK, et al. Inferring time-varying generation time, serial interval, and incubation period distributions for COVID-19. Nat Commun. 2022;13: 7727. doi:10.1038/s41467-022-35496-8

82. Kissler SM, Fauver JR, Mack C, Olesen SW, Tai C, Shiue KY, et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. PLOS Biology. 2021;19: e3001333. doi:10.1371/journal.pbio.3001333

83. Peto T, Affron D, Afrough B, Agasu A, Ainsworth M, Allanson A, et al. COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national

systematic evaluation of sensitivity and specificity for mass-testing. eClinicalMedicine. 2021;36. doi:10.1016/j.eclinm.2021.100924

84. University of Oxford. Oxford University and PHE confirm lateral flow tests show high specificity and are effective at identifying most individuals who are infectious | University of Oxford. 11 Nov 2020 [cited 28 Jan 2024]. Available: https://www.ox.ac.uk/news/2020-11-11-oxford-university-and-phe-confirm-lateral-flow-tests-show-high-specificity-and-are

85. Pung R, Park M, Cook AR, Lee VJ. Age-related risk of household transmission of COVID-19 in Singapore. Influenza and Other Respiratory Viruses. 2021;15: 206–208. doi:10.1111/irv.12809

86. Jørgensen SB, Nygård K, Kacelnik O, Telle K. Secondary Attack Rates for Omicron and Delta Variants of SARS-CoV-2 in Norwegian Households. JAMA. 2022;327: 1610–1611. doi:10.1001/jama.2022.3780

87. Ng OT, Marimuthu K, Koh V, Pang J, Linn KZ, Sun J, et al. SARS-CoV-2 seroprevalence and transmission risk factors among high-risk close contacts: a retrospective cohort study. The Lancet Infectious Diseases. 2021;21: 333–343. doi:10.1016/S1473-3099(20)30833-1

88. Dessie ZG, Zewotir T. Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients. BMC Infectious Diseases. 2021;21: 855. doi:10.1186/s12879-021-06536-3

89. Vardavas CI, Mathioudakis AG, Nikitara K, Stamatelopoulos K, Georgiopoulos G, Phalkey R, et al. Prognostic factors for mortality, intensive care unit and hospital admission due to SARS-CoV-2: a systematic review and meta-analysis of cohort studies in Europe. European Respiratory Review. 2022;31. doi:10.1183/16000617.0098-2022

90. Lu FS, Nguyen AT, Link NB, Molina M, Davis JT, Chinazzi M, et al. Estimating the cumulative incidence of COVID-19 in the United States using influenza surveillance, virologic testing, and mortality data: Four complementary approaches. PLOS Computational Biology. 2021;17: e1008994. doi:10.1371/journal.pcbi.1008994

91. Reed C, Chaves SS, Kirley PD, Emerson R, Aragon D, Hancock EB, et al. Estimating Influenza Disease Burden from Population-Based Surveillance Data in the United States. PLOS ONE. 2015;10: e0118369. doi:10.1371/journal.pone.0118369

92. Centres for Disease Control and Prevention. Estimated COVID-19 burden. 16 Nov 2021 [cited 28 Jan 2024]. Available: https://stacks.cdc.gov/view/cdc/117147

93. Kucharski AJ, Chung K, Aubry M, Teiti I, Teissier A, Richard V, et al. Real-time surveillance of international SARS-CoV-2 prevalence using systematic traveller arrival screening: An observational study. PLOS Medicine. 2023;20: e1004283. doi:10.1371/journal.pmed.1004283

94. Geismar C, Nguyen V, Fragaszy E, Shrotri M, Navaratnam AMD, Beale S, et al. Bayesian reconstruction of SARS-CoV-2 transmissions highlights substantial proportion of negative serial intervals. Epidemics. 2023;44: 100713. doi:10.1016/j.epidem.2023.100713

95. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health. 2020;8: e488–e496. doi:10.1016/S2214-109X(20)30074-7

96. Fyles M, Fearon E, Overton C, null null, Wingfield T, Medley GF, et al. Using a household-structured branching process to analyse contact tracing in the SARS-CoV-2 pandemic. Philosophical Transactions of the Royal Society B: Biological Sciences. 2021;376: 20200267. doi:10.1098/rstb.2020.0267

97. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat Med. 2020;26: 1714–1719. doi:10.1038/s41591-020-1092-0

98. Pung R, Cook AR, Chiew CJ, Clapham HE, Sun Y, Li Z, et al. Effectiveness of Containment Measures Against COVID-19 in Singapore: Implications for Other National Containment Efforts. Epidemiology. 2021;32: 79–86. doi:10.1097/EDE.0000000000001257

99. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. PNAS. 2011;108: 2825–2830. doi:10.1073/pnas.1008895108

100. Cauchemez S, Nouvellet P, Cori A, Jombart T, Garske T, Clapham H, et al. Unraveling the drivers of MERS-CoV transmission. PNAS. 2016;113: 9081–9086. doi:10.1073/pnas.1519235113

101. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, et al. Household Transmission of 2009 Pandemic Influenza A (H1N1) Virus in the United States. New England Journal of Medicine. 2009;361: 2619–2627. doi:10.1056/NEJMoa0905498

102. Bhatt S, Ferguson N, Flaxman S, Gandy A, Mishra S, Scott JA. Semi-mechanistic Bayesian modelling of COVID-19 with renewal processes. Journal of the Royal Statistical Society Series A: Statistics in Society. 2023;186: 601–615. doi:10.1093/jrsssa/qnad030

103. 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;5: e375–e385. doi:10.1016/S2468-2667(20)30133-X

104. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-

19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health. 2020;5: e261–e270. doi:10.1016/S2468-2667(20)30073-6

105. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 2020;395: 689–697. doi:10.1016/S0140-6736(20)30260-9

106. Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science. 2020;369: 846–849. doi:10.1126/science.abc6810

107. Pakkanen MS, Miscouridou X, Penn MJ, Whittaker C, Berah T, Mishra S, et al. Unifying incidence and prevalence under a time-varying general branching process. J Math Biol. 2023;87: 35. doi:10.1007/s00285-023-01958-w

108. Champredon D, Dushoff J, Earn DJD. Equivalence of the Erlang-Distributed SEIR Epidemic Model and the Renewal Equation. SIAM J Appl Math. 2018;78: 3258–3278. doi:10.1137/18M1186411

109. 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. 2021 Oct p. 2021.10.21.21265216. doi:10.1101/2021.10.21.21265216

110. 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;26: 1616–1622. doi:10.1038/s41591-020-1036-8

111. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: An agent-based model of COVID-19 dynamics and interventions. PLOS Computational Biology. 2021;17: e1009149. doi:10.1371/journal.pcbi.1009149

112. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious Diseases. 2020;20: 678–688. doi:10.1016/S1473-3099(20)30162-6

113. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. J R Soc Interface. 2007;4: 879–891. doi:10.1098/rsif.2007.1100

114. Kerr CC, Mistry D, Stuart RM, Rosenfeld K, Hart GR, Núñez RC, et al. Controlling COVID-19 via test-trace-quarantine. Nat Commun. 2021;12: 2993. doi:10.1038/s41467-021-23276-9

115. Cori A, Ferguson NM, Fraser C, Cauchemez S. New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics | American Journal of Epidemiology | Oxford Academic. Am J Epidemiol. 2013;178: 1505–1512. doi:10.1093/aje/kwt133

116. 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;5: 112. doi:10.12688/wellcomeopenres.16006.2

117. Imai N, Dorigatti I, Cori A, Riley S, Ferguson NM. Report 1 - Estimating the potential total number of novel Coronavirus (2019-nCoV) cases in Wuhan City, China | Faculty of Medicine | Imperial College London. 17 Jan 2020 [cited 27 Sep 2021]. Available: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-1-case-estimates-of-covid-19/

118. Bloomberg News. China Covid Cases Surge Infecting 37 Million People a Day, According to Estimate - Bloomberg. 23 Dec 2022. Available: https://www.bloomberg.com/news/articles/2022-12-23/china-estimates-covid-surge-is-infecting-37-million-people-a-day

119. Dushoff J, Park SW. Speed and strength of an epidemic intervention. Proceedings of the Royal Society B: Biological Sciences. 2021;288: 20201556. doi:10.1098/rspb.2020.1556

120. Ryu S, Kim D, Lim J-S, Ali ST, Cowling BJ. Serial Interval and Transmission Dynamics during SARS-CoV-2 Delta Variant Predominance, South Korea - Volume 28, Number 2—February 2022 - Emerging Infectious Diseases journal - CDC. [cited 10 Jul 2022]. doi:10.3201/eid2802.211774

121. Fournet J, Barrat A. Contact Patterns among High School Students. PLOS ONE. 2014;9: e107878. doi:10.1371/journal.pone.0107878

122. Mastrandrea R, Fournet J, Barrat A. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. PLOS ONE. 2015;10: e0136497. doi:10.1371/journal.pone.0136497

123. Génois M, Barrat A. Can co-location be used as a proxy for face-to-face contacts? EPJ Data Sci. 2018;7: 1–18. doi:10.1140/epjds/s13688-018-0140-1

124. Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science. 2015;3: 326–347. doi:10.1017/nws.2015.10

125. Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, et al. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. PLOS ONE. 2013;8: e73970. doi:10.1371/journal.pone.0073970

2 Relative role of border restrictions, case finding and contact tracing in controlling SARS-CoV-2 in the presence of undetected transmission: a mathematical modelling study

In early 2020, countries were implementing a series of outbreak control measures to minimise the introduction and transmission of SARS-CoV-2. These include border restrictions, case finding, contact tracing and other non-targeted population-wide measures such as mask-wearing and physical distancing. However, due to the lack of granular data, most countries have been characterising the effectiveness of outbreak control measures by estimating the changes to the effective reproduction number [1–3]. In this retrospective analysis, I aimed to disentangle the effects of each type of outbreak control measure. I used a comprehensive epidemiological investigation dataset which contains a line list of all the notified COVID-19 cases in Singapore with key information on their travel history, potential sources of infection and date of key events (e.g. arrival (if any), symptoms onset, notification, isolation or quarantine). Incorporating this data with a branching process model, I reconstructed the outbreak dynamics to estimate the effectiveness of each outbreak control measure and the extent of the missed infections at different stages of the COVID-19 pandemic in Singapore.

This paper was published in BMC Medicine in March 2023 [4]. The supplementary information of the paper is in Appendix B.

References

1.   Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584: 257–261. doi:10.1038/s41586-020-2405-7

2.   Liu Y, Morgenstern C, Kelly J, Lowe R, Munday J, Villabona-Arenas CJ, et al. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. BMC Medicine. 2021;19: 40. doi:10.1186/s12916-020-01872-8

3.   Chiu WA, Fischer R, Ndeffo-Mbah ML. State-level needs for social distancing and contact tracing to contain COVID-19 in the United States. Nat Hum Behav. 2020;4: 1080–1090. doi:10.1038/s41562-020-00969-7

4.   Pung R, Clapham HE, Russell TW, Lee VJ, Kucharski AJ. Relative role of border restrictions, case finding and contact tracing in controlling SARS-CoV-2

in the presence of undetected transmission: a mathematical modelling study. BMC Med. 2023;21: 1–17. doi:10.1186/s12916-023-02802-0

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

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 <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006052	**Title**	Ms
First Name(s)	Rachael		
Surname/Family Name	Pung		
Thesis Title	COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore		
Primary Supervisor	Adam J. 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?	BMC Medicine		
When was the work published?	16 March 2023		
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.

Improving health worldwide   **www.lshtm.ac.uk**

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, investigation, visualisation, and writing of the original draft. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as the original author(s) and the source are credited. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-023-02802-0

SECTION E

Student Signature	*Andrael*
Date	7 March 2024

Supervisor Signature	*Ada Kwhali*
Date	13 March 2024

BMC Medicine

RESEARCH ARTICLE

Open Access

Relative role of border restrictions, case finding and contact tracing in controlling SARS-CoV-2 in the presence of undetected transmission: a mathematical modelling study

Rachael Pung[1,2*] , Hannah E. Clapham[3], Timothy W. Russell[2], CMMID COVID-19 Working Group[2], Vernon J. Lee[1,3†] and Adam J. Kucharski[2†]

Abstract

Background Understanding the overall effectiveness of non-pharmaceutical interventions to control the COVID-19 pandemic and reduce the burden of disease is crucial for future pandemic planning. However, quantifying the effectiveness of specific control measures and the extent of missed infections, in the absence of early large-scale serological surveys or random community testing, has remained challenging.

Methods Combining data on notified local COVID-19 cases with known and unknown sources of infections in Singapore with a branching process model, we reconstructed the incidence of missed infections during the early phase of the wild-type SARS-CoV-2 and Delta variant transmission. We then estimated the relative effectiveness of border control measures, case finding and contact tracing when there was no or low vaccine coverage in the population. We compared the risk of ICU admission and death between the wild-type SARS-CoV-2 and the Delta variant in notified cases and all infections.

Results We estimated strict border control measures were associated with 0.2 (95% credible intervals, CrI 0.04–0.8) missed imported infections per notified case between July and December 2020, a decline from around 1 missed imported infection per notified case in the early phases of the pandemic. Contact tracing was estimated to identify 78% (95% CrI 62–93%) of the secondary infections generated by notified cases before the partial lockdown in Apr 2020, but this declined to 63% (95% CrI 56–71%) during the lockdown and rebounded to 78% (95% CrI 58–94%) during reopening in Jul 2020. The contribution of contact tracing towards overall outbreak control also hinges on ability to find cases with unknown sources of infection: 42% (95% CrI 12–84%) of such cases were found prior to the lockdown; 10% (95% CrI 7–15%) during the lockdown; 47% (95% CrI 17–85%) during reopening, due to increased testing capacity and health-seeking behaviour. We estimated around 63% (95% CrI 49–78%) of the wild-type SARS-CoV-2 infections were undetected during 2020 and around 70% (95% CrI 49–91%) for the Delta variant in 2021.

Conclusions Combining models with case linkage data enables evaluation of the effectiveness of different components of outbreak control measures, and provides more reliable situational awareness when some cases are missed.

[†]Vernon J. Lee and Adam J. Kucharski contributed equally to this work.

*Correspondence:
Rachael Pung
rachael.pung@lshtm.ac.uk
Full list of author information is available at the end of the article

BMC

Using such approaches for early identification of the weakest link in containment efforts could help policy makers to better redirect limited resources to strengthen outbreak control.

Keywords Border restrictions, Case finding, Contact tracing, Mathematical modelling, SARS-CoV-2, Undetected

Background

The use of multiple outbreak control measures in the early phases of the COVID-19 pandemic was resource intensive and disruptive, but essential to minimise the loss of lives [1, 2]. Measures such as case finding at the borders and healthcare touchpoints allow health authorities to assess the extent of disease importation and undetected spread in the community. Furthermore, contact tracing around notified cases can identify potential transmission routes and hence new cases [3, 4]. When multiple control measures are implemented together, understanding the effectiveness of each measure enables public health authorities to focus on the most effective measures when resources are limited and to minimise interruption to economic and social activities. Studies typically evaluate the collective effectiveness of country or region-specific COVID-19 outbreak control measures by measuring changes to the reproduction number using overall observed case incidence [1, 5–12] or only focus on the impact of specific interventions using outbreak data [13, 14]. If analysis could disentangle the observed and unobserved transmission dynamics, it would therefore be possible to obtain higher resolution insights on the effects of each outbreak control measure.

Transmission chains from outbreak clusters have been used to characterise the reproduction number of infectious diseases other than COVID-19 and the relative contribution of different transmission routes (e.g. imported or environmental introduction vs community) to the overall spread [15–18]. However, these studies typically do not account for the role of missed infections (e.g. asymptomatic or mildly symptomatic infections) in influencing the effectiveness of outbreak control measures. To our knowledge, the use of data on these transmission linkages to estimate the burden of infection for SARS-CoV-2 at the population level has yet to be documented. The extent of missed infections in the COVID-19 pandemic was commonly assessed via population-wide seroprevalence surveys [19, 20], excess mortality studies [21], random community testing [22] or behavioural surveys [23, 24]. However, during the initial phases of an outbreak of a novel pathogen, serological assays to measure the disease prevalence are generally not available. Moreover, these methods do not provide assessment on the extent of missed cases at the borders. Thus, methods to address these challenges and provide a more complete view of the outbreak are necessary.

With a population of 5.7 million inhabitants, Singapore was one of the first countries to report SARS-CoV-2 infections outside of mainland China at the beginning of the COVID-19 pandemic. The Ministry of Health monitored the daily incidence of imported, and linked and unlinked local COVID-19 cases and collected extensive information on the epidemiological events associated with each case (e.g. time of arrival, symptoms onset, notification, isolation or quarantine). In this study, we reconstructed the pandemic trajectory in Singapore and estimated the effectiveness of various outbreak control measures (Table 1) by combining the observed COVID-19 cases with a mathematical model. As countries redesign surveillance systems for future pandemics, this modelling framework has the potential to inform how the collection of different data fields can shape our understanding of disease transmission in the early phases of a pandemic.

Methods

Data

Cases of COVID-19 (confirmed with a respiratory sample positive for SARS-CoV-2 by PCR [25] were identified through case finding and contact tracing (Table 1). Extensive epidemiological investigations were conducted for each case to establish their exposure history and to classify them as a local linked case if a case has at least one known source of infection or a local unlinked case if a case has an unknown source of infection.

In this study, we used COVID-19 cases notified to the Ministry of Health, Singapore from Jan 23 to Dec 31, 2020, and from Apr 1 to Aug 18, 2021, in Singapore. The former time period precedes the detection and surge in cases infected by SARS-CoV-2 Variants of Concern in Singapore [26], while community spread in the latter time period was dominated by the SARS-CoV-2 Delta variant [27]. Data from Jan to Mar 2021 was not used as the COVID-19 incidence in the community was too low (i.e. less than 5 cases per day) for any meaningful analysis.

For the two time periods of study, all confirmed cases were conveyed to secured isolation facilities and discharged after 21 days from the date of confirmation if assessed to be clinically well, or with sequential negative tests. Cases occurring in persons residing in a foreign-worker dormitory and notified from Apr 7 to Oct 31, 2020, were omitted from the analysis as these dormitories were placed under lockdown for an extended period of time. As workers were subjected to movement

Table 1 Outbreak control measures in Singapore. Observed case data were used to estimate the effectiveness of each measure. Cases are defined as infected individuals that tested positive and are notified, while infections include all notified and missed infected individuals

Control measure (Aims)	Description	Observed data (●) and modelled outputs (◆)
Border control (*Minimise disease introduction into community*)	■ Limiting the number of incoming travellers from countries with ongoing outbreaks ■ Quarantine or restricting movement of incoming travellers	◆ Number of missed imported infections
Case finding (*Targeted testing at known or potential source(s) of infection*)	■ Testing of symptomatic travellers upon arrival or when they developed symptoms during quarantine ■ Testing regime for non-symptomatic travellers	● Imported case data
	■ Testing of suspect cases (e.g. persons with clinical signs and symptoms suggestive of pneumonia or severe respiratory infection, persons with acute respiratory infection and travel history to regions with ongoing outbreak) ■ Routine testing of high-risk populations (e.g. healthcare workers, nursing home residents) ■ Ad-hoc testing during cluster outbreak investigations	● Local unlinked case data ◆ Effectiveness of case finding
Contact tracing (*Targeted testing at potential routes of infection*)	■ Interviewing COVID-19 cases or use of Bluetooth contact tracing devices to identify close contacts ■ Testing of symptomatic contacts ■ Testing of contacts before the end of their quarantine	● Local linked case data ◆ Effectiveness of contact tracing
Use of other non-pharmaceutical interventions and vaccines (*Untargeted community- or population-level preventive measures*)	■ Physical distancing ■ School and venue closure ■ Large-scale population movement restrictions and the corresponding need to work-from-home ■ Population-wide face mask usage ■ Pre-event testing/vaccination ■ Accelerated development and roll-out of COVID-19 vaccines (primary doses and boosters) with priority given to frontline workers and the elderly before progressively offered to younger age groups	◆ Average number of secondary cases generated by a single infectious individual over the course of the entire infectious period (i.e. R)

restrictions, there was a minimal opportunity to interact with the community and hence they were assumed to be incapable of driving community-level transmission. Furthermore, around 0.2% of the confirmed cases occurred in healthcare workers providing care to confirmed cases. As these cases were not community-acquired infections, they were omitted from the analysis.

Transmission model

Using the notified linked and unlinked cases, we fitted a branching process model using a Bayesian framework to estimate the effectiveness of different control measures (Fig. 1, Table 1 and Additional file 1: Table S1), such as (i) border control measures (based on the extent of missed imported infections, ρ), (ii) case finding (ϵ_{cf}), (iii) contact tracing (ϵ_{ct}), (iv) other outbreak control measures (based on the community reproduction number, R), and estimate the incidence of missed COVID-19 infections.

For a single infectious individual, the mean rate at which an individual infects others (i.e. infectiousness) τ time since infection, $\beta(\tau)$, can be expressed as a function of the generation interval, $\omega(\tau)$ and the reproduction number, R:

$$\beta(\tau) = \omega(\tau)R \qquad (1)$$

$\omega(\tau)$ is the probability density function of the time from infection in one case to another and is often approximated using serial intervals (i.e. time from symptom onset in one case to another). We modelled $\omega(\tau)$ as a lognormal distribution with mean 5.9 and standard deviation 2.4, approximated using published estimates of the observed serial interval for COVID-19 during the early stages of the outbreak when the generation interval and the observed serial interval had yet to reduce substantially due to the influence of non-pharmaceutical interventions [28–30]. With the
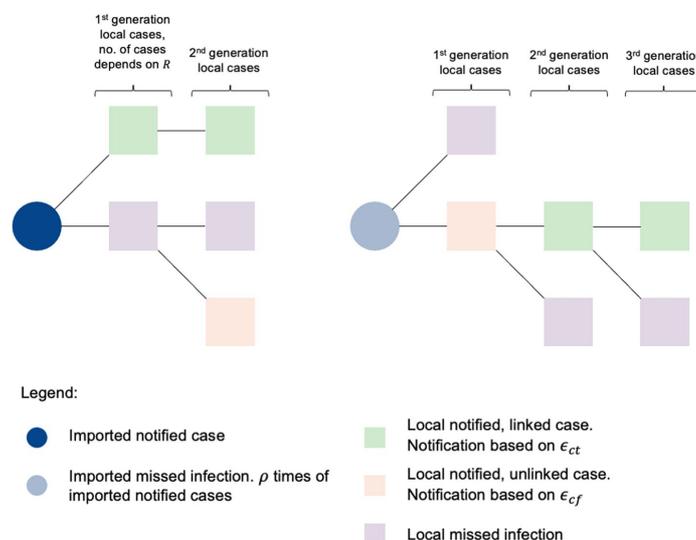
Fig. 1 Branching process model and model parameters

exception of $\omega(\tau)$, all other probability density functions are denoted as f in the subsequent sections.

R is defined as the average number of secondary cases generated by a single infectious individual over the course of the entire infectious period (i.e. no truncation of the infectious period due to individually-targeted measures such as quarantine or isolation). Furthermore, the effects of various outbreak control measures not related to case finding or contact tracing (e.g. social distancing, vaccination) were collectively modelled within R (Table 1).

COVID-19 infections at calendar time t were either notified, n, or missed, m. These infections can be further stratified based on their sources of infection and denoted by subscript im for imported infections, cf for local unlinked infections identified through case finding, and ct for local linked infections identified through contact tracing. Early in the pandemic, COVID-19 was introduced in most countries by the arrival of infectious travellers at time $t-a$ who could be notified to the public health authorities, $n_{im}(t-a)$ or missed, $m_{im}(t-a)$. Beside the time of arrival, the time of symptoms onset, $t-s$, of a notified case is often observed but not the time of infection, $t-\tau$. Estimating the time of infection of notified imported cases would allow us to estimate the potential number of local infections generated by these cases since their time of arrival. Thus, the time series of notified imported cases by the time of infection and arrival is defined as:

$$n_{im}(t-\tau, t-a) = n_{im}(t-a)f_a[(t-\tau)-(t-a)], \qquad for \quad t-\tau \le t-a \tag{2}$$

$$f_a(x) = \int_0^\infty f_s(u)f_{sa}(x-u)\,du \tag{3}$$

where $f_a(x)$ is the probability density function of arriving to a country x time since infection and u is the variable of integration. $f_a(x)$ is derived, by convolving the incubation period for SARS-CoV-2 infection x time since infection, $f_s(x)$, and the observed distribution of time from symptoms onset to arrival, $f_{sa}[(t-a)-(t-s)]$ (Eq. 3). $s-a$ is the time delay to developing symptoms since arrival and $s-a>0$ implies that case was symptomatic before arrival and vice versa. We modelled $f_s(\tau)$ as a lognormal distribution with mean 5.8 days and standard deviation 3.1 days for wild-type SARS-CoV-2 [31] and mean 4 days and standard deviation 0.4 days for the Delta variant [32].

Missed imported infections were modelled to scale by a factor, ρ, of notified imported cases (Eq. 4). Both notified and missed imported infections were capable of generating community infections from their time of arrival to isolation or the end of their infectiousness respectively. Community infections, denoted by subscript c, infected on day t were either notified, $n_c(t)$, through varying effectiveness of case finding and contact tracing or missed, $m_c(t)$ in Eqs. (5) and (6).

$$m_{im}(t - \tau, t - a) = \rho \, n_{im}(t - \tau, t - a) \qquad (4)$$

$$n_c(t) = \epsilon_{ct} \int_0^\infty \int_0^\tau n_{im}(t - \tau, t - a) F_{h'}(\tau)\beta(\tau) \, da \, d\tau + \epsilon_{cf} \int_0^\infty \int_0^\tau m_{im}(t - \tau, t - a)\beta(\tau) \, da \, d\tau = n_{ct}(t) + n_{cf}(t) \quad (5)$$

$$m_c(t) = (1 - \epsilon_{ct}) \int_0^\infty \int_0^\tau n_{im}(t - \tau, t - a) F_{h'}(\tau)\beta(\tau) \, da \, d\tau + \left(1 - \epsilon_{cf}\right) \int_0^\infty \int_0^\tau m_{im}(t - \tau, t - a)\beta(\tau) \, da \, d\tau \qquad (6)$$

The first component of both Eqs. (5) and (6) is the community infections generated by notified imported cases while the second component of both equations is the community infections generated by missed imported infectors. The effectiveness of contact tracing in identifying new secondary cases linked to notified cases and the effectiveness of case finding in identifying new cases that are not linked to any existing cases are ϵ_{ct} and ϵ_{cf}, respectively. $F_{h'}(\tau)$ is the cumulative probability that an imported case is at large in the community τ time since infection and prior to notification (and hence isolation in a hospital or managed facility) (Eq. 7). Using symptomatic cases, we estimate the probability density function of an imported case being isolated x time since infection, $f_{h'}(x)$, by convolving the incubation period for SARS-CoV-2 infection and the observed time from symptoms onset to isolation of imported cases, $f_{sh'}[(t - h') - (t - s)]$ (Eq. 8). $s - h'$ is the time delay to developing symptoms since isolation in imported cases and $s - h' > 0$ implies that the case was symptomatic before isolation and vice versa.

$$F_{h'}(\tau) = 1 - \int_0^\tau f_{h'}(u)du \qquad (7)$$

$$f_{h'}(x) = \int_0^\infty f_s(u)f_{sh'}(x - u) \, du \qquad (8)$$

Subsequent generations of community infections follow the same principles in Eqs. (5) and (6) as follows in Eqs. (9) and (10). $F_h(\tau)$ is the cumulative probability that a local case is at large in the community τ time since infection and prior to notification (and hence isolation in a hospital or managed facility) and derived using the observed time from symptoms onset to isolation in local cases.

nity case $R_n = \int_0^\infty F_h(\tau)\beta(\tau)d\tau$ is lower than that of a missed case $R_m = R = \int_0^\infty \beta(\tau)d\tau$. Overall, the effective reproduction number in the community, R_{eff}, is an aggregate measure of both R_n and R_m whose value corresponds to the dominant eigenvalue of the next generation matrix, K, as follows:

$$K = \begin{bmatrix} (1 - \epsilon_{cf}) \int_0^\infty \beta(\tau) \, d\tau & (1 - \epsilon_{ct}) \int_0^\infty F_h(\tau)\beta(\tau) \, d\tau \\ \epsilon_{cf} \int_0^\infty \beta(\tau) \, d\tau & \epsilon_{ct} \int_0^\infty F_h(\tau)\beta(\tau) \, d\tau \end{bmatrix} \qquad (11)$$

Model fitting

We assumed the infection was first introduced into a naïve population by imported cases and disease transmission was simulated over calendar time through a branching process using Eqs. (5) to (10). Early isolation of notified infected individuals and modelled outbreak control measures such as border controls (ρ), case finding (ϵ_{cf}), contact tracing (ϵ_{ct}), other non-pharmaceutical interventions (R) (Table 1) would influence the trajectory of the notified cases and the expected incidence was fitted using a negative binomial likelihood to the observed daily incidence of linked and unlinked local COVID-19 cases isolated in hospitals or managed facilities (i.e. $i_{ct}(t)$ and $i_{cf}(t)$). The modelled linked and unlinked cases isolated at time t are defined as:

$$h_{ct}(t) = \int_0^\infty n_{ct}(t - \tau)f_h(\tau)d\tau \qquad (12)$$

$$h_{cf}(t) = \int_0^\infty n_{cf}(t - \tau)f_h(\tau)d\tau \qquad (13)$$

We defined the likelihood of observing unlinked and linked cases at the time of isolation as:

$$n_c(t) = \epsilon_{ct} \int_0^\infty n_c(t - \tau)F_h(\tau)\beta(\tau) \, d\tau + \epsilon_{cf} \int_0^\infty m_c(t - \tau)\beta(\tau) \, da \, d\tau \qquad (9)$$

$$m_c(t) = (1 - \epsilon_{ct}) \int_0^\infty n_c(t - \tau)F_h(\tau)\beta(\tau) \, d\tau + \left(1 - \epsilon_{cf}\right) \int_0^\infty m_c(t - \tau)\beta(\tau) \, d\tau \qquad (10)$$

41

$$L_t = P_{nbinom}[\, i_{ct}(t) \mid h_{ct}(t)] \times P_{nbinom}[\, i_{cf}(t) \mid h_{cf}(t)] \tag{14}$$

The final likelihood of the community infections over the course of a period of interest is:

$$L = \prod_t L_t \tag{15}$$

For sensitivity analysis, we assumed that the observed data was not stratified into linked and unlinked cases and the likelihood function was defined as:

$$L_t = P_{nbinom}[\, i_c(t) \mid h_c(t)] \tag{16}$$

Using Eqs. (14) and (16), we could estimate the lower and upper limits on the median number of missed infections respectively as the former assumes no misclassification on the source of infection for a case, while the latter tends to exhibit wider uncertainty as it does not account for the source of infection of a locally infected case. In reality, misclassification could occur during cluster investigation and data processing for a large number of cases, but some information on case linkage would exist and lend support to the analysis if contact tracing and testing of exposed contacts was implemented.

Given the long time series of data available for modelling, we subset the wild-type SARS-CoV-2 and Delta variant notified cases a priori, into different time periods in 2020 and 2021 (Table 2). From Apr 24, 2021, onwards, non-residents with a travel history to India were not allowed entry into Singapore or transit through Singapore in response to the surge in Delta variant cases reported in India [33]. This was extended to include Bangladesh, Nepal, Pakistan and Sri Lanka from 2 May onwards [34]. Following the tightening of border controls, the notified COVID-19 cases among travellers from May 16, 2021, onwards reduced to an average of 5 cases per day and the missed imported infections from this date onwards were assumed to be negligible.

Model fitting was performed using a Markov chain Monte Carlo (MCMC) algorithm with an adaptive multivariate normal proposal distribution [35] and the assumed informative priors are listed in Additional file 1: Table S1. Sensitivity analysis was performed assuming uniform priors. Four chains were run with a burn-in of 5000 iterations and samples were thinned every 50 iterations. Convergence was assessed through visual inspection of the Gelman-Rubin convergence statistic and trace plots. The posterior distribution of the parameters in each time period was estimated via MCMC sampling from 23,200 draws.

Burden of disease and infection
In Singapore, all pneumonia deaths or deaths from unknown causes were subjected to SARS-CoV-2 testing [3, 25]. Hence, the extent of underreporting for SARS-CoV-2 deaths was expected to be low during the study period of interest. The average risk of ICU admission among cases was the proportion of cases admitted into the ICU over all notified cases and the average case fatality ratio was the proportion of deaths among all notified cases. The average risk of ICU admission among all infections and the average infection fatality ratio was also computed using the modelled total infections.

Comparing outbreak metric between using notified cases only and with inclusion of missed cases
We calculated the proportion of unlinked cases over all notified confirmed cases as this metric is commonly used in the COVID-19 pandemic and in previous outbreaks of other infectious diseases to proxy the extent of missed infections [36–42]. Using the modelled missed and notified infections, we derived the level of case ascertainment (i.e. the proportion of notified cases to the total number of infections) and compared both outbreak metrics. All modelled data were presented as the median with 95% credible intervals (CrI).

Table 2 Time periods considered for wild-type SARS-CoV-2 transmission during 2020 and Delta variant transmission during 2021

SARS-CoV-2 lineage	Time period	Description
Wild-type	Jan 18–Feb 29, 2020	Transmission driven by travellers arriving from Wuhan
	Mar 1–Apr 6, 2020	Returning travellers from countries with ongoing outbreaks
	Apr 7–Jun 18, 2020	Increased reopening of national borders
	Jun 19–Jul 12, 2020	Resumption of more local activities
	Jul 13–Dec 31, 2020	Increased reopening of national borders
Delta variant	Apr 1–May 12, 2021	Transmission driven mainly by travellers arriving from India
	May 13–Jun 20, 2021	Tightening of outbreak control measures before relaxation of measures in mid-June
	Jul 1–Jul 17, 2021	Nightclub and fishery port outbreak clusters
	Jul 18–Aug 18, 2021	Tightening of outbreak control measures

Independent model validation

We validated the model outcomes against an independent, cross-sectional population seroprevalence survey conducted from Sep 7 to 31, 2020, with 1578 participants randomly selected from the general population (Chen MI-C, Lim VWX. Updates on the sero-epidemiology of SARS-CoV-2 in Singapore, and reflections on the role of post-vaccine sero-surveillance, unpublished). Serology was performed using commercially available test kits from Roche, Wondfo and GenScript cPass S Protein RBD Neutralization Antibody Detection Kit, a SARS-CoV-2 surrogate virus neutralisation test (sVNT), a pseudovirus-based VNT (pVNT) and an S protein flow-based assay [43–45]. Accounting for the seroconversion probability and IgG detection probability since time of infection, we estimated the number of serology positive cases and compared them against the seroprevalence rate in the general population as follows:

$$\int_{T_s}^{T_e}\int_0^\infty f_T(t)[n_c(t-\tau)+m_c(t-\tau)]\,p_s\,f_p(\tau)\,d\tau dt$$
(17)

where $f_T(t)$ is a uniform probability distribution of being tested on a day from Sep 7 to Oct 31, 2020 (T_s and T_e inclusive of both dates), p_s is the probability of seroconversion [46], $f_p(\tau)$ is the probability of being detected serology positive τ time since infection given seroconversion. We assumed the serology detection probabilities approach 1 after 30 days from time of infection and no decline in immunity was observed up to 11 months post infection [47]. Sensitivity analysis was performed assuming approximately 40% decline in antibody levels 3 months post infection and about 80% decline by 11 months post infection [48, 49]. Observed data were presented as the mean and the 95% confidence intervals (CI) for binomial proportions were computed using Wilson's method [50]. We bootstrapped the difference between the observed and modelled rates and this difference was considered statistically significant if the 95% CI does not contain zero.

Results

Combining multiple data streams with a transmission model, we compared the effectiveness of respective outbreak control measures and epidemiological characteristics for different circulating SARS-CoV-2 variants.

Effectiveness of border control

The earliest measure implemented to minimise the introduction of wild-type SARS-CoV-2, and later also used to delay the Delta variant, was border control measures. Initial measures from Jan 18 to Feb 29, 2020, aimed to reduce the spread of SARS-CoV-2 by infected persons arriving from China. While there was less than 1 notified imported case per day during this period (Fig. 2A), we estimated that there were 0.6 missed imported infections per day (95% credible intervals, CrI 0.2–1) (Fig. 3A) or equivalent to 0.9 missed imported infections per notified case (95% CrI 0.4–2) (Additional file 1: Table S2). From Mar 1 to Apr 6, 2020, there was a surge of 15 notified imported cases per day returning from other countries with ongoing outbreaks (Fig. 2A) and we estimated a median of 7 missed imported infections per day (95% CrI 2–24) (Fig. 3A) or 0.5 missed imported infections per notified imported case (95% CrI 0.1–2) (Additional file 1: Table S2). During this period, border control measures were tightened and incoming travellers were progressively subjected to quarantine in managed institutions. Despite the decline in notified imported cases from Mar 16 to Apr 1, 2020, persistent community transmission prompted a nationwide partial lockdown on Apr 7, 2020 (Fig. 2A–C) where non-essential workers were required to work from home, students transited to home-based learning and non-essential facilities and services ceased operations [51].

Following the partial lockdown, the reopening of borders and hence the risk of disease introduction was carefully balanced against the resumption of community activities and the associated risk of community transmission. From Jul 13 to Dec 31, 2020, there were 7 notified cases per 1,000 travellers, three times higher than the period prior to lockdown (i.e. 2 notified cases per 1,000 travellers from Mar 1 to Apr 6, 2020) but the number of imported cases who were not quarantined upon arrival was kept low at less than 0.1 cases per 1000 arrivals. Furthermore, with the strict quarantine of incoming travellers and continued enforcement of outbreak control measures, the average daily number of missed imported infections declined to 2 cases (95% CrI 0.3–6) from Jul 13 to Dec 31, 2020 (Fig. 3A) or 0.2 missed imported infections per notified imported case (95% CrI 0.04–0.7) (Additional file 1: Table S2) with no signs of a growing outbreak (Fig. 2A–C).

(See figure on next page.)

Fig. 2 Daily incidence of COVID-19 cases in Singapore arising from wild-type SARS-CoV-2 transmission in 2020, **A** notified imported cases who were isolated after testing positive or quarantined upon arrival, **B** notified local linked cases and modelled posteriors, **C** notified local unlinked cases and modelled posteriors, and **D** modelled posteriors for local missed infections. Daily incidence of COVID-19 cases in Singapore arising from SARS-CoV-2 Delta variant transmission in 2021, **E** notified local cases and modelled posteriors and **F** modelled posteriors for local missed infections. Grey-shaded areas in **A**–**F** represent periods with movement and visitor restrictions with darker shades signifying a reduced number of visitors to each household per day. Modelled posterior outbreak metrics for **G** wild-type SARS-CoV-2 transmission in 2020 and **H** Delta variant transmission in 2021

Fig. 2 (See legend on previous page.)

44
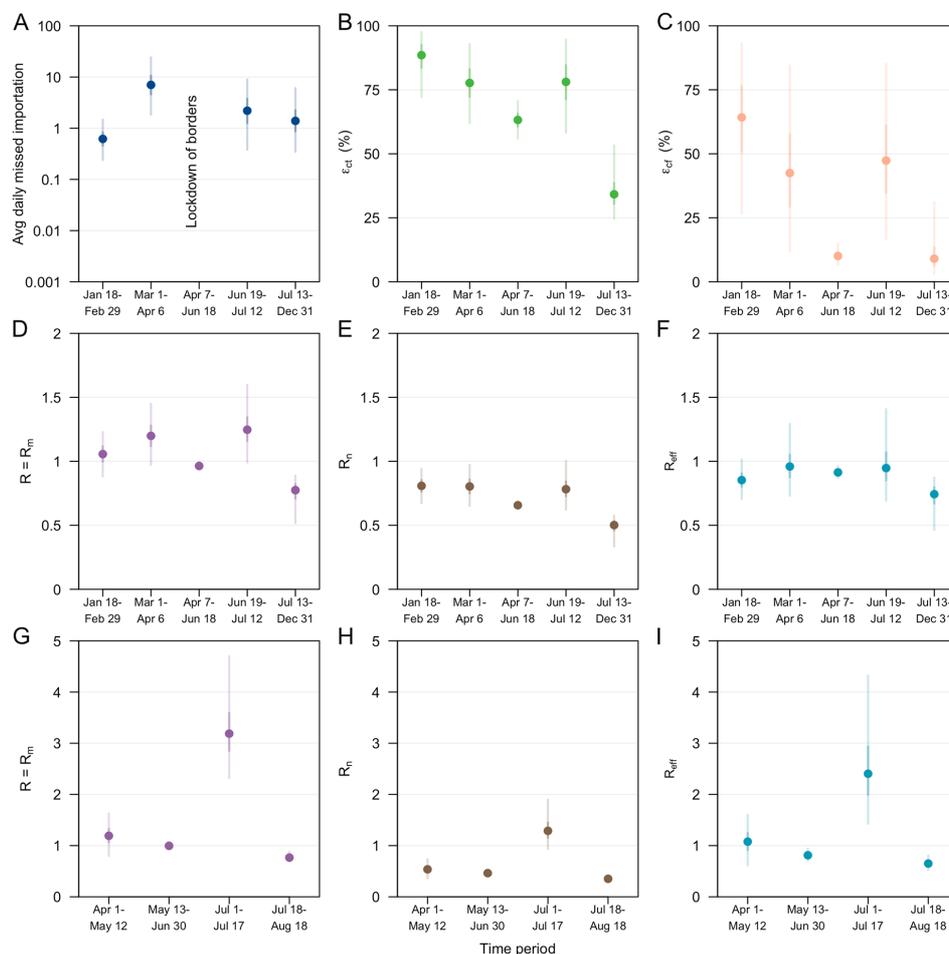
Fig. 3 Model parameter estimates on SARS-CoV-2 transmission. **A** average daily missed imported infections in log scale, **B** effectiveness of contact tracing in detecting a linked case, ϵ_{ct}, **C** effectiveness of case finding in detecting an unlinked case, ϵ_{cf}. Reproduction number for wild-type SARS-CoV-2 in 2020 (**D**–**F**) and Delta variant in 2021 (**G**–**I**). **D**, **G** R or R_m, of a missed COVID-19 case, **E**, **H** R_n, of a notified COVID-19 case, and **F**, **I** effective reproduction number, R_{eff}. Estimates of the posterior median (dot), 50% CrI (dark vertical lines) and 95% CrI (light vertical lines) as shown

From Apr 1 to May 12, 2021, while the country continued to enforce quarantine for the majority of the incoming travellers in managed institutions, there was an average of 14 notified cases per 1000 travellers during this period. This was the highest level in our study window. The surge was attributed to imported cases with travel from India. While notified community COVID-19 cases from Apr 1 to May 12, 2021, were low with an average of six cases per day, the occurrence of increased transmission and COVID-19 clusters at the international airport prompted the tightening of COVID-19 outbreak control measures [46]. Despite imposing a travel ban to all non-residents with a travel history to India from Apr 23, 2021, onwards [27], we estimated 4 missed imported

infections per day (95% CrI 1–26) or 0.3 missed imported infections per notified imported case (95% CrI 0.05–1.3) (Additional file 1: Table S2).

Effectiveness of case finding and contact tracing

We estimated the country's contact tracing system was able to detect over 78% of the secondary infections generated by notified cases (ϵ_{ct}, 95% CrI 62–93%) from Mar 1 to Apr 6, 2020 (Fig. 3B). However, the effectiveness of case finding which depends on the overall testing capacity, the extent of surveillance and the health-seeking behaviour of the population was at 64% (ϵ_{cf}, 95% CrI 27–93%) during the start of the outbreak and declined to 42% (ϵ_{cf}, 95% CrI 12–84%) prior to the partial lockdown in Apr 2020 (Fig. 3C). One week before the partial lockdown, there were an average of 16 unlinked cases per day and we estimated 120 missed infections (95% CrI 25–870) per day signifying substantial gaps in the transmission chains. Consequently, ϵ_{ct} and ϵ_{cf} during the lockdown was lowered to 63% (95% CrI 56–71%) and 10% (95% CrI 7–15%), respectively (Fig. 3B and C).

As social and economic activities progressively resumed from Jun 19, 2020, onwards, we estimated an increase in ϵ_{ct} to 78% (95% CrI 58–94%) and ϵ_{cf} to 47% (95% CrI 17–85%) (Fig. 3B and C). This finding is in line with the broadening of the close contact definition, use of contact tracing devices to facilitate contact tracing, implementation of large-scale swab operations to limit spread from outbreak clusters and increased use of rapid antigen tests for routine surveillance in targeted groups [52–55].

Across all time periods in 2020, ϵ_{cf} exhibits wide credible intervals as a result of some correlation with the factor, ρ, which scales the extent of missed imported infections (Additional file 1: Fig. S1). Similar estimates of ϵ_{ct} and ϵ_{cf} were obtained when using uniform priors for sensitivity analysis (Additional file 1: Figs. S2 and S3).

Community reproduction number

Prior to the partial lockdown, the average number of secondary cases generated by a single infectious individual, R, was estimated to be 1.2 cases (95% CrI 1.0–1.4) from Mar 1 to Apr 6, 2020, but the observed reproduction number among chains of notified cases was lower at 0.8 cases (R_n, 95% CrI 0.7–1.0) due to the reduced amount of time spent in the community while infectious compared to a missed infection (Fig. 3D and E). Overall, R_{eff} was 1.0 cases (95% CrI 0.7–1.3) resulting in a sustained cumulative increase of cases (Fig. 3F). During the partial lockdown in 2020, we estimated R_{eff} to be below 1 at 0.9 cases (95% CrI 0.9–1.0). While this signalled a declining outbreak, it took approximately one month to reach a daily incidence of less than 10 cases (Fig. 2B and C).

From Jan 18 to Jun 18, 2020, the daily number of notified cases in the community was at least 10 cases per day. Using a model fitted against notified cases without stratifying the data into linked and unlinked cases for sensitivity analysis, the median estimates for R were similar to the above main analysis, although wider uncertainty intervals were observed due to the lack of information on case linkage to constraint estimates (Additional file 1: Fig. S4).

Outbreak control measures were tightened from May 16, 2021, onwards and the average daily COVID-19 Delta variant community cases declined to less than 10 from Jun 14 to Jul 30, 2021, with a R of 1.0 cases (95% CrI 0.9–1.1) (Fig. 3G). However, a rapid increase of COVID-19 cases was observed in Jul 2021 and epidemiological investigations pointed to transmissions at nightclubs and at a fishery port [49]. This rapid growth was made possible when R was approximately 3.2 cases (95% CrI 2.3–4.7) but model fitting suggested that this lasted for about 2 weeks from Jul 1 to 17, 2021 (Figs. 2E and 3G). With extensive testing and clamp down of underground nightclubs following detection on Jul 12, 2021, cases were progressively notified over the following week and showed signs of decline prior to the tightening of control measures on Jul 22, 2021. When adjusting for the effect of varying vaccination, the reproduction number across the time periods of study in the Delta variant outbreak was scaled up by 1.2–1.5 times (Additional file 1: Fig. S5). The reproduction number represented the risk arising from other population interventions or human behaviour, in the absence of vaccination and was above 1 as the country progressively reopened and relaxed the outbreak restrictions following an increase in vaccination coverage.

When using a uniform prior for analysis, model fitting for the Delta variant outbreak showed similar outputs to the case incidence and reproduction number from Apr 1 to May 12, 2021, and May 13 to Jun 30, 2021 (Additional file 1: Figs. S6 and S7). However, outputs for the uniform prior diverge from the observed data and the outputs of the informative prior for Jul 1 to Jul 17, 2021, and Jul 18 to Aug 18, 2021—this deviation will be addressed further in the 'Discussion' section.

Burden of disease and infection

Using the incidence of both linked and unlinked cases, our main analysis estimated 730 missed infections (95% CrI 230–3600) (Table 3 and Fig. 2D) from Mar 1 to Apr 6, 2020, which translates to approximately 20 missed infections per day (95% CrI 6–96). During the partial lockdown period and the succeeding period (Apr 7–Jun 18, 2020), the number of missed infections per day decreased to 30 (95% CrI 20–56). As border restrictions

Table 3 Summary of observed data and modelled outputs (median and 95% CrI in parenthesis) by respective time periods in 2020 for wild-type SARS-CoV-2 transmission

Observed data (●) and modelled outputs (◆)	Time period in 2020					
	Overall Jan–Dec	Jan 18–Feb 29	Mar 1–Apr 6	Apr 7–Jun 18	Jun 19–Jul 12	Jul 13–Dec 31
● Imported cases						
Isolated for testing on arrival or quarantined	1653	0	50	5	78	1520
Not quarantined	547	29	497	0	4	17
● Local cases (by time of isolation)						
Linked	1505	65	606	610	113	111
Unlinked	864	20	204	420	107	113
◆ Missed cases	4400 (2400–11,000)	25 (8–100)	730 (230–3600)	2200 (1500–4100)	280 (100–1100)	1100 (360–2800)
◆ Total cases (adjusted by time of infection and missed cases)	7100 (4800–14,000)	130 (90–220)	1900 (1300–4900)	2900 (2200–5100)	590 (350–1500)	1400 (620–3100)
● ICU cases (by time of isolation)	86	13	44	28	1	0
● Deaths (by time of isolation)	22	2	11	9	0	0
◆ Case ICU risk (%)	3.3 (2.5–4.0)	23.2 (16.2–32.4)	4.0 (3.0–5.0)	2.0 (1.5–2.4)	0.3 (0.2–0.4)	0 (0–0)
◆ Infection ICU risk (%)	1.2 (0.6–1.8)	18.2 (10.8–26.3)	2.4 (1.0–3.7)	0.5 (0.3–0.7)	0.2 (0.07–0.3)	0 (0–0)
◆ Case fatality ratio (%)	0.8 (0.6–1.0)	3.8 (2.6–5.3)	1.2 (0.9–1.5)	0.5 (0.4–0.6)	0 (0–0)	0 (0–0)
◆ Infection fatality ratio (%)	0.3 (0.2–0.5)	3.0 (1.7–4.3)	0.7 (0.3–1.1)	0.1 (0.07–0.2)	0 (0–0)	0 (0–0)

were gradually lifted and economic and social activities resumed from Jun 19, 2020, onwards, the daily missed infections remained low at 7 infections (95% CrI 3–20). Overall, we estimated that 4,400 infections (95% CrI 2400–11,000) were missed in 2020 or equivalent to 63% of all infections (95% CrI 49–78%) (Table 3 and Fig. 2D).

Our preceding main analysis incorporated the additional case linkage information provided by case finding and contact tracing (i.e. linked and unlinked cases). When model fitting during sensitivity analysis was performed using the time series of all notified cases without stratification by case linkage, we estimated approximately 1900 infections (95% CrI 600–10,000) were missed prior to the partial lockdown in Apr 2020 or 50 missed infections per day (95% CrI 10–280) (Additional file 1: Table S3). Contrary to the previous model fit, we estimated approximately 130 missed infections per day (95% CrI 80–300) during the partial lockdown, and this was approximately 4 times (95% CrI 4–5) higher than of the previous model fit (Additional file 1: Table S3). We estimated that 15,000 infections (95% CrI 8,400–38,000) were missed in 2020 (Additional file 1: Table S3 and Fig. S8).

Both the main and sensitivity analysis for the wild-type SARS-CoV-2 serve as the lower and upper limit of the modelled missed infections. The former assumed perfect classification of case linkages while the latter was derived without using the case linkage information to constrain the range of parameters that reproduces the modelled outbreak, resulting in wider uncertainty intervals in the estimated missed infections.

From Apr 1 to May 12, 2021, more than 60% of the cases and more than 65% of the population were unvaccinated. Singapore experienced a surge in notified imported cases and consequently missed imported infections. Using all notified Delta variant cases without stratification by case linkage, we estimated that 1,400 community infections were missed (95% CrI 200–15,000) during this period (Table 4 and Fig. 2F). Rapid transmission arising from nightclub clusters and a fishery port followed by extensive case finding measures such as large-scale swab operations resulted in 80 missed infections per day (95% CrI 16–700) from Jul 18 to Aug 18, 2021. Despite the shorter study period for Delta variant transmission as compared to the wild-type SARS-CoV-2, we estimated that there were 12,000 missed infections (95% CrI 4200–75,000), or equivalent to 70% of all infections (95% CrI 49–91%), in a span of about 5 months.

Overall, the estimated case fatality ratio was 0.8% (95% CrI 0.6–1.0%) in 2020 and 0.5 (95% CrI 0.2–0.8%) in Apr–Aug 2021, and remains below 1% as of Nov 2021 (Tables 3 and 4). The infection fatality ratio was 0.3% (95% CrI 0.2–0.5%) in 2020 for wild-type SARS-CoV-2 infections and 0.2% (95% CrI 0.033–0.3%) in 2021 for Delta variant infections. The risk of ICU admission among cases was 3.3% (95% CrI 2.5–4.0%) in 2020 and 0.7% (95% CrI 0.3–1.1%) in 2021 and but these estimates were approximately 3 times higher than the risk of ICU admission among infections at 1.2% (95% CrI 0.6–1.8%) in 2020 and 0.2% (95% CrI 0.04–0.4%) in 2021.

47

Table 4 Summary of observed data and modelled outputs (median and 95% CrI in parenthesis) by respective time periods in 2021 for SARS-CoV-2 Delta variant transmission

Observed data (●) and modelled outputs (◆)	Time period in 2021				
	Overall Apr–Aug	Apr 1–May 12	May 13–Jun 30	Jul 1–Jul 17	Jul 18–Aug 18
● Imported cases					
Isolated for testing on arrival or quarantined	1291	809	270	136	76
Not quarantined	93	34	32	12	15
● Local cases (by time of isolation	4371	196	755	474	2946
◆ Missed cases	12,000 (4200–75,000)	1400 (180–15,000)	1700 (700–11,000)	1400 (270–11,000)	6100 (2600–43,000)
◆ Total cases (adjusted by time of infection and missed cases)	17,000 (8000–84,000)	1700 (420–15,500)	2400 (1500–12,000)	2400 (800–13,000)	9000 (4700–50,000)
● ICU cases (by time of isolation)	36	3	11	3	19
● Deaths (by time of isolation)	25	3	4	1	17
◆ Case ICU risk (%)	0.7 (0.3–1.1)	1.7 (1.0–2.6)	1.3 (1.0–1.8)	0.8 (0.3–1.7)	0.5 (0.2–0.8)
◆ Infection ICU risk (%)	0.2 (0.04–0.4)	0.3 (0.03–1.1)	0.4 (0.08–0.6)	0.3 (0.06–1.0)	0.2 (0.03–0.3)
◆ Case fatality ratio (%)	0.5 (0.2–0.8)	1.4 (0.8–2.0)	0.6 (0.4–0.8)	0.2 (0.09–0.4)	0.5 (0.2–0.8)
◆ Infection fatality ratio (%)	0.2 (0.03–0.3)	0.2 (0.02–0.9)	0.2 (0.03–0.3)	0.08 (0.01–0.2)	0.2 (0.03–0.3)

Comparing outbreak metric between using notified cases only and with inclusion of missed cases

When the effectiveness of detecting linked and unlinked cases declined in March 2020 during the surge of imported cases and further declined during the partial lockdown (Fig. 3B and C), we estimated the proportion of missed infections among all infections increased to 74% (95% CrI 67–82%) (Fig. 2G) between Apr 7 to Jun 18, 2020. This was 1.9 times (95% CrI 1.6–2.3) higher than the proportion of cases that was unlinked at 39% (95% CrI 32–48). The proportion of missed infections among all infections was also 1.3 times (95% CrI 0.7–2.3) higher than the proportion of unlinked cases to all cases from Apr 1 to May 12, 2021, when the Delta variant was the predominant circulating strain (Fig. 2H).

During periods of increased testing during reopening, the estimated proportion of missed cases was low at 47% (95% CrI 26–73%) from Jun 19 to Jul 12, 2020; 0.95 times (95% CrI 0.5–1.6) lower that the proportion of cases that was unlinked which was 49% (95% CrI 38–61%) (Fig. 2G). Similarly, from Jul 12 to Aug 18, 2021, where extensive testing was conducted as part of cluster outbreak investigations, we estimated that 68% of all infections were missed (95% CrI 50–88%) and 0.9 times (95% CrI 0.6–1.3) lower than the proportion of cases that was unlinked at 79% (95% CrI 54–94%) (Fig. 2H).

Independent validation of estimates

While the transmission model was able to reproduce the observed temporal trends, we sought to further validate the model outputs against an independent population-level cross-sectional seroprevalence survey. From Sep 7 to 31, 2020, SARS-CoV-2 antibodies were detected in two out of 1578 participants when subjected to all serological test methods and these participants were also negative for SARS-CoV-1 infection [34]. This translates to an observed seroprevalence of 0.13% (95% confidence intervals, CI 0.03–0.46%). Four other participants had SARS-CoV-2 antibodies detected when twice analysed by the cPass test kit but tested negative on the other serological tests.

Using the linked and unlinked cases in 2020, our model estimates implied a population seroprevalence of 0.05% (95% CI 0.03–0.1%) when assuming no waning immunity up to 11 months after symptoms post infection. When using the notified cases without accounting for their case linkages in 2020 for model fitting in sensitivity analysis, the estimated seroprevalence was revised upwards to 0.13% (95% CI 0.08–0.3%). Both model outcomes were not statistically significantly different from the observed seroprevalence. However, when assuming waning seropositivity 3 months after symptoms onset, the estimated seroprevalence in both models was 0.03% (95% CI 0.02–0.06%) and 0.08% (95% CI 0.05–0.18%).

Discussion

Using the growth patterns in the daily incidence of local linked and unlinked cases, and imported cases with community contact, we reconstructed the incidence of missed infections over time in Singapore. This enabled us to disentangle the effects of targeted measures such as case finding and contact tracing from other population-wide outbreak interventions. Our modelling framework

was able to infer these missed infections without requiring large-scale serological surveys, which are typically challenging to conduct at the start of a pandemic. Such analysis can therefore provide early insights into the effectiveness of respective categories of outbreak control measures, and hence further inform the extent of measures required during different stages of an outbreak.

The changes in the estimated effectiveness of control measures largely coincide with the shifts in outbreak control policies, but there were other likely contributing factors. Changes in human behaviours such as a reduction in health-seeking behaviour coincided with a decline in the effectiveness of case finding, ϵ_{cf}, from 42% in Mar 1 to Apr 6, 2020, to 10% during the lockdown from Apr 7 to Jun 18, 2020 [23]. Furthermore, the interdependence of outbreak control measures can cause the effectiveness of measures to change in tandem. In particular, the contribution of contact tracing towards outbreak control hinges on the extent of case finding. Following the decline in ϵ_{cf} during the lockdown, the effectiveness of contact tracing in identifying new cases declined from 78% in Mar 1 to Apr 6, 2020, to 63% during the lockdown. This observation is also supported by theory—when the effectiveness of isolating cases is low, a slight decrease in the effectiveness of contact tracing can result in a growing outbreak [30]. Collectively, about 75% of the infections were estimated to be missed during the lockdown and this proportion was higher than other time periods due to the lowered effectiveness in both case finding and contact tracing. Thus, by identifying which outbreak control measures were contributing to the growth of an outbreak and the corresponding reasons for its lowered effectiveness, it is possible to address relevant aspects of human behaviour (e.g. promote use of telemedicine as patients feel more comfortable seeing their doctors online [56]; discourage clinic hopping so the same doctor can better assess the need for follow up test [57]).

In both wild-type SARS-CoV-2 and Delta variant outbreaks in Singapore, on average, there was less than 1 death per day. With prolonged periods of low death counts, we reconstructed the underlying outbreak dynamics using the incidence of linked and unlinked cases instead of using reported fatalities [21, 58]. Prior to 2021, the Singapore population was largely unvaccinated and during the Delta variant outbreak about 60% of the population was vaccinated by Aug 2021. Our CFR estimates were less than 1% for the wild-type SARS-CoV-2 and Delta variant outbreak, which was less than the early CFR estimates of around 1.4% for wild-type SARS-CoV-2 and 3 times lower than the CFR estimates for the Delta variant in other studies [59, 60]. The IFR estimates for both outbreaks in Singapore were also less than 0.5%, and in the lower range of IFR estimates as compared to other

countries and regions [58, 61, 62]. While the healthcare system was stretched in both outbreaks, ICU capacity was not exceeded and this helped to keep the number of deaths to a minimum. As deaths observed in small outbreak clusters would not be reflective of the number of deaths that could arise during a large epidemic wave, care is needed in the interpretation of underlying infection dynamics and how these influence measured disease outcomes.

We found that metrics derived from observed data alone do not always accurately reflect the underlying outbreak. Specifically, metrics such as the proportion of unlinked cases among all notified cases are not necessarily representative of the proportion of missed infections among all infections, and policy makers should therefore be careful when drawing conclusions of the latter from the former. This discrepancy is likely to occur because the missed infections have a much higher reproduction number as compared to notified cases, or when a single missed infection is the source of infection for multiple unlinked cases and the outbreak could be misinterpreted as growing or declining slowly in either scenario. In contrast, contact tracing data provides additional information on the source of infection of a case. The collection of such data expends minimal effort yet can help to improve our understanding of the underlying outbreak although misclassification could also affect the interpretation of the outbreak dynamics. Thus, the interpretation of common metrics should be done with a clear understanding of the data collection process. Previous studies have estimated the impact of measures such as border control by assessing correlations between the timing of interventions and national-level case incidence [63], but our results suggest such analysis will not capture the complexity of interacting measures against a background of changing infection detection.

We also found that multiple independent notification datasets and informative priors helped to disentangle the model parameters and achieve more precision in estimates. Unlinked cases were generated by either missed imported or local infections with the former modelled as a factor of the notified imported cases, ρ. As such, the interaction of model parameters results in wide 95% credible intervals for ϵ_{cf} estimates. To improve these estimates, we could further stratify exposure histories of unlinked cases by their interactions with travellers from countries with ongoing outbreak for model fitting. Informative and uniform priors produced a similar set of parameter estimates when there were multiple independent notification data in the SARS-CoV-2 wild-type outbreak in 2020 for model fitting. However, the model output using a uniform prior was different from that of an informative prior for the Delta variant outbreak in Jul

to Aug 2021. Unlike the wild-type SARS-CoV-2 outbreak, model fitting for the Delta variant was based on the time series of cases without accounting for the case linkage. As such, there was limited data to inform the extent of underreporting and hence the number of missed infections. The estimates of R from Jul 1 to 17, 2021 when using the informative prior falls within the lower range of the estimates derived from the uniform prior although both analyses suggest a growing outbreak.

There are some additional limitations to our study. One is that asymptomatic cases were assumed to have a similar distribution of delay from the time of infection to notification as symptomatic cases. To circumvent this, we can study the changes in the trajectory of the cycle threshold values (proxy for viral load) of cases that were tested multiple times over the course of the infection. The infection time of symptomatic and asymptomatic cases can be estimated from their respective viral growth trajectory [64, 65] thereby informing the delay distribution for respective types of cases. Furthermore, we assumed that asymptomatic cases were as infectious as symptomatic individuals, and hence, no stratification of R was modelled as there is no strong evidence to suggest that asymptomatic SARS-CoV-2 infections are less infectious than symptomatic individuals [66, 67]. Our modelled outcomes for wild-type SARS-CoV-2 transmission were able to reproduce independent observations in a separate population-level serological survey and this lends support to our assumption of a homogeneous R among most missed infections.

In addition, the burden of disease and infection estimates were averaged across all age groups, as there was insufficient data to estimate the transmissibility and susceptibility across different age groups in each time period. In our branching process model, we also assumed that each of the four parameters remains constant in a specified time period. As such, we are unable to provide a time-varying measure to characterise the impact of different outbreak detection and control measures that were progressively rolled out in the population at a granular level. Instead, time periods were chosen based on prior knowledge of major policies that would affect at least one of the four model parameters. In particular, from Jul 1 to 17, 2021, the outbreak of COVID-19 cases from a nightclub cluster and fishery port resulted in a reproduction number of more than 1. For cases at the end of this time period, the model assumes that their R is the same as the cases at the start of the same time period. However, as rapid and strict outbreak control measures were implemented around the period of Jul 18, 2021, the R of the cases around this transition period is expected to vary

between the reproduction number estimated for Jul 1 to 17, 2021 and Jul 18 to Aug 18, 2021. With the potential for a larger reproduction number using a uniform prior, the exponential number of new infections generated by cases around the transition period causes the modelled peak outbreak to overshoot the observed peak in the subsequent time period. This further highlights the importance of having multiple independent data on case linkage to better inform the parameter estimates and to infer missed infections.

Conclusions

The SARS-CoV-2 pandemic has generated many new and expanded data streams alongside new ways to reconstruct outbreak dynamics and evaluate the extent of missed infections, even in the presence of high asymptomatic rates and underreporting of cases. Our results show that data on case linkage can help countries evaluate their performance in case finding, contact tracing and the effectiveness of their border restrictions. Relying simply on the interaction of missed and notified infections can introduce unseen heterogeneity into the reproduction number and hence create a false picture of a controlled outbreak. As countries deal with future waves of COVID-19 or plan for pandemics in the future, it will be important to have an integrated surveillance and modelling analysis system that can estimate these crucial undetected transmission events.

Supplementary Information

Additional file 1: Table S1. Mathematical notations. **Table S2.** Notified and modelled missed imported wild-type SARS-CoV-2 infections in 2020. **Table S3.** Summary of observed data and modelled outputs for wild-type SARS-CoV-2 transmission. **Fig. S1.** Contour plots to show the correlation between model parameters. **Fig. S2.** Posterior estimates for model fitted to time series of linked and unlinked SARS-CoV-2 wild type cases in 2020 using informative and non-informative priors. **Fig. S3.** Posterior density of the parameters for model fitted to time series of linked and unlinked SARS-CoV-2 wild type cases in 2020 using informative and non-informative priors. **Fig. S4.** Reproduction number, R of a SARS-CoV-2 wild-type in 2020 and Delta variant case in 2021. **Fig. S5.** Reproduction number, R of a SARS-CoV-2 Delta variant case in 2021 after adjusting for vaccine coverage and vaccine effectiveness and using notified cases with no information of the case linkage for model fitting. **Fig. S6.** Posterior estimates for model fitted to time series of SARS-CoV-2 Delta variant cases (without accounting for case linkage) in 2021 using informative and non-informative priors. **Fig. S7.** Posterior density of the parameters for model fitted to time series of linked and unlinked SARS-CoV-2 Delta variant cases (without accounting for case linkage) in 2021 using informative and non-informative priors. **Fig. S8.** Daily incidence of COVID-19 cases in Singapore arising from SARS-CoV-2 wild-type transmission in 2020. **Fig. S9.** Markov chain Monte Carlo trace plots for parameters modelling wild-type SARS-CoV-2 transmission in 2020. **Fig. S10.** Markov chain Monte Carlo trace plots for parameters modelling SARS-CoV-2 Delta variant transmission in 2021.

Availability of data and materials
All data are available in the manuscript or the supplementary information. The data and code used for our analyses are publicly available at https://doi.org/10.5281/zenodo.7538047.

Declarations

Ethics approval and consent to participate
The study was approved by the London School of Hygiene & Tropical Medicine Observational Research Ethics Committee (ref. 25727). All data and analysis were collected and performed in line with the Infectious Diseases Act in Singapore which permits the collection and publication of surveillance data.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Ministry of Health, Singapore, Singapore. [2]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK. [3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore.

References
1. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584:257–61.
2. Giuntella O, Hyde K, Saccardo S, Sadoff S. Lifestyle and mental health disruptions during COVID-19. Proc Natl Acad Sci U S A. 2021;118:e2016632118.
3. Ng Y, Li Z, Chua YX, Chaw WL, Zheng Z, Er B, et al. Evaluation of the effectiveness of surveillance and containment measures for the first 100 patients with COVID-19 in Singapore — January 2–February 29, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(11):307–11.
4. Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis. 2020;20:1151–60.
5. Liu Y, Morgenstern C, Kelly J, Lowe R, Munday J, Villabona-Arenas CJ, et al. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. BMC Med. 2021;19:40.
6. Chen Y-H, Fang C-T, Huang Y-L. Effect of non-lockdown social distancing and testing-contact tracing during a COVID-19 outbreak in Daegu, South Korea, February to April 2020: a modeling study. Int J Infect Dis. 2021;110:213–21.
7. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature. 2020;585:410–3.
8. Kyrychko YN, Blyuss KB, Brovchenko I. Mathematical modelling of the dynamics and containment of COVID-19 in Ukraine. Sci Rep. 2020;10:19662.
9. Kerr CC, Mistry D, Stuart RM, Rosenfeld K, Hart GR, Núñez RC, et al. Controlling COVID-19 via test-trace-quarantine. Nat Commun. 2021;12:2993.
10. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med. 2020;26:855–60.
11. Pung R, Cook AR, Chiew CJ, Clapham HE, Sun Y, Li Z, et al. Effectiveness of containment measures against COVID-19 in Singapore: implications for other national containment efforts. Epidemiology. 2021;32:79–86.
12. Chiu WA, Fischer R, Ndeffo-Mbah ML. State-level needs for social distancing and contact tracing to contain COVID-19 in the United States. Nat Hum Behav. 2020;4:1080–90.
13. Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health. 2020;74:861–6.
14. Clifford S, Pearson CAB, Klepac P, Van Zandvoort K, Quilty BJ, CMMID COVID-19 working group, et al. Effectiveness of interventions targeting air travellers for delaying local outbreaks of SARS-CoV-2. J Travel Med. 2020;27(5):taaa068.
15. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. PNAS. 2011;108:2825–30.
16. Cauchemez S, Nouvellet P, Cori A, Jombart T, Garske T, Clapham H, et al. Unraveling the drivers of MERS-CoV transmission. PNAS. 2016;113:9081–6.
17. Camacho A, Kucharski AJ, Funk S, Breman J, Piot P, Edmunds WJ. Potential for large outbreaks of Ebola virus disease. Epidemics. 2014;9:70–8.
18. Blumberg S, Lloyd-Smith JO. Inference of R0 and transmission heterogeneity from the size distribution of stuttering chains. PLoS Comput Biol. 2013;9:e1002993.
19. Angulo FJ, Finelli L, Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. JAMA Netw Open. 2021;4:e2033706.
20. Byambasuren O, Dobler CC, Bell K, Rojas DP, Clark J, McLaws M-L, et al. Comparison of seroprevalence of SARS-CoV-2 infections with cumulative and imputed COVID-19 cases: systematic review. PLoS One. 2021;16:e0248946.
21. 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;18:332.
22. Riley S, Ainslie KEC, Eales O, Walters CE, Wang H, Atchison C, et al. Resurgence of SARS-CoV-2: detection by community viral surveillance. Science. 2021;372:990–5.
23. National Centre for Infectious Diseases (NCID), National University of Singapore (NUS), Nanyang Technological University, Singapore (NTU Singapore). NCID, NUS and NTU studies highlight the role of socio-behavioural factors in managing COVID-19. 2020. http://news.ntu.edu.sg/pages/newsdetail.aspx?URL=http://news.ntu.edu.sg/news/Pages/NR2020_May21.aspx&Guid=221cc9a8-f0cd-40ac-a837-c1caf19897b7&Category=All. Accessed 14 Mar 2021.
24. Siette J, Seaman K, Dodds L, Ludlow K, Johnco C, Wuthrich V, et al. A national survey on COVID-19 second-wave lockdowns on older adults' mental wellbeing, health-seeking behaviours and social outcomes across Australia. BMC Geriatr. 2021;21:400.
25. Pung R, Chiew CJ, Young BE, Chin S, Chen MI-C, Clapham HE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. Lancet. 2020;395:1039–46.
26. World Health Organization. SARS-CoV-2 variants: WHO; 2020. https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON305. Accessed 2 June 2021
27. GISAID. GISAID - hCov19 variants. 2021. https://www.gisaid.org/hcov19-variants/. Accessed 28 Sept 2021.
28. Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY. Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. BMC Infect Dis. 2021;21:257.

29. Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369:1106–9.

30. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368:eabb6936.

31. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open. 2020;10:e039652.

32. Zhang M, Xiao J, Deng A, Zhang Y, Zhuang Y, Hu T, et al. Transmission dynamics of an outbreak of the COVID-19 Delta variant B.1.617.2 — Guangdong province, China, May–June 2021. CCDCW. 2021;3:584–6.

33. Ministry of Health, Singapore. Updates on border measures for travellers from India, Westlite Woodlands dormitory cluster and additional precautions for recovered persons. 2021. https://www.moh.gov.sg/news-highl ights/details/updates-on-border-measures-for-travellers-from-india-westl ite-woodlands-dormitory-cluster-and-additional-precautions-for-recov ered-persons. Accessed 29 Oct 2021.

34. Ministry of Health, Singapore. Updates on local situation, border measures for Bangladesh, Nepal, Pakistan, Sri Lanka, and Thailand and precautionary measures to minimise transmission from Tan Tock Seng Hospital cluster. 2021. https://www.moh.gov.sg/news-highlights/details/ updates-on-local-situation-border-measures-for-bangladesh-nepal-pakis tan-and-sri-lanka-thailand-and-precautionary-measures-to-minimise- transmission-from-tan-tock-seng-hospital-cluster. Accessed 29 Oct 2021.

35. Roberts GO, Rosenthal JS. Examples of adaptive MCMC. J Comput Graph Stat. 2009;18:349–67.

36. Government of the Hong Kong Special Administrative Region. COVID-19 testing to intensify: Hong Kong's Information Services Department; 2021. http://www.news.gov.hk/eng/2021/02/20210201/20210201_172308_ 389.html. Accessed 14 Mar 2021

37. New Zealand Government. COVID-19 media conference — 7pm, 14 February 2021 | Unite against COVID-19. 2021. https://covid19.govt.nz/ updates-and-resources/latest-updates/covid-19-media-conference-7pm- 14-february-2021/. Accessed 14 Mar 2021.

38. Channel News Asia. Singapore, Hong Kong to defer air travel bubble launch: CNA; 2020. https://www.channelnewsasia.com/news/singapore/ covid-19-singapore-hong-kong-defer-travel-bubble-december-13675 222. Accessed 14 Mar 2021

39. Oboho IK, Tomczyk SM, Al-Asmari AM, Banjar AA, Al-Mugti H, Aloraini MS, et al. 2014 MERS-CoV outbreak in Jeddah — a link to health care facilities. N Engl J Med. 2015;372:846–54.

40. European Centre for Disease Prevention and Control. Communicable disease threats report, 5-11 July 2015, week 28: European Centre for Disease Prevention and Control; 2015. https://www.ecdc.europa.eu/en/publicatio ns-data/communicable-disease-threats-report-5-11-july-2015-week-28. Accessed 29 Oct 2021

41. Lessler J, Salje H, Van Kerkhove MD, Ferguson NM, Cauchemez S, Rodriquez-Barraquer I, et al. Estimating the severity and subclinical burden of Middle East respiratory syndrome coronavirus infection in the Kingdom of Saudi Arabia. Am J Epidemiol. 2016;183:657–63.

42. Fox M. Ebola cases on the rise again: NBC News; 2015. https://www. nbcnews.com/storyline/ebola-virus-outbreak/ebola-ticks-again-west- africa-n373171. Accessed 29 Oct 2021

43. Bastos ML, Tavaziva G, Abidi SK, Campbell JR, Haraoui L-P, Johnston JC, et al. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. BMJ. 2020;370:m2516.

44. Goh YS, Chavatte J-M, Jieling AL, Lee B, Hor PX, Amrun SN, et al. Sensitive detection of total anti-Spike antibodies and isotype switching in asymptomatic and symptomatic individuals with COVID-19. CR Med. 2021;2(2):100193.

45. Tan CW, Chia WN, Qin X, Liu P, Chen MI-C, Tiu C, et al. A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction. Nat Biotechnol. 2020;38:1073–8.

46. Young BE, Ong SWX, Ng LFP, Anderson DE, Chia WN, Chia PY, et al. Viral dynamics and immune correlates of coronavirus disease 2019 (COVID-19) severity. Clin Infect Dis. 2020. https://doi.org/10.1093/cid/ciaa1280.

47. Dorigatti I, Lavezzo E, Manuto L, Ciavarella C, Pacenti M, Boldrin C, et al. SARS-CoV-2 antibody dynamics and transmission from community-wide serological testing in the Italian municipality of Vo'. Nat Commun. 2021;12:4383.

48. Gallais F, Gantner P, Bruel T, Velay A, Planas D, Wendling M-J, et al. Evolution of antibody responses up to 13 months after SARS-CoV-2 infection and risk of reinfection. EBioMedicine. 2021;71:103561.

49. Post N, Eddy D, Huntley C, van Schalkwyk MCI, Shrotri M, Leeman D, et al. Antibody response to SARS-CoV-2 infection in humans: a systematic review. PLoS One. 2020;15:e0244126.

50. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22:209–12.

51. Ministry of Health, Singapore. Enhancing surveillance and enforcement efforts to keep COVID-19 at bay: News Highlights; 2020. https://www. moh.gov.sg/news-highlights/details/enhancing-surveillance-and-enfor cement-efforts-to-keep-covid-19-at-bay. Accessed 25 Apr 2021

52. Channel News Asia. COVID-19: MOH conducting "swab operations" as part of stepped up efforts: CNA; 2020. https://www.channelnewsasia. com/news/singapore/covid-19-moh-swab-operations-testing-ringfence- 12961220. Accessed 14 Mar 2021

53. Ministry of Health, Singapore. Circuit breaker to minimise further spread of COVID-19. 2020. https://www.moh.gov.sg/news-highlights/details/ circuit-breaker-to-minimise-further-spread-of-covid-19. Accessed 8 Oct 2022.

54. Ministry of Health, Singapore. Roadmap to phase three. 2020. https:// www.moh.gov.sg/news-highlights/details/roadmap-to-phase-three. Accessed 29 Oct 2021.

55. Ministry of Health, Singapore. Safeguarding lives and livelihoods. 2020. https://www.moh.gov.sg/news-highlights/details/safeguarding-lives- and-livelihoods. Accessed 29 Oct 2021.

56. On PA. Telehealth gets a boost from COVID-19 pandemic - CNA. 2020. https://www.channelnewsasia.com/business/telehealth-gets-boost- covid-19-pandemic-706046. Accessed 4 Jan 2023.

57. Cheng I, Ang HM. Many local COVID-19 cases due to "socially irresponsible" behaviour of a few: Health Minister: CNA; 2020. https://www.chann elnewsasia.com/singapore/covid19-coronavirus-irresponsible-behaviour- doctor-hopping-770806. Accessed 21 Jan 2023

58. 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;2:1–13.

59. Sigal A, Milo R, Jassat W. Estimating disease severity of Omicron and Delta SARS-CoV-2 infections. Nat Rev Immunol. 2022;22:267–9.

60. 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;172:577–82.

61. COVID-19 Forecasting Team. Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. Lancet. 2022;399:1469–88.

62. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. Int J Infect Dis. 2020;101:138–48.

63. Shiraef MA, Friesen P, Feddern L, Weiss MA. Did border closures slow SARS-CoV-2? Sci Rep. 2022;12:1709.

64. Singanayagam A, Hakki S, Dunning J, Madon KJ, Crone MA, Koycheva A, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. Lancet Infect Dis. 2022;22:183–95.

65. Kissler SM, Fauver JR, Mack C, Olesen SW, Tai C, Shiue KY, et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. PLoS Biol. 2021;19:e3001333.

66. Stockbridge M, Purver M, Solel T, Jian A, Arikan D, Ovens R, et al. Technical report: in vitro and clinical post-market surveillance of Biotime SARS-CoV-2 lateral flow antigen device in detecting the SARS-CoV-2 Delta variant (B.1.617.2). 2021. https://assets.publishing.service.gov.uk/ government/uploads/system/uploads/attachment_data/file/999867/

52

in-vitro-and-clinical-post-market-surveillance-of-Biotime-SARS-CoV-2-Lateral-Flow-Antigen-Device-in-detecting-the-SARS-CoV-2-Delta-variant-B.1.617.2.pdf. Accessed 30 Aug 2021.

67. McEvoy D, McAloon C, Collins A, Hunt K, Butler F, Byrne A, et al. Relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: a rapid scoping review. BMJ Open. 2021;11:e042354.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.8 Outro

In this chapter, I performed a retrospective analysis of the relative impact of border control, case finding, contact tracing and other population-wide outbreak control measures. In the next chapter, I used a similar dataset and estimated the risk of disease introduction to inform the need for border control measures in real-time.

3   Patterns of infection among travellers to Singapore arriving from mainland China

By 2022, the dynamics of the COVID-19 pandemic varied in different countries and regions. On one end of the spectrum, most countries in Europe and America no longer impose mandatory outbreak control measures. On the other end, countries such as China were still enforcing a strict zero-COVID policy. As such, when China lifted her outbreak control measures in December 2022, given the presence of a large susceptible population, this resulted in a surge of COVID-19 cases. This presented a potential risk for countries with travel links with China — an influx of COVID-19 imported cases from China. In this study, I explored the use of traveller surveillance data to infer the risk of COVID-19 importation to Singapore and to calibrate the outbreak control measures in real-time.

The supplementary information of this study is in Appendix C.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

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 <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006052	**Title**	Ms
First Name(s)	Rachael		
Surname/Family Name	Pung		
Thesis Title	COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore		
Primary Supervisor	Adam J. 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?	
Please list the paper's authors in the intended authorship order:	Rachael Pung, Adam J. Kucharski, Zheng Jie Marc Ho, Vernon J. Lee
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, investigation, visualisation, and writing of original draft.

SECTION E

Student Signature	*[signature]*
Date	7 March 2024

Supervisor Signature	*[signature]*
Date	13 March 2024

3.1 Abstract

In light of the rapid growth of COVID-19 in mainland China in late 2022, countries and regions outside of China have implemented travel restrictions of varying intensity. Surveillance data of symptomatic travellers arriving from mainland China and detected in Singapore served as a proxy for the COVID-19 outbreak in mainland China. Furthermore, this allowed us to ensure that travel-related restrictions commensurate with the current epidemiological situation and risk.

3.2 Introduction

During 2022, many COVID-19 control measures had been relaxed globally, including travel restrictions [1]. Global dynamics during 2020 were typically driven by local control efforts, which varied substantially, but this had gradually transitioned to dynamics driven by the emergence and spread of novel SARS-CoV-2 strains, combined with varying levels of population immunity. However, countries that have suppressed transmission for longer – such as mainland China – only encountered their first large nationwide wave in late 2022. As a result, Omicron epidemic dynamics in these largely unexposed populations were likely to be considerably different to those underway elsewhere in the world.

This heterogeneity creates multiple challenges. First, there was a need to understand the local transmission dynamics in places that had relaxed public health measures and social interactions later than others, and how these outbreaks differ from other waves globally, particularly as local testing protocols change over time. In turn, there was a need to ensure that the global response to these outbreaks, including any travel-related interventions, commensurate with the epidemiological situation and risk. Throughout the COVID-19 pandemic, data on infections identified among travellers provided crucial real-time situational awareness of international outbreaks [2, 3]. To inform planning and response to outbreaks in mainland China in 2022/23, we therefore used data on infections identified in Singapore among travellers from China, and estimated how infection dynamics in the country of origin changed over time.

3.3 Methods

In Singapore, a confirmed case of COVID-19 is defined as an individual who tests positive via an antigen rapid test (ART) or PCR test administered by the healthcare provider. These cases are notified to the Ministry of Health and cases with a travel history in the 5 days prior to diagnosis are classified as imported cases. Since Apr 2022, travellers were no longer required to take an on-arrival test for COVID-19 and only non-fully vaccinated travellers aged 13 and above were required to take a pre-departure test.

The incidence of imported travellers by the date of notification is not equivalent to the outbreak incidence in the country of origin because of the delay from infection to symptoms onset to testing to notification, and travellers demographics and risk of acquiring COVID-19 may not be similar to the local general population. We therefore adjusted the data to estimate the likely incidence of infection among travellers and assumed that traveller infections reflect the shape of the epidemiological curve, including its peak. For the Omicron variant, the mean incubation was 3.4 days (95% CI 2.88–3.96) [4]; in addition, there will be a brief delay between symptom onset and subsequent testing. Therefore, we assumed a mean delay of 5 days from infection to testing among imported cases and shifted the epidemic curve by this delay to estimate the expected number of infections per day [5]. Normalising by arriving traveller volume, we then estimated the incidence of COVID-19 per 1000 daily incoming travellers from mainland China and Hong Kong as a proxy for the outbreak situation in the respective regions. As a sensitivity analysis, we also generated the results using a delay of 4 days from infection to testing (i.e. no delay from symptoms onset to testing).

3.4 Results

From 1 November 2022 to 5 January 2023, the average number of daily arrivals from mainland China was 553 (IQR 395–671). There were 207 imported cases in total identified from mainland China and all were detected within 5 days of arrival. Two cases, aged 80 and above, from mainland China were hospitalised, of which one was admitted into ICU. From 1 December 2022, from travellers' data, we estimated that the outbreak in mainland China grew at a rate of 0.16 per day (i.e. doubling time of 4.3 days) and peaked around 15 Dec 2022 (Figure 3.1A). From 15 Nov 2022 to the peak of the outbreak, we estimated a cumulative attack rate of 14% among travellers, and by the end of 2022, it was 31% (Figure 3.1B).

As a sensitivity analysis, we omitted 3 imported cases from mainland China as these cases had a positive test after 5 or more days since their date of arrival, given they may have been infected outside their country of origin. Using this subset of data, we estimated a lower outbreak growth rate of 0.15 per day (i.e. doubling time of 4.5 days) in mainland China, with the peak occurring on 16 December 2022, and a lower cumulative attack rate of 11% at the height of the outbreak (Supplementary Figure 1).
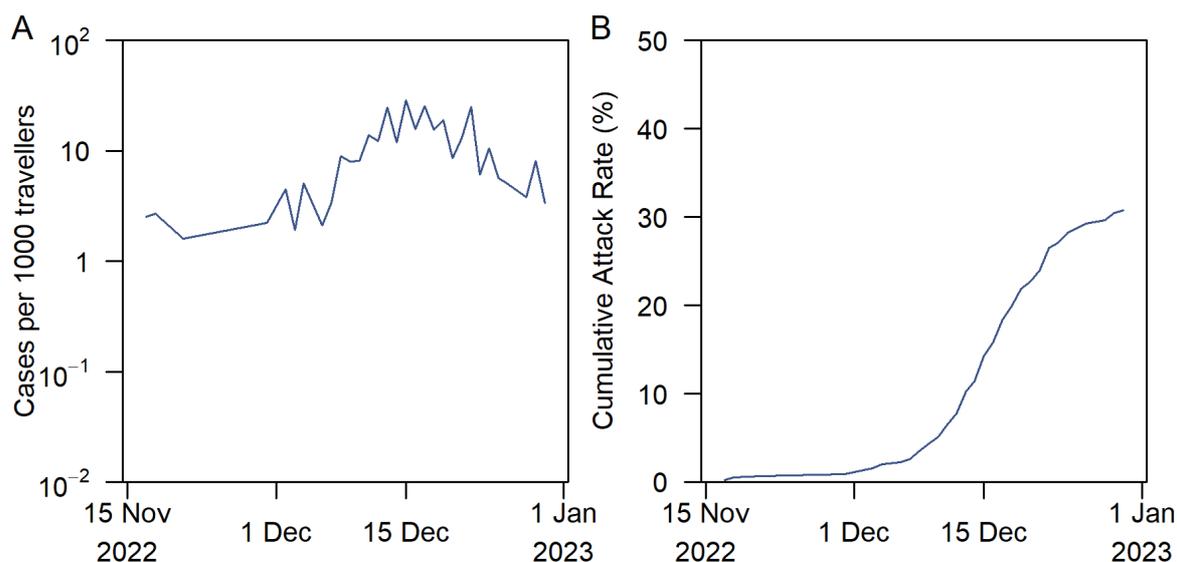
Figure 3.1: COVID-19 outbreak metrics for mainland China. (A) Estimated number of COVID-19 cases per 1000 travellers arriving from mainland China and (B) cumulative attack rate.

3.5 Discussion

Using locally reported cases in Singapore with a known travel history to mainland China, we estimated the number of arriving infections over time from these two regions. Our results showed the benefits of combining epidemiological information, including date of notification and symptoms status, with travel movements to understand importations as well as COVID-19 outbreak dynamics internationally.

We estimated that infection incidence among travellers from mainland China peaked at around 29 cases per 1000 travellers per day (i.e. 2.9%) in mid-December 2022. For comparison, estimates based on community testing in the UK suggested a peak in incidence of just under 1% in early 2022 [6]. This was consistent with the larger initial reproduction number observed in mainland China, which, all things being equal, would typically lead to a shorter epidemic with a larger peak.

There are some limitations to our analysis. Local testing approaches may potentially miss asymptomatic cases, which implies the cumulative attack rate among travellers would be higher in reality. Our estimate of a 30% attack rate up to the end of 2022 is lower than the estimated 75% attack rate for Beijing [7]. However, this difference may also be down to regional differences in mainland China and/or travellers being non-representative of the wider population. There could also be heterogeneity in other demographics, for example, models using traveller incidence which reflect the proportion of travellers from urban cities compared to rural areas. The data we analysed from Singapore also contained no information on patterns of under-ascertainment over time (e.g. as measured by contact tracing and proactive case finding) but if the proportion of infections was reasonably constant during the period

analysed, under-ascertainment would not affect the estimates of the growth rate and time of peak. Testing behaviour among travellers might also change over time, but this was unlikely to be a major issue for Singapore since no travel measures were introduced or removed during the period of interest.

Despite the widespread deployment of travel restrictions by many countries, such as pre-departure and post-arrival testing, in December 2022 in response to the wave in mainland China, the impact on transmission in those countries was likely to be limited. During a growing epidemic, most infections would have been recent, which means that many travellers will be incubating or in the early stage of infection when they are less likely to test positive. Moreover, we estimated that the incidence among travellers from mainland China was already peaking by the time many countries introduced border control measures in mid-December 2022. The level of traveller incidence also suggested that the absolute number of infected travellers would have been many times smaller than the number of locally reported cases during the same period in many countries introducing such measures. This is especially true since flight volumes out of China were low over this period.

As well as missing infections among travellers, testing at the point of departure does not provide situational awareness of the country of arrival. In contrast, general testing of incoming travellers, either at the border for those arriving from countries with ongoing outbreaks [8] or following subsequent symptom onset, as in the analysis presented here, can provide insights into the situation from across the world. As a long-term measure, there could be challenges with estimating importation patterns from locally detected cases if testing behaviours among travellers were to change. Thus, surveillance and control efforts in a country should be tailored to the epidemiological situation and current COVID-19 response objectives in a given country [9], and traveller surveillance will be a good adjunct tool.

3.6 References

1. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus Pandemic (COVID-19). Our World in Data. 2020.

2. 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;20:553–8.

3. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 2020;395:689–97.

4. Wu Y, Kang L, Guo Z, Liu J, Liu M, Liang W. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. JAMA Network Open. 2022;5:e2228008.

5. Petermann M, Wyler D. A pitfall in estimating the effective reproductive number Rt for COVID-19. Swiss Medical Weekly. 2020;150:w20307–w20307.

6. Abbott S, Funk S. Estimating epidemiological quantities from repeated cross-sectional prevalence measurements. 2022;:2022.03.29.22273101.

7. Leung K, Lau EHY, Wong CKH, Leung GM, Wu JT. Estimating the transmission dynamics of Omicron in Beijing, November to December 2022. 2022;:2022.12.15.22283522.

8. Kucharski AJ, Chung K, Aubry M, Teiti I, Teissier A, Richard V, et al. Real-time surveillance of international SARS-CoV-2 prevalence using systematic traveller arrival screening. 2022;:2022.10.12.22280928.

9. Kucharski AJ, Jit M, Logan JG, Cotten M, Clifford S, Quilty BJ, et al. Travel measures in the SARS-CoV-2 variant era need clear objectives. The Lancet. 2022;399:1367–9.

3.7 Outro

Real-time outbreak analysis helps us understand the outbreak dynamics in other countries and prepare for the potential surge in imported cases. Furthermore, real-time outbreak analysis can also help to inform the local outbreak dynamics and the corresponding outbreak control measures required — this will be explored in the next chapter.

4 Serial intervals in SARS-CoV-2 B.1.617.2 variant cases

First detected in India, the SARS-CoV-2 Delta variant (also known as B.1.617.2) resulted in widespread transmission amidst ongoing vaccination efforts. This widespread transmission could be due to a shorter timescale of transmission, the transmissibility of the virus, or both. Thus, establishing the reasons for the growth in cases was necessary to calibrate the outbreak control measures. In this real-time study, I compared the epidemiological data of the wild-type SARS-CoV-2 and Delta variant infectors and infectees to determine if the timescale of infection was reduced.

This paper was published in The Lancet in August 2021 [1]. The supplementary information of the paper is in Appendix D.

References

1. Pung R, Mak TM, Kucharski AJ, Lee VJ. Serial intervals in SARS-CoV-2 B.1.617.2 variant cases. The Lancet. 2021;398: 837–838. doi:10.1016/S0140-6736(21)01697-4

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

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 <u>for each</u> research paper included within a thesis.

<u>SECTION A – Student Details</u>

Student ID Number	2006052	Title	Ms
First Name(s)	Rachael		
Surname/Family Name	Pung		
Thesis Title	COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore		
Primary Supervisor	Adam J. Kucharski		

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

<u>SECTION B – Paper already published</u>

Where was the work published?	The Lancet		
When was the work published?	10 August 2021		
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.

<u>SECTION C – Prepared for publication, but not yet published</u>

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.

Improving health worldwide   **www.lshtm.ac.uk**

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, investigation, visualisation, and writing of the original draft. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as the original author(s) and the source are credited. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)01697-4/fulltext

SECTION E

Student Signature	*[signature]*
Date	7 March 2024

Supervisor Signature	*[signature]*
Date	13 March 2024

Serial intervals in SARS-CoV-2 B.1.617.2 variant cases

The SARS-CoV-2 lineage B.1.617.2, also known as the delta variant, was declared a variant of concern by WHO on the basis of preliminary evidence suggesting faster spread relative to other circulating variants.[1] However, the epidemiological factors contributing to this difference remain unclear. In particular, an increase in observed growth rate of COVID-19 cases could be the result of a shorter generation interval (ie, the delay from one infection to the next) or an increase in the effective reproduction number, R, of an infected individual (ie, the average number of secondary cases generated by an infectious individual), or both.[2] Whereas a shorter generation interval would increase the speed but not the number of individual-level transmissions, a larger value of R would require both faster and wider coverage of outbreak control measures such as vaccination or physical distancing to suppress transmission.

In Singapore, whole-genome sequencing is done for respiratory samples from individuals who tested positive for SARS-CoV-2 by PCR with a cycle threshold of 30 and below. The B.1.617.2 variant was first identified in local cases on April 27, 2021. Despite high levels of adherence to mask wearing and physical distancing in the country,[3,4] clusters of B.1.617.2 variant were detected, and some clusters displayed rapid growth of infections.

We investigated possible drivers of B.1.617.2 variant growth by studying the serial intervals (ie, onset-to-onset delay, a proxy for the generation interval) between pairs of a primary case and a secondary case occurring among household members. Exposure histories were reviewed for all household transmission pairs involving individuals infected with the B.1.617.2 variant and notified between April 27

and May 22, 2021. The B.1.617.2 variant was detected in 97% of the sequenced samples from local cases of COVID-19 identified in this period. Secondary cases with potential exposure to either more than one primary case in the household or to other cases outside the household were omitted from analysis. Households with secondary cases having different symptom onset dates were also omitted from the analysis as we were unable to rule out multiple generations of transmission.

For comparison, we identified household transmission pairs before the partial lockdown in Singapore on April 7, 2020, and applied the same exclusion criteria. This time period precedes the occurrence of the major global SARS-CoV-2 variants and most closely matches the social activity and workplace arrangements in April, 2021,[5] when working from home was not the default. Preliminary analysis showed that the primary cases in this period had a wider range of time from symptom onset to isolation as compared to the B.1.617.2 primary cases (appendix). As such, the following sampling procedure was done to ensure that we matched the number of transmission pairs and the distribution of time from symptom onset to isolation of primary cases. For a given time from symptom onset to isolation of a B.1.617.2 primary case, we randomly sampled, with replacement, the serial intervals of primary cases in the earlier period with matching time from onset to isolation. We then fitted a skewed normal distribution to the sample of serial intervals to account for negative serial intervals arising from pre-symptomatic transmission. The process was repeated 1000 times to obtain the mean and 95% CI of the sample mean, the median, mode, and the difference of these statistics between the B.1.617.2 variant cases and those cases detected before the lockdown.

There were 32 B.1.617.2 variant household transmission pairs, and 63 household transmission pairs identified before April 7, 2020. The median serial interval of the B.1.617.2

variant cases was 3 days, whereas in cases identified before April 7, 2020, the median serial interval was 3 days (95% CI 2 to 4) after matching the time from symptom onset to isolation (figure). The mode of the serial interval was 2 days for B.1.617.2 variant cases and 2·7 days (95% CI –1 to 4) for cases detected before the lockdown. The mean, median, and mode of the serial interval distributions of B.1.617.2 variant cases and the sampled cases before the lockdown was not statistically different (appendix).

This early investigation of recent B.1.617.2 variant cases offers no evidence to support a large difference
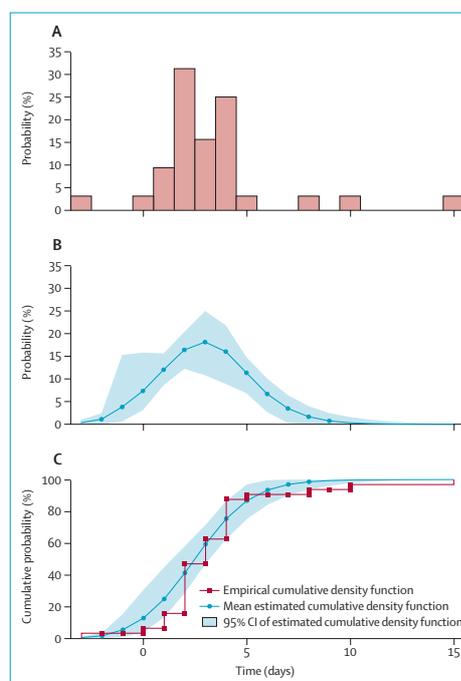
Figure: **Probability mass function of serial interval of SARS-CoV-2 variant B.1.617.2 cases (A), probability density function of serial interval of cases identified before the partial lockdown on April 7, 2020 (B), and empirical cumulative density function of serial intervals and estimated cumulative density function of serial intervals (C)**
Most primary cases had known exposure (or exposures) outside the household and secondary cases do not have the same exposure as the primary case thereby allowing the directionality of infection to be identified. Negative serial intervals, which signify pre-symptomatic transmission, were also included in the analysis.

(ie, >1 day) in serial intervals among the samples studied, which had an exclusion criteria applied to ensure consistency. In turn, this lends support to the hypothesis that the recent rapid growth is potentially driven by an increase in the average number of secondary cases generated by a case infected with the B.1.617.2 variant. Studies with proper control of confounding factors are thus crucial to tease out the key epidemiological factors that facilitate the increased transmissibility of the B.1.617.2 variant. These factors include, but are not limited to, the viral load and shedding dynamics in individuals infected with the B.1.617.2 variant of SARS-CoV-2, the exposure settings, and the vaccination status of infected individuals. Without signs of lowered disease severity for the B.1.617.2 variant, contact tracing and testing around COVID-19 cases, along with vaccination and non-pharmaceutical interventions, continue to remain key SARS-CoV-2 outbreak control measures in the short term.

Rachael Pung, Tze Minn Mak, CMMID COVID-19 working group, Adam J Kucharski, Vernon J Lee
rachaelpung@hotmail.com

Ministry of Health, Singapore (RP, VJL); Centre for the Mathematical Modelling of infectious Diseases (RP, AJK) and Department of Infectious Disease Epidemiology (RP, AJK), London School of Hygiene & Tropical Medicine, London, UK; National Public Health Laboratory, National Centre for Infectious Diseases, Singapore (TMM); Saw Swee Hock School of Public Health, National University of Singapore, Singapore (VJL)

1 WHO. Weekly epidemiological update on COVID-19 - 11 May 2021. May 11, 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19--11-may-2021 (accessed May 23, 2021).

2 Park SW, Bolker BM, Funk S, et al. Roles of generation-interval distributions in shaping relative epidemic strength, speed, and control of new SARS-CoV-2 variants. *medRxiv* 2021; published online May 5. https://doi.org/10.1101/2021.05.03.21256545 (preprint).

3 Lim JM, Tun ZM, Kumar V, et al. Population anxiety and positive behaviour change during the COVID-19 epidemic: cross-sectional surveys in Singapore, China and Italy. *Influenza Other Respir Viruses* 2021; **15:** 45–55.

4 Sim S. More than 6 in 10 Singaporeans likely to continue good hygiene practice after Covid-19: NTU study. Jan 26, 2021. https://www.straitstimes.com/singapore/more-than-6-in-10-singaporeans-likely-to-continue-good-hygiene-practice-after-covid-19-ntu (accessed May 23, 2021).

5 Ministry of Health, Singapore. Expansion of vaccination programme; further easing of community measures. March 24, 2021. https://www.moh.gov.sg/news-highlights/details/expansion-of-vaccination-programme-further-easing-of-community-measures (accessed May 23, 2021).

Towards a European strategy to address the COVID-19 pandemic

Reduction of COVID-19 incidence across Europe in the early spring months of 2021 led to substantial relaxation of restrictions in summer, despite the emergence and spread of the more transmissible SARS-CoV-2 delta variant. As expected, this relaxation led to a renewed increase in incidence. How should Europe act, what strategies should it adopt, and what specific risks should it consider moving forward?[1] These questions become even more pressing, since emerging data indicates the delta variant is more infectious and partially evades immune response. Europe needs a coherent and effective strategy before schools fully reopen and the transmission of SARS-CoV-2 further increases due to seasonality in autumn.

Two opposing strategies are considered: either continue to rapidly lift restrictions, assuming the combination of past natural exposure and current vaccination coverage would allow a high incidence to continue, without overburdening health-care systems; or lift restrictions at the pace of vaccination progress with the core aim to keep incidence low, given this effectively and efficiently controls the pandemic via test-trace-isolate (TTI) programmes.[2,3]

Given immunisation levels as of August, 2021, the first strategy can lead to an incidence of several hundred cases per million per day, whereas the second strategy would require an incidence of well below one hundred cases per million per day. Such a discrepancy of incidence poses considerable friction to European cooperation, economy, and society: high incidence in one country puts the low-incidence strategy in a neighbouring country at risk. Because of this conflict of interest, some countries impose testing and quarantine requirements, hampering international exchange. Thus, either strategy can only work effectively if European countries stop acting as if they could fight the pandemic on their own.

The EU's Digital Covid Certificate (EU DCC) has been introduced to facilitate cross-border travel. However, no vaccine is completely effective at preventing virus transmission. Therefore, the implementation of the EU DCC must be accompanied by systematic evaluation of its contribution to the spread of present and future variants of concern (VOCs).[4] The development of a European strategy for testing travellers and commuters is therefore warranted.[5]

The advantages of low incidence are known and include: (1) less mortality, morbidity, and long COVID; (2) solidarity with those not yet protected; (3) lower risk of new VOCs emerging and spreading; (4) increased feasibility of comprehensive TTI; (5) less workforce in quarantine and isolation, including those in health care; and (6) ensuring schools and childcare remain open during the coming autumn-winter season.[6] In contrast, a high incidence might still overwhelm hospitals and intensive care units in some countries, as estimated in the appendix.

68

4.3 Outro

While real-time analysis in an ongoing outbreak provides an early assessment of the outbreak situation, the robustness of the results can be limited by a small sample size, such as the study in this chapter. In the next chapter, I performed a simulation study to understand how sample size, along with other pathogen and non-pathogen-related factors, could affect our interpretation of the timescales of transmission.

5 Detecting changes in generation and serial intervals under varying pathogen biology, contact patterns and outbreak response

During the pandemic, studies on the generation and serial intervals characterise the timescale of transmission. However, these studies often do not report or adjust for external factors such as the delay from onset-to-isolation, social contact patterns and exponential growth phases [1–5]. These factors are independent of the pathogen biology but potentially affect the time from onset-to-transmission and, hence, the generation and serial intervals. Furthermore, studies that compare these intervals in different outbreak periods seldom report the power to detect changes in the intervals given a study sample size.

Using a simulation framework, I sampled the incubation period of infectors and infectees and modelled stochastic transmission based on the infectiousness period of the infector. By simulating transmission pairs, there is complete knowledge of the modelled generation and serial interval distribution. The modelled serial interval distribution is akin to the observed serial interval distribution when the sample size is large. However, the modelled generation interval distribution is rarely observed and is akin to a theoretical generation interval distribution. In practice, generation intervals are often proxied using serial intervals. Thus, in this study, I estimated the power to detect a change in the mean theoretical generation, the observed serial intervals and the derived generation intervals for a given sample size under varying pathogen and non-pathogen-related factors. This would help inform future outbreak study designs to ensure the robustness of study outcomes.

This paper was accepted by PLOS Computational Biology on March 2024. The supplementary information of the paper is in Appendix E.

References

1. Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109. doi:10.1126/science.abc9004

2. 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. 2021 Oct p. 2021.10.21.21265216. doi:10.1101/2021.10.21.21265216

3. Ryu S, Kim D, Lim J-S, Ali ST, Cowling BJ. Serial Interval and Transmission Dynamics during SARS-CoV-2 Delta Variant Predominance, South Korea -

Volume 28, Number 2—February 2022 - Emerging Infectious Diseases journal - CDC. [cited 10 Jul 2022]. doi:10.3201/eid2802.211774

4.   Zhang M, Xiao J, Deng A, Zhang Y, Zhuang Y, Hu T, et al. Transmission Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2 — Guangdong Province, China, May–June 2021. CCDCW. 2021;3: 584–586. doi:10.46234/ccdcw2021.148

5.   Kang M, Xin H, Yuan J, Ali ST, Liang Z, Zhang J, et al. Transmission dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant infections in Guangdong, China, May to June 2021. Eurosurveillance. 2022;27: 2100815. doi:10.2807/1560-7917.ES.2022.27.10.2100815

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 <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006052	**Title**	Ms
First Name(s)	Rachael		
Surname/Family Name	Pung		
Thesis Title	COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore		
Primary Supervisor	Adam J. 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?	PLOS Computational Biology
Please list the paper's authors in the intended authorship order:	Rachael Pung, Timothy W. Russell, Adam J. Kucharski
Stage of publication	**In press**

Improving health worldwide   **www.lshtm.ac.uk**

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, investigation, visualisation, and writing of original draft.

SECTION E

Student Signature	*Cathrael*
Date	7 March 2024

Supervisor Signature	*Adam Kucharski*
Date	13 March 2024

5.1 Abstract

The epidemiological characteristics of SARS-CoV-2 transmission have changed over the pandemic due to the emergence of new variants. A decrease in the generation or serial intervals would imply a shortened transmission timescale and, hence, outbreak response measures would need to expand at a faster rate. However, there are challenges in measuring these intervals. Alongside epidemiological changes, factors like varying delays in outbreak response, social contact patterns, dependence on the growth phase of an outbreak, and effects of exposure to multiple infectors can also influence measured generation or serial intervals. To guide real-time interpretation of variant data, we simulated concurrent changes in the aforementioned factors and estimated the statistical power to detect a change in the generation and serial interval. We compared our findings to the reported decrease or lack thereof in the generation and serial intervals of different SARS-CoV-2 variants. Our study helps to clarify contradictory outbreak observations and informs the required sample sizes under certain outbreak conditions to ensure that future studies of generation and serial intervals are adequately powered.

5.2 Author summary

Generation and serial intervals quantify the timescale of a transmission process from one person to another. In turn, this informs the speed required to expand outbreak control efforts, especially when we encounter a change in the biological properties of the pathogen. However, shifts in human contact patterns and evolving outbreak response measures can collectively bias the interpretation of these intervals. Using a simulation framework, we estimated the power to detect a difference in these intervals under the influence of multiple factors and investigated the potential for bias in generation and serial interval estimates for COVID-19.

5.3 Introduction

When novel SARS-CoV-2 variants of concern have been identified, a crucial question has been how the epidemiology of the emerging variant relates to the current dominant variants. Novel variants may exhibit multiple phenotypic changes, including changes in the viral load trajectory [1–3], incubation period [4,5], generation interval [5,6] and serial interval [7,8]. Quantifying these epidemiological characteristics is essential to interpret the relative transmissibility of variants of concern and, thus, the potential effectiveness of individual and population-level outbreak control measures. However, comparing specific variants can be challenging owing to changing population-level epidemic dynamics, shifts in human contact patterns and evolving outbreak response measures. In turn, these factors can bias conclusions about the extent to which observed changes in variant dynamics are the result of inherent viral properties, rather than characteristics of the population in which they are spreading. Despite efforts to

compare different aspects of SARS-CoV-2 variant epidemiology, the potential magnitude and direction of such biases in general remain unclear.

Two epidemiological parameters that are particularly important for interpreting growth patterns are the generation and serial intervals. When variant prevalence grows rapidly within a population, it may be the result of increased transmissibility, a shorter delay from one infection to the next, or both. The generation interval is commonly used to define this transmission timescale (i.e. average time between infection of infector and infection of infectee). This interval is a combination of both a host's infectiousness profile since time of infection as well as the timing of social contacts between this primary case and potential infectees. However, because infection times are rarely observed, serial intervals (i.e. time between symptom onset in an infector and an infectee) are often used either as proxies, or to infer the times of infection — and hence the generation interval — over a range of exposure times [9,10]. This can result in several potential biases. Observed serial intervals based on the onset times of infectees are shorter during the exponential phase of an outbreak because transmission events involving most infectees with longer incubation periods have yet to be observed [11]. Furthermore, shorter delays from symptom onset-to-isolation of cases over the course of an outbreak truncates the infectiousness profile and, hence, the serial interval [12]. Large-scale movement restrictions could also influence the relative contribution of household and non-household interactions to transmission and the overall distribution of generation and serial intervals [6,12]. Even if analyses were confined to household contacts only, the timing of contact may not be consistent across the days. As such, broad assumptions such as constant contact over time could potentially diminish the ability to detect differences in the generation and serial intervals between existing and novel variants and, hence, distinguish between a more transmissible variant, and merely a faster one [7].

Using a high-resolution dataset on pre-pandemic human social interactions collected from a large-scale UK study of 469 community participants [13], we parameterised a transmission model of SARS-CoV-2 and other epidemic-prone pathogens to understand factors influencing observed differences in generation and serial intervals during outbreaks. We explored factors including varying viral epidemiological characteristics, isolation strategies, epidemic dynamics, contact patterns between pairs of individuals and within household settings (i.e. competing infectors). Furthermore, we estimated the statistical power to detect these differences between variants and, hence, the potential for bias in variant estimates, using the COVID-19 pandemic as a case study.

5.4 Results

To understand changes in estimated generation and serial intervals, we simulated the incubation period of infector and infectee pairs, and modelled stochastic transmission based on the infectiousness profile of the infector (Figure 5.1A). We simulated 1,000 transmission pairs under varying pathogen biology and outbreak control measures before comparing the modelled serial intervals with those reported in real-life outbreaks to validate the modelling framework.

5.4.1 Influence of pathogen biology and outbreak control measures on observed serial interval

The magnitude of any differences in generation and serial intervals depends on pathogen biology and the transmission process (Figure 5.1A and Table 5.1). At low levels of peak viral load (i.e. average probability of infection per contact pair is less than 25%), we estimated that the median serial interval decreased by less than 0.5 days as probabilities of infection per contact increased (Figure 5.1B and Supplementary Figure 2). These changes were small and approximately linear, as the first-order term in a Poisson process dominates when the force of infection is low.

At higher levels of peak viral load, we estimated that for diseases with a moderate pre-symptomatic period, such as SARS-CoV-2, median serial intervals decreased by 0.7 days when the probability of infection increased from 25% to 50%; and decreased by a further 0.9 days when the probability of infection increased from 50% to 75%. However, for diseases with a short pre-symptomatic phase such as influenza, the decline in the median serial interval was 0.2 days for a probability of infection increasing from 25% to 50% and 0.3 days for an increase from 50% to 75%. Thus, the influence of peak viral loads on serial intervals is greater when viral loads are high with a longer pre-symptomatic infectiousness phase.

Comparing the simulated median serial intervals with the range of observed values in the literature, substantial variation was observed for diseases such as smallpox and SARS-CoV-2 (Figure 5.1B). For smallpox, this variation could be attributed to the long incubation period and duration of infectiousness. For the SARS-CoV-2 wild type and the Delta variant, our analysis suggested that this variation could be attributed to changes in the duration of symptoms onset-to-isolation over the course of the outbreak (Figure 5.1C). Thus, besides pathogen biology, serial intervals are also influenced by population-level outbreak control measures [12], which would need to be controlled for when comparing the epidemiological properties of variants.

We estimated that the difference in the median serial intervals for wild-type SARS-CoV-2 and the Delta variant would not exceed one day across a range of values for symptoms onset-to-isolation, and the interquartile range overlapped considerably

(Figure 5.1C and Supplementary Figure 3). This suggests that there is an inherent epidemiological constraint to detect large serial interval differences for these specific variants, even under very different control scenarios. Published estimates on the serial intervals of SARS-CoV-2 wild type for the respective delay in symptoms onset-to-isolation followed a broadly similar pattern to these model predictions (Figure 5.1C).
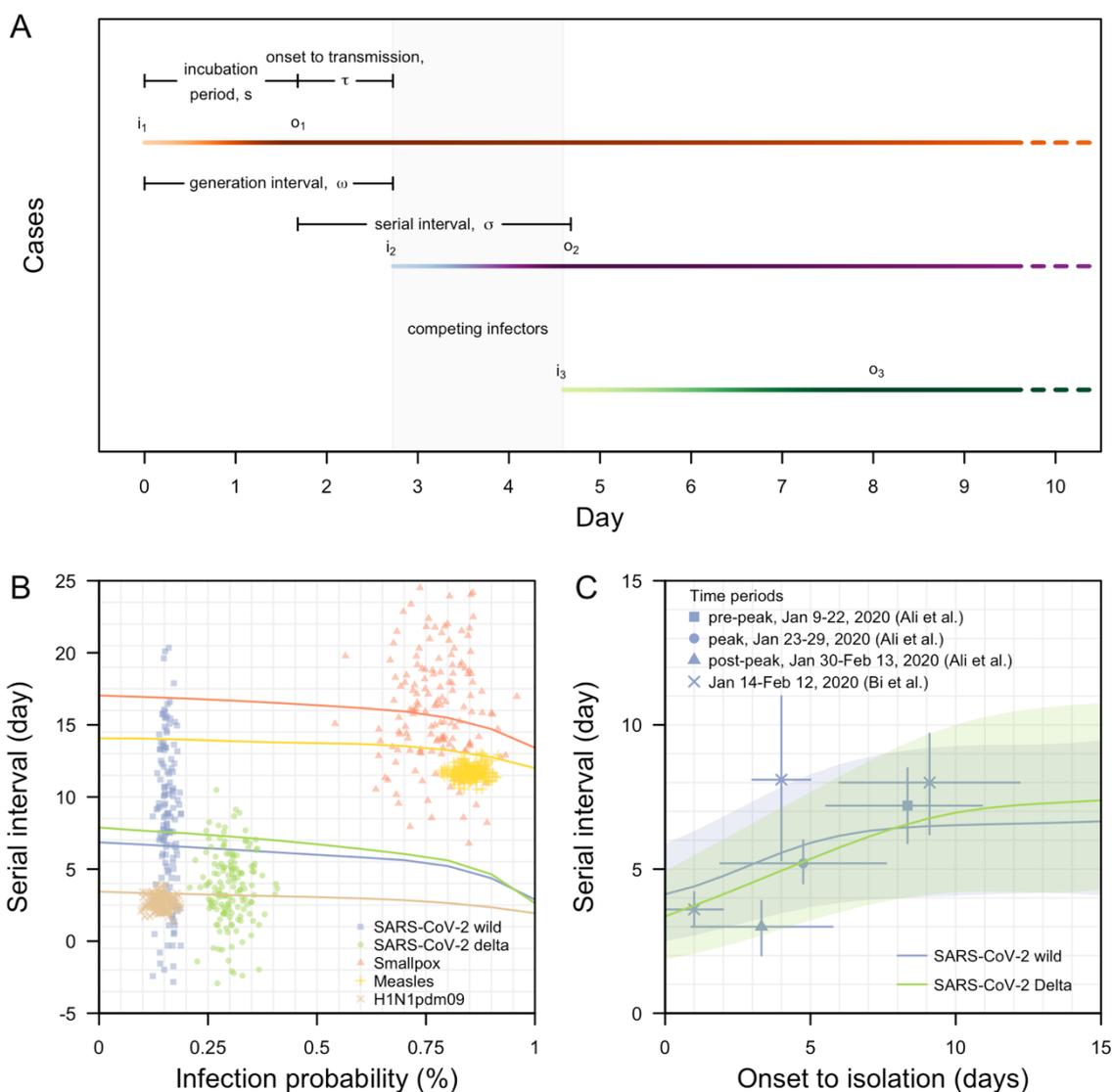
Figure 5.1 Transmission dynamics of infectious diseases. (A) Definitions of epidemiological time intervals and illustration of transmission events on calendar timescales; (B) Modelled (lines) median serial intervals for varying peak infectiousness and hence overall probability of infection for the duration of infectiousness of respective diseases. Range of observed serial interval and attack rate (a proxy for infection probability) for respective diseases in Table 5.1 (points) for comparison; (C) Modelled serial interval for varying delay in case onset-to-isolation in SARS-CoV-2 wild-type and Delta variant with median (lines) and interquartile range (shaded

regions). Observed serial intervals from published studies [12,14] as shown in points (mean) with lines (95% CI).

5.4.2 Power to detect differences in generation and serial interval for transmission pairs

Any measured difference in the mean generation and serial interval based on available data depends on the statistical power of the analysis (i.e. ability to correctly detect a true difference of a given magnitude). To estimate this power, we first need to identify the combination of biological and epidemiological characteristics that would give rise to a particular difference between a reference and an alternative pathogen.

For each combination of biological and epidemiological characteristics, we simulated 1,000 transmission pairs with full knowledge of the time of events (e.g. infection, isolation). In our baseline scenario, we assumed constant outbreak dynamics (i.e. no exponential growth or decay). We then compared the difference in the means of the generation intervals between a reference and an alternative pathogen. We used Welch's t-test to compute the power to detect this difference for a given number of transmission pairs; the same steps were repeated for comparing serial intervals.

In reality, serial intervals are more commonly observed and generation intervals are inferred from these observed serial intervals. Thus, in our three-part inference process, we estimated: (i) the power to detect the theoretical difference in the generation intervals, (ii) the power to detect the observed difference in serial intervals, and (iii) the power to detect the inferred difference in generation intervals.

5.4.2.1 Different incubation period between variants

As a case study, we modelled the reference and alternative pathogen to have a similar peak viral load and duration of shedding post-peak viral load as the Delta- and Alpha-like variants, then extracted the combinations of parameters that gave rise to a one-day reduction in the generation interval of the Delta-like pathogen. Under a scenario of either no isolation or mean symptom onset-to-isolation of 8 days, we estimated the incubation period would need to be 1.6 days shorter for the Delta-like variant to generate a one-day shorter generation interval. When the mean symptom onset-to-isolation was 4 days, the corresponding incubation period was 1.4 days shorter to generate a one-day difference in generation interval (Figure 5.2A).

For a sample size of 100 transmission pairs, we used these extracted characteristics to calculate the corresponding theoretical power to detect a one-day difference in the generation interval. We estimated the power was 32% with no isolation in place, 48% with onset-to-isolation of 8 days and 66% for 4 day delay from onset-to-isolation. As the onset-to-isolation time decreases, the power to detect differences in generation interval increases due to the reduced variance in the generation interval distributions of the reference and alternative pathogens (Supplementary Figure 4). Because more

transmission events in the tail of the distribution are being prevented with a rapid onset to insolation, more of the 100 sampled transmission pairs come from samples near the mean.

As a sensitivity analysis, we also modelled the infectiousness profile using a function derived from the analysis of observed wild-type SARS-CoV-2 transmission pairs by Ferretti et al [9]. The differences in the incubation period for a one-day shorter generation interval in the model by Ferretti et al were comparable to our modelled results (Supplementary Figure 5).

When we performed the same analysis comparing serial intervals between variants, we found that differences in the incubation period for a one-day shorter serial interval in the Delta-like variant were similar to those for the generation intervals. However, the power to detect a one-day difference in the serial intervals was lower. For a sample size of 100 transmission pairs, it was 29% with no isolation, 40% with onset-to-isolation of 8 days and 54% for onset-to-isolation of 4 days (Figure 5.2B). Unlike generation intervals, serial intervals are a combination of biological quantities in two individuals: the incubation period in the infectee and the onset-to-transmission of the infector (Figure 5.1A); these quantities were assumed to be biologically independent in our analysis. On the contrary, the generation interval depends only on the infector's delay from infection-to-transmission. This typically results in a lower variance for the generation interval distribution than the serial interval distribution [15], hence more statistical power.

As generation intervals are rarely observed, serial intervals are often used as a proxy for the delay between generations of infection. In general, the mean of the inferred generation interval is similar to the mean of the observed serial interval when the infector and infectee have the same incubation period distribution [15,16]. The variance of the inferred generation interval is dependent on the variance of the serial intervals, the covariance of the onset-to-transmission and the incubation period of the infector. To understand how the changes in the variance affect the power to detect differences in the inferred generation intervals, we explored two extreme scenarios based on [15].

At one extreme, we assumed that the infectiousness of an infector is dependent on the time of symptoms onset only. Under this assumption, the incubation period of the infector and the time from onset-to-transmission are independent and their sum equates to the generation interval. For the same incubation period distribution in the infector and infectee, this assumption implies that the generation interval is the same as the serial interval (sum of time from onset-to-transmission in infector and incubation period of infectee). Thus, the corresponding power to detect the difference in the inferred generation and observed serial intervals is equivalent (Figure 5.2B).

At the other extreme, we assumed that the infectiousness depends on the time since infection of the infector. As such, the time of transmission is not correlated with the time of symptom onset in the infector, and the variance of the derived generation interval is lower than the observed serial interval. The two assumptions serve as the upper and lower limits to the variance of the inferred generation interval (i.e. lower and upper limits of the power). Under the second assumption, the power to detect a one-day difference in the generation interval for a sample size of 100 was 46% with no isolation, 75% with onset-to-isolation of 8 days and 100% for onset-to-isolation of 4 days (Figure 5.2C). These power values are higher than under the assumption of the lower limit for GI (Figure 5.2B), but in practice, this could also increase the chance of a false positive (i.e. concluding a difference in GI when there is not none).
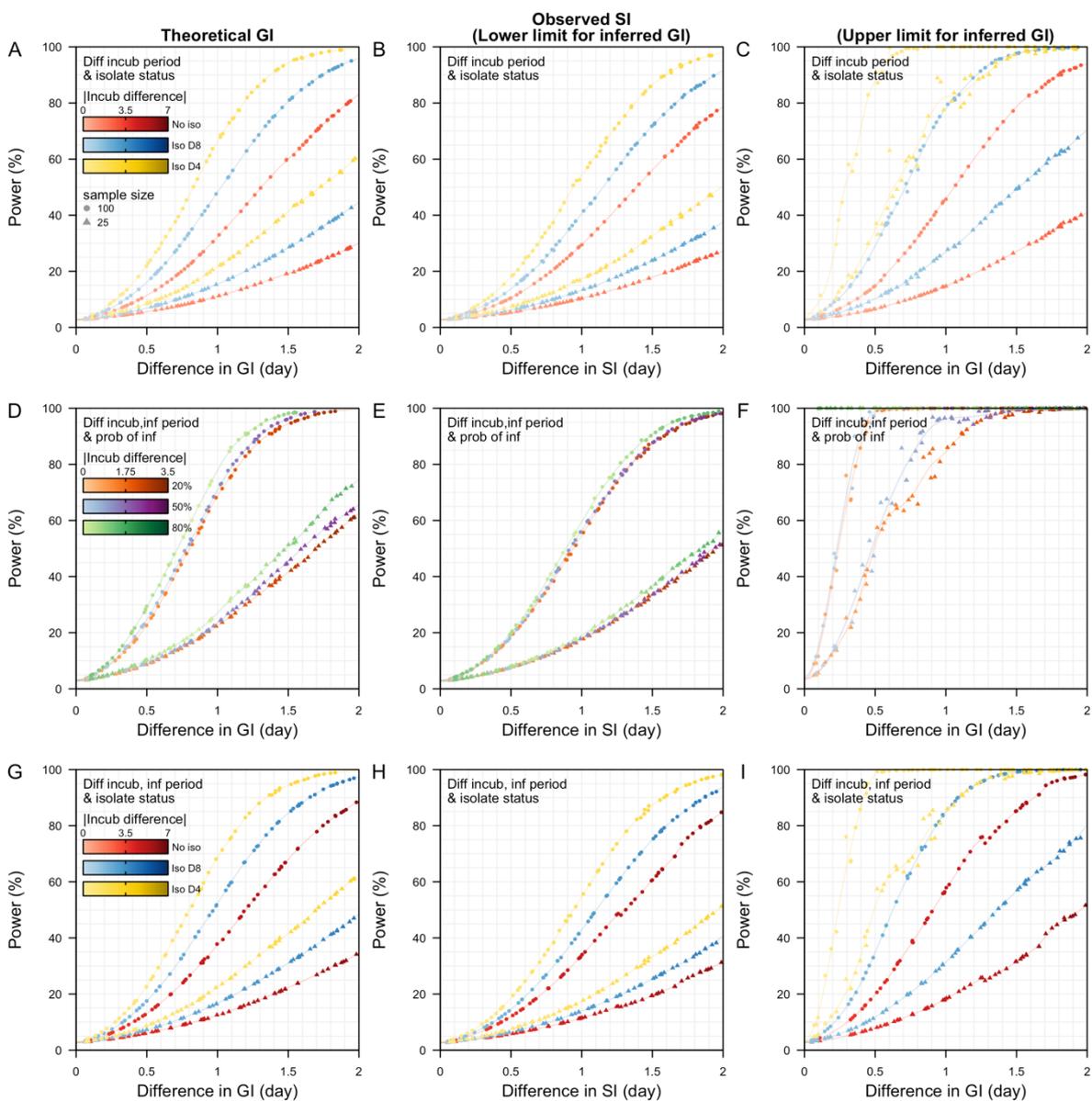
(see captions on next page)

Figure 5.2 Power to detect differences in the generation intervals (GI) and serial intervals (SI) between reference and alternative pathogens. (A-C) Different incubation period between reference and alternative pathogen under the same symptoms onset-to-isolation status of either no isolation, mean symptoms onset-to-isolation of 8 days, or 4 days; (D-F) Different incubation period and longer duration of infectiousness post-peak viral load in reference pathogen under the same mean symptoms onset-to-isolation of 4 days. Peak viral load in reference pathogen is varied, resulting in a probability of infection, p, of either 20%, 50% or 80% when the mean incubation of the reference pathogen was 4 days; (G-I) Different incubation periods and longer duration of infectiousness post-peak viral load in reference pathogen under respective onset-to-isolation. (A,D,G) Theoretical power to detect differences in GI, (B,E,H) power to detect differences in observed SI — lower limit estimates of the theoretical power, (C,F,I) upper limit estimates of the theoretical power.

5.4.2.2 Different incubation period, peak infectiousness and duration of infectiousness

Variants of SARS-CoV-2 can differ by more than one biological characteristic, and different combinations of these characteristics can produce similar differences in the generation and serial intervals. To explore these interactions, we modelled the reference Delta-like variant to have a longer duration of viral shedding (8 days longer) with the same or higher peak infectiousness compared to the alternative wild-type-like pathogen for a range of incubation periods. We did not vary the mean symptoms onset-to-isolation delay (4 days) in order to study the reduction in the generation and serial interval arising from variation in pathogen characteristics only. For the same peak infectiousness (i.e. per-contact probability of transmission for the Delta-like variant equal to 20% with a mean incubation period of 4 days), the incubation period was 1.9 days shorter for the Delta-like variant to give a one-day shorter generation or serial interval. On the contrary, when the peak infectiousness of the Delta-like variant was higher, resulting in a 50% or 80% probability of infection, the corresponding incubation period was 1.3 days or 0.2 days shorter. The theoretical power to detect a one-day difference generation interval was between 70–85% in all three scenarios (Figure 5.2D). Similar differences in the incubation period resulted in a one-day difference in the serial intervals, and the corresponding power was about 50-65% (Figure 5.2E). This serves as the lower limit estimate of the power to detect the same one-day difference in the generation intervals inferred from the observed serial intervals, while the upper limit estimate was 100% (Figure 5.2F). The differences in the incubation period were more pronounced under scenarios of no case isolation (Supplementary Table 1). Even if we account for additional variability in the time of peak infectiousness, allowing it to occur 1–5 days post symptoms onset, the incubation period of the Delta-like variant was 1.3–1.5 days shorter for a 20–50% probability of infection (Supplementary Table 2).

5.4.2.3 Different incubation period and duration of infectiousness

Besides intrinsic differences in biological properties among variants of SARS-CoV-2, vaccination could also shorten the duration of viral shedding and, hence, infectiousness, in vaccinated cases as compared to unvaccinated cases. By modelling the mean duration of infectiousness in an unvaccinated case to be 8 days longer, for a one-day shorter generation interval in the unvaccinated cases, we estimated the incubation period was 5 days shorter in the unvaccinated cases under no case isolation; 2.9 days shorter when the mean symptom onset-to-isolation was 8 days; 1.9 days shorter when the mean onset-to-isolation was 4 days. The corresponding theoretical powers to detect a one-day difference in the generation interval were 37%, 52% and 69% respectively (Figure 5.2G). As compared to previous scenarios (Figure 5.2A and 5.2B), a shorter incubation period in the reference pathogen counteracts the longer shedding profile and narrows the difference in the mean generation interval of both pathogens. The reduction in the incubation period (e.g. 5 days shorter) does not necessarily correspond to the increase in duration of shedding (e.g. 8 days longer).

5.4.2.4 Different contact pattern among household and non-household pairs

Based on measured contact patterns, we also found that frequent contact between household members can result in higher probabilities of infection and earlier infections as compared to non-household contacts for the same pathogen characteristics. For a one-day shorter generation interval among household contacts, the difference in the probability of infection between household and non-household contacts in our baseline scenario was 57% (59% vs 2%) under no isolation; when the mean onset-to-isolation was 8 days, this difference was 65% (68% vs 3%); when the symptoms onset-to-isolation was 4 days, the difference was 74% (78% vs 4%). For a sample size of 100 transmission pairs, the theoretical power to detect a one-day difference in the generation intervals was 33%, 51% and 68%, respectively (Figure 5.3A). Based on previous literature, the probabilities of infection in household pairs are typically estimated to be less than 50% (18, 20). For a 20–40% probability of infection in household contacts, the probability of infection in non-household contacts was 0.6–1% in our analysis. We estimated that the differences in generation and serial intervals among such contacts were 0.2–0.4 days and the corresponding power to detect these differences in the generation and serial intervals were less than 1% (Figure 5.3A-C).

5.4.2.5 Different contact frequency between household pairs

When the frequency of contact is low (e.g. weekly household-type contacts), the timing of the contacts matters more as it determines which portions of the infectiousness period (e.g. start or end) would have the highest concentration of the limited infection opportunities. For the same pathogen, under no case isolation, the frequency of contact can therefore have a considerable impact on transmission risk. In a scenario where the probability of infection was 20% among household members who had daily measured contact with an infectious individual, the corresponding probability of

infection among individuals who had only weekly household contacts dropped to 2%. We estimated that the mean generation and serial intervals were 1.0 and 1.1 days shorter for daily household contacts when there is no case isolation. For a sample size of 100, the theoretical power to detect the differences in the generation and serial intervals was 33% (Figure 5.3D and 5.3E). When the mean duration of symptoms onset-to-isolation was 4 days, the corresponding probability of infection was 64% among household members with daily contact, and 7% for those with weekly contact. The mean generation interval was 0.3 shorter for daily household contacts while the serial interval was 0.2 days longer. The power to detect these differences in both intervals was less than 10% (Figure 5.3D-F).
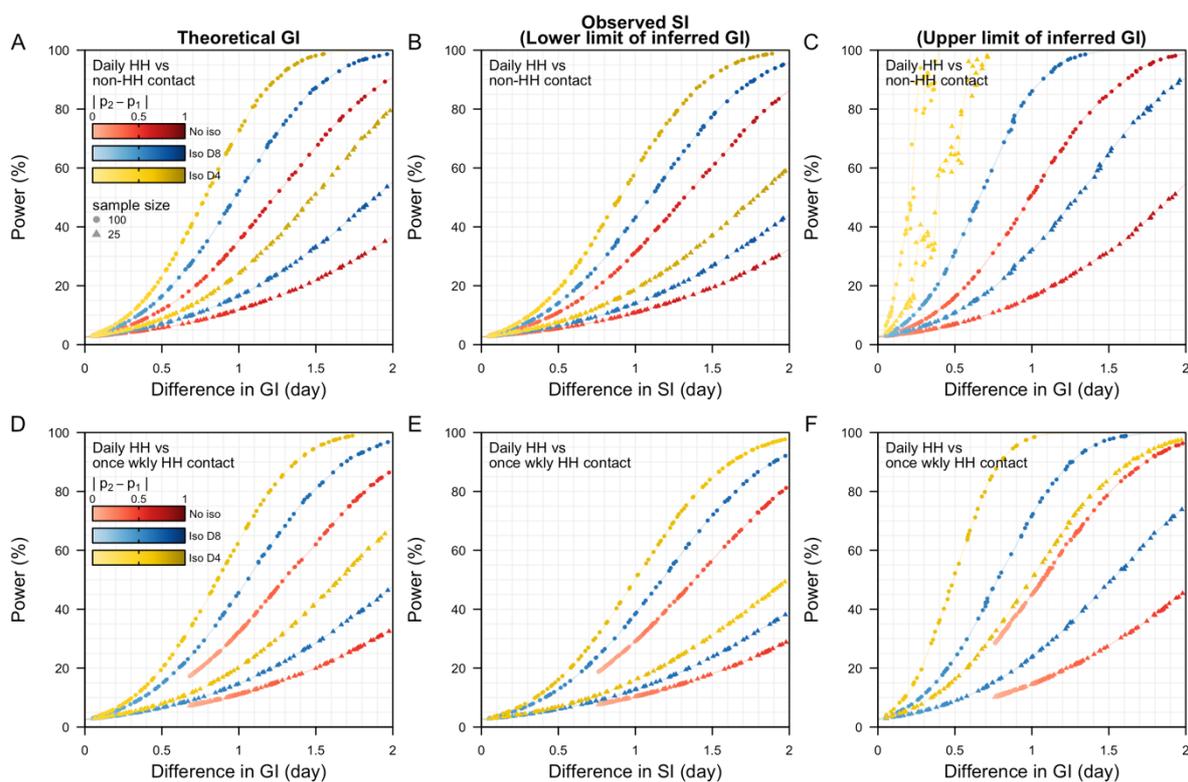
Figure 5.3 Power to detect differences in the theoretical generation intervals (GI), observed serial intervals (SI) and derived GI between reference and alternative pathogen. (A-C) Different contact patterns of either non-household or household contact but same incubation period under respective peak infectiousness and isolation status. Due to the differences in contact frequency, probabilities of infections (p_1 for reference pathogen and p_2 alternative pathogen) are different for both types of contact for the same peak infectiousness; (D-F) Different contact patterns of either daily or weekly household contact but same incubation period under respective peak infectiousness and isolation status. (A,D) Theoretical power to detect differences in GI, (B,E) power to detect differences in observed SI — lower limit estimates of the theoretical power, (C,F) upper limit estimates of the theoretical power.

5.4.2.6 Different epidemic growth dynamics of variants

When a new variant is introduced in a population, the growth rate of this variant and the existing pathogen may differ. Under exponential growth dynamics, the newly observed cases are more likely to be recently infected with shorter incubation periods. The overall incubation period in the population without adjusting for these dynamics will be shorter [11] and potentially bias the measured generation and serial intervals when left unadjusted. Thus, we need to understand the magnitude and direction of this bias.

We modelled a scenario where the reference pathogen has a one-day shorter incubation period but a longer duration of viral shedding (i.e. Delta-like) than the alternative pathogen (i.e. wild-type-like). We also varied the epidemic dynamics of each pathogen. When there was constant growth in both pathogens, the observed mean generation interval was 0.4 days shorter in the reference pathogen. In the absence of bias from the epidemic phase, the true interval would therefore be 0.4 days. When there was exponential growth of 0.2 per day in the reference pathogen and exponential decline of 0.2 per day in the alternative pathogen, the resulting mean generation interval — which is influenced by the combined epidemic process and incubation period distribution – was 2.0 days shorter for the reference pathogen (Figure 5.4A). Under a scenario of constant growth in the reference pathogen but exponential growth of 0.2 per day in the alternative pathogen, the observed mean generation interval was 0.1 days longer in the reference pathogen. Hence, in any analysis, we must simultaneously consider differences in the true generation time and bias from the epidemic phase.

The power to detect a difference In the generation Intervals depends on the extent of overlap in the generation interval distribution of the reference and alternative pathogen. The extent of this overlap is, in turn, affected by the differences in the biological characteristics of each pathogen. However, this overlap also depends on the prevailing outbreak dynamics. Without adjusting for exponential growth and decline dynamics in the reference and alternative pathogen respectively, the extent of overlap in the generation interval distributions of the reference and the alternative pathogen is lesser than in the scenario where both pathogens are at constant incidence (Supplementary Figure 6A and 6B). This accentuates the differences in the mean generation interval, thereby increasing the power to detect this difference and conclude that there exists a non-zero difference between the generation intervals of two pathogens (i.e. lower Type II error) (Figure 5.4B, red square). Furthermore, unadjusted outbreak dynamics can also increase the chance of concluding a difference in the generation intervals when there is none after adjustment (i.e. higher Type I error). However, when we correctly adjusted for the exponential outbreak dynamics, we recovered a similar mean difference in the generation intervals of the reference and alternative pathogen across different combinations of epidemic dynamics and a similar power to detect this difference (Figure 5.4C).

Overall, the sample size of a study should be designed based on the desired power to detect a difference after adjusting for the observed epidemic dynamics. Without adjusting for epidemic dynamics, there is a possibility of accentuating the differences in the generation intervals for a reference pathogen undergoing exponential growth and an alternative pathogen experiencing exponential decline. Consequently, even with a small sample size, there is a large power to detect this biased difference (Figure 5.4B). However, after adjusting for the epidemic dynamics, the power to detect the inherent difference in the generation intervals would be reduced (Figure 5.4C). Thus, it is important to account for epidemic dynamics when planning for the appropriate number of samples for collection under the prevailing or likely outbreak dynamics.
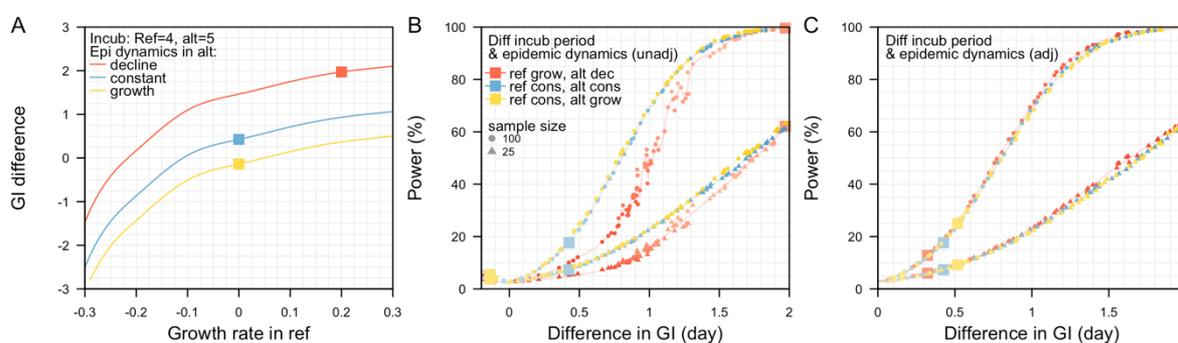

Figure 5.4 Generation intervals, GI, under varying outbreak dynamics. (A) Differences in the generation interval between the reference and alternative pathogen without adjusting for exponential growth or decline outbreak dynamics. Exponential growth of 0.2/day (red), constant outbreak (blue) and exponential decline of 0.2/day (yellow) in alternative pathogen, (B) corresponding power to detect the biased differences in the generation intervals, (C) power to detect differences in the generation intervals after correctly adjusting for exponential outbreak dynamics.

5.4.3 Generation and serial intervals in households with multiple competing infectors

Within household outbreaks, infectors compete for the remaining susceptible individuals, which can influence the dynamics of the observed transmission events in a cluster. For each cluster, we simulated two transmission pairs involving three individuals; the first pair was an index and a secondary case, the second pair was either a secondary and a tertiary case, or the index and another secondary case. We simulated 1,000 clusters and estimated the distribution of the generation and serial intervals over different onset-to-isolation delays.

We estimated that transmission events involving multiple competing infectors resulted in a lower median generation interval as compared to pairwise transmission involving a single infector. The magnitude of this difference increases when the median delay from onset-to-isolation increases (Figure 5.5). For an assumed mean incubation

period of 4 days and onset-to-isolation of 4 days (variance 5 days), the corresponding time from symptoms onset to transmission was 0.9 days (95%CI -3.4–5.8) in pairwise transmission but 0.4 days (95%CI -3.9–5.2) for cluster transmission. Taking the difference between the pairwise and cluster transmission, the mean difference in the generation interval distribution was 0.4 days (95%CI 0.2–0.7) (Figure 5.5B). When the delay from onset-to-isolation was 8 days, the difference in the time from symptoms onset-to-transmission between pairwise and cluster transmission widened, and the mean difference in the generation interval distribution increased to 0.7 days (95% 0.4–1.0). When exposed to multiple infectors, the probability of a susceptible individual being infected in a timestep given no previous infection would increase and hence, reduce the expected time until infection.

The overall incubation period distribution in the modelled transmission events in households with multiple infectors was similar to that in pairwise transmission. In both pairwise and cluster transmission, we modelled the mean incubation period of all primary cases as 4 days (variance 5 days). The mean incubation period of secondary cases with onward transmission was 4.2 days (variance 5.3 days) and the difference in the incubation period between different generations of infectors was 0.2 days (95%CI -0.5–0.03) when infectors were isolated on average 4 days after symptoms onset. Similar outcomes were observed when the delay from onset-to-isolation of infectors was increased to 8 days. As such, while secondary cases with short incubation periods experience earlier onset and peak viral load as compared to the primary cases, they were not observed to transmit more infections to the third susceptible individual to shorten the overall mean incubation period. As the generation and serial intervals are a combination of the time from symptom onset-to-transmission and the incubation period, the shortening of these intervals in a cluster transmission is largely driven by the reduction in the onset-to-transmission rather than the incubation period.

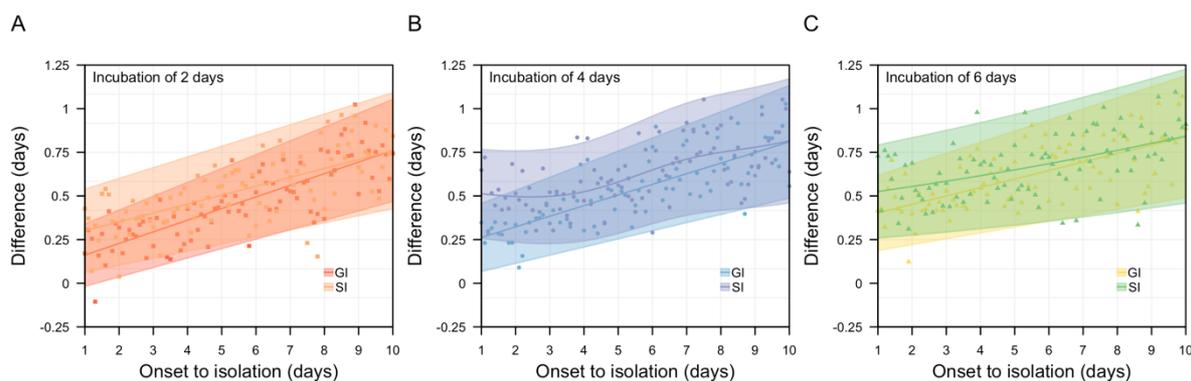

Figure 5.5 Differences in mean generation (GI) and serial (SI) intervals for transmission between pairs (i.e. no competing infector) and triples (i.e. competing infectors) with mean incubation period of (A) 2 days, (B) 4 days and (C) 6 days.

5.5 Discussion

Using a model incorporating high-resolution human interactions, we found that interacting biological and epidemiological processes can have a major impact on the ability to detect changes in observed pathogen generation and serial intervals. Using novel SARS-CoV-2 variants as a case study, we showed that statistical power to estimate differences in the generation or serial intervals between variants can be highly sensitive to factors such as the incubation period and delay from onset-to-isolation. With a large sample size of 100 transmission pairs, the power of studies to detect a one-day change in the generation interval can be 30–70%, depending on the prevailing delay from onset-to-isolation. This power could decline to less than 20% when the sample size is reduced to 25 transmission pairs.

Assuming either a linear or exponential relationship between the generation interval and growth rates [17], if the generation interval decreases by one day (e.g. from 5 days to 4 days), this could result in a 25% increase in growth rate with a reproduction number of 2. In other words, if we compare the initial growth dynamics of the old and new variant, the outbreak trajectory in the latter will double that of the former in about 3 weeks based on the changes in generation intervals only. For SARS-CoV-2 with an initial generation time of about 5–6 days [18,19] and a doubling time of 2-4 days [20], countries have reported taking about 1–3 weeks to expand isolation facilities or testing capacity by at least 2 times at the start of the outbreak [21–23]. Thus, when faced with a novel faster variant, early studies to detect changes in generation intervals, and hence growth rates, may be underpowered. Furthermore, the timescale for expanding healthcare capacity is potentially slower than the outbreak growth rate. Overcoming these challenges would require the implementation of strict population-level outbreak control measures (e.g. physical distancing, mask-wearing) to slow the outbreak at the initial phase, to buy time to expand the healthcare capacity and gather information on the new variant.

Studies with small sample sizes of 30–50 transmission pairs are likely to be underpowered to detect small differences of 1–2 days in the generation or serial intervals [7,24] but our simulation framework allowed us to explore these differences in the absence of biases created during the data collection process. We showed that when the probability of infection is 20–50% and the delay from symptoms onset-to-isolation is 4 days, the corresponding incubation period of the Delta variant would need to be shortened by 1.3–1.9 days to observe a one-day shorter serial interval. When there is no case isolation, a larger difference in the incubation period was required to achieve the same effect. Direct comparison of the incubation periods from different studies suggested that the incubation period of the Delta variant was 0–1.4 days shorter than the wild-type SARS-CoV-2 [15,24–26] and the secondary attack rate of Delta (proxy for probability of infection) ranged from 23.0-37.3% [27]. Taking into consideration the findings from other studies and our modelling outputs, this suggests

that a shorter serial interval of at least one day was not likely to occur between the wild-type SARS-CoV-2 and the Delta variant.

Outcomes from our modelling framework are comparable with other epidemiological studies. The mean generation and serial interval of the Delta household transmission pairs were estimated to be 0.7 days and 1.7 days shorter than that of the Alpha variant [6]. In that study, the mean incubation period of the Delta cases was estimated to be 1.4 days shorter. Based on our modelling framework, the incubation of the Delta variant needs to be about 1–2 days smaller for a one-day reduction in the generation or serial interval, assuming the duration of infectiousness and the peak infectiousness of both variants are similar. Empirical findings from different countries and regions also reported an incubation period of about 4 days for the Delta variant at different time points of the outbreak; a day shorter than the estimated incubation period of the Alpha variant [28,29]. For the same pathogen but different contact frequencies, we estimated small differences in the serial intervals of less than half a day when the probabilities of transmission in non-household members are small. This corroborates with one study estimating an empirical difference of less than 0.5 days between household and non-household members during the peak of the COVID-19 pandemic involving the wild-type SARS-CoV-2 in China in Jan 2020 [12] with a household attack rate of about 20% (a proxy for probability of infection) [30].

While generation and serial interval distributions are shortened due to a decrease in the mean incubation period during an exponential growth phase of an outbreak [11], the occurrence of multiple infectors in a household transmission cluster can also reduce these intervals. This reduction occurs when the time from symptoms onset to infection is shortened. In a modelled cluster with competing infectors, infectors with shorter incubation periods were not observed to preferentially transmit infection to the susceptible individual. Differentiating the reasons for faster outbreak growth is important. If the growth of an outbreak is driven by a true biological reduction in the incubation period, the outbreak control policy would need to focus on rapid and wide contact tracing beyond the household. Exponential growth dynamics may bias our interpretation of the change in a pathogen's incubation period and, hence, changes in the generation and serial intervals, but appropriate adjustments would rectify the bias. On the contrary, if the growth of an outbreak arises from an increase in earlier household transmission, especially during periods of lockdown, control policy would then need to shift towards effective household isolation.

There are some limitations to our study. Firstly, we did not explore the effect of viral dose exposure on the probabilities of transmission over a contact [31]. The duration of a contact can be long and continuous or occur as a series of short contacts with breaks in between. For a continuous contact episode (i.e. a series of consecutive 5-minute contact records), we assumed that the force of infection is summative across the timesteps and constrained the probability of infection among household contacts over the entire period of infectiousness to match the observed secondary attack rates

in households. The lack of in vivo studies on transmission probabilities over contact duration poses a challenge to evaluating dose-response relationships but could be explored in future simulation studies.

Secondly, we did not account for variations in the start and end times of the infectiousness profile, and instead fixed these parameters based on average observed durations of viral growth or decline. Furthermore, the scale factor for the peak infectiousness was not modelled based on a distribution. Accounting for these variations is not likely to affect the mean difference in the generation and serial intervals or the parameters (e.g. mean incubation period) that result in this difference but will lead to reduced power to detect these differences. We estimated the power to detect a difference in the generation intervals with a Welch t-test using the estimated serial intervals for the reference and alternative pathogen. We considered two bounding assumptions about the population-level relationship between the variance of the generation intervals and serial intervals to obtain a plausible range of power values. However, in reality, the inference method for obtaining generation interval distributions could introduce additional uncertainty. If we were to instead try and infer this relationship from individual transmission pairs, it would be important to account for the resulting parameter uncertainty to avoid underestimating the variance of the distributions and, hence, the power.

Thirdly, due to data identifiability issues, the relationship between a pair of contacts used in this dataset was not available and we made a conservative assumption that contact signals within 10m translate to an effective contact. A greater (smaller) radius of detection, would generally lead to more (less) contact episodes between a pair of individuals. We would then expect the scale factor for the peak infectiousness to decrease (increase) in order to achieve the same overall probability of infection for a given range of observed attack rates for a disease. We expect the trends in the overall findings to remain similar but the use of real-world temporal networks with a well-defined edge list between individuals would refine the analysis.

Standardising the contact patterns and effects of non-pharmaceutical interventions to compare changes in pathogen biology and, hence, changes in generation and serial interval in outbreak data is challenging. By simulating known changes in the disease-related factors (e.g. incubation period and duration of infectiousness) and other external factors, we studied how sensitive these intervals were to respective factors. Based on the combination of multiple factors and measured quantities, this helps to clarify contradictory outbreak observations, evaluate the power of detecting such observations and inform future data collection efforts to ensure that studies are well powered.

5.6 Materials and Methods

5.6.1 Contact data

Previously published social contact data recorded interactions among 469 participants over three consecutive days (Thursdays 12 Oct–Saturday 14 Oct, 2017) from 0700–2300 hours each day, as part of the BBC Pandemic study conducted in Haslemere, United Kingdom [13] (Supplementary Figure 1). In the previous published study, participants consented to the collection of their contact data when they downloaded the BBC Pandemic mobile phone application for the purpose of that study. Using secondary data for our analysis, we defined a contact to exist between two individuals when there was a recorded signal in either of their BBC Pandemic mobile phone applications with a GPS distance of at most 10 metres apart in a 5-minute interval. Familial and friendship status were not available in the published individual-level data to avoid re-identification. Thus, we assumed that likely household contacts were represented by pairs of individuals with at least one recorded contact in five out of the six time periods from 0700–0800 hours or 2000–2300 hours over the three days. These time periods are beyond the typical working hours on weekdays before the COVID-19 pandemic [32] and consistency of contact over three consecutive days was assumed to rule out non-household contacts (e.g. commuting) occurring by chance. Based on these assumptions, we identified 54 households with an average size of 2.3, similar to previous survey estimates on household sizes in Haslemere [33], and there were 82 household and 451 non-household contacts.

As the infection process for SARS-CoV-2 typically occurred on timescales lasting more than three days [2,6], we extended the contact between a pair of individuals by randomly sampling their daily contacts over weekdays based on the recorded contact patterns on Thursday and Friday and fixed all weekend contacts based on Saturday. This process was applied to both household and non-household contacts. To study the transmission over once-weekly interactions (e.g. weekly events or meetings), we randomly sampled a day of the week and set all contacts on other days to null.

5.6.2 Infectiousness profile

For each individual, we simulated the start and end time of the infectiousness period, with the time of peak infectiousness for respective diseases relative to the incubation period (Table 5.1). The relative infectiousness ranged from 0 to 1 — normalised relative to the peak infectiousness. We then fitted a cubic Hermite spline through the start, peak and end points of the infectiousness period. We constrained the slope of the spline to be zero at each of the three points (i.e. the first derivative is zero) to simulate the infectiousness profile over the course of the infection. Furthermore, we scaled the splines of the respective diseases such that the probability of infection matches the observed data (Table 5.1).

We used the SARS-CoV-2 Delta variant infectiousness profile in the main analysis to compare differences in the generation and serial interval distributions under changing pathogen biology, contact patterns and outbreak response. For sensitivity analysis, we used the skew-logistic model by Ferretti et al [9] to compare with the findings from our wild-type SARS-CoV-2 spline model. This alternative model concurrently estimates different components of an infectiousness curve (e.g. growth, decline and peak) from observed wild-type SARS-CoV-2 transmission pairs. It assumes a long-tail at the start of the curve for a pathogen with a long incubation period resulting in a longer pre-symptomatic infectious period. However, the model was not updated for subsequent variants. Conversely, the spline model allows for easy parameterisation of each component of the infectiousness curve based on a variant's characteristics derived from separate studies.

Table 5.1 Parameters to model the infectiousness profile of different diseases. References in parenthesis.

Disease	Incubation period	Start of infectiousness	End of infectiousness	Time of peak infectiousness	Secondary attack rate; used as a proxy for average probability of infection per contact pair
SARS-CoV-2 wild	Lognormal mean (log) = 1.62, sd (log) = 0.42 [25]	Gradually increase since time of infection [9,34]	10 days post symptoms onset [27]	At symptoms onset [10,34]	13.2% – 18.2% [27]
SARS-CoV-2 Delta	Weibull shape = 2.23, scale = 4.68 [35]	Gradually increase since time of infection [9,34]	18 days post symptoms onset [36]	At symptoms onset* [9,34]	23.0% – 37.3% [27,37]
Smallpox	Normal mean = 12, sd = 1 [38,39]	Upon symptoms onset [40]	14 days post symptoms onset [41]	Three days post symptoms onset [41]	60.0% – 90.0% [42,43]
Measles	Normal mean = 14, sd = 1.5 [44]	Four days before symptoms onset [44]	Four days post symptoms onset [44]	At symptoms onset [44]	80.0% – 90.0% [44]
Influenza	Normal mean = 2, sd = 0.5 [45,46]	One day before symptoms onset [47,48]	Six days post symptoms onset [47,48]	At symptoms onset [47,48]	11.0% – 18.0% [45]

*Sensitivity analysis elaborated in section on scenarios

5.6.3 Simulating transmission

We simulated the infection of a susceptible individual through a Poisson contact process. In each 5-minute interval contact episode, t, the conditional probability of infection given no prior infection, $p_{inf}(t)$, is defined as:

$$p_{inf}(t) = e^{-\Lambda(t-1)}\left(1 - e^{-\lambda(t)}\right) \tag{5.1}$$
$$\approx \lambda(t)$$

$$\lambda(t) = \beta\, v(t)\, c(t)\, h(t) \tag{5.2}$$

$$\Lambda(t-1) = \sum_{0}^{t-1} \lambda(t) \tag{5.3}$$

where $\lambda(t)$ is the force of infection and is a function of the relative infectiousness, $v(t) \in [0,1]$, scaled by a factor β to constrain the overall probability of infection to be similar to the observed attack rate; the presence or absence of contact between two individuals, $c(t) \in \{0,1\}$, and the current isolation status of the infector, $h(t) \in \{0,1\}$. $\Lambda(t)$ represent the cumulative force of infection up to time t. The first coefficient on the RHS of Equation 1 is the probability of surviving infection up to time step $t-1$ and the second coefficient is the probability of being infected at time step t. The stochastic model then samples the time of infection in each pair of individuals based on p_{inf}. For small values of $\lambda(t)$, Equation 1 approximates to $\lambda(t)$.

Each contact pair has a unique sequence of recorded signals (Supplementary Figure 1) and a corresponding cumulative probability of infection in all 5-minute interval contact episodes over the entire infectiousness period of the infector (i.e. probability of infection per contact pair). We defined the probability of infection to be the average probability of infection per contact pair. Once the simulated transmission occurred between a pair of individuals, there was no further propagation of the infection. We simulated 1,000 transmission pairs for each combination of pathogen and epidemiological characteristics.

Under a scenario of 'competing infectors', we simulated two susceptible individuals being exposed to an index case. Pairwise transmission was modelled and after the first transmission event had occurred, the remaining susceptible individual would subsequently be exposed to an additional infector, thereby acquiring infection from either infectors (Figure 5.1A). This is similar to disease transmission in households and we assumed that all susceptible household members were only exposed to infected cases within the household. Intuitively, infectors with a shorter incubation period are more likely to have earlier infectious contact with existing susceptible household members, thus potentially resulting in shorter generation intervals over the generations. However, infectors with longer incubation periods tend to have longer

pre-symptomatic infectious periods [9,49] and, for the same duration of shedding post symptoms onset, these infectors potentially exert a higher force of infection on the susceptible individuals over the entire duration of infectiousness which could influence the generation intervals over the generations. Thus, we investigated the differences in the generation and serial interval distributions, for transmission in pairs and triples.

5.6.4 Scenarios

For each disease in Table 5.1, we investigated how variations in the scale factor for peak infectiousness, β, would change the probability of infection and serial interval using the spline model. Furthermore, for the wild-type SARS-CoV-2 and the Delta variant, we studied how variations in the delay from symptom onset-to-isolation would vary the serial interval. All transmission events were simulated using household member contact patterns (unless otherwise stated) to achieve a similar probability of infection as the observed secondary attack rate in households. Model outputs were compared against published data on serial intervals and attack rates to ensure they were within the observed range.

5.6.4.1 Pairwise transmission

In the main analysis on pairwise transmission, the incubation period and infectiousness profile of the reference pathogen were based on SARS-CoV-2 Delta variant (Table 5.1). We compared how different scenarios of changing pathogen biology would influence the changes in the generation and serial intervals, and the corresponding power to detect these differences under varying human contact patterns and outbreak responses. Namely, we studied the effects of the following responses: no isolation, average delay of 4 days from symptom onset-to-isolation, average delay of 8 days from symptom onset-to-isolation (Table 5.2). As a sensitivity analysis, we assumed that the time of peak infectiousness occurred 1–5 days after symptoms onset [49,50] instead of at the time of symptom onset (Table 5.1).

Table 5.2 Simulated scenarios and how they relate to observations in the SARS-CoV-2 pandemic.

Scenario	Observations
Different incubation period between reference and alternative pathogen but same peak infectiousness and duration of shedding post-peak infectiousness for respective symptom onset-to-isolation.	SARS-CoV-2 Delta and Alpha variants were reported to have similar peak viral load and duration of shedding after the peak [2] with the former having a shorter incubation period [28].
Different incubation period between reference and alternative pathogen for respective symptom onset-to-isolation. Peak viral load in reference pathogen was varied (β of 0.0005, 0.002, 0.006) resulting in either a 20%, 50% or 80% probability of	SARS-CoV-2 Delta variant was reported to have a longer duration of viral shedding than the wild type but contrasting findings were reported for differences in peak viral load and incubation period [1,3,26]. Furthermore, some studies conducted

infection when the mean incubation of the reference pathogen is 4 days. Peak viral load in alternative pathogen was fixed (β of 0.0005). Duration of shedding post-peak infectiousness for the reference pathogen was 8 days longer.	during the exponential growth phases might not be explicitly adjusted for recent infections (i.e. increased observations of cases with short incubation periods). As such, contrasting findings of serial intervals shortening by 1–2 days [24,26] or no change in these intervals after accounting for earlier case isolation [7,8] were reported.
Different incubation period and shorter duration of infectiousness post-peak viral load in the alternative pathogen under respective symptom onset-to-isolation. Duration of shedding post-peak infectiousness for the reference pathogen was 8 days longer.	Similar peak viral load was reported in both vaccinated and unvaccinated SARS-CoV-2 Delta cases, but the former has a shorter duration of shedding post peak viral load [1,2,36]. Small sample size and infrequent data points for viral growth trajectories as compared to viral decline affected the power to detect differences in viral growth rates [1,2]. This difference, if any, would suggest a different duration of shedding prior to peak viral load thereby affecting the ability to detect cases early and the extent of pre-symptomatic transmissions.
Different contact patterns of either non-household or household contact with the same peak infectiousness and same incubation period under respective symptom onset-to-isolation over a range of peak infectiousness. Due to the differences in contact frequency, probabilities of infections (p_1 and p_2) were different for both types of contact.	Large scale movement restrictions such as lockdowns and work-from-home arrangements would potentially increase the proportion of contacts occurring with household members among all contacts and a corresponding decrease for non-household contacts [51,52]. Thus, for the same pathogen characteristics, the frequent contact in households would alter the probability of infection in each timestep as compared to non-household contacts, thereby altering the generation and serial intervals.
Different contact patterns of either daily or weekly household contact but with the same peak infectiousness and same incubation period under respective symptom onset-to-isolation over a range of peak infectiousness.	

5.6.4.2 Epidemic dynamics

We first studied the difference in the generation interval distribution between a reference and alternative pathogen without adjusting for the bias introduced by varying epidemic dynamics. The reference pathogen had a one-day shorter incubation period and longer duration of viral shedding (i.e. Delta-like) as compared to the alternative

(i.e. wild-type-like). During exponential growth, the incubation periods of the recently infected infectors (also known as the backward incubation period in [5,11]) tend to be shorter than the true incubation period (also known as the forward incubation period in [5,11]). Without adjusting for the outbreak dynamics, the overall observed incubation period will be shortened and, hence, shortening the generation intervals. The relationship between the observed (backward) and the true (forward) incubation period can be expressed as [5]:

$$b(\tau) = \frac{exp(-r\tau)f(\tau)}{\int_0^\infty exp(-rx)f(x)dx} \tag{5.4}$$

where $b(\tau)$ is the backward incubation period and $f(\tau)$ is the forward incubation period τ time since infection, and r is the exponential growth rate if $r > 0$ and exponential decline if $r < 0$. We parameterised our model by $f(\tau)$ to generate the $b(\tau)$ of the infectors and simulate the different outbreak dynamics. We varied the exponential rate r in the reference pathogen from -0.3 to 0.3 in increments of 0.1. We also varied the r in the alternative pathogen to be 0.2, 0 and -0.2 which corresponds to an outbreak with doubling time of 3.5 days, sustained constant outbreak, and half-life of 3.5 days.

In a real-world outbreak, with a given backward incubation period and a known exponential rate, we can adjusted for the bias brought about by the exponential outbreak dynamics and derive the forward incubation period as follow [5]:

$$f(\tau) = \frac{exp(r\tau)b(\tau)}{\int_0^\infty exp(rx)b(x)dx} \tag{5.5}$$

To illustrate the changes in power to detect the mean difference in the generation interval, we simulated a scenario of varying incubation period for a reference and alternative pathogen under (i) exponential growth in a reference pathogen but exponential decline in the alternative pathogen, (ii) constant growth in both reference and alternative pathogen, (iii) constant growth in the reference pathogen and exponential growth in the alternative pathogen.

5.6.4.3 Cluster transmission

For transmission occurring under competing infectors, we simulated 1000 transmission clusters and investigated the changes in the generation and serial intervals for transmission in a cluster of triples as compared to the previous pairwise transmission. We varied the incubation period with an average of 2, 4, or 6 days, under a β value of 0.0005 (scale factor to achieve similar peak infectiousness as SARS-CoV-2 Delta variant) for different delays from symptom onset-to-isolation.

5.6.5 Statistical test

We used a two sample Welch's t-test to estimate the power to detect a difference in means of the interval distributions of the reference and alternative pathogen. The test was two-sided with a significance level of 5%. An equal number of intervals were sampled from each distribution (25 or 100 samples out of 1,000 simulated pairs) and distributions are of unequal variances.

In reality, generation intervals are rarely observed and would need to be inferred using the observed serial intervals [15]. We studied two inference approaches: that the infectiousness of an infector is (i) dependent on the time of symptoms onset of the infector or (ii) dependent on the time of infection of the infector [15]. In each approaches, we would infer the mean and variance of the generation interval (i.e. inferred generation interval) based on the assumptions elaborated in [15] and Supplementary Information. Using this mean and variance, we will perform the Welch's t-test to obtain the estimated power to detect a difference in the means of the inferred generation interval distribution of the reference and the alternative pathogen.

In both approaches, we assumed that the incubation period distribution of the infector and the infectee are independent and identically distribution. As such, the inferred generation interval and observed serial interval have the same mean. The relationship between the variance of the inferred generation interval can be expressed broadly as follows [15] with further elaboration in the supplementary info:

$$Var(G) = Var(S) + 2Cov(P_{ij}, I_i) \tag{5.6}$$

where G is the generation interval, S is the serial interval, P_{ij} is the onset-to-transmission between infector i and infectee j, I_i is the incubation period of infector i. In the first inference approach, we assumed the incubation period of the infector and the time from onset-to-transmission are independent (i.e. $Cov(P_{ij}, I_i) = 0$). Hence, the variance of the inferred generation interval is reduced to:

$$Var(G) = Var(S) \tag{5.7}$$

In this approach, the inferred generation interval has the same mean and variance as the observed serial interval, and is the same as the serial interval.

In the second assumption, the infectiousness profile is independent of the timing of symptoms (i.e. timing of transmission is not correlated with the timing of symptoms, $Cov(G, I_i) = 0$) and the variance of the inferred generation interval can be expressed as:

$$
\begin{aligned}
Var(G) &= Var(S) + 2Cov\big(P_{ij}, I_i\big) \\
&= Var(S) + 2Cov(G - I_i, I_i) \\
&= Var(S) + 2[Cov(G, I_i) - Cov(I_i, I_i)] \\
&= Var(S) - 2Var(I) \\
&\leq Var(S)
\end{aligned}
\tag{5.8}
$$

Thus, the generation interval can be more broadly defined as the sum of the time from infection to infectiousness in the infector and the time from infectiousness to infection. The variance of the inferred generation interval is thus smaller than the observed serial interval.

Both are extreme approaches and formed the basis of inferring the upper and lower limits of the variance of the inferred generation intervals (i.e. lower and upper limits of the power to detect a difference in the inferred generation intervals) [15].

5.7 References

1. Singanayagam A, Hakki S, Dunning J, Madon KJ, Crone MA, Koycheva A, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. The Lancet Infectious Diseases. 2022;22: 183–195. doi:10.1016/S1473-3099(21)00648-4

2. Kissler SM, Fauver JR, Mack C, Tai CG, Breban MI, Watkins AE, et al. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated individuals. 2021 Aug p. 2021.02.16.21251535. doi:10.1101/2021.02.16.21251535

3. Li B, Deng A, Li K, Hu Y, Li Z, Xiong Q, et al. Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant. Virological. 2021; 2021.07.07.21260122. doi:10.1101/2021.07.07.21260122

4. 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;40. doi:10.1016/j.eclinm.2021.101129

5. 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;120: e2221887120. doi:10.1073/pnas.2221887120

6. 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. 2021 Oct p. 2021.10.21.21265216. doi:10.1101/2021.10.21.21265216

7. Pung R, Mak TM, Kucharski AJ, Lee VJ. Serial intervals in SARS-CoV-2 B.1.617.2 variant cases. The Lancet. 2021;398: 837–838. doi:10.1016/S0140-6736(21)01697-4

8. Ryu S, Kim D, Lim J-S, Ali ST, Cowling BJ. Serial Interval and Transmission Dynamics during SARS-CoV-2 Delta Variant Predominance, South Korea - Volume 28, Number 2—February 2022 - Emerging Infectious Diseases journal - CDC. [cited 10 Jul 2022]. doi:10.3201/eid2802.211774

9. Ferretti L, Ledda A, Wymant C, Zhao L, Ledda V, Abeler-Dörner L, et al. The timing of COVID-19 transmission. 2020 Sep p. 2020.09.04.20188516. doi:10.1101/2020.09.04.20188516

10. Sender R, Bar-On YM, Park SW, Noor E, Dushoff J, Milo R. The unmitigated profile of COVID-19 infectiousness. 2021 Nov p. 2021.11.17.21266051. doi:10.1101/2021.11.17.21266051

11. Park SW, Sun K, Champredon D, Li M, Bolker BM, Earn DJD, et al. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers.

Proceedings of the National Academy of Sciences. 2021;118: e2011548118. doi:10.1073/pnas.2011548118

12. Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109. doi:10.1126/science.abc9004

13. Kissler SM, Klepac P, Tang M, Conlan AJK, Gog JR. Sparking "The BBC Four Pandemic": Leveraging citizen science and mobile phones to model the spread of disease. bioRxiv; 2020. p. 479154. doi:10.1101/479154

14. 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;20: 911–919. doi:10.1016/S1473-3099(20)30287-5

15. Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness profile and generation time. Journal of The Royal Society Interface. 2021;18: 20200756. doi:10.1098/rsif.2020.0756

16. Svensson Å. A note on generation times in epidemic models. Mathematical Biosciences. 2007;208: 300–311. doi:10.1016/j.mbs.2006.10.010

17. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci. 2007;274: 599–604. doi:10.1098/rspb.2006.3754

18. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368: eabb6936. doi:10.1126/science.abb6936

19. Hart WS, Abbott S, Endo A, Hellewell J, Miller E, Andrews N, et al. Inference of the SARS-CoV-2 generation time using UK household data. Flegg J, Franco E, Kao RR, Conway E, editors. eLife. 2022;11: e70767. doi:10.7554/eLife.70767

20. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LHK, Fearon E, et al. Challenges in control of COVID-19: short doubling time and long delay to effect of interventions. Philosophical Transactions of the Royal Society B: Biological Sciences. 2021;376: 20200264. doi:10.1098/rstb.2020.0264

21. Chia ML, Him Chau DH, Lim KS, Yang Liu CW, Tan HK, Tan YR. Managing COVID-19 in a Novel, Rapidly Deployable Community Isolation Quarantine Facility. Ann Intern Med. 2020; M20-4746. doi:10.7326/M20-4746

22. Department of Health and Social Care. Testing for coronavirus (COVID-19) will increase to 25,000 a day. In: GOV.UK [Internet]. 18 Mar 2020 [cited 3 Dec 2022]. Available: https://www.gov.uk/government/news/testing-for-coronavirus-covid-19-will-increase-to-25-000-a-day

23. Kim J-H, An JA-R, Min P, Bitton A, Gawande AA. How South Korea Responded to the Covid-19 Outbreak in Daegu. NEJM Catal Innov Care Deliv. 2020;1: 10.1056/CAT.20.0159. doi:10.1056/CAT.20.0159

24. Zhang M, Xiao J, Deng A, Zhang Y, Zhuang Y, Hu T, et al. Transmission Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2 — Guangdong Province, China, May–June 2021. CCDCW. 2021;3: 584–586. doi:10.46234/ccdcw2021.148

25. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open. 2020;10: e039652. doi:10.1136/bmjopen-2020-039652

26. Kang M, Xin H, Yuan J, Ali ST, Liang Z, Zhang J, et al. Transmission dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant infections in Guangdong, China, May to June 2021. Eurosurveillance. 2022;27: 2100815. doi:10.2807/1560-7917.ES.2022.27.10.2100815

27. Madewell ZJ, Yang Y, Longini IM Jr, Halloran ME, Dean NE. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis. JAMA Network Open. 2022;5: e229317. doi:10.1001/jamanetworkopen.2022.9317

28. Grant R, Charmet T, Schaeffer L, Galmiche S, Madec Y, Platen CV, et al. Impact of SARS-CoV-2 Delta variant on incubation, transmission settings and vaccine effectiveness: Results from a nationwide case-control study in France. The Lancet Regional Health – Europe. 2022;13. doi:10.1016/j.lanepe.2021.100278

29. UK Health Security Agency. COVID-19: epidemiology, virology and clinical features. In: GOV.UK [Internet]. 17 May 2022 [cited 11 Aug 2022]. Available: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features

30. Jing Q-L, Liu M-J, Zhang Z-B, Fang L-Q, Yuan J, Zhang A-R, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020;20: 1141–1150. doi:10.1016/S1473-3099(20)30471-0

31. Zhang X, Wang J. Dose-response Relation Deduced for Coronaviruses From Coronavirus Disease 2019, Severe Acute Respiratory Syndrome, and Middle East Respiratory Syndrome: Meta-analysis Results and its Application for Infection Risk Assessment of Aerosol Transmission. Clinical Infectious Diseases. 2021;73: e241–e245. doi:10.1093/cid/ciaa1675

32. YouGov. Over nine in ten not working the usual 9-5 week. 24 Aug 2014 [cited 17 Aug 2022]. Available: https://yougov.co.uk/topics/economy/articles-reports/2018/08/24/over-nine-ten-not-working-usual-9-5-week

33. Chandler D. Haslemere Vision. 2013. Available: https://haslemerevision.org.uk/wp-content/uploads/Evidence-Base-2013-20131030.pdf

34. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26: 672–675. doi:10.1038/s41591-020-0869-5

35. Park SW, Sun K, Abbott S, Sender R, Bar-on Y, Weitz JS, et al. Inferring the differences in incubation-period and generation-interval distributions of the Delta and Omicron variants of SARS-CoV-2. medRxiv; 2022. p. 2022.07.02.22277186. doi:10.1101/2022.07.02.22277186

36. Chia PY, Ong SWX, Chiew CJ, Ang LW, Chavatte J-M, Mak T-M, et al. Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine-breakthrough infections: a multi-center cohort study. 2021 Jul p. 2021.07.28.21261295. doi:10.1101/2021.07.28.21261295

37. Ng OT, Koh V, Chiew CJ, Marimuthu K, Thevasagayam NM, Mak TM, et al. Impact of Delta Variant and Vaccination on SARS-CoV-2 Secondary Attack Rate Among Household Close Contacts. The Lancet Regional Health – Western Pacific. 2021;17. doi:10.1016/j.lanwpc.2021.100299

38. Centres for Disease Control and Prevention. Clinical Disease | Smallpox | CDC. 15 Feb 2019 [cited 10 Aug 2022]. Available: https://www.cdc.gov/smallpox/clinicians/clinical-disease.html

39. Nishiura H. Determination of the appropriate quarantine period following smallpox exposure: An objective approach using the incubation period distribution. International Journal of Hygiene and Environmental Health. 2009;212: 97–104. doi:10.1016/j.ijheh.2007.10.003

40. Centres for Disease Control and Prevention. Transmission | Smallpox | CDC. 15 Feb 2019 [cited 10 Aug 2022]. Available: https://www.cdc.gov/smallpox/transmission/index.html

41. Centres for Disease Control and Prevention. Signs and Symptoms | Smallpox | CDC. 15 Feb 2019 [cited 10 Aug 2022]. Available: https://www.cdc.gov/smallpox/symptoms/index.html

42. De Quadros CC, Morris L, Da Costa EA, Arnt N, Tigre CH. Epidemiology of variola minor in Brazil based on a study of 33 outbreaks. Bull World Health Organ. 1972;46: 165–171.

43. Prevention I of M (US) B on HP and D. SCIENTIFIC BACKGROUND ON SMALLPOX AND SMALLPOX VACCINATION. Scientific and Policy Considerations in Developing Smallpox Vaccination Options: A Workshop Report. National Academies Press (US); 2002. Available: https://www.ncbi.nlm.nih.gov/books/NBK221063/

44. Centres for Disease Control and Prevention. Pinkbook: Measles | CDC. 17 Aug 2021 [cited 10 Aug 2022]. Available: https://www.cdc.gov/vaccines/pubs/pinkbook/meas.html

45. Lessler J, Salje H, Van Kerkhove MD, Ferguson NM, Cauchemez S, Rodriquez-Barraquer I, et al. Estimating the Severity and Subclinical Burden of

Middle East Respiratory Syndrome Coronavirus Infection in the Kingdom of Saudi Arabia. American Journal of Epidemiology. 2016;183: 657–663. doi:10.1093/aje/kwv452

46. Jilani TN, Jamil RT, Siddiqui AH. H1N1 Influenza. StatPearls. Treasure Island (FL): StatPearls Publishing; 2022. Available: http://www.ncbi.nlm.nih.gov/books/NBK513241/

47. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. New England Journal of Medicine. 2009;360: 2605–2615. doi:10.1056/NEJMoa0903810

48. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, et al. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. Am J Epidemiol. 2008;167: 775–785. doi:10.1093/aje/kwm375

49. Ke R, Zitzmann C, Ho DD, Ribeiro RM, Perelson AS. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. Proceedings of the National Academy of Sciences. 2021;118: e2111477118. doi:10.1073/pnas.2111477118

50. Hakki S, Zhou J, Jonnerby J, Singanayagam A, Barnett JL, Madon KJ, et al. Onset and window of SARS-CoV-2 infectiousness and temporal correlation with symptom onset: a prospective, longitudinal, community cohort study. The Lancet Respiratory Medicine. 2022;10: 1061–1073. doi:10.1016/S2213-2600(22)00226-0

51. Gimma A, Munday JD, Wong KLM, Coletti P, Zandvoort K van, 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 Medicine. 2022;19: e1003907. doi:10.1371/journal.pmed.1003907

52. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science. 2020;368: 1481–1486. doi:10.1126/science.abb8001

5.8 Outro

In this chapter, I simulated infectious disease transmission pairs to investigate the power to detect changes in the generation and serial intervals under varying outbreak conditions. Furthermore, I showed how epidemic outbreak dynamics could influence our interpretation of the generation and serial intervals and clarified contradictory findings on serial intervals reported in some studies. In the next chapter, I explored the use of an outbreak simulation model to understand how transmission would occur in a cruise setting and the combination of interventions required for effective outbreak control.

6 Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events

With the introduction of rapid antigen testing and vaccination, the resumption of large-scale economic and social activities was increasingly feasible by the end of 2020 provided that the risk of a Delta variant outbreak was kept low. However, some large-scale economic activities such as conferences, meetings and social activities such as cruises occur over several days, and the contact patterns and, hence, risk of disease transmission in these settings are unclear. Using digital contact tracing devices, I collected contact and location data from passengers and crews in four pilot cruise sailings in Singapore and simulated disease transmission over the contact networks. Furthermore, I modelled different outbreak control measures such as mask-wearing, PCR or rapid antigen testing, and vaccination or combinations of these measures to estimate the overall effectiveness of outbreak control.

This paper was published in Nature Communications in April 2022 [1]. The supplementary information of the paper is in Appendix F.

References

1. 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;13: 1956. doi:10.1038/s41467-022-29522-y

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

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 <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006052	**Title**	Ms
First Name(s)	Rachael		
Surname/Family Name	Pung		
Thesis Title	COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore		
Primary Supervisor	Adam J. 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?	Nature Communications		
When was the work published?	12 April 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.

Improving health worldwide   **www.lshtm.ac.uk**

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, investigation, visualisation, and writing of the original draft. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you the original author(s) and the source are credited. https://www.nature.com/articles/s41467-022-29522-y

SECTION E

Student Signature	*[signature]*
Date	7 March 2024

Supervisor Signature	*[signature]*
Date	13 March 2024

Check for updates

Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events

Rachael Pung [1,2,3✉], Josh A. Firth [4,5], Lewis G. Spurgin [6], Singapore CruiseSafe working group*, CMMID COVID-19 working group*, Vernon J. Lee[1,7] & Adam J. Kucharski[2,3]

The emergence of highly transmissible SARS-CoV-2 variants has created a need to reassess the risk posed by increasing social contacts as countries resume pre-pandemic activities, particularly in the context of resuming large-scale events over multiple days. To examine how social contacts formed in different activity settings influences interventions required to control Delta variant outbreaks, we collected high-resolution data on contacts among passengers and crew on cruise ships and combined the data with network transmission models. We found passengers had a median of 20 (IQR 10–36) unique close contacts per day, and over 60% of their contact episodes were made in dining or sports areas where mask wearing is typically limited. In simulated outbreaks, we found that vaccination coverage and rapid antigen tests had a larger effect than mask mandates alone, indicating the importance of combined interventions against Delta to reduce event risk in the vaccine era.

[1] Ministry of Health, Singapore, Singapore. [2] Centre for the Mathematical Modelling of infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK. [3] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [4] Department of Zoology, University of Oxford, Oxford, UK. [5] Merton College, University of Oxford, Oxford, UK. [6] School of Biological Sciences, University of East Anglia, Norwich, UK. [7] Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. *Lists of authors and their affiliations appear at the end of the paper. ✉email: rachael.pung@lshtm.ac.uk

108

Many countries are resuming domestic activities as vaccination coverage and population immunity against SARS-CoV-2 increases[1–3]. Settings with particularly high contact rates, such as meetings, conferences, exhibitions, and cruises, are also revenue-generating sectors with high pre-pandemic visitor throughput across the world[4,5]. However, the transmission dynamics on real world networks of large-scale events are yet to be fully explored in the COVID-19 era[6]. Furthermore, while pre-COVID-19 studies on human contact networks for understanding the transmission of infections spread by close contacts have analysed various network properties and attempted to reconstruct the social network from contact diaries or digital sensors, they are largely focused in school, healthcare settings or the greater community, with few studies on conferences and business meetings[7–10]. Understanding the risk of outbreaks in these settings and possible outbreak control interventions would enable event planners to gauge the sustainability of their operations and for policy makers to weigh the public health cost against the economic gains. Given breakthrough infections in vaccinated individuals and the spread of the highly transmissible SARS-CoV-2 variants[11–13], countries have employed a range of tools alongside routine vaccination to suppress disease transmission, including vaccine certifications, rapid antigen tests, mask mandates, and digital contact tracing devices[14,15]. Although there have been efforts to estimate infection risk during large events from routine testing data and contact tracing interviews[16], data from contact tracing devices can enable finer-scale assessment of interactions such as the distance and duration of contact depending on the strength and continuity of the Bluetooth signals captured in these devices. Furthermore, these devices overcome the challenges of recall bias and achieve more reliable estimates of the contacts in a network[17].

In Singapore, 'cruises to nowhere' (i.e. cruises that depart and return to the port of origin without other ports of call) began as a safe travelling option during the COVID-19 pandemic with a range of activities and hence setting-specific interactions onboard. We collected contact data from around 1000 crew and 1300 passengers per sailing between November 2020 and February 2021 and analysed the resulting social interaction networks. We then use these contact networks to simulate SARS-CoV-2 Delta variant outbreaks and assess how different combinations of interventions and network formulations influence transmission in a range of settings during a large-scale event.

Results

Characterising social interactions on cruise ships. 3,963,256 contact episodes with 1,846,312 unique contact pairs were recorded during 37-h data collection periods across four separate three-day sailings (see Methods). During the period studied, cruise lines were operating at 50% capacity with a passenger to crew ratio of approximately 1:1 and passengers from different travelling groups were strongly advised to maintain a physical distance of at least one metre from other groups.

The four sailings had a mean of 1304 passengers (range 1142–1682) with a median age of 54 (IQR 35–63) and a mean of 1050 crew (range 1003–1083) spread across eight work departments (Table 1). There was a high density of contacts among passengers, with some clustering of contacts among the crew, although crew members may be required to work with other individuals from the same or different departments, and roles such as housekeeping and galley crew had contacts dispersed across the network (Fig. 1a and Supplementary Fig. 1). The crew was encouraged to form 'work bubbles' as part of COVID-19 workplace interventions (i.e. team of workers that work independently from another team). As a result, on average they

Table 1 Demographics of passengers and department allocation of crew onboard four cruise sailings.

No. of passengers = 1304 (1142–1682)	
Demographics	
Median age across all sailings in years (IQR)	54 (35–63)
Passengers by age group	
<12 years	47 (36–61)
12–29 years	166 (123–285)
30–39 years	184 (99–327)
40–49 years	164 (144–199)
50–59 years	285 (268–317)
60–69 years	314 (274–336)
≥70 years	146 (95–176)
Gender	
Female	676 (602–832)
Male	625 (540–850)

No. of crews = 1050 (1003–1083)	
Department[a]	
Entertainment	77 (73–81)
Food & Beverage (F&B)	179 (171–185)
Galley	214 (208–219)
Gaming	175 (163–187)
Hotel services	84 (77–92)
Housekeeping	123 (114–137)
Marine	154 (148–160)
Security	44 (40–48)

Number of passenger and crew presented as mean with range in brackets, unless specified otherwise.
[a]Details of each department are provided in Supplementary Table 2.

had 10 unique close contacts per day (IQR 6–18), about 50% lower than that of passengers (median 20, IQR 10–36) (Fig. 1b). If the threshold for close contact (defined as a cumulative duration of the interaction of 15 min in our baseline analysis) was relaxed to a shorter duration, the overall median unique close contacts scaled exponentially (Fig. 1c). The strength of each contact (i.e. edge weights) can be further quantified as a function of the duration of the contact (see Methods). Adjusted for the duration of each contact, the median weighted degree in crew was 8.3 (IQR 4.4–13.5), while the median in passengers was 13.9 (IQR 5.6–23.7) (Table 2). Furthermore, passengers had significantly higher connectivity with other highly connected individuals, with a median eigenvector centrality of 0.3 (IQR 0.1–0.5) compared to a median of 0.09 (IQR 0.03–0.2) for the crew (Table 2 and Supplementary Fig. 2).

Analysing the contacts formed during activities. The total number of contacts made by passengers with passengers from other travelling groups varied according to the type of location and the time spent at that location. The total close contacts plateaued at approximately 3 (IQR 2–5) after spending at least 1 h in a food and beverage (F&B) location (Fig. 2a) while the total close contacts were 2 (IQR 1–3) after spending 30 min to 1 h in a sports location and increased to 4 (IQR 2–7) after spending at least 2 h (Fig. 2c).

Over the three-day sailings, a median of 71% (IQR 64–74%) of all the close contact episodes occurring between passengers from different travelling groups occurred in F&B locations of which 23% (IQR 19–26%) and 38% (IQR 31–40%) occurred in the buffet and inclusive restaurants respectively (Fig. 3a, b). 16% (IQR 11–24%) of the close contacts occurred in entertainment areas and 8% (IQR 6–10%) in sports areas (Fig. 3a). Passengers are largely mask-off when dining or engaged in sports and this
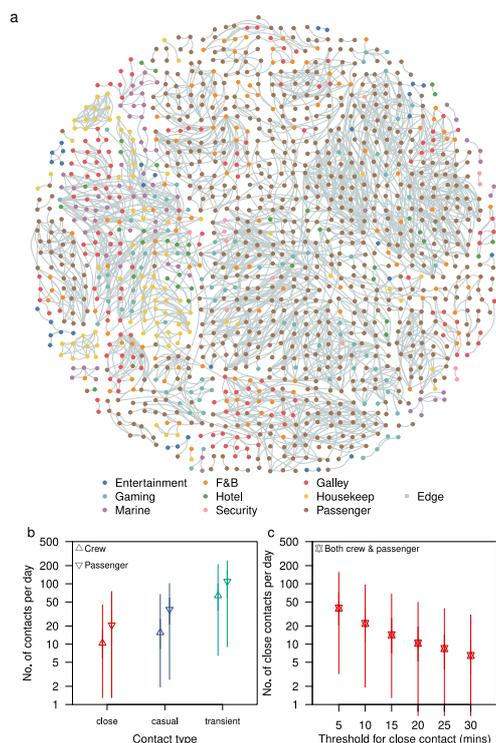
ARTICLE

Table 2 Network properties of passengers and crew onboard four cruise sailings.

Network properties	Passenger	Crew	P-value
Weighted degree	13.9 (5.6–23.7)	8.3 (4.4–13.5)	<2.2 × 10^{-16}
Eigenvector centrality	0.3 (0.1–0.5)	0.08 (0.03–0.2)	<2.2 × 10^{-16}
Clustering coefficient	0.4 (0.3–0.4)	0.3 (0.2–0.4)	<2.2 × 10^{-16}

Two sided Welch's *t*-test was performed and results were presented as median with IQR in brackets.

Fig. 1 Distribution of cruise ship contacts. a Illustrative short-term network dynamics, showing the cumulative network of all contacts that began between 12.00 to 12.05 pm on the second day of a sail and lasted till the end of their contact episode. Edge width and colour intensity are a function of the type of contact (i.e. close, casual and transient). Intra- and inter-cohort contacts are represented by the connection of nodes with the same and different colour respectively. **b** Number of unique close, casual, transient contacts made per day by passenger and crew. **c** Number of close contacts per day for both crew and passengers using different thresholds for the cumulative duration of interaction. Median (shapes), 50% (dark lines), and 95% intervals (light lines) of contacts from 5216 passengers and 4197 crew across four sailings are shown in (**b**) and (**c**).

accounted for 79% (IQR 69–84%) of all close contact episodes, 69% (IQR 57–76%) causal, and 60% (IQR 51–66%) transient contact episodes (Fig. 3a).

Modelling outbreak dynamics and interventions. To examine the spread of SARS-CoV-2 on cruise ships and implications for other large-scale multi-day events, we used the contact data to generate an undirected network with nodes and edges representing individuals and the contact between them respectively. We defined the strength of an edge as a function of the proportion of days with recorded contact over a three-day sail period and the mean daily cumulative contact duration between two individuals to approximate a scenario where the edge weight reached 95% saturation after 3 h of contact (see Methods). This meant that the propensity for transmission increased and stabilised after 3 h of contact, to mimic contacts formed in family gatherings over extended periods of time[18,19].

We extended a community network transmission model[20,21] to simulate SAR-CoV-2 Delta variant transmission over seven days (Table 3), to enable comparison between different interventions during early generations of transmission. We considered interventions including: (i) one-off PCR testing one day before the sailing (to allow for test turnaround time), (ii) rapid antigen testing at the start and halfway through the event, (iii) mask wearing in feasible settings and (iv) vaccination coverage among attendees. In both (i) and (ii) testing interventions, we assumed infected individuals were isolated immediately after a positive test in the main analysis. The sensitivity of the PCR and rapid antigen tests were assumed to vary with viral load modelled according to the Delta variant[12,22–24]. For the mask wearing intervention, we assumed that passengers of different travelling groups would be exposed to each other without a mask during dining, sports activities (e.g. pool and waterslides, rock climbing, basketball, football) or smoking breaks; and would be wearing a mask correctly otherwise. Furthermore, contacts between passengers and crew were assumed to occur with mask-on at all times and crew-crew contacts were assumed to occur without a mask during meals times, workouts or smoking breaks. The proportion of contacts that occurred without a mask were modelled based on the proportion of contacts occurring in F&B and sport settings (Fig. 3a).

Under the baseline scenario with no modelled interventions, with one initial infected individual and assuming that the event lasted for 7 days, we estimated a median of 10 individuals (IQR 3–23) would be infected by the end of the event (Fig. 4a and Supplementary Fig. 3a). Of these, 90% (IQR 84–100) would only develop symptoms after the event. Because presymptomatic transmission was assumed to account for 25% the transmission, more than two generations of infections could sometimes occur during the event (Supplementary Fig. 3b). We estimated that 64% and 17% of the simulated outbreaks involved spillover from passengers to crew and inter-department crew transmission respectively, and we estimated that spillover events first occurred in the 2nd (IQR 2–3) and 3rd (IQR 3–4) generation respectively. Outbreaks with a final size of more than 10 cases occurred in 48% of our simulations (Fig. 4b).

With the introduction of a one-off PCR test one day prior to the start of the cruise, the index case was isolated in 49% of the time, while 5% of the remaining simulations resulted in no transmission due to the stochastic nature of early disease transmission and the structure of the social network (Fig. 4b). As a result, more than half of the simulations had zero secondary cases. The risk of an outbreak of more than 10 cases was reduced to 22% with the PCR intervention. However, with rapid antigen testing at the start and at halfway through the event instead, only 3% of simulations resulted in a large outbreak.

We also modelled passenger-passenger interactions occurring under a mask-off setting ~60% of the time (based on the total transient, casual and close contacts in Fig. 2a) and assumed that
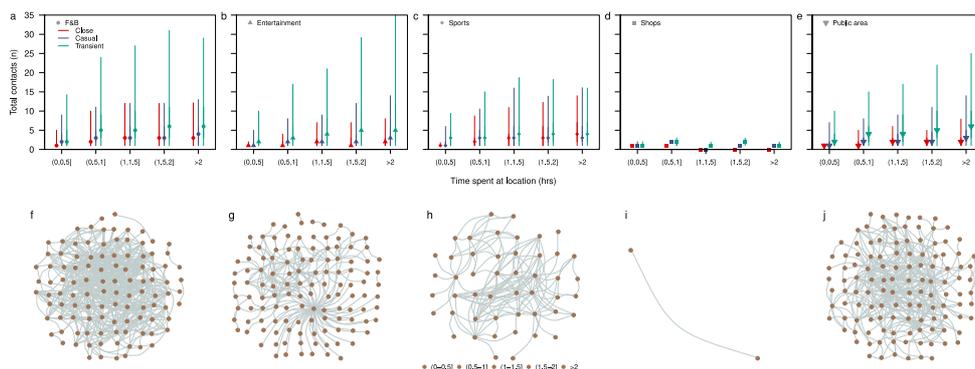
Fig. 2 Number of contacts made over time in respective locations. Contacts made between passengers from different travelling groups per visit to a type of location (**a**–**e**) and a snapshot of contact network at respective locations for 2 h intervals on the second day of the sailing (**f**–**j**). Type of locations are: F&B (**a**, **f**), entertainment (**b**, **g**), sports (**c**, **h**), shops (**d**, **i**) and public areas (**e**, **j**). Median (shapes), 50% (dark lines) and 95% intervals (light lines) of contacts from 5216 passenger and 4197 crew across four sailings are shown in (**a**–**e**). Nodes of different colour intensity represent the time spent in the location by respective passengers in (**f**–**j**).
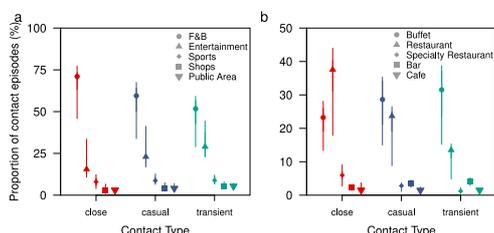
Fig. 3 Type of contact by location of interaction and throughout the sailing. a Proportion of all close, casual and transient contact episodes between passengers of different travelling groups by the location of interaction, **b** proportion of all close, casual, and transient contact episodes between passengers of different travelling groups by respective F&B locations. Median (shapes), 50% (dark lines) and 95% intervals (light lines) are shown.

all passenger-crew interactions occurred while wearing masks and that 30% of crew-crew interactions occurred when eating or working out without mask. Under these conditions and in the absence of other interventions, 22% of the simulations end with a large outbreak size (Fig. 4b). Assuming all individuals onboard the cruise ship were vaccinated (individuals under 12 years of age account for only 2% of the cruise population), 95% of simulations resulted in five or fewer cases (Fig. 4b).

We examined the expected outbreak size under a combination of interventions under the assumption that vaccination confers 50% protection against infection[13,25,26] and 50% lowered infectiousness in a vaccinated but infected individual[27] (Fig. 4c, d). Regardless of the testing strategies applied (i.e. no test, once-off PCR test, rapid antigen testing at the start and halfway through the event), and at any level of vaccine coverage, the addition of a mask-on intervention would further reduce the expected outbreak size by about 54% (IQR 50–59%). Given outbreak size is the cumulative result of individual transmission events, this implies that the overall intervention effectiveness of a mask mandate is substantially less than the assumed mask-on efficacy at the individual level (Table 4)[28–30]. The expected outbreak size in simulations involving rapid antigen testing was

<1 when vaccine coverage was minimally 25% (i.e. the expected number of transmission events was less than the initial number of infected individuals) (Fig. 4c). The expected outbreak size in mask-on, no testing interventions differed from mask-off, once-off PCR testing intervention by <1 case across varying vaccination coverage. The same was observed between a mask-on, once off PCR testing intervention and a mask-off, rapid antigen testing intervention. Compared to the expected outbreak size, the 95th percentile of the outbreak size is approximately three times higher, with the no testing, mask-off and one-off PCR testing, mask-off interventions generating the highest number of cases among all other combinations of interventions (Fig. 4d).

Sensitivity analysis under different assumptions of the edge weights—and hence per-contact risk—showed an increase in the expected outbreak size as the duration required to be defined as a 'maximal contact' (i.e. weight of one) decreases (Supplementary Fig. 4). Across all scenarios of varying testing strategy, vaccination coverage, network assumptions for edge weight, the average reduction in the expected outbreak size between a mask-on and mask-off scenario was 60% (IQR 54–71%). Assuming edge weights vary based on the proportion of days over the entire sailing when interactions were recorded (i.e. a transient contact in a day is as risky as a close contact in a day), the difference in the expected outbreak size between a mask-on, no testing scenario and a mask-off, once-off PCR testing widens to 32 cases (IQR 11–64) (Supplementary Fig. 4c). The difference in the expected outbreak size between a mask-on, once-off PCR testing scenario and a mask-off, rapid antigen testing at the start and halfway through the event differed by 6 cases (IQR 5–19) (Supplementary Fig. 4c). We obtained similar conclusions on the relative effect of different combinations of interventions when we varied assumptions about the extent of vaccine effectiveness and presymptomatic transmission (see Supplementary Information).

In reality, transmission parameters and effectiveness of outbreak interventions exhibit various uncertainties that can act simultaneously (Supplementary Table 1), and contact networks are temporally dynamic as the presence/absence of edges in the network change over time. Accounting for the uncertainty in the transmission process, our results for the expected and 95th percentile of the outbreak size falls between those in simulations assuming 25–50% presymptomatic transmission (Supplementary Figs. 4 and 6). As compared to the main analysis, the risk

111

ARTICLE

Table 3 Parameter values and assumptions.

Parameter	Assumed values	Details and references
Incubation period (days), θ	Lognormal distribution with Mean = 4.4, sd = 1.9	[51]
Adherence to isolation when tested positive (%)	100	For scenarios involving testing only and we assume that there are available cabins for individuals to isolate given that cruises are operating at 50% capacity.
Delay from positive test to isolation (hrs)	No delay	For scenarios involving testing, individuals were isolated once tested positive.
Initial cases among passengers	1	
Scaling parameter, r_{scale}	0.24–0.26	Each network formulation uses one scaling parameter value to calibrate the probability of Delta infection among cabin contacts to be similar to that of household contacts of 20%[34,49,52]. The range of values used across the different network formulations are as shown.

reduction through a mask-on intervention has a wider uncertainty of 40–80% while the adherence to isolation after testing positive could be as low as 60%. As such, both interventions will perform lower but we observed narrower differences in outbreak size for a mask-on, no testing scenario and a mask-off, one-off PCR testing scenario (Supplementary Fig. 8). The lowered adherence to isolation coupled with the possibility of vaccinated infected individuals being as infectious as unvaccinated individuals resulted in a larger outbreak size observed in all testing interventions at low vaccine coverage. This was in spite of the potential for vaccinated susceptible contacts achieving a higher risk reduction against infection of 50–70% which counteracts the reduced effectiveness of the aforementioned interventions. When simulating outbreaks on a dynamic network, we accounted for the heterogeneity in the contact duration over the days and the sequence of contact episodes. As passengers engaged in more activities on the second day on the cruise sailing as compared to the first, the number of contacts and duration is correspondingly higher. A static network that averages out these heterogeneity in contact could thus allow for a higher potential of transmission in earlier stages of the cruise sailing, resulting in a higher 95th percentile as compared to a temporal network (Supplementary Figs. 9 and 10). Nevertheless, it is encouraging to note that the median outbreak size is similar for outbreaks simulated in both a static and temporal network (Supplementary Figs. 9 and 10) as simulations on longer time scales of 7 days were performed using the static network which served as a means of extending the network beyond 3 days.

For context, in real-life cruise operations during 2021, over 80% of the population received two doses of COVID-19 vaccination and a one-off pre-event rapid antign testing was required. No outbreaks occurred on these cruises even when the reported community incidence was 0.7 per 1000 at the height of the outbreak in end of October 2021—approximately 30% lower than that simulated in the model (i.e. one initial infected passenger corresponding to a community incidence of about 1 per 1000).

Discussion

We found that the structure and intensity of contacts over a multi-day cruise have major consequences for outbreak control in different settings, particularly if there are mask-free activities and leaky testing protocols mean infectious individuals are likely to go undetected. Cruises represent an aggregation of different activities including F&B, entertainment, sports, meeting, conference, entertainment and workplace settings. The presence of multi-group passengers and crew from different departments can therefore offers insights into the potential dynamics of different actors in other large-scale multi-day events (e.g. a conference

where there are participants, organising teams, external vendors, front-end and back-end F&B service staff, audiovisual support teams) and resulting implications for control of SARS-CoV-2.

Our social network analysis showed that passengers had a high number of contacts and their contacts typically exhibit high levels of contact with other individuals. As such, any disease transmission would likely be driven by passenger-level interactions rather than crew. In early 2020, this was evident in the sharp rise in the number of COVID-19 passengers with symptom onset before or during the early stages of quarantine onboard the Diamond Princess[31]. While the number of contacts made with other passengers are potentially lowered due to physical distancing and awareness of the pandemic in the studied Singapore setting, the number and type of activities onboard the cruise still means that each passenger forms around 20 unique close contacts per day. Compared to an average of 59 unique close contacts with more than 15 min of interaction in a UK community setting over a 14-day period[32], this was five times higher, further illustrating the intensity of contacts during such events. More than 70% of the close and casual contacts on the cruises occurred in F&B locations where passengers were largely mask-off and thus posing a higher risk of infection and transmission. We observed that the number of close contacts plateaued in F&B settings as the time spent in the location increases. As such, reducing this risk potentially requires more creative use of space to increase the distance between groups of passengers, improve indoor ventilation and encourage more outdoor dining.

With numerous work functions interfacing with passengers, and given the overlapping shifts and closely related job scope between crew (e.g. F&B and galley, hotel services and housekeeping), we found it only took around two generations for the infection to spread from a passenger to a crew and an additional generation of transmission to reach another crew in a different department. For SARS-CoV-2 transmission on the Diamond Princess cruise ship, the earliest onset in crew occurred about 18 days after the onset of the index case[33]. Assuming a generation time of about 5–7 days, this corresponds to a spillover from passengers to the crew after three to four generations of transmission. With about 2.6 times more passengers than crew on the Diamond Princess cruise ship, this could delay the spillover of disease transmission. Crew and event personnel play an important role in ensuring smooth operations and their wellbeing should be accounted for in the plans when reopening events. Hence, besides encouraging crew cohorting, interventions that minimise transmission in passengers would have an indirect effect of protecting the crew.

When applied individually, none of the interventions analysed were capable of reducing the expected outbreak size to be lower
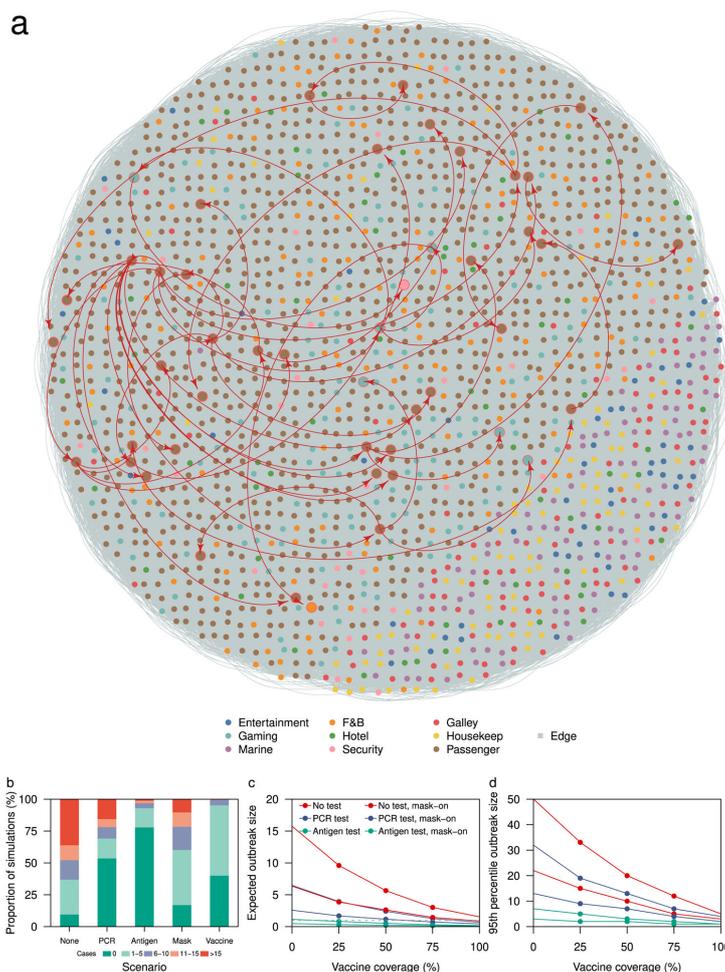
5

112

Fig. 4 Outbreak size under respective interventions. a Cases and contacts in one outbreak simulation with cases represented by an enlarged node and red curved arrows depicting disease transmission, **b** proportion of simulations by respective outbreak size under different interventions, **c** average outbreak size and **d** 95th percentile of outbreak size for different interventions and varying vaccination coverage.

than one; the number of initial infected cases in passengers, equivalent to a community incidence of 1 per 1000 individuals. However, a combination of rapid antigen testing at the start and halfway through the sailing with at least 25% coverage of a vaccine that confers 50% protection against infection and 50% lowered infectiousness would result in the cruise event having fewer onward transmission than the number of initial infectives. This is conditioned on the cases exhibiting Delta variant-like high viral loads with prolonged shedding[12,23,24,34] which improves the sensitivity of rapid antigen tests. While PCR tests have a higher sensitivity than rapid antigen tests at low viral load levels, the tests need to be conducted on land prior to the event due to the turnaround time required and for validity of lab results. This implies that cases who develop symptoms several days after the sailing may not be identified prior to the event, due to

viral loads near the limits of detection, and large outbreaks could occur.

The expected outbreak size under different combinations of interventions was sensitive to the assumptions of the network edge weights. When edges are weighted by the proportion of days with recorded interaction over a three-day sail period, two individuals with transient contact in a day are assumed to have the same risk as two individuals with close contact in a day. This assumption is applicable when the dominant mode of transmission is largely independent of the duration of contact (e.g. environmental or airborne transmission). There were five times more transient interactions than close contacts and these contacts are now equally at risk of infection. Thus, mask wearing would largely help to lower the risk of transmission and acquiring infection, and outperforms PCR testing or even twice antigen

ARTICLE

Table 4 Parameter values for the relative risk of infection, β.

Notation	Vaccination and mask wearing status	Relative risk	Remarks
β^v	Mask-off (i.e. not wearing a mask) and both the infected and susceptible individuals are not vaccinated.	1	No change in probability of infection.
	Mask-off and infected individual i is vaccinated.	0.5	Mean probability of transmitting infection reduces by 50%[27].
	Mask-off and susceptible individual j is vaccinated.	0.5	Mean probability of acquiring infection reduces by 50%[13,25,26].
	Mask-off and both infected individual i and susceptible individual j are vaccinated.	0.25	Mean probability of infection reduces by 75%. Assumes the effect of vaccination on transmission and acquiring infection is independent.
β^{mv}	Mask-on (i.e. wearing a mask) only.	0.2	Mean probability of infection reduces by about 80% when both the infected individual and susceptible contact are wearing a mask[28].
	Mask-on and infected individual i is vaccinated.	0.1	Mean probability of infection reduces by about 90%. Assumes the effect of vaccination and mask wearing on reducing the probability of transmitting infection is independent.
	Mask-on and susceptible individual j is vaccinated.	0.1	Mean probability of infection reduces by 90%. Assumes the effect of vaccination and mask wearing on reducing the probability of acquiring infection is independent.
	Mask-on and both infected individual i and susceptible individual j are vaccinated.	0.05	Mean probability of infection reduces by 95%. Assumes the effect of vaccination and mask wearing on reducing the probability of transmission and acquiring infection is independent.

testing interventions especially when the proportion of pre-symptomatic transmission is high during the early stage of viral shedding. Overall, these models estimate that mask wearing with passengers practising physical distancing could reduce transmission by about 54% under these settings, ~20–30% lower than the effectiveness of wearing a surgical mask in healthcare settings or in public areas after accounting for interactions in mask-off settings when dining or engaged in sports. This findings corroborates with behavioural surveys reporting 1.3–2 times higher risk of being infected when mask wearing in enclosed spaces is not practised[29]. The risk reduction from these model estimates are about five times higher than that reported in a cluster-randomised trial. However, in this trial, proper mask wearing occurred in less than half in the intervention arm, thereby limiting the multiplicative effectiveness of mask wearing in reducing both infection and transmission[30].

Both mask-off, rapid antigen testing and mask-on, once-off PCR testing would help to reduce the risk of disease introduction and further transmission if the index cases successfully escape initial detection. While their differences in the expected outbreak size were less than 10 cases across different assumptions to the edge weights, they bring different outcomes to the passenger experience and operations planning—an extra swab test at the middle of the event versus wearing a mask at all times other than during dining and engaged in sports, logistics to check the test outcomes versus monitoring mask wearing practises, managing false positives versus passengers flouting rules. Pre-event rapid antigen testing has been widely adopted in many large-scale events lasting less than a day and accounted for about 53% reduction in transmission in settings with high levels of social contacts and about 72% reduction after accounting for physical distancing[2]. In a fully susceptible cohort, these models estimate a mask-off, rapid antigen testing intervention at the start and midway of the event would reduce the mean outbreak size by over 90% with the additional reduction largely attributed to the administration of an additional test midway through the event.

One limitation to our study was that we did not model contact tracing around detected cases and the behaviour of contacts who are aware of their potential exposures. Thus, our estimates serve as an upper bound to the potential outbreak size. While cruise lines are trained to trace and quarantine close contacts as part of the pilot reopening, as the ease of rapid testing increases with fast turnaround time, this could serve as a replacement for slower and resource intensive contact tracing in such settings. With pre-symptomatic transmission of SARS-CoV-2 and high levels of transmissibility of the Delta variant, the effectiveness of contact tracing is approaching a point of saturation in many countries[35]. Furthermore, even if the threshold for close contact to be traced is lowered, the corresponding exponential increase in contacts fulfilling this criteria would make it logistically challenging to trace all individuals in a reasonable amount of time. Fully asymptomatic infections—as opposed to presymptomatic infections—were also not considered in the analysis. Should these infections exhibit lowered viral load, the testing interventions would be less likely to detect asymptomatic individuals but any potential for increase in outbreak size would be counteracted by their lowered infectiousness. Currently, there is no strong evidence to suggest that asymptomatic SARS-CoV-2 Delta infections are less infectious than symptomatic individuals[36,37]. Lastly, the accuracy of the data collected is largely dependent on the usage behaviour and the functionality of the device. Passengers are required to carry the contact tracing devices at all times except when engaged in water sports, and this was enforced by crew and external officers. Hence, interactions at the water sports areas may not be well represented but this effect to our analysis is expected to be minor as the cruise line of study required passengers to book these facilities in advance to facilitate crowd control. In a cabin, each passenger's device may not necessarily be placed in a 2 m proximity and the frequent close contact interactions in these settings would not be recorded accurately. However, given that individuals in the cabin would largely continue to interact with each other outside the cabin while carrying the device, this would help to record a large proportion of their close interaction. Furthermore, this limitation is reduced when the probability of infection saturates after a certain level of exposure as is in the case of SARS-CoV-2[38]. Functional issues of the contact tracing device such as drainage of batteries and incomplete data uploading can affect the extent of missing data, but these issues can be minimised with proper training on device usage. In our main analysis, the chosen sailing had more than 97% coverage in both crew and passengers to minimise the impact of missing data on the inference of the outbreak dynamics. Outbreaks were also simulated in three other sailings as part of sensitivity analysis and similar trends in the outbreak trajectory were observed (Supplementary Fig. 7). Despite such limitations, this is one of the few studies with large- and fine-scale data collection from multiple events in one setting and comparison with future

114

studies of similar data collection methods in similar and other settings may help to strengthen our findings and provide better understanding of transmission dynamics under different network structure and disease characteristics.

Given the spread of highly transmissible SARS-CoV-2 variants alongside increasing vaccination coverage, many countries have oscillated between reopening and restrictions of varying extents, in turn affecting the sustainability of economic and social activities. As the pressure to resume large-scale events increases, but the effectiveness of vaccines against infection and transmission remains variable, combining social interaction data with models such as the one presented here can enable an improved data-driven assessment of the risk of transmission arising from planned activities and the potential reduction offered by the continuation and implementation of non-pharmaceutical interventions.

Methods

Ethics statement. Information was provided and consent was obtained from all participants in the study before the digital contact tracing device recorded any data. The study was approved by the London School of Hygiene & Tropical Medicine Observational Research Ethics Committee (ref. 25727).

Data. Each cruise sailing lasts for three days—departing at 7 pm on the first day and arriving at 8am on the last day, and only contacts during this period were studied. Embarking and disembarking begins and ends at approximately 12 noon on both days and devices are stored together prior to issuance or after collection. As such, data prior to departure and after arrival were not used, as the recorded data may be an artefact of devices being stored together.

All individuals onboard a cruise are issued a digital contact tracing device with a unique device identification number. These devices are calibrated based on signal strength to broadcast omnidirectional Bluetooth signals to other devices within a 2 m radius every 14.9 s followed by an omnidirectional scan of nearby signals lasting for 0.1 s. Each scan record captures the timestamp of the signal exchange and the identification number of the interacted device. After every five-minute interval, the records of 30 unique devices with the highest signal strength in each device are then stored. The stored records are then uploaded to a server on land. Further data processing is required to determine the duration of contact between two individuals. If there are two or more records with consecutive difference(s) of less than five minutes, the duration of the contact is the difference between the last and first timestamp in the series of records.

For each cruise sailing, we collected a de-identified manifest with the device identification number and details of the device holder (passenger or crew; for crew: department of the crew (Supplementary Table 2); for passengers: cabin number, keycard number, age, gender). The cruise ship can be demarcated into different areas based on the activities in a location (i.e. type of location: food and beverage (F&B), entertainment, shops, sports, public areas) and all passengers were required to tap-in using their keycards upon entering a new area onboard the cruise ship. We also collected a de-identified list of entry records with each record capturing the keycard number, location and timestamp of entry.

Using the three data sources (i.e. contact data, de-identified manifest and de-identified location records), we categorised the contacts between each dyad into one of four contact groups, g, namely (i) passenger-passenger contact from within the same travelling group (i.e. passengers in the same cabin or having a cumulative contact duration of more than 5 h over 3 days), (ii) passenger-passenger contact from different travelling groups, (iii) crew-crew contact, and (iv) passenger-crew contact. Five hours was selected as a conservative definition for travelling groups, given that this is considerably longer than an average meal duration and more than 99% of the cumulative contact duration (i.e. sum of all contact episodes) between passengers from different cabins were less than this duration.

We further classified a contact episode in a location into close, casual and transient types of contact if the cumulative duration of contact was at least 15 min, at least 5 min but less than 15 min, and less than 5 min respectively in a 2 m radius[39–41]. For each individual in each type of location, we estimated the number of different types of contacts (i.e. close, casual and transient) with passengers from different travelling groups over time spent in the location. Across the sailings, for each type of contact, we estimated the proportion of contacts occurring at a type of location over all types of location.

Social network construction. We performed a preliminary social network analysis and estimated the weighted degree distribution (number of contacts made per individual with each contact weighted by the duration of contact, to be elaborated), the distribution of the clustering coefficient (a measure of the triadic linkage among individuals[42]) and individuals' eigenvector centrality (a measure of direct and indirect centrality within a network) of passengers and crew in respective departments in each sailing. We performed a Welch's t-test to evaluate each network property for passengers against that for crew and p-values < 0.05 were

considered statistically significant. While the mean and interquartile range (IQR) of each estimate fluctuate across sailings, the 95% range of the estimates exhibit substantial overlap (Supplementary Fig. 1). Due to these similarities, we selected contact data collected over a single focal sailing with 1208 passengers and 1032 crew to construct the social network for simulating disease transmission for the primary analysis. However, we also carried out Supplementary analysis whereby simulations were also performed on all other sailings, and used this to ensure consistency in the percentage reduction in outbreak size for various outbreak interventions across the different sailings (Supplementary Fig. 7).

In the main analysis, we generated an undirected network with the strength of an edge weighted as a value between 0 and 1 based on the proportion of days with recorded contact over a three-day sail period and the exponent transformation of the mean daily cumulative contact duration between two individuals as follows:

$$w_{ij} = c_{ij}(1 - e^{-\bar{d}_{ij}\sigma}) \tag{1}$$

where w_{ij} is the weight of a contact between individuals i and j, c_{ij} is the proportion of days with recorded contact and $\bar{d}_{ij}$ is the mean daily cumulative contact duration expressed hours. σ is a scalar of 0.5 to approximate a scenario where the edge weight reaches 95% saturation after 3 h of contact ($w_{ij} \to 1$). As a sensitivity analysis, we explored other weightings for the network edges; similar to the above but 95% saturation to the same level of infection risk after 1 h of contact, or based on the proportion of days over the entire sailing with recorded contact only. These scenarios depict how risk of infection increases based on contact duration as observed in SARS-CoV-2 outbreaks in settings of poor ventilation[43,44] or transmission driven by a highly transmissible pathogen onboard cruises such as norovirus[45].

Incorporating c_{ij} implicitly extends the contact networks as the contact data was collected over a 3-day sail but the transmission was simulated over a longer timescale of seven days to quantify the differences in outbreak trajectory for events lasting more than 3 days. Nevertheless, we have also performed sensitivity analysis using the actual temporal network to understand how the correlation of contact duration and sequence of contact events could potentially influence the outbreak.

Transmission model. We simulated SARS-CoV-2 Delta variant transmission on the above generated social contact network by extending the individual-based models developed by Firth et al. and Hellewell et al. (Table 3)[20,21].

For each simulation, we assume that the disease is introduced by one passenger who could be infected up to 14 days prior to the event, with equal probability on any of the days but the onset of the index case would only occur between the start (i.e. day 1) and the end (i.e. day 7) of the event. The distribution of the symptoms onset date, S, on respective day of the event, d, is as follows:

$$S(d) = \int_{-13}^{0} I(\delta)\theta(d + |\delta|)d\delta \tag{2}$$

where δ is the day of infection prior to the event (i.e. $\delta = 0$ represents the day before the start of event), $I(\delta)$ is the probability of being infected on any of the 14 days prior to the event and is fixed at $1/14$, θ is the incubation period distribution with $d + |\delta|$ representing the time since infection on the respective day of the event.

Currently, all crew are required to be tested weekly and are largely confined to the cruise except during periods of shore leave, thereby reducing the risk of disease introduction by crew. Each day, the model searches for susceptible individuals in contact with the infected cases who are not isolated and infection from infector i to susceptible individual j occurs based on the following probability:

$$P_{i \to j}(d) = 1 - e^{-\Delta d \lambda_{i \to j}(d)} \tag{3}$$

where Δd is the modelled time step of one day, and $\lambda_{i \to j}(d)$ is the force of infection between infector i and susceptible individual j on day d expressed as:

$$\lambda_{i \to j}(d) = w_{ij}f(d|\mu_i, \alpha_i, \omega_i)r_{scale}\beta, \text{ for } \beta \in \{\beta^v, \beta^{mv}\} \tag{4}$$

where $f(d|\mu_i, \alpha_i, \omega_i)$ is the probability density function that represents the infectiousness of the infector on day d. We assumed a skew normal distribution with location parameter μ_i set based on the infector's day of onset, a slant parameter α_i and a scale parameter ω_i adjusted such that 25% of the infections occurred prior to symptom onset. As there is substantial uncertainty in the proportion of presymptomatic transmission for SARS-CoV-2[46], for sensitivity analysis, we considered a scenario where about 50% of transmission occurred prior to symptom onset. With a skewed normally distributed infectiousness profile centred based on the day of the symptoms onset, this ensures that the majority of the infections occurred around the time of symptoms onset[47,48].

While an edge weight has a maximum value of 1, infection between two individuals over the entire duration of infectiousness of the infected individual is not guaranteed. As such, we multiplied the force of infection with a scaling factor, r_{scale}, and this parameter was calibrated such that the mean probability of infection of a susceptible individual staying in the same cabin as an infected case is approximately 20% assuming exposure in the cabin and during all shared activities throughout the entire duration of infectiousness, similar to the household attack rates for SARS-CoV-2 Delta variant cases[34,49]. β_{ij} is the relative risk of infection

ARTICLE

depending on vaccination status and mask wearing behaviours, and is parameterised to reduce the probability of infection according to Table 4.

Interventions. In the testing interventions, the sensitivity of the tests were assumed to vary with viral load. We assumed PCR is 100% sensitive for cycle threshold (Ct) values (a measure of viral load) below 35 and rapid antigen tests are 94.5% sensitive for Ct values below 25 and lowered sensitivities as the Ct values increases[22]. The viral load trajectory was modelled in relation to the Delta variant, rising above the limits of test detection three days before symptoms onset with prolonged shedding post symptoms onset[12,23,24].

For the mask wearing intervention, the expected weight of the contact between individuals i and j of contact group g are then modified based on the intervention of mask wearing and vaccination as follows:

$$\bar{\lambda}_{ijg}(t) = w_{ij} \int_{t-1}^{t} f(u;\mu_i, \alpha_i, \omega_i) du\, r_{scale} \left[(1 - m_g)\beta^v + m_g\beta^{mv} \right] \quad (5)$$

where m_g is the probability that the contact between any pairs of individual of a contact group g occurs while wearing a mask. β_v and $\beta_{m,v}$ are the relative risk of infection based on the vaccination status of the infector and infectee (Table 2).

For each intervention or combination of interventions, we ran 1000 simulations. We estimated the incidence by the day of infection, the number of cases in each generation, and the expected final outbreak size. All analyses were done in R version 4.0.4[50].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data are available in the manuscript or the supplementary information. The data used for our analyses is publicly available at https://doi.org/10.5281/zenodo.6009027

Code availability

The code used for our analyses is publicly available at https://doi.org/10.5281/zenodo.6009027

References

1. Pavli, A. & Maltezou, H. C. COVID-19 vaccine passport for safe resumption of travel. *J. Travel Med.* **28**, taab079 (2021).
2. GOV.UK, D. for D., Culture, Media and Sport. *Events Research Programme: Phase I Findings.* https://www.gov.uk/government/publications/events-research-programme-phase-i-findings/events-research-programme-phase-i-findings (GOV.UK, 2021).
3. The Lancet Microbe. Vaccine certificates: does the end justify the means? *Lancet Microbe* **2**, e130 (2021).
4. Singapore Tourism Board. Third consecutive year of growth for Singapore tourism sector in 2018. https://www.stb.gov.sg/content/stb/en/media-centre/media-releases/third-consecutive-year-of-growth-for-singapore-tourism-sector-in-2018.html (2019).
5. Curley, A., Garber, R., Krishnan, V. & Tellez, J. For corporate travel, a long recovery ahead. https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/for-corporate-travel-a-long-recovery-ahead (2020).
6. Environment Modelling Group & Department for Digital, Culture, Media and Sport. *EMG and DCMS: Science Framework for Opening up Group Events, 16 March 2021.* https://www.gov.uk/government/publications/emg-and-dcms-science-framework-for-opening-up-group-events-16-march-2021 (GOV.UK, 2021).
7. Hoang, T. et al. A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology* **30**, 723–736 (2019).
8. Cattuto, C. et al. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE* **5**, e11596 (2010).
9. Klepac, P., Kissler, S. & Gog, J. Contagion! The BBC four pandemic—the model behind the documentary. *Epidemics* **24**, 49–59 (2018).
10. Salathé, M. et al. A high-resolution human contact network for infectious disease transmission. *Proc. Natl Acad. Sci. USA* **107**, 22020–22025 (2010).
11. Li, B. et al. Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant. *Nat. Commun.* **13**, 460 (2022).
12. Ong, S. W. X. et al. Clinical and virological features of SARS-CoV-2 variants of concern: a retrospective cohort study comparing B.1.1.7 (Alpha), B.1.315 (Beta), and B.1.617.2 (Delta). *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciab721 (2021).
13. Elliott, P. et al. REACT-1 round 13 final report: exponential growth, high prevalence of SARS-CoV-2 and vaccine effectiveness associated with Delta variant in England during May to July 2021. http://spiral.imperial.ac.uk/handle/10044/1/90800 (2021).
14. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
15. O'Connell, J. & O'Keeffe, D. T. Contact tracing for Covid-19—a digital inoculation against future pandemics. *N. Engl. J. Med.* **385**, 484–487 (2021).
16. Smith, J. A. E. et al. Public health impact of mass sporting and cultural events in a rising COVID-19 prevalence in England. *Epidemiol. Infect.* **150**, e42 (2022).
17. Eames, K., Bansal, S., Frost, S. & Riley, S. Six challenges in measuring contact networks for use in modelling. *Epidemics* **10**, 72–77 (2015).
18. Yong, S. E. F. et al. Connecting clusters of COVID-19: an epidemiological and serological investigation. *Lancet Infect. Dis.* **20**, 809–815 (2020).
19. Liu, Y., Eggo, R. M. & Kucharski, A. J. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* **395**, e47 (2020).
20. Firth, J. A. et al. Using a real-world network to model localized COVID-19 control strategies. *Nat. Med.* **26**, 1616–1622 (2020).
21. Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
22. Dinnes, J. et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst. Rev.* https://doi.org/10.1002/14651858.CD013705.pub2 (2021).
23. Chia, P. Y. et al. Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine-breakthrough infections: a multi-center cohort study. https://doi.org/10.1101/2021.07.28.21261295 (2021).
24. Kissler, S. M. et al. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated individuals. https://doi.org/10.1101/2021.02.16.21251535 (2021).
25. Nanduri, S. Effectiveness of Pfizer-BioNTech and Moderna vaccines in preventing SARS-CoV-2 infection among nursing home residents before and during widespread circulation of the SARS-CoV-2 B.1.617.2 (Delta) variant—National Healthcare Safety Network, March 1–August 1, 2021. *Morb. Mortal. Wkly. Rep.* **70**, 1163–1166 (2021).
26. Fowlkes, A. Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection among frontline workers before and during B.1.617.2 (Delta) variant predominance—Eight U.S. Locations, December 2020–August 2021. *Morb. Mortal. Wkly. Rep.* **70**, 1167–1169 (2021).
27. Harris, R. J. et al. Effect of vaccination on household transmission of SARS-CoV-2 in England. *N. Engl. J. Med.* **385**, 759–760 (2021).
28. Chu, D. K. et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* **395**, 1973–1987 (2020).
29. Yapp, R., Willis, Z. & Jones, J. Coronavirus (COVID-19) Infection Survey technical article: analysis of populations in the UK by risk of testing positive for COVID-19, Office for National Statistics, September 2021, https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/analysisofpopulationsintheukbyriskoftestingpositiveforcovid19september2021 (2021).
30. Abaluck, J. et al. Impact of community masking on COVID-19: a cluster-randomized trial in Bangladesh. *Science* **375**, eabi9069 (2022).
31. Expert Taskforce for the COVID-19 Cruise Ship Outbreak. Epidemiology of COVID-19 outbreak on cruise ship quarantined at Yokohama, Japan, February 2020. *Emerg. Infect. Dis.* **26**, 2591–2597 (2020).
32. Keeling, M. J., Hollingsworth, T. D. & Read, J. M. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J. Epidemiol. Commun. Health* **74**, 861–866 (2020).
33. Nishiura, H. Backcalculating the incidence of infection with COVID-19 on the diamond princess. *J. Clin. Med.* **9**, 657 (2020).
34. Kang, M. et al. Transmission dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant infections in Guangdong, China, May to June 2021. *Eurosurveillance* **27**, 2100815 (2022).
35. Reardon, S. How the Delta variant achieves its ultrafast spread. *Nature* https://doi.org/10.1038/d41586-021-01986-w (2021).
36. Stockbridge, M. et al. Technical Report: In vitro and clinical post-market surveillance of Biotime SARS-CoV-2 Lateral Flow Antigen Device in detecting the SARS-CoV-2 Delta variant (B.1.617.2). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/999867/in-vitro-and-clinical-post-market-surveillance-of-Biotime-SARS-CoV-2-Lateral-Flow-Antigen-Device-in-detecting-the-SARS-CoV-2-Delta-variant-B.1.617.2.pdf (2021).
37. Abbott. *Evaluating Delta and other COVID Variants to Ensure Test Effectiveness.* https://www.abbott.com/corpnewsroom/diagnostics-testing/monitoring-covid-variants-to-ensure-test-effectiveness.html (Abbott, 2021).
38. Zhang, X. & Wang, J. Dose-response relation deduced for coronaviruses from coronavirus disease 2019, severe acute respiratory syndrome, and middle east

116

ARTICLE

respiratory syndrome: meta-analysis results and its application for infection risk assessment of aerosol transmission. *Clin. Infect. Dis.* **73**, e241–e245 (2021).

39. Government of the United Kingdom. *Guidance for Contacts of People with Confirmed Coronavirus (COVID-19) Infection Who Do Not Live with the Person*. https://www.gov.uk/government/publications/guidance-for-contacts-of-people-with-possible-or-confirmed-coronavirus-covid-19-infection-who-do-not-live-with-the-person/guidance-for-contacts-of-people-with-possible-or-confirmed-coronavirus-covid-19-infection-who-do-not-live-with-the-person (GOV.UK, 2021).

40. Ministry of Health, Singapore. MOH | FAQs - Confirmed Cases and Contact Tracing. https://www.moh.gov.sg/covid-19/general/faqs-confirmed-cases-and-contact-tracing (2021).

41. Centres for Disease Control and Prevention. *Public Health Guidance for Community-Related Exposure*. https://www.cdc.gov/coronavirus/2019-ncov/php/public-health-recommendations.html (Centers for Disease Control and Prevention, 2020).

42. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101**, 3747–3752 (2004).

43. Lu, J. et al. COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. *Emerg. Infect. Dis.* **26**, https://doi.org/10.3201/eid2607.200764 (2020).

44. Shen, Y. et al. Community outbreak investigation of SARS-CoV-2 transmission among bus riders in Eastern China. *JAMA Intern. Med.* **180**, 1665–1671 (2020).

45. Isakbaeva, E. T. et al. Norovirus transmission on cruise ship. *Emerg. Infect. Dis.* **11**, 154–157 (2005).

46. Buitrago-Garcia, D. et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med.* **17**, e1003346 (2020).

47. Ferretti, L. et al. The timing of COVID-19 transmission. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.04.20188516 (2020).

48. Lehtinen, S., Ashcroft, P. & Bonhoeffer, S. On the relationship between serial interval, infectiousness profile and generation time. *J. R. Soc. Interface* **18**, 20200756 (2021).

49. Public Health England. *SARS-CoV-2 Variants of Concern and Variants Under Investigation in England Technical Briefing 14*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/991343/Variants_of_Concern_VOC_Technical_Briefing_14.pdf (2021).

50. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).

51. Zhang, M. et al. Transmission dynamics of an outbreak of the COVID-19 Delta variant B.1.617.2—Guangdong Province, China, May–June 2021. *CCDCW* **3**, 584–586 (2021).

52. Ng, O. T. et al. Impact of Delta variant and vaccination on SARS-CoV-2 secondary attack rate among household close contacts. *Lancet Reg. Health West Pac.* **17**, 100299 (2021).

Singapore CruiseSafe working group
Annie Chang[8], Jade Kong[8], Jazzy Wong[8], Ooi Jo Jin[8], Deepa Selvaraj[1], Dominique Yong[1], Jocelyn Lang[1] & Abilash Sivalingam[9]

[8]Singapore Tourism Board, Singapore, Singapore. [9]Government Technology Agency, Singapore, Singapore.

CMMID COVID-19 working group
Simon R. Procter[2], Stefan Flasche[2], William Waites[2], Kiesha Prem[2], Carl A. B. Pearson[2], Hamish P. Gibbs[2], Katharine Sherratt[2], C. Julian Villabona-Arenas[2], Kerry L. M. Wong[2], Yang Liu[2], Paul Mee[2], Lloyd A. C. Chapman[2], Katherine E. Atkins[2], Matthew Quaife[2], James D. Munday[2], Sebastian Funk[2], Rosalind M. Eggo[2], Stèphane Huè[2], Nicholas G. Davies[2], David Hodgson[2], Kaja Abbas[2], Ciara V. McCarthy[2], Joel Hellewell[2], Sam Abbott[2], Nikos I. Bosse[2], Oliver Brady[2], Rosanna C. Barnard[2], Mark Jit[2], Damien C. Tully[2],

117

ARTICLE

Graham Medley[2], Fiona Yueqian Sun[2], Christopher I. Jarvis[2], Rachel Lowev, Kathleen O'Reilly[2], Sophie R. Meakin[2], Akira Endo[2], Frank G. Sandmann[2], W. John Edmunds[2], Mihaly Koltai[2], Emilie Finch[2], Amy Gimma[2], Alicia Rosello[2], Billy J. Quilty[2], Yalda Jafari[2], Gwenan M. Knight[2], Samuel Clifford[2] & Timothy W. Russell[2]

118

6.7 Outro

Disease transmission models, including the current study, often assume that contact patterns are static. In the next chapter, I used the same cruise sailing dataset and contact network data from other settings to explore the implications of time-varying contact patterns on disease transmission, outbreak control and modelling.

7 Temporal contact patterns and the implications for predicting superspreaders and planning of targeted outbreak control

As the population progressively acquired immunity during the COVID-19 pandemic, policymakers were interested to know if and when herd immunity can be achieved. In other words, when can the country relax existing outbreak control measures without causing another surge in COVID-19 cases. Studies have shown that contacts in a population are often heterogeneous. Thus, herd immunity can be achieved when a small number of highly connected individuals acquire immunity through natural infection or vaccination [1–3]. However, these studies assumed that social contact patterns are static, and the generalisability of these findings to real-world contact patterns that change over time remains unclear.

Using the contact network data collected in Chapter 6 and other previously published literature [4–10], the focus of this chapter is to study the fundamental properties of temporal networks and understand the implications of outbreak control measures. At a sub-population level, I developed a metric to characterise the retention of contacts in a temporal network over consecutive timesteps relative to a fully static and fully dynamic network. I applied the metric to contact networks on cruises, a community, high schools, a hospital and workplaces. Furthermore, I explored the type of contacts that are likely to be retained to determine which subpopulations are more likely to exhibit consistent contacts and thus influence the type of outbreak controls required. Moving from a subpopulation-level to an individual-level, I also analysed the number of contacts made by each individual in each timestep to understand the consistency of an individual in displaying high levels of connectivity over the study period. I aimed to determine the feasibility of identifying potential 'superspreaders' ahead of time for targeted outbreak control. Lastly, I also estimated the repetition of contacts over the observed days to determine the potential overestimation in outbreak resource planning if we assume independent contacts made by an individual each day.

The supplementary information of this study is in Appendix G.

References

1.   Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science. 2020;369: 846–849. doi:10.1126/science.abc6810

2.   Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the

design of control programs. Proceedings of the National Academy of Sciences. 1997;94: 338–342. doi:10.1073/pnas.94.1.338

3.   Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438: 355–359. doi:10.1038/nature04153

4.   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;13: 1956. doi:10.1038/s41467-022-29522-y

5.   Kissler SM, Klepac P, Tang M, Conlan AJK, Gog JR. Sparking "The BBC Four Pandemic": Leveraging citizen science and mobile phones to model the spread of disease. bioRxiv; 2020. p. 479154. doi:10.1101/479154

6.   Mastrandrea R, Fournet J, Barrat A. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. PLOS ONE. 2015;10: e0136497. doi:10.1371/journal.pone.0136497

7.   Fournet J, Barrat A. Contact Patterns among High School Students. PLOS ONE. 2014;9: e107878. doi:10.1371/journal.pone.0107878

8.   Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, et al. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. PLOS ONE. 2013;8: e73970. doi:10.1371/journal.pone.0073970

9.   Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science. 2015;3: 326–347. doi:10.1017/nws.2015.10

10.   Génois M, Barrat A. Can co-location be used as a proxy for face-to-face contacts? EPJ Data Sci. 2018;7: 1–18. doi:10.1140/epjds/s13688-018-0140-1

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

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 <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	2006052	**Title**	Ms
First Name(s)	Rachael		
Surname/Family Name	Pung		
Thesis Title	COVID-19 Transmission Dynamics and Implications for Outbreak Control in Singapore		
Primary Supervisor	Adam J. 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?	Journal of The Royal Society Interface
Please list the paper's authors in the intended authorship order:	Rachael Pung, Josh A. Firth, Timothy Russell, Tim Rogers, Vernon J. Lee, Adam J. Kucharski
Stage of publication	**Not yet submitted**

Improving health worldwide   **www.lshtm.ac.uk**

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, investigation, visualisation, and writing of original draft.

SECTION E

Student Signature	*[signature]*
Date	7 March 2024

Supervisor Signature	*[signature]*
Date	13 March 2024

123

7.1 Abstract

Epidemic models often heavily simplify the dynamics of human-to-human contacts, but the resulting bias in outbreak dynamics – and hence requirements for control measures – remains unclear. Even if high-resolution temporal contact data were routinely used for modelling, the role of this temporal network structure towards outbreak control is not well characterised. We address this by assessing dynamic networks across varied social settings in three ways. Firstly, we characterised the distribution of retained contacts over consecutive timesteps by developing a novel metric, the "retention index", which accounts for the change in the number of contacts over consecutive timesteps on a normalised scale with the extremes representing fully static and fully dynamic networks. Secondly, we described the repetition of contacts over the days by estimating the frequency of contact pairs occurring over the study duration. Thirdly, we distinguish the difference between 'superspreader' and infectious individuals driving 'superspreading events' by estimating the connectivity of an individual (i.e. individual has high connectivity in a timestep if he accounts for 80% of the contacts in the timestep) and the frequency of exhibiting high connectivity. Using 11 networks from 5 settings studied over 3–10 days, we estimated that more than 80% of the individuals in most settings were highly connected for only short periods. This suggests a challenge to identify superspreaders, and more individuals would need to be targeted as part of outbreak interventions to achieve the same reduction in transmission as predicted from a static network. Taking into account repeated contacts over multiple days, we estimated simple resource planning models might overestimate the number of contacts made by an infector by 20%–70%. In workplaces and schools, contacts in the same department accounted for most of the retained contacts. Hence, outbreak control measures would be better off targeting specific sub-populations in these settings to reduce transmission. In contrast, no obvious type of contact dominated the retained contacts in hospitals, so reducing the risk of disease introduction is critical to avoid disrupting the interdependent work functions. This study identified key epidemiological properties of temporal network that potentially shapes outbreak dynamics and illustrate the need for incorporating such properties in outbreak simulations.

7.2 Significance

Directly transmitted infectious diseases spread through social contacts that can change over time. Modelling studies have largely focused on simplifying these contact patterns to predict outbreaks but the assumptions on contact patterns may bias results and, in turn, conclusions on the effectiveness of control measures. An ongoing challenge is, therefore, how to measure key properties of complex and dynamic networks, to facilitate the development of network disease simulation models which ensures that outbreak analysis is transparent and interpretable in the real-world context. To address this challenge, we analysed 11 networks from 5 different settings

and developed new metrics to capture crucial epidemiological features of these networks. We showed that there is an inherent difficulty in identifying individual 'superspreaders' reliably in most networks. In addition, the key types of individuals driving transmission vary across settings, thus requiring different outbreak control measures to reduce disease transmission or the risk of introduction. Simple models to mimic disease transmission in temporal networks may not capture the repeated contacts over the days, and hence could incorrectly estimate the resources required for outbreak control. Our study characterised temporal network data in epidemiologically relevant ways and is a step towards developing simplified contact networks to capture real-world contact patterns for future outbreak simulation studies.

7.3 Background

Directly transmitted infections spread through human social contacts, but the dynamic and often high-dimensional nature of these networks has historically made them difficult to measure and interpret. As a result, epidemic models often implicitly approximate complex dynamic networks with simpler contact processes, including static networks [1,2], branching processes [3] and compartmental models [4]. These relatively simpler models of disease transmission have been well-studied (Figure 7.1), but it remains unclear how they compare with real-life temporal social networks, which exhibit a mix of repeated and occasional contacts [5,6]. As such, the assumptions in these simpler models could bias model outputs that are crucial for epidemic planning and response, from estimating the required resources for contact tracing and testing programmes to assessing the impact of social distancing measures and vaccine coverage [7–9].
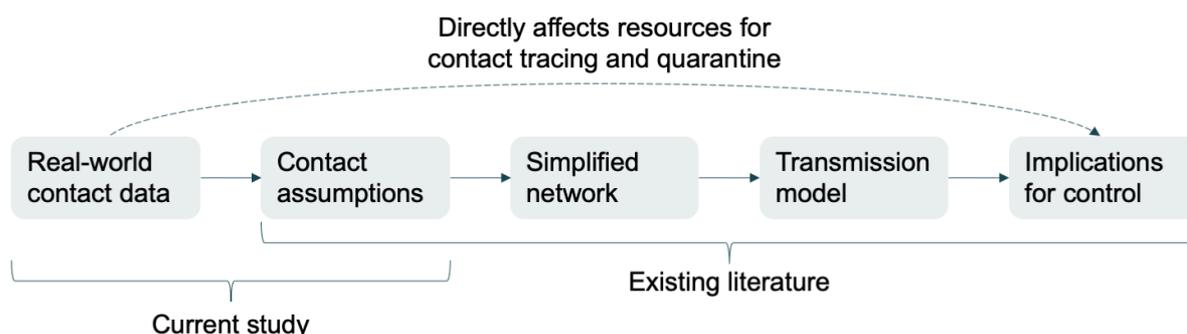
Figure 7.1 Different components of contact network studies and how they influence outbreak control measures

There has been recent progress in the collection of dynamic contact network data via proximity sensors [10,11] or mobile devices [12]. The automated nature of such data collection enabled large-scale deployment for contact tracing during the COVID-19 pandemic [11,13]. These devices work by exchanging radio frequency identification (RFID) signals within a calibrated distance, enabling us to monitor contacts and map the emerging network structure. This can – in theory – enable us to interpret the

transmission process on temporal networks. However, in practice, most studies still tend to simplify the temporal network structure by extending static network properties, which depend on characteristics such as population sizes [6], making it hard to compare findings across studies. Furthermore, it can be challenging to tease out the effects of different network features on the transmission dynamics in models [6,14,15]. Finally, temporal contact data in some studies was collected through self-reported contact dairies, which may be prone to recall bias [5,14,16]. With the extensive data collected from automated devices, this is increasingly an opportunity to better compare contact structures and hence, the implications for key transmission processes.

Using real-world temporal social data from over 4 million contact events collected across five settings (cruises, community, schools, hospital and workplaces), we quantified the impact of dynamic contacts on key epidemiological metrics driving person-to-person transmission across these varied social settings. As well as examining the range of bias introduced by common simplifying assumptions, we identify the extent to which it is possible to identify individuals linked to superspreading events reliably. To characterise time-varying properties of the real-life networks, we developed a new metric – the retention index – that allows complex dynamic networks to be summarised and compared in an epidemiologically meaningful manner.

7.4   Methods

7.4.1 Temporal contact network data

We collated temporal contact network data from previously published studies across different settings, with contacts recorded using proximity sensors or mobile devices (Table 7.1). These devices were calibrated to record contacts between pairs of individuals within a specified radius on cruises and in a community or, alternatively face-to-face interactions in high schools, a hospital and workplaces. The radius approach is omnidirectional, while the face-to-face methods record a contact when the sensors face each other. For each network, we performed preliminary analysis to identify common types of contact, contact durations, and delays before the next contact occurs between a pair of individuals (Table 7.1). Contact data from the cruises were recorded in 15-second intervals, while in all other networks, contacts were recorded in 5-minute- or 20-second intervals.

To analyse the network properties, we first needed to choose a timescale for defining a 'contact' within each dataset. In our main analysis, we set the length of the timestep for each network based on the median delay in contact. The timestep was set at 15-min, or 1-hr for subsequent sensitivity analysis. We also performed additional sensitivity analysis, assuming the directed contact networks in the non-cruise settings were undirected. For the high school, hospital and workplace networks, a small timestep (e.g. 20-sec) would result in few repeated contacts over consecutive timesteps because the median delay between contact events was higher than the

contact duration (Table 7.1). As such, the main analysis considered the contact patterns based on timesteps defined for each network, while our sensitivity analysis standardised the timesteps across all networks. A contact is defined to occur within a timestep if it lasts for at least the median contact duration for respective networks (Table 7.1). At one theoretical extreme, networks may exhibit no variation over time, resulting in a static network, where the contacts remain the same over consecutive timesteps; at the other extreme, we have fully dynamic networks, where every individual's contacts are drawn randomly at each timesteps (Figure 7.2). When simulating the fully dynamic network across consecutive timesteps, we retained the degree distribution of each individual observed in a timestep but randomly rewired their contacts. This ensures that the fully dynamic network has the same degree distribution as the static network of the same timestep.

Figure 7.2 Contacts made by an individual of interest (brown, centre) in a single timestep with contacts retained from the previous timestep (blue), contacts that were not retained from the previous timestep (grey with black outline) and new contacts in current timestep (red) for (A) fully static; (B) temporal; and (C) fully dynamic network.
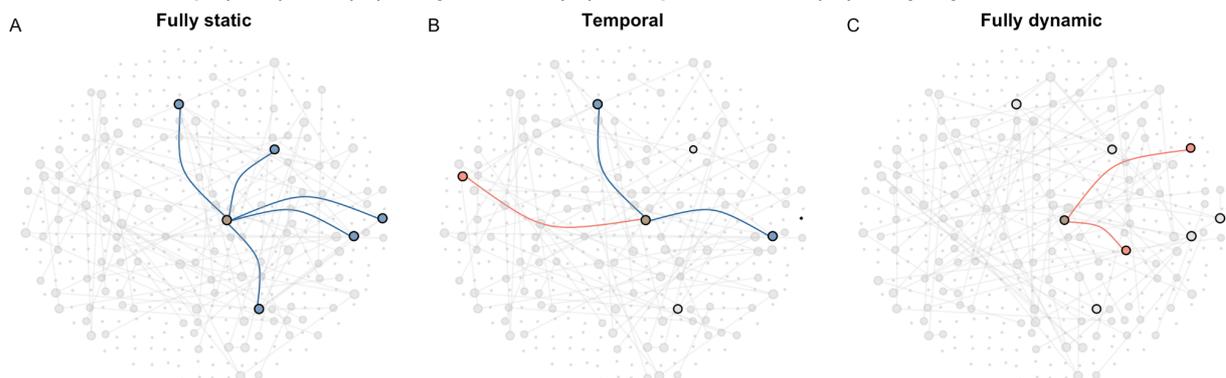
Table 7.1 Characteristics of real-world contact networks

Network setting	Study date, observed days	Types of contact	Median contact duration (sec)	Median delay in contact (sec)	Remarks (references)
Cruises, Singapore	Nov 2020, 3d Nov 2020, 3d Jan 2021, 3d Feb 2021, 3d (i.e. four sailings with two in Nov 2020)	P: passenger C: crew P-P (same cabin) P-P (different cabin) C-C (same department) C-C (different department) P-C	900 for all four sailings	900 for all four sailings	COVID-19 restrictions onboard. Undirected network [11]
Community, Haslemere, UK	Oct 2017, 3d	Household Non-household	300	600	No data before 0700 hrs and after 2300 hrs. Directed network [12]
High Schools, Marseilles, France	Dec 2011, 4d Nov 2012, 7d Dec 2013, 5d	Classmates Non-classmates	20 for all three high school	140 120 100	No data over weekends. Directed network [16,17]
Hospital, Lyon, France	Dec 2010, 5d	Same department Different department	20	140	Directed network [18]
Workplaces, France	Jun 2013, 10d 2015, 10d	Same department Different department	20 for both workplaces	220 120	No data over weekends. Directed network [19,20]

7.4.2 Contact retention

To explore how contacts were retained and changed over time, we defined the distribution of the number of retained contacts, r, over consecutive timesteps, t and $t+1$, in the network as follows:

$$P(r_{t+1}) = \sum_{k_{t+1}=0}^{N-1} \sum_{k_t=0}^{N-1} P(r_{t+1} \mid k_t, k_{t+1}) P(k_{t+1} \mid k_t) P(k_t), \qquad r_{t+1} \leq k_t, k_{t+1} \qquad (7.1)$$

where k_t is the number of contacts (i.e. degree) in timestep t and N is the number of individuals in a network. The maximum possible number of contacts an individual could make is N-1. For static or fully dynamic networks, where contacts are either fixed or made at random, $P(r_{t+1} \mid k_t, k_{t+1})$ of equation (1) is replaced with the binomial distribution as follows:

$$P(r_{t+1}) = \sum_{k_{t+1}=0}^{N-1} \sum_{k_t=0}^{N-1} \frac{k}{r_{t+1}} p^{r_{t+1}} (1-p)^{k_t - r_{t+1}} P(k_{t+1} \mid k_t) P(k_t) \qquad (7.2)$$

where k is the minimum of k_t and k_{t+1} and p is the binomial probability of preserving a contact between a pair of individuals. For static networks, $p = 1$ and equation (2) simplifies as follows

$$\begin{aligned} P(k_{t+1} \mid k_t) &= 1, & k_{t+1} &= k_t \\ &= 0, & k_{t+1} &\neq k_t \end{aligned} \qquad (7.3)$$

$$P(r_{t+1}) = P(k_t) \qquad (7.4)$$

For fully dynamic networks with randomly made links, $p = \frac{k_{t+1}}{N-1}$ and equation (2) is expressed as follows

$$P(k_{t+1} \mid k_t) = P(k_{t+1}) \qquad (7.5)$$

$$P(r_{t+1}) = \sum_{k_{t+1}=0}^{N-1} \sum_{k_t=0}^{N-1} \frac{k}{r_{t+1}} p^{r_{t+1}} (1-p)^{k_t - r_{t+1}} P(k_{t+1}) P(k_t) \qquad (7.6)$$

By definition, we expect the highest mean number of retained contacts to be observed in static networks, $\bar{r}_{stat}$, and the lowest in fully dynamic networks, $\bar{r}_{dyna}$. To quantify the mean number of retained contacts in our collated temporal networks, $\bar{r}_{temp}$ we computed a scaled metric, defined as the 'retention index', as follows:

$$\bar{r} = \frac{\bar{r}_{temp} - \bar{r}_{dyna}}{\bar{r}_{stat} - \bar{r}_{temp}} \tag{7.7}$$

This metric (retention index) provides a standardised measure of where a network lies between the two theoretical extremes. If $\bar{r} \to 1$, the temporal network reflects a fully static (and hence fully predictable) structure; when $\bar{r} \to 0$, the temporal network reflects a fully dynamic (and hence non-predictable) structure.

7.4.3 Epidemiological metrics

If contacts are retained over consecutive timesteps, it will result in a longer duration of continuous contact and, hence, a higher risk of transmission. Under the assumption that infection does not change the individual's contact patterns (e.g. for an infection that exhibits substantial asymptomatic or pre-symptomatic transmission), clustering of retained contacts would also limit further disease transmission by an infector if the contact is already infected. To identify predictors of contact retention over consecutive timesteps, we estimated the proportion of repeated contacts occurring for each type of contact (Table 7.1). Besides evaluating the retention of contacts over consecutive timesteps, we can also evaluate the repetition of contacts over different days by estimating the frequency distribution of contact encounters in days among all the contact pairs.

We also assessed the bias introduced when assuming independence of contacts over the days. To do this, we estimated the difference between the cumulative unique contacts from the start to the day of interest, and the sum of unique contacts each day from the start to the day of interest. We estimated the relative difference in contacts to generalise the findings across different studies with different population sizes.

7.4.4 Extent of superspreaders and superspreading events

We defined potential 'superspreaders' as individuals frequently identified to account for the top 80% of the contacts made or contact duration over the observed period (see example in next paragraph). We also define potential 'superspreading events' to be transmission driven by individuals less frequently identified to account for the top 80% of the contacts or contact duration over the observed period. The latter group of individuals typically forms few contacts. However, for a small proportion of the time, they have many or prolonged contacts and could disproportionately account for many transmission events in that time if they were infectious [21,22].

In each timestep, we identified the individuals accounting for the top 80% of contacts or contact duration (i.e. highly connected individuals). The minimum and maximum proportion of timesteps that an individual was identified in this top group could range between 0 to 1. For each incremental proportion of time, we estimated the proportion of the population identified for the corresponding time. To illustrate the extent of

transmission events driven by superspreader or superspreading dynamics, we plot the cumulative proportion of the population identified for at least a given proportion of time. For example, we might identify a certain proportion of the population to be highly connected in at least half of the number of observed timesteps. In this example, we could label this group as 'superspreaders'. On the other hand, we might identify a certain proportion of the population to be highly connected but only in less than half of the number of observed timesteps. We could label this group as individuals who drive 'superspreading events'.

To provide context of how the real-world networks compare with static and fully dynamic networks when visualising our results, we simulated a homogenous and an overdispersed network over different timesteps to estimate the above metrics. In a homogenous network, expected 80% of the population accounts for 80% of the contacts (i.e. $p_{80} = 0.8$), while in an overdispersed network, this is less than 80% of the population (in this study, we used 50%, i.e. $p_{80} = 0.5$). For a static network, regardless of a homogeneous or an overdispersed network, the same proportion of the population was identified across all timesteps by definition. For a fully dynamic network of varying timesteps, the proportion of the population identified for each incremental proportion of time is approximately p_{80} raised to the power of s, where s is the number of timesteps corresponding to the proportion of time.

7.5 Results

7.5.1 Contact retention

We found considerable variation in the retention index, $\bar{r}$, across different networks and over time. For example, cruise networks exhibited an $\bar{r}$ of 0.59 (IQR 0.52–0.81). This study was conducted under strict COVID-19 physical distancing and social gathering restrictions onboard the cruises (Figure 7.3A). As a result, most of the repeated contacts occurred among passengers who shared the same cabin and, hence, were in the same travelling group and crew members of the same department (Figure 7.3B). We estimated an $\bar{r}$ of less than 0.5 in only 12–24% of the observed timesteps for the four cruise sailings, indicating that in a given time period, contacts are much more likely to be retained rather than new contacts being made. Between 30–60% of these timesteps with lower $\bar{r}$ occurred between 1200-1400 hrs and 1800-2000 hrs across the four cruise sailings. Passengers were likely to be engaged in dining during these periods and previous work showed that dining settings promote social contact [11]. The seating arrangements or the movement patterns (e.g. buffet counters) facilitate increased mixing and interaction between passengers of different cabins (Figure 7.3B and Supplementary Figure 1). High values of $\bar{r}$ were also observed at the start and end of each day, the result of contact between passengers from the same cabin.

Pre-pandemic community networks from the UK exhibited an even higher $\bar{r}$ of 0.73 (IQR 0.65–0.81). About 40% of the contacts occurred before 0900 hrs and after 1700 hrs when the individual is likely to be at home with household contacts (Figure 7.3A and B, and Supplementary Figure 1). In contrast, networks from schools, a hospital and workplaces showed lower $\bar{r}$ of 0.58 (IQR 0.44–0.69), 0.49 (IQR 0.36–0.64) and 0.50 (IQR 0.33–0.61) respectively. In these networks, $\bar{r}$ was below 0.5 for about half of the observed duration and changes in $\bar{r}$ did not exhibit any time trends, unlike the cruise or community networks (Supplementary Figure 1). Moreover, at low and at high values of $\bar{r}$, there was no apparent variation in the type of retained contacts. We estimated that contacts made between classmates or individuals of the same department form the majority of the contacts in each timestep for the high school network, about 60% for the hospital network and about 80% for the workplace networks. We observed similar proportions among the retained contacts (Figure 7.3B).

The overall patterns in our analysis remained unchanged when we performed sensitivity analyses around choice of timestep and contact definition. We obtained similar results when assuming undirected contacts in the non-cruise settings (Supplementary Figure 2), although when using fixed timesteps of 15-min or 1-hr for all networks, the overall median $\bar{r}$ of all networks was slightly lower than the main analysis. However, $\bar{r}$ in both the cruise and community networks remained higher than networks from schools, a hospital and workplaces (Supplementary Figure 3 and 4).

Figure 7.3 Contacts patterns in different network settings, (A) ridgeline plot showing distribution of contact retention index, $\bar{r}$, over consecutive timesteps, (B) proportion of the type of retained contacts for respective $\bar{r}$.
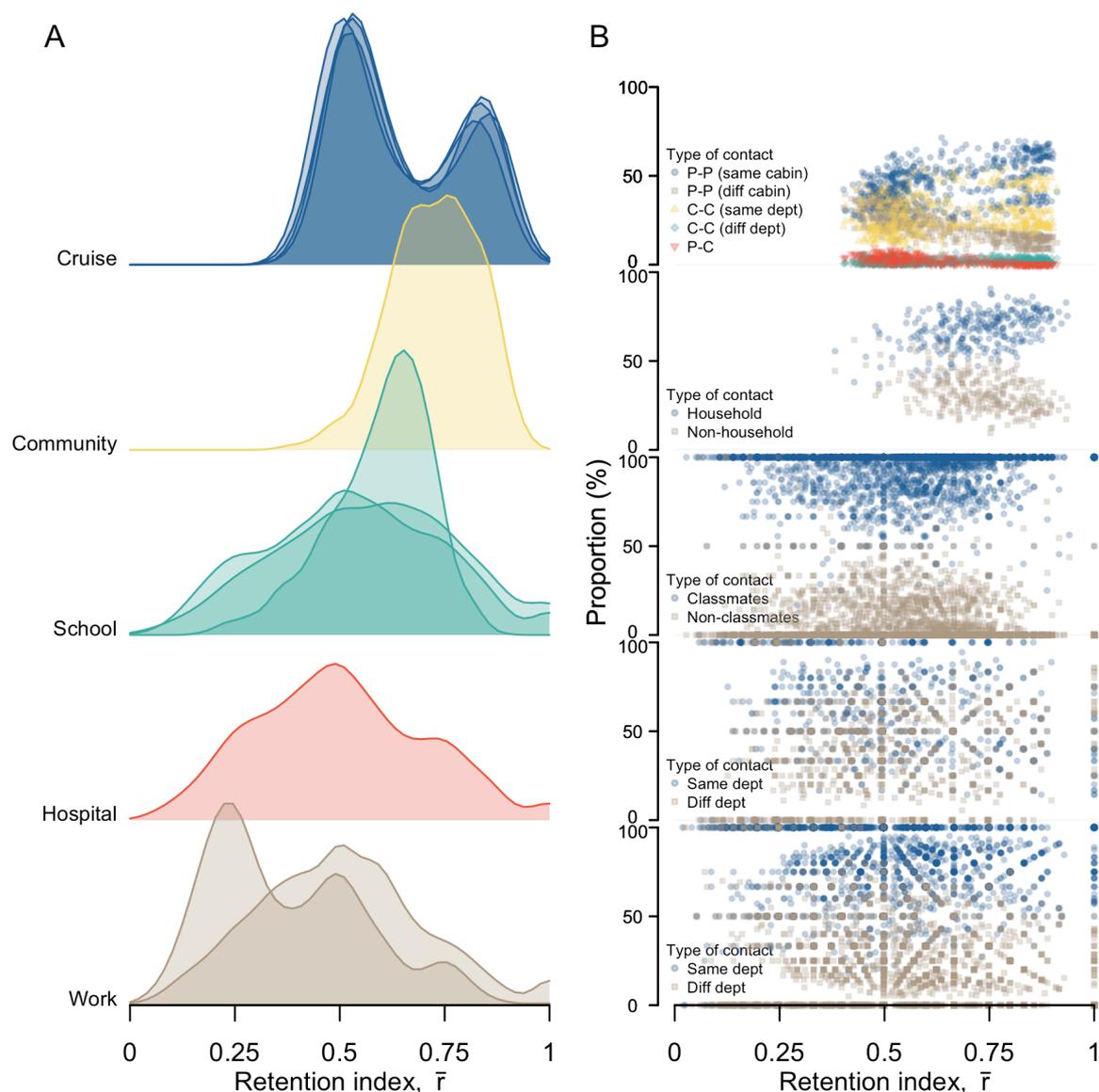
7.5.2 Epidemiological metrics

Although a longer study duration will in theory increase the probability of observing a repeated contact over multiple days, there was some agreement across different networks on the proportion of total measured contacts that occurred in one day out of all days in respective network studies. For studies conducted over three days, the proportion of total contacts that occurred over one-day was 86% (range 83–87%) in the cruises and 82% in the community (Figure 7.4A). For studies conducted over longer durations of up 10 days of recorded contacts, the proportion of total contacts recorded in a given day was 57% (range 52–60%) in the high schools, 51% in the

hospital and 47% (range 38–55%) in the workplace networks (Figure 7.4A). Across all the networks, over 75% of the contacts either occurred over one-day only or were repeated for less than half the study duration (Figure 7.4A).

When planning outbreak control measures such as contact tracing, we need to consider the number of unique contacts made per infected individual. If we did not account for repeated contacts over the days and instead assumed the measured number of daily contacts would be made independently each day, we could overestimate the number of unique contacts. With the exception of the community network, we found that we would overestimate the unique contacts by 13–35% across all networks after three days of observation under this independence assumption (Figure 7.4B). For longer study duration in the schools, this difference between the total and unique contacts was 71% (IQR 35%–110%) after seven days; for workplaces, the difference rose to 73% (IQR 33%–130%) after ten days (Figure 7.4B).

Figure 7.4 Contact pairs over the study duration in different networks, (A) cumulative distribution of contact encounters in days in pairs of contact. Study duration varied across networks and was normalised. For networks with the same study duration, such as the four cruises and three workplace networks, the distribution was represented by the median (lines) and range (shaded region). For networks with different study durations, such as the three high school networks, or a single network study, such as the community and hospital networks, the distribution of each network study was illustrated, (B) Median (shapes) and range (lines) of the relative difference in the number of unique contacts.
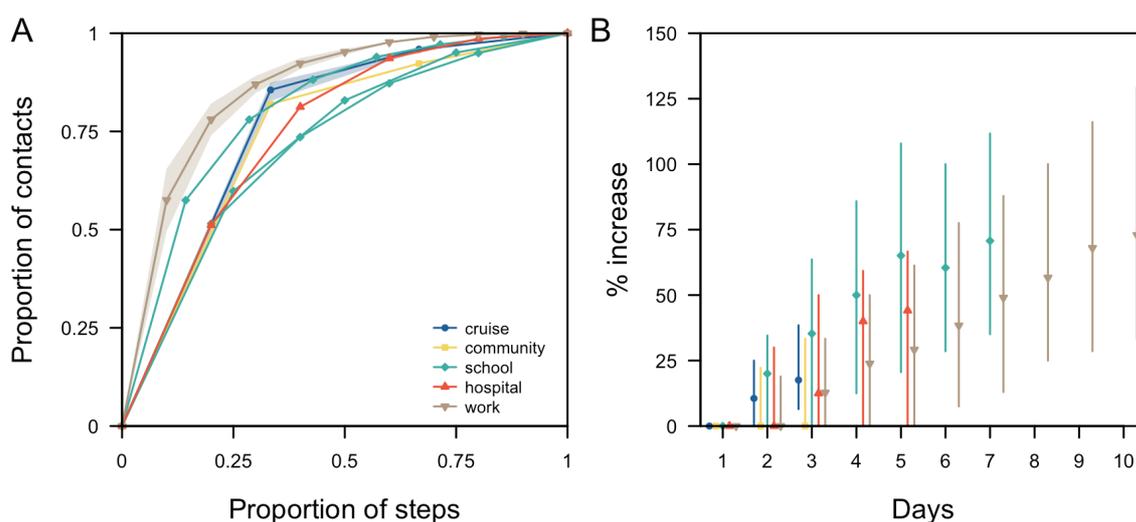
7.5.3 Extent of superspreaders and superspreading events

Depending on the level of overdispersion of individual-level contacts in a network and the duration of observation, our ability to correctly predict highly connected individuals in a given time period will vary. For a homogenous static network, 80% of the population accounts for 80% of the contacts made. As such, 80% of the population would be identified as highly connected across all the timesteps while the remaining 20% of the population would never be identified in this group (Figure 7.5, dotted lines). For a fully dynamic homogeneous network with 25 timesteps, 80% of the population accounts for 80% of the contacts in each timestep. Given changes in the network structure over the timesteps, only 40% of the population would be identified for at least half the total number of timesteps (Figure 7.5, dashed lines). For a fully dynamic overdispersed network with 10 timesteps, 50% of the population accounts for 80% of the contacts in each timestep. Consequently, only 5% of the population would be identified in at least half the observations (Figure 7.5, dot-dashed lines). We found that as networks transition from homogeneous to overdispersed, and as the duration of observation increases, the proportion of highly connected individuals that can be identified consistently is reduced.

Real-world networks with higher levels of contact retention had a higher probability of correctly predicting frequent, highly connected individuals but these individuals only accounted for less than 30% of the population. These are individuals who account for the top 80% of the contact episodes for at least half of the number of observed timesteps (i.e. potential 'superspreaders', top left region of each panel in Figure 7.5). In real-world cruise contact networks, 26% (range 22%–29%) of the population were predicted to fall into this potential 'superspreader' category. The remaining population are individuals who have high connections but for short periods of time only. These are individuals who are likely to drive 'superspreading events' (i.e. bottom right region of each panel in Figure 7.5). In particular, 44% (range 40%–48%) of the population were identified for less than a quarter of the observed timesteps (Figure 7.5A). In the community network, 9% of the population would be predicted to be potential 'superspreaders' while 81% of the population are likely to drive 'superspreading events' for less than a quarter of the time (Figure 7.5A). The proportion of the population identified as potential 'superspreaders' was less than 5% in the high school, hospital and workplace networks; the majority of the individuals would, if anything, drive 'superspreading events' instead (Figure 7.5A). Similar trends were observed when analysing the proportion of the population that accounted for the top 80% of the contact duration (Figure 7.5B).

Figure 7.5 Proportion of 'superspreaders' and 'superspreading events' in respective networks, estimated based on (A) contact episodes or (B) contact duration. For reference, grey lines represent homogeneous static network (dotted), homogeneous dynamic network in 25 timesteps (dashed) and overdispersed dynamic network in 10 timesteps (dot dashed). Cutoff marks for the proportion of individuals in the cruise networks who were highly connected for more than half the total number of timesteps (triangle) and those who were highly connected for less than a quarter of the time (dot) as shown.
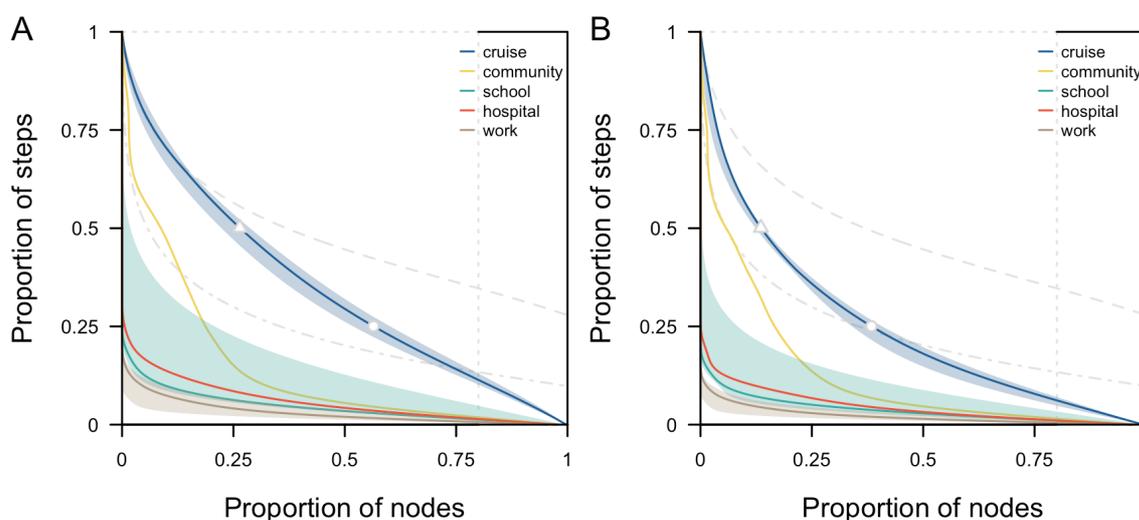
7.6 Discussion

Using real-world contact data collected from a variety of settings over different days and population sizes, we assessed the key structural properties of temporal networks that drive transmission processes and, hence, influence the effectiveness of outbreak control measures. We estimated that most individuals in each social context had high levels of connectivity with others for less than a quarter of the study duration. Contact retention and the type of contacts driving this retention varied across settings, emphasising the need for tailored outbreak analysis and control strategies for different settings.

In our analysis, we compared the properties of the real-world temporal networks relative to static and fully dynamic networks, normalised by the population size. This enabled us to contextualise our findings and allow for appropriate comparison across different networks. In particular, our study highlighted an inherent difficulty in predicting 'superspreaders' over time across different settings [5]. In cruise data, the high level of consistency in identifying highly connected individuals (i.e. 26% of the population identified to account for the top 80% of the contacts in more than half the total observed

timesteps) was likely influenced by the prevailing COVID-19 restrictions onboard during the study. Passengers and crews were encouraged to remain within their travel or working groups and to practise physical distancing from other groups [11]. However, the level of consistency in identifying highly connected individuals was generally low in all other networks. More than 80% of the population was identified to be highly connected for only a short period of the study duration. Targeting small groups of infectious individuals with high levels of connectivity has been shown to, in theory, produce an effective and efficient reduction in transmission, but such studies were largely based on static networks [23,24]. In contrast, our study showed that if we were to sample a network for a few days or a short period of time, and target individuals with high measured connectivity, this level of connectivity would generally turn out to be much lower if data collection were to be repeated in the near future. As such, when designing interventions to identify potential 'superspreaders', we would need to target a greater number of individuals than basic theory from static networks suggests in order to achieve the same reduction in transmission.

When an outbreak occurs, outbreak control policies often target subpopulations rather than individuals given the lack of information on contact patterns [15]. Across most social settings we analysed, contacts between individuals in the same social group (e.g. same cabin, department or school class) dominated interactions, even if retention of these contacts was variable. For high schools and workplaces, we estimated low contact retention even when most of these contacts were formed between individuals of the same class. This result corroborates previous findings indicating low levels of repeated contact among household contacts for those residing in dormitories [14].

When implementing outbreak control policies, our results suggest it is important to consider if the priority is to reduce disease introductions, or reduce transmission if introduced to a locality, and thus, which is the appropriate individuals or subpopulations to target with restrictions. In schools and workplaces, the majority of close contacts were from individuals of the same department or class, implying that targeted rather than school- or workplace-wide closures could still help to minimise disruption to activities. This would be particularly relevant if disease prevalence in the wider population is low and the likelihood of introductions to other departments or classes is low. In contrast, for settings such as hospitals, contacts from both the same (e.g. nurse-nurse contacts) or different (e.g. patient-nurse contacts) departments are likely to be retained over consecutive timesteps. This higher proportion of contacts between different departments is expected given the multi-faceted roles of healthcare workers [18]. Thus, more stringent measures to reduce the risk of nosocomial outbreaks starting is highly important to avoid disruptions to hospital functions.

While the use of detailed contact data to plan quarantine measures can provide an upper limit on the resources required [7,9], our results suggest the occurrence of repeated contacts would mean that simple analysis, based on cross-sectional data collection that assumes independence of contacts, would generally overestimate the

resources required for contact tracing each case. With the occurrence of pre-symptomatic transmission for SARS-CoV-2 [25,26] and delays from symptoms onset to testing to isolation [27,28], contact tracing would involve the identification of cases over 3–11 days and repeated contacts arising from regular daily activities would imply that the actual contacts made over this period are 20–70% lower than the sum of all the contact episodes recorded independently on each day.

There are some limitations to our study. First, we focused on the network and epidemiological metrics between pairs of contacts. We did not study the changes in clustering on temporal networks and overlay the dynamics of infectiousness profiles on these networks. As such, this limits our ability to make conclusions on the impact of temporal contacts on outbreak size, time to outbreak extinction and herd immunity thresholds. Nevertheless, the current study is a first step in characterising temporal networks. Our 'retention index', $\bar{r}$, quantifies the retention of contacts in temporal networks relative to static and highly dynamic networks. Furthermore, we analysed the type of contact pairs that are likely to be retained and highlighted the implications to control measures. Future studies could extend this metric to account for higher-order network properties. This would allow us to better understand the impact of time-varying contacts on disease transmission and study the feasibility of using simpler static networks or compartmental models.

Second, different devices were used to measure the networks in different studies. They could either detect face-to-face interactions or RFID signals from all directions. As each device has a different calibration, the measured differences between the networks can be an outcome of the data collection process or due to inherent differences in the context setting. As such, in the main analysis, we defined the contact duration and delay between contacts based on the characteristics of each network (Table 7.1). In our sensitivity analysis, we standardise the duration and delay. The changes in $\bar{r}$ for different networks were similar in both analyses. Hence, the impact of the device setting on the overall observed contact patterns was not expected to be significant.

Thirdly, real-life contact typically exists in an open population, and thus, not every contact was captured in these network studies. If these missed contacts were to occur in specific sub-populations this may result in a shift in the proportion of retained contact types. Furthermore, the level of connectivity in missed contacts is unknown. As such, our analysis could over- or underestimate the proportion of 'superspreaders' and 'superspreading events'. However, our findings would remain valid if we assume that the missingness is independent of the level of connectivity and can occur in any subpopulation.

Our analysis highlights the difficulty in identifying highly connected individuals unless real-world contacts are surveyed at high resolution over several days. However, we

did find more consistency in contact patterns among specific settings and social groups. Hence, outbreak control measures that target key settings or at-risk subpopulations are likely to be more effective than targeting specific individuals if currently available data approaches continue to be used. Comparing the dynamics of such real-world temporal networks and corresponding outbreak data would further advance our understanding of the risk of different contacts in practice.

7.7 References

1.   Newman MEJ. Spread of epidemic disease on networks. Phys Rev E. 2002;66: 016128. doi:10.1103/PhysRevE.66.016128

2.   Keeling MJ, Eames KTD. Networks and epidemic models. Journal of The Royal Society Interface. 2005;2: 295–307. doi:10.1098/rsif.2005.0051

3.   Grassly NC, Fraser C. Mathematical models of infectious disease transmission. Nat Rev Microbiol. 2008;6: 477–487. doi:10.1038/nrmicro1845

4.   Keeling MJ, Rohani P. Modelling Infectious Diseases in Humans and Animals. Princeton University Press; 2008.

5.   Eames K, Bansal S, Frost S, Riley S. Six challenges in measuring contact networks for use in modelling. Epidemics. 2015;10: 72–77. doi:10.1016/j.epidem.2014.08.006

6.   Holme P, Saramäki J. Temporal networks. Physics Reports. 2012;519: 97–125. doi:10.1016/j.physrep.2012.03.001

7.   Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. The Lancet Infectious Diseases. 2020;20: 1151–1160. doi:10.1016/S1473-3099(20)30457-6

8.   Tupper P, Boury H, Yerlanov M, Colijn C. Event-specific interventions to minimize COVID-19 transmission | PNAS. Proc Natl Acad Sci U S A. 2020;117: 32038–32045. doi:10.1073/pnas.2019324117

9.   Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious Diseases. 2020;20: 678–688. doi:10.1016/S1473-3099(20)30162-6

10.   Cattuto C, Gauvin L, Panisson A, Quaggiottto M, Tizzoni M, Barrat A, et al. SocioPatterns. [cited 9 Aug 2023]. Available: http://www.sociopatterns.org/datasets/

11.   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;13: 1956. doi:10.1038/s41467-022-29522-y

12.   Kissler SM, Klepac P, Tang M, Conlan AJK, Gog JR. Sparking "The BBC Four Pandemic": Leveraging citizen science and mobile phones to model the spread of disease. bioRxiv; 2020. p. 479154. doi:10.1101/479154

13.   Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. Nature. 2021;594: 408–412. doi:10.1038/s41586-021-03606-z

14.   Read JM, Eames KTD, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. Journal of The Royal Society Interface. 2008;5: 1001–1007. doi:10.1098/rsif.2008.0013

15.   Eames KTD, Read JM, Edmunds WJ. Epidemic prediction and control in weighted networks. Epidemics. 2009;1: 70–76. doi:10.1016/j.epidem.2008.12.001

16.   Mastrandrea R, Fournet J, Barrat A. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. PLOS ONE. 2015;10: e0136497. doi:10.1371/journal.pone.0136497

17.   Fournet J, Barrat A. Contact Patterns among High School Students. PLOS ONE. 2014;9: e107878. doi:10.1371/journal.pone.0107878

18.   Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, et al. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. PLOS ONE. 2013;8: e73970. doi:10.1371/journal.pone.0073970

19.   Génois M, Barrat A. Can co-location be used as a proxy for face-to-face contacts? EPJ Data Sci. 2018;7: 1–18. doi:10.1140/epjds/s13688-018-0140-1

20.   Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science. 2015;3: 326–347. doi:10.1017/nws.2015.10

21.   Siddique H. "Super-spreader" brought coronavirus from Singapore to Sussex via France. The Guardian. 10 Feb 2020. Available: https://www.theguardian.com/world/2020/feb/10/super-spreader-brought-coronavirus-from-singapore-to-sussex-via-france. Accessed 26 Sep 2023.

22.   Yusef D, Hayajneh W, Awad S, Momany S, Khassawneh B, Samrah S, et al. Large Outbreak of Coronavirus Disease among Wedding Attendees, Jordan - Volume 26, Number 9—September 2020 - Emerging Infectious Diseases journal - CDC. [cited 26 Sep 2023]. doi:10.3201/eid2609.201469

23.   Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proceedings of the National Academy of Sciences. 1997;94: 338–342. doi:10.1073/pnas.94.1.338

24.   Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438: 355–359. doi:10.1038/nature04153

25.   Wei WE. Presymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. MMWR Morb Mortal Wkly Rep. 2020;69. doi:10.15585/mmwr.mm6914e1

26.   Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368: eabb6936. doi:10.1126/science.abb6936

27.   Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109. doi:10.1126/science.abc9004

28.   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;20: 911–919. doi:10.1016/S1473-3099(20)30287-5

8   Discussion

This thesis was motivated by a real situational need to understand SARS-CoV-2 epidemiology and transmission dynamics and inform COVID-19 outbreak control measures in Singapore during the acute phase of the pandemic. Furthermore, as part of future pandemic preparedness and outbreak control planning, I have performed a retrospective analysis on the effectiveness of outbreak control measures, explored the impact of pathogen and non-pathogen-related factors on the power of a study, and leveraged technological developments and data collected from the pandemic to understand human contact patterns. A diverse range of statistical and mathematical modelling tools were used as appropriate depending on the type of questions answered.

8.1   Summary of key findings

The findings of the six studies performed in this thesis can be classified into three broad themes as follows:

8.1.1 Surveillance of cases with unknown sources of infection helped in the evaluation of missed infections and the effectiveness of outbreak control measures

Relative role of border restrictions, case finding and contact tracing in controlling SARS-CoV-2 in the presence of undetected transmission: a mathematical modelling study

In Chapter 2, I performed a retrospective analysis of the COVID-19 outbreak in Singapore in 2020 and 2021. Using data on notified imported cases, local linked and unlinked cases (i.e. cases with known and unknown infectors), I developed a renewal equation model which incorporated the observed and latent transmission process. This allowed me to reconstruct the observed outbreak trajectory in Singapore and estimate the number of missed infections. Then, I used a Bayesian adaptive MCMC method to estimate model parameters on (i) the extent of missed imported infections modelled proportionately to the notified imported cases, (ii) the effectiveness of case finding, (iii) the effectiveness of contact tracing and (iv) the impact of overall population-level outbreak controls measures estimated by the effective reproduction number, R_t. Based on the combination of the parameters, the modelled notified cases were fitted to the observed data. I also compared two different outbreak metrics: the proportion of linked cases to all notified cases and the proportion of notified cases to all infections (i.e. infections comprise notified cases and missed infections). The former proportion measure was used in Singapore to proxy the extent of observing an outbreak, but its validity under different outbreak growth phases was unclear. Using a

seroprevalence survey, I validated the modelled estimates of missed COVID-19 infections in Singapore in 2020.

I estimated a decline in the effectiveness of case finding from 64% (95% CrI 27–93%) in January to February 2020, to 42% (95% CrI 12–84%) in March to April 2020. In the same periods, the effectiveness of contact tracing also declined from 89% (95% CrI 72–98%) to 78% (95% CrI 62–93%). This corresponding decline in the effectiveness of contact tracing corroborates with other theoretical studies on the relationship between case finding and contact tracing in achieving outbreak control [1]. More importantly, the study in Chapter 2 was a proof-of-concept that data on the source of infection can provide a more accurate understanding of the effectiveness of targeted control measures.

With individual-level case data, I derived representative estimates on the transmissibility and case ascertainment rates for the COVID-19 pandemic in Singapore. The total observed number of imported cases from July to December 2020 was about three times higher than that reported from March to April 2020. However, the estimated effective reproduction number, R_t, from July to December 2020 was 0.7 (95% CrI 0.5–0.9), lower than R_t from March to April 2020 of 1.0 (95% CrI 0.7–1.3). This was because most of the notified cases were imported cases quarantined in managed institutions, and the resulting onward transmission to the local community was minimal. My estimated R_t was at least 30% lower than the estimates derived by other external research groups as they did not have granular data on the transmission patterns for imported cases in Singapore and only relied on the overall reported cases from public sources [2]. I have also estimated a lower case ascertainment rate of 26% (95% CrI 18–34%) from April to June 2020 compared to the estimated case ascertainment rate of more than 75% in another study conducted by external researchers using publicly available data [3]. My estimated case ascertainment rate was also 2–3 times lower than the proportion of linked cases to all notified cases that was commonly used as a proxy for the extent of observing an outbreak.

Estimates of the case fatality ratio (CFR) and infection fatality ratio (IFR) depend on the severity of the pathogen, the population's characteristics and the healthcare system's capacity. By integrating surveillance data with a mathematical model, I estimated a CFR of 0.8 (95% CrI 0.6–1.0) for the wild-type SARS-CoV-2 outbreak in Singapore in 2020 and 0.5 (95% CrI 0.2–0.8) for the Delta variant outbreak in 2021. The corresponding IFR was 0.3 (95% CrI 0.2–0.5) and 0.2 (95% CrI 0.03–0.3) respectively. Early estimates of the wild-type CFR from other country-specific studies was 1.4 [3,4], about twice the estimated CFR for Singapore in the current study. Furthermore, systematic review and meta-analysis presented Europe-centric IFR estimates of 0.8 in 2020 and only featured a small number of studies from Asia with an IFR of 0.5, and both IFRs were higher than the IFR estimated for Singapore [5].

The study outlines how country-specific data analysis is necessary to inform policymakers on the behaviour of a population and the severity of an outbreak. Furthermore, data collection on the potential source of infection provided additional information on how the cases were identified over time and was useful for evaluating the effectiveness of outbreak control measures.

8.1.2 Real-time analyses of outbreak data could be biased in either direction by several non-pathogen-related factors and adjustment is necessary to interpret changes in pathogen biology

Patterns of infection among travellers to Singapore arriving from mainland China

In Chapter 3, I used a subset of the individual-level data presented in Chapter 2, specifically data on the COVID-19 cases among travellers arriving from mainland China, to estimate the country's outbreak trajectory in December 2022. For Chapter 2, I used a renewal equation model to reconstruct both the observed and unobserved outbreaks driven by imported and local infections in the Singapore population. However, the renewal equation model was not used in Chapter 3 as testing was limited in China during that period and there was insufficient local data from China to reconstruct their outbreak trajectory [6]. As such, I used a statistical method and adjusted the incidence from imported cases by the total travel volume and potential reporting delay to obtain a representative understanding of China's outbreak situation. I also assessed the need for border control measures against travellers arriving from mainland China. In general, local transmission dynamics are influenced by the number of imported infectors and the onward transmission generated by imported and local infectors. While Singapore had a high level of vaccine coverage in late 2022, a surge in imported cases could affect our healthcare capacities.

Following the lifting of COVID-19 restrictions in mainland China in 2022, I estimated that the outbreak peaked on 15 December 2022 and a cumulative attack rate of 31% among travellers by the end of December. However, countries outside China reacted to the surge of cases by imposing pre-departure or on-arrival testing almost 2–3 weeks after 15 December 2022 [7]. My study showed how the surveillance and testing of symptomatic travellers can enhance our understanding of the outbreak trajectory in other countries when information in that locality is limited. This ensures that countries do not impose measures indiscriminately and that measures commensurate with the local risk of transmission.

Serial interval in SARS-CoV-2 B.1.617.2 variant cases

In Chapter 4, I compared the serial intervals among household transmission pairs during the wild-type SARS-CoV-2 outbreak in 2020 and the Delta (i.e. B.1.617.2) variant outbreak in 2021. The differences in the serial intervals served as a proxy to determine if the time from one infection to another (i.e. generational interval) was

reduced. With the emergence of new COVID-19 variants, faster growth in observed case incidence could be attributed to a shorter timescale of infection, an increase in cases in each generation, or both. In both instances, faster outbreak responses are required, but the latter also calls for wider control measures. After controlling for the delays from onset-to-isolation, the type of transmission pairs and prevailing population-level movement restrictions, I detected a reduction of 0.04 day (95%CI -1–1) in the median serial interval of the Delta variant cases compared to its previous wild-type SARS-CoV-2 ancestor. This finding did not show a reduction of more than one-day in the serial interval and suggested that the rapid growth in cases was attributed to the increase in transmissibility of the Delta variant. Given the small sample size of the real-time analysis, this motivated my following chapter to understand how different factors can affect the power to detect changes in pathogen biology.

Detecting changes in generation and serial intervals under varying pathogen biology, contact patterns and outbreak response

In Chapter 5, I used high-resolution contact data previously collected from members of a community in Haslemere, United Kingdom, to simulate disease transmission between pairs of individuals. With the increase in transmission of the Delta variant in 2021, changing speed of outbreak response, social contact patterns and growth rate of the outbreak in different countries, this led to contradictory findings on the changes in the serial intervals for the Delta variant relative to the wild-type SARS-CoV-2; either a 1–2 day reduction in the serial interval [8,9] or no difference [10], similar to the findings in Chapter 4. Using a branching process simulation model, I modelled different scenarios with changes in the pathogen biology (e.g. incubation period, infectiousness profile) and external factors (e.g. delay from onset-to-isolation, exponential growth, frequency of contact) to determine the power to detect a given change in the generation and serial interval for a given sample size.

The generation and serial intervals are a function of the incubation period of the infector and infectee respectively, and the time from symptoms onset to transmission of the infector. Changes to the infectiousness profile, such as the time to peak infectiousness, would change the transmission probabilities over the infectious period non-linearly. As such, a one-day reduction in the time to peak infectiousness would not bring about a corresponding one-day reduction in the generation or serial interval, assuming all factors remain unchanged. Therefore, a simulation framework allowed me to investigate the interplay between each factor. I considered different infectiousness profiles for the Delta variant by assuming different peak infectiousness and time-to-peak infectiousness based on the literature [11–14]. For a decrease of 0–1.4 days in the incubation period of the Delta variant reported in the literature [8,9,15,16], using the simulation framework, I found that a one-day reduction in the serial interval of the Delta variant was unlikely. This provided support to the findings in Chapter 4.

When comparing pathogens with different biological characteristics, I showed how unadjusted exponential outbreak dynamics could accentuate differences in generation interval distribution. This would increase the power to detect a difference and conclude that a non-zero difference exists between the generation intervals of two pathogens (i.e. lower Type II error). The unadjusted dynamics also increase the chance of concluding a difference in the generation intervals even when there is none after adjustment (i.e. higher Type I error).

Besides exponential growth dynamics, I also examined how exposure to multiple household infectors is more likely to reduce the onset-to-isolation and, hence, shorten the generation and serial intervals. Under periods of lockdown, the frequency and duration of household contacts would increase. As such, the transmission dynamics among household members can influence the overall population-level observations on the generation and serial intervals. Hence, it is important to determine if the observed changes in these intervals arise from changes in the pathogen biology, external factors or reporting artefacts arising from exponential growth.

Generation intervals are rarely observed, and serial intervals are often used as a proxy. Under this simulation framework, I provided estimates of (i) the power to detect the theoretical difference in the generation intervals, (ii) the power to detect the observed difference in serial intervals, and (iii) the power to detect the inferred difference in generation intervals. Overall, I showed that a sample size of at least 100 would provide 30–70% power to detect a one-day change in the generation or serial interval, depending on the prevailing delay from onset-to-isolation. The faster the onset-to-isolation, the smaller the variance in the generation interval distributions, thereby increasing the power to detect any differences. These findings help inform future study designs for comparing the changes in pathogen biology when novel variants emerge.

8.1.3 High-resolution contact data helped calibrate outbreak control measures to each contact setting

Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events

In Chapter 6, I studied the contact patterns in high-resolution cruise contact networks — a proxy for large-scale multi-day events. Vaccination and rapid antigen testing can reduce SARS-CoV-2 transmission at large-scale events. However, pilot studies at concerts, football matches, and nightclubs [17] typically do not last more than a day, and studies were not designed to quantify the duration of contact and hence, the risk of infection. Using Bluetooth contact tracing devices, I collected and analysed contact and location records from over 2,000 passengers and crew on each of the four cruise sailings. I estimated that an average passenger had 20 (IQR 10–36) close contacts while an average crew had 10 (IQR 6–18) close contacts. This information was crucial

for cruise operators to plan for sufficient isolation and quarantine facilities and crew duty schedules if there were COVID-19 cases onboard the cruise. Furthermore, this work demonstrated that nearly 80% of the passenger interactions occurred in dining or sports areas, and individuals generally do not wear masks in these settings.

With data on the frequency and duration of contact, and the proportion of time spent in areas where mask-wearing was limited, I simulated SARS-CoV-2 Delta variant transmission over seven days. This simulation framework helped to evaluate the effectiveness of different outbreak control measures and allowed policymakers to consider the possible options for a desired outbreak outcome. The complete network data and the interactions involving various types of individuals increased the realism of the simulation in event settings involving multiple actors (e.g. conferences, business meetings). After accounting for the mask-down interactions, I estimated that mask-wearing interventions would reduce the outbreak size by about 50%. Infection with the Delta variant resulted in high viral load and hence high sensitivity of the PCR and rapid antigen tests during early stages of the infection and this was incorporated within the model [18,19]. As such, I estimated that even with a low vaccine coverage among passengers and crew, implementing regular rapid antigen testing resulted in outbreaks with an average size of less than one new case onboard. When implementing a combination of interventions, different combinations produced similar outcomes. Wearing a mask and no testing strategy generated similar expected outbreak sizes as a strategy with no mask-wearing but once-off PCR testing. Also, a mask-wearing and once-off PCR testing strategy generated similar expected outbreak sizes as a strategy with no mask-wearing but two rapid antigen tests; one before departure and one during midway of the event.

Temporal contact patterns and the implications for predicting superspreaders and planning of targeted outbreak control

Finally, in Chapter 7, I studied the characteristics of real-world temporal networks and their impact on disease transmission and outbreak control. I used the temporal contact data from the previous cruise analysis in Chapter 6, along with temporal contact data collected in other studies conducted in a community in the UK, high schools, hospitals and workplaces in France. As directly transmitted infectious diseases spread through contacts that can change over time, a key question is whether we can reliably identify potential 'superspreaders' for pre-emptive targeted outbreak control measures. In this study, I defined a 'superspreader' as a highly connected person who accounts for the top 80% of contacts in a time step and consistently displays such high connections in at least half the time steps of a study. Furthermore, I defined a 'superspreading event' to be driven by a person who displayed a high number of connections for less than one-quarter of the time steps in a study. Across most of the networks, I estimated that less than 10% of the population was consistently identified to be highly connected and were potential 'superspreaders' if infectious. On the other hand, more than 80% were highly connected for short periods and could drive 'superspreading events' if

infectious. Overall, this suggested an inherent difficulty in reliably identifying superspreaders.

Furthermore, another characteristic of temporal networks is the correlation of contacts over consecutive days or consecutive timesteps. Firstly, I studied the extent of repeated contact over the days. Repeated contacts over the days implied that the total unique number of contacts made by an individual was about 20-70% lower than the contact episodes over a typical contact tracing period. In this study, I showed how high-resolution data can help minimise the extent of overestimating outbreak resources for testing or quarantine of close contacts. As such, contact surveys such as POLYMOD or real-time surveys like CoMix could incorporate data collection on the frequency of close contacts in days to inform outbreak planning [20–22]. Secondly, I developed a new metric to measure contact retention over consecutive time steps. This metric measures the change of contacts in the temporal network relative to a fully static and a fully dynamic network. Also, I normalised the metric by the population size, allowing for a meaningful comparison of the networks across different settings and population sizes. As contact retention increases, the risk of transmission increases. For the cruise and community networks, I estimated a high level of contact retention across the study period, and household members accounted for most of the retained contacts. For the high school, hospital and workplace networks, I estimated a wide distribution in the level of contact retention. In the schools and workplaces, contacts of the same department formed most of the retained contacts. In contrast, contacts from the same and different departments were equally likely to be retained among the hospital contacts. In other words, transmission in a hospital setting is more likely to spread through multiple departments. These differences in contact retention patterns will significantly impact the outbreak control strategy for each setting — to minimise transmission within a subpopulation or minimise introduction altogether in a given setting.

8.2 Strengths

8.2.1 Real-time data analysis during the pandemic

As new SARS-CoV-2 variants of concern (VOC) were detected in late 2020, quantifying changes in the pathogen's epidemiology was necessary to determine if existing outbreak control measures need to be tightened to reduce transmission. Furthermore, these findings were potentially useful for other countries where the VOC had yet to spread. My analysis in Chapter 4 on the comparison of the serial intervals of the wild-type SARS-CoV-2 and the Delta variant identified no significant changes in these intervals after controlling for the time from onset-to-isolation, the type of transmission pairs, and the outbreak control policies in the periods of comparison. As such, the growing outbreak was attributed to higher transmissibility, an important finding that corroborated with studies suggesting higher viral loads in cases infected with the Delta variant [14,19].

Surveillance of imported cases in most countries was reduced over the pandemic after a substantial proportion of the population had been vaccinated, infected or both. In late 2022, China relaxed the country's outbreak control measures, resulting in a surge of COVID-19 cases locally. However, the extent of case ascertainment was unclear, given accounts of the limited testing capacity [6]. Consequently, many countries responded with strict border control policies against incoming travellers from mainland China without assessing the contribution of imported cases to the local outbreak dynamics [7]. Our study in Chapter 3 provided a situational assessment to the Ministry of Health, Singapore, on 18–29 December 2022 regarding the outbreak in mainland China. Based on the incidence among imported cases from China, I found that the outbreak peaked on 15 December 2022. Thus, Singapore did not impose any pre-departure testing requirements on travellers arriving from mainland China and continued to monitor all imported cases and their country of arrival. To our knowledge, there were no other reports from other countries documenting the incidence of COVID-19 among travellers from mainland China during late 2022.

Large-scale multi-day events were interrupted during the pandemic to reduce the risk of superspreading events but inevitably disrupted social and economic activities. Furthermore, the definition of prolonged contact was revised over the pandemic, with the threshold lowered from 30 minutes or more to 15 minutes or more [23,24]. Thus, when resuming large-scale events, we would need to determine the increase in the number of close contacts and the risk of infection between close contacts under such events to account for the changes in the close contact definition. This would ensure sufficient resources were in place to manage acute outbreaks arising from these events. Using digital contact tracing devices, our data collection onboard four pilot cruise sailings in Chapter 6 was not affected by recall bias and provided an invaluable understanding of the number of close contacts made by different types of individuals (e.g. passengers versus crew members). This allowed the cruise operators to plan for sufficient isolation rooms and assess the number of potential individuals required for quarantine should there be a COVID-19 case onboard. Furthermore, data collection on the location of interactions indicated that nearly 80% of the contacts occurred in dining or sports areas, thereby allowing the cruise operators to implement pre-booking and crowd control measures in subsequent cruise sailings.

8.2.2 Explaining counterintuitive outbreak outcomes

Interpreting the outbreak using observed data alone can create a false picture of a controlled outbreak because of delays in reporting and underreporting. Modelling was used in this thesis to highlight the potential pitfalls and the need to question the observed outbreak patterns. In Singapore, the proportion of linked cases (i.e. cases with known source(s) of infection) among all notified cases was commonly reported and served as a proxy for the extent of observing the outbreak. However, in Chapter 2, I showed that the above outbreak metric, based on observed data alone, was a poor

indicator of the proportion of notified cases among all infections during periods of exponential growth or decline in an outbreak. This discrepancy occurs as contact tracing may effectively identify most cases around an index case. However, if the effectiveness of case finding in identifying these index cases is poor, the outbreak can continue to grow uncontrollably in the background. Thus, it is important to consider the latent transmission process when assessing the state of an epidemic. Public health agencies could report (internally) the inferred proportion of missed infections among all infections in addition to reports on the observed cases.

Changes in one epidemiological characteristic may not always lead to a linear change in another characteristic. For example, the serial interval is a function of the incubation period of the infectee and the time from symptoms onset to transmission of the infector. The former is often observed, while the latter depends on the infectiousness profile, which can be proxied by the viral shedding profile. A reduction (increment) of one day in the mean incubation period is expected to reduce (increase) the mean serial interval by one day, assuming all other factors remain unchanged. On the other hand, changes to the viral shedding profiles change the probabilities of transmission non-linearly. Furthermore, the probability of transmission from the time of symptoms onset is also affected by the times from onset-to-isolation. As such, changes to the infectiousness profile do not always result in a corresponding change in the serial intervals. In Chapter 5, I simulated and compared the Delta variant and wild-type SARS-CoV-2 serial intervals. With a reported reduction in the incubation period of the Delta variant by 0–1.4 days in other outbreak studies, I showed this would not reduce the serial interval by more than one day. Sensitivity analysis was performed using different infectiousness profiles (i.e. peak infectiousness, time to peak infectiousness) and different delays from onset-to-isolation of the Delta variant outbreak to ensure the robustness of the finding.

When we observe an increase in imported cases, the country of departure could either be experiencing sustained exponential growth, or the outbreak is about to – or has already – peaked. My findings in Chapter 3 showed that when countries implemented strict pre-departure testing for travellers arriving from mainland China in late 2022, the outbreak was already on the decline despite local reports of sustained increase in hospitalisation arising from delayed adverse infection outcomes [25]. The observed imported cases had varied symptoms onset-to-arrival profiles, and hence, it is important to adjust for these differences to understand the underlying outbreak dynamics to ensure that we are ahead of the outbreak curve.

8.2.3 Access to granular data and appropriate model design to draw insights

Outbreak control policies in a country are dynamic, and the cases reported on public-facing platforms often omit granular details on the events associated with each case. With access to individual-level data and awareness of the operational changes to

outbreak control policies, I could evaluate the delay distributions from infection to arrival, testing or isolation in Chapter 2. The estimates on the effectiveness of the outbreak control measures were thus specific to Singapore. That said, my findings on how the effectiveness of contact tracing relies on case finding corroborated with other theoretical studies [1].

Furthermore, with a strict outbreak control policy, Singapore experienced prolonged periods of low COVID-19 death counts of less than five per month for more than six months in 2020 and 2021. The reported COVID-19 deaths accounted for the test outcomes of all deaths from unknown causes. Hence, the extent of underreported COVID-19 deaths was expected to be low. Studies conducted by other external research groups used the number of reported deaths in Singapore to determine the case ascertainment rate [3]. Due to the low number of reported deaths, the case ascertainment in that study was estimated to be more than 75%. However, my study in Chapter 2 used the number of unlinked cases and incorporated a latent transmission process to assess the extent of underreporting. I estimated a much lower case ascertainment rate of 26% during the partial lockdown in Singapore in 2020. This corroborated with a local behavioural survey showing reduced health-seeking behaviour during the same period [26].

My analysis in Chapter 2 was also able to stratify the cases by their travel history and account for the movement history of travellers. From mid-2020 to mid-2021, travellers were quarantined in dedicated facilities to minimise spillover transmission into the community. This allowed Singapore to resume air travel, and the number of imported cases in the second half of 2020 was more than twice that in the first half of 2020. However, the number of reported non-dormitory cases in the community in the second half of 2020 was about a third compared to the first half of 2020. Hence, when evaluating the effective reproduction number, I could exclude imported cases with no known exposure to the community. Compared to other studies that used the overall reported COVID-19 cases in Singapore [2], I estimated the effective reproduction number to be less than 1 by the end of 2020.

8.2.4 International collaboration

This thesis fostered collaboration between policymakers and modellers in Singapore and the United Kingdom. Decolonisation of research was achieved in two ways. Firstly, core modelling capabilities in the Ministry of Health, Singapore, were enhanced during this study. Country-specific models were developed to capture setting-specific data and outbreak control policies. Furthermore, I uploaded all code and derived data for the modelling studies onto my GitHub public repositories to allow other researchers from other countries to adapt the studies with local data sources.

Secondly, this thesis clarified the key public health questions to be addressed in-house or locally and the required tools. Outsourcing key modelling questions during periods

of public health crisis may be unsustainable as challenges in data sharing, lack of contextual understanding and absence of dedicated analytical capacities create a bottleneck in the decision-making process. Instead, fostering strong relations between external modelling teams helps facilitate exchanges of information and ideas, and generic tools can be developed based on public health agencies' needs during peacetime, which will be discussed in Section 8.3.3.2.

With globalisation and climate change, no country is spared from the threat of emerging infectious diseases. In addition, geographical distance and proximity for disease transmission is greatly reduced with global travel. Thus, building core modelling capacities in a locality and sustaining communications with international partners is increasingly important to ensure timely situational updates.

8.3   Limitations and future work

8.3.1 Enhancement of surveillance systems for data collection

8.3.1.1 Data on asymptomatic rates or exposure histories

By integrating surveillance data with modelling, I identified key challenges in the model inference process arising from the need for independent data sources. These are vital areas to improve in the surveillance data collection process. In Chapter 2, a local unlinked case (i.e. a case with an unknown infector) could acquire the infection from a missed local infector or a missed imported individual. This secondary case was subsequently identified due to case finding measures (e.g. testing of suspect cases). Missed imported infections were modelled proportionately to the notified imported cases. As such, observed unlinked cases were used to infer both the extent of missed imported infections and the effectiveness of case finding. The correlation of parameters resulted in slower convergence of the MCMC chains and increased the number of iterations required to achieve an effective step size of about 5,000. Information from other studies on the extent of asymptomatic infections, such as serological studies or intensive screening of sub-populations, would provide informed priors on the extent of the missed imported infections. Furthermore, stratifying the local unlinked cases by their exposure history to travellers provides additional information for model fitting to determine the potential of acquiring infection from a missed imported or local infector. To achieve this stratification, we need to collect information on the case's occupation either through case interviews or by building data pipelines to extract details from other official manpower and statistics data sources.

8.3.1.2 Data on the risk of infection for a given contact

In Chapter 6, I calibrated the risk of transmission per contact between persons residing in the same cruise cabin based on the probability of Delta transmission in household members over the entire duration of infectiousness from different studies performed by others (i.e. for a given duration of contact over the infectiousness period, the risk of

infection was 20%) [9,27,28]. As such, the study's findings were dependent on this assumption and were not calibrated based on the risk of infection onboard the cruise due to the lack of cruise-related transmission data in Singapore. Given that most of the contacts onboard the cruise were between passengers from different cabins or passengers and crew, future outbreak projection studies should perform sensitivity analyses by calibrating the risk of transmission based on non-household contacts. Data on infection events, the duration and proximity of contact and time of exposure since the symptom onset of the case was used to evaluate the risk of transmission between non-household contacts in one study in the United Kingdom [29]. With extensive digital contact tracing and epidemiological investigations performed in Singapore, a similar analysis could be performed, and the findings would be useful for other disease simulation network models.

8.3.2 Enhancement of model design

8.3.2.1 Incorporating level of protection against infection or symptoms

The model in Chapter 2 was designed to evaluate the relative role of border control, case finding and contact tracing at the early stages of an outbreak. However, it was not intended to estimate the effectiveness of these measures as the population's immunity increased. The level of protection against infection and symptoms, and the changes in the level of infectiousness among vaccinated or previously infected persons were several model parameters to incorporate. Studies showed that neutralising antibody titres correlate with protection against symptomatic and severe SARS-CoV-2 infection (i.e. correlate of protection) [30–32], although findings should be interpreted with caution depending on the type of assays used and the means of calibrating these assays (e.g. relative to the wild-type SARS-CoV-2, relative to severe cases) [33]. Population-wide seroprevalence surveys or the testing of sera samples from healthy individuals (e.g. from blood donation or health screenings) would help establish baseline antibody levels in a (sub)population. The corresponding level of protection against symptomatic or severe SARS-CoV-2 could be estimated before incorporating these outcomes into the model in Chapter 2. On the other hand, we can also estimate the risk of infection given an antibody level against known variants by collecting sera samples and nasopharyngeal swabs in patients with acute respiratory infection (ARI). Antibodies take about 14–20 days to peak post-symptom onset [34]. Thus, collecting sera samples near symptom onset allows us to estimate the baseline antibody titres and correlate them with infection outcomes [32]. These findings can then be extrapolated to model the level of protection in the wider population.

8.3.2.2 Incorporating viral circulation patterns from other pathogens

As countries relaxed their COVID-19 restrictions, such as mask-wearing and physical distancing, this led to the co-circulation of SARS-CoV-2 and other respiratory viruses such as influenza, respiratory syncytial virus and seasonal coronaviruses. Consequently, the circulation of SARS-CoV-2 in Singapore is increasingly out of

phase with influenza. The outbreak modelled in this thesis were specific to SARS-CoV-2 and did not account for the effects of co-circulation of other respiratory viruses. As such, to future-proof these models, we would need to account for the population's behaviour, including potential reduction in social contacts when an individual falls sick, and immunity against SARS-CoV-2 and other non-specific innate immunity generated by other competing viruses [35,36].

8.3.2.3 Ensemble modelling

In this thesis, I developed models for real-time analysis of the epidemiology of SARS-CoV-2 and evaluated the effectiveness of different outbreak control measures. However, integrating these models with other short-term forecasting tools is the next step to enhance our surveillance system. While a seroprevalence survey was used to validate the estimated missed infections in Chapter 2, wastewater testing and comparison with the viral loads against the reported cases can provide a timely assessment of the extent of the case ascertainment [37,38]. Another method involves random testing of individuals regardless of symptoms, and population surveys on health-seeking behaviour can reveal the proportion of symptomatic individuals and the proportion of symptomatic individuals who sought medical attention. When combining this information with other proxies on human contact patterns (e.g. mobility data, transactional data), there is potential to improve the short-term forecasted outcomes to plan for a possible surge in healthcare demands.

8.3.3 Pandemic preparedness planning

8.3.3.1 Development of data infrastructures

In this thesis, I have used COVID-19 as a case study. While future pandemics may not be of the same characteristics as SARS-CoV-2 (i.e. Disease X [39]), the COVID-19 pandemic highlighted several key questions to be answered at the start of an outbreak and the required data to be collected: (i) how do we perform case finding effectively and this depends on the modes of transmission and extent of asymptomatic transmission, (ii) is contract tracing feasible and this is determined by the extent of pre-symptomatic and asymptomatic transmission, (iii) what forms of non-pharmaceutical interventions (NPIs), other than case finding and contact tracing, is required and this is informed by tracking the changes in the effective reproduction number. Most of these questions require additional data on the epidemiology of the pathogen (e.g. incubation period, infectiousness profile, generation or serial interval). Training contact tracers to collect this information and data analysts to process this data from multiple sources is critical. Furthermore, depending on the type of public health questions to be outsourced to external modelling groups, data pipelines and data sharing agreements would need to be in place to facilitate outbreak analytics. Overall, data governance and data engineering considerations should be discussed and developed during peacetime and stress-tested before the next pandemic. Thus, it is essential to list the key questions of interest to policymakers at different stages of

the pandemic and the level of complexity, as this will determine the amount of lead time and resources required.

8.3.3.2 Development of modularised codes and documentation

The network simulation model in Chapter 6 was an extension of the other studies [40,41]. The model allows researchers to input real-world contact data for outbreak simulation and evaluate combinations of interventions for outbreak control. A potential extension of this work is to modularise the codes so that different components can be used for different analyses and to document these codes with easy walkthroughs for interested users.

Besides developing tools for network simulation models, other toolkits could focus on:
- (i)   how to devise testing strategies (e.g. at the borders, in at risk-populations, using a combination of different tests)
- (ii)   how to report the incubation period, generation or serial intervals under varying external factors (e.g. exponential growth phase, early isolation, contact patterns)
- (iii)   how to perform data augmentation in generation or serial interval analyses using pairwise transmission data [1,42,43]. My findings on the Delta variant serial intervals in Chapter 4 were limited to observed serial intervals only, and I did not perform data augmentation to evaluate the potential changes in the generation interval distribution.
- (iv)   how to model the infectiousness profile. I have used a spline function in Chapter 5 and a skewed normal distribution in Chapter 6, but functional forms such as the Hill function [11] and piecewise linear regression [44] could be considered for future model fitting or simulation.
- (v)   how to devise a sampling strategy under varying pathogen characteristics and external factors to detect changes in the pathogen's characteristics, similar to the sampling framework in Chapter 5.

8.4   Concluding remarks

This thesis explored different modelling techniques to investigate changes in the SARS-CoV-2 epidemiology, perform real-time analysis of the outbreak trajectory, and retrospectively evaluate the effectiveness of outbreak control measures. Looking beyond COVID-19, this thesis also served as a reflection on the data types, the modelling techniques and the study frameworks to consolidate, improve, or prepare for future pandemics.

8.5 References

1.   Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al.
     Quantifying SARS-CoV-2 transmission suggests epidemic control with digital
     contact tracing. Science. 2020;368: eabb6936. doi:10.1126/science.abb6936

2.   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;5: 112.
     doi:10.12688/wellcomeopenres.16006.2

3.   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 Medicine. 2020;18: 332. doi:10.1186/s12916-020-
     01790-9

4.   Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates
     of the severity of coronavirus disease 2019: a model-based analysis. The
     Lancet Infectious Diseases. 2020;20: 669–677. doi:10.1016/S1473-
     3099(20)30243-7

5.   Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of
     published research data on COVID-19 infection fatality rates. International
     Journal of Infectious Diseases. 2020;101: 138–148.
     doi:10.1016/j.ijid.2020.09.1464

6.   Bloomberg News. China Covid Cases Surge Infecting 37 Million People a
     Day, According to Estimate - Bloomberg. 23 Dec 2022. Available:
     https://www.bloomberg.com/news/articles/2022-12-23/china-estimates-covid-
     surge-is-infecting-37-million-people-a-day

7.   Reuters. UK, France add to COVID restrictions on Chinese travellers.
     Reuters. 30 Dec 2022. Available: https://www.reuters.com/world/china/covid-
     rules-travellers-china-rolled-out-around-world-2022-12-29/. Accessed 13 Dec
     2023.

8.   Zhang M, Xiao J, Deng A, Zhang Y, Zhuang Y, Hu T, et al. Transmission
     Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2 —
     Guangdong Province, China, May–June 2021. CCDCW. 2021;3: 584–586.
     doi:10.46234/ccdcw2021.148

9.   Kang M, Xin H, Yuan J, Ali ST, Liang Z, Zhang J, et al. Transmission
     dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant
     infections in Guangdong, China, May to June 2021. Eurosurveillance.
     2022;27: 2100815. doi:10.2807/1560-7917.ES.2022.27.10.2100815

10. Ryu S, Kim D, Lim J-S, Ali ST, Cowling BJ. Serial Interval and Transmission
     Dynamics during SARS-CoV-2 Delta Variant Predominance, South Korea -
     Volume 28, Number 2—February 2022 - Emerging Infectious Diseases
     journal - CDC. [cited 10 Jul 2022]. doi:10.3201/eid2802.211774

11.   Ke R, Zitzmann C, Ho DD, Ribeiro RM, Perelson AS. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. Proceedings of the National Academy of Sciences. 2021;118: e2111477118. doi:10.1073/pnas.2111477118

12.   Hakki S, Zhou J, Jonnerby J, Singanayagam A, Barnett JL, Madon KJ, et al. Onset and window of SARS-CoV-2 infectiousness and temporal correlation with symptom onset: a prospective, longitudinal, community cohort study. The Lancet Respiratory Medicine. 2022;10: 1061–1073. doi:10.1016/S2213-2600(22)00226-0

13.   Singanayagam A, Hakki S, Dunning J, Madon KJ, Crone MA, Koycheva A, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. The Lancet Infectious Diseases. 2022;22: 183–195. doi:10.1016/S1473-3099(21)00648-4

14.   Li B, Deng A, Li K, Hu Y, Li Z, Shi Y, et al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Nat Commun. 2022;13: 460. doi:10.1038/s41467-022-28089-y

15.   Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness profile and generation time. Journal of The Royal Society Interface. 2021;18: 20200756. doi:10.1098/rsif.2020.0756

16.   McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open. 2020;10: e039652. doi:10.1136/bmjopen-2020-039652

17.   Environment Modelling Group, Department for Digital, Culture, Media and Sport. EMG and DCMS: Science framework for opening up group events, 16 March 2021. In: GOV.UK [Internet]. 23 Apr 2021 [cited 21 Sep 2021]. Available: https://www.gov.uk/government/publications/emg-and-dcms-science-framework-for-opening-up-group-events-16-march-2021

18.   Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database of Systematic Reviews. 2021 [cited 12 Jun 2021]. doi:10.1002/14651858.CD013705.pub2

19.   Kissler SM, Fauver JR, Mack C, Tai CG, Breban MI, Watkins AE, et al. Viral Dynamics of SARS-CoV-2 Variants in Vaccinated and Unvaccinated Persons. New England Journal of Medicine. 2021;385: 2489–2491. doi:10.1056/NEJMc2102507

20.   Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLOS Computational Biology. 2017;13: e1005697. doi:10.1371/journal.pcbi.1005697

21. Gimma A, Munday JD, Wong KLM, Coletti P, Zandvoort K van, 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 Medicine. 2022;19: e1003907. doi:10.1371/journal.pmed.1003907

22. 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;5: e74. doi:10.1371/journal.pmed.0050074

23. Ministry of Health, Singapore. Revision of Suspect Case Definition for Coronavirus Disease 2019 (COVID-19). 2020. Available: https://www.ams.edu.sg/view-pdf.aspx?file=media%5C5460_fi_558.pdf&ofile=REVISIONOFSUSPECTCASEDEFINITIONFORCORONAVIRUSDISEASE2_C-97A-2020_1.pdf

24. Low Y. Covid-19: MOH updates 'close contact' criteria on website, after man was quarantined following "cumulative exposure" to cases over several days - TODAY. 22 Jun 2021 [cited 15 Dec 2023]. Available: https://www.todayonline.com/singapore/covid-19-moh-updates-close-contact-criteria-man-quarantined-after-exposure-bukit-merah

25. Nidumolu J, Farge E, Goh B, Kasolowsky R, Lewis B. WHO says China releases COVID hospital data after reporting gap. Reuters. 5 Jan 2023. Available: https://www.reuters.com/world/china/who-says-china-reports-218019-new-covid-cases-beginning-2023-2023-01-05/. Accessed 15 Feb 2024.

26. National Centre for Infectious Diseases (NCID), National University of Singapore (NUS), Nanyang Technological University, Singapore (NTU Singapore). NCID, NUS and NTU Studies Highlight the Role of Socio-Behavioural Factors in Managing COVID-19. 21 May 2020 [cited 14 Mar 2021]. Available: http://news.ntu.edu.sg/pages/newsdetail.aspx?URL=http://news.ntu.edu.sg/news/Pages/NR2020_May21.aspx&Guid=221cc9a8-f0cd-40ac-a837-c1caf19897b7&Category=All

27. Ng OT, Koh V, Chiew CJ, Marimuthu K, Thevasagayam NM, Mak TM, et al. Impact of Delta Variant and Vaccination on SARS-CoV-2 Secondary Attack Rate Among Household Close Contacts. The Lancet Regional Health – Western Pacific. 2021;17. doi:10.1016/j.lanwpc.2021.100299

28. Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England Technical briefing 14. 2021 Jun. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/991343/Variants_of_Concern_VOC_Technical_Briefing_14.pdf

29. Ferretti L, Wymant C, Petrie J, Tsallis D, Kendall M, Ledda A, et al. Digital measurement of SARS-CoV-2 transmission risk from 7 million contacts. Nature. 2024;626: 145–150. doi:10.1038/s41586-023-06952-2

30. Cromer D, Steain M, Reynaldi A, Schlub TE, Khan SR, Sasson SC, et al. Predicting vaccine effectiveness against severe COVID-19 over time and against variants: a meta-analysis. Nat Commun. 2023;14: 1633. doi:10.1038/s41467-023-37176-7

31. Cromer D, Steain M, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralising antibody titres as predictors of protection against SARS-CoV-2 variants and the impact of boosting: a meta-analysis. The Lancet Microbe. 2022;3: e52–e61. doi:10.1016/S2666-5247(21)00267-6

32. Nilles EJ, Paulino CT, Aubin M de S, Duke W, Jarolim P, Sanchez IM, et al. Tracking immune correlates of protection for emerging SARS-CoV-2 variants. The Lancet Infectious Diseases. 2023;23: 153–154. doi:10.1016/S1473-3099(23)00001-4

33. Liu K-T, Han Y-J, Wu G-H, Huang K-YA, Huang P-N. Overview of Neutralization Assays and International Standard for Detecting SARS-CoV-2 Neutralizing Antibody. Viruses. 2022;14: 1560. doi:10.3390/v14071560

34. Borremans B, Gamble A, Prager K, Helman SK, McClain AM, Cox C, et al. Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset. Malagón T, Davenport MP, Mina MJ, editors. eLife. 2020;9: e60122. doi:10.7554/eLife.60122

35. Nickbakhsh S, Mair C, Matthews L, Reeve R, Johnson PCD, Thorburn F, et al. Virus–virus interactions impact the population dynamics of influenza and the common cold. Proceedings of the National Academy of Sciences. 2019;116: 27142–27150. doi:10.1073/pnas.1911083116

36. Wu A, Mihaylova VT, Landry ML, Foxman EF. Interference between rhinovirus and influenza A virus: a clinical data analysis and experimental infection study. The Lancet Microbe. 2020;1: e254–e262. doi:10.1016/S2666-5247(20)30114-2

37. Watson LM, Plank MJ, Armstrong BA, Chapman JR, Hewitt J, Morris H, et al. Improving estimates of epidemiological quantities by combining reported cases with wastewater data: a statistical framework with applications to COVID-19 in Aotearoa New Zealand. medRxiv; 2023. p. 2023.08.14.23294060. doi:10.1101/2023.08.14.23294060

38. Carlson SJ, Innes RJ, Howard ZL, Baldwin Z, Butler M, Dalton CB. FluTracking: Weekly online community-based surveillance of influenza-like illness in Australia, 2019 Annual Report. Commun Dis Intell (2018). 2023;47. doi:10.33321/cdi.2023.47.14

39. World Health Organization. WHO R&D Blueprint for Epidemics. [cited 15 Feb 2024]. Available: https://www.who.int/teams/blueprint/who-r-and-d-blueprint-for-epidemics

40. 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;26: 1616–1622. doi:10.1038/s41591-020-1036-8

41. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health. 2020;8: e488–e496. doi:10.1016/S2214-109X(20)30074-7

42. Ferretti L, Ledda A, Wymant C, Zhao L, Ledda V, Abeler-Dörner L, et al. The timing of COVID-19 transmission. 2020 Sep p. 2020.09.04.20188516. doi:10.1101/2020.09.04.20188516

43. Geismar C, Nguyen V, Fragaszy E, Shrotri M, Navaratnam AMD, Beale S, et al. Bayesian reconstruction of SARS-CoV-2 transmissions highlights substantial proportion of negative serial intervals. Epidemics. 2023;44: 100713. doi:10.1016/j.epidem.2023.100713

44. Hellewell J, Russell TW, Matthews R, Severn A, Adam S, Enfield L, et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. BMC Medicine. 2021;19: 106. doi:10.1186/s12916-021-01982-x

A Supplementary Material Ethics 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

Miss Rachael Pung
LSHTM

30 March 2021

Dear Rachael

Submission Title: Modelling COVID-19 transmission dynamics in Singapore

LSHTM Ethics Ref: 25727

Thank you for responding to the Observational Committee Chair's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

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 as revised, 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
Investigator CV	CV Rachael	18/03/2021	V1
Consent form	1. Infectious Diseases Act	18/03/2021	V1
Consent form	2. Cruise Privacy Policy	18/03/2021	v1
Local Approval	1. Infectious Diseases Act	18/03/2021	V1
Protocol / Proposal	Research proposal V4 Ethics	18/03/2021	V1
Investigator CV	Kucharski_CV	21/03/2021	1
Other	GCP_Kucharski	21/03/2021	1
Local Approval	study_1_approval	29/03/2021	V1
Local Approval	study_2_approval	29/03/2021	V1
Covering Letter	Cover Letter_ethics	29/03/2021	V1

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,

Page 1 of 2

APPENDIX A: ETHICS APPROVAL

J Whitworth

Professor Jimmy Whitworth
Chair

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

Improving health worldwide

B   Supplementary Material Chapter 2

Additional File 1: Relative role of border restrictions, case finding and contact tracing in control SARS-CoV-2 in the presence of undetected transmission: a mathematical modelling study

Rachael Pung[1,2]*, Hannah E. Clapham[3], Timothy W. Russell[2], CMMID COVID-19 Working Group[2], Vernon J. Lee[1,3]**, Adam J. Kucharski[2]**

[1]Ministry of Health, Singapore
[2]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine
[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore

*For correspondence rachael.pung@lshtm.ac.uk
**These authors contributed equally to this work.

Table S1 Mathematical notations

Category	Notation	Description
Time	τ or x	Time since infection
	t	Calendar time
	a	Duration since arrival; $t - a$ is time of arrival
	s	Duration since symptoms onset; $t - s$ is the time of symptoms onset
	T_s / T_e	Start / end time of cross-sectional population seroprevalence survey from Sep 7 to Oct 31, 2020
Observed incidence	$i_{cf}(t)$	Unlinked cases isolated at time t
	$i_{ct}(t)$	Linked cases isolated at time t
Probability	$\omega(\tau)$	Probability density function (PDF) of the time from infection in one case to another (i.e. generation interval). Approximated by the serial interval (i.e. time from symptom onset in one case to another). Modelled as a lognormal distribution with mean 5.9 days and standard deviation 2.4 days [1–3].
	$f_a(\tau)$	PDF of arriving to a country τ time since infection, derived from the convolution of $f_s(\tau)$ and $f_{sa}(s - a)$
	$f_s(\tau)$	PDF of symptom onset τ time since infection (i.e. incubation period). Modelled as a lognormal distribution with mean 5.8 days and standard deviation 3.1 days for wild-type SARS-CoV-2 [4] and mean 4 days and standard deviation 0.4 days for Delta variant [5].
	$f_{sa}(s - a)$	PDF of time from arrival to symptoms onset using observed data from symptomatic imported cases.
	$f_{h'}(\tau)$	PDF of an imported case being isolated τ time since infection, derived from the convolution of $f_s(\tau)$ and $f_{sh'}(s - h')$. $f_h(\tau)$ for local cases.
	$f_{sh'}(s - h')$	PDF of time from symptoms onset to isolation in an imported case using observed data from symptomatic imported cases. $f_{sh}(s - h)$ for local cases.
	$F_{h'}(\tau)$	Cumulative probability that an imported case is at large in the community τ time since infection and prior to notification. $F_h(\tau)$ for local cases.
	$f_p(\tau)$	PDF of being seropositive τ time since infection given seroconversion. Assumed serology detection probabilities approach 1 after 30 days from time of infection and no decline in immunity up to 11 months post infection [6]. Sensitivity analysis was performed assuming approximately 40% decline in antibody levels 3 months post infection and about 80% decline by 11 months post infection [7, 8].
	$f_T(t)$	Uniform PDF of being tested on a day between T_s and T_e (both days inclusive)
	p_s	Probability of seroconversion = 0.87 [9].
	P_{nbinom}	PDF for negative binomial distribution
Category	Notation	Description

Unknown parameters to be modelled	ρ	Scale parameter to model the missed imported infections as a factor of the notified imported cases Lognormal(1,1) prior assumed (i.e. the number of missed imported infections is not more than 10 times of notified imported cases in 90% of the time but higher proportions of missed infections is still possible under this sampling framework).
	ϵ_{cf}	Effectiveness of case finding Beta(3.1,3.1) prior assumed (i.e. effectiveness of case finding lies in the range of 10-90% for about 99% of the time to prevent sampler from being stuck at the tail ends of the probability range)
	ϵ_{ct}	Effectiveness of contact tracing Beta(3.1,3.1) prior assumed (i.e. effectiveness of contact tracing lies in the range of 10-90% for about 99% of the time to prevent sampler from being stuck at the tail ends of the probability range)
	R	Reproduction number or the average number of secondary cases generated by a single infectious individual over the course of the entire infectious period (i.e. no truncation of the infectious period due to quarantine or isolation). Analogous to the reproduction number of a missed infection, R_m. Lognormal(1,0.5) prior assumed (i.e. reproduction number is not more than 5 in 90% of the time but higher values are still possible under this sampling framework).
Derived parameters	$\beta(\tau)$	Mean rate at which an infected person infects others (i.e. infectiousness) τ time since infection
	R_n	Reproduction number of a notified case; $R_n \leq R_m$
	R_{eff}	Effective reproduction number; $R_n \leq R_{eff} \leq R_m$
	K	Next-generation matrix
	L	Likelihood function
Modelled incidence	$n_{im}(t)$ / $m_{im}(t)$	Notified / Missed imported infections infected at time t
	$n_c(t)$ / $m_c(t)$	Notified / Missed local infections infected at time t
	$n_{cf}(t)$	Unlinked cases infected at time t (i.e. notified local cases with unknown sources of infection)
	$n_{ct}(t)$	Linked cases infected at time t (i.e. notified local cases with known sources of infection)
	$h_{cf}(t)$	Unlinked cases isolated at time t
	$h_{ct}(t)$	Linked cases isolated at time t

Table S2 Notified and modelled missed imported wild-type SARS-CoV-2 infections in 2020

Observed data (●) and modelled outputs (◆)	Time period					
	Jan 18 – Feb 29, 2020	Mar 1 – Apr 6, 2020	Apr 7 – Jun 18, 2020	Jun 19 – Jul 12, 2020	Jul 13 – Dec 31, 2020	Apr 1 – May 12, 2021
SARS-CoV-2 lineage	Wild-type					Delta variant
● Notified imported cases	29	547	5	82	1,537	843
◆ Missed imported infections	30 (10–60)	260 (88–2,900)	0	50 (10–200)	200 (60–1,000)	200 (50–1,200)
◆ Missed infections per notified imported case	0.9 (0.4–2)	0.5 (0.1–2)	0	0.6 (0.1–3)	0.2 (0.04–0.8)	0.3 (0.05–1)

Table S3 Summary of observed data and modelled outputs (median and 95%CI in parenthesis) by respective time periods in 2020 for wild-type SARS-CoV-2 transmission without using case linkage information for model fitting

Observed data (●) and modelled outputs (◆)	Time period in 2020					
	Overall Jan – Dec	Jan 18 – Feb 29	Mar 1 – Apr 6	Apr 7 – Jun 18	Jun 19 – Jul 12	Jul 13 – Dec 31
● Imported cases						
Isolated for testing on Arrival or quarantined	1,653	0	50	5	78	1,520
Not quarantined	547	29	497	0	4	17
● Local cases (by time of isolation						
Linked	1,505	65	606	610	113	111
Unlinked	864	20	204	420	107	113
◆ Missed cases	15,000 (8,400–38,000)	80 (20–300)	1,900 (600–10,000)	9,300 (5,600–22,500)	1,300 (600–3,100)	1,700 (900–2,900)
◆ Total cases (adjusted by time of infection and missed cases)	17,000 (11,000–41,000)	200 (100–500)	3,200 (1,600–12,000)	10,000 (6,300–24,000)	1,600 (800–3,400)	1,900 (1,100–3,200)
● ICU cases (by time of isolation)	86	13	44	28	1	0
● Deaths (by time of isolation)	22	2	11	9	0	0
◆ Case ICU risk (%)	3.1 (2.2–4.0)	22.6 (14.9–31.5)	3.6 (2.6–4.8)	1.7 (1.1–2.5)	0.4 (0.3–0.5)	0 (0–0)
◆ Infection ICU risk (%)	0.5 (0.2–0.8)	12.5 (5.1–21.5)	1.4 (0.4–2.9)	0.2 (0.07–0.2)	0.06 (0.03–0.1)	0 (0–0)
◆ Case fatality ratio (%)	0.8 (0.6–1.0)	3.7 (2.4–5.1)	1.1 (0.8–1.5)	0.4 (0.3–0.6)	0 (0–0)	0 (0–0)
◆ Infection fatality ratio (%)	0.1 (0.05–0.2)	2.0 (0.8–3.5)	0.4 (0.1–0.9)	0.04 (0.02–0.06)	0 (0–0)	0 (0–0)
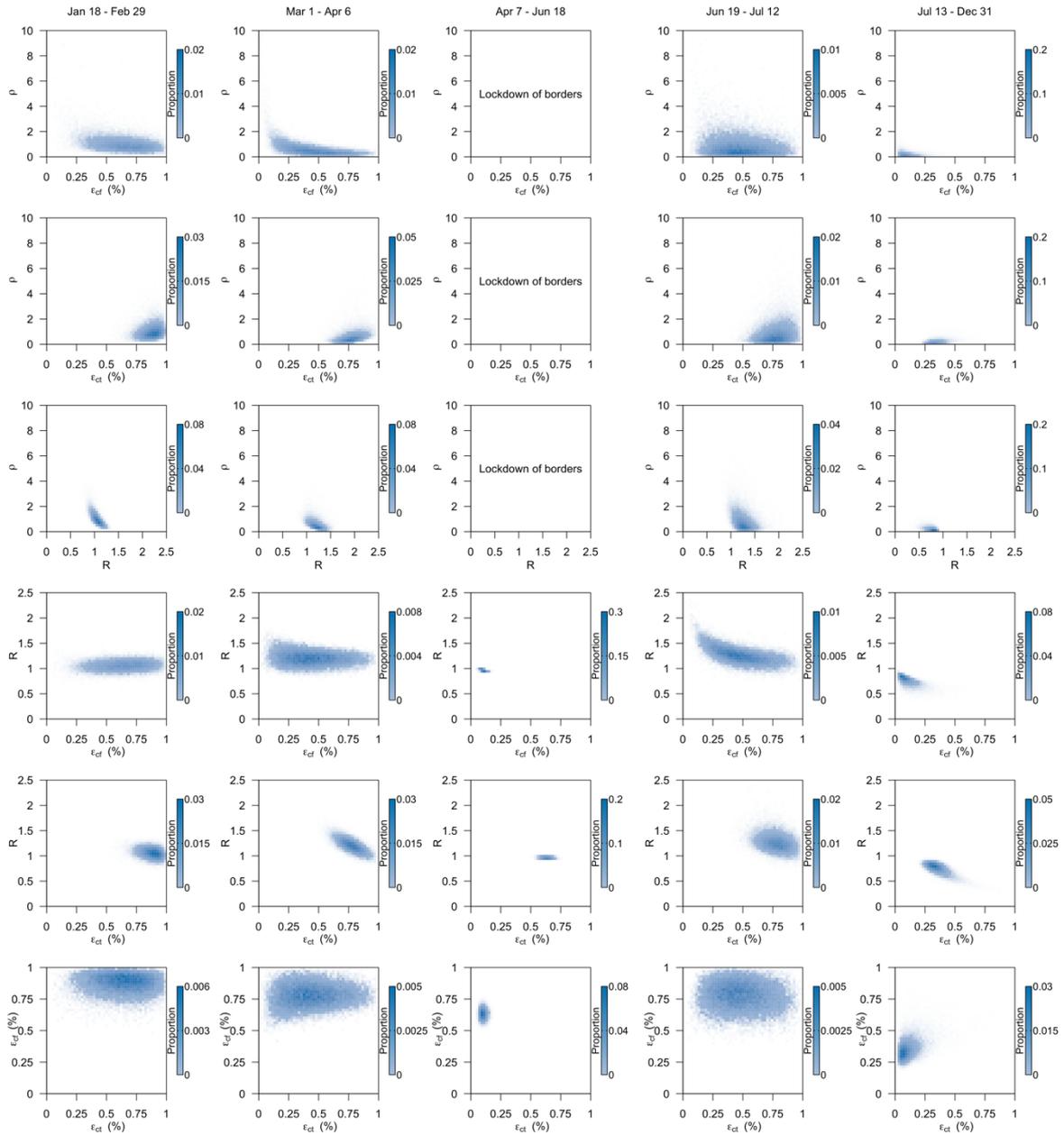
Figure S1 Contour plots to show the correlation between model parameters. Model parameters were discretised along a range of values and the proportion of posterior samples that falls within a set of values in each pair of parameters was evaluated to derive the plots.
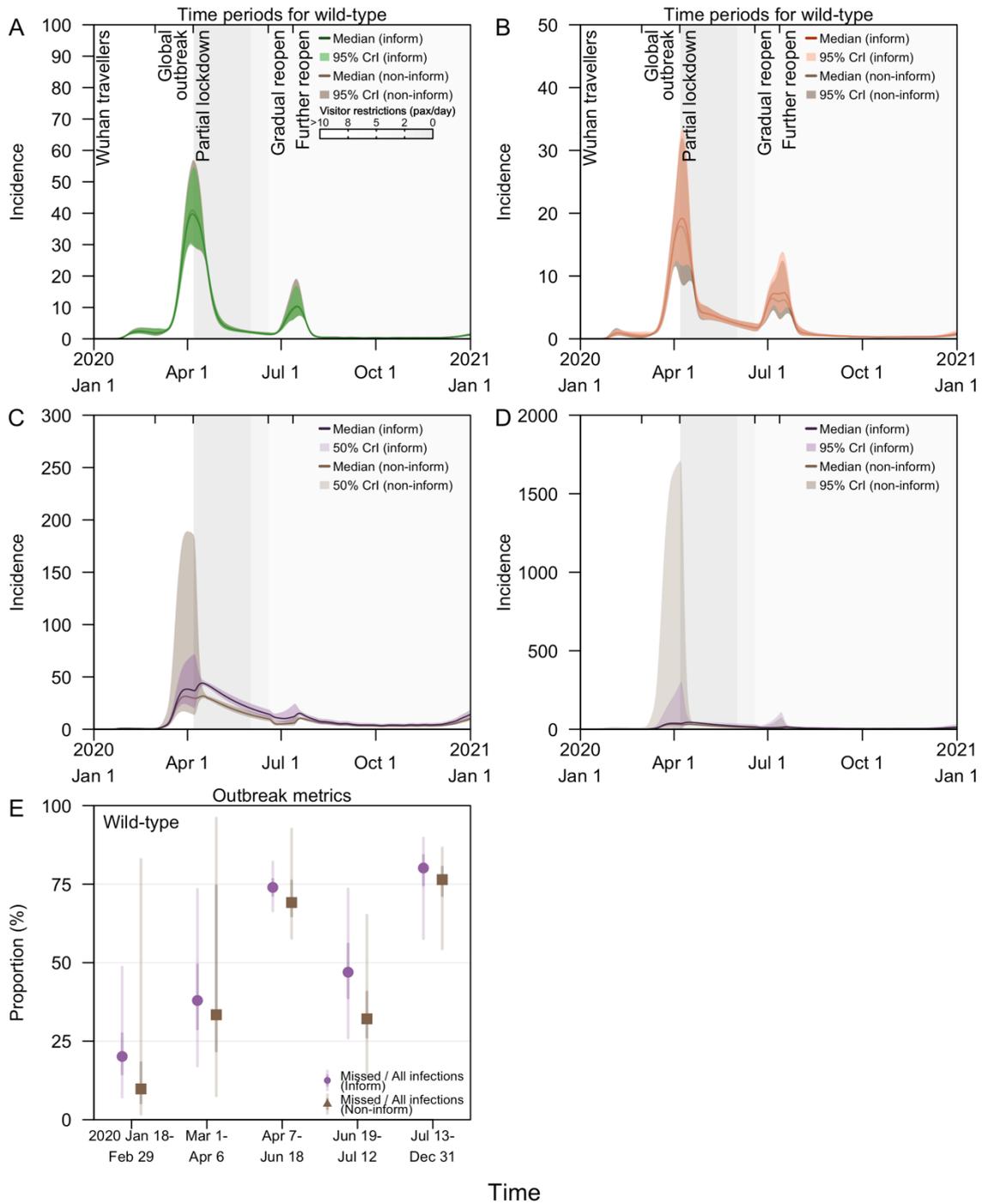
Figure S2 Posterior estimates for model fitted to time series of linked and unlinked SARS-CoV-2 wild type cases in 2020 using informative and non-informative priors. (A) Incidence of linked cases, (B) incidence of unlinked cases, (C) incidence of missed cases with 50% CI, (D) incidence of missed cases with 95%CI, (E) proportion of missed infections to all infections.
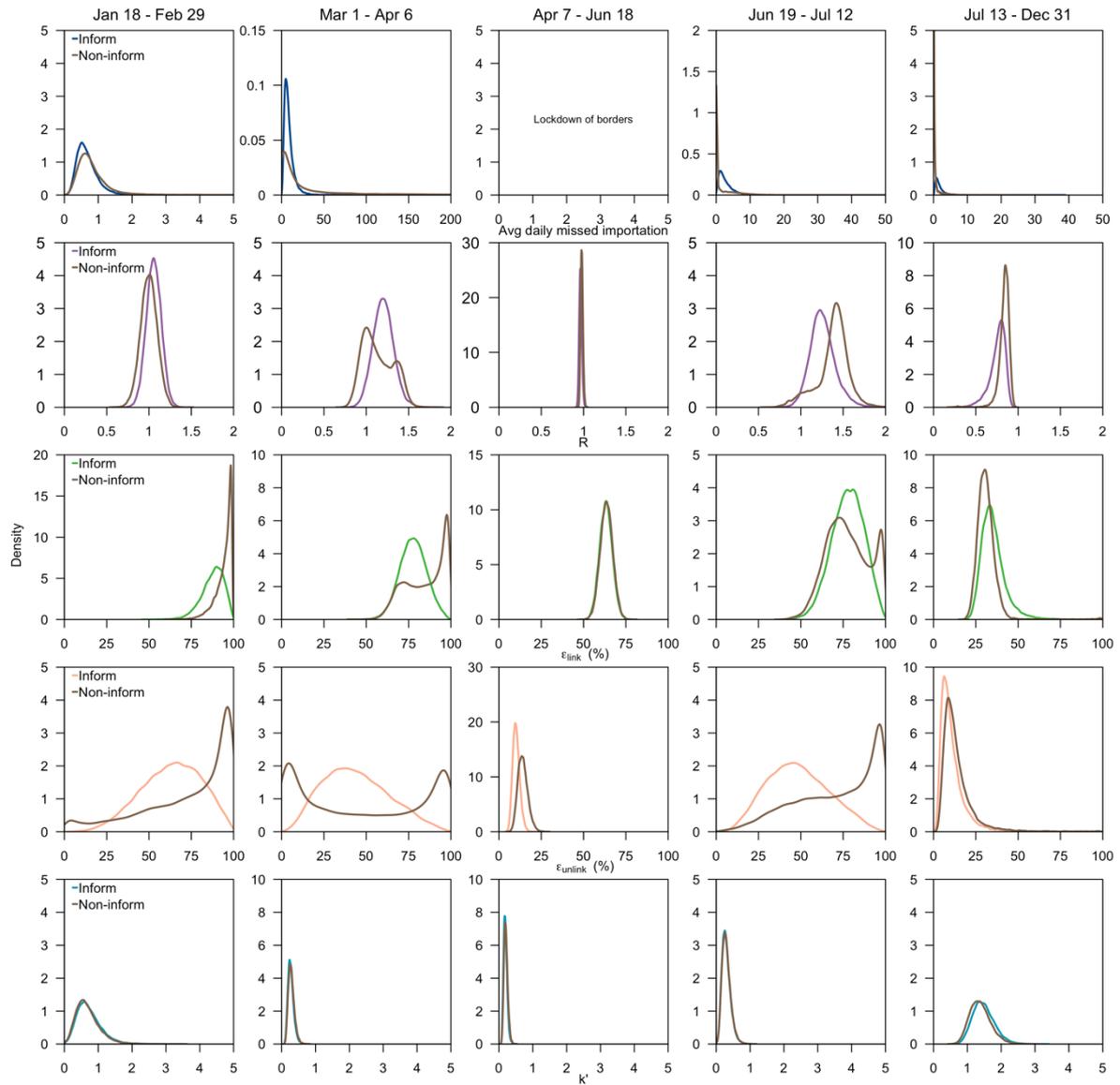
Figure S3 Posterior density of the parameters for model fitted to time series of linked and unlinked SARS-CoV-2 wild type cases in 2020 using informative (blue: average missed imported cases per day, purple: reproduction number, green: effectiveness of contact tracing, orange: effectiveness of contact tracing, turquoise: dispersion parameter) and non-informative priors (brown).
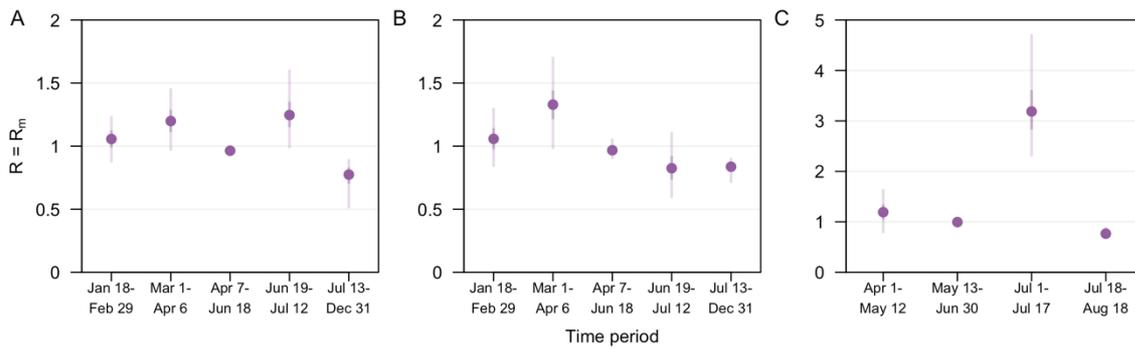
Figure S4 Reproduction number, R of a SARS-CoV-2 (A,B) wild-type in 2020 and (C) Delta variant case in 2021. (A) using linked and unlinked notified cases for modelling fitting and (B,C) using notified cases with no information of the case linkage for model fitting. Posterior median (dot), 50% CI (dark vertical lines and 95% CI (light vertical lines) as shown.
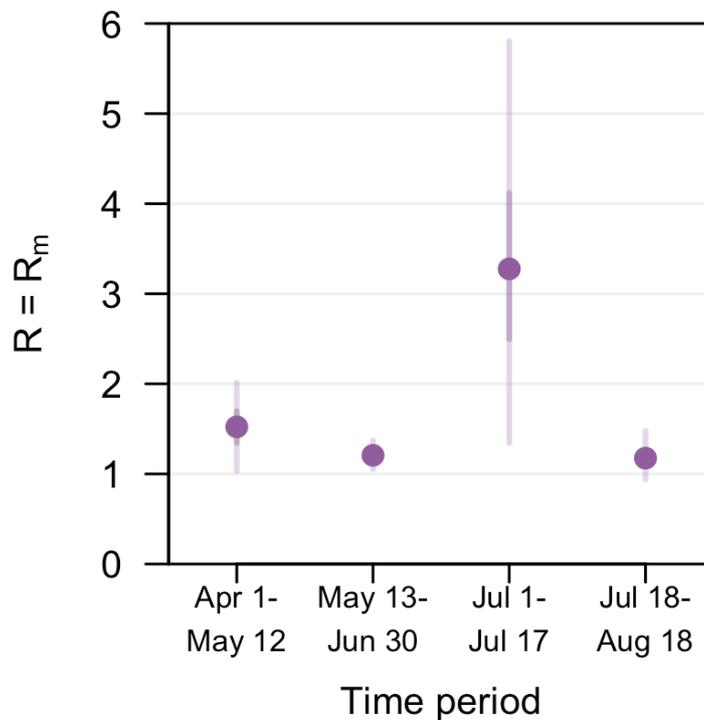
Figure S5 Reproduction number, R of a SARS-CoV-2 Delta variant case in 2021 after adjusting for vaccine coverage and vaccine effectiveness and using notified cases with no information of the case linkage for model fitting. Posterior median (dot), 50% CI (dark vertical lines and 95% CI (light vertical lines) as shown.
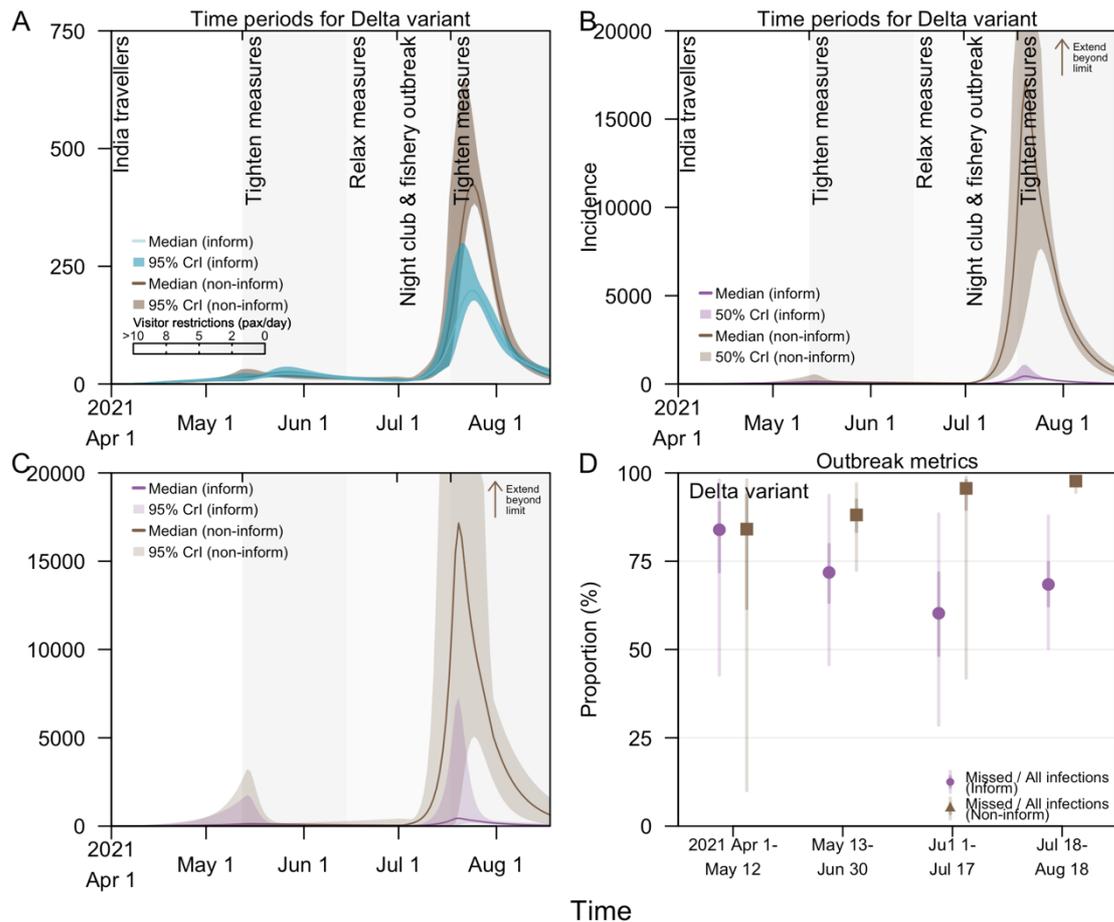
Figure S6 Posterior estimates for model fitted to time series of SARS-CoV-2 Delta variant cases (without accounting for case linkage) in 2021 using informative and non-informative priors. (A) Incidence of cases, (B) incidence of missed cases with 50%CI, (C) incidence of missed cases with 95%CI, (D) proportion of missed infections to all infections.
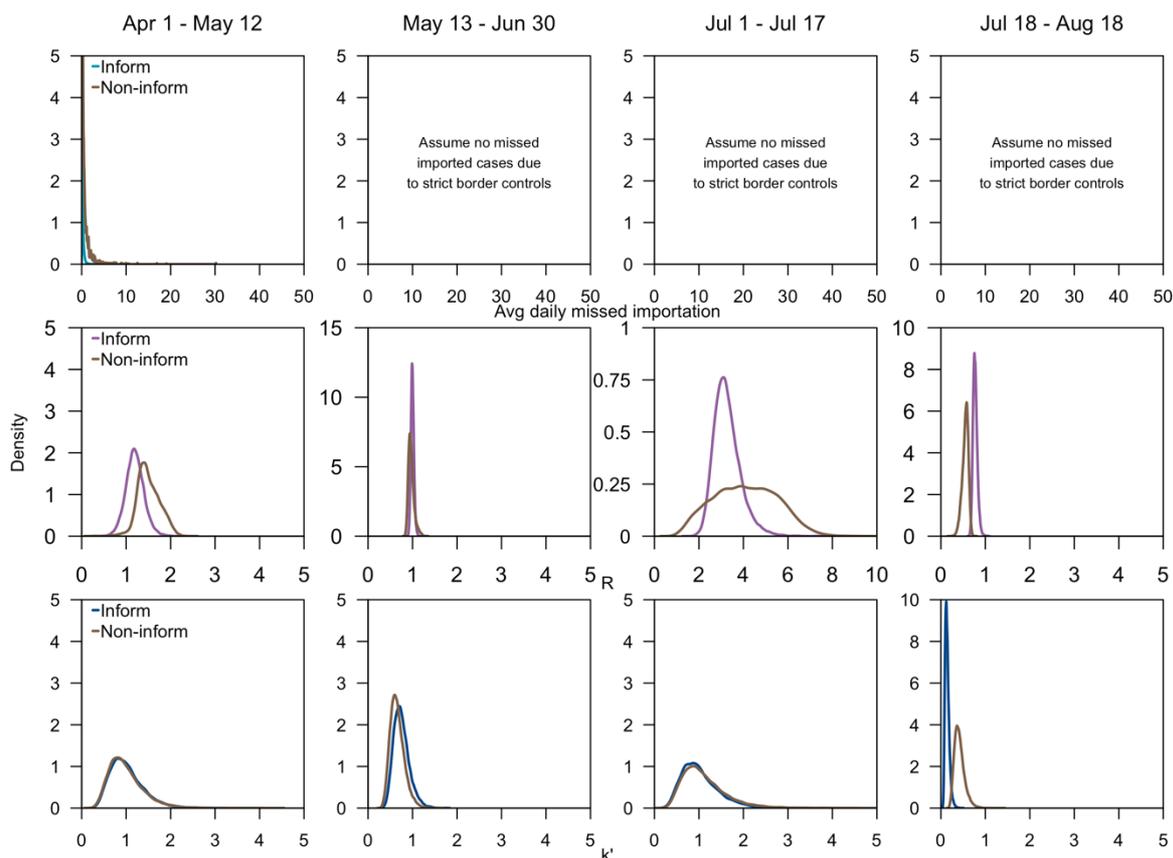
Figure S7 Posterior density of the parameters for model fitted to time series of linked and unlinked SARS-CoV-2 Delta variant cases (without accounting for case linkage) in 2021 using informative (turquoise: average missed imported cases per day, purple: reproduction number, blue: dispersion parameter) and non-informative priors (brown).
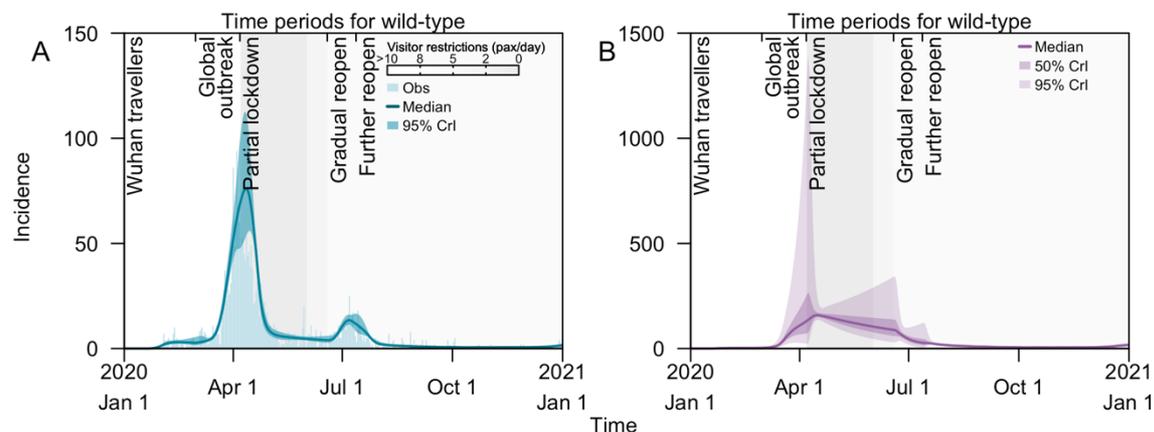
Figure S8 Daily incidence of COVID-19 cases in Singapore arising from SARS-CoV-2 wild-type transmission in 2020, (A) notified local cases and modelled posteriors, (B) modelled posteriors for local missed infections. Grey shaded areas represents periods with movement and visitor restrictions with darker shades signifying reduced number of visitors to each household per day.
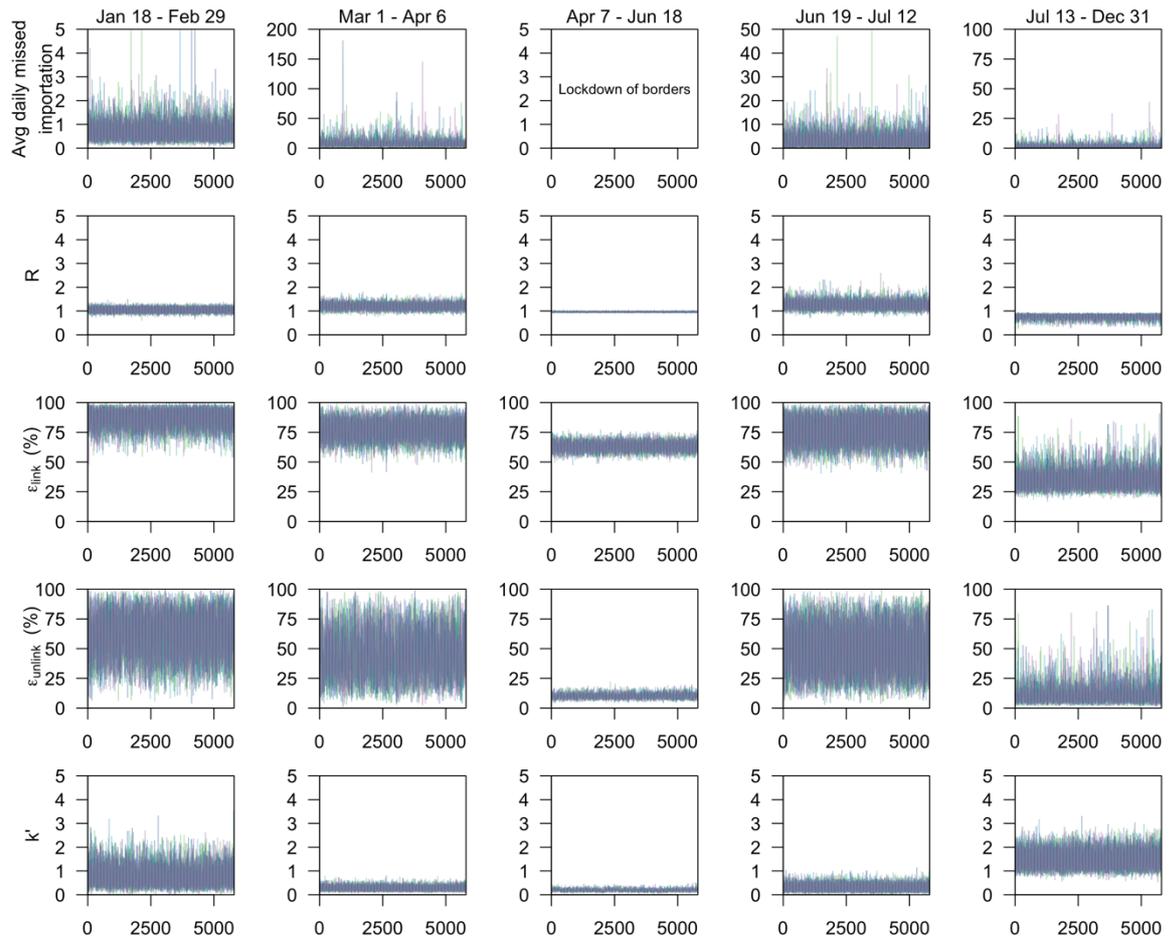
Figure S9 Markov chain Monte Carlo trace plots for parameters modelling wild-type SARS-CoV-2 transmission in 2020. Different lines represent different MCMC chains.
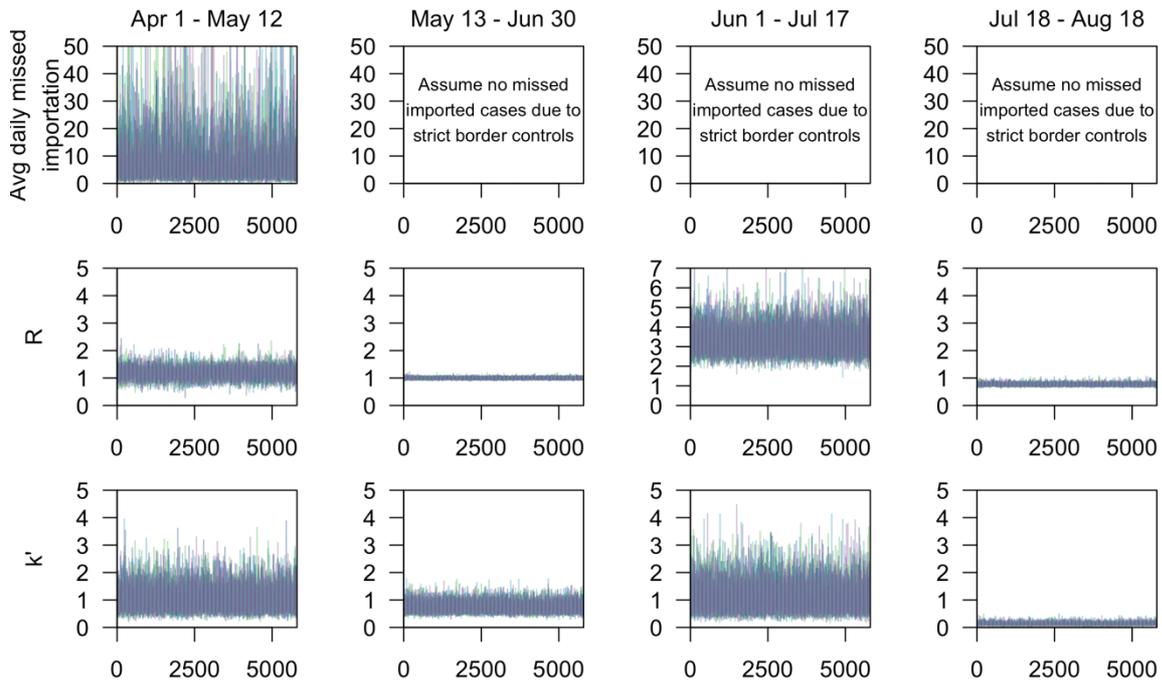
Figure S10 Markov chain Monte Carlo trace plots for parameters modelling SARS-CoV-2 Delta variant transmission in 2021. Different lines represent different MCMC chains.

C   Supplementary Material Chapter 3

Patterns of infection among travellers to Singapore arriving from mainland China

Rachael Pung[1,2]*, Adam J. Kucharski[2], Zheng Jie Marc Ho[1,3], Vernon J. Lee[1,3]

[1] Ministry of Health, Singapore
[2] Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine
[3] Saw Swee Hock School of Public Health, National University of Singapore, Singapore

*For correspondence rachael.pung@lshtm.ac.uk
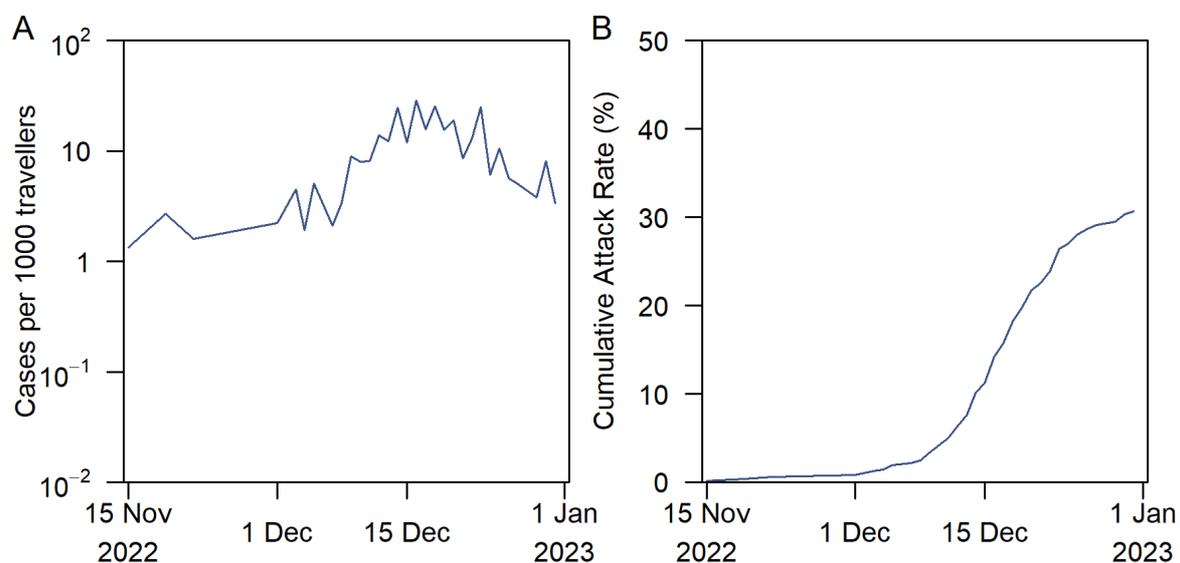
Supplementary Figure 1: COVID-19 outbreak metrics for mainland China. (A) Estimated number of COVID-19 cases per 1000 travellers arriving from mainland China and (B) cumulative attack rate.

D   Supplementary Material Chapter 4

Serial intervals observed in SARS-CoV-2 B.1.617.2 variant cases

Rachael Pung[1,2], Tze Minn Mak[3], CMMID COVID-19 working group*, Adam J Kucharski[2], Vernon J. Lee[1,4]

[1] Ministry of Health, Singapore
[2] Centre for the Mathematical Modelling of infectious Diseases, and Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London
[3] National Public Health Laboratory, National Centre for Infectious Diseases, Singapore
[4] Saw Swee Hock School of Public Health, National University of Singapore, Singapore
*Listed at the end of supplementary
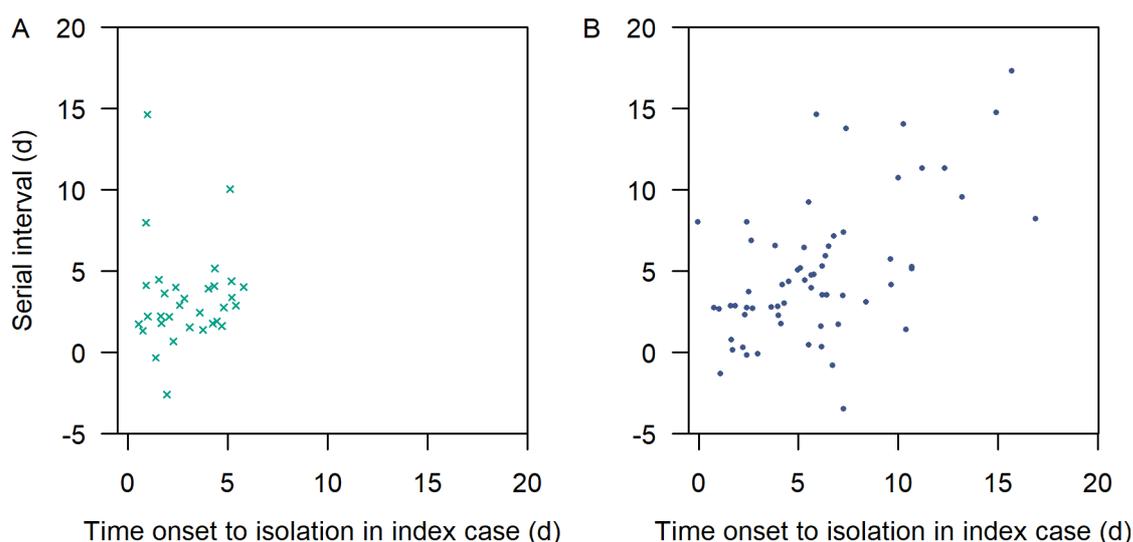
Supplementary Figure 1 Time from onset to isolation in primary case against the serial interval in household pairs (a) in recent B.1.617.2 cases, (b) in cases identified prior to the partial lockdown in Apr 7, 2020.
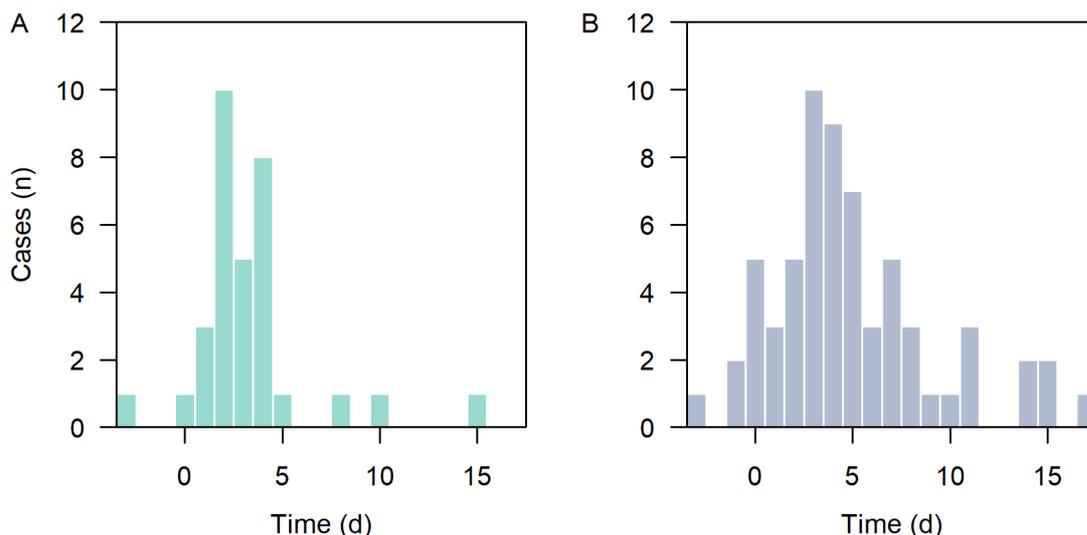
Supplementary Figure 2 Serial interval of household transmission pairs. (a) B.1.617.2 cases, (b) in cases identified prior to the partial lockdown in Apr 7, 2020 without adjusting for time from onset to isolation in primary case.

Supplementary Table 1 Descriptive statistic of the serial interval distributions of B.1.617.2 cases and of the sampled cases prior to the partial lockdown in Apr 7, 2020.

Descriptive statistic	B.1.617.2 cases	Cases prior to Apr 7, 2020* (95% CI)	Difference (95% CI)^
Mean	3.3	3 (2.3–3.8)	0.23 (-0.54–0.96)
Median	3	3 (2–4)	0.036 (-1–1)
Mode	2	2.7 (-1–4)	-0.69 (-2–3)

* 32 out of 63 transmission pairs were sampled and a skewed normal distribution was fitted to obtain the mean, median and mode. This process was repeated 1000 times to obtain the mean and 95%CI of each descriptive statistic as displayed.
^ Derived by taking the observed statistic in the B.1.617.2 cases minus the sampled statistic.

CMMID COVID-19 working group

E   Supplementary Material Chapter 5

Detecting changes in generation and serial intervals under varying pathogen biology, contact patterns and outbreak response

Rachael Pung[1,2]*, Timothy W. Russell[2], Adam J. Kucharski[2]

[1]Ministry of Health, Singapore
[2]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine

*Corresponding author: rachael.pung@lshtm.ac.uk

Supplementary Figure 1 Contact sequence at different times of day. Each line represents a pair of individuals with recorded contact in a five-minute time interval (blue bars). Data was not collected from 2300-0700 hours.
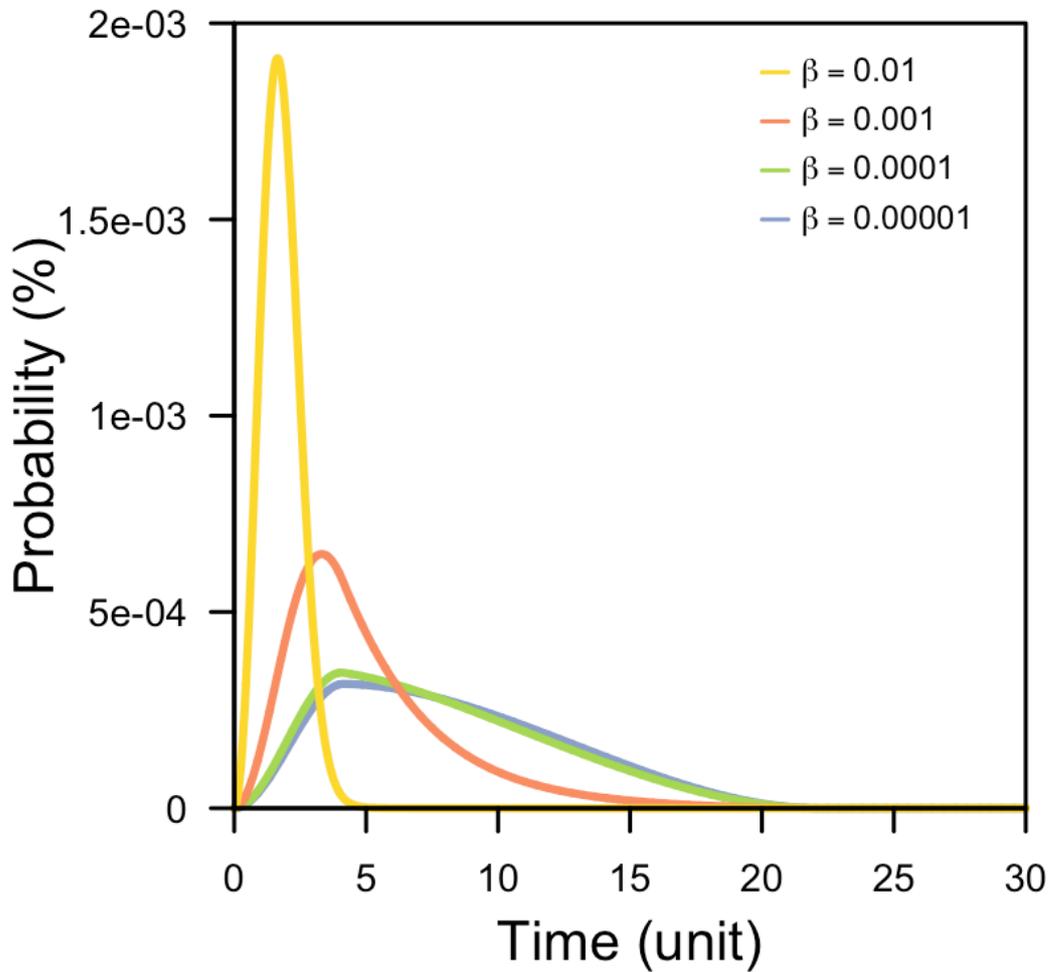
Supplementary Figure 2 Probability of infection at different timestep. Peak infectiousness profile varied based on β scale factor which influences the time of peak probability of infection. At low values of β, the distribution of probabilities is approximately similar over time.
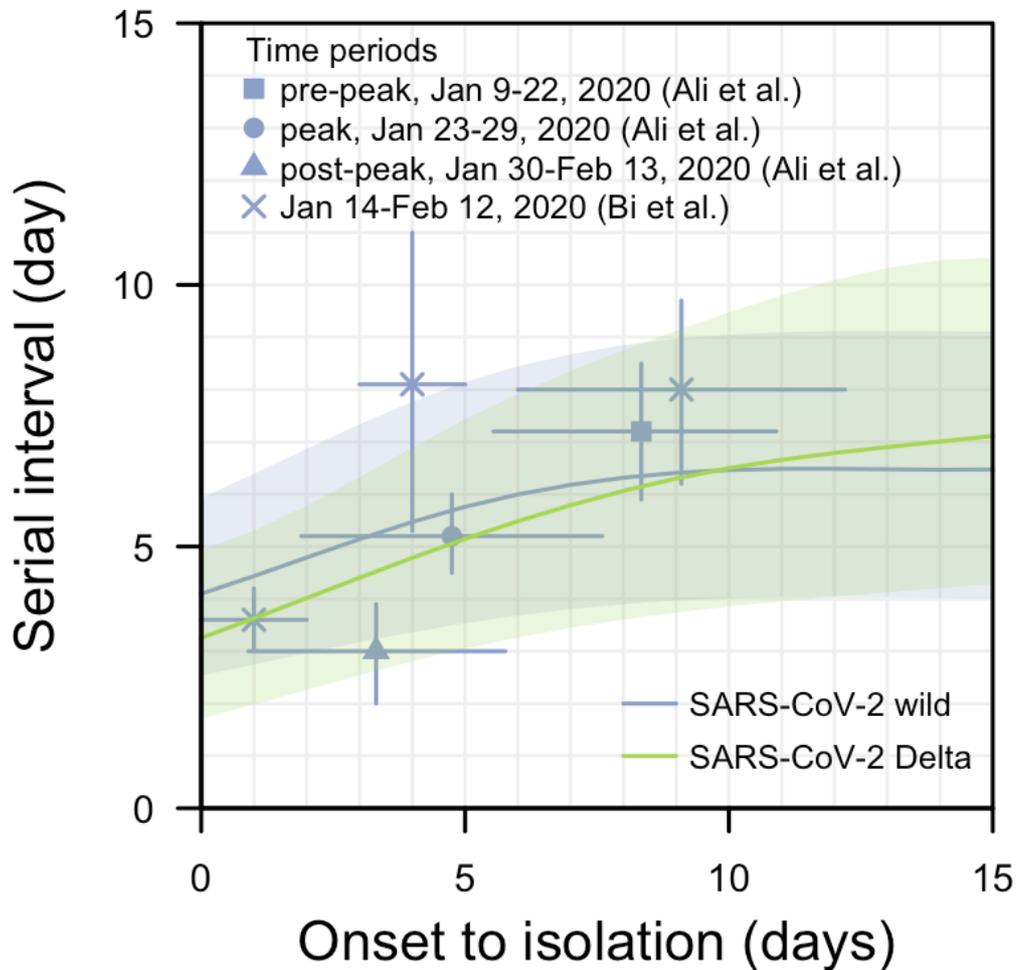
Supplementary Figure 3. Modelled serial interval for varying delay in case onset-to-isolation in SARS-CoV-2 wild type and Delta variant with median (lines) and interquartile range (shaded regions). We assumed higher peak infectiousness for the Delta variant and the attack rate for the Delta was twice that of the wild type in the absence of isolation. Observed serial intervals from published studies [1,2] as shown in points (mean) with lines (95% CI).
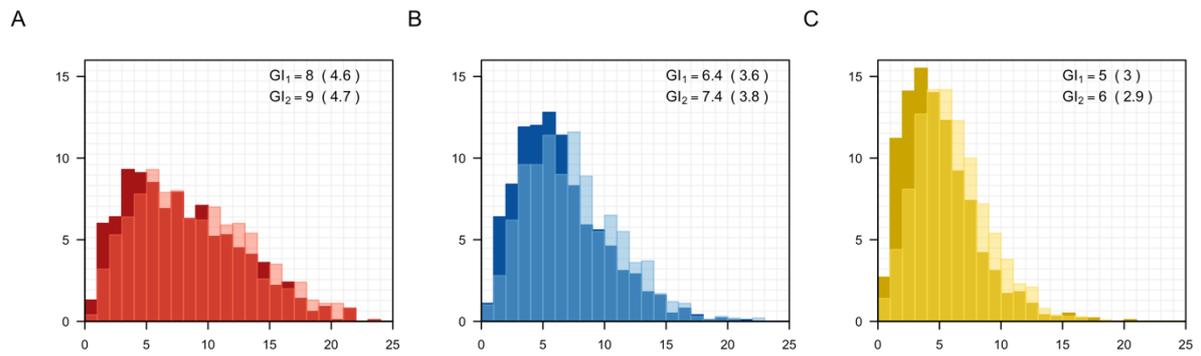
Supplementary Figure 4 Histogram and mean (sd) of generation intervals in reference (dark, GI_1) and alternative (light, GI_2) pathogen under (A) no isolation, (B) case isolation on average 8 days post symptoms onset, (C) case isolation on average 4 days post symptoms onset.
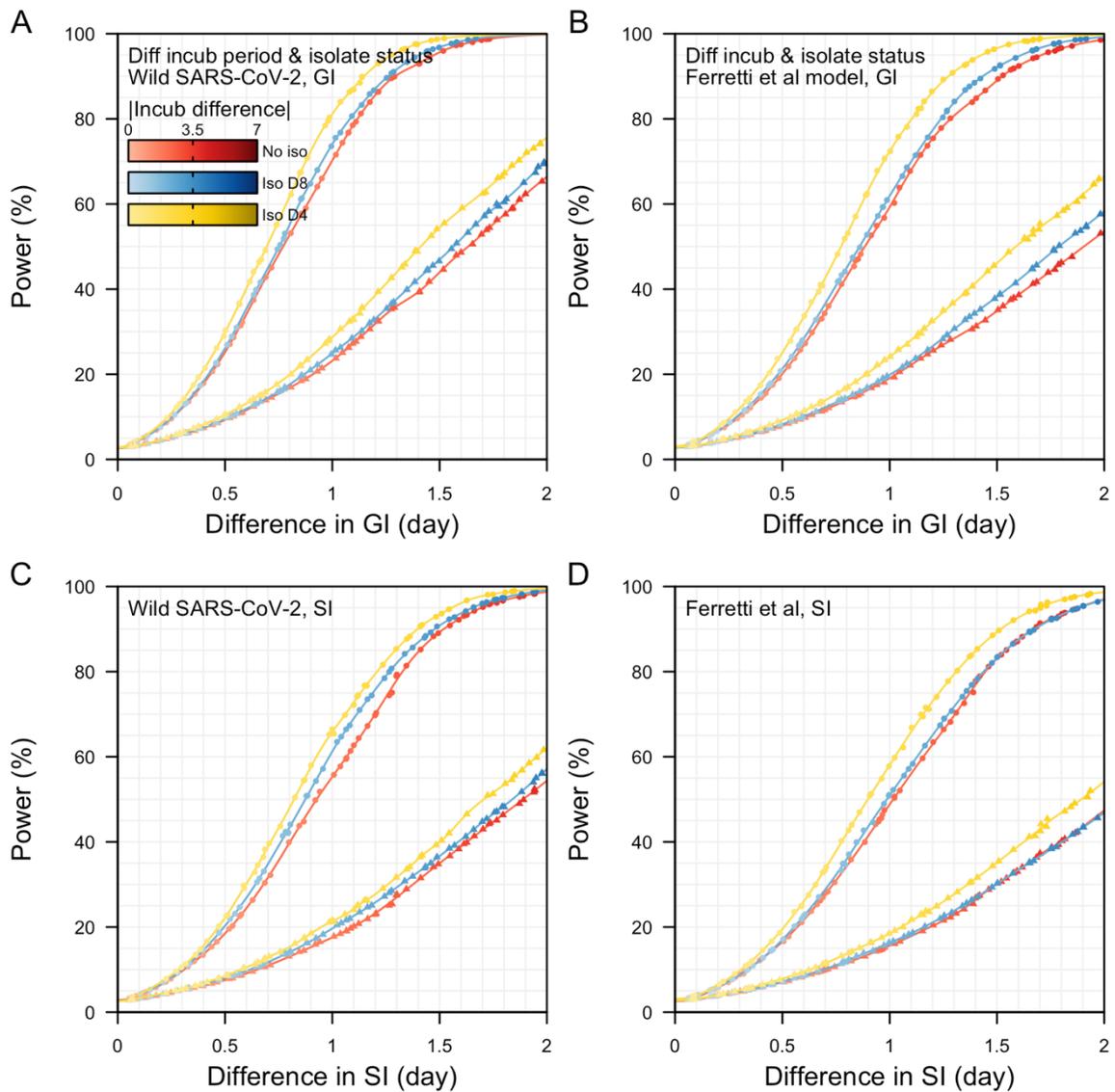
Supplementary Figure 5 Power to detect differences in generation (GI) and serial (SI) intervals between reference and emerging pathogen. (A,C) Different incubation period between reference and alternative pathogen under same symptoms onset-to-isolation status of either no isolation, mean symptoms onset-to-isolation is 8 days, or 4 days using spline model; (B,D) similar to (A,C) but using skew logistic model by Ferretti et al.
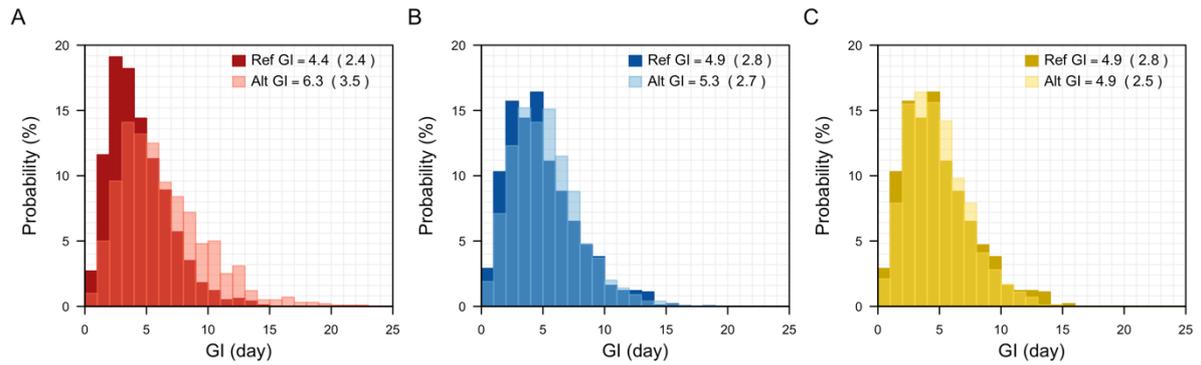
Supplementary Figure 6 Generation intervals, GI, without adjusting for bias introduced by different epidemic dynamics for refence pathogen with a one-day shorter incubation period and longer shedding duration as compared to alternative pathogen. (A) exponential growth in reference pathogen but exponential decline in alternative pathogen, (B) constant growth in both reference and alternative pathogen, (C) constant growth in reference pathogen but exponential growth in alternative pathogen.

Supplementary Table 1 Differences in mean incubation period of reference (Delta-like) and alternative (wild type-like) pathogen for a one day difference in mean generation interval when sample size is 100. Power to detect this difference in incubation period as shown in brackets. Peak infectiousness of Delta-like reference pathogen was scaled by β with values of 0.0005, 0.002, 0.006. The corresponding probability of infection was 20%, 50% and 80% when the mean incubation period was 4 days and peak infectiousness coincided with symptoms onset. Peak infectiousness of wild type SARS-CoV-2-like alternative pathogen is scaled by β of 0.0005. Duration of infectiousness after the peak infectiousness is 8 days shorter in the reference pathogen.

Onset-to-isolation		4 days post onset on average	No isolation
Probability of infection of Delta-like reference pathogen	20%	1.9 (71%)	5.0 (39%)
	50%	1.3 (73%)	2.9 (45%)
	80%	0.2 (85%)	0.0 (64%)

Supplementary Table 2 Differences in mean incubation period of reference (Delta-like) and alternative (wild type-like) pathogen for a one day difference in mean generation interval when sample size is 100. Power to detect this difference in incubation period as shown in brackets. Peak infectiousness of Delta-like reference pathogen was scaled by β with values of 0.0005, 0.002, 0.006. The corresponding probability of infection was 20%, 50% and 80% when the mean incubation period was 4 days and peak infectiousness occurs between 0-2 days prior to symptoms onset. Peak infectiousness of wild type SARS-CoV-2-like alternative pathogen is scaled by β of 0.0005. Duration of infectiousness after the peak infectiousness is 8 days shorter in the reference pathogen. Negative difference indicates that the incubation period of reference pathogen is larger than alternative pathogen.

Onset-to-isolation		4 days post onset on average	No isolation
Probability of infection of Delta-like reference pathogen	20%	1.5 (72%)	4.9 (35%)
	50%	1.3 (74%)	2.6 (38%)
	80%	0.4 (81%)	-1.1 (56%)

Elaborations on Equation 5 in main text

The relationship of the generation and serial interval can be expressed as follow [3,4]:

$$S_{ij} = P_{ij} + I_j$$
$$G_{ij} = P_{ij} + I_i$$
$$G_{ij} = S_{ij} + I_i - I_j$$

where S_{ij} is the serial interval between infector i and infectee j, P_{ij} is the onset-to-transmission, I_i and I_j are the incubation period of infector i and infectee j and G_{ij} is the generation interval.

The variance of the serial interval can be expressed as:

$$Var(S) = Cov(S, S)$$
$$= Cov(P_{ij} + I_j, P_{ij} + I_j)$$
$$= Cov(P_{ij}, P_{ij}) + 2Cov(P_{ij}, I_j) + Cov(I_j, I_j)$$

Assuming that the onset-to-transmission in the infector is independent with the incubation period of the infectee (i.e. $Cov(P_{ij}, I_j) = 0$), thus,

$$Var(S) = Var(P_{ij}) + Var(I_j)$$

The variance of the generation interval can be expressed as:

$$Var(G) = Cov(G, G)$$
$$= Cov(P_{ij} + I_i, P_{ij} + I_i)$$
$$= Cov(P_{ij}, P_{ij}) + 2Cov(P_{ij}, I_i) + Cov(I_i, I_i)$$
$$= Var(P_{ij}) + 2Cov(P_{ij}, I_i) + Var(I_i)$$

Assuming that the incubation period distribution of the infector and infectee is independent and identically distributed, thus

$$Var(I_i) = Var(I_j)$$

and

$$Var(G) = Var(S) + 2Cov(P_{ij}, I_{ij})$$

References

1.   Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369: 1106–1109. doi:10.1126/science.abc9004

2.   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;20: 911–919. doi:10.1016/S1473-3099(20)30287-5

3.   Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness profile and generation time. Journal of The Royal Society Interface. 2021;18: 20200756. doi:10.1098/rsif.2020.0756

4.   Svensson Å. A note on generation times in epidemic models. Mathematical Biosciences. 2007;208: 300–311. doi:10.1016/j.mbs.2006.10.010

F   Supplementary Material Chapter 6

Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events

Rachael Pung*, Josh A Firth, Lewis G Spurgin, Singapore CruiseSafe working group, CMMID COVID-19 working group, Vernon J Lee, Adam J Kucharski

*Correspondence to: rachael.pung@lshtm.ac.uk
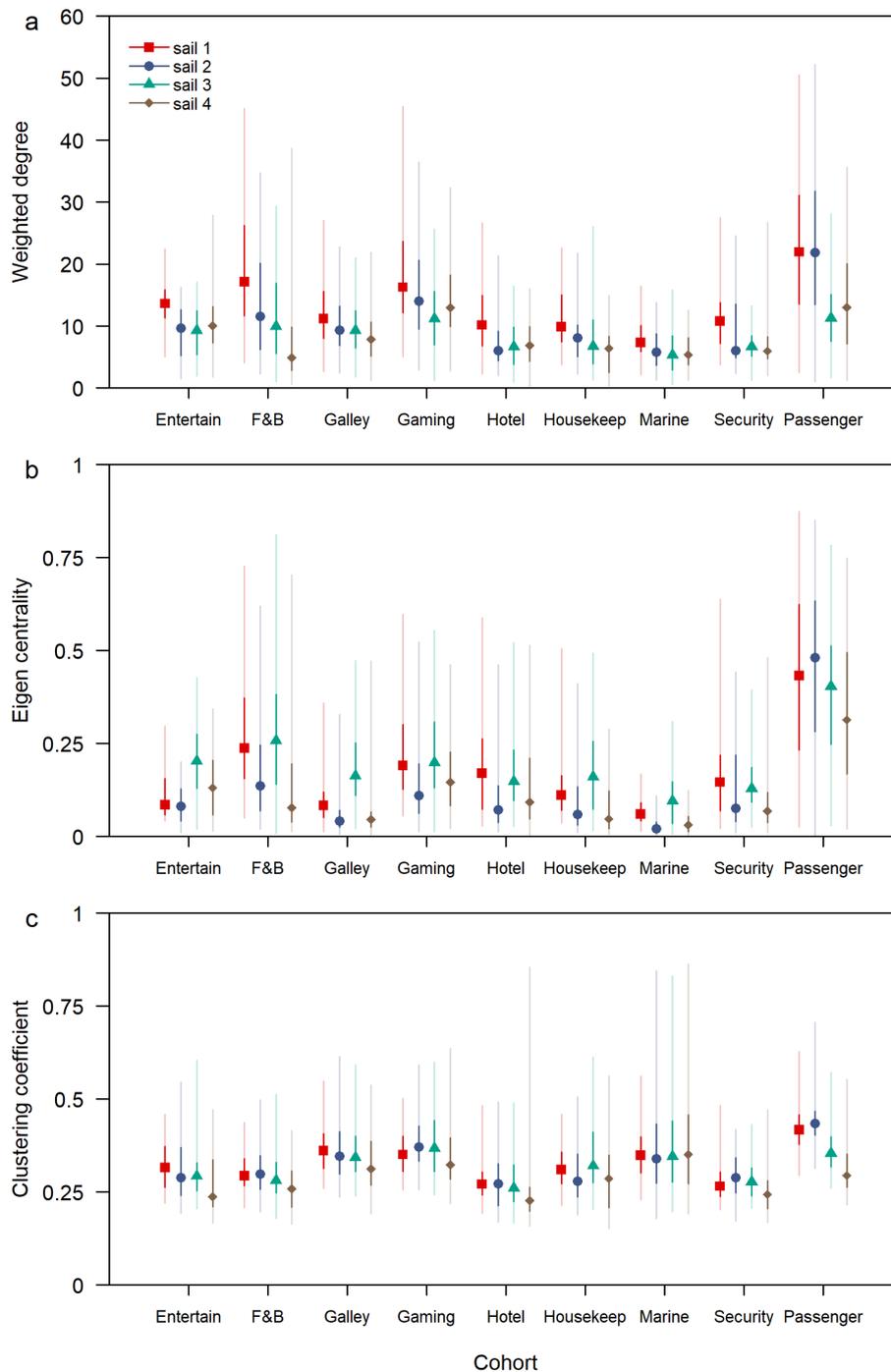
Supplementary Figure 1 Social network analysis over four cruise sailings. (a) Weighted degree, (b) eigenvector centrality, (c) clustering coefficient of crew and passengers of each sailing. Colours represent the cruise departure date and the median (shapes), 50% (dark lines) and 95% intervals (light lines) network property measures from 5,216 passenger and 4,197 crew across 4 sailings are shown. Weights were assigned based on exponent transformation of the mean daily cumulative duration of interaction between two individuals (see Methods)
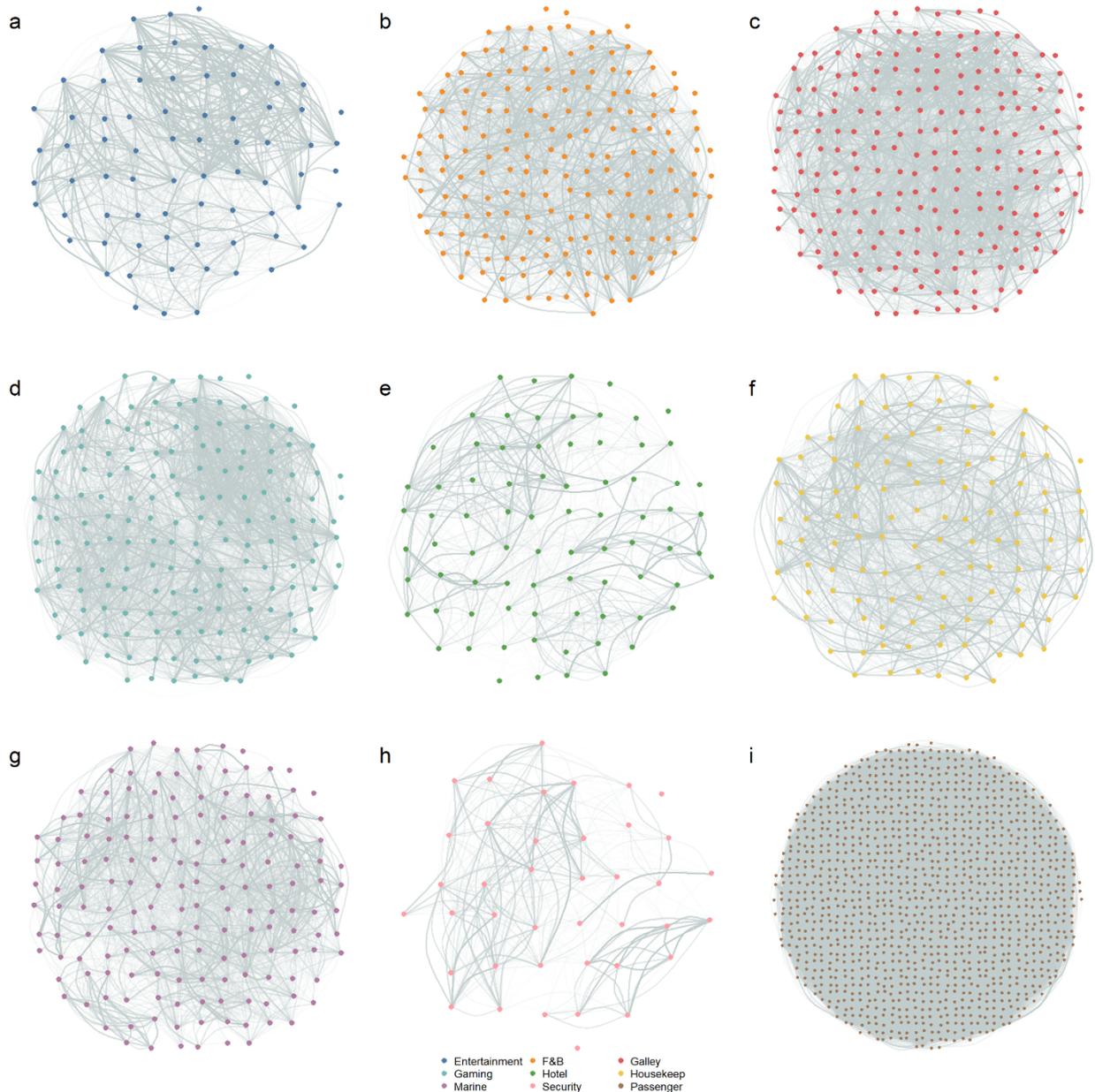
Supplementary Figure 2 Static intra-cohort contacts throughout the entire sailing, with crew from entertainment (a), F&B (b), galley (c), gaming (d), hotel services (e), housekeeping (f), marine (g), security and surveillance (h) departments and passengers (i). In addition, there were 77,107 unique pairs of crew contacts from different cohorts and 70,360 unique pairs of crew and passenger contacts but these links were not represented in this figure. Edge width and colour intensity of the edges correspond to the weights of a contact with the highest colour intensity as shown in the legend. Edge weights are a function of the proportion of days with recorded contact over a three-day sail period and the exponent transformation of the mean daily cumulative contact duration between two individuals.
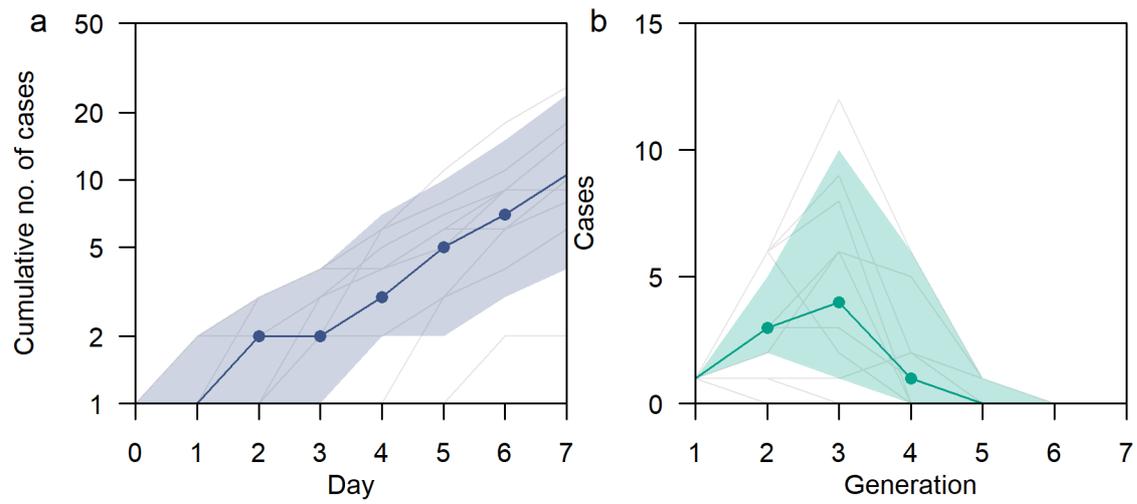
Supplementary Figure 3 (a) Cumulative cases by day of exposure and (b) number of cases in respective generations in the baseline scenario. Median (dots) and 95% intervals (shaded region) and outbreak trajectory for 10 selected simulations (grey lines) are shown.
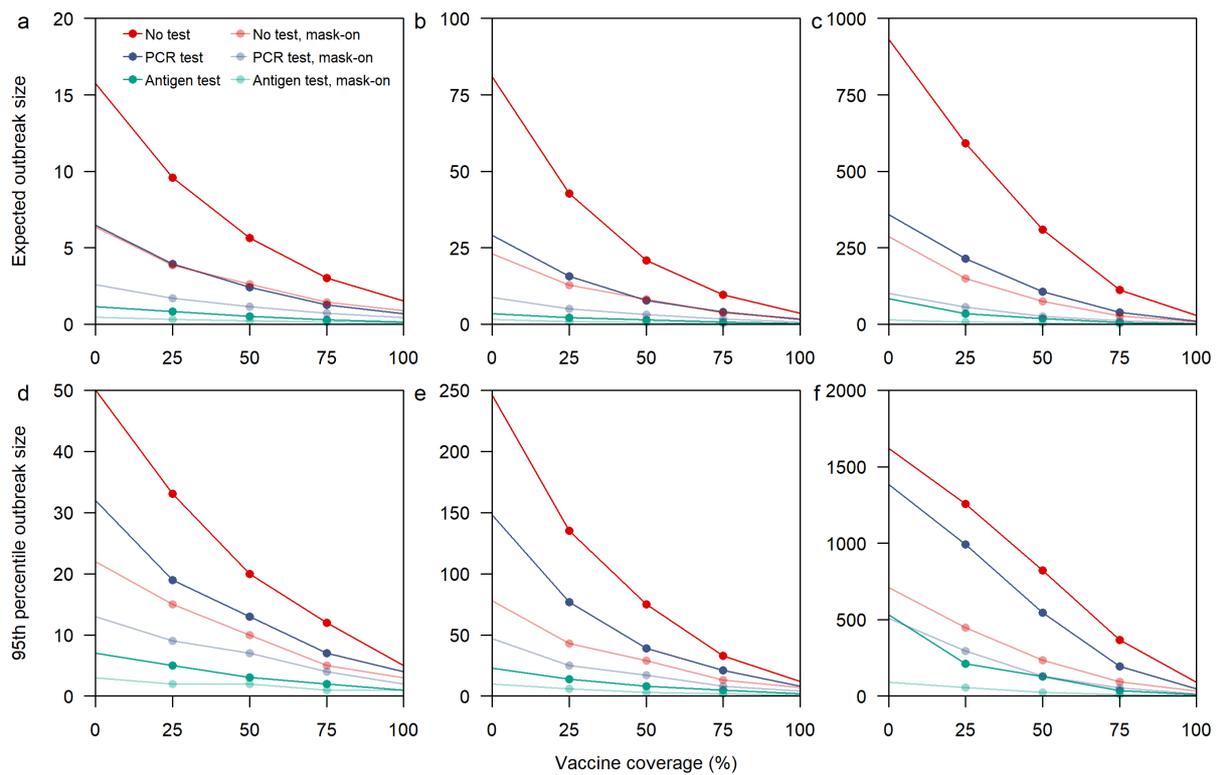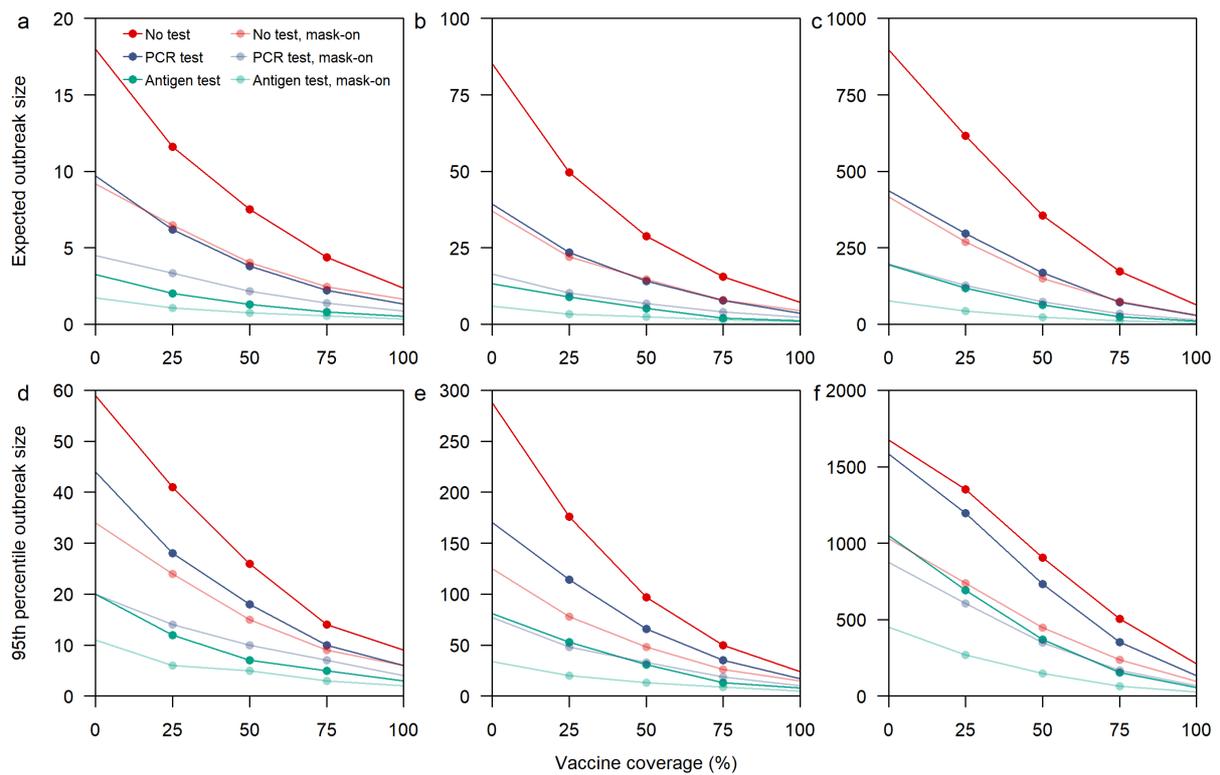
Supplementary Figure 4 Average and 95th percentile in outbreak size for varying interventions, vaccination coverage and assumption on network edge. Vaccines were assumed to confer 50% protection against infection and 50% lowered infectiousness for breakthrough infections in vaccinated individuals. Pre-symptomatic transmission was modelled to occur in 25% of the infections. (a, d) Edge weights vary based on the proportion of days with recorded interaction over a three-day sail period and duration of contact with weights increasing with days of interaction and contact time but reaches 95% saturation after 3 hours of contact, (b, e) same as (a, d) but reaches 95% saturation after 1 hour of contact, (c, f) edge weights vary based on proportion of days with recorded interaction.
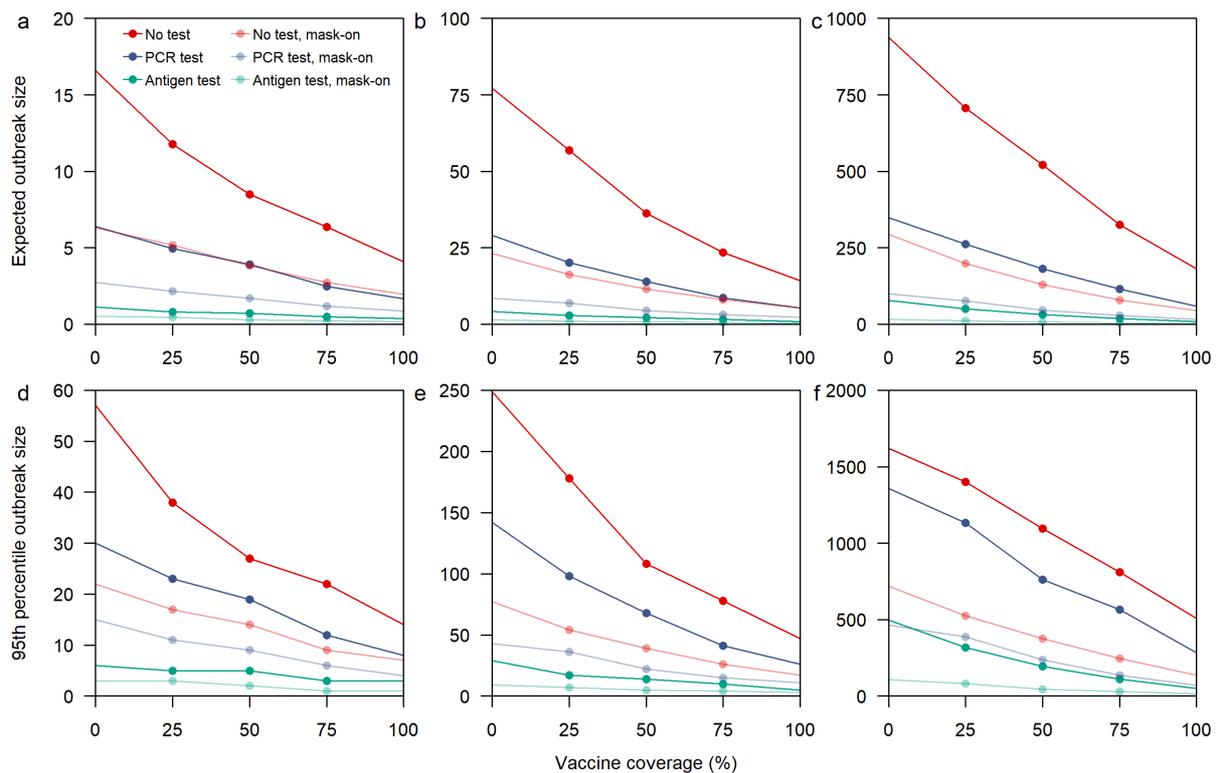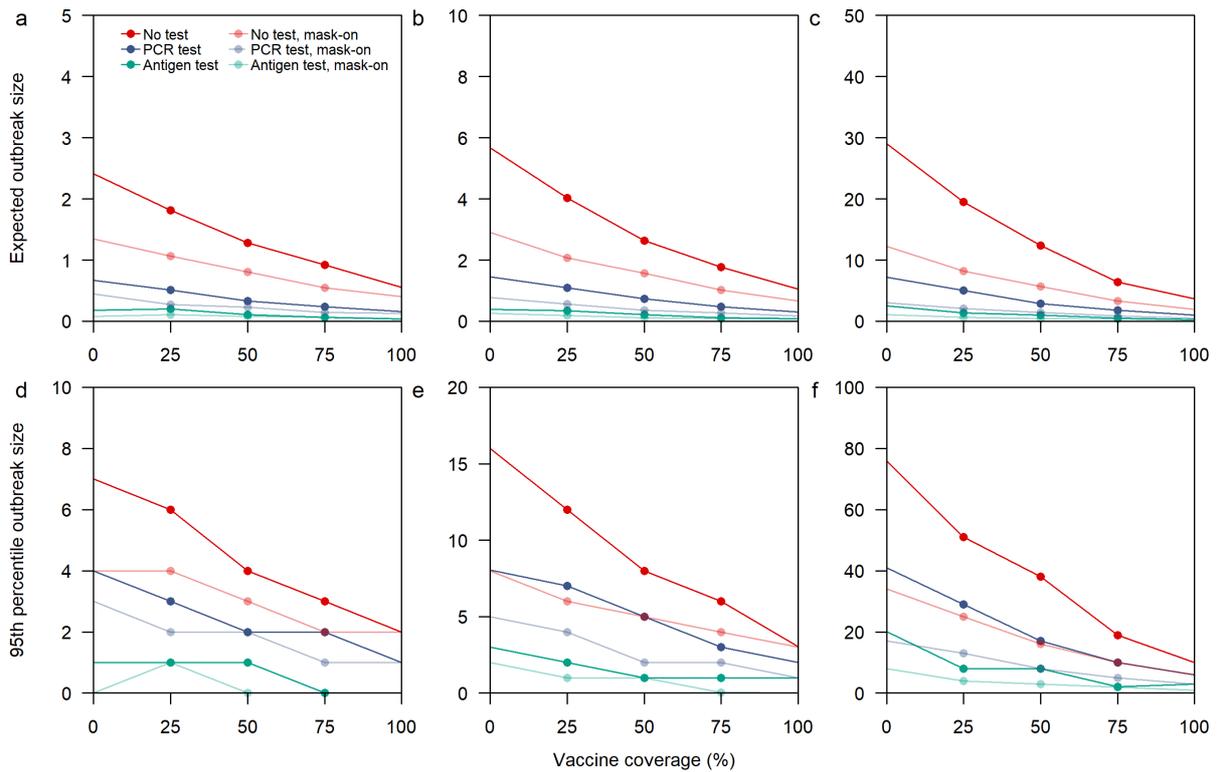
Supplementary Figure 5 Average and 95th percentile in outbreak size for varying interventions, vaccination coverage and assumption on network edge. Vaccine was assumed to confer 50% protection against infection <u>but no</u> lowered infectiousness. Pre-symptomatic transmission was modelled to occur in 25% of the infections. (a, d) Edge weights vary based on the proportion of days with recorded interaction over a three-day sail period and duration of contact with weights increasing with days of interaction and contact time but reaches 95% saturation after 3 hours of contact, (b, e) same as (a, d) but reaches 95% saturation after 1 hour of contact, (c, f) edge weights vary based on proportion of days with recorded interaction.

Relative to supplementary fig. 4, the expected outbreak size of all simulations increased across all vaccination coverage but the trend of outbreak size across varying coverage and differences between different combinations of interventions remains relatively unchanged
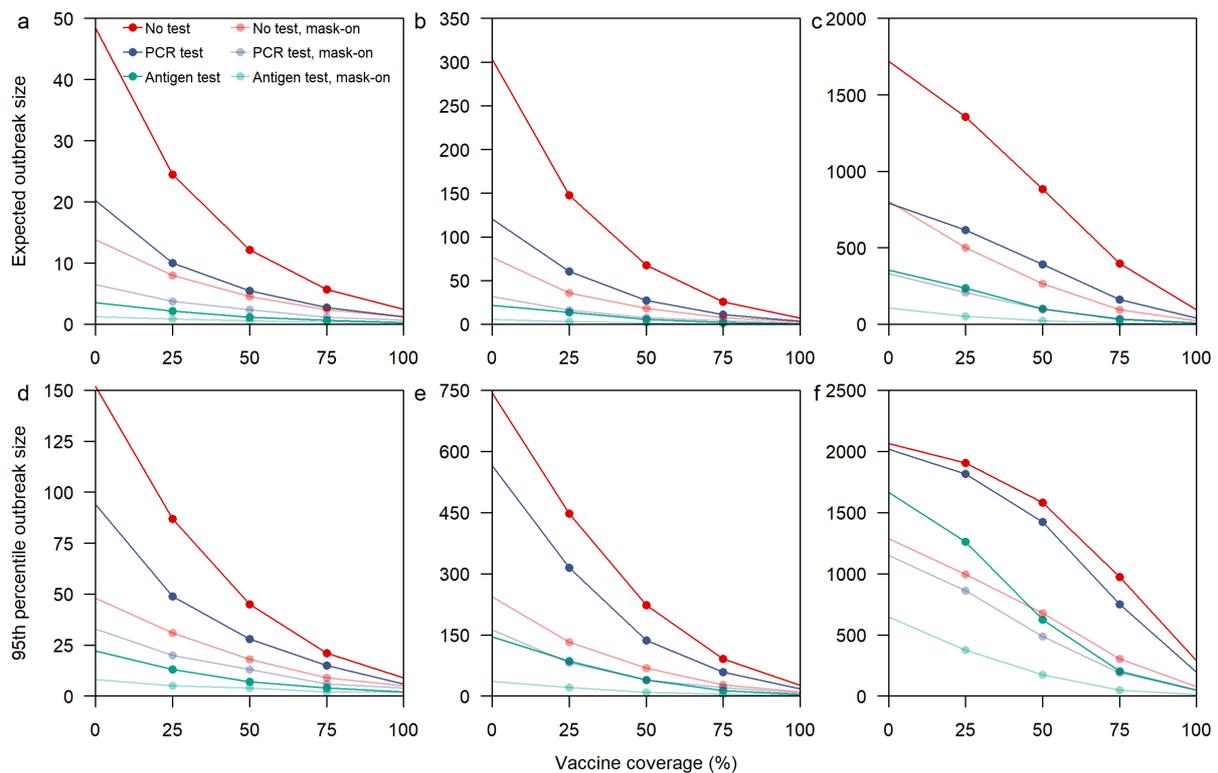
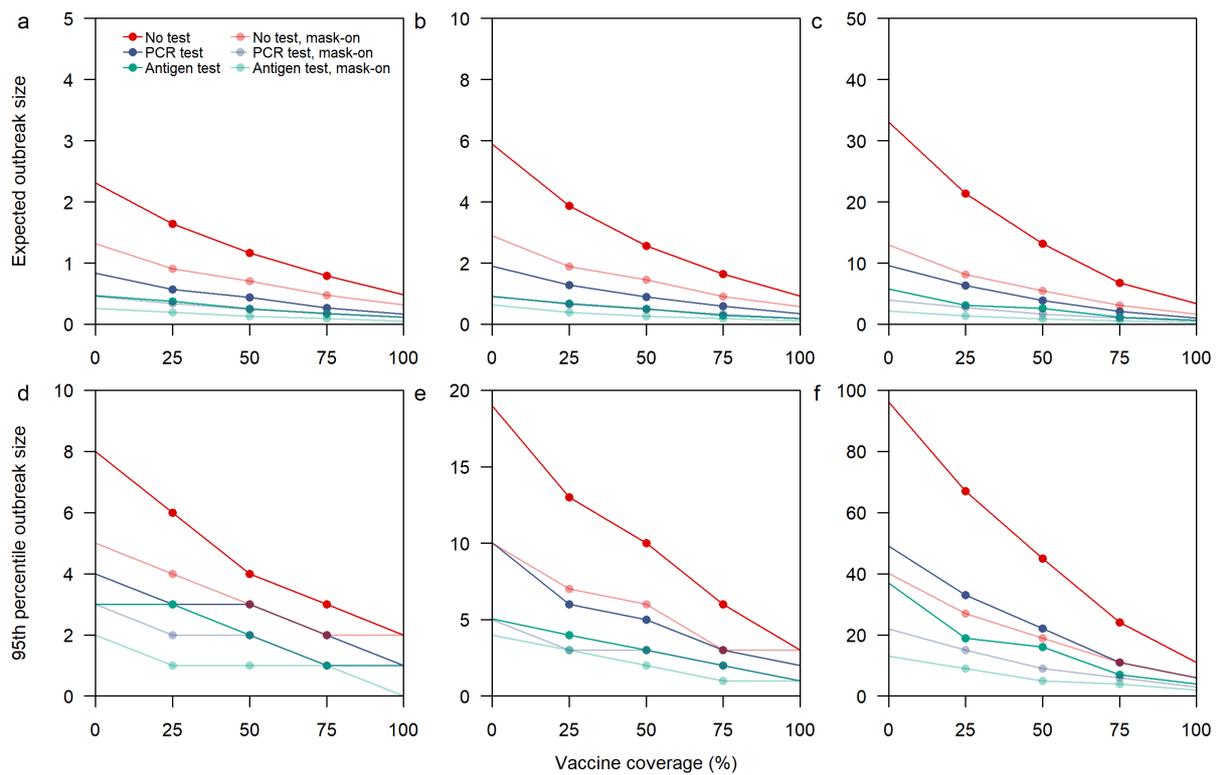**Supplementary Figure 6** Average and 95th percentile in outbreak size for varying interventions, vaccination coverage and assumption on network edge. Vaccine was assumed to confer 50% protection against infection and 50% lowered infectiousness. Pre-symptomatic transmission was modelled to occur in <u>50%</u> of the infections. (a, d) Edge weights vary based on the proportion of days with recorded interaction over a three-day sail period and duration of contact with weights increasing with days of interaction and contact time but reaches 95% saturation after 3 hours of contact, (b, e) same as (a, d) but reaches 95% saturation after 1 hour of contact, (c, f) edge weights vary based on proportion of days with recorded interaction.

Relative to supplementary fig. 4, individuals with onset late into the event were able to generate more infections and drove up the expected outbreak sizes. Furthermore, the differences between a mask-off, once off PCR intervention and a mask-on baseline intervention widens with the former having lowered potential in identifying cases prior to the event. At low or no vaccine coverage, the 95th percentile outbreak size under mask-on interventions was lower than that for mask-off interventions with the latter approaching an outbreak size of 90%.
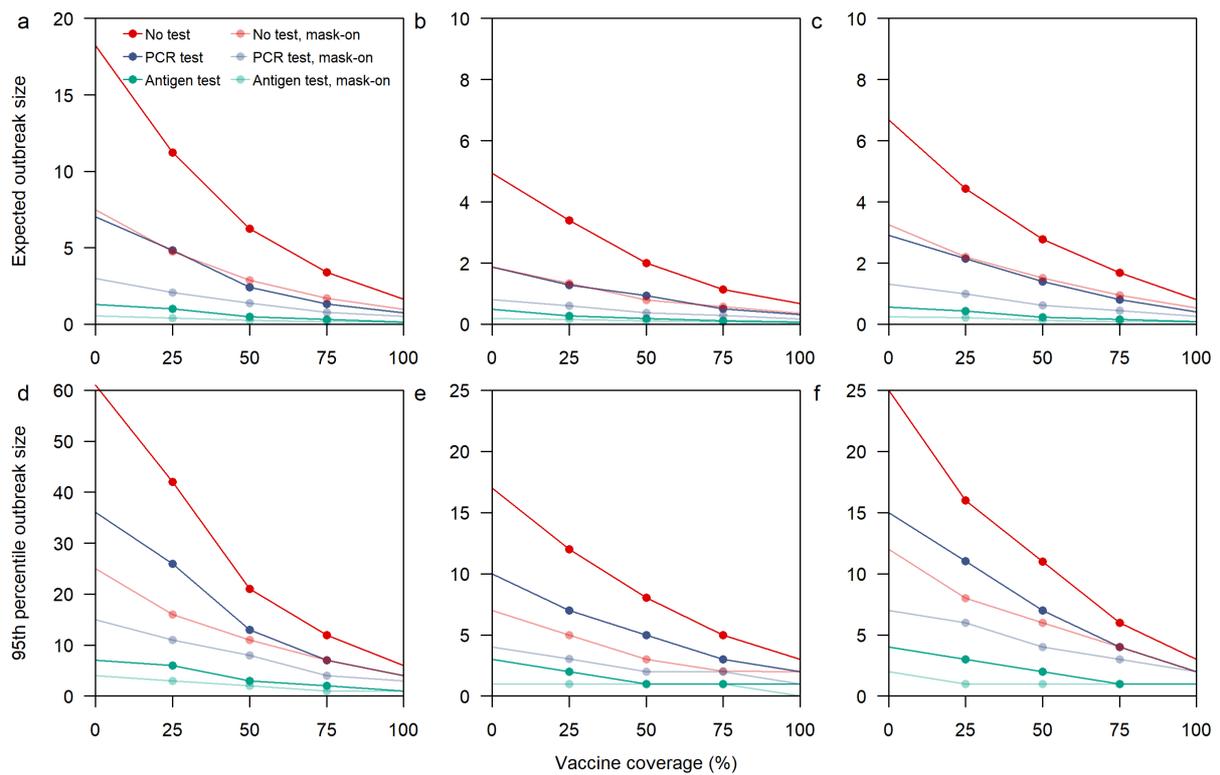
Supplementary Figure 7 Average and 95th percentile in outbreak size for varying interventions, vaccination coverage and for <u>different cruise sailings</u> on second (a, d), third (b, e) and fourth (c, f) sailing. General similarities in the results were found across all sailings (as potentially expected given their similarities in network structure). Vaccine was assumed to confer 50% protection against infection and 50% lowered infectiousness. Pre-symptomatic transmission was modelled to occur in 25% of the infections. Edge weights vary based on the proportion of days with recorded interaction over a three-day sail period and duration of contact with weights increasing with days of interaction and contact time but reaches 95% saturation after 3 hours of contact.

Supplementary Figure 8 Average and 95th percentile in outbreak size for varying interventions and vaccination coverage for outbreaks simulated based on a range of uncertainty in parameter values detailed in Table S1. (a, d) Daily edge weights vary based on the duration of contact with weights increasing with contact time but reaches 95% saturation after 3 hours of contact, (b, e) same as (a, d) but reaches 95% saturation after 1 hour of contact, (c, f) edge weights of 1 when interaction is recorded.

Supplementary Figure 9 Average and 95th percentile in outbreak size for varying interventions and vaccination coverage for outbreaks simulated on the temporal network for 3 days of sailing. Vaccines were assumed to confer 50% protection against infection and 50% lowered infectiousness for breakthrough infections in vaccinated individuals. Pre-symptomatic transmission was modelled to occur in 25% of the infections. (a, d) Daily edge weights vary based on the duration of contact with weights increasing with contact time but reaches 95% saturation after 3 hours of contact, (b, e) same as (a, d) but reaches 95% saturation after 1 hour of contact, (c, f) edge weights of 1 when interaction is recorded.

Supplementary Figure 10 Average and 95th percentile in outbreak size for varying interventions and vaccination coverage for outbreaks simulated on the static network for 3 days of sailing. Vaccines were assumed to confer 50% protection against infection and 50% lowered infectiousness for breakthrough infections in vaccinated individuals. Pre-symptomatic transmission was modelled to occur in 25% of the infections. (a, d) Daily edge weights vary based on the duration of contact with weights increasing with contact time but reaches 95% saturation after 3 hours of contact, (b, e) same as (a, d) but reaches 95% saturation after 1 hour of contact, (c, f) edge weights of 1 when interaction is recorded.

Supplementary Table 1. Parameter uncertainty for Fig S8, assuming uniform distribution across the assumed values

Parameter	Assumed values	Details and references
Pre-symptomatic transmission	25-50%	[1]
Adherence to isolation when tested positive	60-100%	For scenarios involving testing only, we assume that there are available cabins for individuals to isolate given that cruises are operating at 50% capacity. Lower bound based on self-reported adherence to isolation in the UK[2].
Relative risk of transmission by mask-off vaccinated, infected individual i	50-100%	Mean probability of transmitting infection reduces by 0-50%[3,4].
Relative risk of acquiring infection by mask-off vaccinated, susceptible individual j	30-50%	Mean probability of acquiring infection reduces by 50-70%[4–9].
Relative risk of transmission when both infected individual i and susceptible individual j are wearing a mask	20-60%	Mean probability of infection reduces by about 40-80% when both the infected individual and susceptible contact are wearing a mask[10].

Supplementary Table 2. Work functions of respective crew department

Department	Work functions
Entertainment	Cruise shows, live entertainment
Food & Beverage	From-end consumer facing food and beverages services
Galley	Back-end non-consumer facing galley, provision, stewarding
Gaming	Casinos, sports, arcade
Hotel	Hotel admin, front desk, embarkation training, spa, finance, IT, retail
Housekeep	Housekeeping, laundry
Marine	Deck, safety, security, medical, engineers, technicians, contractors
Security	Surveillance, security

Supplementary Note 1: CMMID COVID-19 Working Group funding

Supplementary References

1. Buitrago-Garcia, D. *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med* **17**, e1003346 (2020).

2. Office for National Statistics. Coronavirus (COVID-19) latest insights - Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/lifestyle (2021).

3. Harris, R. J. *et al.* Effect of Vaccination on Household Transmission of SARS-CoV-2 in England. *New England Journal of Medicine* **385**, 759–760 (2021).

4. Singanayagam, A. *et al.* Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The Lancet Infectious Diseases* **0**, (2021).

5. Elliott, P. *et al. REACT-1 round 13 final report: exponential growth, high prevalence of SARS-CoV-2 and vaccine effectiveness associated with Delta variant in England during May to July 2021.* http://spiral.imperial.ac.uk/handle/10044/1/90800 (2021).

6. Nanduri, S. Effectiveness of Pfizer-BioNTech and Moderna Vaccines in Preventing SARS-CoV-2 Infection Among Nursing Home Residents Before and During Widespread Circulation of the SARS-CoV-2 B.1.617.2 (Delta) Variant — National Healthcare Safety Network, March 1–August 1, 2021. *MMWR Morb Mortal Wkly Rep* **70**, 1163–1166 (2021).

7. Fowlkes, A. Effectiveness of COVID-19 Vaccines in Preventing SARS-CoV-2 Infection Among Frontline Workers Before and During B.1.617.2 (Delta) Variant Predominance — Eight U.S. Locations, December 2020–August 2021. *MMWR Morb Mortal Wkly Rep* **70**, 1167–1169 (2021).

8. Bernal, J. L. *et al.* Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. *medRxiv* 2021.05.22.21257658 (2021) doi:10.1101/2021.05.22.21257658.

9. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* **383**, 2603–2615 (2020).

10. Chu, D. K. *et al.* Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* **395**, 1973–1987 (2020).

G   Supplementary Material Chapter 7

Temporal contact patterns and the implications for predicting superspreaders and planning of targeted outbreak control

Rachael Pung[1,2,3], Josh A Firth[4,5], Timothy Russell[2,3], Tim Rogers[6], Vernon J Lee[1,7,8], Adam J Kucharski[2,3]

[1] Ministry of Health, Singapore
[2] Centre for the Mathematical Modelling of infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK
[3] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[4] Department of Biology, University of Oxford, Oxford, UK
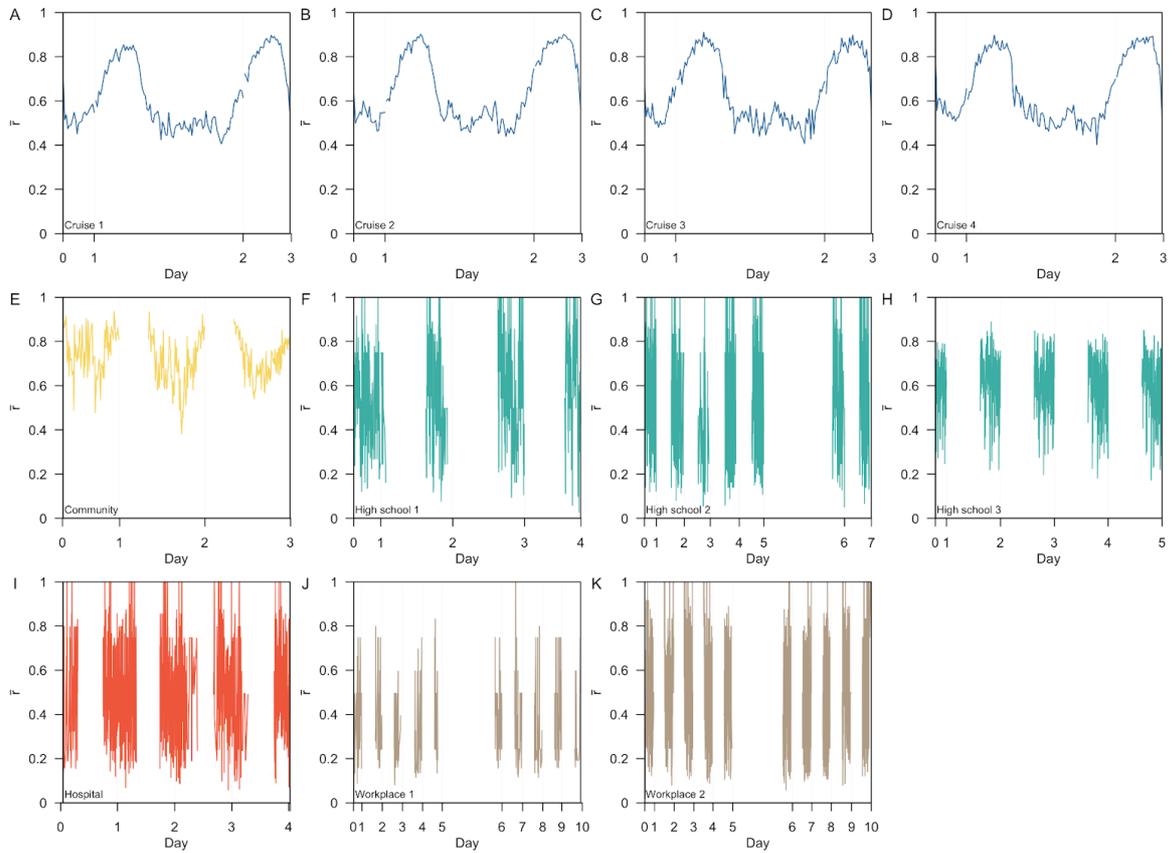[5] Merton College, University of Oxford, Oxford, UK
[6] Department of Mathematical Sciences, University of Bath, Bath, UK
[7] National Centre for Infectious Diseases, Singapore
[8] Saw Swee Hock School of Public Health, National University of Singapore, Singapore

Corresponding author: rachael.pung@lshtm.ac.uk

Supplementary Figure 1 Changes in contact retention index, $\bar{r}$, over time for (A-D) four cruise networks, (E) one community network, (F-H) three high school networks, (I) one hospital network, (J-K) two workplace networks. The duration of observation for each day is not necessarily the same across all studies.

Supplementary Figure 2 Contacts patterns in different settings for contacts assuming undirected contacts in all networks, (a) distribution of contact retention index, $\bar{r}$, over consecutive timesteps, (b) proportion of each type of contact retained for respective $\bar{r}$

Supplementary Figure 3 Contacts patterns in different settings for contacts formed in a fixed time window of 1-hr, (a) distribution of contact repetition, $\bar{r}$, over consecutive timesteps, (b) proportion of each type of contact retained for respective $\bar{r}$
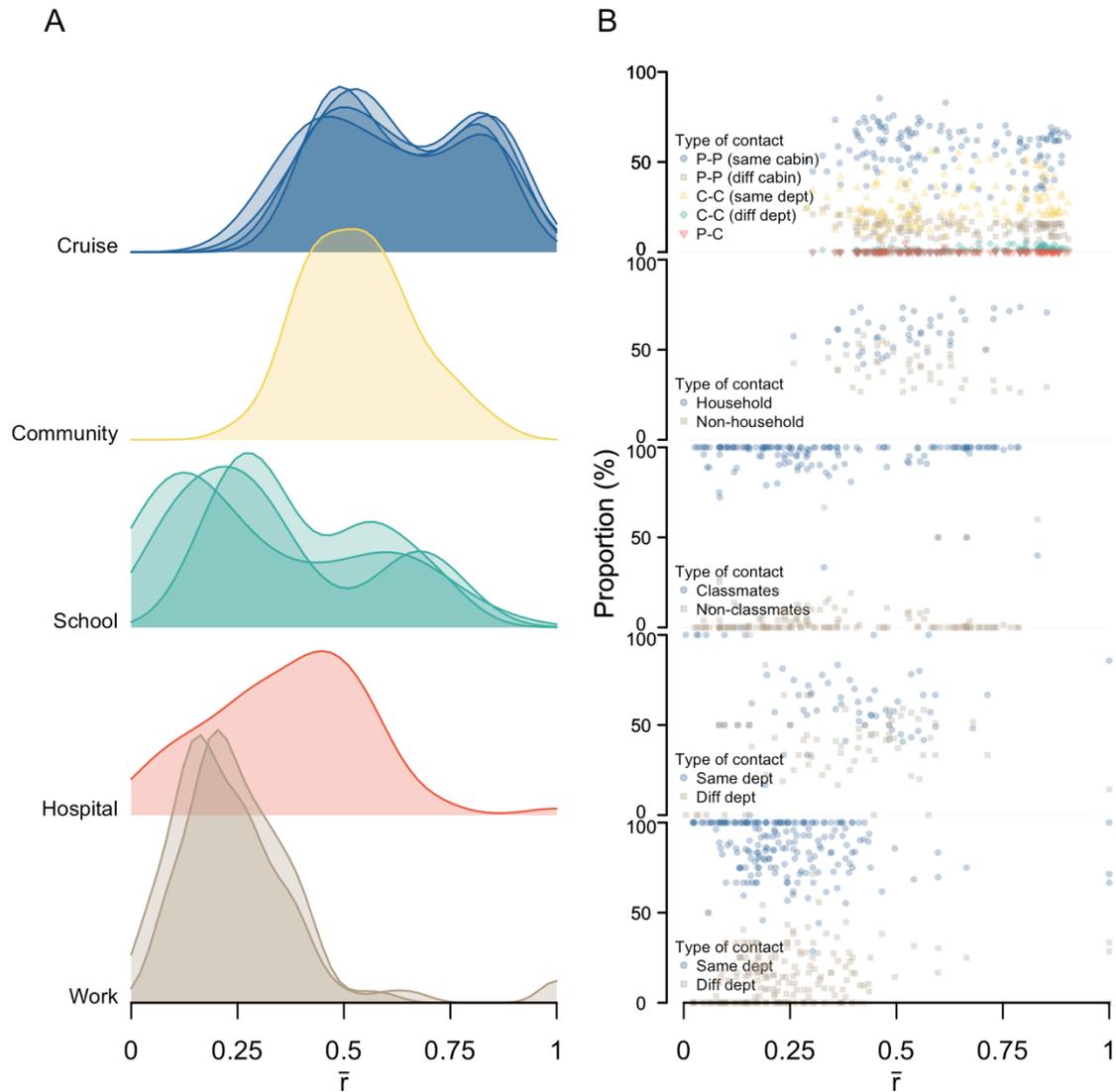
Supplementary Figure 4 Contacts patterns in different settings for contacts formed in a fixed time window of 15-min, (a) distribution of contact repetition, $\bar{r}$, over consecutive timesteps, (b) proportion of each type of contact retained for respective $\bar{r}$
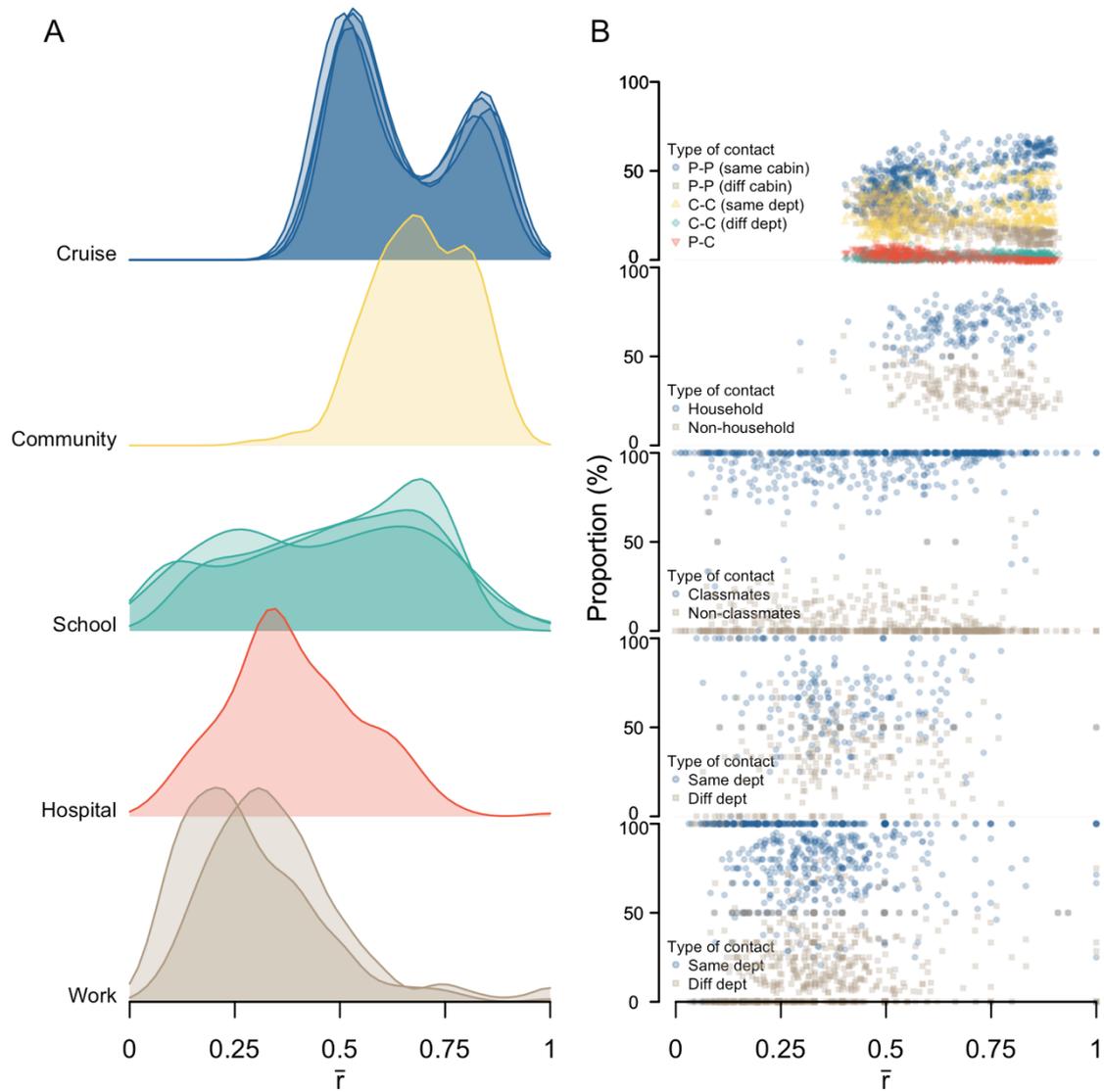
Supplementary Figure 5 Proportion of 'superspreaders' and 'superspreading events' in respective networks (coloured). Individuals that account for the top 80% of the contact (A) episodes or (B) duration in a day were identified.