UNIVERSITÁ DI TRENTO
DOTTORATO DI RICERCA IN MATEMATICA

Addressing Public Health Challenges Through Epidemiological Models

Tailored Mathematical Approaches for the Control of Infectious Diseases

DOCTORAL THESIS

Author
Alfredo De Bellis

Supervisor (s)
Dr. Giorgio Guzzetta, Prof. Andrea Pugliese

PhD in Mathematics, XXXVII cycle

Acknowledgements

I would like to sincerely thank my research group at FBK for providing an excellent environment where I had the opportunity to learn and grow during the past three years. In particular, I am deeply grateful to my supervisor, Dr. Giorgio Guzzetta. Your approach to science, your talent for finding solutions, and your ability to guide and support have been, and will continue to be, a profound source of inspiration for me. I could not have asked for a better supervisor. I also want to express my gratitude to my co-advisor Prof. Andrea Pugliese for the inspiring conversations we shared on the projects.

A special thanks goes also to the group at the UMCU for welcoming me during my visiting period and making me feel like a part of the team. I am especially grateful to Ganna Rozhnova, whose mentorship throughout our project has been invaluable.

I would also like to extend my gratitude to all the collaborators from research institutes and public health organizations who contributed to the projects that form the basis of this thesis. Your expertise and insights were fundamental to the success of the research and have greatly enriched my understanding of the field.

Finally, I wish to acknowledge all the researchers I have had the chance to engage with during various scientific events. The stimulating exchanges with you have broadened my perspectives and had a significant impact on my work.

Table of Contents

List of figures

List of tables

INTRODUCTION

Infectious diseases represent a growing challenge in our increasingly interconnected world. Globalization, urbanization, environmental and climatic changes have all accelerated the spread of pathogens, making it more difficult to predict and contain outbreaks. Public health must therefore expand its range of tools to face these rising threats. The recent COVID-19 pandemic underscored the critical importance of mathematical modeling, as governments and health authorities relied heavily on model estimations to guide life-saving interventions. Mathematical models help estimate key epidemiological parameters, shed light on the mechanisms of transmission, and assess the risk of spread across populations. Additionally, they allow the retrospective evaluation of the effectiveness of interventions and the exploration of future scenarios based on prospective public health measures. To ensure that models are sufficiently robust to provide guidance for decision making, they need to accurately reflect the complexities of real-world pathogen transmission, and to adapt to rapidly evolving scenarios. The close interaction between modelers and public health practitioners can provide insights that are critical for monitoring disease trends, assessing transmission risks, and making informed decisions, augmenting the intrinsic value of epidemiological data.

Background

Although today's mathematical models are highly sophisticated and often capable of capturing the intricacies of disease spread, this level of refinement is a relatively recent achievement in modern computational epidemiology [1]. Over the past century, this evolving field has made remarkable progress, moving from early, simplified models to more customized and realistic representations that are tailored to the specific heterogeneities of each problem [2]. This transition, from a theoretical exploration of general transmission laws to practical, data-driven applications, has made models increasingly relevant for addressing real-world challenges in public health.

The main breakthrough in our understanding of infectious diseases transmission came with the pioneering work of Kermack and McKendrick in 1927 [3]. Their foundational model was the first systematic attempt to mathematically describe the dynamics of infectious diseases, laying the groundwork for modern epidemiological modeling. The well-known compartmental models, such as the SIR (Susceptible-Infected-Recovered) model [4], derive almost directly from their work and remain key frameworks in epidemiology. These models capture the essential features of epidemics at the population level, introducing critical concepts, such as the basic reproduction number R_0 [5], which quantifies the transmission potential of infectious agents within a population [6].

Despite their influence and utility, basic compartmental models are limited by certain simplifying assumptions. In their classical form, individuals in a population are indistinguishable except for their health status, either susceptible, infected, or recovered.

This assumption of homogeneous mixing, which implies that contact between individuals occurs randomly and uniformly, has been largely employed for its simplicity and proved to be reasonable when modeling large populations of individuals with relatively similar characteristics. However, homogeneous mixing is most often an oversimplification of the intricate real-world interactions [7]. To reflect the complexity and heterogeneity of actual populations more sophisticated models were needed.

In this perspective, the 1980s represented a pivotal moment in epidemiological modeling, with important studies highlighting the need to bridge the gap between theoretical models and real-world data [8]. Among these studies, the seminal work of Anderson and May [9,10] was particularly impactful in pointing out the importance of incorporating biological realism into mathematical models and revealed the necessity of accounting for various types of heterogeneity. Over the past decades, this has led to a growth in research aimed at creating more refined models, integrating data-driven insights, and including stochastic elements to better capture the random nature of infectious disease outbreaks [11].

One of the critical advancements has been the inclusion of age-structured models, driven by demographic data. In particular, stratification of populations by age has proven essential for modeling the dynamics of childhood diseases like measles or mumps [12,13,14,15], as age may correlate with susceptibility to infection, immune response, and also influence contact patterns [16]. Beyond age, other forms of heterogeneity are equally important. For example, the HIV epidemic in the 1980s highlighted the need to account for variability in infectiousness, as individuals exhibit different transmission potentials over the course of their infection [17,18], progressing from the acute to the chronic stage, and eventually to AIDS. Sexual contact networks exhibit significant variability, with a small core group of individuals with high contact rates playing a disproportionate role in sustaining the epidemic [19,20,21]. The general concept of 'superspreaders', i.e., a minority of individuals responsible for the majority of transmission events, extends to various infectious diseases [22]. This phenomenon underscores the importance of accounting for the structure of contact networks, which led to the interest toward network-based epidemiological models over the past few decades [22].

Spatial dynamics represents an additional factor of heterogeneity in the spread of infectious diseases, particularly in vector-borne diseases like Zika and dengue [24,25]. For such pathogens, transmission is influenced not only by the mobility of human populations but also by the spatial variability in vector abundance. One type of model that may explicitly include the spatial dimension are metapopulation models, which have been developed to capture both local transmission and the movement of human populations between regions [26].

The complexity of real-world systems presents significant challenges in model development. While it is possible to include a wide range of heterogeneities, such as contact networks or spatial structures, this often comes with trade-offs. More detailed models require a higher granularity of data, a larger number of parameters to estimate, and significantly increased computational costs. Individual-based models, for example, simulate the dynamics of individuals and their interactions in great detail, tracking characteristics of each individual in a population over time [27,28]. These models allow for the representation of diverse transmission settings, such as households, schools, and

workplaces [29], as well as specific scenarios like hospital transmission [30] or sexual networks [31]. Individual-based models make it possible to evaluate intervention strategies tailored to individuals (e.g., contact tracing [32] or ring vaccination) which would be difficult to assess with standard compartmental models. However, these models are computationally intensive and may not always be necessary or practical to implement.

An alternative approach that provides both a high level of detail and computational efficiency is offered by non-generative Bayesian inference models, which can be applied to individual line-list data. Leveraging spatial, temporal, or genetic information included in such data, these models can be used to probabilistically reconstruct transmission chains (i.e., infector-infectee pairs). They have been employed, for example, to study outbreaks of Ebola [33], chikungunya [34], dengue [25], and others [35,36]. These models do not support scenario simulations but have proven effective to extract key epidemiological insights on transmission dynamics.

In general, identifying the key characteristics of a pathogen's transmission dynamics can be challenging, requiring modelers to carefully balance accuracy with practicality. This balance is not only critical for retrospective analyses but becomes particularly vital during infectious disease emergencies, where rapid, data-driven decisions are essential.

Thesis in context

In my PhD research, I focused on addressing current and impactful public health questions by employing a range of models tailored to the specific research questions, heterogeneities in pathogen transmission, and available data. This work was conducted in close collaboration with local health authorities and research institutes, who provided essential data, contributed to their interpretation, and helped to define the research questions. Throughout these studies, there was continuous bidirectional feedback with the stakeholders, allowing us to refine and adjust the models and analyzed scenarios to ensure more robust and relevant results.

The first part of the thesis (Chapters 1 and 2) focuses on the COVID-19 pandemic, specifically the period between 2021 and early 2022, which saw the emergence of several SARS-CoV-2 variants of concern (VOCs). A critical public health need during this period was the rapid characterization of the epidemiological features of these variants, which was necessary for timely adjustments to control measures and policies.

One key epidemiological metric is the generation time, defined as the interval between infection events within a transmission pair. The distribution of generation times may be interpreted as a proxy for infectiousness over time and influences decisions about the duration of quarantine and isolation [37]. However, accurately estimating the generation time is extremely challenging since it is not directly observable. A related quantity that can be measured directly, using data from epidemiological investigations performed during contact tracing activities, is the serial interval. This is defined as the time between the onset of symptoms in the infector and the infectee within a transmission pair. In absence of better measures, the serial interval is often used as a proxy for the generation time. However, this approximation may be inaccurate when incubation periods have a high individual variability and when pre-symptomatic transmission is important, as is the

case for SARS-CoV-2 [38,39]. In some rare cases, the times of exposure of infectors and infectees can also be determined within epidemiological investigations when the exposure episodes are clear and well-defined in time, leading to a direct measure of the generation time. However, even in these rare cases, the measure may be heavily biased with respect to the desired quantity. Indeed, studies have shown that competition among infectors for available susceptible individuals and depletion of the susceptible pool can artificially shorten the observed ("realized") generation time, especially for highly transmissible pathogens [40]. This happens in particular for transmission settings with small populations such as households, which are often those targeted by epidemiological investigations. For example, it has been showed in [40], through a network-based transmission model, that realized generation times within a household are shorter than in the general community due to competition effects and the rapid depletion of susceptibles. Therefore, direct measures of serial intervals and generation times may be affected by the context of transmission, the intensity of pathogen transmissibility, and the presence of control measures and may not represent the generation time that would be observed in the general population ("intrinsic" to the pathogen), which better represent a proxy of the infectiousness of hosts over time [41].

To address the challenge of estimating the intrinsic generation time, the Italian national institute of health (Istituto Superiore di Sanità) [42], in collaboration with the health authority of the province of Reggio Emilia [43], provided us with extensive household contact tracing data for SARS-CoV-2 cases. The datasets included individual line-list data from 6,272 cases across 2,240 households for the Alpha variant, 3,452 cases from 1,305 households for the Delta variant, and 23,122 cases across 8,903 households for the Omicron variant, with information on symptom onset dates, test results, vaccination status, and quarantine times for each individual. We applied a Bayesian inference model to these data, in order to reconstruct transmission chains and estimate the intrinsic generation time for the Alpha, Delta, and Omicron variants. These estimates aimed to fill a significant gap in existing research. Up until this point, few studies had investigated the generation time for these variants [44,45,46,47], most of which estimated context-specific realized generation times.

Chapter 1 will be based on the publication: Manica M, Litvinova M, De Bellis A, et al. Estimation of the incubation period and generation time of SARS-CoV-2 Alpha and Delta variants from contact tracing data. Epidemiology & Infection. 2023;151:e5.

Chapter 2 will be based on the publication: Manica M, De Bellis A, Guzzetta G, et al. Intrinsic generation time of the SARS-CoV-2 Omicron variant: An observational study of household transmission. The Lancet Regional Health–Europe. 2022;19.

The second project presented in this thesis (Chapter 3) is part of the "Healthy Sailing" initiative [48], a collaborative effort involving leading scientists in maritime transport epidemiology and general travel medicine, as well as port health authorities and cruise line companies. This project is a research and innovation action aimed at improving the quality of passenger shipping services by providing an evidence base for the effectiveness of proposed measures to prevent, mitigate, and manage infectious diseases on board passenger ships.

Cruise ships, as semi-closed, crowded environments where international travelers

frequently interact, are particularly vulnerable to the spread of infectious diseases. One of the major health threats in such settings is gastrointestinal outbreaks, primarily caused by norovirus [49]. During a norovirus outbreak, isolation after diagnosis is one of the key interventions, with different protocols implemented in different cruise ships and conditions. The impact of the duration of isolation and of timely diagnosis remains poorly quantified. Our study aimed to fill this gap to provide insights in optimal strategies for outbreak management onboard.

The University of Thessaly, in collaboration with the travel company Celestyal Cruises, provided a line-list of 121 individual epidemiological records from a large outbreak on a cruise ship, including information on symptom onset, diagnosis date, and number and type of symptom episodes before diagnosis. Using a Bayesian model, we reconstructed transmission chains, explicitly accounting for the daily routines of passengers and crew members (e.g., nighttime spent in cabins, port visits, and interactions in public areas), and incorporating diagnostic delays (i.e., the time that elapses between infection and diagnosis). We then used a branching process model that stochastically generated infections onboard to explore various intervention scenarios, such as faster diagnoses and alternative isolation protocols. This allowed us to assess the potential impact of these strategies on minimizing virus transmission. To date, mathematical modeling literature of norovirus outbreaks on cruise ships include only one previous study estimating the effectiveness of isolation [50], based on the simplifying and unrealistic assumption of immediate isolation of cases upon development of symptoms.

The work presented in Chapter 3 is in the final stage of preparation, and it will be submitted soon for publication.

The final work presented in this thesis (Chapter 4) reports the results of a collaboration I undertook during my visiting period at the University Medical Center Utrecht (UMCU) [51] in partnership with the National Institute for Public Health and the Environment (RIVM) in the Netherlands [52]. This collaboration emerged from a summer school organized by UMCU mathematical modeling of infectious diseases group, where I attended as a student. The research focused on modeling the population-level impact of introducing a potential HIV cure on the dynamics of the HIV epidemic among men who have sex with men (MSM) in the Netherlands.

Despite encouraging epidemiological trends, HIV continues to pose a significant burden on the MSM population in high-income countries. HIV prevalence among MSM remains disproportionately high compared to the general population, with infection rates persisting despite the availability of highly effective interventions such as the antiretroviral therapy (ART) and pre-exposure prophylaxis (PrEP) [53,54]. While ART treatments have transformed HIV from a fatal disease to a manageable chronic condition, individuals living with HIV still experience lifelong health challenges and a reduced quality of life. On the other hand, PrEP has significantly reduced HIV transmission, but elimination goals are still far from being achieved, even in contexts of relatively high coverage.

The global search for an HIV cure is an ongoing priority in biomedical research, with notable progress and frequent breakthroughs in clinical trials [55]. However, there is limited research assessing how the introduction of a cure would influence HIV

epidemiology [56,57], particularly in high-income countries. In our modeling work, we evaluated the population-level impact of introducing different types of HIV cures, using simulations to explore the conditions under which a cure could either accelerate or potentially reverse the current decline in HIV transmission within the MSM population. We employed a deterministic compartmental model that accounted for different stages of the disease and stratified individuals by sexual risk groups, based on heterogeneous sexual contact rates within the MSM population in the Netherlands. The model was calibrated and validated using behavioral survey data on sexual contacts provided by social scientists at UMCU, along with epidemiological records and data on times to diagnosis from the RIVM.

The work presented in Chapter 4 is in the final stage of preparation, and it will be submitted soon for publication.

I contributed as first author in the studies presented in Chapters 3 and 4, as co-first author in Chapter 2, and as one of the main co-authors in Chapter 1. In all the projects, I contributed by developing and implementing models and conducting analyses.

Additionally, I am a co-author on another study on real-time forecasting of norovirus outbreaks onboard cruise ships (in the final stage of preparation) and main author in two further ongoing studies, on COVID-19 transmission onboard of cruise ships and on *methicillin-resistant Staphylococcus Aureus* (MRSA) transmission in a paediatric hospital. These studies will be briefly discussed in the Conclusion chapter.

CHAPTER 1

Intrinsic generation time of Alpha and Delta SARS-CoV-2 variants

Introduction

The second year of the COVID-19 pandemic has been characterized by the global emergence of several lineages which were able to replace circulating ones thanks to their increased transmissibility [58]. In particular, 2021 saw the sequential rise and fall of two variants of concern, Alpha and Delta, the latter of which has been rapidly outpaced by Omicron around the end of 2021. Compared to ancestral strains, scarce quantitative information is available on several variant-specific epidemiological quantities, among which the incubation period (i.e., the time elapsed between the date of infection and symptom onset) and the generation time (i.e., the time elapsed between the date of infection of a primary case and that of a secondary case). These two quantities are especially important to define the duration of isolation for infectious individuals and of quarantines for close contacts and travelers, as well as protocols for community-based interventions such as contact tracing activities [59,60,61] and class/school closures [60 ,62]. The knowledge of the generation time distribution also informs the estimation of the net reproduction number (i.e., the average number of new cases generated by an infectious case at a given time of the epidemic) [37], which is a key indicator for monitoring epidemic outbreaks and defining population-level measures, such as physical distancing and movement restrictions [63].

The incubation time is mostly a biologically determined parameter since it depends on virus characteristics and virus-host immunological and pathological interactions. On the other hand, the generation times that occur in a population depend on the interactions between infectious individuals and their contacts, and therefore may be subject to specific epidemiological conditions in which they are measured, including individual behaviors, environmental determinants, and control measures put in place [40]. For example, observable generation times within a household are generally shorter than in the general community due to competition effects and the rapid depletion of susceptibles [40]. A distinction is therefore necessary between “realized” distributions of the generation time, which are actually occurring in specific networks of contacts, and the “intrinsic” distribution, i.e., the one that is expected in the general population in absence of control interventions and local network dynamics [41]. The intrinsic generation is less sensitive to the transmissibility conditions of the epidemiological setting under study. Here, we applied a Bayesian inference approach to COVID-19 contact tracing data from the province of Reggio Emilia, Italy, during 2021 to estimate the distribution of incubation

periods and generation times (both intrinsic and realized) for SARS-CoV-2 variants Alpha and Delta.

Methods

Data

Contact tracing activities were carried out in the province of Reggio Emilia, Italy throughout the duration of the pandemic to mitigate the spread of SARS-CoV-2. Identified SARS-CoV-2 cases occurring in the province were confirmed via a Polymerase Chain Reaction (PCR) assay, reported in real time to the public health service of the Reggio Emilia local health authority and isolated at home until a negative PCR test result and for a maximum of 21 days. During the study period all antigenic positive tests were confirmed with PCR. All cases were contacted via telephone to identify their close contacts. A close contact was defined as a person who stayed in the same room with a confirmed case without a face mask, or for more than 15 minutes at less than 2 meters, between 2 days before and 10 days after symptom onset (for symptomatic cases) or diagnosis (for asymptomatic infections). Contacts were tested and quarantined at home for 10 days, if they had a negative PCR test result at that date, or for 14 days without testing [64]. All household members of a case were quarantined until a negative test after the end of the isolation period for the index case. Compliance with at least one of the tests proposed by the public health service was 97.0% during the study period (March-October 2021).

Table 1-1. Descriptive statistics of SARS-CoV-2 cases in the household datasets for Alpha and Delta variants.

	ALPHA	**DELTA**
Period	March 1 – April 30, 2021	August 1 – October 31, 2021
Number of cases	6272	3452
Clinical outcome (%):		
Symptomatic	3591 (57.3%)	2680 (77.6%)
Asymptomatic	2681 (42.7%)	772 (22.4%)
Gender (%):		
Male (%)	3058 (48.8%)	1685 (48.8%)
Female (%)	3214 (51.2%)	1767 (51.2%)
Age group (%):		
0-15 years old	1279 (20.4%)	886 (25.7%)
16-44 years old	2236 (35.7%)	1181 (34.2%)
45-64 years old	1915 (30.5%)	948 (27.5%)
65+ years old	842 (13.4%)	437 (12.6%)
Vaccination status at the end of the period (%):		
1 dose	222 (3.54%)	226 (6.5%)
2 doses	78 (1.24%)	1386 (40.2%)
3 doses	0 (0.0%)	8 (0.2%)
None	5972 (95.2%)	1832 (53.1%)
Number of households	2240	1305
Average household size (2.5 to 97.5 percentile range)	2.97 (2 – 6)	2.83 (2 – 5)

Data on test results, symptom onset date (if applicable), and setting of likely transmission were collected for all identified cases and their contacts and were linked to individual records on vaccination history (first, second, and booster doses). The date of the last reported contact with any known case within a cluster (date of last exposure), as

uncovered by epidemiological investigations, was also collected. Appropriate data quality checks were conducted in strict collaboration with the Reggio Emilia local health authority to minimize missing information and accurately define household clusters. A household cluster was defined as households with at least two positive individuals with a diagnosis spaced less than 25 days apart.

Since genomic information on the variant was not available, we conservatively defined two time periods where circulation of SARS-CoV-2 in the Region was almost exclusively attributable (at least ~90% prevalence) to variant Alpha (March 1 – April 30, 2021) and to Delta (August 1 – October 31, 2021) [65]. Statistics of the corresponding datasets are summarized in Table 1-1.

Estimation of the incubation period

For the estimation of the incubation period, we selected all symptomatic cases with a date of diagnosis within either of the two periods defined for Alpha or Delta. For each case, the date of the last negative PCR test T_N and the date of last exposure T_L were used to set the limits for the earliest and last exposure, respectively. We note that the dates of test results included in the database are always referring to tests taken in response to a positivity of a contact and not tests taken autonomously by contacts for other reasons. We excluded all cases for which either date was unavailable or for which the condition $T_N \leq T_L \leq T_S$, where T_S is the date of symptom onset, did not hold. The resulting sample for the estimation contains 193 observations for Alpha and 89 for Delta. We used the generalization of the Wilcoxon-Mann-Whitney test for interval-censored data to compare the empirical data in the two samples. Two parametric distributions (gamma and Weibull) were fit to the interval-censored empirical data on the time between likely infection and the symptom onset [66, 67] using a maximum likelihood optimization. The best fit was selected based on the minimum Akaike information criterion [68]. Confidence Intervals (CI) for the parameters of estimated distributions were obtained from the 2.5 to 97.5 percentile range of estimates over 10,000 bootstrap samples for censored data. See Appendix A for further details on the method and for sensitivity analyses on estimation criteria.

Estimation of the generation time and of the serial intervals

For the estimation of the generation time, we selected only household clusters for which all dates of diagnosis were included in either of the two periods defined for Alpha or for Delta. To reduce the possibility of missed diagnoses in the households due to false negative test results, we further selected households for which undiagnosed members had at least two negative test results. Figure 1-1 shows a schematization of an illustrative household cluster, with the corresponding dates of infection, symptom onset, diagnosis, and negative tests for individuals, as well as relevant intervals to be estimated. We adopted a Bayesian inference model for the reconstruction of transmission links in households already applied for the estimation of the generation time of the Omicron variant [69,70]. The model exploits the temporal information on SARS-CoV-2 infections recorded in the dataset to probabilistically identify, for every case, the likely source of infection (from outside the household or from a specific household member).

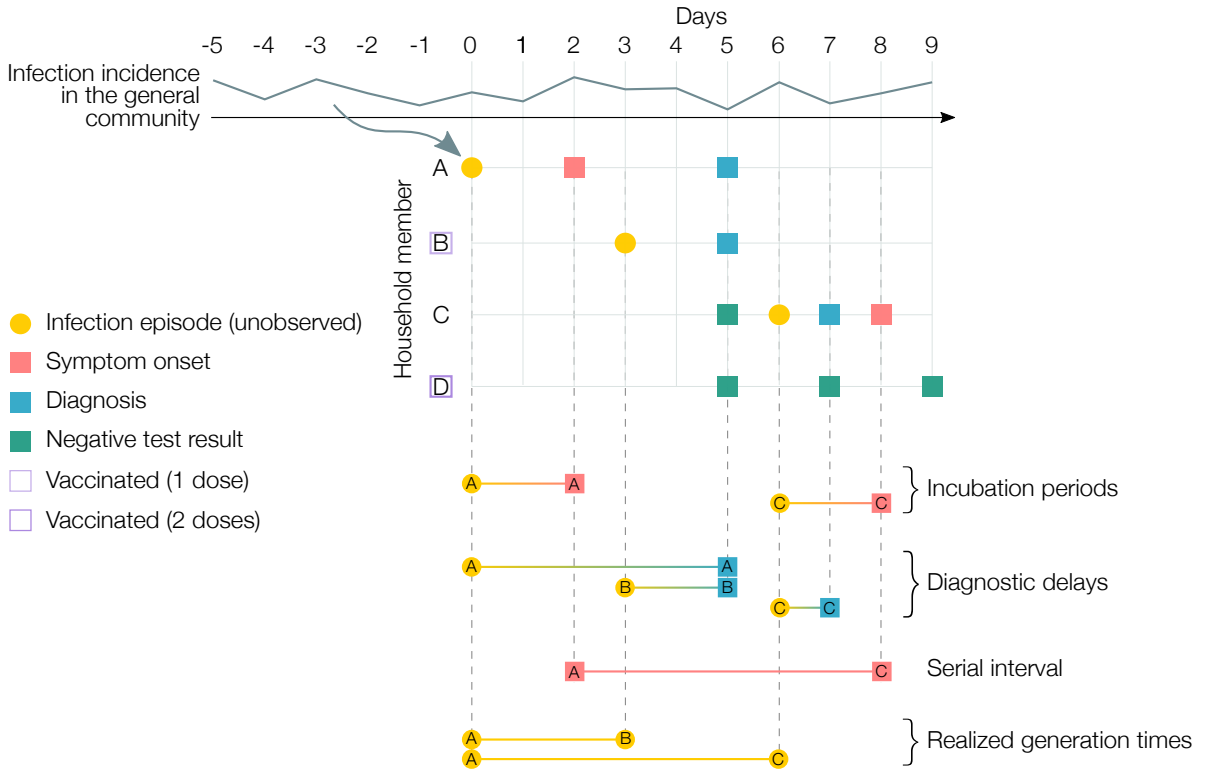

Figure 1-1. Illustrative example of actual transmission dynamics in household clusters. A household with 4 members, of which A was infected outside the household (in the general community) at day 0 and then transmitted to cases B (asymptomatic) and C (symptomatic), while D remained uninfected. B and D were vaccinated with 1 and 2 doses respectively. A hypothetical epidemic curve in the general community, representing the external force of infection on household members, is reported on top of the graph. Circles indicate unobserved events; squares indicate observed events. Examples of the temporal intervals of interest for the estimates of this work are reported in the bottom part of the figure. Note that for the household serial interval and the realized household generation time, the source of infection (whether from outside the household or from a household member, and, in the latter case, which household member) is also unobserved and needs to be probabilistically reconstructed. Pre-symptomatic transmission and negative serial intervals are also possible but have not been included in this example for the sake simplicity. The intrinsic generation time is not displayed as it represents the distribution of generation times among infections occurring in the general population in a fully susceptible population [41].

We assumed the parameters for the generation time to be Gamma-distributed. These parameters are simultaneously calibrated via a Markov Chain Monte Carlo approach where the likelihood of the observed data is defined mechanistically through the computation of the force of infection to which all individuals are subject over time. The force of infection takes into account the SARS-CoV-2 incidence in the general community, and the individual dates of infection and vaccination. We imputed the date of infection for each symptomatic case by subtracting from the date of symptom onset a random sample from the estimated discretized distribution of the incubation period. We then computed the diagnostic delay distribution as the distribution of delays between dates of infection and diagnosis in symptomatic individuals. Successively, we imputed the date of infection for asymptomatic cases by subtracting from the date of diagnosis a random sample from the estimated diagnostic delay distribution. For both symptomatic and asymptomatic individuals, the probability of an imputed date of infection was weighted by the probability of false negative test results based on available test dates and results for each individual. We repeated 100 times the sampling of infection dates and re-calibrated on each resampling the Bayesian model. 95% CrI for the estimated parameters were obtained from the resulting pooled distributions. All technical details for the model and calibration are reported in the Appendix. Using this Bayesian approach, we could estimate at the same time both the parameters of the intrinsic generation time and the

distributions of the realized household generation time and household serial interval. For each set of imputed infection dates and sample from the joint posterior distribution of parameters, we reconstructed likely transmission chains (i.e., the source of infection for each case). The distribution of realized generation times was obtained by the differences between infection dates in each inferred infector-infectee pair; correspondingly, the distribution of household serial intervals was obtained from the differences between symptom onset dates in each inferred pair of symptomatic infector-infectee.

We evaluated the robustness of model results against six sensitivity analyses (SA) encoding different assumptions in the model. In SA (a), we relaxed the assumption of the baseline model that symptomatic and asymptomatic individuals have the same distribution of diagnostic delays by considering an alternative method for inferring the date of infection of asymptomatic individuals. In SA (b) and (c), we imputed dates of infection using two alternative distributions of the incubation period previously estimated for ancestral SARS-CoV-2 lineages [67,71]. In SA (d) we assumed a halved transmissibility for asymptomatic individuals [72]. In SA (e), we considered the possible protection from previous infection in a fraction of undiagnosed household members. Finally, in SA (f), we considered a negligible adherence of household members to quarantine (i.e., the probability of being infected outside the household was unchanged upon onset of quarantine for household members).

Results

The best fit for the distributions of the incubation period was a gamma distribution, with a mean of 4.9 days for Alpha (95% Confidence Intervals of the mean, CI, 4.4-5.4; 2.5 to 97.5 percentile range of the mean distribution 1-12 days) and of 4.5 days for Delta (95% CI, 4.0-5.0; 2.5 to 97.5 percentile range of the mean distribution 1-10 days) (Figure 1-2 and Table 1-2). The differences between empirical distributions of incubation periods for Alpha and Delta variants were not statistically significant (Wilcoxon-type test p-value 0.45). Unsurprisingly, the estimate for the incubation period was longer (mean: 7.3-7.4 days for Alpha and 6.2-6.3 days for Delta) and had a larger uncertainty when including in the estimation those cases for which the date of earliest exposure was unknown (Appendix A), supporting the importance of considering only data samples for which information on the time window of exposure is more compelling.

The resulting estimated distribution of delays between infection and diagnosis (used to assign infection dates for asymptomatic individuals) had a mean of 7.14 days (2.5 to 97.5 percentile range: 3-15 days) for the Alpha variant (Table 1-2). The mean intrinsic generation time estimated for Alpha was 7.12 days (95% CrI of the mean: 6.27-8.44 days) and the mean realized household generation time was 4.41 days (95%CrI of the mean: 4.26-4.58 days) (Figure 1-3 and Table 1-2). The mean household serial interval was 2.43 days (95%CrI of the mean: 2.29-2.58 days), with 47.8 % (95%CrI: 45.5-49.8%) of transmission events being pre-symptomatic (i.e., secondary cases transmitted by cases who would develop symptoms after the transmission event). Sensitivity analyses yielded similar results, with the mean intrinsic generation time ranging between 6.22 and 7.77 days (Figure 1-4), the mean realized household generation time ranging between 4.09 and 5.08 days, and the mean household serial interval ranging between 2.14 and 2.53 days (Appendix A).

Table 1-2. Estimates for the incubation period, diagnostic delay, intrinsic and realized generation time, and household serial intervals. Reported parameters of shape and scale for the incubation period and intrinsic generation time refer to a gamma distribution. The mean distribution indicates the distribution obtained using the mean value estimated for the parameters.

		ALPHA	DELTA
INCUBATION PERIOD	mean (95%CrI) [days]	4.9 (4.4-5.4)	4.5 (4.0-5.0)
	2.5 to 97.5 percentile range of the mean distribution [days]	1-12	1-10
	shape mean (95%CrI)	3.08 (2.56-3.86)	4.43 (3.26-6.70)
	scale mean (95%CrI)	1.58 (1.24-1.93)	1.01 (0.65-1.43)
DIAGNOSTIC DELAY	mean (2.5 to 97.5 percentile range) [days]	7.1 (3-15)	7.1 (3-14)
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.12 (6.27-8.44)	6.52 (5.54-8.43)
	2.5 to 97.5 percentile range of the mean distribution [days]	1-18	1-17
	shape mean (95%CrI)	2.53 (2.27-3.21)	2.49 (2.14-2.97)
	scale mean (95%CrI)	2.83 (2.28-3.39)	2.63 (2.19-3.21)
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	4.41 (4.26-4.58)	4.05 (3.87-4.24)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.43 (2.29-2.58)	2.74 (2.62-2.88)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	47.8 (45.5-49.8)	50.9 (48.4-53.0)

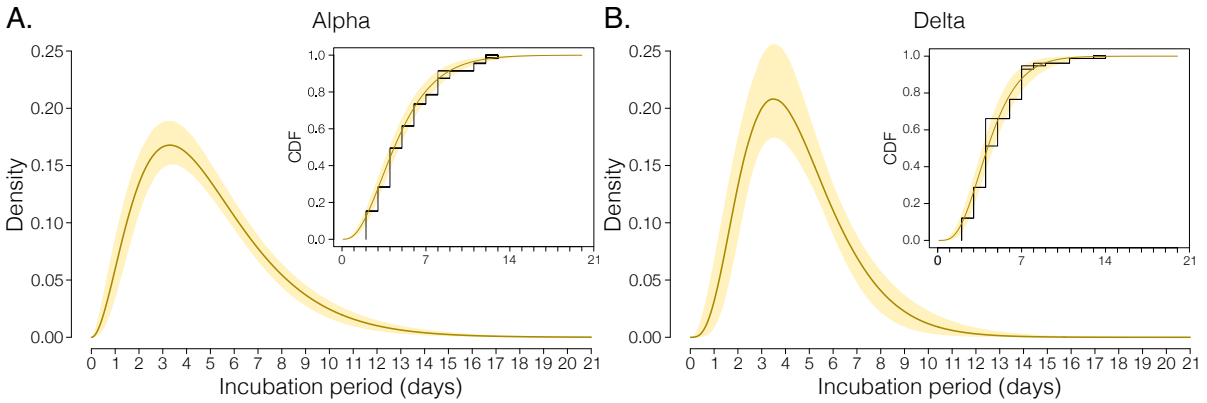

*Figure 1-2. Estimation of the incubation period for the Alpha and Delta SARS-CoV-2 variants. (**A**) Probability density function (PDF) of the estimated distribution of incubation period for Alpha variant with 95% CI based on nonparametric bootstrap resampling of the distribution parameters (10 000 samples). Line: mean PDF; shaded area: bootstrapped pointwise 95% CI. The inset shows the cumulative distribution function (CDF) of the empirical distribution (black line) where rectangles represent areas of non-unique empirical distribution function, CDF of the distribution fitted to interval-censored data (line) and bootstrapped pointwise 95% CI on probabilities (shaded area). (**B**) Same as (**A**), but for Delta variant.*

The estimated distribution of delays between infection and diagnosis had a mean of 7.12 days for the Delta variant (2.5 to 97.5 percentile range: 3-14 days) (Table 1-2). The mean intrinsic generation time estimated for Delta was 6.52 days (95%CrI of the mean: 5.54-8.43 days) and the mean realized household generation time was 4.05 days (95%CrI of the mean 3.87-4.24 days) (Figure 1-3 and Table 1-2). The mean household serial interval was 2.74 days (95%CrI of the mean 2.62-2.88 days), with 50.9% (95%CrI: 48.4-53.0%) of transmission events being pre-symptomatic. Sensitivity analyses yielded similar results, with the mean intrinsic generation time ranging between 5.95 and 7.38 days

(Figure 1-4), the mean realized household generation time ranging between 3.84 and 4.66 days, and the mean household serial interval ranging between 2.28 and 2.76 days (Appendix A).

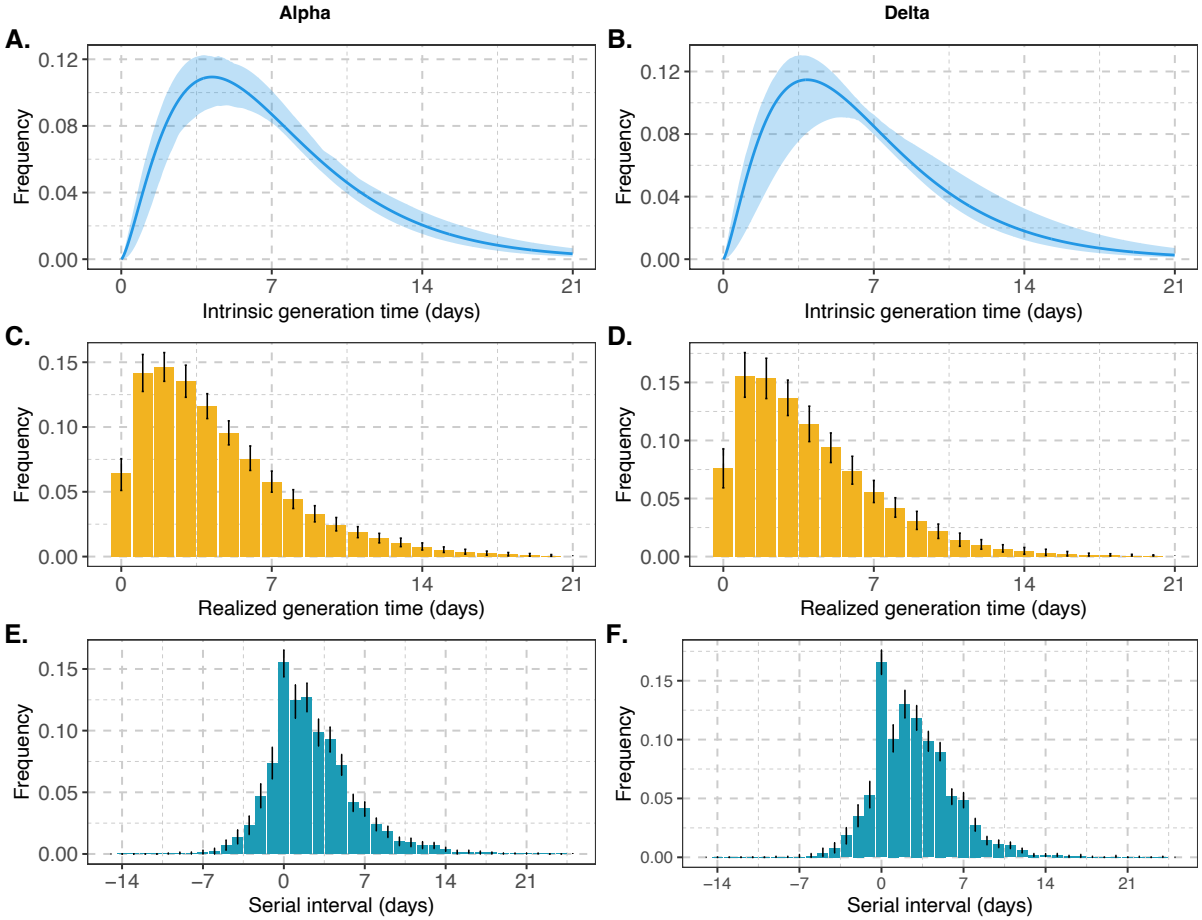

*Figure 1-3. Estimates of generation times and household serial intervals for the Alpha and Delta variants. (**A**) Distribution of the intrinsic generation time for the Alpha variant; solid line: mean estimate; shaded area: 95% CrI; (**B**) same as (**A**), but for Delta. (**C**) Distribution of the realised household generation time for the Alpha variant; bars: mean estimate over all reconstructed transmission chains; vertical lines: 95% CrI across all reconstructed transmission chains; (**D**) same as (**C**), but for Delta. (**E**) Distribution of the household serial interval for the Alpha variant; bars: mean estimate over all reconstructed transmission chains; vertical lines: 95% CrI across all reconstructed transmission chains; (**F**) same as (**E**), but for Delta.*

Discussion

We estimated the distribution of the incubation period and generation time for SARS-CoV-2 Alpha and Delta variants by analyzing comprehensive data collected during contact tracing activities in the province of Reggio Emilia, Italy, throughout 2021. We found no statistical difference for the duration of the incubation period for Alpha (mean: 4.9 days) and Delta (mean: 4.5 days) variants. Both estimates are very close to those reported for the same variants in a recent extensive meta-analysis (Alpha: 5.0 days; Delta: 4.4 days) [73] and in line (albeit slightly shorter) with those obtained for the ancestral lineages [66,71,74,75]. We did not evaluate the dependence between age and incubation periods for Alpha and Delta variants, which was previously evaluated for ancestral lineages [76].

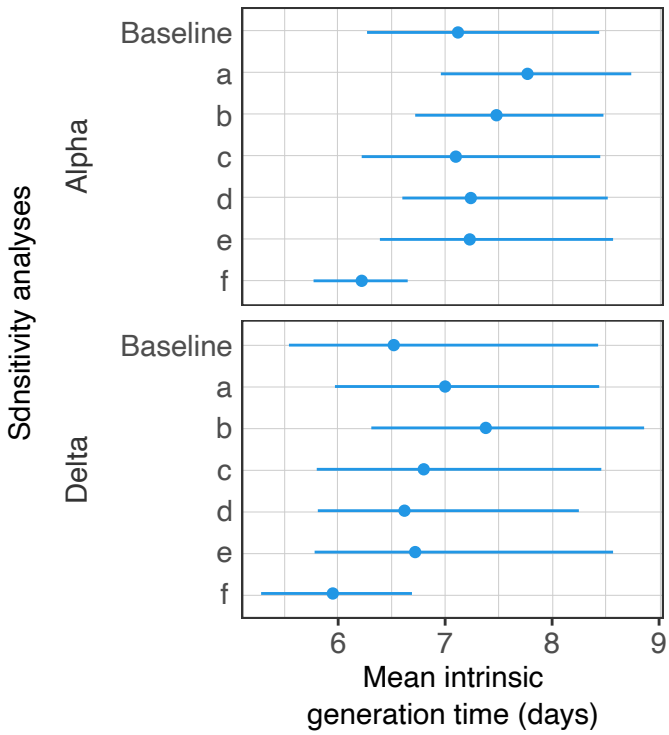

Figure 1-4. Estimates of the mean intrinsic generation time for the Alpha and Delta variants under different assumptions with respect to the baseline model: (a) uses an alternative method of imputation for the dates of infection in asymptomatic cases; (b) and (c) use different distributions of the incubation period, taken from previous estimates for ancestral SARS-CoV-2 lineages; (d) assumes a halved transmissibility for asymptomatic individuals; (e) considers the possibility of protection from previous natural infection in a fraction of undiagnosed individuals and (f) assumes no compliance of household members to quarantine. Full details on sensitivity analysis are reported in the Appendix A.

Our estimates of the mean generation time for both Alpha (mean: 7.12 days) and Delta (mean: 6.52 days) are compatible with previous estimates for ancestral lineages [77,78,79,80,81], including a previous estimate for Italy of 6.68 days [82]. We also found comparable household serial intervals between Alpha (mean: 2.43 days) and Delta (mean: 2.74 days) with similar proportions of pre-symptomatic transmissions (about 50% for both variants). Previous estimates on both ancestral lineages and Alpha and Delta variants are highly variable (due to the high sensitivity of these parameters to epidemiological conditions of the study settings) and ranged between 1.8 days and 7.5 days [66,67,71,79,83] for the serial interval and between 13% and 65% [66,67,75] for the proportion of pre-symptomatic transmission.

Estimates of the intrinsic generation time may depend on epidemiological specificities of the geographical setting from which the data are collected, as well as by the inference method. For instance, a study conducted in England estimated, using a different approach, a shorter mean intrinsic generation time for the Alpha (5.5 days) and Delta variants (4.7 days) [44]. Given its potential sensitivity to local factors, we point out the need to obtain country-specific estimates of the distribution of the generation time. For what concerns Italy, this study suggests the adequacy of epidemiological analyses (i.e., computation of reproduction numbers; modeling estimates) performed by assuming a distribution of the generation time similar to ancestral lineages.

A main strength of this work consists in the very large population-based dataset that comprehensively covers household clusters observed in the province of Reggio Emilia.

The protocol for tracing and testing contacts was the same in the Alpha and Delta periods. Thanks to efforts by public health officials, a high compliance to testing was achieved, with only 3% of individuals refusing to be tested; in addition, all household members of cases were tested on the same date of the first diagnosis in the household. To minimize the possibility that our data contain clusters due to other variants, we selected two periods where Alpha and Delta were largely dominant [66]. However, for the Alpha period a residual circulation (7-8% prevalence) of the Gamma variant was detected in the Emilia-Romagna region [65,84]. The estimates of the intrinsic generation time can be compared across periods with different vaccination coverage since the model includes susceptibility and transmissibility variations according to the individual's vaccination history. A limitation in the estimation of the incubation period was the use of censored interval data under the assumption that the date of infection was bounded by the date of last negative test and the date of last exposure, as both dates may suffer of potential biases. In some cases, the date of last negative test may be a too stringent limit for the date of first exposure, as a test can provide false-negative results if performed in the days immediately successive to the date of infection. On the other hand, the date of last exposure may be incorrectly recorded if a case broke from isolation/quarantine and did not report further contacts to the tracing team for fear of administrative fines. These biases may have an impact on the estimation of incubation periods. However, the closeness of our estimates to results of a recent extensive meta-analysis [73] we and the high level of collaboration and trust between the population and the contact tracing teams within this study (as witnessed by the high level of compliance to offered tests) suggest that such biases may be mild in our data.

A limitation of the model for the estimation of the generation time is its reliance on assumptions for the dates of infection of infected individuals. Ideally, these could be inferred as nuisance parameters in the model, but this is computationally unfeasible with the large number of cases within this study. Therefore, dates of infection were imputed multiple times based on the distribution of the incubation period [25]. The same intrinsic limitation of the unobservability of infection times is shared by all transmission chain reconstruction models, but there are now several examples where these models have been proven to correctly identify the transmission dynamics of infectious outbreaks [25,33,34,85,86]. Estimates were substantially robust with respect to different methods of imputation and different distributions of the incubation period (Figure 1-4 and Appendix A). Thus, potential biases in the estimate of the incubation period reported above are not expected to propagate to the generation time. A specific limitation of this study was the lack of information about previous SARS-CoV-2 infection in undiagnosed individuals. In the main analysis we assumed that all undiagnosed individuals did not have a pre-existing protection from natural immunity. However, in a sensitivity analysis, we show that assuming full protection from previous infection in a fraction of undiagnosed individuals hardly affects our results (Appendix A). Another specific limitation is that we assumed 100% compliance to quarantine protocols (i.e., that household members quarantined after diagnosis of another member could only be infected within the household). A sensitivity analysis where quarantines of household members are not considered (i.e., 0% compliance) yielded similar results to the ones illustrated in the main analysis (Appendix A).

Results from this study suggest that the length of the incubation period and generation time for Alpha and Delta variants were comparable to those of the ancestral lineages.

These findings provide support to the recommendations of adopting duration of quarantine, isolation, and contact tracing operations similar to those for the ancestral lineage. This work also confirms the suitability of the adopted method for estimating the incubation periods and generation times on further emerging variants of concern, provided that high-quality contact tracing data are available.

Appendix A

Estimation of the incubation period

For the estimation of the incubation period, we considered observations on individuals who were contacts of an index case and who later became symptomatic and diagnosed with SARS-CoV-2. Among these, we selected cases having a diagnosis in either of the selected study period for Alpha (1021 cases) or Delta (519 cases). The date of symptom onset and the date of last exposure was available for all symptomatic cases. For each case, the potential incubation period was bounded by the date of the latest negative test result before the diagnosis (earliest possible exposure) and by the date of the last exposure. By considering the latest negative test as a proxy for the earliest possible exposure, we assume the test to have a perfect sensitivity, i.e., we neglect possible negative false results which are especially probable in the earliest days after infection. We excluded 172 Alpha cases and 53 Delta cases for which the information on exposure were conflicting (e.g., last negative test successive to the last reported exposure), 298 Alpha cases and 173 Delta cases for which the date of last exposure was successive to symptom onset, and 358 Alpha cases and 204 Delta cases for which only a date of last exposure was available, obtaining 193 Alpha cases and 89 Delta cases for the main analysis (see Figure 1-5 for the sample selection). The resulting censored intervals of the possible incubation periods are reported for all cases in Figure 1-6.

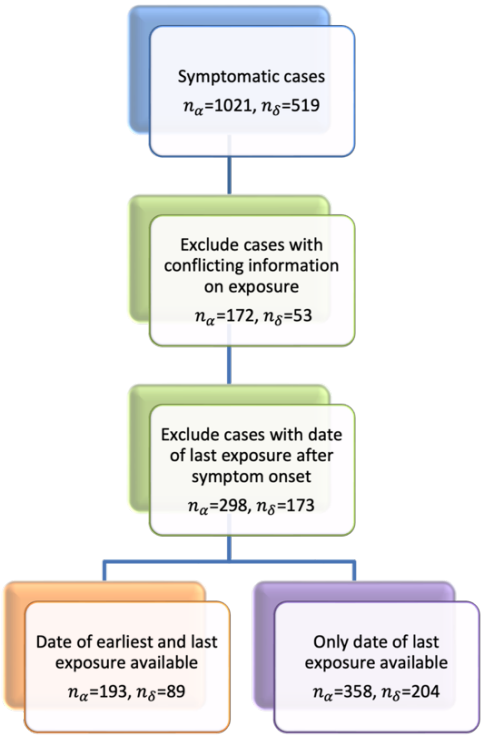

Figure 1-5. Workflow of sample selection. Gray boxes represent exclusion steps. The green box shows the sample used for the main analysis. The yellow box shows the additional sample used for a sensitivity analysis. n_α represents the sample size for Alpha variant, n_δ represents the sample size for Delta variant.

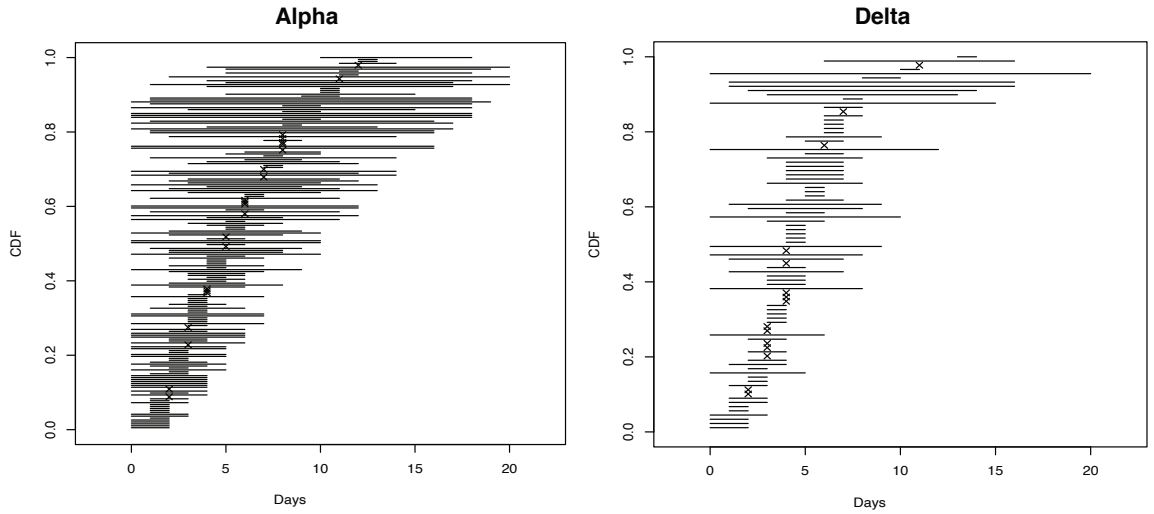

Figure 1-6. Incubation period censored data. Interval censored and non-censored observations for each case are ordered by their mid-points.

We estimated both Gamma and Weibull distributions using the censored data. Maximum likelihood estimations of the distribution parameters were calculated by using the *fitdistrplus* package in R. Direct optimization of the log-likelihood is performed using general-purpose optimization based on Nelder–Mead, quasi-Newton algorithm for both Gamma and Weibull distributions. Nonparametric bootstrap resampling was used to simulate uncertainty in the parameters of the estimated distributions. Results of the estimation procedure described in the main text are presented in Table 1-3.

Table 1-3. Estimated distribution of the incubation period. SD: Standard Deviation. AIC: Akaike Information Criterion.

Variant	Distribution	Parameters: mean (SD)	Mean distribution (days)			AIC score
			Mean	SD	2.5 to 97.5 percentile range	
Alpha (N=193)	Gamma	shape = 3.08 (0.39), rate = 0.63 (0.084)	4.9	2.8	1.0 – 11.7	506.9
	Weibull	scale = 5.52 (0.27), shape = 1.83 (0.13)	4.9	2.8	0.7 – 11.3	510.7
Delta (N=89)	Gamma	shape = 4.43 (0.76), rate = 0.99 (0.18)	4.5	2.1	1.3 - 9.6	261.3
	Weibull	scale = 5.09 (0.30), shape = 2.10 (0.18)	4.5	2.2	0.9 - 9.5	267.1

Sensitivity analysis

As a first sensitivity analysis, we added the observations with only the date of last exposure being available (see Figure 1-5). For cases with unknown date of earliest possible exposure, we set the maximum boundary of the incubation period to 21 days before the symptom onset. Figure 1-7 shows the censored data used in this estimation. This increased both the sample size and the uncertainty regarding the earliest possible exposure, naturally increasing the average of the estimated incubation period (Table 1-4). This shows the importance of considering only data samples for which information on the time window of exposure is more compelling. Then, we repeated the main analysis and the sensitivity analysis above after selecting cases falling within the Alpha or Delta period based on the date of symptom onset rather than on the date of diagnosis, obtaining similar results to the corresponding analyses above.

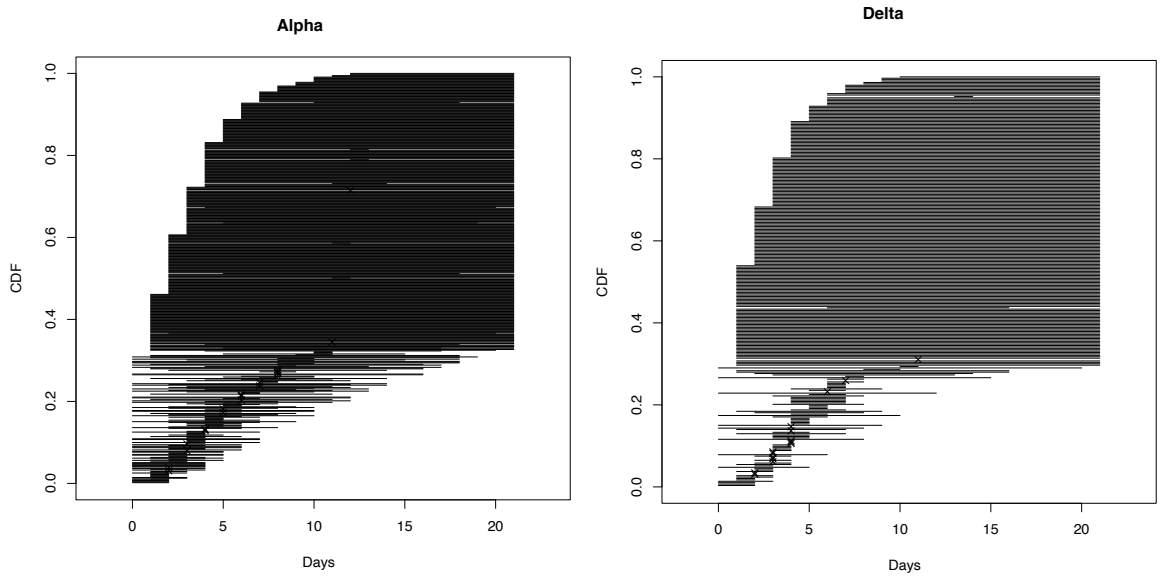

Figure 1-7. Incubation period censored data, including observations with no information on earliest possible exposure (sensitivity analysis A). Interval censored and non-censored observations for each case are ordered by their mid-points.

Table 1-4. Estimated distribution of incubation period in sensitivity analyses. SD: Standard Deviation. AIC: Akaike Information Criterion.

Variant	Distribution	Parameters: mean (SD)	Mean distribution (days)			AIC score
			Mean	SD	2.5 to 97.5 percentile range	
A) **Date of the last exposure available (independently from availability of earliest exposure)**						
Alpha (N=551)	Gamma	shape = 3.75 (0.31), rate = 0.51 (0.05)	7.3	3.8	1.9 - 16.6	770.9
	Weibull	scale = 8.33 (0.29), shape = 2.23 (0.12)	7.4	3.5	1.6 - 15.0	764.7
Delta (N=293)	Gamma	shape = 4.58 (0.60), rate = 0.73 (0.12)	6.3	2.9	2.0 - 13.2	376
	Weibull	scale = 7.07 (0.33), shape = 2.49 (0.19)	6.3	2.7	1.7 - 12.0	380.2
B) **As main analysis, but cases are assigned to variant via date of symptom onset**						
Alpha (N=187)	Gamma	shape = 3.12 (0.40), rate = 0.65 (0.087)	4.8	2.7	1.0 - 11.4	495.5
	Weibull	scale = 5.45 (0.27), shape = 1.84 (0.13)	4.8	2.7	0.7 - 11.1	499.2
Delta (N=89)	Gamma	shape = 4.70 (0.81), rate = 1.04 (0.19)	4.5	2.1	1.4 - 9.4	255.9
	Weibull	scale = 5.10 (0.30), shape = 2.14 (0.19)	4.5	2.2	0.9 - 9.3	263
C) **As sensitivity analysis A), but cases are assigned to variant via date of symptom onset**						
Alpha (N=546)	Gamma	shape = 3.71 (0.31), rate = 0.50 (0.049)	7.4	3.8	1.9 - 16.5	769
	Weibull	scale = 8.45 (0.29), shape = 2.21 (0.12)	7.5	3.6	1.5 - 15.4	763.1
Delta (N=286)	Gamma	shape = 4.75 (0.64), rate = 0.76 (0.12)	6.2	2.9	1.9 - 12.8	367.7
	Weibull	scale = 7.04 (0.32), shape = 2.51 (0.19)	6.3	2.7	1.6 - 12.0	373.3

Imputation of dates of infection

The task of reconstructing transmission chains must overcome the intrinsic limitation of the unobservability of transmission chains. We use available evidence to probabilistically impute plausible infection dates for all SARS-CoV-2 cases in our dataset. We combine observed dates of symptom onset, diagnosis, and negative test results with available knowledge on incubation periods and the probability of testing positive over time for infected individuals.

First, we impute the dates of infection for all symptomatic cases. Let T_D be the date of diagnosis (when the individual tested positive), $T_{N,n}$ the date of the n-th negative test before diagnosis, and T_S the date of symptom onset; we define the following probability P_I of being infected on day T_I to be proportional to the product of three probabilities:

- the probability of having an incubation period equal to $T_S - T_I$ days;
- the probability of testing positive at the date of diagnosis given infection at day T_I;
- the probability of testing negative (including false negatives) at all the dates of negative tests given infection at day T_I;

This can be summarized by the following equation:

$$P_I(T_I) = f(T_D - T_I) \cdot \prod_n \left[1 - f\left(T_{N,n} - T_I\right)\right] \cdot P_S(T_s - T_I) \quad (1.1)$$

where $f(t)$ is the probability of a SARS-CoV-2 case of testing positive after a time t since infection and $P_S(t) = \int_t^{t+1\ day} p_S(\tau)d\tau$ is the discretized version of the probability density function of the incubation period $p_S(t)$. For $p_S(t)$ we use the average variant-specific estimate from contact tracing data in Reggio Emilia defined by the algorithm above as a baseline, and two previous alternative estimates on ancestral lineages [67,71] as sensitivity analyses (see Section S5-b and S5-c). For $f(t)$, we use a previously estimated piecewise logistic function with one breakpoint [87], also discretized at intervals of one day. For each symptomatic case, a time of infection T_I is sampled from $P_I(t)$; note that this sampling allows for possible false negative results in dates $T_{N,n}$. The sample is repeated K = 100 times.

For asymptomatic cases, we cannot use the information on the incubation period given that no date of symptom onset is defined. Therefore, we use the imputed dates of infection for symptomatic cases to define a distribution of diagnostic delays $P_D(x)$, defining the probability of being diagnosed after x days from infection. An empirical approximation of $P_D(x)$ will be given, for any x, by the fraction of all instances across the K stochastic samples for which the diagnostic delay $T_R = T_D - T_I$ is equal to x. A gamma function p_D for the probability density function of the diagnostic delay is then fitted to the empirical distribution using a maximum likelihood approach and then discretized as above to obtain $P_D(x) = \int_x^{x+1\ day} p_D(\tau)d\tau$. The infection date of asymptomatic cases can then be sampled from the following probability

$$P_I(T_I) = f(T_D - T_I) \cdot \prod_n \left[1 - f\left(T_{N,n} - T_I\right)\right] \cdot P_D(T_D - T_I) \quad (1.2)$$

Equation (1.2) has the same rationale as that of Equation (1.1), except that instead of the

incubation period term we consider the probability of having a diagnostic delay equal to $T_D - T_I$, assuming that the distribution of diagnostic delays for asymptomatic cases is the same as for symptomatic cases. Because this assumption cannot be tested, we use as a sensitivity analysis an alternative method where only the probabilities of negative and positive tests are used to define the P_I (see Section S5-a). The sampling of infection times is repeated K times also for asymptomatic cases.

Assuming the imputation of incubation periods is correct, we obtain that 12.6% (95%Credible Intervals, CrI: 11.6-13.7%) of negative tests is a false negative result for the Alpha variant and 18.1% (95%CrI: 16.5-19.5%) for the Delta variant. This result was used to define the criterium of inclusion for households where undiagnosed cases have at least two negative tests, in order to reduce the fraction of undiagnosed positive cases to negligible levels (1.6% for Alpha and 3.3% for Delta) for the purpose of this analysis. Figure 1-8 reports the estimated empirical and fitted distributions of diagnostic delays for variants Alpha and Delta.

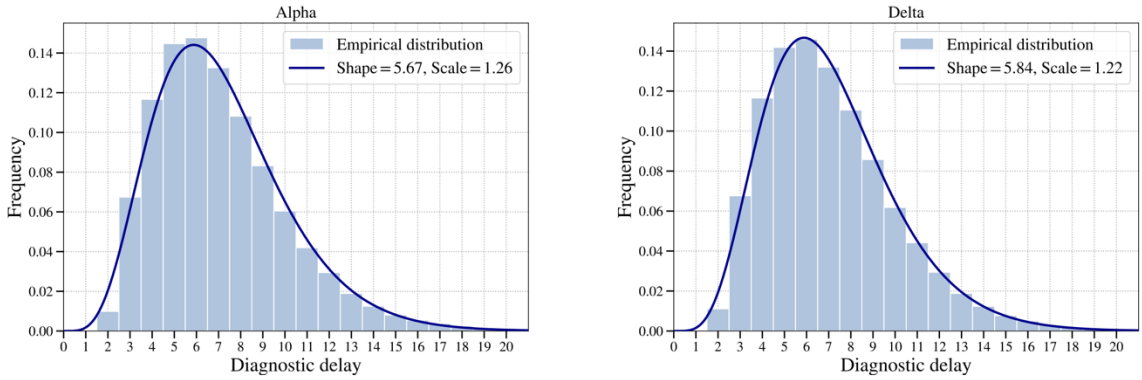

Figure 1-8. Empirical and fitted distribution of the diagnostic delay, P_D, estimated from symptomatic cases. Left: Alpha variant; Right: Delta variant. The histograms represent the empirical distribution given the imputed infection times for symptomatic individuals. The curves represent the fit of gamma functions.

Estimation of the generation time distribution and inference of transmission links

The model adopted in this work extends the approach previously proposed in [69]. We assumed that, at any time t, a susceptible individual j within a household is exposed to a force of infection composed of two components:

$$\lambda_j(t) = \lambda_j^o(t) + \lambda_j^h(t) \tag{1.3}$$

where $\lambda_j^o(t)$ represents the force of infection from the general community outside the household, and $\lambda_j^h(t)$ represents the one from infected members inside the household; this considers the possibility that an individual can be infected either within the household by one of its members or in the general community.

We now specify the two components. First, we describe the force of infection at time t from the general community $\lambda_j^o(t)$ as the forces of infection exerted at day t on the individual j by all potential infectors in the general community. We define the force of infection from the general community as given by the sum of the individual forces of infection from all cases that were infected at any day z before t. The force of infection from each candidate infector was proportional to the relative susceptibility of j according to his vaccination status and to the probability of infecting t-z days after infection; we also

considered the possibility that j was in quarantine/isolation at home at day t and therefore could not have contacts with the general community. Therefore, we define $\lambda_j^O(t)$ as follows:

$$\lambda_j^O(t) = \sum_{z \in 0..t} \psi \, Inc(z) \, \chi_j(t) \Gamma(t - z; a, b) q_j(t) \tag{1.4}$$

where:
- ψ is an unknown transmission rate from the general community;
- $Inc(z)$ is the number of newly infected cases at time z outside the household of j
- $\chi_j(t)$ represents the relative susceptibility of individual j and changes over time t depending on the dates of vaccination of j;
- $\Gamma(t; a, b)$ represents the distribution of the intrinsic generation time at day t after infection, for which we assumed a discretized Gamma distribution with scale a and shape b; in particular, given $g(t;\ a, b)$ the continuous Gamma probability distribution, $\Gamma(t; a, b) = \int_t^{t+1} g(\tau; a, b) d\tau$.
- the term $q_j(t)$ is an on/off function that is 0 when the household of j is in quarantine and 1 otherwise. For each household, a quarantine of 14 days is started after the first diagnosis and reinstated for a further 14 days every time there is a new diagnosis after the previous quarantine has ended.

Since $Inc(z)$ is unknown due to underreporting, we considered the epidemic curve by date of symptom onset for the province of Reggio Emilia in the Italian integrated surveillance system [113,63], $I(z)$, which is proportional to $Inc(z)$ via an unknown reporting parameter u: $I(z) = u \, Inc(z)$. Thus, Equation (1.4) becomes:

$$\lambda_j^O(t) = \sum_{z \in 0..t} \frac{\psi}{u} \, I(z) \, \chi_j(t) \Gamma(t - z; a, b) q_j(t) \tag{1.5}$$

Because ψ and u cannot be estimated at the same time due to their collinearity, we estimate a single free parameter α that is a scaling factor accounting for both underreporting of cases and the transmissibility from the general community. This allows to make the problem tractable at the cost of losing the interpretability on the estimated value of α.

Similarly, we describe the force of infection at time t from the general community $\lambda_j^h(t)$ as the sum of all forces of infection exerted on the individual j by each household members with an earlier date of infection. The force of infection $\lambda_{j,i}^h(t)$ from one household member i was proportional to the relative transmissibility of i at time t (according to the vaccination status of i), the relative susceptibility of j at time t (according to the vaccination status of j), and to the probability of transmitting t-$T_{I,i}$ after the date of infection $T_{I,i}$. Therefore, we define $\lambda_j^h(t)$ as:

$$\lambda_j^h(t) = \sum_{i \in H_j} \lambda_{j,i}^h(t) = \sum_{i \in H_j} \beta \rho_i(t) \chi_j(t) \Gamma\big(t - T_{I,i}; a, b\big) \tag{1.6}$$

where:

- i is an index running over the set H_j of infected household members of individual j;
- $\rho_i(t)$ represents the relative transmissibility of individual i, and changes over time t depending on the dates of vaccination of i;
- β is a free parameter scaling the transmissibility inside households.

For the relative susceptibility, we assumed that each dose may reduce the susceptibility to a given value 14 days after inoculation; protection of each dose starts to wane immediately, following an exponential function, increasing again the susceptibility over time. When the booster dose (third dose) is administered, we assume no waning (note that the booster dose started to be administered in Italy towards the end of the Delta study period, with only 3 individuals in our data having received it):

$$\chi_j(t) = \begin{cases} 1 & if\ t < t_{v,1} + 14 \\ 1 - \eta^{(1)}\, e^{-w(t-t_{v,1}-14)} & if\ t_{v,1} + 14 \le t < t_{v,2} + 14 \\ 1 - \eta^{(2)} e^{-w(t-t_{v,2}-14)} & if\ t_{v,2} + 14 \le t < t_{v,3} + 14 \\ 1 - \eta^{(3)} & if\ t \ge t_{v,3} + 14 \end{cases} \quad (1.7)$$

where $t_{v,d}$ is the date of vaccination dose d, $\eta^{(d)}$ are the initial effectiveness of dose d (i.e., 14 days after vaccination) against the considered variant, and w is the waning rate of vaccine protection. Estimates of vaccine effectiveness and waning rate were obtained from a large-scale retrospective cohort study on the Italian population [89,90] and reported in Table 1-5.

Table 1-5. Parameters for vaccine effectiveness and waning.

Parameter	Unit	Alpha	Delta
Initial effectiveness of dose 1 $\eta^{(1)}$	%	49.2	49.4
Initial effectiveness of dose 2 $\eta^{(2)}$	%	81.9	80.2
Effectiveness of the booster dose $\eta^{(3)}$	%	-	80.2
Waning rate *w*	days^{-1}	0	1/227

For the relative transmissibility, we assumed a reduction by ρ = 50% after 14 days from the first dose [91,92]:

$$\rho_i(t) = \begin{cases} 1 & if\ t < t_{v,1} + 14 \\ \rho & if\ t \ge t_{v,1} + 14 \end{cases} \quad (1.8)$$

The model assigns a source of infection k_j for all cases by choosing from either a generic source outside the household or from an infectious household member in H_j, with probability proportional to the contribution of each source to the total force of infection $\lambda_j(T_{I,j})$ at the time $T_{I,j}$ at which j was infected. The probability for an individual j of being infected at time $T_{I,j}$, L_j, is given by the product of the probability of being infected by the assigned source of infection on that day (P_j) and the probability of not being infected up until that day (Q_j). The overall likelihood of the observations given parameter set $\theta = (\alpha, \beta, a, b)$ and the assigned sources of infection k_j is given by the product of individual probabilities L_j:

$$L(\theta, k_j) = \prod_j L_j = \prod_j P_j Q_j \tag{1.9}$$

where

$$P_j = \begin{cases} \lambda_j^o(\mathrm{T_{I,j}}) & \text{if } k_j \text{ is outside the household} \\ \lambda_{j,i}^h(\mathrm{T_{I,j}}) & \text{if } k_j \text{ is household member } i \\ 1 & \text{if } j \text{ is uninfected} \end{cases} \tag{1.10}$$

For infected individuals, Q_j is the probability that j has not been infected until $t_{I,j}$, namely $Q_j = e^{-\int_0^{t_{I,j}} \lambda_j(t)dt}$. For uninfected individuals, it is the probability that j has never been infected, $Q_j = e^{-\int_0^{\infty} \lambda_j(t)dt}$.

We estimated the unknown parameters θ and the source of infection k_j for all cases using a Monte Carlo Markov Chain (MCMC) procedure. We considered uninformative prior distribution for all parameters (α: Uniform(10^{-8}, 10^{-4}); β: Uniform(0.1, 4); a: Uniform(0.1, 5); b: Uniform(0.1, 5). At each step, all parameters in θ are updated using reversible normal jumps. Z=500 samples from the posterior distributions obtained by the MCMC for each of the K=100 samples were pooled together to obtain the final parameter distribution and the distribution of the sources of infection for each case. Each sample of the joint distribution of the sources of infection constitutes a possible reconstructed transmission chain. Although the low values of prior bounds of α might suggest that importations from outside are limited, we also explored larger parameter regions. However, such initial explorations converged to degenerate distributions. The final posterior distribution of α had a mean of $7 \cdot 10^{-5}$ (95%CrI $6 \cdot 10^{-5} - 8 \cdot 10^{-5}$), that correspond to approximately 20% of infections being attributed to external sources for both the Alpha and Delta variants (see the next section).

The model for the inference of transmission links was implemented in C using GSL libraries (version 2.6) and compiled with GCC (version 4.2.1).

The code used to run the model is available at: figshare.com/IGT_alpha_delta .

Additional results of the baseline model

Table 1-6 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-6. Statistics on the posterior distributions of parameters for the intrinsic generation time in the baseline model.

Alpha	Shape variance	0.04
	Scale variance	0.08
	Covariance	-0.032
Delta	Shape variance	0.05
	Scale variance	0.07
	Covariance	-0.024

Statistics on reconstructed transmission links

Given the set of 50,000 reconstructed transmission chains, it is possible to compute

descriptive statistics on the number of infections acquired within or outside the household accounting for household size (Table 1-7). We obtained that the average per-household number of infections contracted from the general community was 1.18 (95%CrI 1.16 – 1.22) during the Alpha period and 1.11 (95%CrI 1.08 – 1.14) during the Delta period. The average number of secondary infections generated by a positive case was 0.64 (0.63 – 0.65) during the Alpha period and 0.60 (0.59 – 0.61) during the Delta period. Table 1-7 shows how the model reconstructed transmission links within households with different numbers of cases.

Table 1-7. Statistics for the model-based reconstruction of transmission links in households by number of SARS-CoV-2 cases. Reported numbers are the average and their 95% CrI, in bold the total number of households in the sample.

	Number	***%***	***Number***	***%***
Households with 2 SARS-CoV-2 cases	***1158***	**100**	***748***	**100**
- Both infected in the general community	*94 (75-121)*	8 (6-10)	*48 (35-63)*	6 (5-8)
- One infected the other	*1064 (1034-1083)*	92 (89-94)	*700 (685-713)*	94 (92-95)
Households with 3 SARS-Cov-2 cases	***611***	**100**	***338***	**100**
- All infected in the general community	*2 (0-6)*	0 (0-1)	*1 (0-3)*	0 (0-1)
- One transmission, 2 infected in the general community	*75 (58-94)*	12 (9-15)	*31 (20-42)*	9 (6-12)
- Two transmissions, same infector (1 generation)	*248 (221-277)*	41 (36-45)	*143 (121-167)*	42 (36-49)
- Two transmissions, different infectors (2 generations)	*286 (251-317)*	47 (41-52)	*163 (139-188)*	48 (41-56)
Households with 4 or more SARS-Cov-2 cases	***471***	**100**	***219***	**100**
- All infected in the general community	*0 (0-1)*	0 (0-0)	*0 (0-0)*	0 (0-0)
- One transmission	*58 (42-77)*	12 (9-16)	*29 (19-41)*	13 (9-19)
- Two transmissions	*241 (221-261)*	51 (47-55)	*117 (104-129)*	53 (47-59)
- Three or more transmissions	*171 (136-207)*	36 (7-44)	*73 (52-94)*	33 (24-43)

Stability of the attributed source of infection

For each case, we considered the distribution of the sources of infection attributed by the model through the reconstructed chains of transmission and evaluated its stability. We categorized cases according to whether its source of infection was consistently (i.e., more than 75% of the times over the Z sampling of infector and K sampling of infectious dates) attributed to:

- the same household member;
- transmission within household but from different potential infectors;
- transmission in the general community.

The setting of transmission was uncertain (less than 75% consistency in attribution) in about 40% of cases (Figure 1-9) in both the Alpha and Delta periods. This generally happened when two or more cases in a household had close diagnosis dates, so that either could have been infected in the general community and then transmitted to the other, or both could have been infected in the general community, depending on the assigned dates

of infection.

Sensitivity analyses

We performed six sensitivity analyses (SA) to test the robustness of model results against different model assumptions. The first three SA (*a-c*) impact on the main unknown of the data, i.e. the imputed infectious periods of cases; the fourth (*d*) considers a reduced transmissibility for asymptomatic individuals; the fifth (*e*) evaluates the possibility that a fraction of undiagnosed individuals were fully protected from infection from previous natural immunity: the sixth (*f*) assumes that any effort to quarantine positive cases would not impact the force of infection from outside the household (i.e., $q(t) = 1$ for any value of t in Equation (1.5)).

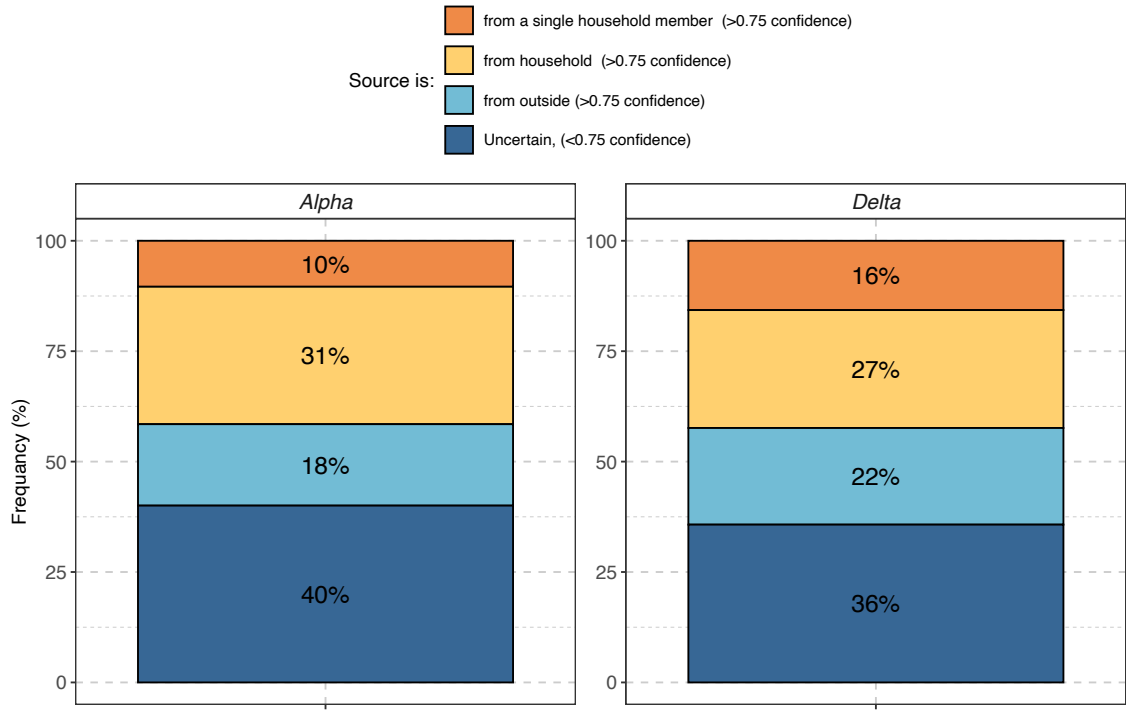

Figure 1-9. Consistency in the attribution of the infector or the infector setting. The stacked barchart represents the proportion of individuals that were consistently (more than 75% of the times across Z sampling of sources and K sampling of infectious dates) or inconsistently attributed to either category.

Imputation of dates of infection in asymptomatic cases (a)

In the baseline method for the imputation of dates of infection, we implicitly assumed that symptomatic and asymptomatic cases have the same diagnostic delay distribution. We assess the impact of this assumption by considering an alternative method where the date of infection of asymptomatic individuals was assigned only on the basis of information on diagnostic date and negative test results. In this additional procedure the factor depending on P_D is removed from Equation (1.2), resulting in:

$$P(j) = f(t_D - j) \cdot \prod_n [1 - f(n - j)] \qquad (1.11)$$

Table 1-8 and Figure 1-10 show that results obtained in this sensitivity analysis are in line with the baseline.

Table 1-8. Estimates for the intrinsic and realized generation time and serial intervals using an alternative method for the imputation of infection dates for asymptomatic individuals.

		ALPHA	DELTA
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.77 (6.96-8.74)	7 (5.97-8.44)
	shape mean (95%CrI)	2.43 (2.12-3.15)	2.33 (2.03-2.78)
	scale mean (95%CrI)	3.22 (2.62-3.65)	3.02 (2.4-3.69)
REALIZED GENERATION TIME	mean (95%CrI) [days]	5.08 (4.87-5.33)	4.39 (4.22-4.59)
SERIAL INTERVAL	mean (95%CrI) [days]	2.53 (2.37-2.72)	2.76 (2.64-2.88)

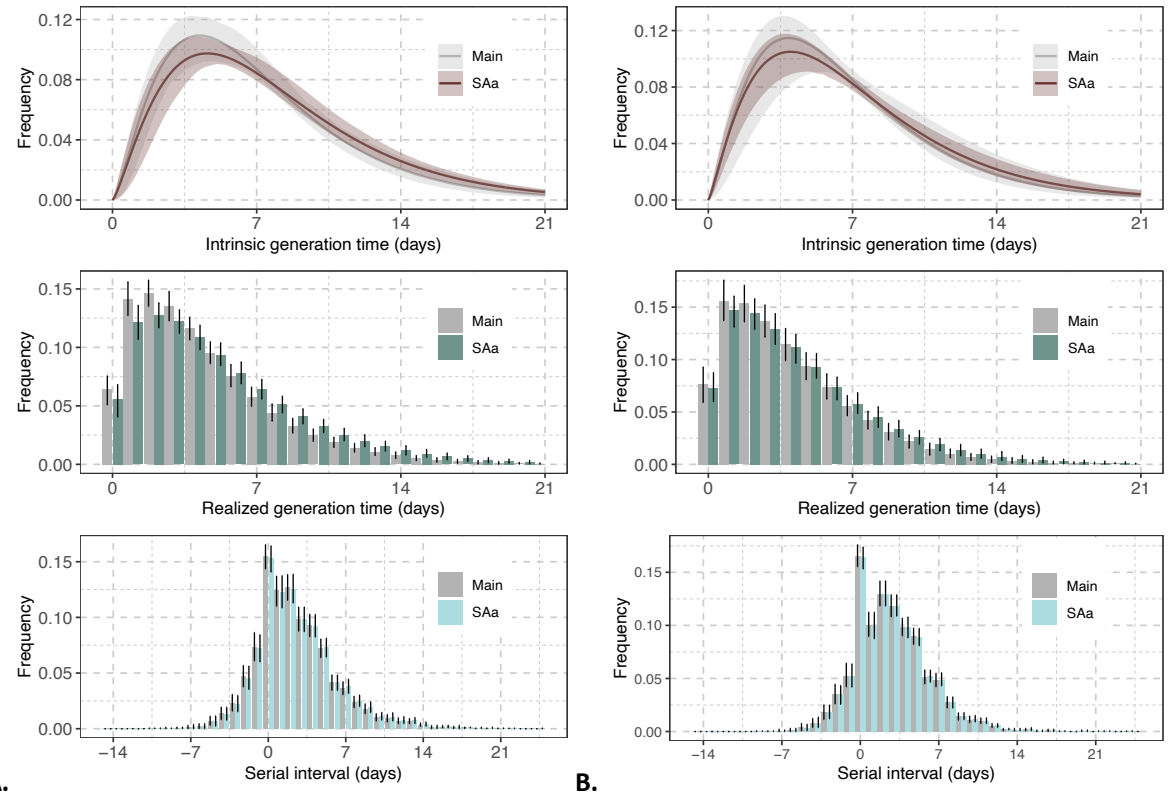

Figure 1-10. Comparison between baseline analysis and results obtained with sensitivity analysis a), using an alternative method for the imputation of infection dates for asymptomatic individuals. ***A****. Alpha variant.* ***B****. Delta Variant.*

Table 1-9 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-9. Statistics on the posterior distributions of parameters for the intrinsic generation time in sensitivity analysis a).

Alpha	Shape variance	0.06
	Scale variance	0.07
	Covariance	-0.048
Delta	Shape variance	0.03
	Scale variance	0.11
	Covariance	-0.036

Distribution of the incubation period – I (b)

In this sensitivity analysis, we reassigned infectious dates according to the baseline method, but using a different probability density function of the incubation period P_S in Equation (1.1). We considered a gamma-distributed estimate for Ps with shape 2.08 and

scale 3.03 as derived for ancestral lineages in [67]. Table 1-12 and Figure 1-11 show that results obtained in this sensitivity analysis are in line with the baseline.

Table 1-10. Estimates for the intrinsic and realized generation time and serial intervals using an alternative distribution of incubation periods estimated for ancestral lineages in [67].

		ALPHA	DELTA
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.48 (6.72-8.48)	7.38 (6.31-8.86)
	shape mean (95%CrI)	2.46 (2.19-3.03)	2.39 (2.06-2.92)
	scale mean (95%CrI)	3.05 (2.58-3.5)	3.1 (2.58-3.64)
REALIZED GENERATION TIME	mean (95%CrI) [days]	4.76 (4.60-4.92)	4.66 (4.45-4.88)
SERIAL INTERVAL	mean (95%CrI) [days]	2.14 (1.98-2.32)	2.28 (2.12-2.43)

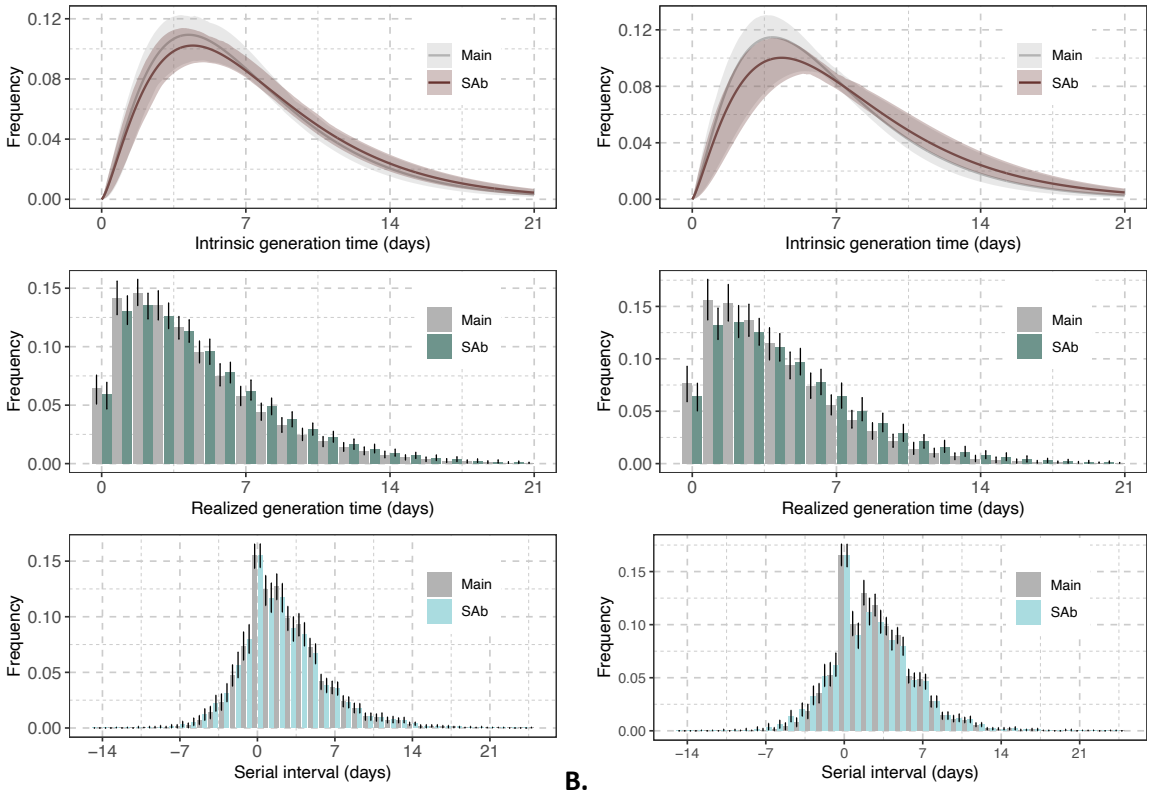

*Figure 1-11. Comparison between baseline analysis and results obtained with sensitivity analysis b), using an alternative distribution of incubation periods estimated for ancestral lineages in [S2]. **A**. Alpha variant. **B**. Delta Variant.*

Table 1-11 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-11. Statistics on the posterior distributions of parameters for the intrinsic generation time in sensitivity analysis b).

Alpha	Shape variance	0.03
	Scale variance	0.06
	Covariance	-0.026
Delta	Shape variance	0.05
	Scale variance	0.08
	Covariance	-0.031

Distribution of the incubation period – II (c)

Similarly to SA b), we considered a further alternative for the gamma-distributed estimate for Ps with shape 4.23 and scale 1.23, as derived for ancestral lineages in [69]. Table 1-12 and Figure 1-12 show that results obtained in this sensitivity analysis are in line with the baseline.

Table 1-12. Estimates for the intrinsic and realized generation time and serial intervals using an alternative distribution of incubation periods estimated for ancestral lineages in [69].

		ALPHA	DELTA
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.1 (6.22-8.45)	6.8 (5.8-8.46)
	shape mean (95%CrI)	2.51 (2.25-3.07)	2.45 (2.11-2.94)
	scale mean (95%CrI)	2.84 (2.35-3.42)	2.79 (2.27-3.46)
REALIZED GENERATION TIME	mean (95%CrI) [days]	4.39 (4.25-4.56)	4.22 (4.05-4.40)
SERIAL INTERVAL	mean (95%CrI) [days]	2.48 (2.35-2.62)	2.61 (2.46-2.72)

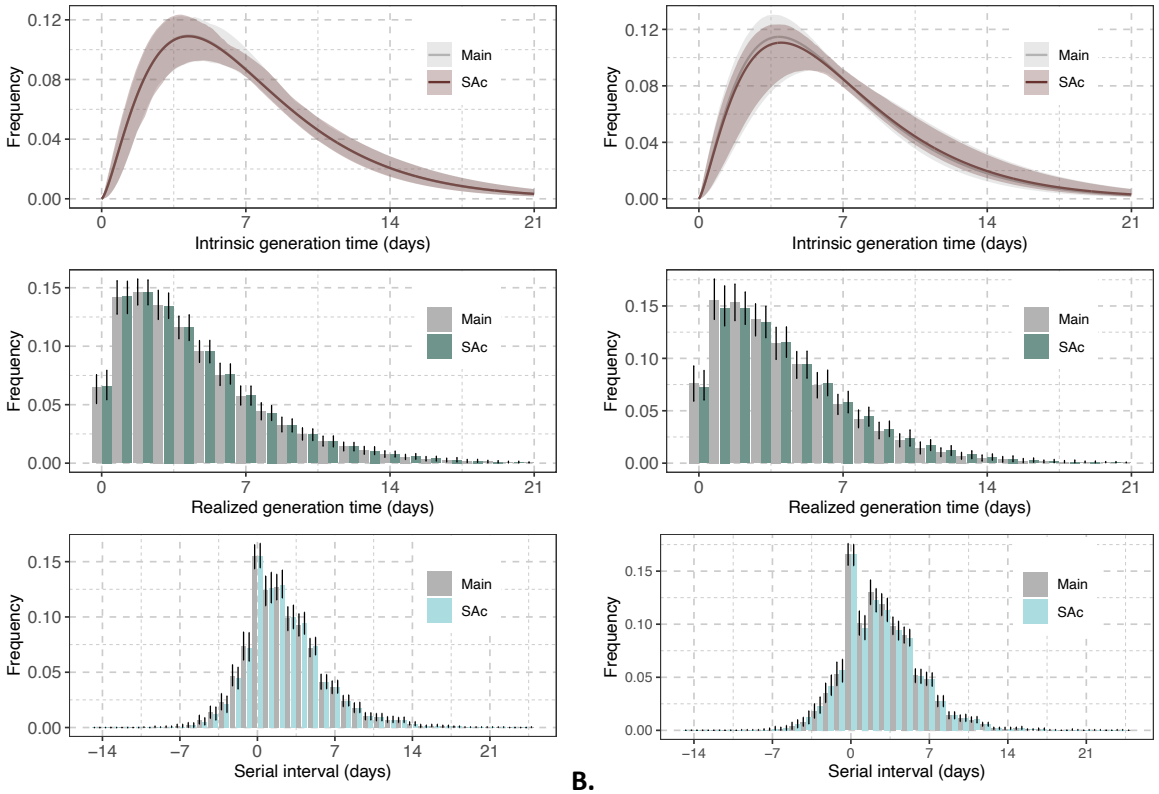

*Figure 1-12. Comparison between baseline analysis and results obtained with sensitivity analysis c), using an alternative distribution of incubation periods estimated for ancestral lineages in [69]. **A**. Alpha variant. **B**. Delta Variant.*

Table 1-13 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-13. Statistics on the posterior distributions of parameters for the intrinsic generation time in sensitivity analysis c).

Alpha	Shape variance	0.04
	Scale variance	0.08
	Covariance	-0.026
Delta	Shape variance	0.04
	Scale variance	0.10
	Covariance	-0.036

Reduced transmissibility for asymptomatic individuals (d)

In this sensitivity analysis, we consider a halved transmissibility for asymptomatic individuals [72] by modifying Equation (1.8) as follows:

$$\rho_i(t) = \begin{cases} \varphi_i & if\ t < t_{v,1} + 14 \\ \rho\varphi_i & if\ t \geq\ t_{v,1} + 14 \end{cases} \quad (1.12)$$

where φ_i is 1 if I is symptomatic and 0.5 if asymptomatic. Table 1-14 and Figure 1-13 show that results obtained in this sensitivity analysis are in line with the baseline.

Table 1-14. Estimates for the intrinsic and realized generation time and serial intervals using a halved transmissibility for asymptomatic individuals.

		ALPHA	**DELTA**
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.24 (6.6-8.52)	6.62 (5.81-8.25)
	shape mean (95%CrI)	2.56 (2.17-3.03)	2.43 (2.15-2.77)
	scale mean (95%CrI)	2.85 (2.39-3.43)	2.74 (2.29-3.36)
REALIZED GENERATION TIME	mean (95%CrI) [days]	4.51 (4.37-4.67)	4.06 (3.92-4.22)
SERIAL INTERVAL	mean (95%CrI) [days]	2.49 (2.37-2.62)	2.73 (2.62-2.85)

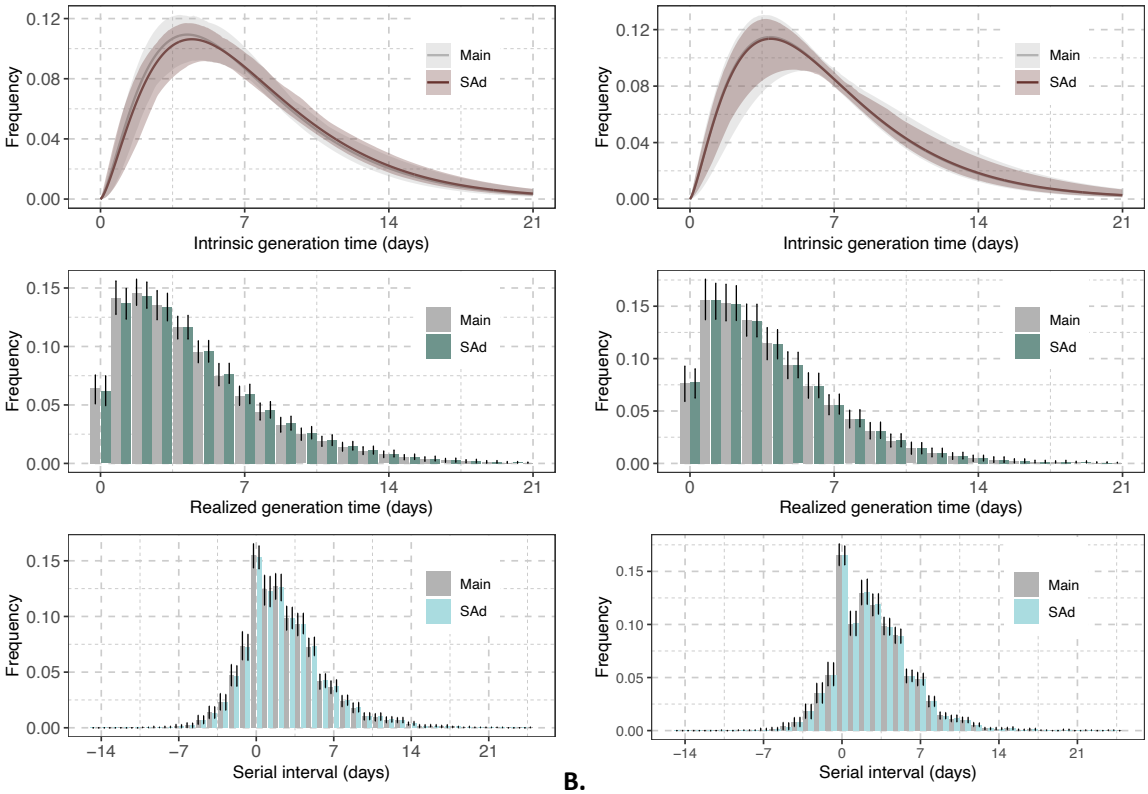

*Figure 1-13. Comparison between baseline analysis and results obtained with sensitivity analysis d), using a halved transmissibility for asymptomatic individuals. **A**. Alpha variant. **B**. Delta Variant.*

Table 1-15 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-15. Statistics on the posterior distributions of parameters for the intrinsic generation time in sensitivity analysis d).

Alpha	Shape variance	0.04
	Scale variance	0.07
	Covariance	-0.041
Delta	Shape variance	0.03
	Scale variance	0.07
	Covariance	-0.021

Protection from previous infection in a fraction of undiagnosed household members (e)

In this sensitivity analysis, we assume that a fraction of individuals who were undiagnosed were not susceptible to infection due to immunity conferred by previous SARS-CoV-2 infection. Using previous estimates of the cumulative SARS-CoV-2 attack rate in Italy before the Alpha and the Delta waves [93], we assume that 15% of undiagnosed household cases during the Alpha period and 20% of undiagnosed household cases during the Delta period were immune. These cases were randomly sampled and removed from set of j for each of the Z repetitions of the MCMC procedure. The absence of these cases impacts on the component of Q_j of the likelihood in Equation (1.9). Table 1-16 and Figure 1-14 show that results obtained in this sensitivity analysis are in line with the baseline.

Table 1-16. Estimates for the intrinsic and realized generation time and serial intervals when assuming that 15% of undiagnosed cases in the Alpha period and 20% of undiagnosed cases in the Delta period were protected from infection via natural immunity from previous infection.

		ALPHA	DELTA
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.23 (6.39-8.57)	6.52 (5.54-8.43)
	shape mean (95%CrI)	2.48 (2.26-2.87)	2.45 (2.13-2.87)
	scale mean (95%CrI)	2.92 (2.45-3.44)	2.75 (2.29-3.33)
REALIZED GENERATION TIME	mean (95%CrI) [days]	4.41 (4.27-4.56)	4.06 (3.89-4.25)
SERIAL INTERVAL	mean (95%CrI) [days]	2.43 (2.29-2.58)	2.75 (2.63-2.89)

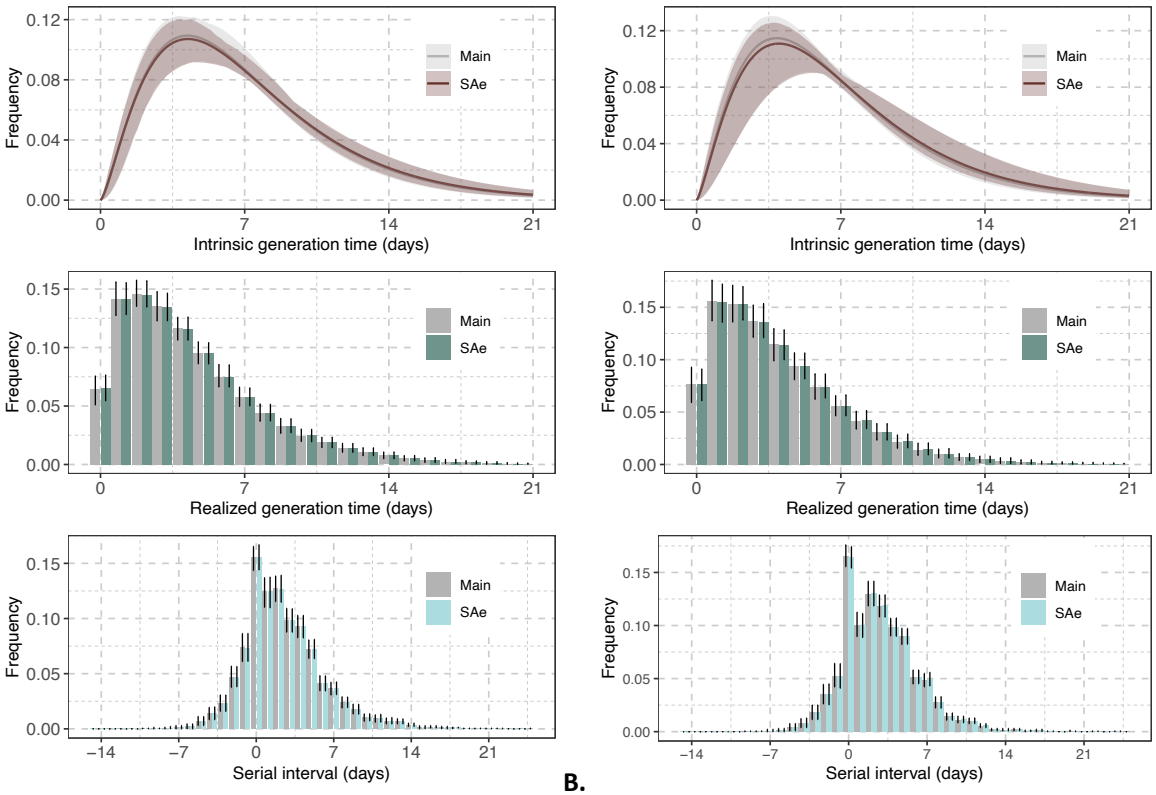

*Figure 1-14. Comparison between baseline analysis and results obtained with sensitivity analysis e), assuming that 15% of undiagnosed cases in the Alpha period and 20% of undiagnosed cases in the Delta period were protected from infection via natural immunity from previous infection. **A**. Alpha variant. **B**. Delta Variant.*

Table 1-17 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-17. Statistics on the posterior distributions of parameters for the intrinsic generation time in sensitivity analysis e).

Alpha	Shape variance	0.03
	Scale variance	0.07
	Covariance	-0.024
Delta	Shape variance	0.04
	Scale variance	0.08
	Covariance	-0.022

No protection from infection outside the household during quarantine (f)

In this sensitivity analysis, we assume that the imposed quarantine period after the first

positive diagnosis would not impact the force of infection from outside the household (i.e., $q(t) = 1$ for any value of t in Equation (1.5)). Table 1-18 and Figure 1-15 show that results obtained in this sensitivity analysis are in line with the baseline.

Table 1-18. Estimates for the intrinsic and realized generation time and serial intervals when assuming no protection from outside infection during the quarantine period.

		ALPHA	DELTA
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.22 (5.77-6.65)	5.95 (5.28-6.69)
	shape mean (95%CrI)	2.48 (2.24-2.75)	2.41 (2.15-2.73)
	scale mean (95%CrI)	2.51 (2.18-2.89)	2.48 (2.06-2.95)
REALIZED GENERATION TIME	mean (95%CrI) [days]	4.09 (3.96-4.21)	3.84 (3.70-3.97)
SERIAL INTERVAL	mean (95%CrI) [days]	2.25 (2.1-2.38)	2.61 (2.5-2.72)

Table 1-19 shows statistics on the posterior distributions of parameters for the intrinsic generation time.

Table 1-19. Statistics on the posterior distributions of parameters for the intrinsic generation time in sensitivity analysis f).

Alpha	Shape variance	0.02
	Scale variance	0.03
	Covariance	-0.021
Delta	Shape variance	0.02
	Scale variance	0.05
	Covariance	-0.025

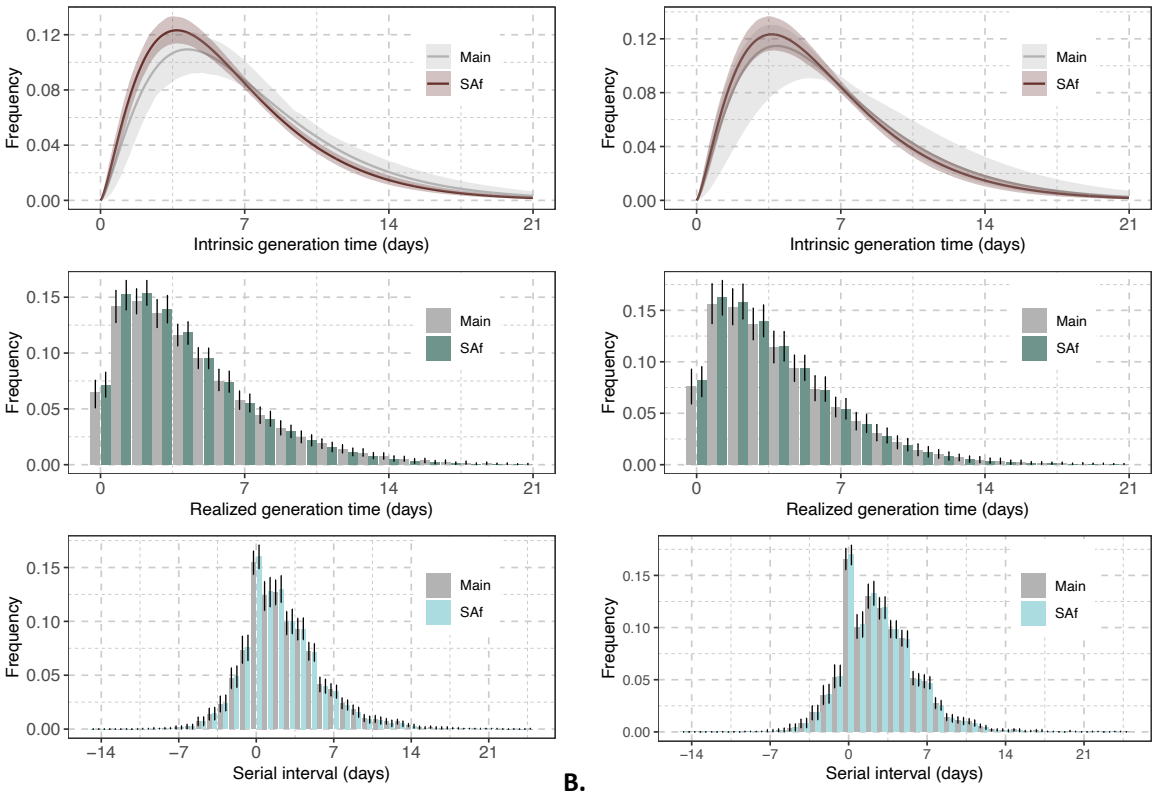

*Figure 1-15. Comparison between baseline analysis and results obtained with sensitivity analysis f), assuming no protection from outside infection during the quarantine period. **A**. Alpha variant. **B**. Delta Variant.*

CHAPTER 2

Intrinsic generation time of the Omicron SARS-CoV-2 variant

Introduction

The SARS-CoV-2 Omicron variant emerged at the end of 2021 and was able to completely replace the dominant variant Delta with a swiftness that was unprecedented compared to previously emerged lineages [94]. The exceptional fitness of Omicron is likely due to a combination of competitive advantages, including an increased transmissibility [95] and the ability to escape the immune response both from natural infection and vaccination [96,97]. The takeover of Omicron was accompanied by peak incidence values that were several times higher than the previous record in most countries of the world.

Early after its emergence, studies have indicated that Omicron may have shorter incubation period [98,99,100] and serial interval [98,99,101]. This observation, although based on preliminary evidence, has led many countries to abbreviate the duration of quarantine and isolation [102,103] in the attempt to contain the negative effect of high simultaneous absenteeism on the economy. However, whether the reduction in incubation periods and observed serial intervals reflects a reduction of the generation time, i.e., the time that elapses between the infection episode of an infector and of its infectee, is still to be determined. The generation time is an important parameter for monitoring and modeling infectious diseases. For example, if quarantine and isolation mandates are based on the generation time (as a proxy of infectiousness over time), an underestimation of this parameter would imply an early release of individuals when they still have a high probability to infect others, resulting in a reduced effectiveness of the intervention; on the other hand, an overestimation would imply an unnecessarily lengthy limitation of individual freedoms and increased absenteeism from school and workplaces, with impacts on the economy and the society. Biased estimates of the generation time also impact on the accuracy of the estimate of the net reproduction number, which is a key quantity for epidemiological surveillance, often used also as a parameter for deciding interventions. Finally, the generation time is largely used in scientific research (including mathematical modeling), for applications such as evaluating the effectiveness of interventions, understanding epidemiological dynamics, reconstructing transmission chains.

A direct measure of the generation time cannot be obtained empirically because the infection episodes in a chain of transmission are generally unobservable. Even when some certainty can be attributed to the dates of infection for pairs of infectors-infectee via detailed epidemiological investigation, the observed ("realized") generation times may be biased by the specific transmissibility conditions, including the structure of the study population's contact network, individual behaviors, environmental determinants, and control measures put in place [40]. For example, it is known that the generation time realized in households is remarkably shortened with respect to the one observed in the general community, due to the depletion of susceptible individuals and the competition of simultaneously infectious individuals to find susceptible household members to infect [40]. In contrast to the realized generation time, occurring in a realistic network of contacts, the "intrinsic" generation time represents the generation time that would be observed in a fully susceptible, homogenously mixed population [41]. The intrinsic generation time is therefore less dependent on the specific conditions of the epidemiological setting from which it is inferred but must be estimated with the use of quantitative inference techniques.

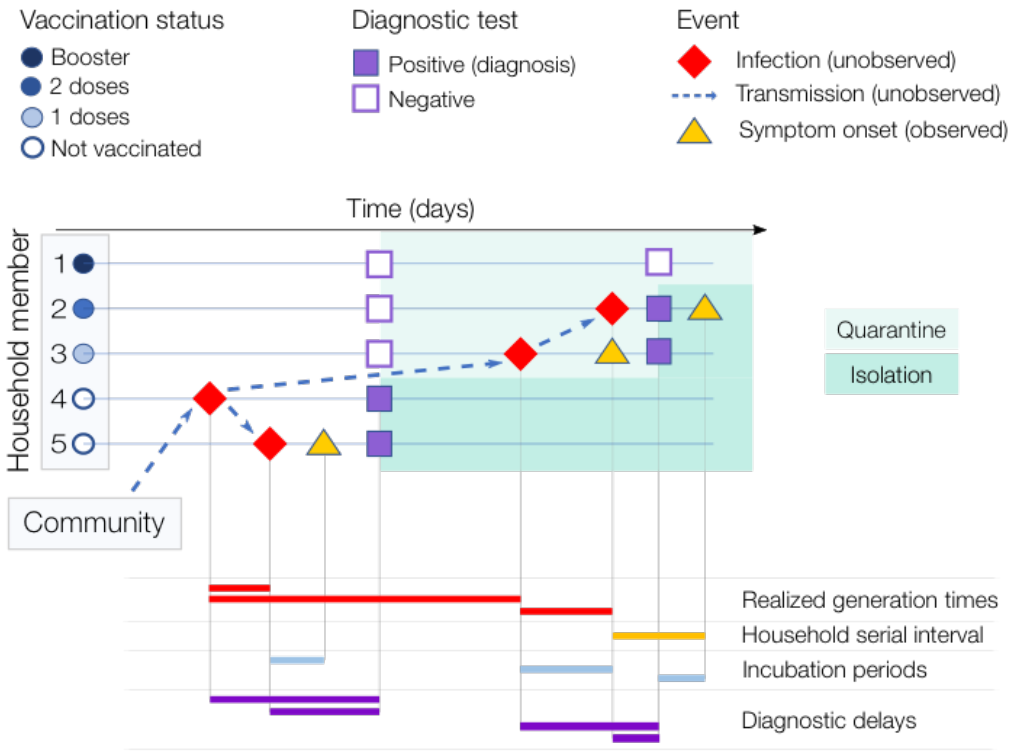

Figure 2-1. Illustrative example of a household cluster. A household with 5 members, of which #4 (asymptomatic) was infected outside the household (in the general community) and then transmitted to cases #5 and #3 (both symptomatic). Case #3 infected #2 while #1 remained uninfected. . #3, #2 and #1 were vaccinated with 1 dose, 2 doses, and 2 doses + booster respectively. In the bottom part of the figure, we show examples of the temporal intervals of interest for this work. Note that for the household serial interval and the realized household generation time, the source of infection (whether from outside the household or from a household member, and, in the latter case, which household member) is also unobserved and needs to be probabilistically reconstructed. The intrinsic generation time is not displayed as it represents the distribution of generation times among infections occurring in the general population in a fully susceptible population [41].

In this study, we collected the data for 23,122 SARS-CoV-2 infected individuals clustered in 8,903 households as determined from contact tracing operations in Reggio Emilia, Italy, between January 1st and January 31st, 2022. We then leveraged a Bayesian inference approach to estimate the distribution of the generation time (both intrinsic and

realized), realized household serial interval, and contribution of pre-symptomatic transmission for SARS-CoV-2 Omicron variant.

Methods

Data

To mitigate the spread of SARS-CoV-2, contact tracing activities were carried out in the province of Reggio Emilia, Italy, throughout the duration of the pandemic. Identified COVID-19 cases occurring in the province were confirmed via a Polymerase Chain Reaction (PCR) assay, reported in real time to the public health service of the Reggio Emilia local health authority, and isolated at home until either a negative PCR test result or 21 days were elapsed. During the study period (from January 1st to January 31st, 2022), both antigenic positive tests and PCR tests were considered for COVID-19 diagnosis. Household members were tested and quarantined at home. After 5 days (for individuals who were fully vaccinated more than 4 months before the date of contact) or 10 days (for unvaccinated individuals), a PCR test was performed and in the case of a negative result their quarantine was ended. If the test was not taken, household members were quarantined for 14 days, as per national guidelines [104]. Contacts with a booster dose or complete vaccination cycle or recovered from a previous SARS-CoV-2 infection in the 4 months before the date of contact were not subject to quarantine and were tested only upon development of symptoms. Compliance with at least one of the tests proposed by the public health service was 97.4% during the study period.

Data on test results and symptom onset dates (if applicable) for all identified cases and their contacts were linked to individual records on vaccination history (first, second, and booster dose). Appropriate data quality checks were conducted in strict collaboration with the Reggio Emilia local health authority to minimize missing information and accurately define household clusters. A household cluster was defined as households with at least two positive individuals with a diagnosis spaced less than 14 days apart.

A random sample of new diagnoses was characterized for the viral variants of SARS-CoV-2. Viral RNA was extracted from nasopharyngeal swab and specimens were screened by a commercial multiplex RealTime PCR assay (SARS-CoV-2 Variants I and II Allplex kit, Seegene; Seoul, South Korea) detecting L452R, W152C, K417T, K417N, E484Q, E484K, and N501Y mutations and HV69/70 deletion, and able to identify SARS-CoV-2 variants Alpha, Beta, Gamma, Delta, Epsilon, Omicron BA.1 and Omicron BA.2.

Ethics

The collection of data used for this manuscript (surveillance and contact tracing data) is compulsory in Italy according to national laws on infectious diseases. The COVID-19 Italian National Working group on Bioethics has stated that consensus for the collection of this data in the context of the COVID-19 emergency is not mandatory (Rapporto ISS COVID-19 n. 34/2020), based on Guideline 12 of the WHO on ethical issues in public health surveillance. The legal ordinance n. 640 of February 28 2020, explicitly declares Istituto Superiore di Sanità as entitled to collect data for COVID-19 surveillance and contact tracing and that such data can be used and shared, upon anonymization, to advance scientific knowledge on this new disease.

Estimation of the incubation period

The incubation period was estimated using data from a superspreading event occurred on November 26, 2021, in Norway, where 81 individuals were infected with the Omicron variant at a company's Christmas dinner, 80 of which became symptomatic [98]. We fitted a gamma distribution to the empirical distribution of incubation periods, and a nonparametric bootstrap resampling to assess uncertainty in the parameters (see Appendix B for details).

Table 2-1. Descriptive statistics of SARS-CoV-2 cases in the household dataset.

Period	**JANUARY 1 – 31, 2022**
Number of cases	23,122
Clinical outcome (%):	
Symptomatic	9,637 (41.7%)
Asymptomatic	13,465 (58.3%)
Gender (%):	
Male (%)	11,142 (48.2%)
Female (%)	11,980 (51.8%)
Age group (%):	
0-15 years old	6,138 (25.6%)
16-44 years old	9,396 (39.1%)
45-64 years old	5,952 (24.8%)
65+ years old	2,532 (10.5%)
Vaccination status at the end of the period (%):	
1 dose	1,132 (4.9%)
2 doses	10,175 (44.0%)
3 doses	4,651 (20.1%)
None	7,164 (31.0%)
Number of households	8,903
Mean household size (95% quantile)	2.70 (2 – 5)

Estimation of the generation time and serial interval

For the estimation of the generation time, we selected only household clusters for which all dates of diagnosis were included between January 1 and January 31, 2022. To reduce the possibility of missed diagnoses in the households, we further selected households for which undiagnosed members had at least one negative test result. We extended a Bayesian inference model for the reconstruction of transmission links in households [69,105]. The model exploits the temporal information on SARS-CoV-2 infections recorded in the dataset to probabilistically identify, for every infection, the likely source of infection (from outside the household or from a specific household member). Parameters for the generation time, which we assume to be gamma-distributed, are simultaneously calibrated via a Markov Chain Monte Carlo approach where the likelihood of the observed data is defined mechanistically through the computation of the force of infection to which all individuals are subject over time. The force of infection includes information on the temporal incidence of cases in the general population, on the date of infection and vaccination history for any individual and on previous infection from other variants. For each symptomatic case, the date of infection was imputed by subtracting the time of symptom onset by a randomly sampled incubation period from the estimated distribution. The imputed dates of infection for symptomatic individuals defined a distribution of delays between infection and diagnosis (diagnostic delay distribution), which was used to impute the date of infection of asymptomatic individuals starting from their date of diagnosis. For both symptomatic and asymptomatic individuals, we set to zero the probability of imputed dates of infection that preceded the latest negative test

result. The sampling of infection dates was repeated 100 times and the Bayesian model was re-calibrated on each resampling. Credible intervals (CrI) for the estimated parameters were obtained from the 95% percentile of the resulting pooled distributions.

The inferred transmission links allowed us to estimate, in addition to parameters of the intrinsic generation time, the distribution of the realized household generation time. We also estimated the distribution of the household serial interval from the difference of symptom onset dates in each infector-infectee pair (as inferred by the model) where both are symptomatic. Figure 2-1 schematizes a potential household cluster, with an indication for each individual of the dates of infection, symptom onset, diagnosis and negative tests, and summarizes the relevant quantities for the purpose of the study. A full description of the Bayesian inference model is available in the Appendix B.

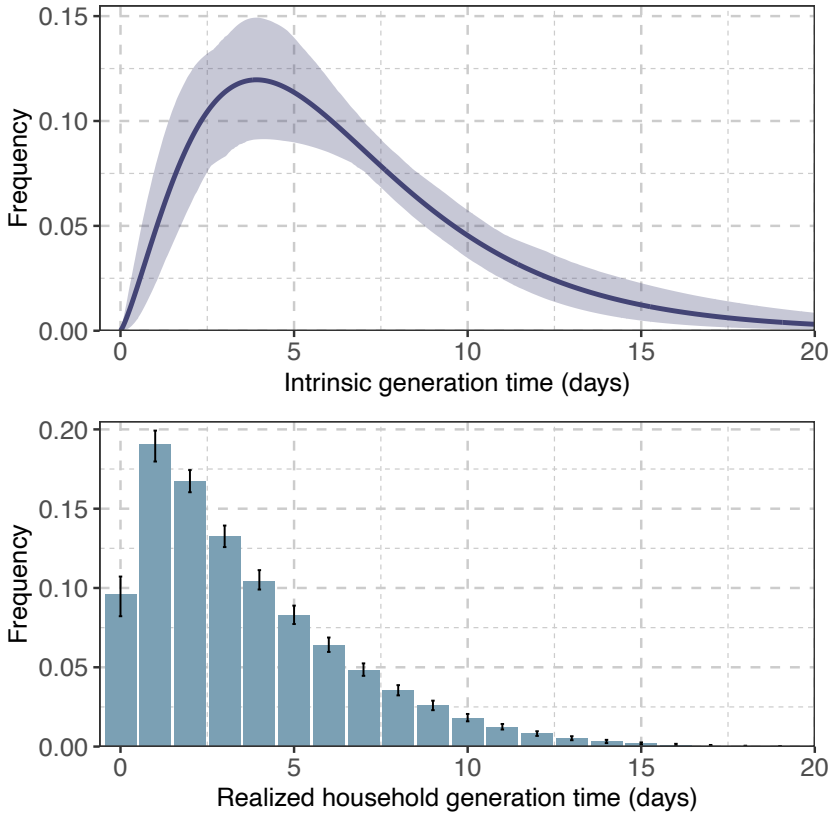

Figure 2-2. Estimates of the generation time for the Omicron variant. A) Distribution of the intrinsic generation time; solid line: mean estimate; shaded area: 95% CrI; B) Distribution of the realized household generation time; bars: mean estimate; vertical lines: 95% CrI.

Sensitivity analysis

To test the robustness of our results, we conducted an extensive set of sensitivity analyses where we considered: a) the subset of 380 households (1,127 cases in total) for which a case was genotyped as Omicron; b) the subset of 1,148 households (2,770 cases in total) for which all individuals were unvaccinated; c) the estimated incubation period of the Delta variant (mean: 4.5 days; standard deviation: 2.1 days) [105] to reassign the imputed infectious dates (baseline: Omicron variant with mean 3.5 days and standard deviation: 1.2 days); d) a diagnostic delay for asymptomatic cases that was 50% longer than that for symptomatic cases, implemented by increasing the shape of the gamma distribution by 50% (mean: 7.58 days; standard deviation: 1.61 days); e) similar to d), but implemented by changing the scale of the gamma distribution by 50% (mean: 7.58 days; standard

deviation: 1.97 days); f) the possibility of false negatives for negative test results when imputing infection dates; g) a halved transmissibility for asymptomatic individuals (baseline: equal to symptomatic individuals); h) a halved transmissibility for vaccinated individuals (baseline: equal to unvaccinated individuals); i) a scenario in which any effort to quarantine positive cases would not impact the force of infection from outside the household, which corresponds to the extreme case where there is 0% compliance to the policy (baseline: 100% compliance); j) that previous infection from other variants provides no cross-protection against Omicron (baseline: 56% cross-protection [106]).

Results

The study considered 8,903 households with mean size of 2.7 (standard deviation: 1.05, 95% quantile: 2–5) and a total of 23,122 diagnosed infections diagnosed between January 1st and January 31st, 2022. Of these, 9,637 (41.7%) were symptomatic and 11,980 (51.8%) were among women (see Table 2-1). A significant proportion of cases included in the study were unvaccinated (7,164, corresponding to 31%) and only 4,651 (20%) had received a booster dose before the end of the study period, compared to national statistics on the vaccination status of the Italian population on January 31 (17% unvaccinated, 56% with a booster dose [107]). Further descriptive statistics on the data are provided in Table 2-1.

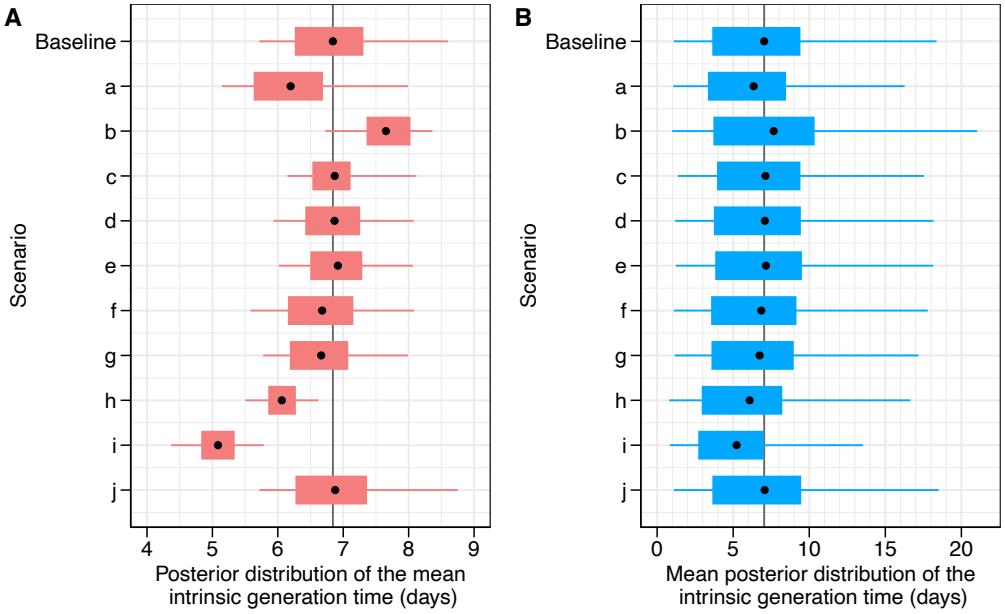

Figure 2-3. Estimates of generation times for the Omicron variant under different sensitivity analyses. A) Posterior distributions of the mean intrinsic generation time; B) Mean distributions of the intrinsic generation time. Point: mean value; box: interquartile range; whiskers: 95% CrI. The labels on the y-axis represent the performed sensitivity analysis to evaluate the robustness of baseline model results against different model assumptions where we consider: a) only households genotyped as Omicron; b) only household composed of unvaccinated individuals; c) an incubation period for Omicron with the same distribution as previous estimates for Delta (mean: 4.5 days; standard deviation: 2.1 days) [77]; d) a prolonged diagnostic delay for asymptomatic individuals (mean: 7.58 days, standard deviation: 1.61 days); e) a prolonged diagnostic delay for asymptomatic individuals (mean: 7.58 days, standard deviation: 1.97 days); f) the possibility of false negative tests; g) a halved transmissibility for asymptomatic individuals; h) a halved transmissibility for vaccinated individuals; i) a scenario where any effort to quarantine positive cases would not impact the force of infection from outside the household; j) previous infection from other variants provides no cross-immunity against Omicron infection.

From the analysis of symptom onset data of 74 individuals participating to a superspreading event in Norway [98], we estimated a gamma-distributed incubation

period of mean: 3.49 days (standard deviation: 1.20 days, 95%CI: 3.19−3.77), see Table 2-2. By leveraging the estimated distribution of incubation period, we estimated a distribution of delays between infection and diagnosis having a mean of 5.05 days (standard deviation: 1.31 days, 95% quantile: 3−7 days) (Table 2-2) for symptomatic subjects. By applying the Bayesian inference model, we estimated a mean intrinsic generation time of 6.84 days (95% CrI of the mean: 5.72−8.60 days) and a mean realized household generation time of 3.59 days (95% CrI of the mean: 3.55−3.60 days) (Figure 2-2 and Table 2-2).

Table 2-2. Estimates for the incubation period, diagnostic delays, intrinsic and realized generation time, and household serial intervals of SARS-CoV-2 Omicron Variant. Reported parameters of shape and scale for the incubation period and intrinsic generation time refer to a gamma distribution. Estimates of the incubation period are derived from the analysis of 80 participants to a single superspreading event in Norway. Data taken from [98].

	MEAN (95%CRI) [DAYS]	**3.49 (3.19-3.77)**
INCUBATION PERIOD	95% quantile of the mean distribution [days]	2-6
	shape mean (95%CrI)	8.50 (6.14-13.20)
	scale mean (95%CrI)	0.41 (0.25-0.68)
	standard deviation of the mean distribution [days]	1.20
DIAGNOSTIC DELAY FOR SYMPTOMATIC INDIVIDUALS	mean (95% quantile) [days]	5.05 (3-7)
	standard deviation [days]	1.31
INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.84 (5.72-8.60)
	95% quantile of the mean distribution [days]	1-17
	shape mean (95%CrI)	2.39 (2.01-3.34)
	scale mean (95%CrI)	2.95 (1.81-4.25)
	standard deviation of the mean distribution [days]	4.48
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.59 (3.55-3.60)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.38 (2.30-2.47)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	51.1 (45.5-55.7)

The robustness of these estimates was tested against several sensitivity analyses regarding selected subsets of the sample (sensitivity analyses a and b), alternative imputation methods for infection times (c-f) and alternative modeling assumptions (g-j; see Appendix B for full details). All sensitivity analyses yielded comparable results with respect to the distribution of the intrinsic generation time (Figure 2-3), with 95% confidence intervals broadly overlapping with the baseline estimate, except for a significantly shorter mean estimate (5.09 days) obtained when assuming no compliance to household quarantines. The longest mean realized household generation time (3.96 days) was estimated under the assumption of an incubation period equal to the one estimated for Delta, while the shortest (3.24 days) was estimated when considering only unvaccinated individuals.

The mean household serial interval in the baseline analysis was 2.38 days (95%CrI of the mean: 2.30−2.47 days), with 51.1% (95%CrI: 45.5−55.7%) of transmission episodes being pre-symptomatic (i.e., secondary cases transmitted by cases who would develop symptoms after the transmission episode). The mean household serial intervals estimated in sensitivity analysis ranged between 1.89 and 2.38 days (Appendix B), while the mean proportion of pre-symptomatic transmission ranged between 51% and 59%, comparable to the baseline estimate.

Discussion

We analyzed comprehensive data collected during contact tracing activities on over 23,000 SARS-CoV-2 cases distributed in about 9,000 households from the province of Reggio Emilia, Italy, between January 1 and 31, 2022. Our estimate of the mean generation time (mean: 6.8 days) is compatible with previous estimates for ancestral lineages [77] (including a previous estimate for Italy of 6.7 days [82]). Existing estimates for Alpha and Delta were in the same range in a study similar to the present one on the same study population in Italy (6.0 and 6.6 days respectively [105]). We also found a mean household serial interval of 2.38 days, shorter than previous estimates for Delta of 2.56 on a similar study population [105]. Available studies have suggested a shorter generation time of Omicron (between 50% and 80% the one of Delta) using population-level data on the growth rate of Omicron relative to Delta in Denmark [45] and United Kingdom [46]. An analysis of 43 infector-infectee pairs from contact tracing data in Hong Kong [47] estimated a mean realized generation time of 2.38 days (95% confidence interval 2.01−2.80) under very strict control measures (population-wide screenings and quarantine imposed to both contacts and contacts of contacts) that are known to reduce the realized generation time. Generalization of epidemiological estimates to different geographic contexts and conditions (and therefore their direct comparison) always needs to be made with caution. We believe that the provided estimate may be representative for places with similar socio-economic conditions and housing structure. For example, the mean number of residents per housing unit is similar between Reggio Emilia and the rest of Italy (Reggio Emilia: 2.44; Italy: 2.42; national range across NUTS1 aggregations: 2.31−2.67) and similarly for the mean number of residents per room (Reggio Emilia: 0.55; Italy: 0.57; national range across NUTS1 aggregations: 0.54−0.63) [108]. However, other factors such as seasonality in transmission, mitigation measures, testing efficiency, or the progression of the vaccination campaign may make direct comparison of estimates performed at different times problematic, even when they come from the same study population.

The result that the intrinsic generation time of the Omicron variant in Italy is not significantly shorter than previous lineages may be surprising with respect to the intuition suggested by repeated observations of shorter incubation periods [98,99,100] and serial intervals [98,99,101] (the latter also confirmed by this study). Realized serial intervals in households and other small-population settings such as schools, workplaces, hospital wards and nursing homes, may be a biased proxy for the intrinsic generation time since they depend strongly on the epidemiological conditions of the study population [41,40]; in particular, they tend to be shorter when transmissibility is higher (as in the case of the Omicron variant) because the competition for susceptible individuals is stronger [40]. For example, simulating transmission in households through a simple generative model where we imposed a mean generation time of 6.9 days, the

mean realized generation time in households turned out to be 4.7 days because of this competition effect (see Appendix B). On the other hand, the incubation period only reflects a clinical condition (development of symptoms) that is known to be poorly correlated to infectiousness for COVID-19, given the large proportion of infections transmitted by asymptomatic and pre-symptomatic individuals. The duration of viral shedding is likely a better biological proxy of the intrinsic generation time, as it is more closely related to the intrinsic infectiousness of an infected individual [109]. Several studies have found a similar duration of viral shedding for Omicron and other variants [110,111,112], in agreement with the conclusions from our study.

A main strength of this work consists in the very large population-based dataset that comprehensively covers household clusters observed in the province of Reggio Emilia. Public health officials made efforts to have high compliance to testing policies (97.4% of individuals who were offered a test accepted at least once), including testing all household members of cases at the date of the first diagnosis in the household. However, the following limitations should be taken into consideration to interpret our results. First, the main analysis relied on cases diagnosed during the study period. The prevalence of the Omicron variant in the region (Emilia Romagna) was 80% among infections diagnosed on January 3, 2022, rising to 99% among infections diagnosed on January 17 and 31 [65]; more specifically, over 97% of all infections diagnosed in Italy on January 31 were classified as Omicron sublineage BA.1 [65]. Therefore, it cannot be excluded that a minority of cases in our sample belonged to other variants. However, a sensitivity analysis performed on 380 households (1,127 cases) for which a case was genotyped as Omicron yielded compatible results. Second, the model relies on assumptions for the dates of infection of infected individuals; nonetheless, estimates were substantially robust with respect to different imputations of the dates of infection (i.e., by using the incubation period estimated for Delta, different distributions of the diagnostic delay for asymptomatic individuals, and allowing the possibility of false negative test results, see Figure 2-3 and Appendix A). The same intrinsic limitation of the unobservability of infection times is shared by all transmission chain reconstruction models, but there are now several examples where these models have been proven to correctly identify the transmission dynamics of infectious outbreaks [25,86]. It is also important to stress that for traced contacts, the detection of symptoms was done at the time of diagnosis; as such, if symptoms appeared in the days following the positive swab, the infected individual was recorded as asymptomatic. Another specific limitation is that compliance to quarantine protocols is unknown; we assumed 100%, i.e., that household members quarantined after diagnosis of another member could only be infected within the household. If compliance was imperfect in the considered population, infected household members may have contracted the infection from the general community, especially considering the very high incidence observed in January 2022 in Italy. A sensitivity analysis where quarantines of household members are not considered (i.e., 0% compliance) yielded a significantly shorter mean intrinsic generation time (5.1 vs 6.8 days), because in this case longer generation times that were attributed to potential household infectors in the baseline analysis are preferentially attributed to an importation from the general community. As a result, the mean estimate of the intrinsic generation time may be shorter than the baseline if compliance to quarantine decreased during the period when Omicron was dominant. Lower compliance to quarantine is possible because the lower severity of the Omicron wave in Italy and a general relaxation of control measures induced a lower perception of risk. Considering the ability of Omicron to escape the immune response

from vaccination, we assumed no reduction in transmissibility for vaccinated individuals. However, relaxing such assumption by halving the transmissibility for vaccinated individuals yielded shorter yet comparable results in the mean estimate of the intrinsic generation time (6.1 days). Additional sensitivity analysis investigating uncertainties on the transmissibility of asymptomatic individuals, or on the absence of cross-protection from previous infections with other lineages did not affect the main results significantly (Appendix B).

In conclusion, we produced robust estimates of the length of the intrinsic generation time for Omicron in Emilia Romagna, Italy, suggesting limited variations with respect to ancestral lineages or variants Alpha and Delta obtained in the same country, and providing useful insights for further characterizing the transmission patterns of the SARS-CoV-2 Omicron variant and for policy evaluation.

Appendix B

Estimation of the incubation period

We fitted a gamma distribution to the empirical distribution of incubation periods observed during a superspreading event (a company dinner) occurred on November 26, 2021, in Norway [98]. A nonparametric bootstrap resampling was performed to assess uncertainty in the parameters. In the epidemiological study, 81 of the tracked participants to the dinner developed infection with SARS-CoV-2, and 80 of these were symptomatic. We excluded 6 cases who had symptom onset before the event and considered the empirical distribution of the remaining 74 cases for the baseline estimate. As a sensitivity analysis, we excluded 14 further cases whose samples had not been sequenced (as they could have been infected with a different lineage) and 5 who had come back from travels abroad (of which one from South Africa) within one week before the dinner (as they could have contracted the variant abroad).

The resulting fits for the two analyses are substantially overlapping (Figure 2-4). We obtained a mean incubation period for Omicron of 3.49 days (standard deviation: 1.20, 95% bootstrap confidence interval, bCI: 3.19-3.77 days) in the baseline analysis (blue line) and of 3.45 days (standard deviation: 1.20 days, 95% bCI: 3.13-3.80) in the sensitivity analysis. The algorithm for the fit was implemented in python (version 3.9.7) using package scipy (version 1.7.1).

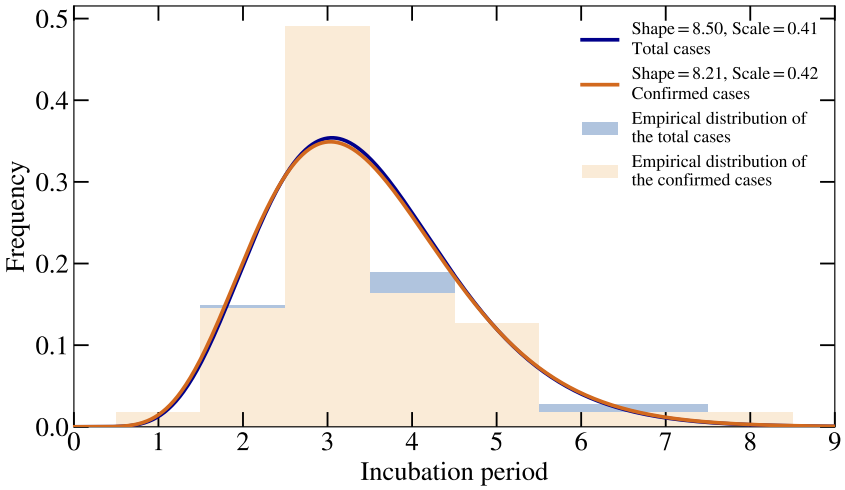

Figure 2-4. Empirical and fitted distribution of the incubation period for variant Omicron, using data from [98].

Imputation of infection dates

The task of reconstructing transmission chains must overcome the intrinsic limitation of the unobservability of transmission chains. We use available evidence to probabilistically impute plausible infection dates for all SARS-CoV-2 cases in our dataset. We combine observed dates of symptom onset, diagnosis, and negative test results with available knowledge on incubation periods and the probability of testing positive over time for infected individuals.

First, we impute the dates of infection for all symptomatic cases. Let T_D be the date of diagnosis (when the individual tested positive), T_S the date of symptom onset, T_N the date of the last negative test before symptom onset; we define the following probability P_I of being infected on day T_I:

$$P_I(T_I) = P_S(T_s - T_I) \cdot H(T_I - T_N) \tag{2.1}$$

where $P_S(t)$ is the probability density function of the incubation period and $H(x)$ is the Heaveside step function (i.e., $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x \geq 0$). For $P_S(t)$ we use the estimate above as a baseline, and an alternative estimate on the delta lineage [105] as sensitivity analysis (see below). For each symptomatic case, a time of infection is sampled from $P_I(t)$ and the date of infection T_I is obtained by rounding to the closest integer. The sample is repeated K = 100 times.

For asymptomatic cases, we cannot use the information on the incubation period given that no date of symptom onset is defined. Therefore, we use the imputed dates of infection for symptomatic cases to define a distribution of diagnostic delays $P_D(x)$, defining the probability of being diagnosed after x days from infection. An empirical approximation of $P_D(x)$ will be given, for any x, by the fraction of all instances across the K stochastic samples for which the diagnostic delay $T_R = T_D - T_I$ is equal to x. A gamma function is then fitted to the empirical distribution using a maximum likelihood approach to obtain $P_D(x)$. The infection date of asymptomatic cases can then be sampled from the following probability:

$$P_I(T_I) = P_D(T_D - T_I) \cdot H(T_I - T_N) \tag{2.2}$$

assuming that the distribution of diagnostic delays for asymptomatic cases is the same as for symptomatic cases. Figure 2-5 reports the estimated empirical and fitted distributions of diagnostic delays for variant Omicron.

The algorithm for the imputation of dates of infection was implemented in python (version 3.9.7) using packages *numpy* (version 1.20.3) and *scipy* (version 1.7.1).

Inference of transmission links

The model adopted in this work extends the approach previously proposed in [69,105]. We assumed that, at any time t, a susceptible individual j within a household is exposed to a force of infection composed of two components:

$$\lambda_j(t) = \lambda_j^o(t) + \lambda_j^h(t) \tag{2.3}$$

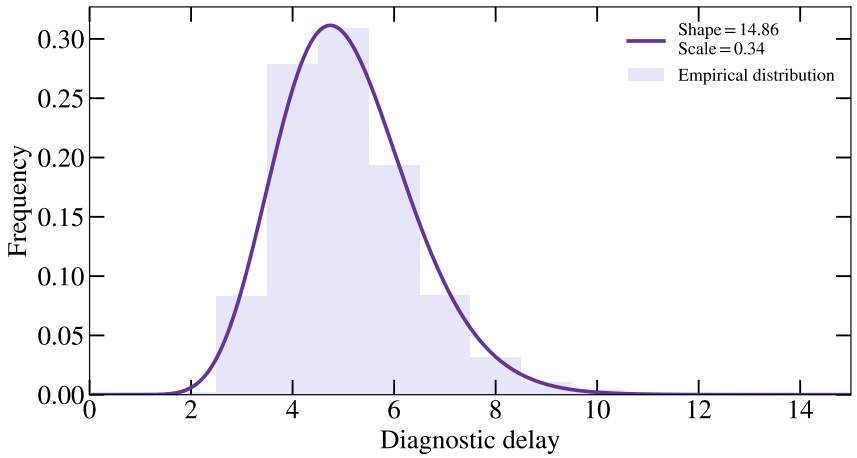

Figure 2-5. Empirical and fitted distribution of the diagnostic delay P_D for variant Omicron, estimated from symptomatic cases. The histogram represents the empirical distribution given the imputed infection times for symptomatic individuals. The curve represents the fitted gamma function.

where $\lambda_j^o(t)$ represents the force of infection from the general community outside the household, and $\lambda_j^h(t)$ represents the one from infected members inside the household.

We define:

$$\lambda_j^o(t) = \sum_{z \in 0..t} \alpha I(z) \chi_j(t) \Gamma(t - z; a, b)\, q_j(t) \tag{2.4}$$

where:

- α is a free parameter scaling the transmissibility in the general community;
- $I(z)$ is proportional to the number of newly infected cases at time z outside the household of j, obtained from epidemic curves by date of symptom onset for the province of Reggio Emilia in the Italian integrated surveillance system [63,113];
- $\chi_j(t)$ represents the relative susceptibility of individual j and changes over time t depending on the dates of vaccination of j and on cross-protection given by previous infection with other SARS-CoV-2 lineages;

- $\Gamma(t; a, b)$ represents the distribution of the intrinsic generation time at day t after infection, for which we assumed a discretized Gamma distribution with scale a and shape b; in particular, given $g(t; a, b)$ the continuous Gamma probability distribution, $\Gamma(t; a, b) = \int_t^{t+1} g(\tau; a, b) d\tau$;
- $q_j(t)$ is an on/off function that is 0 when the household of j is in quarantine and 1 otherwise. For each household, a quarantine of 14 days is started after the first diagnosis and reinstated for a further 14 days every time there is a new diagnosis after the previous quarantine has ended.

In this analysis (and similarly in the previous chapter), we assumed that each household member has an independent probability of acquiring infection from outside the household. While this simplification was necessary to ensure tractable and interpretable estimates, it may not account for correlated exposures [114], such as multiple household members being exposed during shared external activities.

In addition, we define $\lambda_j^h(t)$ as:

$$\lambda_j^h(t) = \sum_{i \in H_j} \lambda_{j,i}^h(t) = \sum_{i \in H_j} \beta \chi_j(t) \Gamma\left(t - T_{I,i}; a, b\right) \tag{2.5}$$

where:
- i is an index running over the set H_j of infected household members of individual j;
- β is a free parameter scaling the transmissibility inside households.

For the relative susceptibility, we assumed that a vaccine dose starts to be protective 14 days after inoculation:

$$\chi_j(t) = \begin{cases} 1 & if\ t < t_{v,1} + 14 \\ (1 - \chi^{(1)}\, e^{-w_1(t - t_{v,1} - 14)}) & if\ t_{v,1} + 14 \le t < t_{v,2} + 14 \\ (1 - \chi^{(2)} e^{-w_2(t - t_{v,2} - 14)}) & if\ t_{v,2} + 14 \le t < t_{v,3} + 14 \\ (1 - \chi^{(3)} e^{-w_3(t - t_{v,3} - 14)}) & if\ t \ge t_{v,3} + 14 \end{cases} \tag{2.6}$$

Where $t_{v,d}$ is the date of vaccination dose d, $\chi^{(d)}$ are the initial effectiveness of dose d (i.e., 14 days after vaccination) against the considered variant, and w_d is the waning rate of vaccine protection for dose d. Estimates of vaccine effectiveness and waning rate were obtained from a large-scale test-negative case–control study [115] and reported in Table 2-2. To reproduce at best the vaccines administered in Italy, we considered the effectiveness estimated for a COMIRNATY vaccine for the main schedule and a heterologous booster with Spikevax vaccine. The waning rate assumed in this chapter is based on a preliminary, unpublished version of [90]. The value of w_d here is higher since Omicron had a higher ability to escape vaccine protection, as marked by reported increase in breakthrough infections. The authors of [90] later refined their results by incorporating additional data into their meta-analysis, published after this paper.

Table 2-2. Parameters for vaccine effectiveness and waning.

Parameter	Unit	Dose	Value
Initial effectiveness (14 days after dose)	%	1	0
		2	76.3
		3	76.5
Waning rate	$days^{-1}$	1	0
		2	1/74.5
		3	1/195.3

For individuals who had a previous infection with a different lineage, we assumed a cross-protection η=56% [106] before he receives a complete primary vaccination cycle (first two doses), and a non-waning protection equal to the one conferred by the booster dose after completion of the primary cycle [106]:

$$\chi_j(t) = \begin{cases} 1-\eta & if\ t < t_{v,2} + 14 \\ 1-\chi^{(3)} & if\ t \geq t_{v,2} + 14 \end{cases} \quad (2.7)$$

The model assigns a source of infection k_j for all cases by choosing from either a generic source outside the household or from an infectious household member in H_j, with probability proportional to the contribution of each source to the total force of infection $\lambda_j(T_{I,j})$ at the time $T_{I,j}$ at which j was infected. The overall likelihood of the observations given parameter set $\theta = (\alpha, \beta, a, b)$ and the assigned sources of infection k_j is given by:

$$L(\theta, k_j) = \prod_j P_j Q_j \quad (2.8)$$

where

$$P_j = \begin{cases} \lambda_j^o(\mathrm{T_{I,j}}) & \text{if } k_j \text{ is outside the household} \\ \lambda_{j,i}^h(\mathrm{T_{I,j}}) & \text{if } k_j \text{ is household member } i \\ 1 & \text{if } j \text{ is uninfected} \end{cases} \quad (2.9)$$

For infected individuals, Q_j is the probability that j has not been infected until $T_{I,j}$, namely $Q_j = e^{-\int_0^{T_{I,j}} \lambda_j(t)dt}$. For uninfected individuals, it is the probability that j has never been infected, $Q_j = e^{-\int_0^{\infty} \lambda_j(t)dt}$.

We estimated the unknown parameters θ and the source of infection k_j for all cases using a Monte Carlo Markov Chain (MCMC) procedure. At each step, all parameters in θ are updated using reversible normal jumps. Z=500 samples from the posterior distributions obtained by the MCMC for each of the K=100 samples were pooled together to obtain the final parameter distribution and the distribution of the sources of infection for each case.

The model for the inference of transmission links was implemented in C using GSL libraries (version 2.6) and compiled with GCC (version 4.2.1).

The code used to run the model is available at: figshare.com/IGT_omicron .

Table 2-3. Statistics for the model-based reconstruction of transmission links in households by number of SARS-CoV-2 cases.

	Number	***%***
Households with 2 SARS-CoV-2 cases	**5,579**	**100**
- Both infected in the general community	642 (540-770)	11.5 (9.7-13.8)
- One infected the other	4,937 (4,809-5,039)	88.5 (86.2-90.3)
Households with 3 SARS-Cov-2 cases	**2,080**	**100**
- All infected in the general community	18 (9-29)	0.8 (0.4-1.4)
- One transmission, 2 infected in the general community	349 (300-410)	16.8 (14.4-19.7)
- Two transmissions, same infector (1 generation)	838 (792-884)	40.3 (38.1-42.5)
- Two transmissions, different infectors (2 generations)	875 (795-944)	42.0 (38.2-45.4)
Households with 4 or more SARS-Cov-2 cases	**1,244**	**100**
- All infected in the general community	0 (0-2)	0 (0-0.2)
- One transmission	174 (144-208)	14 (11.6-16.7)
- Two transmissions	637 (604-670)	51.2 (48.6-53.9)
- Three or more transmissions	432 (372-494)	34.8 (29.9-39.7)

Statistics on reconstructed transmission links

The mean per-household number of infections contracted from the general community was 1.15 (95%CrI 1.13 – 1.18). The mean number of secondary infections generated by a positive case was 0.56 (95%CrI 0.55 – 0.57). Table 2-3 shows how the model reconstructed transmission links within households with different numbers of cases.

Stability of the attributed source of infection

For each case, we considered the distribution of the sources of infection attributed by the model and evaluated its stability. We categorized cases according to whether its source of infection was consistently (i.e., more than 75% of the times over the Z sampling of infector and K sampling of infectious dates) attributed to:

- The same household member;
- transmission within household but from different potential infectors;
- transmission in the general community.

The setting of transmission was uncertain (less than 75% consistency in attribution) in about 40% of cases (Figure 2-6). This generally happened when two or more cases in a household had close diagnosis dates, so that either could have been infected in the general community and then transmitted to the other, or both could have been infected in the general community, depending on the assigned dates of infection.

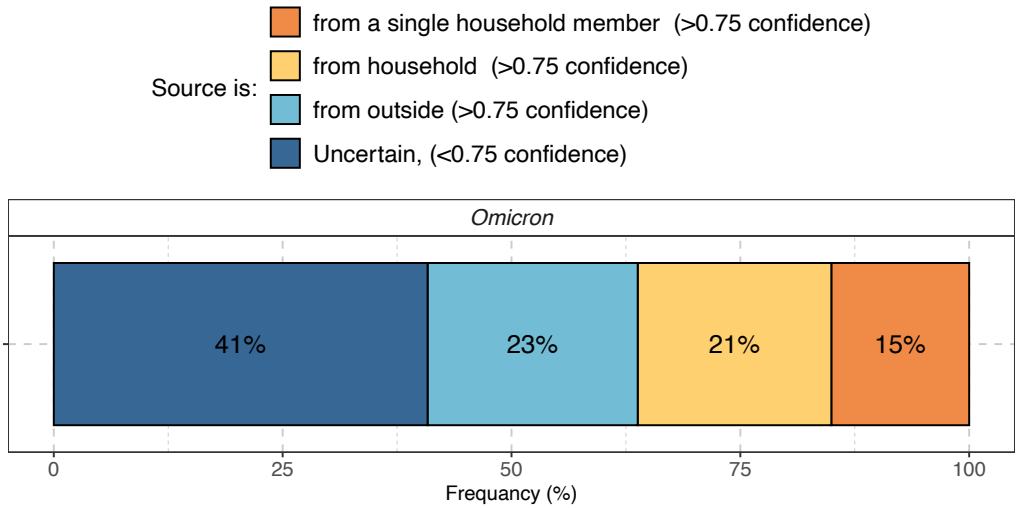

Figure 2-6. Consistency in the attribution of the infector or the infector setting. The stacked barchart represents the proportion of individuals that were consistently (more than 75% of the times across Z sampling of sources and K sampling of infectious dates) or inconsistently attributed to either category.

Sensitivity Analyses

We performed ten sensitivity analyses (SA) to test the robustness of model results against different model assumptions. The first SA (a) applies the baseline model to a subset of 380 households (1,127 cases in total) for which a case was genotyped as Omicron; the second (b) applies the baseline model to a subset of 1,148 households (2,770 cases in total) for which all individuals were unvaccinated; the third to sixth (c, d, e, f) impact on the main unknown of the data, i.e., the imputation of infection times. In particular, the third SA (c) assumes an incubation period equal to one previously estimated for Delta [105] (mean: 4.5 days and standard deviation 2.1 days) to reassign the imputed infectious dates (baseline: Omicron variant with mean 3.5 days and standard deviation 1.2 days); the fourth (d) considers a distribution of the diagnostic delay for asymptomatic cases that is 50% longer than that for symptomatic cases, implemented by increasing the shape of the gamma distribution by 50% (mean: 7.58 days; standard deviation: 1.61 days); the fifth (e) is similar to (d), but implemented by changing the scale of the gamma distribution by 50% (mean: 7.58 days; standard deviation: 1.97 days); Figure 2-7 shows a comparison of the diagnostic delay distributions used in SA (d) and (e). The sixth SA (f) allows false negatives test results when imputing infection dates; the seventh to tenth SA (g, h, i, j) consider slightly different model assumptions. In particular, the seventh (g) considers a reduced transmissibility for asymptomatic individuals; the eighth (h) considers a reduced transmissibility for vaccinated individuals, the ninth (i) assumes that any effort to quarantine positive cases would not impact the force of infection from outside the household (i.e., $q(t) = 1$ for any value of t in Equation (2.4)) which corresponds to a 0% compliance to the policy; the tenth (j) evaluates the possibility that previous infection from other variants provides no residual natural immunity.

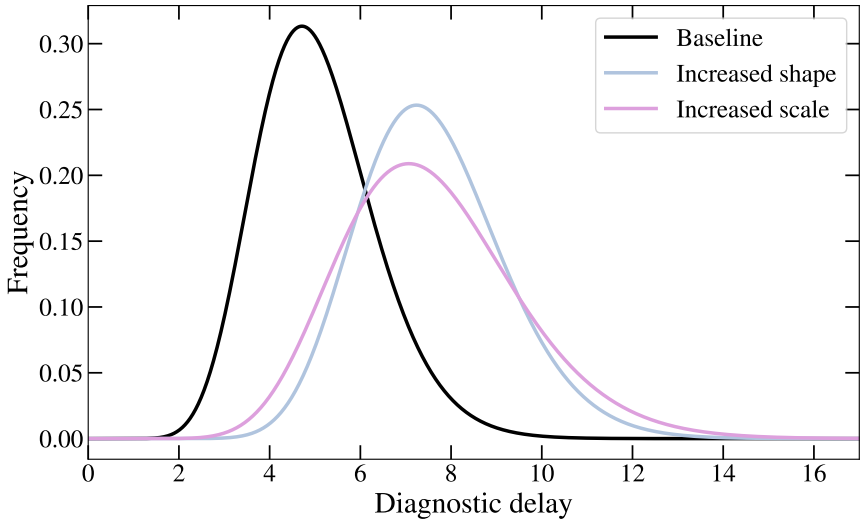

Figure 2-7. Diagnostic delays distributions for asymptomatic individuals. "Baseline" represents the distribution used in the main analysis, equal to the one estimated for asymptomatic cases; "Increased shape" is the distribution used in SA (d); "Increased scale" is the one used in SA (e).

Only households with confirmed omicron genotype (a)

In this sensitivity analysis we considered only the subset of households for which one case has been classified as Omicron after genotyping (380 households, 1,127 cases in total). Results are reported in Table 2-4.

Table 2-4. Estimates for the intrinsic and realized household generation time and household serial intervals when considering only households with a confirmed Omicron genotype.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.20 (5.14-7.99)
	shape mean (95%CrI)	2.51 (2.1-3.58)
	scale mean (95%CrI)	2.53 (1.6-3.64)
	Standard deviation of the mean distribution [days]	3.95
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.39 (3.36-3.4)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.27 (2.1-2.44)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	55.6 (51.8-59.2)

Only households including unvaccinated individuals (b)

In this sensitivity analysis we considered only the subset of households where all individuals were unvaccinated (1,148 households; 2,770 cases in total). Results are reported in Table 2-5.

Table 2-5. Estimates for the intrinsic and realized household generation time and household serial intervals when considering only households where all individuals were unvaccinated.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	7.65 (6.73-8.36)
	shape mean (95%CrI)	2.08 (1.89-2.35)
	scale mean (95%CrI)	3.69 (3.23-4.01)
	Standard deviation of the mean distribution [days]	5.31

REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.24 (3.23-3.26)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	1.89 (1.77-2.01)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	57.4 (54.8-60.2)

Only households including unvaccinated individuals (c)

In this sensitivity analysis we considered an alternative distribution of the incubation period, previously estimated for the Delta variant, with mean 4.5 days (shape: 4.43; scale: 1.01) [105]. Results are reported in Table 2-6.

Table 2-6. Estimates for the intrinsic and realized household generation time and household serial intervals using an alternative distribution of the incubation period.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.87 (6.15-8.12)
	shape mean (95%CrI)	2.81 (2.09-3.49)
	scale mean (95%CrI)	2.53 (1.80-3.59)
	Standard deviation of the mean distribution [days]	4.15
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.96 (3.96-3.97)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.02 (1.96-2.08)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	58.8 (53.8-63.6)

Longer diagnostic delay for asymptomatic cases (increased shape of gamma distribution) (d)

In this sensitivity analysis we increased by 50% the diagnostic delay of asymptomatic cases. For asymptomatic cases, we considered a gamma function that has the same scale parameter as that estimated for symptomatic cases, and a shape parameter that is 1.5 times that estimated for symptomatic cases. Results are reported in Table 2-7.

Table 2-7. Estimates for the intrinsic and realized household generation time and household serial intervals using a longer diagnostic delay (increased shape) for asymptomatic cases.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.87 (5.94-8.07)
	shape mean (95%CrI)	2.51 (2.01-3.27)
	scale mean (95%CrI)	2.83 (1.86-3.76)
	Standard deviation of the mean distribution [days]	4.39
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.68 (3.67-3.68)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.28 (2.22-2.33)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	55.3 (50.4-59.9)

Longer diagnostic delay for asymptomatic cases (increased scale of gamma distribution) (e)

In this sensitivity analysis we increased by 50% the diagnostic delay of asymptomatic cases. For asymptomatic cases, we considered a gamma function that has the same shape parameter as that estimated for symptomatic cases, and a scale parameter that is 1.5 times that estimated for symptomatic cases. Results are reported in Table 2-8.

Table 2-8. Estimates for the intrinsic and realized household generation time and household serial intervals using a longer diagnostic delay (increased scale) for asymptomatic cases.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.92 (6.01-8.07)
	shape mean (95%CrI)	2.58 (2.02-3.3)
	scale mean (95%CrI)	2.77 (1.88-3.76)
	Standard deviation of the mean distribution [days]	4.36
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.79 (3.79-3.8)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.29 (2.23-2.34)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	55 (49.9-59.6)

False negative test allowed when imputing infection dates (f)

In this sensitivity analysis we allowed for false negative test results when imputing infection dates, i.e. we removed the factor $H(T_I - T_N)$ in Equations 1 and 2. Results are reported in Table 2-9.

Table 2-9. Estimates for the intrinsic and realized household generation time and household serial intervals when allowing for false negative test when imputing infection dates.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.68 (5.58-8.08)
	shape mean (95%CrI)	2.42 (2.04-3.3)
	scale mean (95%CrI)	2.83 (1.8-3.75)
	Standard deviation of the mean distribution [days]	4.34
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.4 (3.39-3.4)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.23 (2.18-2.29)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	54 (49.4-58.1)

Reduced transmissibility for asymptomatic individuals (g)

In this sensitivity analysis, we consider a halved transmissibility for asymptomatic individuals [116] by multiplying Equation (2.5) by an individual transmissibility φ_i, where φ_i is 1 if *i* is symptomatic and 0.5 if asymptomatic. Results are reported in Table 2-10.

Table 2-10. Estimates for the intrinsic and realized household generation time and household serial intervals using a halved transmissibility for asymptomatic individuals.

INTRINSIC GENERATION TIME	**MEAN (95%CRI) [DAYS]**	**6.66 (5.78-7.99)**
	shape mean (95%CrI)	2.55 (2.05-3.22)
	scale mean (95%CrI)	2.65 (2.06-3.42)
	Standard deviation of the mean distribution [days]	4.19
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.58 (3.57-3.58)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.36 (2.29-2.42)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	51.3 (46.4-55.6)

Reduced transmissibility for vaccinated individuals (h)

In this sensitivity analysis, we consider a halved transmissibility for vaccinated individuals [116] by considering φ_i = 0.5 for vaccinated and φ_i = 1 for unvaccinated individuals. Results are reported in Table 2-11.

Table 2-11. Estimates for the intrinsic and realized household generation time and household serial intervals using a halved transmissibility for vaccinated individuals.

INTRINSIC GENERATION TIME	mean (95%CrI) [days]	6.06 (5.5-6.62)
	shape mean (95%CrI)	2.09 (1.9-2.3)
	scale mean (95%CrI)	2.91 (2.45-3.39)
	Standard deviation of the mean distribution [days]	4.20
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.46 (3.45-3.47)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.26 (2.22-2.32)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	53.0 (48.6-56.8)

No protection from infection outside the household during quarantine (i)

In this sensitivity analysis, we assume that the imposed quarantine period after the first positive diagnosis would not impact the force of infection from outside the household (i.e., $q(t) = 1$ for any value of t in Equation (2.4)). Results are reported in Table 2-12.

Table 2-12. Estimates for the intrinsic and realized household generation time and household serial intervals when assuming no protection from outside infection during the quarantine period.

INTRINSIC GENERATION TIME	**MEAN (95%CRI) [DAYS]**	**5.09 (4.37-5.78)**
	shape mean (95%CrI)	2.44 (2.09-4.21)
	scale mean (95%CrI)	2.14 (1.06-2.7)
	Standard deviation of the mean distribution [days]	3.30
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.39 (3.38-3.4)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.22 (2.16-2.28)
PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	53.6 (48.7-57.5)

No protection from previous infection (j)

In this sensitivity analysis, we assume that the previous infection would not impact the susceptibility to Omicron infection of an individual. Results are reported in Table 2-13.

Table 2-13. Estimates for the intrinsic and realized household generation time and household serial intervals when assuming no protection from previous infection.

INTRINSIC GENERATION TIME	**MEAN (95%CRI) [DAYS]**	**6.88 (5.72-8.75)**
	shape mean (95%CrI)	2.37 (2-3.29)
	scale mean (95%CrI)	2.99 (1.83-4.35)
	Standard deviation of the mean distribution [days]	4.52
REALIZED HOUSEHOLD GENERATION TIME	mean (95%CrI) [days]	3.59 (3.55-3.59)
HOUSEHOLD SERIAL INTERVAL	mean (95%CrI) [days]	2.38 (2.3-2.46)

PRE-SYMPTOMATIC TRANSMISSION	mean (95%CrI) [%]	51.1 (45.3-55.7)

Epidemiological context

In the Emilia Romagna region, where our study area is located, Omicron became dominant in the second half of December 2021 (Table 2-14), causing a large upsurge of cases that subsided by the end of January (Figure 2-8). A second smaller wave occurred since early March, likely due to the expansion of sublineage BA.2 (Table 2-14 and Figure 2-8). The selected study period allows to include a large majority of cases from the first Omicron wave while minimizing the risk of including multiple sublineages with different transmissibility in the data, which would require a significant complication of the adopted model.

Table 2-14. Prevalence of the Omicron variant in genomic surveillance surveys conducted within the Emilia Romagna region, December 2021-April 2022. Data from [65].

Relative prevalence	**2021**		**2022**				
	Dec 6	Dec 20	Jan 3	Jan 17	Jan 31	Mar 7	Apr 4
Omicron	0.6%	16.5%	79.4%	99%	98.7%	100%	100%
of which BA.1	NA*	NA*	NA*	97%**	97%**	41.7%	6.4%
BA.2	NA*	NA*	NA*	3%**	3%**	58.3%	94.6%

* Sublineages were not genotyped on December 6, December 20 and January 3.
** Sublineages were genotyped at the national level on January 17 and 31.

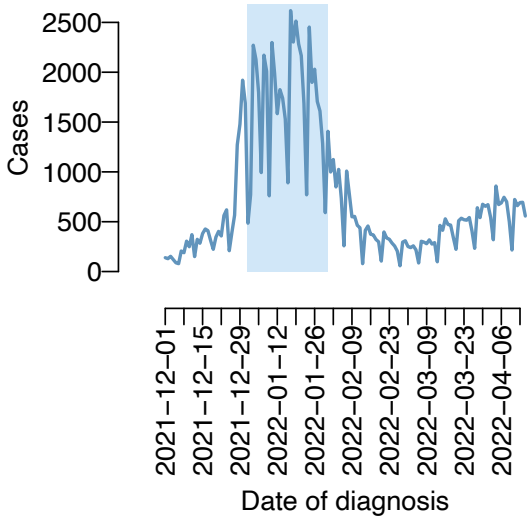

Figure 2-8. Daily number of diagnosed cases in the province of Reggio Emilia between December 1, 2021 and April 15, 2022. The study period of this work (January 1-31, 2022) is highlighted in light blue. Data from the Integrated Surveillance System of the Istituto Superiore di Sanità [63,113].

Simulation of within-household transmission

To illustrate the difference between the intrinsic and realized generation time, we built a simple computational model capable of generating infections within households while accounting for the possibility of acquiring infections from the community. First, we built a population of individuals distributed in households according to Italian statistics on household size [117]. Then, we consider each individual to be subject to the force of infection defined in Equations 3-5, where I(z) is the number of symptomatic cases in the province of Reggio Emilia and parameters have been set to arbitrary values. In particular, we imposed $a = 2.3$ and $b = 3$ for the gamma-distribution of the generation time,

corresponding to a mean intrinsic generation time of 6.9 days (standard deviation: 4.6 days). Every time an infection within the household occurs, we record the date of infection and the infector; for infections imported from the community we only recorded the date of infection. In this way, we can reconstruct the distribution of the realized generation times in household.

The model is implemented stochastically with a time step of 1 hour and applied to a population of 2,000 households for a duration of 60 days, where day 1 represented December 16, 2021, and day 60 represented February 14, 2022. After running the model, we filtered households with at least two cases, with less than 14 days elapsing between two consecutive infections, and with all diagnoses comprised in the period January 1-31, 2022. We obtained a mean realized generation time in households of 4.72 days (standard deviation: 0.09 days); 95%CrI of the mean: 4.54-4.90 days), i.e., significantly shorter than the intrinsic one, under parameters for transmissibility ($\alpha = 5 \cdot 10^{-6}$ and $\beta = 0.5$) that resulted in an attack rate of 70% and in a rate of 1.13 imported cases per household.

CHAPTER 3

Transmission dynamics of Norovirus on cruise ships

Introduction

Norovirus is a main cause of outbreaks of gastroenteric infections (GI) on cruise ships [118]. Norovirus can spread explosively on-board, either through an environmental reservoir (contaminated food, water or surfaces), or through person-to-person transmission. In turn, person-to-person transmission can happen directly during close social contacts, or via the ingestion of virus particles aerosolized through vomitus or feces [119].

In addition to direct consequences on travelers' health, outbreaks of norovirus may cause the disruption of holidays for affected passengers and co-travelers, temporary shortages of crew members for routine cruise operations, and high costs related to the implementation of infection control measures for cruise companies, with a remarkable economic burden [120]. If we exclude the temporary effect of the COVID-19 pandemic on the cruise passenger market [121], the number of passengers worldwide has seen a steady and rapid expansion since 2004 [122], suggesting a potentially higher future impact of norovirus outbreaks on the tourism industry in the near future.

Standards for hygiene and plans for prevention and management measures (PMM) have been defined by multiple international institutions [123,124,125]. Pre-embarkation screening, syndromic surveillance on-board, isolation of infected individuals, application of environmental decontamination and education of crew and passengers on hand washing and rapid reporting of symptoms have been identified among the main preventative measures [123,124,125]. Outbreak management measures may include active surveillance, enhanced disinfection of public surfaces, discontinuation of self-service restaurants, and social distancing [123,124,125].

A better understanding of the dynamics of norovirus spread on-board is critical to assess the effectiveness of control interventions. In this work, we analyze an individual line-list of GI cases from a large outbreak on a cruise ship [126] to probabilistically reconstruct transmission chains on-board, assess the importance of diagnostic delays on transmission and evaluate the effectiveness of alternative case isolation scenarios.

Methods

We considered data from an outbreak of 121 GI cases, reported over a 7 days voyage in a cruise ship calling Mediterranean ports [126]. Data consisted in a line-list of reported cases with information on case gender, role (passenger or crew member), cabin number, time of symptom onset, date of diagnosis, type of symptoms (vomiting or diarrhea) and number of vomiting and diarrhea episodes until diagnosis. For 65 of 121 cases there was a PCR confirmation of norovirus presence in stools, therefore we assumed that all 121 GI cases were caused by norovirus. The cruise carried 1229 passengers and 487 crew members (9.7% attack rate among passengers, 0.4% among crew, 7.1% overall). Summary information for the analyzed outbreak is reported in Table 3-1, and the number of cases by time of symptom onset (panel A) and by date of diagnosis (panel B) are shown in Figure 3-1.

We applied a Bayesian model to probabilistically reconstruct transmission chains (i.e., who infected whom) [69,70,105]. For each susceptible individual on board, the model defines a force of infection (FOI) exerted over time by infected individuals, considering the possibility of cases infected before first embarkation, of infection among people sharing cabins during night-time, transmission among travelers on board of the ship, and acquisition of infection during visits in ports of calling. The FOI is a function of the generation time, modeled as a Gamma distribution with a mean of 87.6 hours [127], and includes increased transmission rates for cases who reported vomiting (2.12-fold increase), diarrhea (1.39-fold), or both (2.95-fold) [128]. The model explicitly represents the isolation protocol observed on board, by assuming that diagnosed individuals are immediately confined in their cabins for 72 hours. Isolation results in the impossibility to transmit to people other than cabin members. The model assigns likely infectors by selecting which potential index case contributed the most to an individual's FOI at their time of infection.

Table 3-1. Summary information for the considered outbreak data.

Total number of cases		**121**
Symptoms (%)		
	Only vomit	1 (0.8%)
	Only diarrhea	25 (20.7%)
	Both	95 (78.5%)
	None	0 (0%)
Gender (%)		
	Male	54 (44.6%)
	Female	67 (55.4%)
Role (%)		
	Passengers	119 (98.3%)
	Crew members	2 (1.7%)

The model was calibrated using a Markov Chain Monte Carlo procedure with Metropolis-Hastings sampling algorithm; free model parameters were the transmission rates for each route of infection, the average prevalence of norovirus in port communities, and the unknown infection times of each case. Since the data reported the date but not the time

of diagnosis, model calibration was repeated Z=50 times after imputing for all reported cases their time of diagnosis (in hours). The time was sampled uniformly over the date of diagnosis, excluding night hours (between midnight and 8am) and times preceding the time of symptom onset if this occurred on the same date as diagnosis. M=2500 parameter values and reconstructed transmission chains were sampled from the posterior distributions of the Z calibration procedures and pooled together to obtain the final results. To evaluate results with respect to the potential underreporting of cases, we run sensitivity analyses where we re-calibrated the model assuming levels of underreporting of 23% and 40% [129,130]. Full model specifications are provided in the Appendix C.

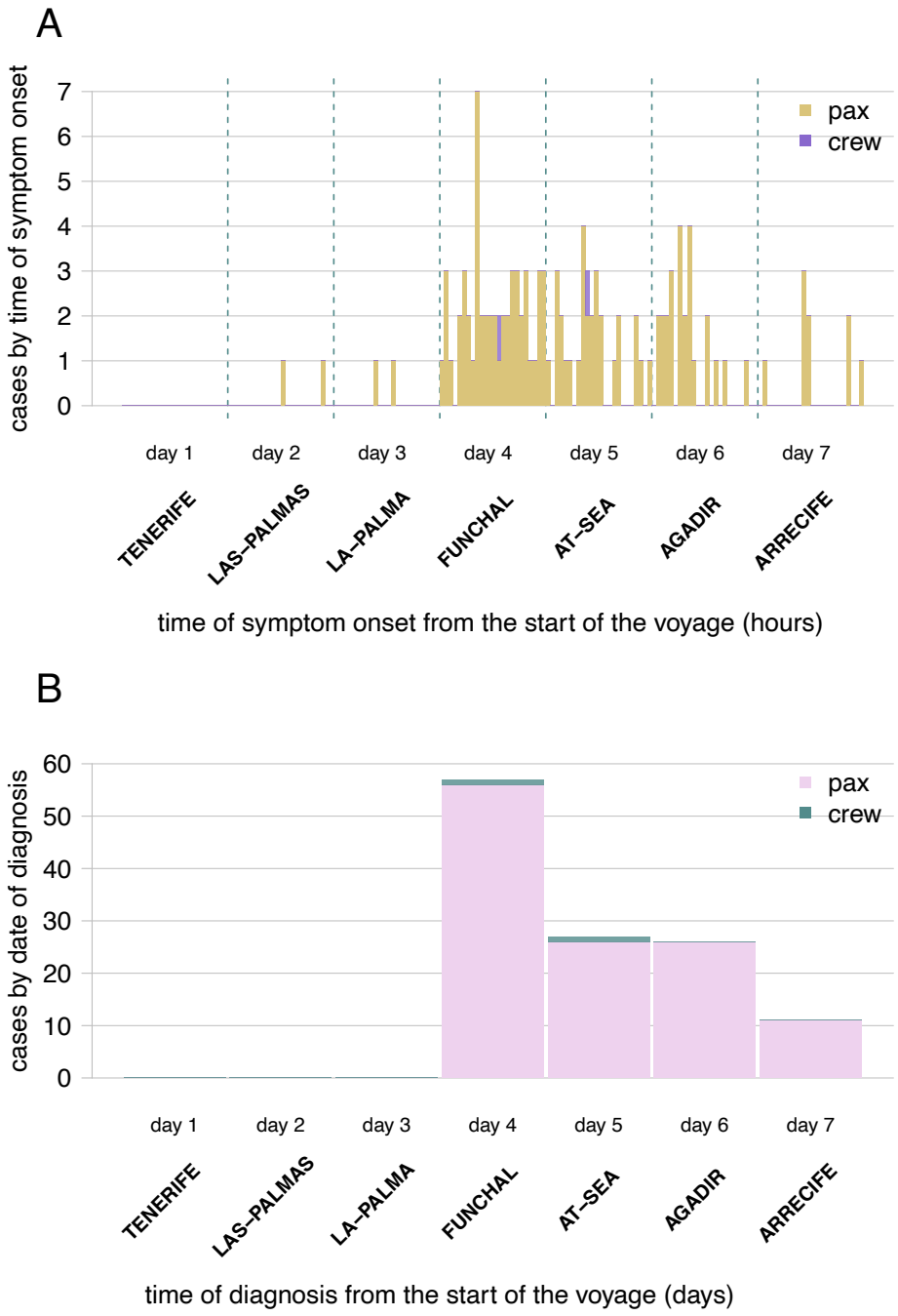

*Figure 3-1. Epidemic curves, disaggregated by passengers (pax) and crew members (crew). **A**. Hourly time series by symptom onset. **B**. Daily time series by date of diagnosis.*

To assess the overall effectiveness of alternative isolation protocols and the impact of diagnostic delays, we additionally developed a branching process model [131], where each infected individual is assumed to generate a theoretical number of cases sampled from a negative binomial distribution with mean proportional to the basic reproduction number R_0 and overdispersion ν. The time of theoretical secondary infections was distributed according to the distribution of the generation time [127]; infections whose sampled time would fall within the isolation period of the infector, or after the end of the cruise, were discarded from the realized infections. The branching process model does not explicitly distinguish cabin transmission or importation of cases during tours.

We assumed that two cases were infected before embarkation (based on results from the reconstructed transmission chains), and their symptom onset time was fixed at those of the first two symptomatic cases in the dataset; their time of infection was established by subtracting an incubation period sampled from a log-normal distribution with mean 32.6 hours [132], until the assigned infection time preceded embarkation. For each case transmitted on-board, the date of symptom onset was assigned by adding to the time of infection an incubation period sampled from the same distribution. The time of diagnosis (corresponding to the time of isolation start) was assigned for all infections by adding a diagnostic delay sampled from an empirical distribution (determined from observed outbreak data in the baseline analysis); infections for which the assigned diagnosis time fell after the end of the cruise were considered as unreported. The time at which isolation ended for each case was assigned by adding a fixed duration of isolation (72 hours in the baseline analysis) to the time of diagnosis.

The observed outbreak may be interpreted as an individual realization of a stochastic process, which the branching process aims to represent. Therefore, we calibrated free model parameters (R_0 and ν) by using a particle filtering approach [133] weighing particles (model trajectories) by the mean squared error (MSE) between the modeled and observed epidemic curves, and exploring parameters by grid search. The best-fitting parameter sets were then used to simulate seven alternative intervention scenarios. In the first four, isolation of diagnosed cases was reduced from 72 hours (baseline scenario, i.e. the strategy actually implemented onboard) to 48 and 24 hours or was not done altogether. In two other scenarios we maintained the 72 hours isolation, but the empirical diagnostic delays after symptoms were either increased or reduced by 50%. Finally, we considered a "perfect" isolation scenario, where for all cases isolation was immediate after the development of first symptoms and lasted until the end of the cruise. Model outputs were the relative reduction in the total number of cases observed in the cruise compared to a scenario with no interventions, and the effective reproduction number, measured as the mean number of secondary cases realized in the cruise by infections occurring within day 2 of voyage. Considering only infections from the early days of voyage allows to reduce the right-censoring effects due to infections that would occur after the end of the cruise. Full details on the branching process model and calibration are reported in Appendix C.

Results

In the model-reconstructed transmission chains for the considered outbreak, 4.5% (95% credible interval, CrI: 1.7%-8.3%) of infections on average occurred before embarkation, 7.3% (95%CrI: 2.5%-10.7%) were transmitted in the cabin and 88.0% (95%CrI: 82.7%-

93.4%) occurred in public spaces of the ship. Acquisition during visits at ports was estimated to be negligible.

Based on the reconstructed transmission chains, the effective reproduction number (i.e., the number of secondary cases caused by a single infectious individual) declined over time for cases infected later, together with their delay between the time of infection and that of diagnosis, hereafter termed "infection diagnostic delay" (Figure 3-2). Cases already carrying the virus before embarkation on the cruise ship caused an average of about 7.6 secondary infections each (95%CrI of 4.4-14.5). Individuals infected during the first 24 hours of cruise went on to cause about 3.0 (95%CrI 1.5-5.2) secondary cases each; the effective reproduction number decreased briskly in the following days, and crossed the epidemic threshold of 1 for infections that occurred at after 24 hours since the start of the cruise.

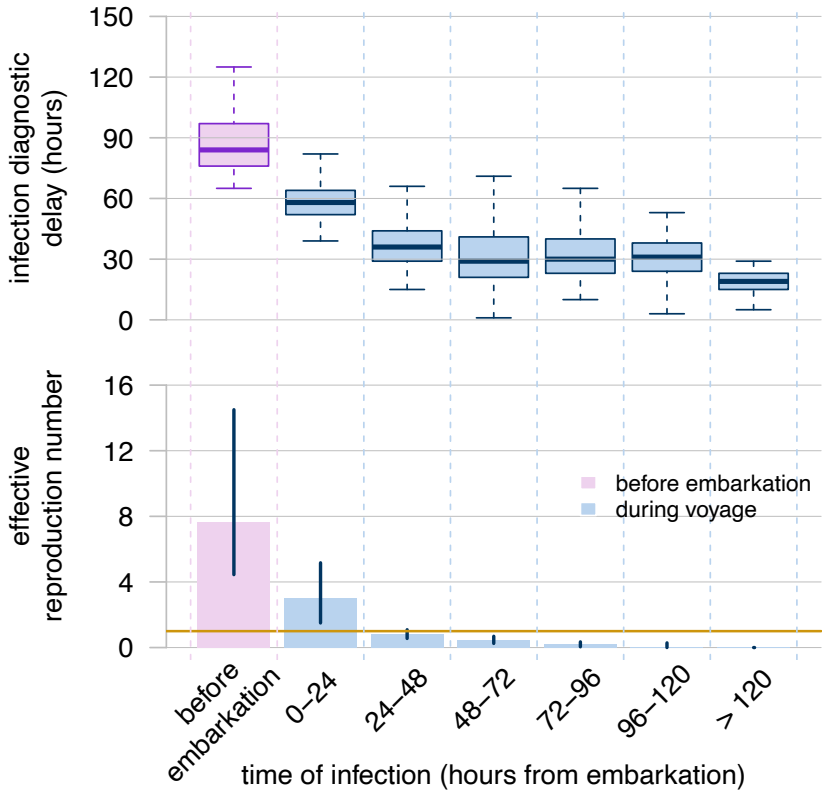

*Figure 3-2. Reproduction number and infection diagnostic delay over the course of the outbreak. Boxplots (**top**) show the distribution of infection diagnostic delays (in hours) for cases infected on each day of the voyage. The bar plot (**bottom**) represents the* estimated effective reproduction number for cases infected on each day of the voyage. The bars indicate the mean estimate, while the error bars indicate the 95%CrI of the mean over all the reconstructed chains.

The model estimates that 60% (95%CrI: 55%-65%) of cases did not transmit the infection further, while 11% of individuals went on to cause 3 or more secondary cases (Figure 3-3A). Fitting negative binomial distributions to the offspring distributions associated to model-reconstructed transmission chains resulted in a mean overdispersion parameter of 0.42 (95%CrI: 0.28-0.62), indicating the existence of superspreading individuals disproportionately contributing to transmission. This result is confirmed in Figure 3-3B, displaying the cumulative proportion of cases caused by infectors ranked by their number of secondary cases; the figure shows that the top 20% of infectors were responsible for 77% (95%CrI: 71%-83%) of all cases, in accordance with the Pareto rule [134], and the top 10% was responsible for 57% (95%CrI: 48%-

65%) of all cases.

Figure 3-4 shows a disaggregation of the model-estimated mean number of secondary cases by presence of vomit among symptoms, and by infection diagnostic delay. The mean number of secondary cases generated by cases who experienced vomiting episodes was estimated by the model at 1.02 (95%CrI: 0.88-1.15), significantly higher than the corresponding value for cases with no vomiting (mean 0.70, 95%CrI: 0.28-1.28; Student's t-test p-value << 10-6) (Figure 3-4A). The estimated mean number of secondary cases disaggregated by infection diagnostic delay increased from 0.39 (95%CrI: 0.27-0.51) for infections diagnosed within 2 days, to 1.89 (95%CrI: 1.13-2.81) for delays between 2 and 3 days, and to 5.13 (95%CrI: 3.14-8.25) for delays larger than 3 days (Figure 3-4B).

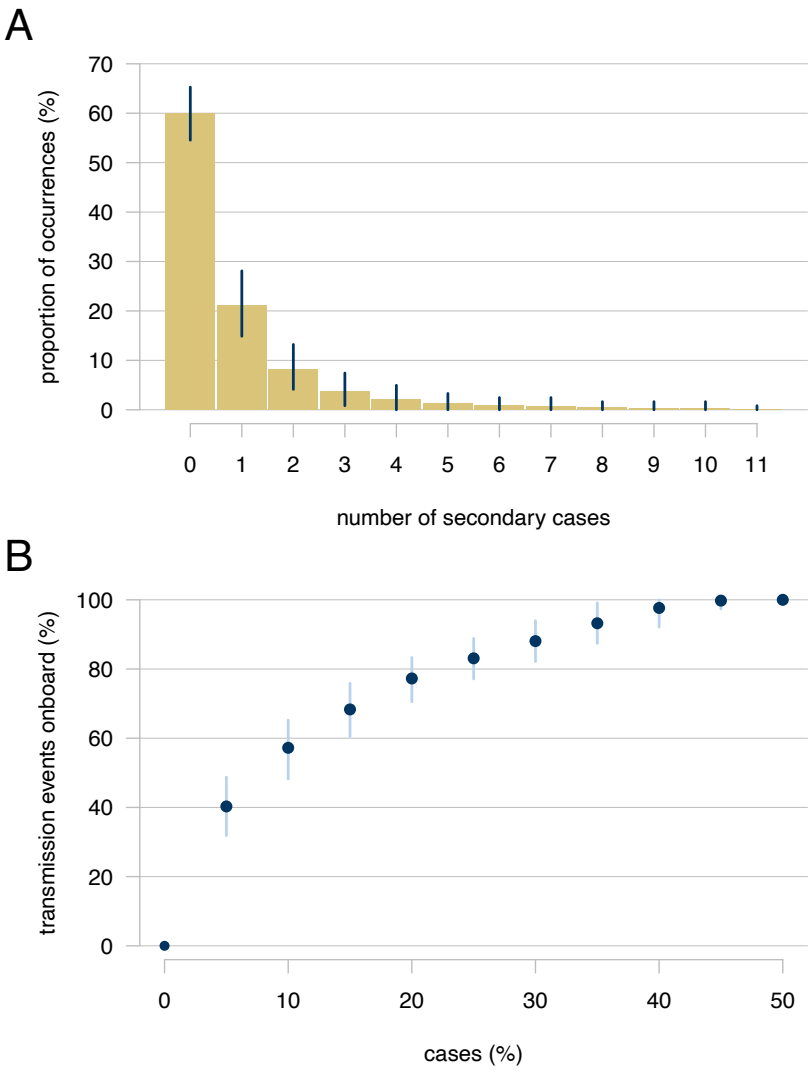

Figure 3-3. Superspreading in the considered outbreak. ***A.*** *Distribution of the number of secondary cases generated by infected individuals. Bars indicate mean values over 125,000 reconstructed transmission chains, while error bars indicate the 95%CrI;* ***B.*** *Cumulative proportion of secondary cases ranked by infectors. The points indicate the average value, while the error bars indicate the 95%CrI over all the reconstructed chains.*

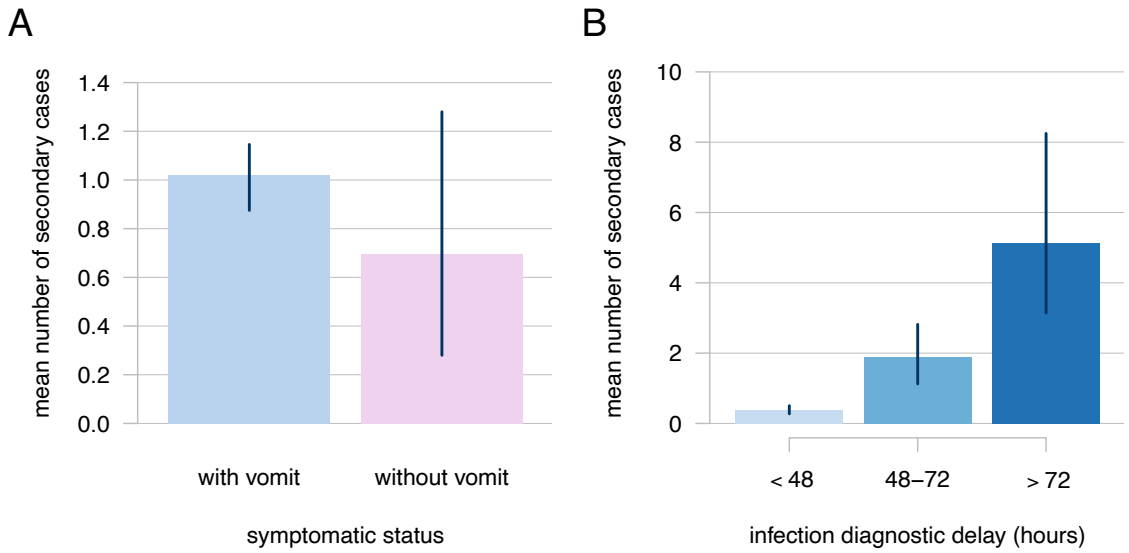

*Figure 3-4. **A** Number of secondary cases generated by individuals with and without vomiting episodes; **B** Number of secondary cases stratified by infection diagnostic delay. Bars indicates mean values; error bars indicate the 95%CrI of the mean over all the reconstructed chains.*

The model estimated a mean number of secondary cases for the top 10% of infectors of 5.64 (95%CrI: 4.83-6.42), 12.5-fold higher than the remainder of cases (0.45; 95%CrI: 0.38-0.54) (Figure 3-5A). The mean diagnostic delay for the top 10% of infectors was 83 hours (95%CrI: 70-96 hours), much larger than the one for the remainder of cases (47 hours, 95%CrI: 44-50 hours) (Figure 3-5B). The top 10% of infectors experienced gastrointestinal symptoms (vomiting or diarrhea) with a halved frequency (mean 4.8, 95%CrI: 3.5-6.4 per day) compared to the remainder of infected individuals (mean 9.8, 95%CrI: 9.1-10.6 per day) (Figure 3-5C).

All findings remained robust when considering a potential underreporting of cases up to 40% (see Appendix C).

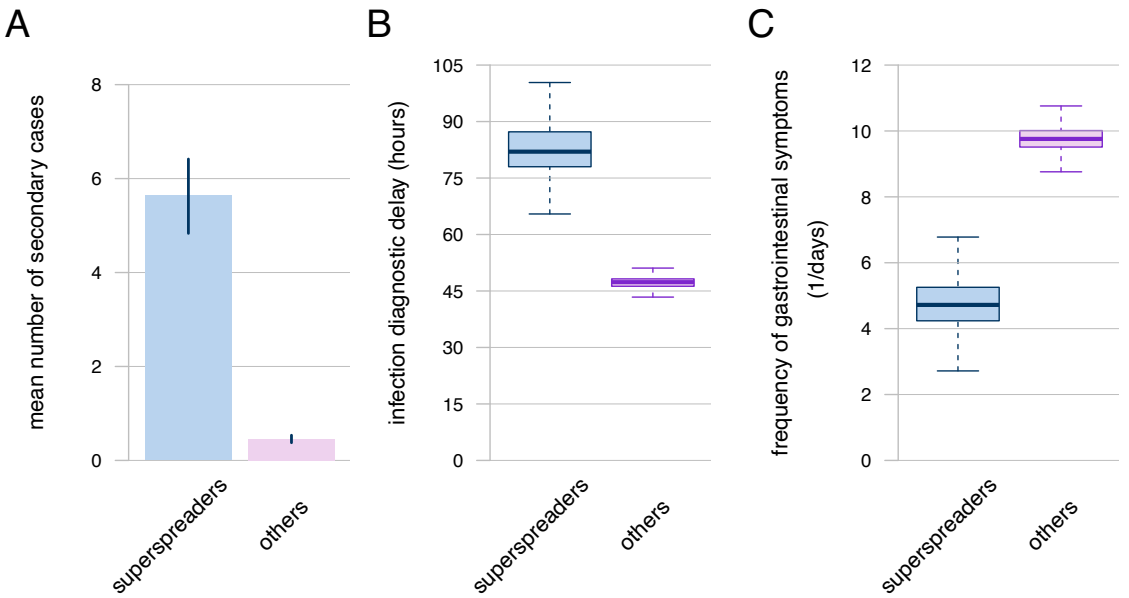

*Figure 3-5. Characteristics of superspreaders (top 10% of infectors) compared to other cases. **A** Mean number of secondary cases. Bars indicate the mean estimate, error bars indicate the 95%CrI across the 125000 reconstructed chains. **B** Infection diagnostic delays. **C** Frequency of gastrointestinal symptoms per day. Boxplots represent the mean, IQR and 95%CrI of the mean across the 125000 reconstructed chains.*

The calibration of the branching process model to outbreak data resulted in an estimated value for the basic reproduction number R_0 of 15.6 (95%CrI 11.5-20.4) and for the overdispersion in the offspring distribution ν of 0.13 (95%CrI 0.10-0.18) (see Appendix C). We found that, even in absence of any control strategies, the effective reproduction number on board decreased to 9.8 (95% Prediction Interval (PI) of the mean: 7.1-12.7) due to the short duration of the cruise, since a fraction of secondary cases would be transmitted by infectious individuals after disembarkation (Figure 3-6A). The isolation of cases in their cabins for 24, 48 and 72 hours further reduced the reproduction number to 7.8 (95%PI: 5.3-10.5), 6.0 (95%PI: 4.2-7.9) and 4.9 (95%PI: 3.0-7.1) respectively. Reductions and increases by 50% in observed diagnostic delays after symptoms (maintaining isolation for 72 hours as in the baseline) would have a mild impact on the effective reproduction number. Even in a perfect control scenario where all cases are diagnosed and isolated instantaneously after symptom onset and until the end of the cruise, the reproduction number would still be significantly above the epidemic threshold at 3.8 (95%PI of the mean: 2.1-5.7). Perfect isolation would have led to averting on average 89% of cases that would occur in absence of isolation (Figure 3-6B). We estimate that the implemented control scenario with 72-hours isolation of a case after diagnosis averted about 71% of the potential cases. The proportion of averted cases goes down to 60% in the case of 72-hours isolation but with 50% longer symptom diagnostic delays, and up to 81% with 50% shorter symptom diagnostic delays. Reducing the isolation duration to 48 and 24 hours respectively would avoid 67% and 50% of cases compared to a no-intervention scenarios. The probability of observing an outbreak, which is defined as a cruise where the number of total cases exceeds 2% of the passenger population, was above 90% in all scenarios, except those envisioning a marked reduction of diagnostic delays (Figure 3-6C).

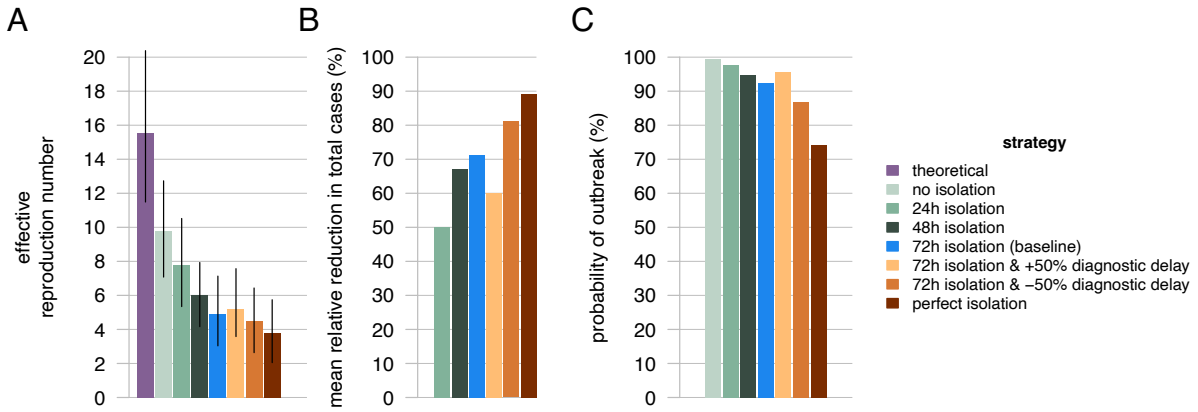

*Figure 3-6. Impact of alternative control scenarios. **A** Effective reproduction number. For the theoretical basic reproduction number, the bar represents the mean and error bar represent the 95%CrI of the posterior distribution estimated. For scenarios, bars indicate mean values over 100 simulations and error bars indicate the 95%PI of the mean. **B** Mean relative reduction in the total expected cases compared to a scenario with no interventions. **C** Probability of an outbreak (defined as >2% of passengers infected).*

Discussion

We reconstructed likely transmission chains during a large norovirus outbreak on a cruise ship calling Mediterranean ports, providing insights in the transmission patterns on board. We estimated that the majority of cases occurred in public spaces on the cruise ship (almost 90%), with limited transmission in cabins and a negligible number of infections acquired during tours in visited ports. We identified a strong degree of

superspreading, with 20% of patients being responsible for almost 80% of the infections diagnosed on board. Delays between infection and diagnosis emerged as a key driver of outbreak dynamics. The association of longer diagnostic delays with higher numbers of secondary cases corresponds to the increased opportunities to transmit the infection to others through social contacts on the ship before being isolated in their cabins for 72 hours. The top 10% of infectors, who transmitted on average 12.5 times more cases than the rest of the infected individuals, were characterized by diagnostic delays that were on average twice the ones for the rest of cases. The halved frequency of gastrointestinal symptoms in the top 10% of infectors compared to the rest of cases may explain why these individuals had longer delays in health-seeking behavior, indicating that the difficulty in timely identifying mild cases may play a critical role in transmission dynamics. Awareness of the existence of norovirus spread on-board after the first diagnosed cases, diffused among passengers via information campaigns on board and word-of-mouth, may have led to the reduction over time of diagnostic delays, facilitating outbreak management. Results were robust when considering the possibility of a non-negligible fraction of infections remaining underreported at the end of the cruise [129,130].

Simulations based on a branching process model estimated the basic reproduction of norovirus on board of the cruise at about 15, a value that is compatible with high values estimated in other crowded settings [50,127]. This theoretical value does not correspond to the actual average number of secondary infections ("effective reproduction number") on board because a proportion of all infections would occur after the end of voyage duration, which was 7 days in the considered cruise. Longer voyages are likely subject to higher effective reproduction numbers. The implemented protocol of isolation for 72 hours after diagnosis reduced the estimated effective reproduction number to 4.9 (95%PI 3.0-7.1). This value is consistent with the 95%CrI of 4.4-14.5 for individuals who carried the virus at embarkation and of 1.5-5.2 for individuals infected on the first day, estimated by the transmission chain reconstruction model.

We estimated that the implemented isolation protocol has avoided about 71% of potential cases that would be observed in its absence. Choosing protocols with shorter durations of isolation would have avoided significantly less cases, while halving diagnostic delays after symptoms might further improve the effectiveness of the protocol, increasing the proportion of avoided cases to 80%. Even in absence of diagnostic delays and with permanent isolation of cases, the reproduction number would still remain largely above the epidemic threshold of 1 and a >70% probability of observing a cumulative attack rate among passengers higher than 2%. Previous modeling studies [50] have estimated the effectiveness of isolation in reducing the outbreak size, but did not consider the effect of diagnostic delays, assuming sudden isolation of cases.

This study has several limitations. First, we did not have sufficient data to explicitly model potential environmental or foodborne transmission. Although the literature suggests direct person-to-person transmission is predominant [50], environmental contamination could also play a role in norovirus outbreaks and requires further investigation. Second, the present analysis is based on a single outbreak, albeit characterized by a large cumulative incidence (9.7% among passengers), and caution should be exercised when generalizing the findings to other settings or outbreaks. The unique conditions of a cruise ship may not fully represent other environments where norovirus spreads. Our finding

that the risk of importation from port visits is negligible may not hold for other geographical contexts with higher endemic norovirus prevalence and less strict hand hygiene and food handling norms. Given the small number of cases occurring in crew, we approximated transmission as occurring entirely in the passenger population and we could not study the effect of isolation separately on the passenger and crew populations.

Our findings confirm the difficulty of halting norovirus transmission on board, and the fundamental role of timely isolation of cases to control transmission in confined settings like cruise ships. The low frequency of symptoms and long diagnostic delays in superspreaders suggest the importance of passenger education towards seeking immediate health assistance when experiencing gastrointestinal symptoms on board, to reduce the probability of outbreaks and their cumulative incidence.

Appendix C

Inference of transmission pairs

At any time t, we assume that each susceptible individual j is exposed to a force of infection (FOI):

$$\lambda_j(t) = \lambda_j^C(t) + \lambda_j^S(t) + \lambda_j^O(t) \tag{3.1}$$

where $\lambda_j^C(t)$ represents the FOI from infected cabinmates, $\lambda_j^S(t)$ represents the FOI from infected individuals onboard the ship outside cabin (in public spaces), and $\lambda_j^O(t)$ represents the FOI from infected individuals offboard during visits at ports.

The FOI from infected cabinmates is defined as:

$$\lambda_j^C(t) = \sum_{i \in C_j} \lambda_{ji}^C(t) = \sum_{i \in C_j} \beta_C \rho_i(t) \Gamma(t - \tau_i; a, b)\, \Theta\left(h \leq h_C^{(end)}\right) \tag{3.2}$$

where:

- C_j is the set of individuals who share the same cabin as individual j;
- β_C is a free model parameter scaling the transmissibility inside cabins;
- $\rho_i(t)$ is the relative increase in infectivity of the potential infector i based on their symptoms status at time t: $\rho_i(t) = \rho_i^v = 2.120$ if the individual i had at least one vomit episode, $\rho_i(t) = \rho_i^d = 1.390$ if the individual i had at least one diarrhea episode, $\rho_i(t) = \rho_i^v \rho_i^d = 2.947$ if the individual i experienced both kinds of symptomatic episodes [128], $\rho_i(t) = 1$ if the individual i did not have symptoms at time t;
- $\Gamma(x; a, b)$ is the discretized distribution of the generation time, i.e. the time that elapses between infection episodes of an infector-infectee pair, assumed to be distributed as a Gamma with shape $a = 3.35$ and rate $b = 0.92$ (mean generation time: 3.65 days) [127];
- τ_i represents the time of infection of individual i;
- $\Theta(X)$ is a Heaviside step function that is 1 when the condition X is true, and 0 otherwise. It is devised in such a way that the FOI within the cabin occurs only between midnight and 08:00, when passengers are assumed to withdraw in their cabins for sleeping;

- h represents the hour of the day and is calculated by using the modulo operation on the total time t, with respect to a 24-hour cycle. In other words, $h = t_{mod\ 24}$. For example, if $t = 26$ then $h = 26_{mod\ 24} = 2$, corresponding to 02:00 in the morning. Time t is initialized at $t = 0$, which corresponds to the beginning of the first day of travel at midnight (00:00);
- $h_C^{(end)}$ =08:00 is the hour at which we assume passengers to leave the cabin.

The FOI in public spaces accounts for the contribution from infected individuals (both cabinmates and others) during daytime. This FOI is only applicable to a potential infector-infectee pair if neither individual is in quarantine at the time being considered and at least one of the following condition holds: (i) it is a day of navigation at sea, i.e. without a stop in a port; (ii) the time falls before or after times of disembarkment for tour; (iii) both the infector and the infected stay on board of the ship despite the possibility of disembarking to visit the port destination.

The FOI from public spaces of the ship is defined as:

$$\lambda_j^S(t) = \sum_{i \neq j} \lambda_{ji}^S(t) = \sum_{i \neq j} \beta_S \phi_{ij} \rho_i \Gamma(t - \tau_i; a, b) \cdot \Big(1 - q_j(t)\Big)\Big(1 - q_i(t)\Big) \cdot \Big\{\Big[\big(1 - \delta(t)\big) + \delta(t)\big(1 - p_j\big)(1 - p_i)\Big]\Theta\Big(h > h_O^{(start)}\Big)\Theta\Big(h \leq h_O^{(end)}\Big) + \Theta\Big(h > h_C^{(end)}\Big)\Theta\Big(h \leq h_O^{(start)}\Big) + \Theta\Big(h > h_O^{(end)}\Big)\Big\} \tag{3.3}$$

where:

- β_S is a free parameter scaling the transmissibility onboard in public spaces of the ship;
- $\phi_{ij} = \begin{cases} 1 & \text{if } i \notin C_j \\ \phi & \text{if } i \in C_j \end{cases}$ and $\phi \geq 1$ is a free parameter accounting for increased relative transmissibility in public areas among cabinmates, who may be more likely to contribute more to the FOI of a cabinmate even when outside the cabin, since they are more likely to spend time in close contact;
- $q_j(t)$ is a Kronecker delta-function accounting for quarantine of individual j over time: it is one if individual j is quarantined in cabin at time t and zero otherwise.
- $\delta(t)$ is delta-function that equals one if the day associated to the time t is a day of port stop of the cruise, and zero otherwise;
- p_j is the probability that the individual j goes on the daily visit at ports. We assume that $p_j = 90\%$, if j is a passenger and $p_j = 0\%$ if j is a crew member (i.e., the disembarkment of crew members for visiting ports is considered negligible);
- $h_O^{(start)}$=10:00 and $h_O^{(end)}$=18:00 are the hours at which we assume the port visits start and end, respectively.
- $\Theta(X)$ is devised in such a way that the FOI in public spaces occurs only between 08:00 and 10:00 or between 18:00 and midnight, when passengers are assumed to be outside cabins. During days without port stop, $\Theta(X)$ allows the FOI in public spaces to occur also between 10:00 and 18:00;

The FOI from the general port population accounts for the contribution from infected individuals encountered offboard during the designated post visit hours. This FOI is only relevant if the individual j visits a port and is not in quarantine, and is defined as:

$$\lambda_j^O(t) = \beta_P \alpha p_j \delta(t) \left(1 - q_j(t)\right) \Theta\left(h > h_O^{(start)}\right) \Theta\left(h \leq h_O^{(end)}\right) \quad (3.4)$$

where α is a free parameter representing the relative prevalence of norovirus in the general port population and β_P is a free parameter scaling the transmissibility offboard during post visits. $\Theta(X)$ is devised in such a way that the FOI during visits occurs only between 10:00 and 18:00.

A schematics of a typical day schedule is illustrated in Figure 3-7.

The model assigns a source of infection k_j for all cases by choosing from sources of one of the three settings of transmission (an infector within cabins, an infector in public spaces onboard, or a generic source of infection in offboard port visits) with probability π_{ji} proportional to the contribution of each source to the total FOI at the time τ_j at which j was infected:

$$\pi_{ji} = \begin{cases} \lambda_{ji}^C(\tau_j), & \text{for all infectors } i \text{ in the cabin} \\ \lambda_{ji}^S(\tau_j), & \text{for infectors } i \text{ in public spaces of the ship} \\ \lambda_j^O(\tau_j), & \text{for transmission offboard during port visits} \end{cases}$$

The set of sources of infections k_j represents the probabilistically reconstructed transmission chain for a given set of parameters and imputed infection times.

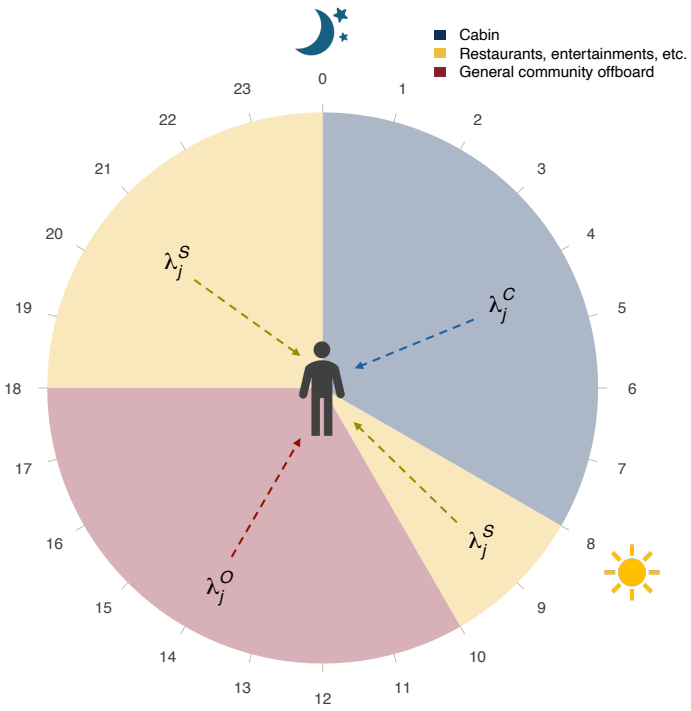

Figure 3-7. Instance of daily schedule and exposure to force of infections on days with tours offboard.

Initialization of times of diagnosis and infection

Since for each GI case on the ship only the date of diagnosis is known (measured in days from the cruise start) and not the specific time of diagnosis, we first impute for each case

j the time of diagnosis d_j by sampling uniformly across the date of diagnosis, excluding night hours (between midnight and 8am).

For each case j, we then initialize an imputed time of infection τ_j by sampling values with probability $P_I(\tau)$ of being infected on the time step τ, defined as:

$$P_I(\tau) = P_s(s - \tau) \tag{3.5}$$

where $P_s(x)$ is the discretized (on time steps of length 1 hour) and normalized probability distribution of the incubation period, i.e., the probability to develop symptoms after a time x since infection. The continuous probability density function of the incubation period used to build $P_s(x)$ is a log-normal with parameters: $\mu_s = 0.182$ days and $\sigma_s = 0.494$ days (mean incubation period: 1.36 days) [132].

The index case (or cases) that brought the virus on board and started the outbreak must have been infected either before embarkation or during the first port visits. In order to impute the index case(s), we compare the minimum value among the imputed infection times, τ_m, with the time of first embarkation on the ship, T_{emb} (assumed to be at 19: 00 on the first day), and with the time of first port visit, T_f. If $\tau_m < T_{emb}$, then the index case was imported at first embarkation on the ship and no further operation is needed. Any other case with time of infection before T_{emb}, besides the one corresponding to τ_m, was considered as imported. If $\tau_m < (T_{emb} + T_f)/2$, i.e. the minimum imputed time of infection is closer to first embarkation than to the first port visit, then we still assume that the index case was also imported at first embarkation, and we forcedly reassign the time of infection for the index case as $T_{emb} - 1$. All other cases are assumed to be infected during the cruise, including during visits at ports. If $\tau_m > (T_{emb} + T_f)/2$, we forcedly reassign all times of infection comprised between $(T_{emb} + T_f)/2$ and T_f to T_f, under the assumption that they were all index cases infected during the first port visit.

Calibration

After imputing the times of diagnosis and the starting points for the times of infection of all cases, we estimated the free model parameters $(\alpha, \beta_c, \beta_s, \beta_P, \phi)$, the unknown times of infection τ_j and the source of infection k_j for each case j, using a Monte Carlo Markov Chain (MCMC) procedure. The MCMC calibration was then repeated for Z=50 times different model initializations.

The overall likelihood of the observed cases, given the set of parameters $\omega = (\alpha, \beta_c, \beta_s, \beta_P, \phi)$, the times of infection τ_j, and the sources of infection k_j, is given by:

$$\mathcal{L}(\omega, \tau_j, k_j) = \prod_j P_j Q_j W_j I_j \tag{3.6}$$

where P_j represents the likelihood that j was infected by k_j:

$$P_j = \begin{cases} \lambda^C_{jk_j}(\tau_j), & \text{if transmission occurred in the cabin} \\ \lambda^S_{jk_j}(\tau_j), & \text{if transmission occurred in public spaces of the ship} \\ \lambda^O_j(\tau_j), & \text{if transmission occurred offboard during port visits} \\ 1, & \text{if } j \text{ is uninfected} \end{cases} \quad (3.7)$$

Q_j represents the likelihood that j was not infected:

$$Q_j = \begin{cases} e^{-\int_0^{\tau_j} \lambda_j(t)dt}, & \text{if } j \text{ is infected} \\ e^{-\int_0^{\infty} \lambda_j(t)dt}, & \text{if } j \text{ is uninfected} \end{cases} \quad (3.8)$$

The factor W_j is the contribution to the likelihood of the incubation period for individual j, that is $W_j = P_s(s_j - \tau_j)$, where s_j is the known time of symptom onset and τ_j is the inferred time of infection. Finally, $I_j = \alpha$ if the individual j has been imported from outside before embarkation, and $I_j = 1 - \alpha$ if the infection of j occurred onboard or if j remained uninfected during the whole voyage.

At each step, all parameters (ω and infection times τ_j) are updated using reversible normal jumps. The MCMC algorithm was run for 400000 iterations, with the first 300000 discarded as burn-in. We discarded (assigning a zero likelihood) all the realizations with infection times that involve a situation with no possibility of transmission (e.g., the infection time of a given individual is sampled during night hours, but there are no other cabinmates infected yet). M=2500 samples were drawn from the posterior distributions obtained by the MCMC. Results from the $M \times Z = 125000$ MCMC were pooled together to obtain the final parameter distribution and the distribution of the sources of infection for each case. Figure 3-8 reports the posterior distributions of estimated parameters.

The code used to run the model will be available upon publication of the paper at: github.com/alfredodebellis .

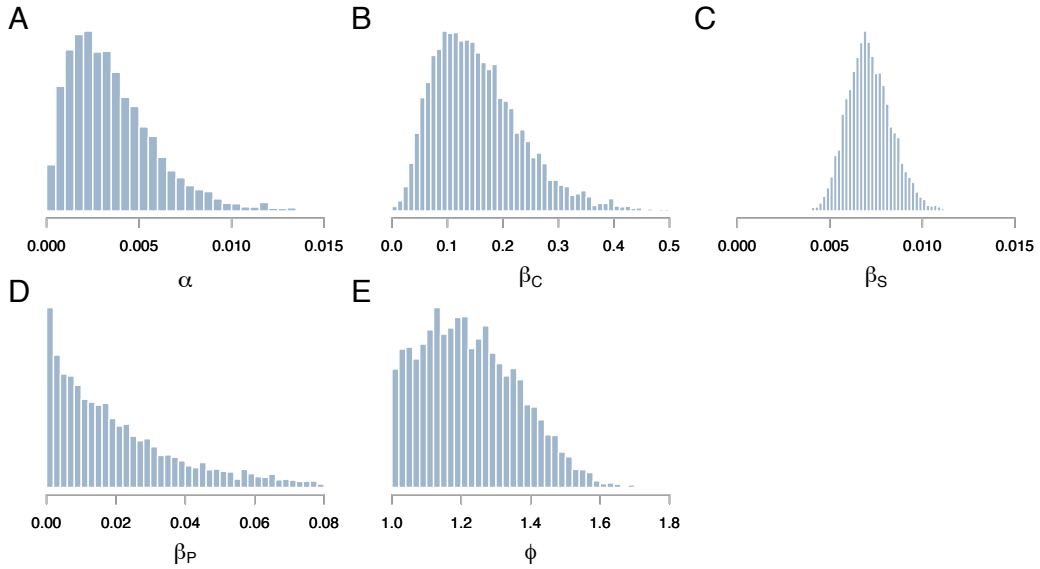

*Figure 3-8. Pooled posterior distributions of free parameters of the model obtained by MCMC. **A** norovirus prevalence in the general population; **B** transmission rate within cabins; **C** transmission rate in public spaces; **D** transmission rate offboard during port visits; **E** increased relative transmissibility in public spaces for cabinmates.*

Sensitivity analyses with underreporting

Available estimates suggest that between 23% and 40% of cases may go undetected in a norovirus outbreak on cruise ships [129,130]. In the baseline analysis, we implicitly assumed that all infections in the outbreak have been reported. In this section we test the robustness of our conclusions when accounting for underreporting. To tackle this question, we apply the same model by applying it to a line list of cases where an additional number of underreported cases is synthetically generated in such a way that the total proportion of underreported cases is 23% or 40%. For each synthetically generated case, the symptom onset date is assigned with multinomial probability based on the observed time series by symptom onset. The hour of symptom onset is assigned with uniform probability throughout the day of symptom onset. We sampled whether the individual had vomiting or diarrhea with probabilities given by the proportions of cases with vomiting ($p_v = 79\%$) or diarrhea ($p_d = 99\%$) observed during the cruise among reported cases.

The figures below demonstrate that our main conclusions from the baseline analysis remain consistent, even when underreporting is included into the model. Figure 3-9 shows trends similar to the baseline for both the distribution of secondary cases generated by infectors and the proportion of secondary cases ranked by infectors.

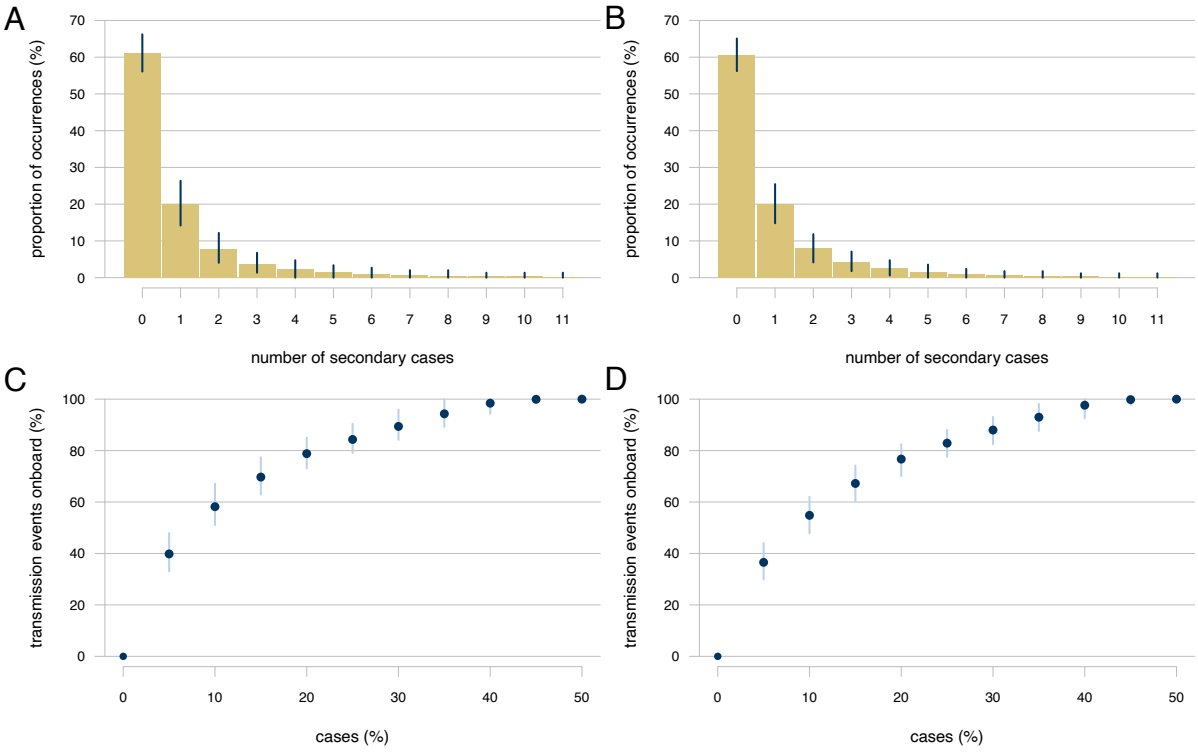

*Figure 3-9. **Top**: Distribution of the number of secondary cases generated by infectors. The bars indicate the average value, while the error bars indicate the 95%CrI over all the reconstructed chains. **A** Underreporting of 23%; **B** Underreporting of 40%; **Bottom**: Cumulative proportion of secondary cases ranked by infectors. The points indicate the average value, while the error bars indicate the 95%CrI over all the reconstructed chains. **C** Underreporting of 23%; **D** Underreporting of 40%.*

We define the "effective infectious period" as the interval during which individuals may transmit the infection. For reported cases, this corresponds to the time between the infection and diagnosis, while for underreported cases it is the time between the infection and the end of the cruise. Figure 3-10 shows that the mean number of secondary cases generated by infectors increases with a longer effective infectious period, consistent with

findings from the baseline analysis. Figure 3-11 illustrates the differences in the effective infectious period and the mean number of secondary cases between superspreaders (top 10% infectors) and other cases. Consistent with results from the baseline analysis, both the effective infectious period and the mean number of secondary cases are significantly higher for superspreaders. Finally, Figure 3-12 shows, in line with the baseline analysis, a decreasing trend in the distributions of the effective infectious period over time of infection of cases (from embarkation), along with a decaying distribution of the time-varying reproduction number.

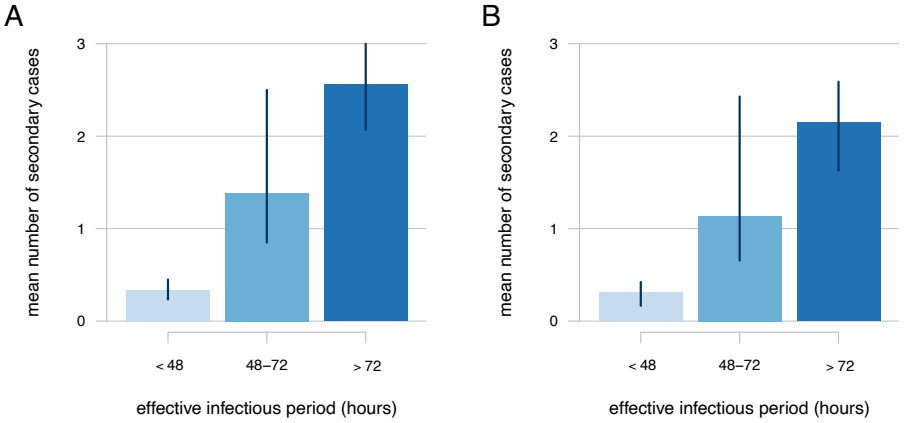

Figure 3-10. Number of secondary cases stratified by effective infectious period. The bars indicate the average value, while the error bars indicate the 95%CrI over all the reconstructed chains. ***A*** *Underreporting of 23%;* ***B*** *Underreporting of 40%.*

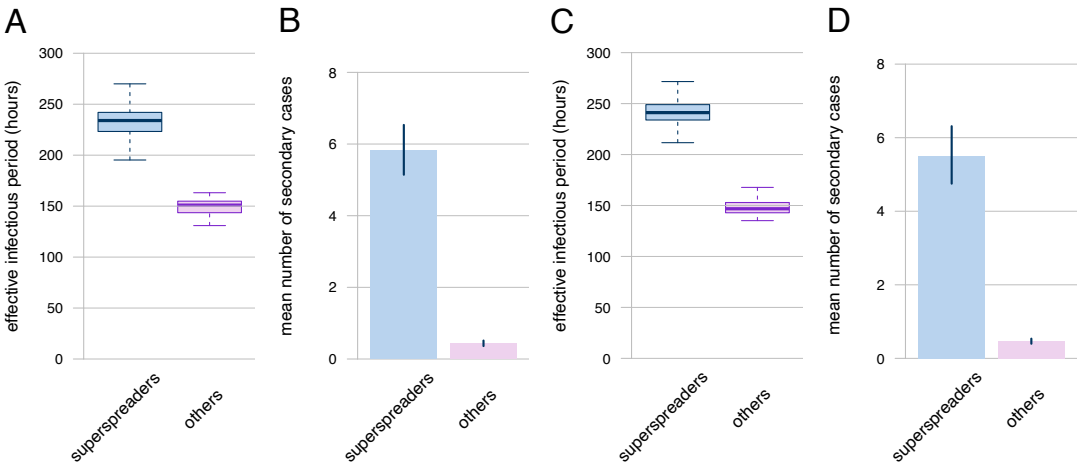

Figure 3-11. Characteristics of superspreaders compared to other infectors. ***A,C*** *Boxplot of effective infectious period distribution of superspreaders (top 10% of infectors) and of other cases;* ***B,D*** *Average number of secondary cases generated by superspreaders (top 10% of infectors) and by other cases. The bars indicate the average value, while the error bars indicate the 95%CrI over all the reconstructed chains.* ***A,B*** *Underreporting of 23%;* ***C,D*** *Underreporting of 40%. We estimate that the average percentage of not diagnosed superspreaders is 27% (95%CrI: 7% - 47%) and 43% (95%CrI: 23% - 63%) when underreporting is 23% and 40%, respectively.*

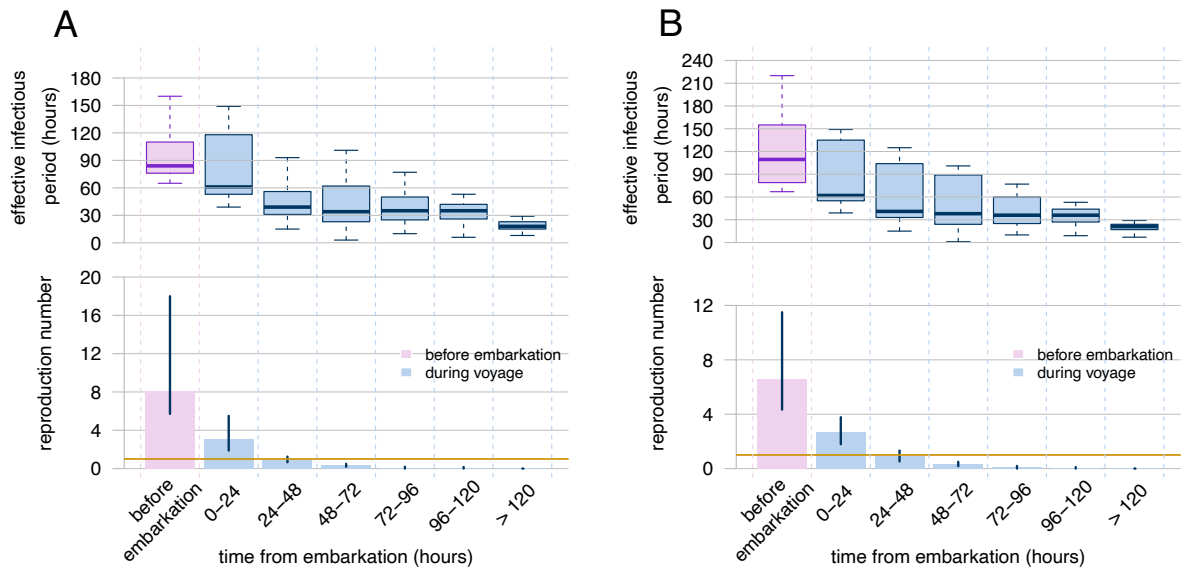

Figure 3-12. Reproduction number and diagnostic delay over the course of the outbreak. The box plots show the distribution of effective infectious period (in hours) for cases infected each day of the voyage (top). The bar plot (bottom) represents the estimated reproduction number for cases before and during the voyage. The bars indicate the average value, while the error bars indicate the 95%CrI over all the reconstructed chains. ***A*** *Underreporting of 23%;* ***B*** *Underreporting of 40%.*

Branching process

Model description

To assess the overall effectiveness of alternative isolation protocols and the impact of diagnostic delays, we implemented a stochastic branching process model [131] that allows to simulate the generation of secondary cases in the considered outbreak under different assumptions. In the branching process, each infector generates a random number of secondary infections, which is sampled from an offspring distribution. The mean of this distribution corresponds to the basic reproduction number (R_0). Based on the evidence of superspreading that we found via transmission chain reconstruction, we selected a negative-binomial distribution for the offspring distribution, with a mean of R_0 and an overdispersion parameter ν [22]. The branching process was applied to the total population carried on the ship, with no distinction between passengers and crew. Within this population, we assumed homogeneous mixing and disregarded heterogeneities in transmissibility by symptom status.

In the data, two individuals had symptom onset on day 2 of the voyage. Accordingly, we initialized the model with two index cases, assuming these cases were imported from outside the population. Their infection times were sampled from the incubation period distribution considered before and constrained to occur before embarkation (since a negligible number of individuals was infected in port visits, according to the reconstructed chains.

Then, for each newly infected individual i_t (where $i_t = 1, \dots, I_t$), we sample the number of secondary cases J_{i_t} they generate from the negative-binomial offspring distribution:

$$J_{i_t} \sim NegBin(R_0 \cdot \rho(t), \nu) \tag{3.9}$$

where $\rho(t)$ represents the proportion of susceptible individuals at time t, computed as the difference between the total population size, $N_{tot} = N_{pax} + N_{crew} = 1229 + 487 =$

1716, and the cumulative number of infections before time *t,* $N_I(t)$*,* rescaled by N_{tot}.

For each secondary case j_{i_t} (where $j_{i_t} = 1, \dots, J_{i_t}$), we assign an infection time by sampling from the generation time distribution; infections occurring during the infector's isolation period, which begins upon diagnosis and lasts for a time that depend on the considered scenario, were discarded and considered as averted by the intervention; those occurring after the end of the cruise were also discarded. As a result, the actual number of realized infections was always lower than the theoretical number of secondary cases sampled from the offspring distribution even in scenarios without isolation. For each infection that was not discarded, we assign a time of symptom onset by adding to the time of infection a sample from the incubation period; and we assign a time of diagnosis by adding to the time of symptom onset a sample from a diagnostic delay distribution which depends on the day of symptom onset of the case (see Section "Empirical diagnostic delay distribution" below) and on the isolation scenario. Secondary infections for which diagnosis occurred after the end of the voyage were considered as unreported but could contribute to transmission on-board.

Calibration

To calibrate the branching process model to the observed time-series of norovirus cases by symptom onset and diagnosis, we used a combination of grid search and particle filtering [133]. The two free model parameters were the basic reproduction number R_0 and the overdispersion of the offspring distribution ν.

For particle filtering, we used a discretized version of the branching process model described above, with a time step of Δt = 1 hour. For a given parameter set, the performance of the model was evaluated at each time step using a mean-squared error (MSE) function, comparing the number of reported cases in simulated trajectories against the observed time series, aggregated by time of symptom onset over 4-hour intervals and by date of diagnosis over 24-hour intervals. At each aggregated time of analysis (i.e., $t_A^S = 0 - 4\ hours, 4 - 8\ hours, 8 - 12\ hours$, …, for trajectories by time of symptom onset and $t_A^R = 0 - 24\ hours, 24 - 48\ hours, 48 - 72\ hours$, …, for trajectories by date of diagnosis), the MSE of a particle w is defined by:

$$\mathrm{MSE}_w^{\mathrm{S}}(t_A^S) = \mathrm{MSE}_w^{\mathrm{S}}(t_A^S) = \left(X_w^S(t_A^S) - D^S(t_A^S)\right)^2 \tag{3.10}$$

and

$$\mathrm{MSE}_w^{\mathrm{R}}(t_A^R) = \mathrm{MSE}_w^{\mathrm{R}}(t_A^R) = \left(X_w^R(t_A^R) - D^R(t_A^R)\right)^2 \tag{3.11}$$

where $X_w^S(t_A^S)$ is the trajectory of particle w by time of symptom onset; $X_w^R(t_A^R)$ is the trajectory of the particle w by date of diagnosis; $D^S(t_A^S)$ is the time series of observed cases by time of symptom onset; $D^R(t_A^R)$ is the time series of observed cases by date of diagnosis. The resulting total score of particle w was given by $1/\mathrm{MSE}_w$, where:

$$\mathrm{MSE}_w = \sum_{t_A^S} \mathrm{MSE}_w^{\mathrm{S}}(t_A^S) + \sum_{t_A^R} \mathrm{MSE}_w^{\mathrm{R}}(t_A^R) \tag{3.12}$$

The particle filtering algorithm consists of two steps:

1) At each time step t, W=1000 particles were generated and the M=100 with the best scores (i.e., the 100 lowest MSE_w) were selected;
2) From the top M particles, W particles were sampled with replacement, with probabilities weighted by their score. These W particles were then used to reinitialize the model to generate the next time step ($t + 1$).

The frequent resampling of particle filtering can lead to the so-called degeneracy problem, where only a few particles dominate the simulation, reducing the variability necessary to accurately represent the outbreak dynamics [133]. To prevent this and ensure sufficient diversity among the particles, we repeated the particle filtering process S=5 times, each time starting from different random seeds.

The selected M particles from all S procedures were pooled together and the final score for the given parameter set was given by their average mean squared error:

$$\text{MSE} = \frac{1}{S}\sum_{s}\frac{1}{M}\sum_{m}\text{MSE}_m \tag{3.13}$$

To estimate parameter values, we evaluated the MSE for values of R_0 between 8 and 24, with steps equal to 0.25, and values of v from 0.1 to 0.6, with steps equal to 0.02, using a grid search (i.e., testing all combination of parameter values, corresponding to a total of 65x26=1690 parameter sets explored). Figure 3-13 illustrates the values of the MSE across the explored parameter sets. Considering the K=100 parameter sets with lowest MSE, the estimated average basic reproduction number is 15.6 (95%CrI 11.5-20.4) and the estimated average overdispersion is 0.13 (95%CrI 0.10-0.18). Figure 3-14 shows the fit of the $M \times S \times K$ trajectories against observed data.

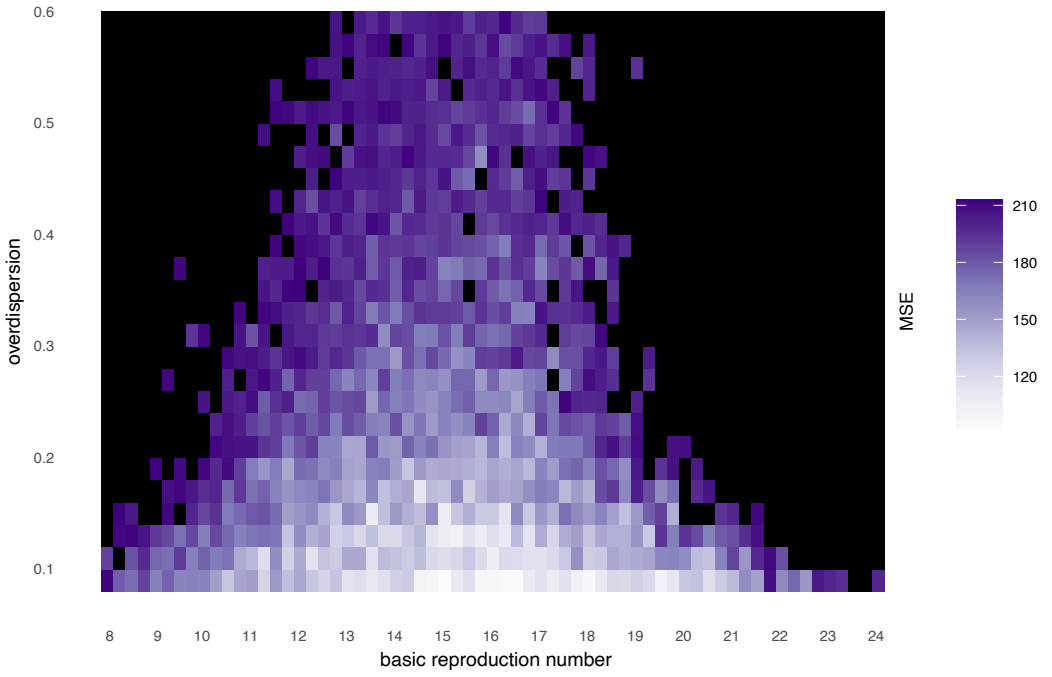

Figure 3-13. Grid of mean squared error (MSE) values across parameter sets for basic reproduction number and overdispersion. Lower MSE values indicate better fits of the model to observed data.

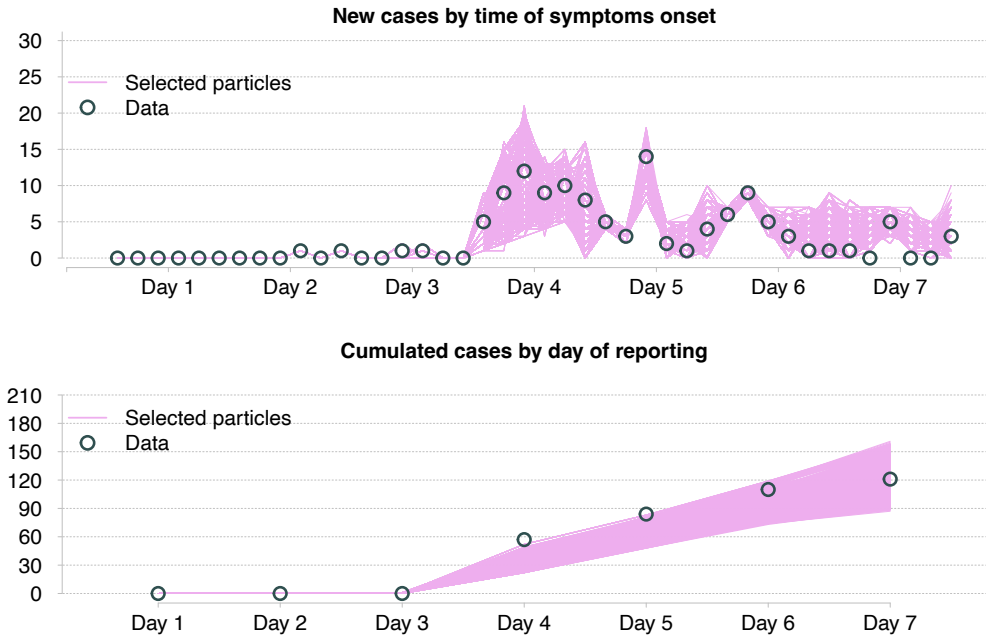

Figure 3-14. Comparison of pooled selected particles from the branching process model for the top 100 parameter sets with actual data for the time series. ***Top****: New cases by time of symptom onset;* ***Bottom****: Cumulative cases by reporting day (day of diagnosis).*

Empirical diagnostic delays distributions

The diagnostic delay from symptom onset refers to the interval between the onset of symptoms and the diagnosis of cases. Since the available data only included the day of diagnosis, not the specific time, the probability distribution of the diagnostic delay was estimated by imputing for all cases a diagnostic time sampled uniformly over the day, under the constraint that diagnoses occurred during daytime (between 08:00 and 23:59) and always after the onset of symptoms (since no norovirus screening was implemented on the ship). The imputation process was repeated N_A=50 times, and a distribution of diagnostic delays was obtained by pooling together diagnostic delays from the N_A imputations. Because cases with later symptom onset tended to have smaller diagnostic delays, due to the right censoring imposed by the end of the cruise, we estimated different diagnostic delay distributions based on the day of symptom onset (Figure 3-15).

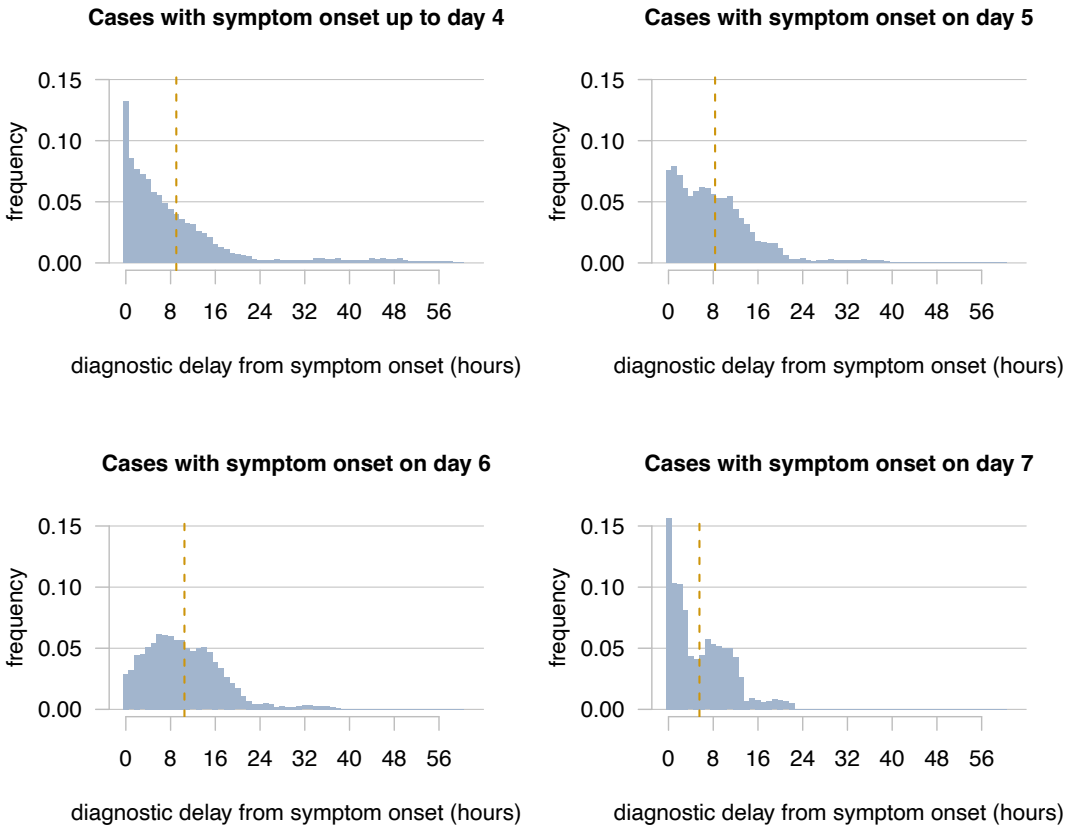

Figure 3-15. Empirical diagnostic delay distribution, disaggregated by day of symptom onset. Vertical dashed lines indicate mean values of distributions.

Output analysis

A schematic description of scenarios considered is presented in Table 3-2. Results for all scenarios were obtained by running L=100 stochastic iterations with the top K parameter sets.

We computed three main outputs for each scenario: the effective reproduction number, the relative change in the total number of cases observed during the cruise, and the probability of an outbreak.

Table 3-2. Synopsis of scenarios.

Name	Description	Duration of isolation	Diagnostic delay distribution
S_0	baseline	72h	empirical
S_1	no isolation	0h	empirical
S_2	24h isolation	24h	empirical
S_3	48h isolation	48h	empirical
S_4	baseline isolation, increased diagnostic delay	72h	increased by 50%
S_5	baseline isolation, reduced diagnostic delay	72h	reduced by 50%
S_6	perfect isolation	until end of cruise	no delay (immediate diagnosis upon symptoms)

The effective reproduction number was computed by averaging the mean number of secondary infections caused by cases that were either imported before embarkation or infected within the first two days of the voyage. This choice is based on the observation that the number of secondary cases tends to decline sharply for infections occurring later in the cruise due to the right truncation caused by the end of the cruise (see Figure 3-16 for the reproduction number disaggregated by time of infection of the cases).

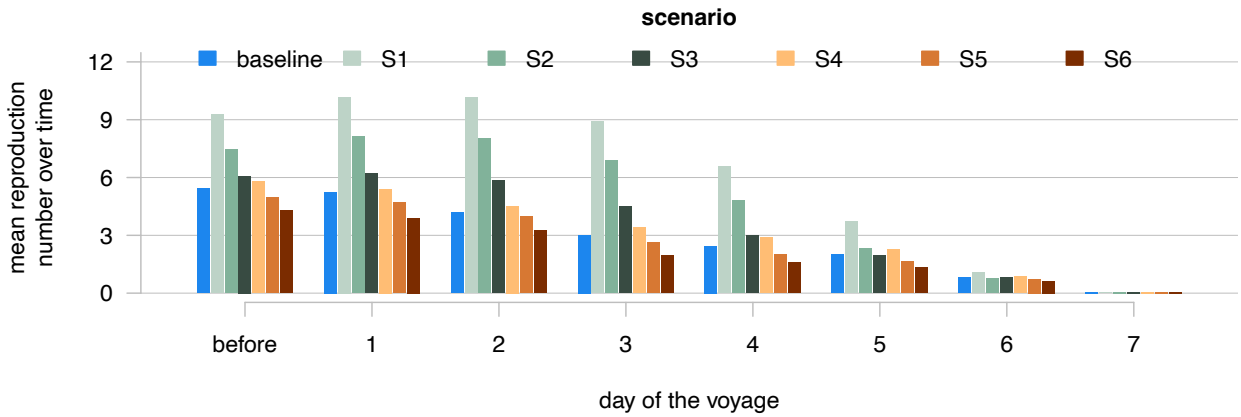

Figure 3-16. Mean time-dependent reproduction number R_t across alternative scenarios. Bars indicate mean values over 100 simulations.

For each scenario X, the relative change Δ_X in the total number of cases was calculated as the percentage change in the mean number of cases expected for the considered scenario, $\overline{C^X}$ (calculated over the $L \times K$ simulations), compared to the mean number of cases $\overline{C_{S1}}$ expected for scenario S1, where no isolation was implemented:

$$\Delta_X = 100 \cdot \frac{\overline{C_{S1}} - \overline{C_X}}{\overline{C_{S1}}} \tag{3.14}$$

where $\overline{C_X} = \frac{1}{K}\sum_k \frac{1}{L}\sum_l C_{X,k,l}$ and $C_{X,k,l}$ represents the number of cases generated in the l-th simulation with the k-th parameter set in scenario X.

The probability of an outbreak was determined as the proportion of the $L \times K$ simulations where the number of reported cases was larger than 2% of the total passenger population, corresponding to a threshold outbreak size of $S_{thr} = 0.02 \cdot N_{pax} = 25$ cases. The threshold was referred only to the passengers population as these represented the vast majority of reported cases in the considered data.

CHAPTER 4

Epidemiological impact of a prospective HIV cure

Introduction

Over the past decades, significant progress has been made toward achieving the United Nations Sustainable Development Goal of ending the HIV epidemic by 2030 [135]. Successes in curbing HIV transmission are particularly evident among men who have sex with men (MSM) in several European countries, including the Netherlands [136, 137,138]. HIV incidence in this population has been declining due to public health interventions such as pre-exposure prophylaxis (PrEP) [139] and test-and-treat [140].

Despite encouraging epidemiological trends, HIV continues to pose a substantial burden for the MSM population. HIV prevalence among MSM is disproportionately high compared to the general population [53,54]. HIV infection is lifelong, and MSM with HIV continue to experience impaired quality of life despite effective ART [141,142,143,144]. Compared to uninfected controls, MSM with HIV have worse physical and mental health due to chronic comorbidities, physical impairments, and social factors such as stigma, and negative interpersonal and social experiences [141,142,143]. Emerging biomedical technologies, such as an HIV cure, could improve the well-being of people with HIV in various life domains, similar to the impact of ART availability in the past [145]. Therefore, developing an HIV cure is considered a global priority [146,147,148].

Biomedical research and clinical trials for curative HIV interventions are advancing rapidly [55,149]. To date, five patients have been cured of HIV through invasive stem cell transplants [150], but a scalable cure has yet to be developed. The consensus is that an acceptable and scalable cure for HIV will most likely require a combination of strategies targeting different aspects of HIV infection [146,148,151]. An elimination, or sterilizing, cure, aiming for the complete removal of HIV from the body, represents the ultimate treatment goal. However, HIV post-treatment control (PTC), also known as a functional cure, where the virus remains suppressed below detectable levels without ongoing ART, could be a more attainable goal. Strategies that are currently being developed aim at both an elimination cure (gene editing, latency-reversing agents combined with immune stimulators, T-cell therapy, therapeutic vaccines, etc.) [146,151,152] and at PTC ('block and lock', CCR5 inactivation, and broadly neutralizing antibodies, etc.)

[151,153,154,155].

A tool commonly used to guide drug development, known as the target product profile (TPP), has been developed to align stakeholders around a set of attributes for a potential HIV cure [148]. The TPP outlines important characteristics of a cure such as target population, clinical efficacy, protection from re-infection, and time until viral rebound. The characteristics of a cure intervention could have major consequences not only for the quality of life of people affected by HIV but also for virus transmission at the population level. It is often hypothesized that a cure could aid in controlling the epidemic [146,147,155]. However, there is little research investigating whether and how a cure would affect HIV transmission dynamics. In a PTC scenario with a risk of viral rebound, individuals could become infectious again. In an elimination scenario, if a cure does not confer immunity, cured individuals could still be vulnerable to re-infection.

From the early days of the HIV pandemic, mathematical modeling played an important role in advancing our understanding of HIV epidemiology, evaluating the impact of interventions, and shaping public health strategies [156,157,158]. For example, the study by Granich et al. [140] demonstrated the effectiveness of test-and-treat in reducing transmission, significantly influencing the Joint United Nations Programme on HIV/AIDS policy to end the AIDS epidemic by 2030. Understanding how emerging HIV technologies, such as a potential cure, may affect the future course of the epidemic is key to their future successful population-based implementation. We investigated this question using a transmission model fitted to sexual behavior and epidemiological data for MSM in the Netherlands. We evaluated the impact of introducing PTC and elimination cure and explored under which assumptions regarding cure characteristics these cure scenarios may accelerate or reverse the observed decline of the HIV epidemic. Finally, we discussed the implications of our findings for the population-based implementation of cure interventions.

Results

Dynamics without cure

The model was fitted to data from the HIV Monitoring Foundation on the number of new HIV diagnoses and to the estimated number of undiagnosed cases from 2017 till 2022 (Figure 4-1) [136]. The model reproduced well the declining trends in new annual diagnoses and in the number of undiagnosed cases. The number of new cases per 100,000 persons per year was estimated to decrease from 74 (95% Prediction Interval (PI) 33–127) in 2017 to 39 (95%PI 14–67) in 2022, with a slower decline later. The estimated average time to diagnosis was 28 months (95% Credible Interval (CrI) 24–33) months, consistent with the Dutch Monitoring Foundation data [136]. The model predicted that the proportions of diagnoses within 6 months, between 6 to 12 months, and more than 12 months since HIV infection were 22% (95%PI 20%–24%), 14% (95%PI 13%–16%), and 63% (95%PI 60%–66%), respectively, also aligning with the data from the Dutch Monitoring Foundation [136]. The estimated HIV prevalence was almost constant, around 6%–7% throughout the considered period, which is in the range observed for MSM in Western Europe [53]. The model was additionally validated using independent data, not included in the model fitting, on the numbers of individuals on PrEP and ART, new imported cases on ART, and ART coverage among all infected individuals from 2017 till 2022 (Figure 4-12).

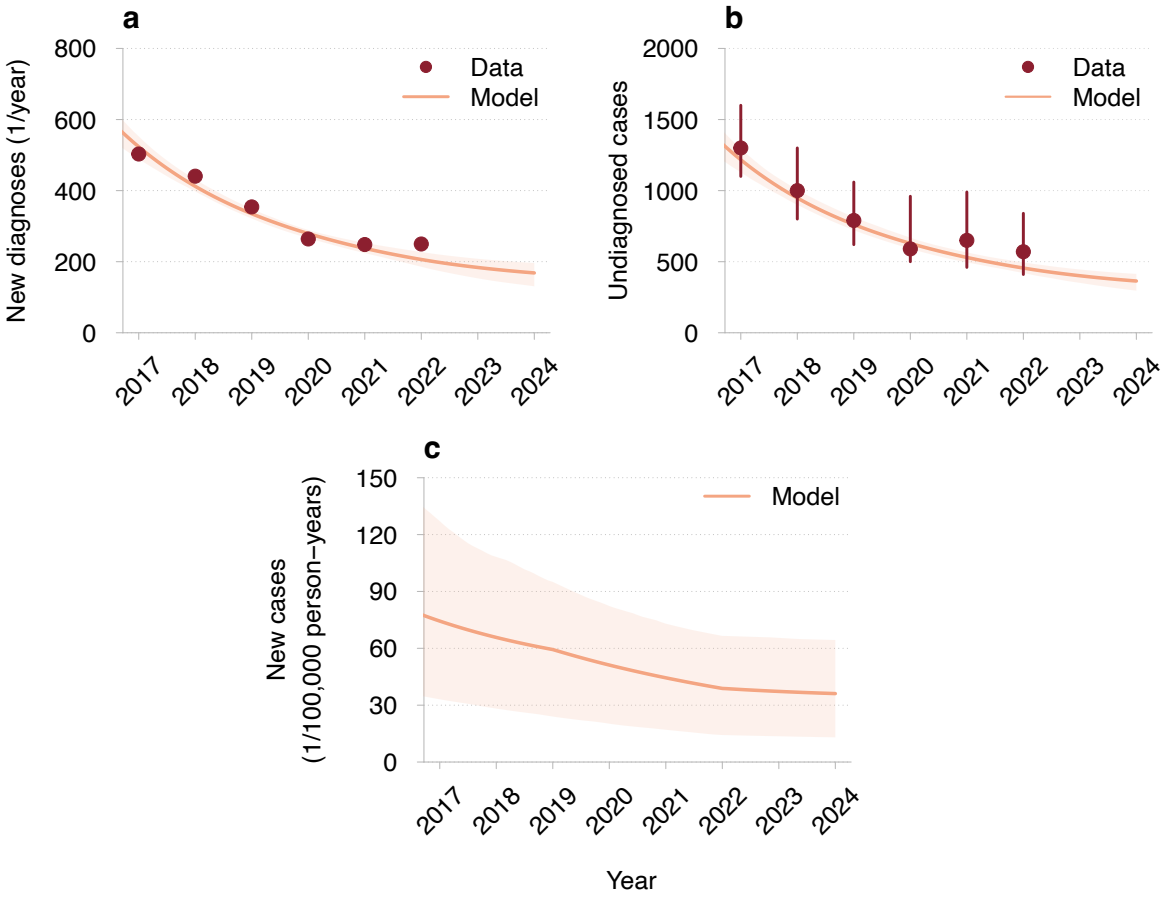

*Figure 4-1. Model fit to HIV surveillance data for MSM in the Netherlands. (**a**) New diagnoses, (**b**) estimated number of undiagnosed cases, and (**c**) new cases. The red dots and the error bars correspond to the mean estimates and the 95% confidence intervals in the data from the HIV Monitoring Foundation [2]. The mean trajectories estimated from the model are shown as orange lines. The orange shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution.*

Cure scenarios

Our analysis was guided by the TPP, which outlines the minimum and optimum characteristics of a cure intervention [148]. We developed a transmission model for the PTC and elimination scenarios (see Figure 4-6 and Methods). In both scenarios, the cure was targeted at virologically suppressed individuals on ART, aligning with the minimum requirement for the target population as specified in the TPP [148]. From an HIV dynamics perspective, the key difference between PTC and elimination is that, in the PTC scenario, the HIV reservoir within the individual is not completely removed, allowing for the possibility of viral rebound and potential onward transmission. In contrast, in the elimination scenario, all HIV including the rebound-competent reservoir, is fully removed, meaning the individual is cured but remains susceptible to re-infection. Individuals who experienced a viral rebound in the PTC scenario or re-infection in the elimination scenario were diagnosed with a delay that depended on the assumptions about viral load monitoring and testing of cured individuals. When exploring a prospective cure, many characteristics are not known in advance, such as efficacy (the proportion of individuals effectively cured), the annual uptake (the proportion of individuals cured each year among all eligible), or the diagnostic delay (the time between re-infection or rebound and diagnosis) for cured individuals. In both cure scenarios, we explored annual uptakes of 10%, 50% and 90%, and efficacy of 20% and 90%,

corresponding to the minimum and optimum values in the TPP [148]. Moreover, we considered three monitoring strategies, where the average (across all HIV stages) diagnostic delay for individuals who experienced a viral rebound or re-infection was (i) 28 months [136], reflecting the current standard of care, (ii) 3 months, corresponding to the testing interval for PrEP users [159], and (iii) two weeks, as frequently used in analytical treatment interruptions studies [160]. A cure intervention was assumed to be introduced in 2026 and to achieve maximum uptake within 3 years. We compared the epidemic trajectories for the two cure scenarios and the model without a cure, projecting outcomes until 2034. A detailed description of the model equations, parameters, and assumptions is provided in Methods, Figure 4-6, and Table 4-3.

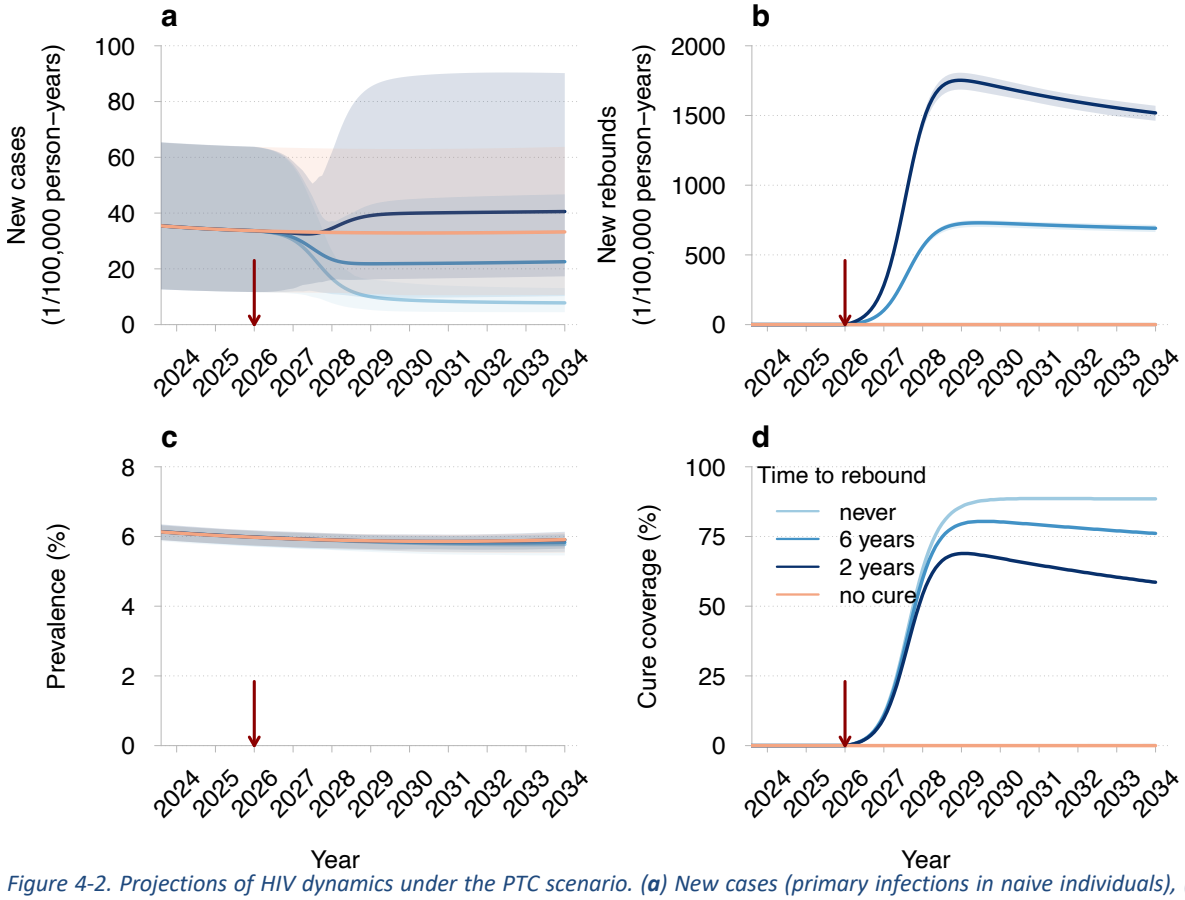

*Figure 4-2. Projections of HIV dynamics under the PTC scenario. (**a**) New cases (primary infections in naive individuals), (**b**) new rebounds (new cases of viral rebound in cured individuals), (**c**) prevalence (proportion of individuals with HIV), and (**d**) cure coverage (proportion of cured individuals among all eligible) for different times until viral rebound. The red vertical arrows indicate the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. Different shades of blue correspond to different times until viral rebound. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90%, annual cure uptake of 90%, and diagnostic delay of cured individuals who experience a viral rebound of 3 months.*

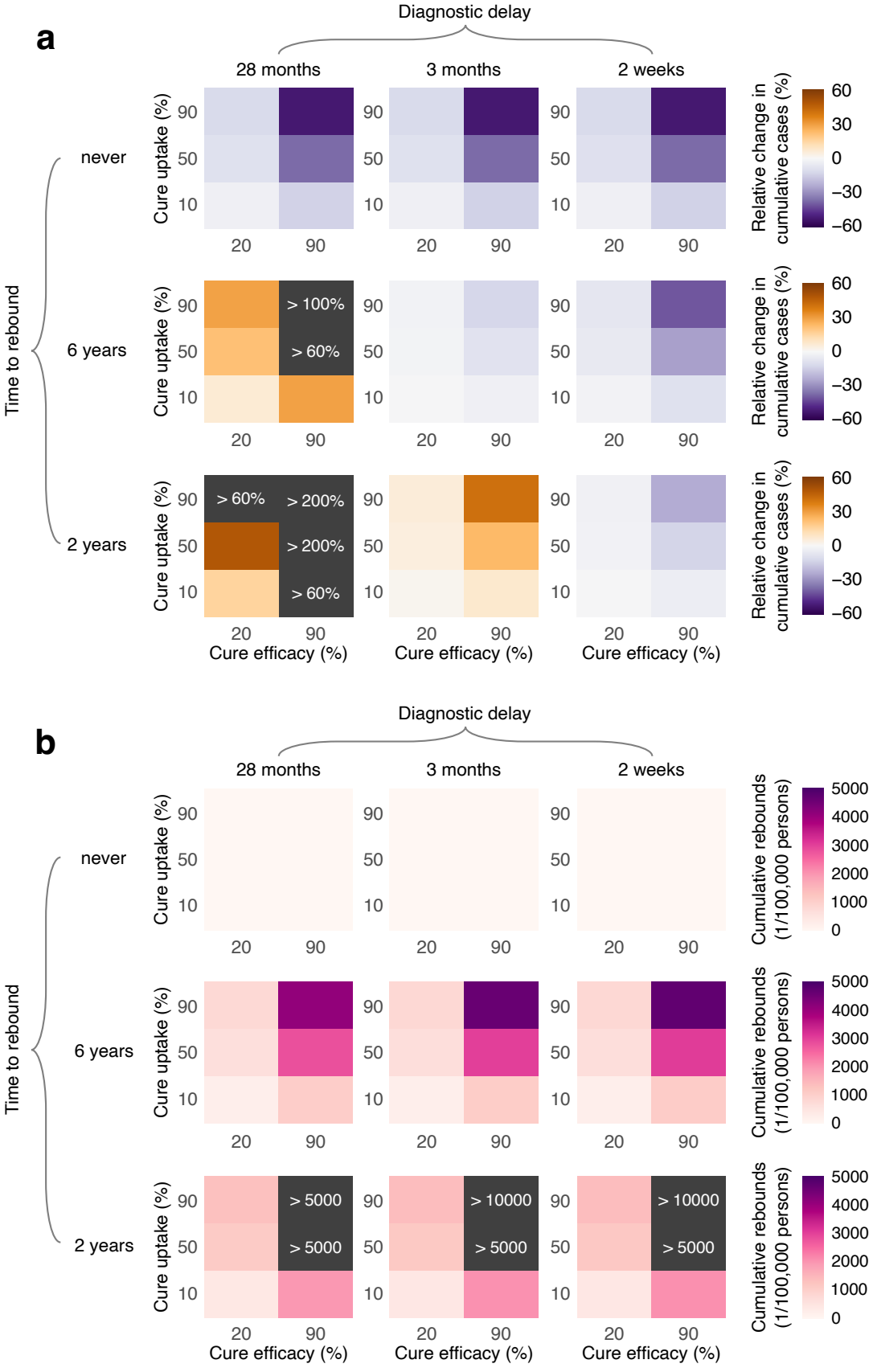

Figure 4-3. Impact of PTC on HIV dynamics under varying PTC characteristics. (a) Mean change in cumulative cases relative to the no-cure scenario and (b) mean cumulative rebounds from the introduction of PTC in 2026 to the end of the simulation in 2034.

Dynamics for PTC scenario

We first explored how the dynamics under the PTC scenario might unfold based on different assumptions about the time to viral rebound in cured individuals. We simulated the model with average rebound times of 2 years and 6 years, as well as the possibility of perfect PTC (i.e., no rebound). The extremes (2 years to rebound and no rebound) represent the minimum and optimal requirements outlined in the TPP [148]. We assumed 90% cure efficacy and 90% annual cure uptake, along with a 3-month diagnostic delay for rebound cases. Projections of the epidemic dynamics for these parameters are shown in Figure 4-2. Perfect PTC, with no risk of rebound, resulted in the fewest new cases among naive individuals compared to the no-cure scenario, 8 (95%PI 4–13) versus 33 (95%PI 11–64) per 100,000 persons per year in 2034. If the time to viral rebound was 6 years, the estimated number of new cases decreased to 23 (95%PI 10–47) per 100,000 persons per year, but the number of new rebounds was high and greatly exceeded the number of new cases in naive individuals. For a shorter rebound time of 2 years, new cases increased to 41 (95%PI 17–90) per 100,000 persons per year, and rebounds increased even further. The prevalence remained nearly constant regardless of the rebound time, and cure coverage was lower for shorter rebound times.

Impact of PTC under varying PTC characteristics

We further systematically compared the impact of PTC under varying cure efficacy, uptake, time to rebound, and diagnostic delay from its introduction in 2026 to 2034 (Figure 4-3). Perfect PTC consistently resulted in fewer cases in naive individuals, regardless of diagnostic delay (Figure 4-3, top rows). The largest and smallest reductions in cumulative cases over this period, compared to the no-cure scenario, were observed at the extreme values of cure uptake and efficacy, 52% average reduction for 90% uptake and 90% efficacy, and 3% average reduction for 10% uptake and 20% efficacy. Intermediate reductions in cases were predicted for all other parameter combinations. Whether imperfect PTC increased or decreased cases in naive individuals from 2026 to 2034, compared to the no-cure scenario, depended on diagnostic delay and rebound time (Figure 4-3, middle and bottom rows). In the absence of a specific monitoring strategy to diagnose rebounds (i.e., with the current standard of care, a 28-month diagnostic delay), the introduction of PTC consistently led to more cases. Larger increases in cases were predicted for shorter rebound time, and higher PTC uptake and efficacy. Conversely, with frequent monitoring of cured individuals (i.e., a 2-week diagnostic delay), cumulative cases were consistently reduced compared to the no-cure scenario. The largest average reductions in cumulative cases were between 41% for a rebound time of 6 years and 23% for a rebound time of 2 years. However, an estimated 4,695 (95%PI 4,485–4,847) and 11,310 (95%PI 10,836–11,653) rebounds per 100,000 persons were predicted to occur alongside these maximum case reductions. For comparison, only 561 (95%PI 184–1,074) cases per 100,000 persons were estimated during the same period in the no-cure scenario.

Dynamics for elimination scenario

Like in the PTC scenario, the dynamics under the elimination scenario assumed a 90% cure efficacy. The introduction of the elimination cure led to a consistent reduction in the number of new cases in naive individuals, with values varying based on cure uptake (Figure 4-4). For 90% uptake, the estimated number of new cases decreased to 8 (95%PI 4–13) per 100,000 persons per year in 2034, which is comparable to the outcome of perfect PTC. Similarly, for lower uptakes, new cases also declined, with 9 (95%PI 5–14)

new cases for a 50% uptake and 22 (95%PI 9–39) new cases for a 10% uptake. Notably, across the entire range of cure uptakes considered, the estimated number of re-infections in cured individuals remained low, approaching only about 2 re-infections per 100,000 persons per year in 2034. Furthermore, unlike in the PTC scenario, HIV prevalence markedly dropped after the introduction of the elimination cure, reaching 3.60% (95%PI 3.38%–3.74%) for a 10% uptake and falling to less than 1% for uptakes above 50%. The model also indicated that for lower uptakes, no equilibrium in HIV dynamics was yet observed.

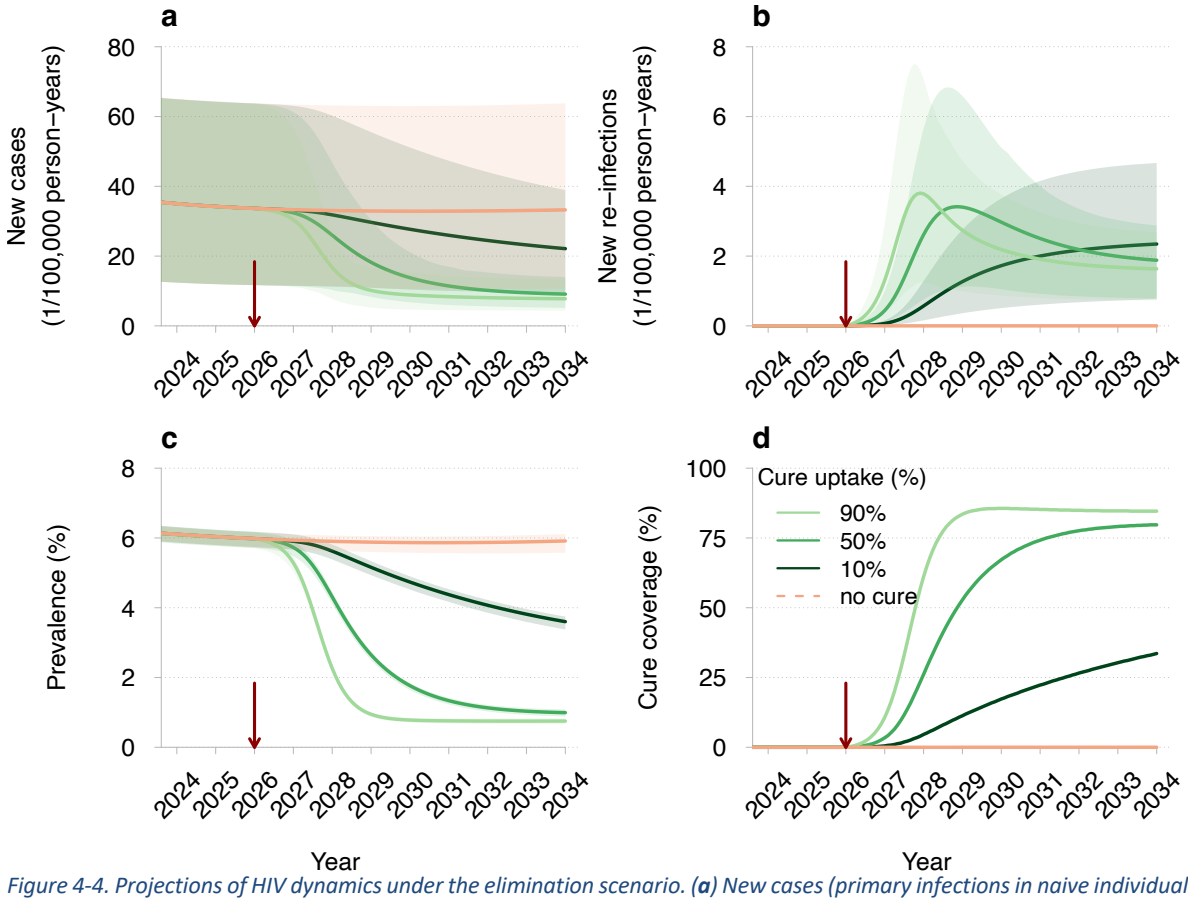

*Figure 4-4. Projections of HIV dynamics under the elimination scenario. (**a**) New cases (primary infections in naive individuals), (**b**) new re-infections (secondary infections in cured individuals), (**c**) prevalence (proportion of individuals with HIV), and (**d**) cure coverage (proportion of cured individuals among all eligible) for different cure uptakes. The red vertical arrows indicate the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. Different shades of green correspond to different cure uptakes. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90% and diagnostic delay of cured individuals who experience re-infection of 3 months.*

Impact of elimination under varying cure characteristics

A systematic comparison of the impact of the elimination cure under varying efficacy, uptake, and diagnostic delay of re-infected individuals is shown in Figure 4-5. Similar to perfect PTC, the elimination cure consistently reduced cases in naive individuals, regardless of diagnostic delay. Like before, the largest and smallest reductions in cumulative cases over the 2026-2034 period were observed at the extreme values of uptakes and efficacy, with the reduction ranging from 3% to 52% for all parameter combinations explored. For all diagnostic delays, the estimated number of re-infections remained low, not exceeding 16 re-infections per 100,000 persons throughout the entire period.

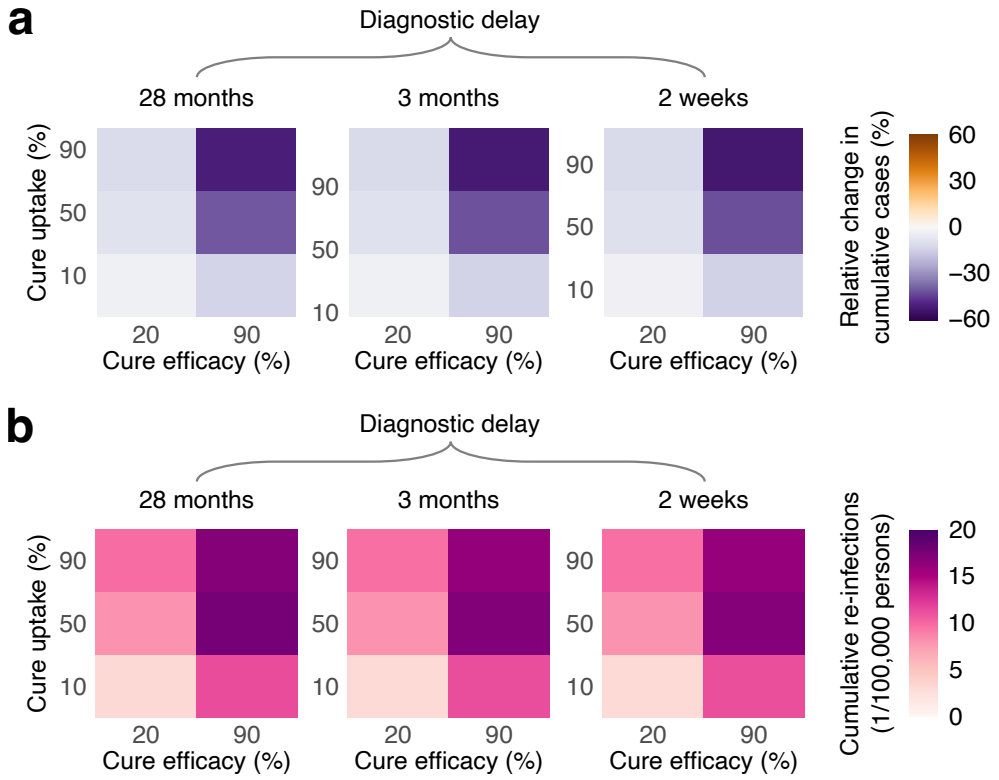

*Figure 4-5. Impact of the elimination cure on HIV dynamics under varying cure characteristics. (**a**) Mean change in cumulative cases (primary infections in naive individuals) relative to the no-cure scenario and (**b**) mean cumulative re-infections (secondary infections in cured individuals) from the introduction of the cure in 2026 to the end of the simulation in 2034. The color bar scale is the same as that in Figure 4-3 for direct comparison.*

Robustness and sensitivity analyses

Our results remained robust across different years of cure introduction. Potential behavioral changes among MSM in response to a cure based on survey data did not qualitatively affect the overall outcomes. Specifically, eight years post-cure, the number of new HIV cases was still lower for the elimination cure compared to the no-cure scenario, though this was not necessarily the case for imperfect PTC. Sensitivity analyses regarding the infectivity of individuals after viral rebound indicated that our projections for the PTC scenario were the most optimistic. In contrast, the results for the elimination cure were consistent across varying assumptions about the infectivity of re-infections. Further details on the sensitivity analyses can be found in the Appendix D.

Methods

The mathematical transmission model was calibrated using behavioral, epidemiological, and surveillance data specific to the MSM population in the Netherlands. Parameter estimates were obtained by fitting the model to observed trends in new HIV diagnoses and the estimated number of undiagnosed cases from 2017 to 2022. This calibrated model was then used to explore the potential impact of hypothetical HIV cure scenarios on the epidemic dynamics at population level.

Data

Sexual contact rates data were taken from the survey described in [145]. Within the survey, the participants were asked questions about their current sexual behavior (number of new partners in the last six months) and preventive behavior (condom use habits), as well as potential changes after the introduction of the two different cure

scenarios introduced above. Out of the participants surveyed, a total of N = 529 MSM provided responses regarding changes in sexual and preventive behaviors concerning casual sexual contacts for both current circumstances and hypothetical cure scenarios.

Data on the monthly number of PrEP users over time, available from June 2019 to April 2022, were taken from [161]. All epidemiological data, such as new diagnoses per year, new imported cases already on treatment, cases on treatment, proportion of diagnoses, and secondary estimates such as the number of undiagnosed cases, HIV incidence and HIV prevalence, were taken from the Dutch HIV Monitoring Foundation reports [136]. We considered yearly data from 2015 to 2022.

Transmission model

The model was implemented in Rstudio (version 2022.12.0+353) using a system of ordinary differential equations for the number of individuals in different compartments describing the state of individuals. The population was stratified by the average number of new sexual partners per year into sexual risk groups l, where $l = 1, \dots, n$ with $l = 1$ and $l = n$ corresponding to the lowest and the highest risk groups, respectively.

Susceptible individuals can be infected at time-dependent rate $J(t)$ (i.e., force of infection) by having sexual interactions with the infectious population. Upon infection, they enter the acute HIV stage. As the disease progresses, these individuals transition through various stages, at different constant rates ρ_k, from the chronic stage to the AIDS stage. In particular, $k = 1$ represents the acute stage, $k = 2$ the chronic stage, $k = 3$ the AIDS stage. We have also included a non-sexually active AIDS stage ($k = 4$) to account for individuals who are no longer sexually active due to the severity of their condition. From this last stage, individuals may die at rate ρ_4 because of AIDS. Throughout this disease progression, individuals may be diagnosed at different constant rates, τ_k, depending on their current stage k. Once diagnosed, they may either die because of HIV-related complications, at rate ρ_D, or begin ART, at rate η. Once on ART, individuals can still die from HIV-related causes at rate γ. We assume that the infectivity of diagnosed and treated individuals is the same. Susceptible individuals may opt in and out of PrEP programs, at rates $k^{\text{on}}(t)$ and k^{off}. PrEP reduces the risk of infection, offering protection with an effectiveness of $1 - \Omega$. Individuals who get infected while on PrEP have reduced transmissibility and are diagnosed more rapidly (at rate τ_P) due to regular visits associated with PrEP usage. Individuals can enter or leave the sexually active population over time. Susceptible individuals are recruited into the population at rate βN_0, and each person enters a specific risk group l with probability q_l. The natural outflow from all compartments occurs at rate μ. In addition, importation of infected individuals (both undiagnosed or already on treatment) is also modeled. Specifically, the importation of individuals already on treatment occurs at a time-dependent linearly increasing rate $M_A t$, while undiagnosed infected individuals are imported at a constant rate M_I. Imported cases are assigned to risk group l with probability Q_l, and to HIV stage k with probaility p_k. When a cure is introduced, individuals on ART become eligible for this treatment. Depending on the cure's efficacy e, they may either be successfully cured at rate $e\alpha(t)$, moving to the cured compartment, or experience a cure failure, at rate $(1 - e)\alpha(t)$, remaining on ART until death at rate γ.

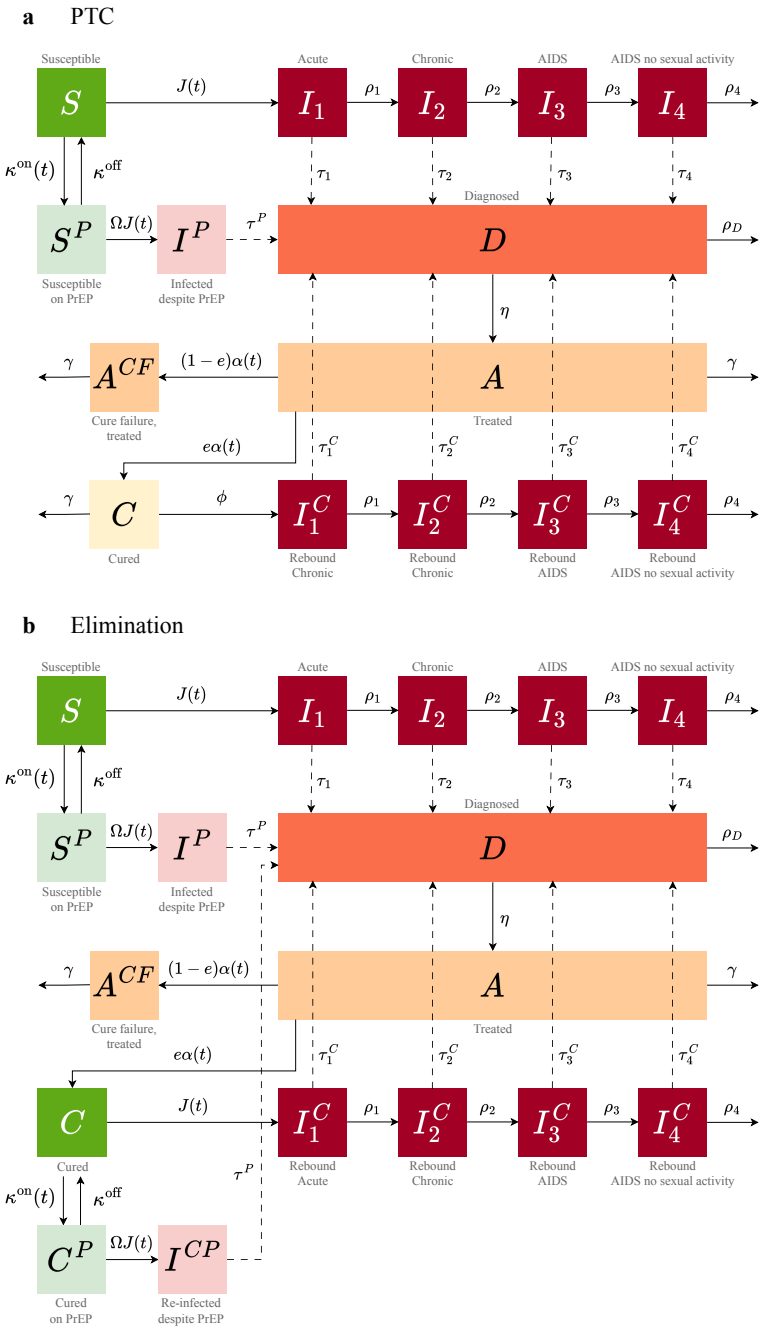

Figure 4-6. Schematic of the transmission model with cure. Recruitment into and exit from the sexually active population are not shown. A detailed description of the model equations, parameters, and assumptions for (a) post-treatment control and (b) elimination cure is given in the Supplementary Material.

In the post-treatment control (PTC) scenario, cured individuals may experience a viral rebound at rate φ. Upon rebound, they re-enter the chronic stage, bypassing the acute stage, and continue through the disease progression at rates ρ_k. They can be re-diagnosed at rates τ_k^C, depending on the specific stage k, or die from AIDS at rate ρ_4. As with individuals on ART, cured individuals may also die at rate γ. In the elimination scenario, cured individuals are assumed to revert to a susceptible state, where they can either be re-infected or start a PrEP program. Upon re-infection, they enter the acute stage, and they follow the same disease progression and diagnosis patterns as in the PTC scenario.

The system of differential equations of the models is reported in the Appendix D, together with the table with the description of all the parameters of the model (Table 4-3). The

schematics of PTC and elimination models are illustrated in Figure 4-6a and Figure 4-6b, respectively.

Model calibration

Sexual partner change rate

We used survey data to estimate the number of new sexual partners per year by sexual risk group, weighted on condom use, in the MSM population. A full explanation of the estimation is reported in Appendix D.

Diagnostic delay

We set diagnostic rates for the AIDS stages 3 and 4, τ_3 and τ_4 respectively, such that the average waiting time before being diagnosed with AIDS is 1 month. On the other hand, diagnostic rates for the acute and chronic stages, τ_1 and τ_2 respectively, remain to be reconstructed. We calibrate those, prior to the fitting of the other model parameters, by means of a simple stochastic model simulating the dynamics of disease progression and diagnoses. We describe the procedure in Appendix D.

Estimation of parameters

The parameters of our model to be calibrated were: the probability of transmission per sexual partner λ; the mixing parameter ω (assortativity of contacts by risk group); the infectivity of diagnosed/treated individuals ε; the number of newly imported infected undiagnosed individuals per year M_I; and an additional number of undiagnosed individuals at initialization to allow the model to burn-in U_0; the infected population fraction used to split initialized and imported infected cases into risk group l, with $l = 1, \ldots, n$, Q_l. To calibrate those, we used an Approximate Bayesian Computation (ABC) approach based on Latin Hypercube Sampling (LHS). The detailed procedure to obtain approximations of the posterior distributions of all the estimated parameters is presented in Appendix D.

HIV cure rollout

We modeled the cure rollout by using a time-dependent cure uptake rate, assumed to be equal for all risk groups, $\alpha(t)$, that grows logistically since the start of the rollout until a maximum uptake value αmax within 3 years.

Model outcomes

The model outcomes were annual new cases (primary infections in naive individuals) and the change in cumulative cases compared to the no-cure scenario from 2026 to 2034. For PTC, annual and cumulative viral rebounds, while for the elimination cure, annual and cumulative re-infections (secondary infections in cured individuals) were also estimated over the same period. HIV prevalence was calculated as the ratio of individuals in specific compartments and the total population size. For PTC, the relevant compartments were $I_k, I^P, D, A, A^{CF}, C$, and I_k^C $(k = 1,2,3,4)$, with cured individuals included because their HIV reservoir is not completely removed after PTC. For the elimination cure, the compartments considered were $I_k, I^P, D, A, A^{CF}, I^{CP}$, and I_k^C $(k = 1,2,3,4)$. Cure coverage was calculated as the proportion of cured individuals among all eligible (A, A^{CF}, and C compartments).

Sensitivity analyses

In the sensitivity analyses, we explored the impact on model projections of (i) the year of cure introduction, (ii) risk compensation following cure introduction, and (iii) infectivity of individuals after rebound and re-infection. The results of these analyses are provided in Appendix D.

Discussion

In light of rapidly advancing HIV cure research [151], it is essential to anticipate the potential population-level impact of cure implementation before effective interventions become available. From a public health perspective, an effective cure intervention should align with the United Nations Sustainable Development Goal of ending the HIV epidemic [53]. Indeed, there is a common assumption that the successful introduction of a cure will contribute to controlling the HIV epidemic [146,147,155]. In the context of concentrated epidemics in high-income countries, where HIV incidence is already low, this assumption implies that the implementation of a cure should not disrupt the decline in HIV infections achieved under the current standard of HIV care and prevention.

Our analyses revealed that introducing either type of cure could alter the course of the HIV epidemic. The impact differed between perfect PTC and elimination cure, which both have the potential to accelerate the current decline in HIV incidence, and imperfect PTC, which could reverse that progress. This divergence in the potential outcomes of cure scenarios underscores the importance of evaluating their public health impacts, which, for now, can only be done through mathematical modeling.

The model projections indicate that while perfect PTC consistently resulted in a decrease in HIV incidence (by up to 50% for optimal characteristics), imperfect PTC could drastically increase new infections and undermine HIV control. Notably, the increase in incidence was observed for plausible parameter combinations. For example, a 3-month diagnostic delay, which matches the testing interval for PrEP users [159], and the TPP's minimum requirement of 2 years before relapse [148] might be achievable in a real-world setting. HIV incidence under imperfect PTC increased in the model because individuals who experienced viral rebound became infectious and could transmit the virus before being diagnosed. This led to a counterintuitive effect where higher incidence was observed for higher cure uptakes and efficacies. The more individuals were cured, the more rebound cases occurred, leading to an increase in new infections. This effect could be mitigated by developing PTC interventions with longer times to rebound or through stricter monitoring. However, even if imperfect PTC reduced HIV incidence, managing numerous rebound cases could pose a challenge in the real world and lead to adverse health outcomes for patients. MSM with HIV have also raised concerns about imperfect PTC because of fear of transmitting the virus to their partners [145]. Frequent rebound episodes would require ongoing efforts for rapid viral load monitoring and timely diagnosis of rebounds to maximize the public health benefits of any PTC intervention.

In contrast, the elimination scenario presents a more optimistic outlook for HIV dynamics. Our findings suggest that elimination cure would always reduce HIV incidence among naive individuals, with projected new infections decreasing to a comparable level seen for perfect PTC (50% for optimal characteristics). Importantly, the elimination scenario would maintain a low number of re-infections, which is relevant for MSM particularly vulnerable to HIV acquisition. The main reason for few re-infections under

the elimination cure and a potentially large number of rebounds under imperfect PTC is that the rate of re-infections depends on the infectious population, while rebounds occur at a constant rate among those previously cured.

It is recognized that mathematical modeling can help advance HIV cure research [162]. However, few studies have modeled the population-level impact of cure strategies. To our knowledge, this is the first study to assess the potential impact of an HIV cure on an epidemic concentrated among MSM in a Western country. We ensured the reliability of our model to accurately project HIV dynamics without a cure by performing inference on key model parameters and conducting validation tests on multiple datasets. Unlike other studies assessing interventions for HIV control [161,163,164], our model incorporated importations of new HIV infections from abroad, which is crucial for capturing the openness of the MSM population in low-incidence settings. The strength of our modeling approach under cure scenarios lies in its ability to project HIV dynamics without requiring precise information on the biological mechanisms underlying a cure. Instead, our analyses were guided by the TPP, using the full range of acceptable values for several cure characteristics [148].

While our findings are particularly relevant for similar epidemics among MSM in Western countries, they align qualitatively with two previous modeling studies for a generalized epidemic in heterosexual populations in Africa [56,57]. These studies, though not formally fitted to data, also suggested that a PTC-like intervention without sustained viral suppression could lead to an increase in HIV incidence. However, our results differ from [57], as we found that the timing of cure introduction did not alter the outcomes of cure scenarios. This discrepancy may be explained by the fact that the HIV epidemic in the Netherlands was estimated to be close to reaching a low stable level of incidence, unlike the more dynamic epidemic in South Africa.

Additionally, there is concern that the introduction of a cure could shift perceptions of HIV risk, severity, and prevention, as observed when ART became widely available [165]. Our study is the first to incorporate actual survey data on potential behavioral changes among MSM following the introduction of a cure into sensitivity analyses. Our findings suggest that risk compensation could further reduce the effectiveness of imperfect PTC, while having minimal impact on the outcomes of elimination cure. Therefore, real-world cure interventions may need to be accompanied by additional prevention strategies to address the potential rise in risky behaviors.

Our study has several limitations. First, we used a classical deterministic compartmental model rather than an equivalent stochastic approach. Given the large population size modeled, stochastic effects, such as the random extinction of the epidemic, are likely to have minimal impact in this context. Additionally, as our goal was not to provide precise forecasts but rather to identify broad epidemiological trends, we believe the average behavior captured by our deterministic model sufficiently reflects the population-level transmission dynamics. However, stochastic models generally offer a more realistic representation of reality by incorporating variability. Developing a stochastic version of our model could enhance its realism and robustness, representing a potential direction for future work. Another assumption inherent to our ODE model is the use of an exponential distribution for the time until viral rebound. While this distribution may not fully capture the biological variability of relapse timing, it provides a straightforward

framework for analysis, especially in the absence of empirical data. This simplification allows for conceptual comparisons across cure scenarios without introducing unnecessary model complexity. Second, in line with the call for equitable HIV cure solutions [166,167,168,169], our analysis did not focus on targeting cure strategies to specific population groups based on behavior or HIV status. Instead, in agreement with the inclusion criteria for many HIV cure trials [170] and the minimum requirement in the TPP [148], our model assumes that a cure is administered to MSM on ART. Since most MSM with HIV in the Netherlands are diagnosed and receive effective treatment [136], this assumption likely has minimal impact on our findings. Third, despite the formulation of the TPP, uncertainty remains around several biological parameters for cured individuals. Based on data from ART interruption studies, we assumed that the infectivity of individuals following viral rebound would be similar to that observed during the chronic stage of HIV infection [171,172]. Conversely, for individuals re-infected after elimination cure, we assumed their infectivity would resemble the acute stage of infection, similar to what occurs with superinfection by a different HIV subtype [173] or re-infection with hepatitis C virus [174,175]. While these assumptions are biologically plausible, they remain hypothetical. Our sensitivity analyses demonstrated that high infectivity could undermine the effectiveness of imperfect PTC and should be carefully considered when developing this cure strategy.

In summary, our study suggests that the elimination cure has the potential to reduce new HIV infections, contribut- ing to the United Nations Sustainable Development Goal of ending the HIV epidemic, while imperfect PTC could increase infections if rebounds are not carefully monitored. These findings emphasize the importance of strategically developing and implementing cure interventions to maximize their benefits for individual and public health.

Appendix D

The code used to run the model will be available upon publication of the paper at: github.com/alfredodebellis .

Calibration of diagnostic delay

We implemented a simple stochastic model where we simulate the state progression of Z = 1000 individuals initialized at the acute stage. We sampled N_τ = 10000 uniform values of τ_1 and τ_2 in the intervals [1/12, 2] and [1/12, 1], respectively. We set diagnostic rates for the AIDS stages 3 and 4, τ_3 and τ_4 respectively, such that the average waiting time before being diagnosed with AIDS is 1 month. Then, for each sampled pair, we ran the model described in Algorithm 3-1. We define a relative error threshold $\epsilon = 10\%$ and accept all the pairs such that the proportions of diagnoses happened before 6 months, between 6 and 12 months, and after 12 months after infection match (with a tolerance given by ϵ) available data. The output of the reconstructed proportions of diagnosed per period is reported in Figure 4-7. In Figure 4-8 we show the distributions of τ_1, τ_2.

Algorithm 4-1. Pseudocode for the stochastic model to build the distribution of diagnostic delays

$\Delta \leftarrow$ Empty distribution of diagnostic delays
for $i \leftarrow 1$ to Z **do**
  $\delta \leftarrow 0$
  $k \leftarrow 1$
  while $k \leq 4$ **do**
    $T_k^p \leftarrow$ random exponential progression time with rate ρ_k
    $T_k^d \leftarrow$ random exponential diagnostic delay with rate τ_k
    if $T_k^p < T_k^d$ **then**
      next k
    else
      for $k' \leftarrow 1$ to $(k-1)$ **do**
        $\delta \leftarrow \delta + T_{k'}^p + T_k^d$
      end for
    end if
  end while
  Append δ to Δ
end for

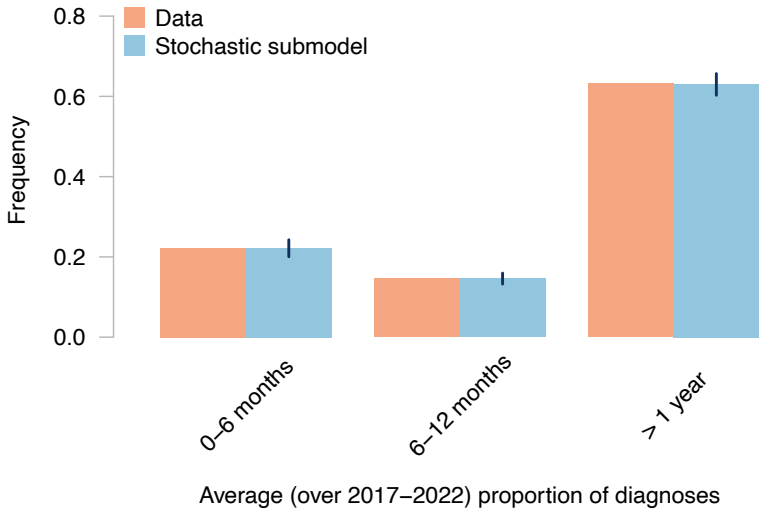

Figure 4-7. Model validation against data on proportion to diagnoses within different time ranges. The orange rectangles represent data from the HIV monitoring foundation. The blue rectangles and the error bars represent the mean estimates and the 95% credible intervals estimated by the stochastic model in Algorithm 3-1 after calibration of the diagnostic delays

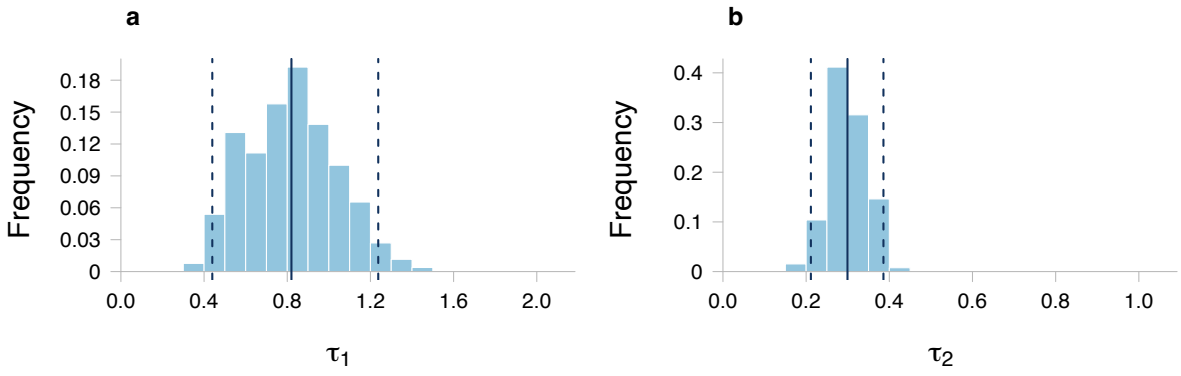

*Figure 4-8. Estimated diagnosis rates. **a** Histograms of diagnosis rates for acute and **b** chronic HIV stages fitted via the approximate Bayesian computation framework. The vertical continuous line indicates the mean value. The vertical dashed lines indicate the bounds of the 95% credible interval.*

Model equations

Post-treatment control

The equations for the numbers of individuals in risk group $l, l = 1,\dots,n$, who are susceptible (S_l), susceptible on PrEP (S_l^P), infected but not diagnosed (I_{lk}) in stage k, k = 1 (acute), 2 (chronic), 3 (AIDS), 4 (AIDS without sexual activity), infected while on PrEP (I_l^P), diagnosed (D_l), treated on ART (A_l), cured and virally suppressed without ART (C_l), not effectively suppressed after cure and thus still on ART (A_l^{CF}), infectious due to viral rebound l after a cure failure (I_{lk}^C) read as follows

$$
\begin{aligned}
\frac{dS_l(t)}{dt} &= \beta N_0 q_l + k_l^{off} S_l^P(t) - \big(\mu + J_l(t) + k_l^{on}(t)\big)S_l(t)\\
\frac{dS_l^P(t)}{dt} &= k_l^{on}(t)S_l(t) - \big(\mu + \Omega J_l(t) + k_l^{off}\big)S_l^P(t)\\
\frac{dI_{l1}(t)}{dt} &= M_I Q_l p_1 + J_l(t)S_l(t) - (\mu + \rho_1 + \tau_1)I_{l1}(t)\\
\frac{dI_{lk}(t)}{dt} &= M_I Q_l p_k + \rho_{k-1} I_{lk-1}(t) - (\mu + \rho_k + \tau_k)I_{lk}(t)\\
\frac{dI_l^P(t)}{dt} &= \Omega J_l(t)S_l^P(t) - (\mu + \tau_l^P)I_l^P(t)\\
\frac{dD_l(t)}{dt} &= \sum_k \tau_k I_{lk}(t) + \tau_l^P I_l^P(t) + \sum_k \tau_k^C I_{lk}^C(t)\\
&\qquad - (\mu + \rho_D + \eta)D_l(t)\\
\frac{dA_l(t)}{dt} &= M_A Q_l t + \eta D_l(t) - \big(\mu + \gamma + \alpha(t)\big)A_l(t)\\
\frac{dC_l(t)}{dt} &= e\alpha(t)A_l(t) - (\mu + \gamma + \phi)C_l(t)\\
\frac{dA_l^{CF}(t)}{dt} &= (1-e)\alpha(t)A_l(t) - (\mu + \gamma)A_l^{CF}(t)\\
\frac{dI_{l1}^C(t)}{dt} &= \phi C_l(t) - (\mu + \rho_1 + \tau_1^C)I_{l1}^C(t)\\
\frac{dI_{lk}^C(t)}{dt} &= \rho_{k-1} I_{lk-1}^C(t) - (\mu + \rho_k + \tau_k^C)I_{lk}^C(t)
\end{aligned}
\tag{4.1}
$$

where $k = 2,3,4$ and $l = 1,\dots,n$.

The force of infection in risk group l is written as

$$
J_l(t) = \lambda c_l \sum_{l'} M_{ll'}(t)\left\{\varepsilon^P \frac{I_{l'}^P(t)}{N_{l'}(t)} + \varepsilon \frac{D_{l'}(t) + A_{l'}(t) + A_{l'}^{CF}(t)}{N_{l'}(t)} + \sum_k \frac{h_k I_{l'k}(t) + h_k^C I_{l'k}^C(t)}{N_{l'}(t)}\right\}
\tag{4.2}
$$

$N_l(t)$ is the population size of risk group l

$$N_l(t) = S_l(t) + S_l^P(t) + \sum_k [I_{lk}(t) + I_{lk}^C(t)] + I_l^P(t) + D_l(t) + A_l(t) + A_l^{CF}(t) + C_l(t) \tag{4.3}$$

and $M_{ll'}(t)$ is the mixing matrix

$$M_{ll'}(t) = \omega \frac{c_{l'} N_{l'}(t)}{\sum_{l''} c_{l''} N_{l''}(t)} + (1-\omega)\delta_{ll'} \tag{4.4}$$

where $\delta_{ll'} = 1$ if $l = l'$ and $\delta_{ll'} = 0$ otherwise.

Elimination

The equations for the numbers of individuals in risk group $l, l = 1, \ldots, n$, who are naive susceptible (S_l), naïve susceptible on PrEP (S_l^P), cured and susceptible after complete elimination of HIV (C_l), susceptible after cure and on PrEP (C_l^P), primarily infected but not diagnosed (I_{lk}), re-infected after a cure (I_{lk}^C) in stage k, k = 1 (acute), 2 (chronic), 3 (AIDS), 4 (AIDS without sexual activity), infected while on PrEP (I_l^P), re-infected after a cure while on PrEP (I_l^{CP}), diagnosed (D_l), treated on ART (A_l), cured and virally suppressed without ART (C_l), not effectively suppressed after cure and thus still on ART (A_l^{CF}), are given by

$$\begin{aligned}
\frac{dS_l(t)}{dt} &= \beta N_0 q_l + k_l^{off} S_l^P(t) - \big(\mu + J_l(t) + k_l^{on}(t)\big) S_l(t) \\
\frac{dS_l^P(t)}{dt} &= k_l^{on}(t) S_l(t) - \big(\mu + \Omega J_l(t) + k_l^{off}\big) S_l^P(t) \\
\frac{dC_l(t)}{dt} &= e\alpha(t) A_l(t) + k_l^{off} C_l^P(t) \\
&\quad - \big(\mu + J_l(t) + k_l^{on}(t)\big) C_l(t) \\
\frac{dC_l^P(t)}{dt} &= k_l^{on}(t) C_l(t) - \big(\mu + \Omega J_l(t) + k_l^{off}\big) C_l^P(t) \\
\frac{dI_{l1}(t)}{dt} &= M_I Q_l p_1 + J_l(t) S_l(t) - (\mu + \rho_1 + \tau_1) I_{l1}(t) \\
\frac{dI_{lk}(t)}{dt} &= M_I Q_l p_k + \rho_{k-1} I_{lk-1}(t) - (\mu + \rho_k + \tau_k) I_{lk}(t) \\
\frac{dI_{l1}^C(t)}{dt} &= J_l(t) C_l(t) - (\mu + \rho_1 + \tau_1^C) I_{l1}^C(t) \\
\frac{dI_{lk}^C(t)}{dt} &= \rho_{k-1} I_{lk-1}^C(t) - (\mu + \rho_k + \tau_k^C) I_{lk}^C(t) \\
\frac{dI_l^P(t)}{dt} &= \Omega J_l(t) S_l^P(t) - (\mu + \tau_l^P) I_l^{CP}(t) \\
\frac{dI_l^{CP}(t)}{dt} &= \Omega J_l(t) C_l^P(t) - (\mu + \tau_l^P) I_l^P(t) \\
\frac{dD_l(t)}{dt} &= \sum_k \tau_k I_{lk}(t) + \tau_l^P \left(I_l^P(t) + I_l^{CP}(t)\right) + \sum_k \tau_k^C I_{lk}^C(t) \\
&\quad - (\mu + \rho_D + \eta) D_l(t)
\end{aligned} \tag{4.5}$$

$$\frac{dA_l(t)}{dt} = M_A Q_l t + \eta D_l(t) - \big(\mu + \gamma + \alpha(t)\big)A_l(t)$$
$$\frac{dA_l^{CF}(t)}{dt} = (1-e)\alpha(t)A_l(t) - (\mu+\gamma)A_l^{CF}(t)$$

where $k = 2,3,4$ and $l = 1,\ldots,n$.

The force of infection in risk group l is written as

$$J_l(t) = \lambda c_l \sum_{l'} M_{ll'}(t)\left\{\varepsilon^P \frac{I_{l'}^P(t) + I_{l'}^{CP}(t)}{N_{l'}(t)} + \varepsilon \frac{D_{l'}(t) + A_{l'}(t) + A_{l'}^{CF}(t)}{N_{l'}(t)} + \sum_k \frac{h_k I_{l'k}(t) + h_k^C I_{l'k}^C(t)}{N_{l'}(t)}\right\} \quad (4.6)$$

$N_l(t)$ is the population size of risk group l:

$$N_l(t) = S_l(t) + S_l^P(t) + C_l(t) + C_l^P(t) + \sum_k [I_{lk}(t) + I_{lk}^C(t)] + I_l^P(t) + I_l^{CP}(t) + D_l(t) + A_l(t) + A_l^{CF}(t) \quad (4.7)$$

and $M_{ll'}(t)$ is the mixing matrix as above.

Sexual partner change rates

We used survey data to estimate the number of new sexual partner per year, weighted on condom use, in the MSM population. For each risk group, we define as sexual contact rate change c_l the product of the number of new partners times a binary condom use variable equal to zero if the individual reported to always using condoms and equal to one otherwise. After building the empirical distribution of casual sexual contact rate change, we added, for each individual, zero, one or two steady partners via multinomial probabilities $p_0^S = 0.425$, $p_1^S = 0.483$ and $p_2^S = 0.092$, estimated from data in the current scenario concerning only individuals that were recruited through the Aids Care Service (ACS) program [176]. We modeled the empirical histogram of observed values of c_l under each scenario using a Weibull distribution. The CDFs are illustrated in Figure 4-9, with vertical dashed lines indicating the intervals defining all the risk groups' rates. We assume that the proportions of the population, q_l, belonging to each risk groups will not change after the introduction of a cure. The results obtained represents the rates per six months. To obtain yearly estimates, we multiply the rates by 1.5 [177].

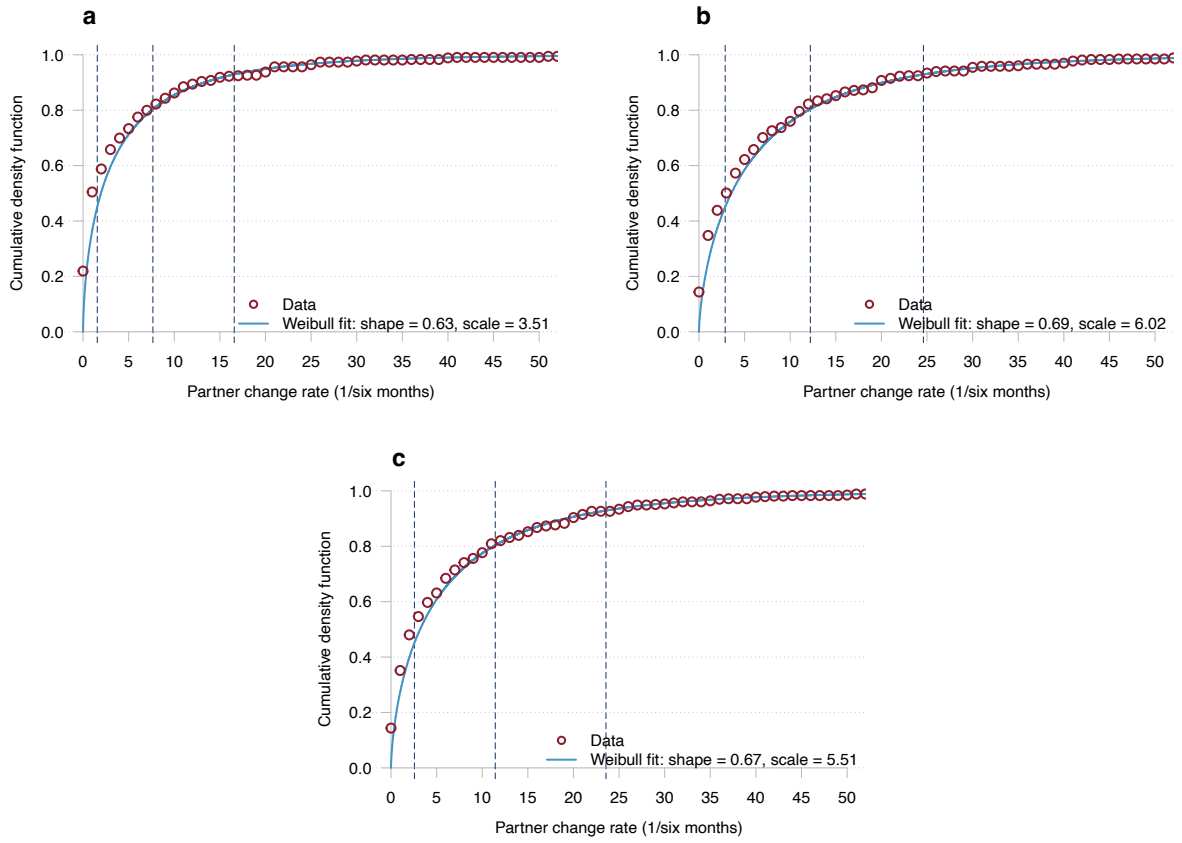

Figure 4-9. Cumulative density function of sexual partner change rates. ***a*** *Current scenario,* ***b*** *PTC cure scenario and* ***c*** *elimination cure scenario. The red circles correspond to the empirical distribution data from the survey. The blue line corresponds to the Weibull distribution that better fit the data. The parameters of the fitted Weibull are reported in the legend of each panel.*

Calibration of free parameters

The free parameters of our model to be calibrated were: the probability of transmission per sexual partner λ; the mixing parameter ω (groups assortativity of contacts); the infectivity of diagnosed/treated individuals ε; the number of newly imported infected undiagnosed individuals per year MI; and an additional number of undiagnosed individuals at initialization to allow the model to burn-in U_0; the infected population fraction Q_l used to split initialized and imported infected cases into risk group l, with $l = 1, \ldots, n$. To calibrate those, we used an Approximate Bayesian Computation (ABC) approach based on Latin Hypercube Sampling (LHS). The procedure consists of two sequential steps. First, we sampled from L = 40000 parameters sets, generated through the LHS based on the prior ranges reported in Table 4-3, for which the condition $Q_1 < Q_2 < Q_3 < Q_4$ holds. After normalizing all the Q_l such that $\sum_l Q_l = 1$, we sampled parameters from the generated list and simulated the epidemic dynamics for each set. The performance of each simulation was evaluated by comparing the simulated new diagnosed per year and the number of undiagnosed individuals with the data time series. We accepted a parameter set if it met the following criteria: i) the simulated yearly diagnoses fell within 10% of the observed data points, and ii) the annual number of undiagnosed cases remained within the uncertainty intervals of the observed data throughout the calibration period (2017-2022). Additionally, we constrained the estimated number of imported undiagnosed cases per year to be less than upper bound of the HIV incidence in 2022, estimated by the Dutch Monitoring Foundation, which is 150 new cases. We accepted 300 parameter sets, and then we sampled 100 accepted sets to produce cure scenarios with variability.

The fitted curves are reported in Figure 4-1. The empirical distribution of accepted parameters and the Pearson correlations between parameters are showed in Figure 4-10 and Figure 4-11, respectively. The model is validated on the number of PrEP users and the number of newly imported individuals who are already on treatment. These two quantities were also used to calibrate a time-dependent functional form of the PrEP uptake and the rate of importation from abroad of treated individuals (see next section for details). We show such validations in Figure 4-12a and Figure 4-12b. We assume that, after 2022, the number of PrEP users remains at a maximum capacity of $N_P^{max} = 10^4$. We validated our model calibration further through the number of treated individuals and the proportion of treated out of the infected individuals over time (Figure 4-12c and Figure 4-12d). An additional validation of our model is reported Figure 4-13, where we show how the average proportions, throughout the time window of the simulation, of infected individuals per risk group remain almost constant.

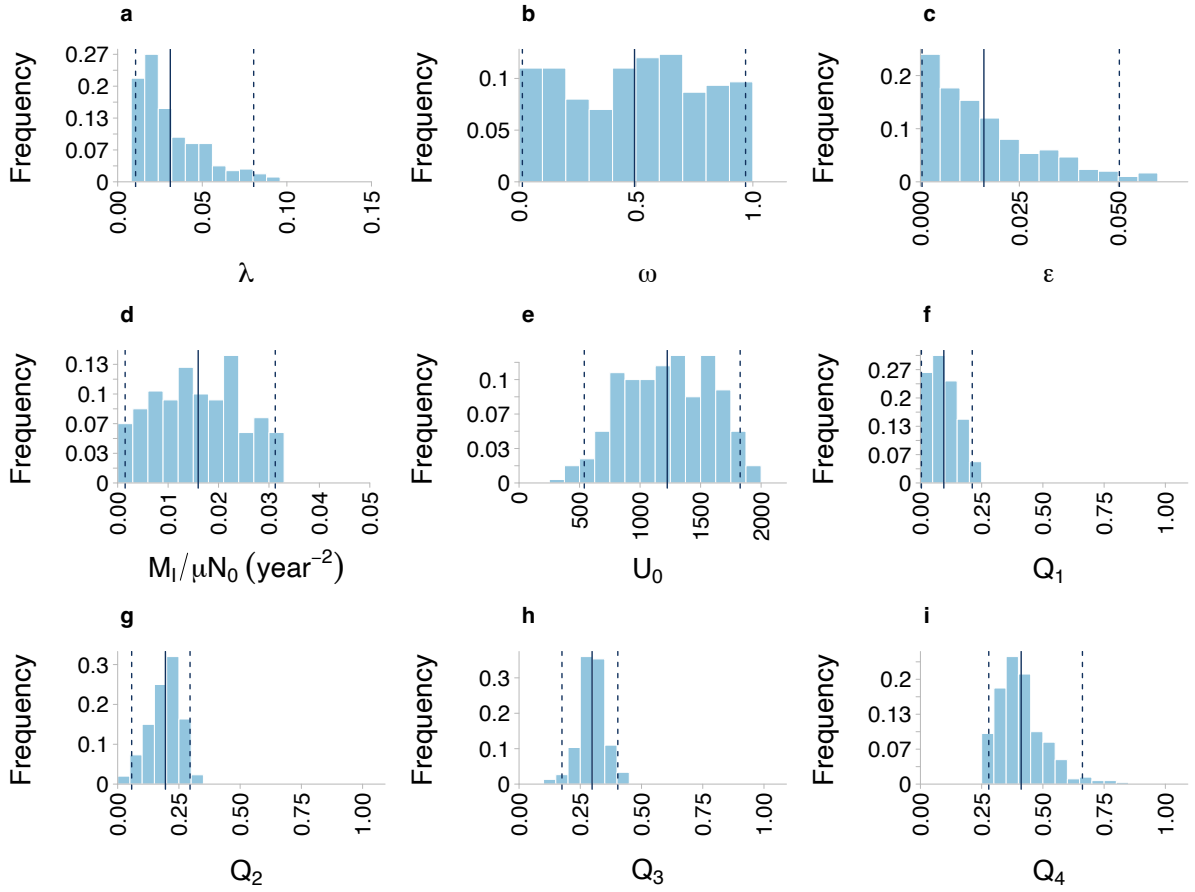

*Figure 4-10. Estimated free parameters. Histograms of all the parameters of the model fitted via the approximate Bayesian computation framework. **a** Transmission probability per partner; **b** mixing (assortativity) parameter; **c** infectivity of diagnosed and treated individuals; **d** number of newly imported undiagnosed cases rescaled by the birth rate and the population size; **e** number of not detected cases at the initialization of the model in 2015; **f** probability to be in the risk group 1 of imported cases; **g** probability to be in the risk group 2 of imported cases; **h** probability to be in the risk group 3 of imported cases; (i) probability to be in the risk group 4 of imported cases. The vertical continuous line indicates the mean value. The vertical dashed lines indicate the bounds of the 95% credible interval.*

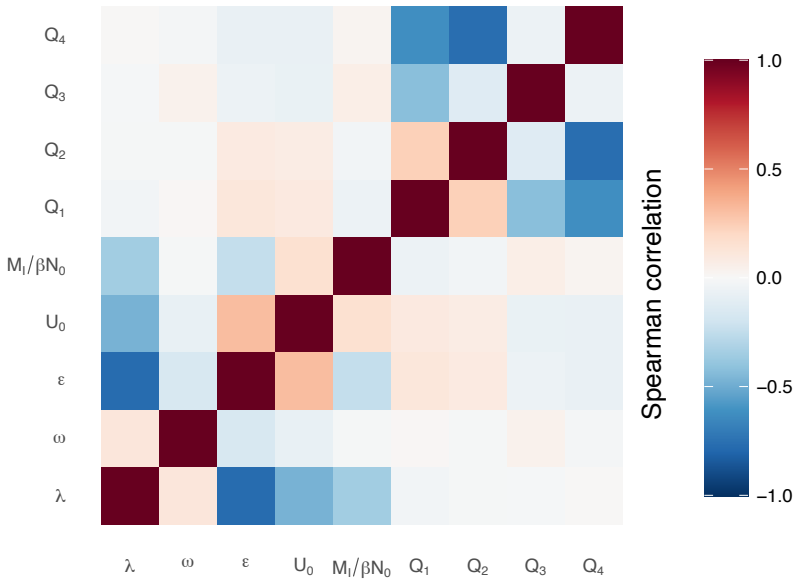

Figure 4-11. Spearman correlation for each pair of parameters fitted via the approximate Bayesian framework.

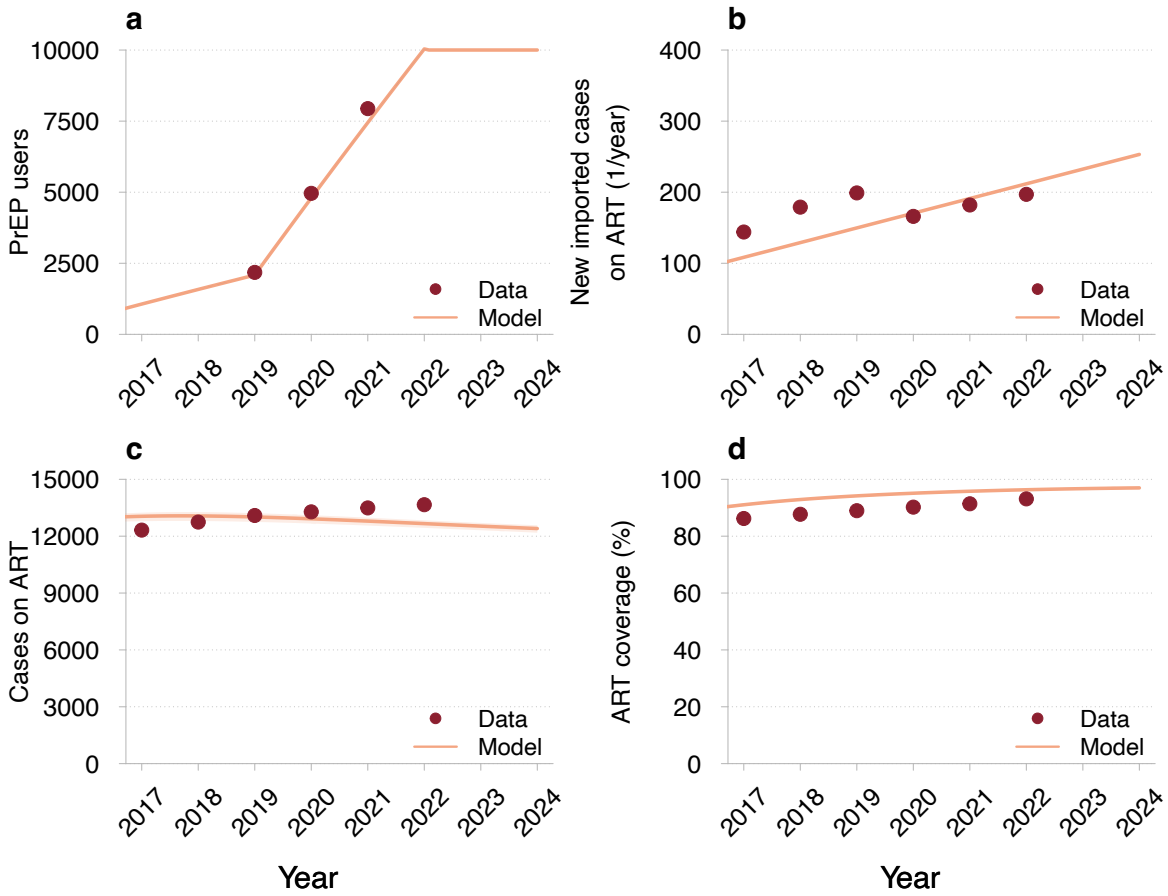

*Figure 4-12. Model validation against HIV surveillance data for MSM in the Netherlands. **a** Number of individuals on PrEP, **b** new imported cases on ART, **c** number of individuals on ART, and **d** ART coverage among all infected individuals. The red dots and the error bars correspond to the mean estimates and the 95% confidence intervals in the data from **a** the national STI surveillance database and **b**–**d** the HIV Monitoring Foundation. The mean trajectories estimated from the model are shown as orange lines. The orange shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution.*

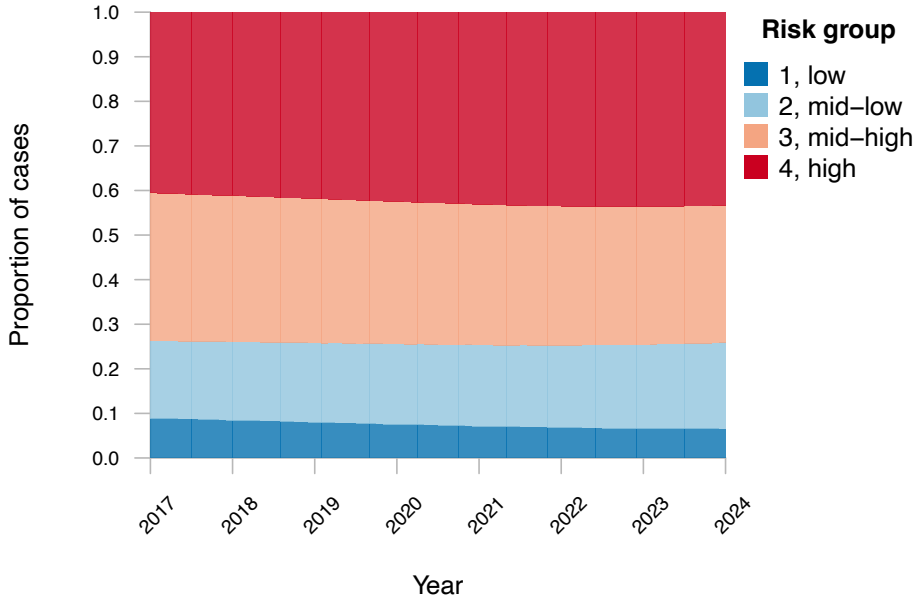

Figure 4-13. Proportions of infected individuals per risk group over time. Proportions, throughout the time window of the fitting simulations, of infected individuals for each of the four risk group.

Linear assumptions

Based on observed data, we assume a linear increase of the annual number of individuals on PrEP and of the number of new imported cases already on treatment per year. We then fitted two linear models to estimate the slopes, M_A and M_P, of the lines to drive the related compartment variables in the compartmental model. While MA enters directly the equation of the model (see Equations (4.1) and (4.5)) for the compartment A, further assumptions should be made for PrEP users. Data on PrEP users start from 2019 and end in 2021. However, based on the fact that pilot trials on PrEP were ongoing even before, we model PrEP enrolling already from 2015. Hence, we split data in two regions, namely 2015-2019 and 2019-2021, and estimated two different slopes, M_P^i and M_P^f , respectively. Since we assume that only risk groups 3 and 4 are eligible for PrEP, we weighted M_P^i and M_P^f with population fractions q_3 and q_4, thus leading to four different values that are $M_{P,3}^i$, $M_{P,3}^f$, $M_{P,4}^i$, and $M_{P,4}^f$. To use the linearity assumption of PrEP users in our model, we take the equation for the derivative of S_l^P and we focus on the case where the outflux from the compartment that is not due to PrEP dropout is negligible, that is:

$$\frac{dS_l^P(t)}{dt} \approx \kappa_l^{on}(t)S_l(t) - \kappa_l^{off}S_l^P(t) \tag{4.8}$$

A linear growth of S^P translate in a constant derivative, $dS^P/dt \approx M_P$. Inserting this in the Equation (4.8), we obtain that the PrEP uptake rate, for each risk group l, should take the following form:

$$\kappa_l^{on}(t) \approx \frac{M_{P,l}^X + \kappa_l^{off}S_l^P(t)}{S_l(t)} \tag{4.9}$$

where $X = i$ before 2019 and $X = f$ after.

Tables

Table 4-1. Description of cure scenarios. The scenarios are reproduced from our previous qualitative study on the perceived impact of an HIV cure by people living with HIV and key populations vulnerable to HIV in the Netherlands [144]. PTC = post-treatment control.

Type of scenario	Description

PTC	Imagine that this treatment has been extensively tested, is safe and now available for everyone living with HIV. The purpose of this treatment is to suppress HIV in the body long-term without ART.
	• After this treatment, the immune system suppresses HIV.
	• If HIV is suppressed without HIV medication, then:
	  ○ HIV is still present in the body.
	  ○ There is a small chance that HIV will become active again, which is why it is necessary to have blood tests done. Every six months up till three years after the treatment.
	  ○ Taking HIV medication is not necessary anymore.
	  ○ It is not possible to contract HIV again or transfer HIV.
	  ○ PrEP or condom use is not necessary to prevent HIV.

Elimination	Imagine that this treatment has been extensively tested, is safe and now available for everyone living with HIV.
	• This treatment removes HIV from the body.
	• Blood tests are used to determine whether the treatment is successful. Every 6 months, up till 3 years after the treatment.
	• If HIV is successfully removed from the body, then:
	  ○ It is not necessary to take HIV medication (ART).
	  ○ There is no chance of HIV becoming active again.
	  ○ It is not possible to transfer HIV.
	  ○ It is not possible to contract HIV again. This treatment does not provide immunity to HIV.
	  ○ It is recommended to use PrEP and/or condoms.
	  ○ It is still advised to test for HIV regularly.

Table 4-2. Scenarios tested and relation to TTP. Different parameter values and modeling decisions represent various scenarios that have been tested and are compared to the minimum and optimum values outlined in the Target Product Profile (TPP) [148], there named as time of remission, clinical efficacy, and protection from reinfection. In our PTC scenario, full protection is assumed, while the elimination scenario assumes no protection.

	Minimum TTP	Optimum TTP	
Varied characteristics			
Efficacy, $\boldsymbol{e}$	20%	90%	
Time to rebound, $\boldsymbol{1/\phi}$	2 years	6 years	never
Protection from reinfection	None	Full	
Fixed characteristics			
Target population	Virologically suppressed on ART	-	

Table 4-3. Summary of the model parameters.

Description (unit)	Notation	Value	Reference
Duration of stage $\boldsymbol{k}$ in undiagnosed individuals (years)	$1/\rho_k$	$1/\rho_1 = 0.14, 1/\rho_2 = 8.44,$ $1/\rho_3 = 1.18, 1/\rho_4 = 1.32$	[178-182]
Probability of being in stage $\boldsymbol{k}$ (proportional to the duration of stage $\boldsymbol{k}$)	p_k	$p_1 = 0.01, p_2 = 0.76,$ $p_3 = 0.11, p_4 = 0.12$	-
Duration of survival of diagnosed individuals (years)	$1/\rho_D$	11	[178,180,183]
Duration of survival of treated individuals (years)	$1/\gamma$	61	[178,180,183]
Infectivity in stage $\boldsymbol{k}$ of undiagnosed individuals	h_k	$h_1 = 0.62, h_2 = 0.12,$ $h_3 = 0.64, h_4 = 0$	[178,181,182]
Infectivity of individuals infected on PrEP	ε_P	$h_1/2$	assumed
Total initial population size	N_0	210,000	[164]
Initial susceptible population fractions of risk group $\boldsymbol{l}$	q_l	$q_1 = 0.45, q_2 = 0.35,$ $q_3 = 0.13, q_4 = 0.07$	[184]
Rate of entrance in the sexually active population (1/year)	β	$1/45$	[184]
Duration of sexual life (years)	$1/\mu$	45	[184]
Diagnostic delay in AIDS stages, 3 and 4 (years)	$1/\tau_{3,4}$	$1/12$	assumed
Treatment delay after diagnosis in risk group $\boldsymbol{l}$ (years)	$1/\eta$	0.125	[185]
PrEP effectiveness	$1-\Omega$	0.86	[139,186]
Duration on PrEP in risk group $\boldsymbol{l}$ (years)	$1/\kappa_l^{off}$	5	[164]
PrEP uptake rate in risk group $\boldsymbol{l}$ (1/year)	$\kappa_l^{on}(t)$	-	Eq. (9)
Diagnosis rate for individuals in risk group $\boldsymbol{l}$ infected on PrEP (1/year)	τ_l^P	4	[159]
Diagnostic delay (95% CrI) for individuals in acute stage (years)	$1/\tau_1$	$1.31\ (0.81-1.53)$	estimated
Diagnostic delay (95% CrI) for individuals in chronic stage (years)	$1/\tau_2$	$3.42\ (2.59-3.74)$	estimated
Average number of new partners in risk group $\boldsymbol{l}$ in the current situation (1/six months)	c_l	$c_1 = 0.52, c_2 = 3.81,$ $c_3 = 11.16, c_4 = 28.56$	estimated
Average number of new partners in risk group $\boldsymbol{l}$ after introducing PTC cure (1/six months)	c_l	$c_1 = 1.03, c_2 = 6.42,$ $c_3 = 17.15, c_4 = 40.02$	estimated
Average number of new partners in risk group $\boldsymbol{l}$ after introducing elimination cure (1/six months)	c_l	$c_1 = 0.90, c_2 = 5.90,$ $c_3 = 16.25, c_4 = 39.04$	estimated
Average transmission probability (95% CrI) per partnership	λ	$0.031\ (0.011\ -\ 0.080)$	estimated
Average mixing parameter (95% CrI)	ω	$0.494\ (0.014\ -\ 0.969)$	estimated

Average infectivity (95% CrI) of diagnosed/treated individuals	ε	$0.016\ (4.081 \times 10^{-4} - 5.028 \times 10^{-2})$	estimated
Infected population fraction in risk group 1	Q_1	$0.096\ (0.004 - 0.212)$	estimated
Infected population fraction in risk group 2	Q_2	$0.195\ (0.057 - 0.295)$	estimated
Infected population fraction in risk group 3	Q_3	$0.298\ (0.175 - 0.403)$	estimated
Infected population fraction in risk group 4	Q_4	$0.411\ (0.279 - 0.661)$	estimated
Cure effectiveness (%)	e	Varied	-
Maximal cure uptake in risk group $\boldsymbol{l}$ (%)	$\bar{\alpha}_l^{max}$	Varied	-
Maximum cure uptake rate in risk group $\boldsymbol{l}$ (1/year)	α_l^{max}	$ln[1 - \bar{\alpha}_l^{max}/100]$	-
Cure uptake rate in risk group $\boldsymbol{l}$	$\alpha_l(t)$	Logistic grow within 3 years	[139]
Cure uptake in risk group $\boldsymbol{l}$ (%)	$\bar{\alpha}_l(t)$	$\{1 - exp[-\alpha_l(t)]\} \cdot 100$	-
Diagnostic delay for rebound (PTC)/reinfected (elimination) cases in stage $\boldsymbol{k}$ (years)	$1/\bar{\tau}_k^C$	Varied	-
Duration of viral suppression without ART in the PTC scenario (years)	$1/\phi$	Varied	-

Table 4-4. Uniform priors of free parameters.

Parameter	Interval
$\boldsymbol{\tau_1}$	$[0.01, 2]$
$\boldsymbol{\tau_2}$	$[0.01, 1]$
$\boldsymbol{\lambda}$	$[0.01, 0.15]$
$\boldsymbol{\omega}$	$[0, 1]$
$\boldsymbol{\varepsilon}$	$[0, 0.06]$
$\boldsymbol{M_I/(\beta N_0)}$	$[0, 0.05]$
$\boldsymbol{U_0}$	$[0, 2000]$
$\boldsymbol{Q_l}$, $\boldsymbol{l = 1, \ldots, n}$	$[0, 1]$

Sensitivity analyses figures

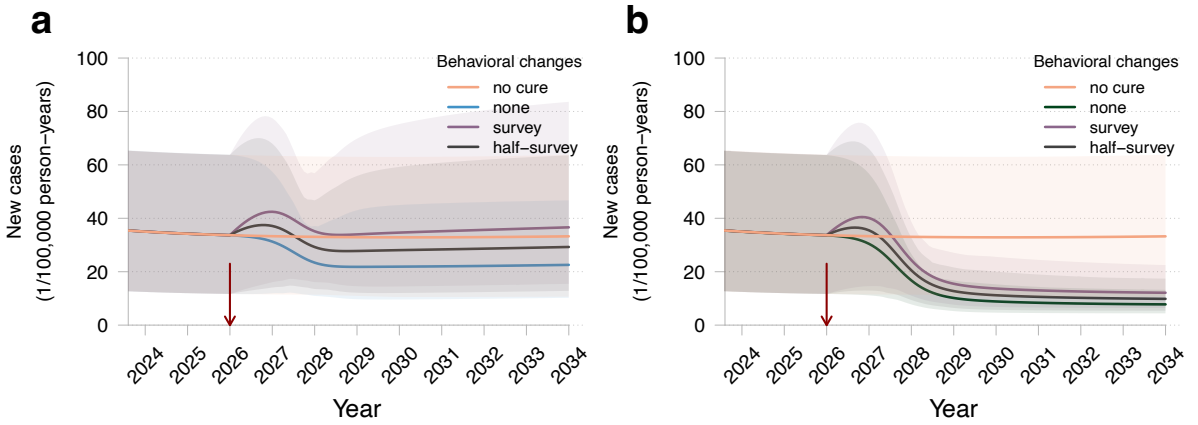

Figure 4-14. Impact of the cure dynamics on the HIV dynamics when behavioral changes are included. New cases (primary infections in naive individuals) with different types of behavioral changes introduced (different colors). The red arrow indicates the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90%, cure uptake of 90%, diagnostic delay of cured individuals who experience a viral rebound of 3 months. ***a*** *PTC cure scenario with time to rebound of 6 years;* ***b*** *elimination cure scenario.*

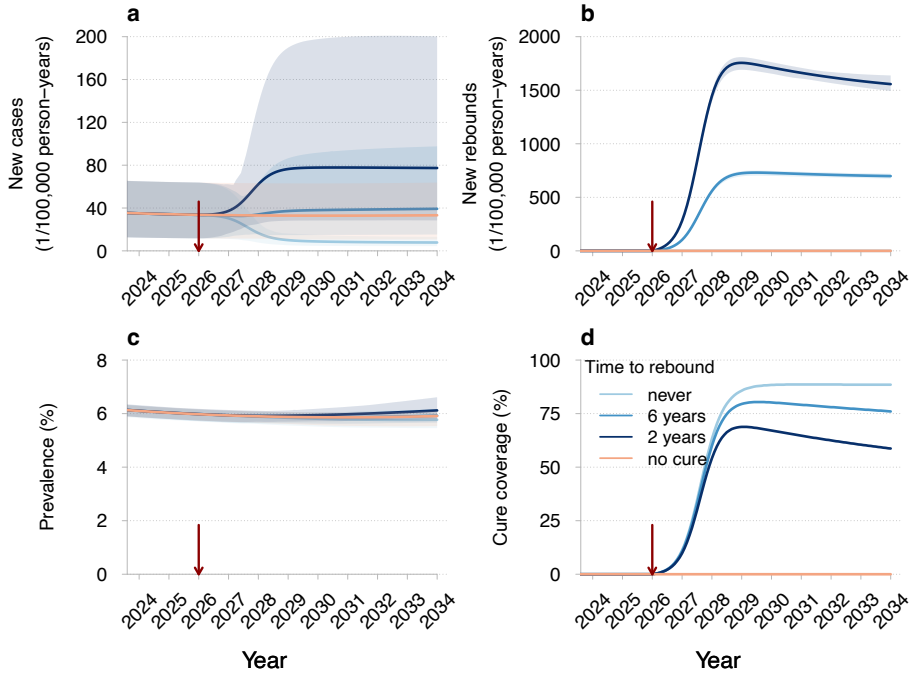

Figure 4-15. Impact of the PTC cure on the HIV dynamics if the first stage after rebound is acute. ***a*** *New cases (primary infections in naive individuals),* ***b*** *new rebounds (new cases of viral rebound in cured individuals),* ***c*** *prevalence (proportion of individuals with HIV), and* ***c*** *cure coverage (proportion of cured individuals among all eligible) as a function cure uptake. The red arrows indicate the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. Different shades of blue correspond to different times until viral rebound. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90%, cure uptake of 90% and diagnostic delay of cured individuals who experience a viral rebound of 3 months.*

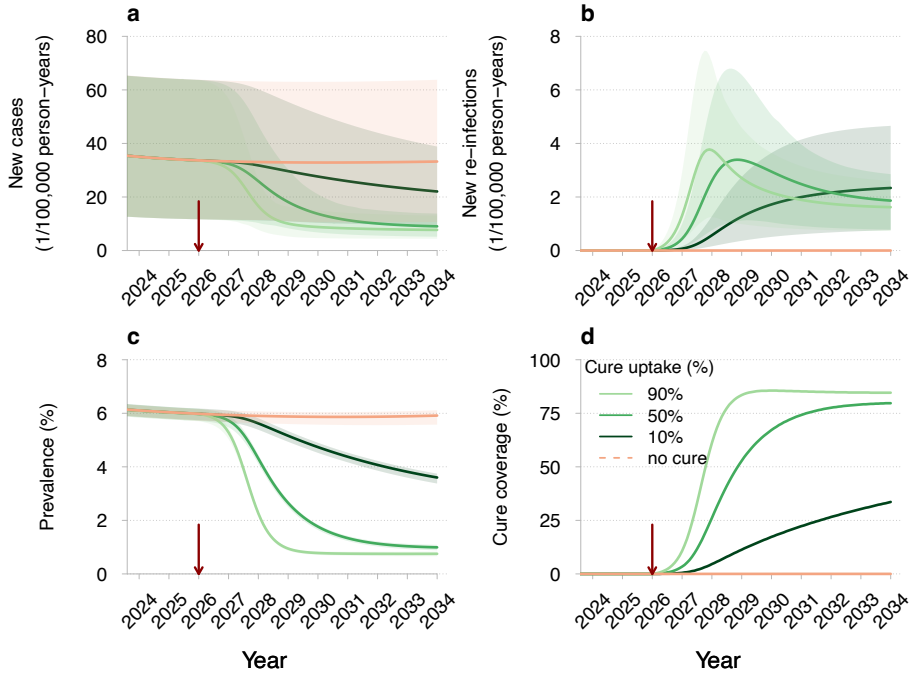

*Figure 4-16. Impact of the elimination cure on the HIV dynamics if the first stage after re-infection chronic. **a** New cases (primary infections in naive individuals), **b** new re-infections (infections in cured individuals), **c** prevalence (proportion of individuals with HIV), and **c** cure coverage (proportion of cured individuals among all eligible) as a function cure uptake. The red arrows indicate the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. Different shades of green correspond to different cure uptakes. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90% and diagnostic delay of cured individuals who experience re-infection of 3 months.*

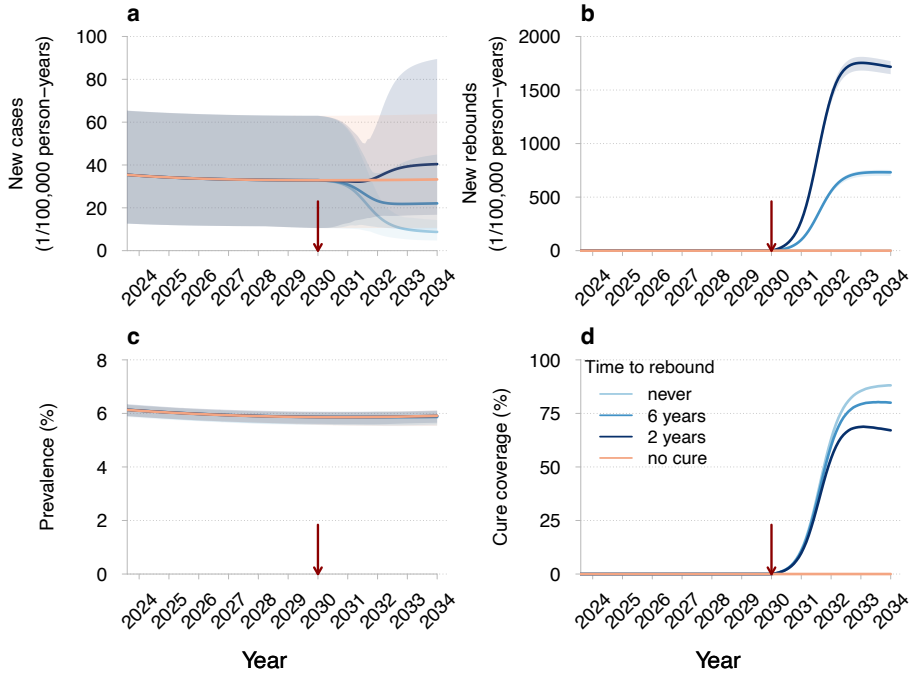

*Figure 4-17. Impact of the PTC cure on the HIV dynamics if the cure is introduced 4 years later, in 2030. **a** New cases (primary infections in naive individuals), **b** new rebounds (new cases of viral rebound in cured individuals), **c** prevalence (proportion of individuals with HIV), and **c** cure coverage (proportion of cured individuals among all eligible) as a function cure uptake. The red arrows indicate the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. Different shades of blue correspond to different times until viral rebound. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90%, cure uptake of 90% and diagnostic delay of cured individuals who experience a viral rebound of 3 months.*

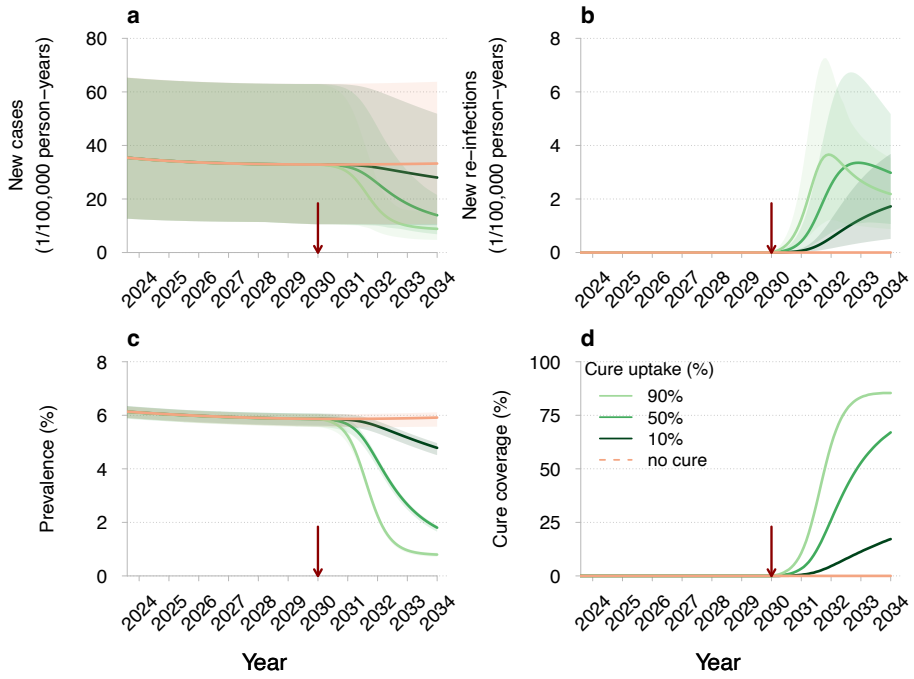

Figure 4-18. Impact of the elimination cure on the HIV dynamics if cure is introduced 4 years later, in 2030. ***a*** *New cases (primary infections in naive individuals),* ***b*** *new re-infections (infections in cured individuals),* ***c*** *prevalence (proportion of individuals with HIV), and* ***c*** *cure coverage (proportion of cured individuals among all eligible) as a function cure uptake. The red arrows indicate the cure introduction. The mean trajectories from the model are shown as solid lines. The shaded regions correspond to 95% prediction intervals based on 100 samples from the joint posterior parameter distribution. Different shades of green correspond to different cure uptakes. The projections of the model without a cure are shown in orange. Parameters: cure efficacy of 90% and diagnostic delay of cured individuals who experience re-infection of 3 months.*

CONCLUSION

In this thesis, I addressed specific public health questions by studying three distinct infectious disease scenarios, each involving different transmission mechanisms caused by different viruses: airborne transmission (SARS-CoV-2), faecal-oral transmission (norovirus), and sexual transmission (HIV). By applying context-specific, data-driven modeling techniques, such as Bayesian probabilistic models, stochastic branching processes, and stratified ODE models, I reconstructed transmission dynamics and assessed the potential impact of public health interventions. These models were tailored to the unique characteristics of each disease and its affected population, in order to provide actionable insights in response to the research questions raised by stakeholders.

In this chapter, I will summarize the main results of these studies, discuss their scientific and practical impact and highlight the critical role of collaborations with data providers and public health stakeholders.

The first two chapters focus on estimating the generation time of the Alpha, Delta, and Omicron variants of SARS-CoV-2, a key parameter for understanding the epidemiology of any pathogen. Knowledge of the generation time distribution is useful for estimating the reproduction number, the primary metric for monitoring epidemic progression [187,188]. It also informs the development of mathematical models used to understand transmission dynamics, evaluate interventions, and forecast epidemic trends. Additionally, from a public health perspective, it plays a practical role in guiding decisions about contact tracing protocols, quarantine durations, and isolation measures.

To provide reliable estimates of the intrinsic generation time, we used household contact tracing data from the Reggio-Emilia province in Italy. We applied on these data a Bayesian inference model to reconstruct transmission chains from line-list epidemiological data. Our estimates of the mean intrinsic generation time for all variants considered ranged from 6 to 6.8 days, consistent with those for the ancestral lineage of SARS-CoV-2. This finding retrospectively supported the continuation of the same contact tracing, quarantine, and isolation protocols used for the ancestral lineage also for the Alpha and Delta waves. However, during the Omicron phase, isolation and quarantine durations were reduced, partly due to the large volume of cases, that would have led to excessive numbers of people being isolated for extended periods, partly to the observation that Omicron generally caused milder illness, and finally due to the belief that Omicron had a shorter generation time, as suggested by several studies at the time. A study from Hong Kong based on contact tracing data estimated the realized generation time in a context with strict control measures that are known to bias downward this estimate [47]. Additional studies had provided preliminary and indirect measures of the generation time using mathematical models describing trajectories of variant frequencies from population-level data in Denmark and the United Kingdom [45,46] (the latter of which

was eventually not published). Both of them suggested a reduced value compared to Delta. Our mathematical model based on household contact tracing data and accounting for transmission heterogeneity provided a more direct and robust estimate of the intrinsic generation time while still reproducing the shorter realized generation times and serial intervals. These results influenced subsequent modeling studies published by independent research groups, achieving a significant scientific impact and reinforcing the importance of accurate generation time estimates in shaping public health strategies (e.g., [189], [190]).

Direct interactions with Reggio Emilia local health authorities were essential for clarifying data collection methods, identifying biases, and defining the appropriate filtering criteria. These discussions also provided a clear understanding of the contact tracing and isolation protocols in place, which were explicitly integrated into our models. Similarly, collaboration with the Istituto Superiore di Sanità offered invaluable insights into nationwide distancing measures, helping refine the accuracy of models, guiding the interpretation of the findings, and their appropriate communication to the public.

In the third chapter, I focused on modeling norovirus transmission aboard cruise ships, where confined environments amplify the risk of large-scale outbreaks. These outbreaks not only pose significant health risks but also have considerable economic implications, as cruise operators face high costs associated with infection control measures. Mathematical modeling studies on norovirus transmission in such settings remain extremely limited. To date, we identified only one other study that evaluated the effectiveness of isolation in reducing outbreak size [50]. The study estimated that the reduction in the number of cases following complete and permanent isolation immediately after symptom onset was between 85% and 97%, close to our estimate of "perfect isolation" at 91%. Our model was able to further evaluate the effectiveness of more realistic isolation protocols and to evaluate the impact of diagnostic delays. In particular, we found that, in the presence of diagnostic delays, an isolation protocol of 72 hours for each diagnosed case (as implemented during the considered outbreak) prevented 71% of potential cases. Halving diagnostic delays could further decrease the size of outbreaks by increasing the effectiveness to 80%.

Our findings also revealed the presence of norovirus superspreaders who contributed disproportionately to transmission. Superspreaders were associated with significantly longer diagnostic delays and lower frequency of symptoms, suggesting that the latter induced the patients to postpone health-seeking, thereby allowing for more occasions to transmit the infection before isolation. This result highlights the importance of passenger education towards seeking immediate health assistance when experiencing gastrointestinal symptoms on board, to reduce the probability of outbreaks and their cumulative incidence.

Once again, there were extensive interactions with the University of Thessaly for understanding the specificities of transmission onboard cruise ships and identify potential biases in data collection, such as underreporting. Their input was additionally essential for interpreting the significance of results within the domain of maritime transport epidemiology, which has its own peculiarities and practices.

The results of this study have also laid the foundation for additional research, in which I

am involved as a co-author. This new work continues to explore norovirus outbreak dynamics on cruise ships, with a particular focus on the performance of automated real-time forecasting models. In recent years, the task of forecasting infectious disease has been a growing field of investigation, with applications to endemic diseases such as seasonal influenza [191] or emerging infections such as Ebola [192] or COVID-19 [193]. However, while research efforts have been mostly focused on open populations, with the aim of supporting national or regional public health authorities, our study is the first to address the problem of forecasting on semi-closed populations. In this work, we assess the performance of a real-time forecasting system projecting norovirus cases on board cruise ships, in order to evaluate the feasibility of an integration of forecasts within the ship's surveillance system software. By comparing different prediction models, we found that the best forecasting performance came from a model that incorporates both diagnostic delays before case isolation and superspreading phenomena. Specifically, by integrating overdispersion in the distribution of secondary cases, as estimated in Chapter 3, we demonstrated that up to 78% of data points were more accurately predicted by models accounting for transmission heterogeneity caused by superspreaders, compared to a standard baseline used to benchmark forecasts. These findings might have practical implications for cruise line companies. Having demonstrated the accuracy of simple predictive models paves the way for integrating them into syndromic surveillance systems onboard, with the potential to enhance and improve outbreak monitoring, management and prevention. This integration is currently being prototyped by engineers of the National&Kapodistrian University of Athens within the Healthy Sailing project.

In the fourth chapter, I focused on the population-level impact of introducing a potential HIV cure among men who have sex with men (MSM) in the Netherlands. Research on how a prospective HIV cure might influence transmission dynamics at the population level is still limited. We identified only two studies that explored this issue focusing on African settings or heterosexual populations [56,57], leaving a gap in understanding how a potential cure might affect the epidemic dynamics within MSM communities in high-income countries. To fill this gap, we employed a compartmental deterministic model that incorporated heterogeneity in sexual activity and infectiousness profiles, calibrated using national behavioral and epidemiological data. Our model-based evaluation indicated that caution is needed in the future deployment of HIV cures. Specifically, we found that a cure that would temporarily suppress HIV in the host, but is subject to failure and viral rebound, could disrupt the current declining trend in HIV incidence observed in the Netherlands. In contrast, a cure that would completely eliminate the virus from the host would consistently reduce HIV cases despite the possibility of reinfection for cured individuals, leading to a reduction of up to 50% in cumulative incidence over the next decade.

Our findings qualitatively align with previous modeling studies focused on generalized HIV epidemics in heterosexual populations in Africa [56,57], which similarly indicated that interventions lacking sustained viral suppression could result in an increase in HIV incidence.

Also in this case, close collaboration with researchers from diverse backgrounds at both RIVM and UMCU was essential in refining the model and ensuring its relevance to real-world public health contexts. The RIVM provided key epidemiological data, while UMCU social scientists offered insights into behavioral patterns within the MSM community,

improving the accuracy of our model's assumptions. Additionally, interactions with medical experts, both from RIVM and UMCU, improved our understanding of HIV treatments, guiding the selection and simulation of cure scenarios.

In addition to the work presented in Chapters 1-4, I have been leading the implementation of two further research projects that are ongoing and have not yet yielded results substantial enough to be included in this thesis. Both studies involve the use of individual-based models (IBM), a class of models particularly well-suited for simulating complex population settings and capturing individual behaviors, allowing for the assessment of individually targeted intervention strategies.

The first was the development of an IBM designed to replicate social and epidemiological dynamics aboard cruise ships in the context of COVID-19 outbreaks in the fall of 2020. By simulating the daily activities of passengers and crew members, such as their cabin routines, mealtimes, participation in onboard entertainment, and port visits, this model provides a highly detailed representation of the population movements and gathering in different spaces, also accounting for restrictions existing at the time. We leverage data provided by the MSC cruise line company, including ship layouts, cabin divisions, capacity and opening hours of entertainment venues onboard, and we integrate epidemiological dynamics of SARS-CoV-2 transmission and testing. This model aims to understand the transmission dynamics onboard and to assess the effectiveness of various intervention strategies, such as individual isolation and closure of venues, with the perspective of contributing to optimize the management of COVID-19 outbreaks and other respiratory viruses on board.

Additionally, I developed another IBM to replicate the dynamics of patients within a pediatric hospital. This model was built using detailed data directly provided by the hospital, including movement patterns between wards, admissions and discharges, and the distributions of durations of stay in the hospital. The model is applied to the study of the transmission dynamics of *methicillin-resistant Staphylococcus aureus* (MRSA) [194], a significant threat in healthcare settings. We additionally incorporated data on diagnostic testing, isolation protocols, treatment regimens and patient medical procedures that are relevant for MRSA transmission. This work aims to provide insights on the impact of patient transfer networks across wards and infection control measures (e.g., testing strategies at admission, isolation and cohorting), and fits in the broader scope of understanding and contrasting the spread of antimicrobial-resistance (AMR).

In summary, all the work performed within the thesis shares the common objective of addressing practical research questions to tackle real-world public health challenges. Close collaboration with data providers and domain experts, who were also the main beneficiaries of the research outcomes, was essential throughout my work. This collaborative approach is crucial in modern epidemiology [2] and has increased the robustness of the presented models and the relevance of produced results through a deeper, otherwise unattainable, understanding of the dynamics at play.

In recent years, the frequency of new epidemic threats due to emerging and re-emerging pathogens has increased, driven by a range of complex and interconnected factors. Globalization, with its unprecedented levels of international travel, facilitates the rapid spread of pathogens across borders [195]. Urbanization, which results in more densely

populated environments, provides fertile ground for diseases to spread [196]. Growing socio-economic inequalities increase the risk of disease exposure and limit access to healthcare in vulnerable regions, possibly hampering the containment of epidemics at the source [197]. Global warming is expanding the habitats of vectors like mosquitoes, contributing to the rise of vector-borne diseases in previously unaffected geographic areas [198,199]. Other anthropogenic environmental changes such as the disruption of natural habitats affect the interface between humans and wildlife, favoring their encounters and increasing the probability of zoonotic spillover events and the adaptation of new pathogens to humans [200]. A specific emerging threat is constituted by the progressive spread of AMR which risks disrupting decades of progress in the treatment of bacterial and fungal infections [201]. As public health authorities worldwide face these new challenges, the incorporation of epidemiological modeling in surveillance and decision making is increasingly becoming a strategic tool. To be effective, models must capture the intricate dynamics of disease transmission across diverse populations and settings, as well as the specificities of available data for the considered research question. The tailored approach adopted in this thesis, integrating diverse expertise and data within modeling activities, may result in more reliable insights for guiding policymakers toward better informed decisions.

BIBLIOGRAPHY

1. Heesterbeek H, Anderson RM, Andreasen V, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015;347(6227):aaa4339.

2. Kretzschmar M. Disease modeling for public health: added value, challenges, and institutional constraints. Journal of public health policy. 2020;41(1):39-51.

3. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character. 1927;115(772):700-721.

4. Murray JD. Mathematical biology: I. An introduction. Springer Science & Business Media. 2007.

5. Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of mathematical biology. 1990;28:365-382.

6. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton university press. 2011.

7. Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences. 2009;106(36):15274-15278.

8. Heesterbeek JAP, Roberts MG. How mathematical epidemiology became a field of biology: a commentary on Anderson and May (1981) "The population dynamics of microparasites and their invertebrate hosts". Philosophical Transactions of the Royal Society B: Biological Sciences. 2015;370(1666):20140307.

9. Anderson RM. Infectious diseases of humans: dynamics and control. Cambridge University Press. 1991.

10. Anderson RM, May RM. Directly transmitted infections diseases: control by vaccination. Science. 1982;215(4536):1053-1060.

11. Grenfell BT, Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. Epidemiology & Infection. 1985;95(2):419-436.

12. Schenzle D. An age-structured model of pre- and post-vaccination measles transmission. Mathematical Medicine and Biology: A Journal of the IMA. 1984;1(2):169–191.

13. Bokler B. Chaos and complexity in measles models: a comparative numerical study. Mathematical Medicine and Biology: A Journal of the IMA. 1993;10(2):83-95.

14. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American journal of epidemiology. 2006;164(10):936-944.

15. van Boven M, Kretzschmar M, Wallinga J, O'Neill PD, Wichmann O, Hahné S. Estimation of measles vaccine efficacy and critical vaccination coverage in a highly vaccinated population. Journal of the Royal Society Interface. 2010;7(52):1537-1544.

16. Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine. 2008;5(3):e74.

17. Quinn TC, Wawer MJ, Sewankambo N, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. New England journal of medicine. 2000;342(13):921-929.

18. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. Proceedings of the National Academy of Sciences. 2007;104(44):17441-17446.

19. Anderson R, May RM. Epidemiological parameters of HI V transmission. Nature. 1988:333:514–519.

20. Hethcote HW, Yorke JA, Hethcote HW, Yorke JA. Modeling gonorrhea in a population with a core group. Gonorrhea Transmission Dynamics and Control. 1984;32-48.

21. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Åberg Y. The web of human sexual contacts. Nature. 2001;411(6840):907–908.

22. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355-359.

23. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Reviews of modern physics. 2015;87(3):925-979.

24. Zhang Q, Sun K, Chinazzi M, et al. Spread of Zika virus in the Americas. Proceedings of the national academy of sciences. 2017;114(22):e4

25. Guzzetta G, Marques-Toledo CA, Rosà R, Teixeira M, Merler S. Quantifying the spatial spread of dengue in a non-endemic Brazilian metropolis via transmission chain reconstruction. Nature communications. 2018;9(1):2837.

26. Apolloni A, Poletto C, Ramasco JJ, Jensen P, Colizza V. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. Theoretical Biology and Medical Modelling. 2014;11:1-26.

27. Eubank S, Guclu H, Anil Kumar VS, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004;429(6988):180-184.

28. Germann TC, Kadau K, Longini Jr IM, Macken CA. Mitigation strategies for pandemic influenza in the United States. Proceedings of the National Academy of Sciences. 2006;103(15):5935-5940.

29. Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. Proceedings of the Royal Society B: Biological Sciences. 2010;277(1681):557-565.

30. Di Ruscio F, Guzzetta G, Bjørnholt JV, et al. Quantifying the transmission dynamics of MRSA in the community and healthcare settings in a low-prevalence country. Proceedings of the National Academy of Sciences. 2019;116(29):14599-14605.

31. Guzzetta G, Marziano V, Mammone A, et al. The decline of the 2022 Italian mpox epidemic: Role of behavior changes and control strategies. Nature Communications. 2024;15(1):2283.

32. Kretzschmar ME, Rozhnova G, Bootsma MC, van Boven M, van de Wijgert JH, Bonten MJ. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. The Lancet Public Health. 2020;5(8):e452-e459.

33. Lau MS, Dalziel BD, Funk S, McClelland A, Tiffany A, Riley S, Grenfell BT. Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. Proceedings of the National Academy of Sciences. 2017;114(9):2337-2342.

34. Salje H, Lessler J, Paul KK, et al. How social structures, space, and behaviors shape the spread of

infectious diseases using chikungunya as a case study. Proceedings of the National Academy of Sciences. 2016;113(47):13420-13425.

35. Hampson K, Dushoff J, Cleaveland S, Haydon DT, Kaare M, Packer C, Dobson A. Transmission dynamics and prospects for the elimination of canine rabies. PLoS Biology. 2009;7(3):e1000053.

36. Lemey P, Hong SL, Hill V, et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. Nature communications. 2020;11(1):5110.

37. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B: Biological Sciences. 2007;274(1609):599-604.

38. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine. 2020;172(9):577-582.

39. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLoS medicine. 2020;17(9):e1003346.

40. Liu QH, Ajelli M, Aleta A, Merler S, Moreno Y, Vespignani A. Measurability of the epidemic reproduction number in data-driven contact networks. Proceedings of the National Academy of Sciences. 2018;115(50):12680–12685.

41. Champredon D, Dushoff J. Intrinsic and realized generation intervals in infectious-disease transmission. Proceedings of the Royal Society B. 2015;282(1821):20152026.

42. Istituto Superiore di Sanità. https://www.iss.it/

43. Azienda USL di Reggio Emilia. https://www.ausl.re.it/

44. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S. Generation time of the Alpha and Delta SARS-CoV-2 variants: an epidemiological analysis. The Lancet Infectious Diseases. 2022;22(5):603-610.

45. Ito K, Piantham C, Nishiura H. Estimating relative generation times and relative reproduction numbers of Omicron BA.1 and BA.2 with respect to Delta in Denmark. medRxiv. 2022. https://www.medrxiv.org/content/10.1101/2022.03.02.22271767. Accessed 16 June 2022.

46. Abbott S, Sherratt K, Gerstung M, Funk S. Estimation of the test to test distribution as a proxy for generation interval distribution for the Omicron variant in England. medRxiv. 2022.

47. Mesfin Y, Chen D, Bond H, et al. Epidemiology of infections with SARS-CoV-2 Omicron BA.2 variant in Hong Kong, January-March 2022. medRxiv. 2022. https://www.medrxiv.org/content/10.1101/2022.04.07.22273595. Accessed 16 June 2022

48. Healthy Sailing. https://healthysailing.eu/

49. De Graaf M, van Beek J, Koopmans MP. Human norovirus transmission and evolution in a changing world. Nature Reviews Microbiology. 2016;14(7):421-433.

50. Towers S, Chen J, Cruz C, et al. Quantifying the relative effects of environmental and direct transmission of norovirus. Royal Society open science. 2018;5(3):170602.

51. University Medical Center Utrecht. https://www.umcutrecht.nl/en

52. Rijksinstituut voor Volksgezondheid en Milieu. https://www.rivm.nl/en

53. The Joint United Nations Programme on HIV/AIDS (UNAIDS). UNAIDS Key Population Atlas; 2024. Available from: https://kpatlas.unaids.org/dashboard.

54. The Joint United Nations Programme on HIV/AIDS (UNAIDS). Global HIV & AIDS statistics — Fact sheet; 2024. Available from: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

55. Schou MD, Søgaard OO, Rasmussen TA. Clinical trials aimed at HIV cure or remission: new pathways and lessons learned. Expert Review of Anti-infective Therapy. 2023;21(11):1227–1243.

56. Dimitrov DT, Kiem HP, Jerome KR, Johnston C, Schiffer JT. A curative regimen would decrease HIV prevalence but not HIV incidence unless targeted to an ART-naïve population. Scientific reports. 2016;6(1):22183.

57. Beacroft L, Hallett TB. The potential impact of a "curative intervention" for HIV: a modelling study. Global Health Research and Policy. 2019;4:1–8.

58. WHO. Tracking SARS-CoV-2 variants. Available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/

59. ECDC. Contact tracing: public health management of persons, including healthcare workers, who have had contact with COVID-19 cases in the European Union – third update, 18 November 2020. Stockholm: ECDC;2021.

60. ECDC. Contact tracing in the European Union: public health management of persons, including healthcare workers, who have had contact with COVID-19 cases – fourth update, 28 October 2021.

61. CDC. Operational Considerations for Adapting a Contact Tracing Program to Respond to the COVID-19 Pandemic in non-US Settings. Last updated September 22, 2021. Available at: https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/operational-considerations-contact-tracing.html

62. CDC. Considerations for Case Investigation and Contact Tracing in K-12 Schools and Institutions of Higher Education (IHEs). Available at: https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/contact-tracing.html

63. Riccardo F, Guzzetta G, Urdiales AM, Del Manso M, Andrianou CD, Bella A, et al. COVID-19 response: effectiveness of weekly rapid risk assessments, Italy. Bulletin of the World Health Organization. 2022;100(2):161.

64. Djuric O, Larosa E, Cassinadri M, et al. Surveillance, contact tracing and characteristics of SARS-CoV-2 transmission in educational settings in Northern Italy, September 2020 to April 2021. Plos one. 2022;17(10):e0275667.

65. Istituto Superiore di Sanità. Monitoraggio delle varianti del virus SARS-CoV-2 di interesse in sanità pubblica in Italia. Available at https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti-indagini-rapide. Accessed 16 June 2022.

66. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. The Lancet Infectious Diseases. 2020;20(7):793-802.

67. Hu S, Wang W, Wang Y, Litvinova M, Luo K, Ren L, et al. Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. Nature communications. 2021;12(1):1-11.

68. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel. 1986;81(10.5555):26853.

69. Guzzetta G, Minosse C, Pisapia R, et al. Household transmission and disease transmissibility of a large HAV outbreak in Lazio, Italy, 2016–2017. Epidemics. 2019;29:100351.

70. Manica M, De Bellis A, Guzzetta G, et al. Intrinsic generation time of the SARS-CoV-2 Omicron variant: An observational study of household transmission. The Lancet Regional Health-Europe. 2022;19:100430.

71. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine. 2020;382:1199-1207.

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

73. 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(8).

74. Tian S, Hu N, Lou J, Chen K, Kang X, Xiang Z, et al. Characteristics of COVID-19 infection in Beijing. Journal of Infection. 2020;80(4):401-406.

75. 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(8):911-919.

76. Tan WY, Wong LY, Leo YS, Toh MP. Does incubation period of COVID-19 vary with age? A study of epidemiologically linked cases in Singapore. Epidemiology & Infection. 2020;148.

77. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368(6491).

78. 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(17):2000257.

79. Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID-19 symptom onset highlights the importance of continued contact tracing. Elife. 2021;10.

80. Lau YC, Tsang TK, Kennedy-Shaffer L, Kahn R, Lau EHY, Chen D, et al. Joint Estimation Of Generation Time And Incubation Period For Coronavirus Disease (Covid-19). The Journal of Infectious Diseases. 2021;224(10):1664-1671.

81. Challen R, Brooks-Pollock E, Tsaneva-Atanasova K, Danon L. Meta-analysis of the severe acute respiratory syndrome coronavirus 2 serial intervals and the impact of parameter uncertainty on the coronavirus disease 2019 reproduction number. Statistical Methods in Medical Research. 2021;09622802211065159.

82. Cereda D, Manica M, Tirani M, et al. The early phase of the COVID-19 epidemic in Lombardy, Italy. Epidemics. 2021;37:100528.

83. Buchholz U, Schulze K, An der Heiden M. Household clusters reveal household-and variant-specific properties of SARS-CoV-2. Epidemiology & Infection. 2022;7:1-9.

84. Stefanelli P, Trentini F, Guzzetta G, Marziano V, Mammone A, Sane Schepisi M, et al. Co-circulation of SARS-CoV-2 Alpha and Gamma variants in Italy, February and March 2021. Eurosurveillance. 2022;27(5):2100429.

85. Vazquez-Prokopec GM, Montgomery BL, Horne P, Clennon JA, Ritchie SA. Combining contact tracing with targeted indoor residual spraying significantly reduces dengue transmission. Science advances. 2017;3(2).

86. Guzzetta G, Vairo F, Mammone A, et al. Spatial modes for transmission of chikungunya virus during a large chikungunya outbreak in Italy: a modeling analysis. BMC Medicine. 2020;18(1):1–10.

87. Hellewell J, Russell TW, The SAFER Investigators and Field Study Team, 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.

88. Riccardo F, Guzzetta G, Mateo Urdiales A, Del Manso M, Andrianou XD, Bella A, et al. COVID-19 response: effectiveness of weekly rapid risk assessments, Italy. Bulletin of the World Health Organization. 2022;100(2):161.

89. Fabiani M, Puopolo M, Morciano C, Spuri M, Alegiani SS, Filia A, et al. Effectiveness of mRNA vaccines and waning of protection against SARS-CoV-2 infection and severe covid-19 during predominant circulation of the delta variant in Italy: retrospective cohort study. BMJ. 2022;376.

90. Menegale F, Manica M, Zardini A, et al. Evaluation of waning of SARS-CoV-2 vaccine–induced immunity: a systematic review and meta-analysis. JAMA Network Open. 2023;6(5):e2310650-e2310650.

91. Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G. Effect of Vaccination on Household Transmission of SARS-CoV-2 in England. New England journal of medicine. 2021;385(8): 759-760.

92. Lipsitch M, Kahn R. Interpreting vaccine efficacy trial results for infection and transmission. Vaccine. 2021;39(30):4082-4088.

93. Marziano V, Guzzetta G, Mammone A, et al. The effect of COVID-19 vaccination in Italy and perspectives for living with the virus. Nature Communications. 2021;12:7272.

94. Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature. 2022;603:679–686.

95. Hui KP, Ho JC, Cheung MC, et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. Nature. 2022;603:1–5.

96. Cele S, Jackson L, Khoury DS, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. Nature. 2022;602(7898):654–656.

97. Planas D, Saunders N, Maes P, et al. Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. Nature. 2022;602 (7898):671–675.

98. Brandal LT, MacDonald E, Veneti L, et al. Outbreak caused by the SARS-CoV-2 Omicron variant in Norway, November to December 2021. Eurosurveillance. 2021;26(50):2101147.

99. Song JS, Lee J, Kim M, et al. Serial intervals and household transmission of SARS-CoV-2 Omicron variant, South Korea, 2021. Emerging Infectious Diseases. 2022;28(3):756.

100. Jansen L, Tegomoh B, Lange K, et al. Investigation of a SARS-CoV-2 B.1.1.529 (Omicron) variant cluster—Nebraska, November–December 2021. Morbid Mortal Week Rep. 2021;70(5152):1782.

101. Backer JA, Eggink D, Andeweg SP, et al. Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021. Eurosurveillance. 2022;27(6):2200042.

102. Walensky RP. CDC updates and shortens recommended isolation and quarantine period for general population: media statement for immediate release: Monday, December 27, 2021. Available at https://www.cdc.gov/media/releases/2021/s1227-isolation-quarantine-guidance.html. Accessed 16 June 2022.

103. European Center for Diseases Control. ECDC updates its guidance regarding quarantine and isolation considering the rapid spread of Omicron in the EU/EEA. Available at https://www.ecdc.europa.eu/en/news-events/ecdc-updates-guidance-regarding-quarantine-and-isolation-considering-spread-of-omicron. Accessed 16 June 2022.

104. Djuric O, Larosa E, Cassinadri M, et al. Secondary transmission of SARS-CoV-2 in educational settings in Northern Italy from September 2020 to April 2021: a population-based study. medRxiv; https://doi.org/10.1101/2021.09.03.21263061. Accessed 16 June 2022.

105. Manica M, Litvinova M, De Bellis A, et al. Estimation of the incubation period and generation time of SARS-CoV-2 Alpha and Delta variants from contact tracing data. Epidemiology&Infection. 2023;151:e5.

106. Altarawneh HN, Chemaitelly H, Hasan MR, et al. Protection against the Omicron variant from previous SARS-CoV-2 infection. New England Journal of Medicine. 2022;386:1288–1290.

107. Struttura Commissariale per l'Emergenza Covid-19. Open data on COVID-19 vaccination in Italy. Available at https://github.com/italia/covid19-opendata-vaccini. Accessed 16 June 2022.

108. Istituto Nazionale di Statistica (ISTAT). Censimento Popolazione Abitazioni. Stanze occupate da persone residenti. Available at http://dati-censimentopopolazione.istat.it/Index.aspx?DataSetCode=DICA_STANZE. Accessed 16 June 2022.

109. He X, Lau EH, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature Medicine. 2020;26(5):672–675.

110. Boucau J, Marino C, Regan J, et al. Duration of viable virus shedding in SARS-CoV-2 Omicron variant infection. medRxiv. 2022. https://www.medrxiv.org/content/10.1101/2022.03.01.22271582. Accessed 16 June 2022.

111. Torjesen I. Covid-19: Peak of viral shedding is later with Omicron variant, Japanese data suggest. BMJ. 2022;376.

112. Hay JA, Kissler SM, Fauver JR, et al. Viral dynamics and duration of PCR positivity of the SARS-CoV-2 Omicron variant. medRxiv. 2022. https://www.medrxiv.org/content/10.1101/2022.01.13.22269257. Accessed 16 June 2022.

113. Riccardo F, Ajelli M, Andrianou XD, Bella A, Del Manso M , Fabiani M, et al. COVID-19 working group. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. Eurosurveillance. 2020;25(49):2000790.

114. van Boven M, van Dorp CH, Westerhof I, et al. Estimation of introduction and transmission rates of SARS-CoV-2 in a prospective household study. PLoS computational biology. 2024;20(1):e1011832.

115. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, Gower C, Kall M, Groves N, O'Connell AM, Simons D. Covid-19 vaccine effectiveness against the omicron (B. 1.1. 529) variant. New England Journal of Medicine. 2022;386(16):1532-1546.

116. McEvoy D, McAloon C, Collins 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.

117. ISTAT. Popolazione e famiglie. Available at https://www.istat.it/it/popolazione-e-famiglie?dati

118. Jenkins KA, Vaughan GH Jr, Rodriguez LO, Freeland A. Acute gastroenteritis on cruise ships - maritime illness database and reporting system, United States, 2006-2019. MMWR Surveill Summ. 2021;70(6):1-19.

119. de Graaf M, Villabruna N, Koopmans MP. Capturing norovirus transmission. Curr Opin Virol. 2017;22:64-70.

120. Lopman BA, Steele D, Kirkwood CD, Parashar UD. The Vast and Varied Global Burden of Norovirus: Prospects for Prevention and Control. PLoS Medicine 2016;13(4).

121. Phocuswright. U.S. cruise market report 2021-2025. Rutherford: Phocuswright; 2022. Available from: https://www.phocuswright.com/Travel-Research/Market-Overview-Sizing/US-Cruise-Market-Report-2021-to-2025

122. Cruise Lines International Association (CLIA). 2019 cruise trends & industry outlook. Washington: CLIA; 2019. Available from: https://cruising.org/-/media/eu-resources/pdfs/CLIA%202019-Cruise-Trends--Industry-Outlook

123. EU SHIPSAN ACT Joint Action (20122103). European manual for hygiene standards and communicable disease surveillance on passenger ships. 2nd ed. ISBN 978-960-99647-3-9. Brussels: European Commission; 2016. Available from: https://www.shipsan.eu/Home/EuropeanManual.aspx

124. Centers for Disease Control and Prevention (CDC). Vessel Sanitation Program Operations Manual 2018. Atlanta: US Department of Health and Human Services, CDC; 2018. Available from: www.cdc.gov/vessel-sanitation/media/files/vsp_operations_manual_2018-508.pdf

125. World Health Organization (WHO). Guide to ship sanitation. 3rd ed. Geneva: WHO; 2011. Available from: https://iris.who.int/bitstream/handle/10665/43193/9789241546690_eng.pdf?sequence=1

126. Mouchtouri VA, Verykouki E, Zamfir D, Hadjipetris C, Lewis HC, Hadjichristodoulou C, the EU SHIPSAN ACT partnership. Gastroenteritis outbreaks on cruise ships: contributing factors and thresholds for early outbreak detection. Euro Surveill. 2017;22(45):16-00576.

127. Heijne JC, Teunis P, Morroy G, et al. Enhanced hygiene measures and norovirus transmission during an outbreak. Emerging Infectious Diseases. 2009;15(1):24-30.

128. Adams NL, Young D, Gastañaduy PA, et al. Quantifying the roles of vomiting, diarrhea, and residents vs. staff in norovirus transmission in U.S. nursing home outbreaks. PLoS Computational Biology. 2020;16(10).

129. Wikswo ME, Cortes J, Hall AJ, et al. Disease Transmission and Passenger Behaviors during a High Morbidity Norovirus Outbreak on a Cruise Ship. Clinical Infectious Diseases. 2021; 52(9):1116–1122.

130. Chimonas MA, Vaughe GH, Andre Z, et al. Passenger behaviors associated with norovirus infection on board a cruise ship--Alaska, May to June 2004. Journal of Travel Medicine. 2008;15(3):177-83.

131. Harris TE. The theory of branching processes. Vol. 6. Berlin: Springer. 1963.

132. Götz HM, Ekdahl K, Lindbäck J, de Jong B, Hedlund KO, Giesecke J. Clinical spectrum and transmission characteristics of infection with Norwalk-like virus: Findings from a large community outbreak in Sweden. Clinical Infectious Diseases. 2001;33(5):622-628.

133. Doucet, A, deFreitas, N, Gordon N. An Introduction to Sequential Monte Carlo Methods. New York: Springer. 2001.

134. Juran JM. Quality-Control Handbook. McGraw-Hill. 1951.

135. The Joint United Nations Programme on HIV/AIDS (UNAIDS). Understanding fast-track accelerating action to end the AIDS epidemic by 2030; 2015. Available from: https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf.

136. van Sighem A, Wit F, Boyd A, Smit C, Jongen V, Koole J. Monitoring Report 2023. Human Immunodeficiency Virus (HIV) Infection in the Netherlands. Amsterdam: Stichting hiv monitoring; 2023. Available from: https://www.hiv-monitoring.nl.

137. Brizzi F, Birrell PJ, Kirwan P, Ogaz D, Brown AE, Delpech VC, et al. Tracking elimination of HIV transmission in men who have sex with men in England: a modelling study. The Lancet HIV. 2021;8(7).

138. Palk L, Gerstoft J, Obel N, Blower S. A modeling study of the Danish HIV epidemic in men who have sex with men: travel, pre-exposure prophylaxis and elimination. Scientific Reports. 2018;8(1):16003.

139. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. The Lancet. 2016;387(10013):53–60.

140. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet. 2008;373(9657):48–57.

141. Langebeek N, Kooij KW, Wit FW, Stolte IG, Sprangers MAG, Reiss P, et al. Impact of comorbidity and ageing on health-related quality of life in HIV-positive and HIV-negative individuals. AIDS. 2017;31(10).

142. Galit Zeluf-Andersson LNSJHPM Lars E Eriksson, Ekstr¨om AM. Beyond viral suppression: the quality of life of people living with HIV in Sweden. AIDS Care. 2019;31(4):403–412.

143. van Bilsen WPH, Zimmermann HML, Boyd A, Davidovich U. Burden of living with HIV among men who have sex with men: a mixed-methods study. The Lancet HIV. 2020;7(12).

144. de los Rios P, Okoli C, Castellanos E, Allan B, Young B, Brough G, et al. Physical, emotional, and psychosocial challenges associated with daily dosing of HIV medications and their impact on indicators of quality of life: findings from the positive perspectives study. AIDS and Behavior. 2021;25(3):961–972.

145. Romijnders KAGJ, de Groot L, Vervoort SCJM, Basten MGJ, van Welzen BJ, Kretzschmar ME, et al. The perceived impact of an HIV cure by people living with HIV and key populations vulnerable to HIV in the Netherlands: A qualitative study. Journal of Virus Eradication. 2022;8(1):100066.

146. Ndung'u T, McCune JM, Deeks SG. Why and where an HIV cure is needed and how it might be achieved. Nature. 2019;576(7787):397–405.

147. Dybul M, Attoye T, Baptiste S, Cherutich P, Dabis F, Deeks SG, et al. The case for an HIV cure and how to get there. The Lancet HIV. 2021;8(1).

148. Lewin SR, Attoye T, Bansbach C, Doehle B, Dubé K, Dybul M, et al. Multi-stakeholder consensus on a target product profile for an HIV cure. The Lancet HIV. 2021;8(1).

149. Group TA. Research Toward a Cure Trials; 2024. Available from: www.treatmentactiongroup.org/cure/trials.

150. Khan A, Paneerselvam N, Lawson BR. Antiretrovirals to CCR5 CRISPR/Cas9 gene editing-A paradigm

shift chasing an HIV cure. Clinical Immunology. 2023; p. 109741.

151. Deeks SG, Archin N, Cannon P, et al. Research priorities for an HIV cure: international AIDS society global scientific strategy 2021. Nature medicine. 2021;27(12):2085–2098.

152. Pitman MC, Lau JSY, McMahon JH, Lewin SR. Barriers and strategies to achieve a cure for HIV. The Lancet HIV. 2018;5(6).

153. Vansant G, Bruggemans A, Janssens J, Debyser Z. Block-and-lock strategies to cure HIV infection. Viruses. 2020;12(1):84.

154. Bailon L, Mothe B, Berman L, Brander C. Novel approaches towards a functional cure of HIV/AIDS. Drugs. 2020;80(9):859–868.

155. Landovitz RJ, Scott H, Deeks SG. Prevention, treatment and cure of HIV infection. Nature Reviews Microbiology. 2023;21(10):657–670.

156. Garnett GP. An introduction to mathematical models in sexually transmitted disease epidemiology. Sex Transm Infect. 2002;78(1):7–12.

157. Giddings R, Indravudh P, Medley GF, Bozzani F, Gafos M, Malhotra S, et al. Infectious disease modelling of HIV prevention interventions: a systematic review and narrative synthesis of compartmental models. PharmacoEconomics. 2023;41(6):693–707.

158. Hallett TB, Menzies NA, Revill P, Keebler D, Borquez A, McRobie E, et al. Using modeling to inform international guidelines for antiretroviral treatment. Aids. 2014;28:S1–S4.

159. Nederlandse multidisciplinaire richtlijn. Pre-expositie profylaxe (PrEP) ter preventie van hiv; 2022. Available from: https://www.soaaids.nl/files/2022-07/20220711-PrEP-richtlijn-Nederland-versie-3-update-2022.pdf.

160. Stecher M, Claßen A, Klein F, Lehmann C, Gruell H, Platten M, et al. Systematic review and meta-analysis of treatment interruptions in human immunodeficiency virus (HIV) type 1–infected patients receiving antiretroviral therapy: implications for future HIV cure trials. Clinical Infectious Diseases. 2020;70(7):1406–1417.

161. Reitsema M, Wallinga J, van Sighem AI, Bezemer D, van der Valk M, van Aar F, et al. Impact of varying pre-exposure prophylaxis programs on HIV and Neisseria gonorrhoeae transmission among MSM in the Netherlands: a modelling study. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2022; p. 10–1097.

162. Freedberg KA, Possas C, Deeks S, Ross A, Rosettie KL, Di Mascio M, et al. The HIV cure research agenda: the role of mathematical modelling and cost-effectiveness analysis. Journal of virus eradication. 2015;1(4):245–249.

163. Rozhnova G, Heijne J, Bezemer D, Van Sighem A, Presanis A, De Angelis D, et al. Elimination prospects of the Dutch HIV epidemic among men who have sex with men in the era of preexposure prophylaxis. Aids. 2018;32(17):2615–2623.

164. Nichols BE, Boucher CA, van der Valk M, Rijnders BJ, van de Vijver DA. Cost-effectiveness analysis of preexposure prophylaxis for HIV-1 prevention in the Netherlands: a mathematical modelling study. The Lancet Infectious Diseases. 2016;16(12):1423–1429.

165. Jansen IA, Geskus RB, Davidovich U, Jurriaans S, Coutinho RA, Prins M, et al. Ongoing HIV-1 transmission among men who have sex with men in Amsterdam: a 25-year prospective cohort study. AIDS. 2011;25(4).

166. Lo B, Grady C, on Ethics of the International AIDS Society WG, et al. Ethical considerations in HIV cure research: points to consider. Current Opinion in HIV and AIDS. 2013;8(3):243–249.

167. Dub´e K, Sylla L, Dee L, Taylor J, Evans D, Bruton CD, et al. Research on HIV cure: mapping the ethics landscape. PLoS medicine. 2017;14(12):e1002470.

168. Dub´e K, Kanazawa J, Taylor J, Dee L, Jones N, Roebuck C, et al. Ethics of HIV cure research: an unfinished agenda. BMC medical ethics. 2021;22:1–14.

169. Grossman CI, Ross AL, Auerbach JD, Ananworanich J, Dub´e K, Tucker JD, et al. Towards multidisciplinary HIV-cure research: integrating social science with biomedical research. Trends in microbiology. 2016;24(1):5–11.

170. Ismail SD, Pankrac J, Ndashimye E, Prodger JL, Abrahams MR, Mann JF, et al. Addressing an HIV cure in LMIC. Retrovirology. 2021;18(1):21.

171. Pilcher CD, Joaki G, Hoffman IF, Martinson FE, Mapanje C, Stewart PW, et al. Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection. Aids. 2007;21(13):1723–1730.

172. Castagna A, Muccini C, Galli L, Bigoloni A, Poli A, Spagnuolo V, et al. Analytical treatment interruption in chronic HIV-1 infection: time and magnitude of viral rebound in adults with 10 years of undetectable viral load and low HIV-DNA (APACHE study). Journal of Antimicrobial Chemotherapy. 2019;74(7):2039–2046.

173. Van der Kuyl AC, Cornelissen M. Identifying HIV-1 dual infections. Retrovirology. 2007;4:1–12.

174. Lambers FAE, Prins M, Thomas X, Molenkamp R, Kwa D, Brinkman K, et al. Alarming incidence of hepatitis C virus re-infection after treatment of sexually acquired acute hepatitis C virus infection in HIV-infected MSM. AIDS. 2011;25(17).

175. Heffernan A, Cooke GS, Nayagam S, Thursz M, Hallett TB. Scaling up prevention and treatment towards the elimination of hepatitis C: a global mathematical model. The Lancet. 2019;393(10178):1319–1329.

176. Aids Care Service programs; 2024. Available from: https://www.aidscareservice.org/programs.

177. Rozhnova G, Heijne JC, Basten M, den Daas C, Matser A, Kretzschmar M. Impact of sexual trajectories of men who have sex with men on the reduction in HIV transmission by pre-exposure prophylaxis. Epidemics. 2019;28:100337.

178. Kretzschmar ME, Schim van der Loeff MF, Birrell PJ, De Angelis D, Coutinho RA. Prospects of elimination of HIV with test-and-treat strategy. Proceedings of the National Academy of Sciences USA. 2013;110:15538–15543.

179. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action Concerted Action on SeroConversion to AIDS and Death in Europe . Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Lancet. 2000;355:1131–1137.

180. Birrell PJ, Presanis AM, De Angelis D, Collaboration TC. Multi-state models of HIV progression in homosexual men: an application to the CASCADE collaboration. MRC Biostatistics Unit; 2012.

181. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. Journal of Infectious Diseases. 2008;198:687–693.

182. Bellan SE, Dushoff J, Galvani AP, Meyers LA. Reassessment of HIV-1 acute phase infectivity:

accounting for heterogeneity and study design with simulated cohorts. PLoS Medicine. 2015;12:e1001801.

183. The CASCADE Collaboration Concerted Action on SeroConversion to AIDS and Death in Europe . Survival after introduction of HAART in people with known duration of HIV-1 infection. Lancet. 2000;355:1158–1159.

184. Rozhnova G, van der Loeff MFS, Heijne JCM, Kretzschmar ME. Impact of Heterogeneity in Sexual Behavior on Effectiveness in Reducing HIV Transmission with Test-and-Treat Strategy. PLoS Comput Biol. 2016;12:e1005012.

185. Dijkstra M, van Rooijen MS, Hillebregt MM, Van Sighem A, Smit C, Hogewoning A, et al. Decreased time to viral suppression after implementation of targeted testing and immediate initiation of treatment of acute human immunodeficiency virus infection among men who have sex with men in Amsterdam. Clinical Infectious Diseases. 2021;72(11):1952–1960.

186. Molina JM, Capitant C, Spire B, Pialoux G, Cotte L, Charreau Iea. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. The New England Journal of Medicine. 2015;373:2237–2246.

187. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. American journal of epidemiology. 2013;178(9):1505-1512.

188. Thompson RN, Stockwin JE, van Gaalen RD, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics. 2019;29:100356.

189. Cai, J., Deng, X., Yang, J., Sun, K., Liu, H., Chen, Z., ... & Yu, H. (2022). Modeling transmission of SARS-CoV-2 omicron in China. Nature medicine, 28(7), 1468-1475

190. Fosch A, Aleta A, & Moreno Y. Characterizing the role of human behavior in the effectiveness of contact-tracing applications. Frontiers in Public Health. 2023;11:1266989.

191. Biggerstaff M, Alper D, Dredze M, et al. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. BMC Infectious Diseases. 2016;16:1–10.

192. Viboud C, Sun K, Gaffey R, et al. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics. 2018;22:13–21.

193. Sherratt K, Gruson H, Johnson H, et al. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. Elife. 2023;12:e81916.

194. Lowy FD. Staphylococcus aureus infections. New England Journal of Medicine. 1998;339:520–532.

195. Saker L, Lee K, Cannito B, Gilmore A, Campbell-Lendrum DH. Globalization and infectious diseases: a review of the linkages. WHO. 2004.

196. Neiderud CJ. How urbanization affects the epidemiology of emerging infectious diseases. Infection ecology & epidemiology. 2015;5(1):27060.

197. Tizzoni M, Nsoesie EO, Gauvin L, Karsai M, Perra N, Bansal S. Addressing the socioeconomic divide in computational modeling for infectious diseases. Nature communications. 2022;13(1):2897.

198. WHO. Vector-borne diseases. https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases

199. WHO. Environment. Climate Change and Health. https://www.who.int/teams/environment-climate-change-and-health/emergencies/disease-outbreaks

200. European Parliament. Think Tank. The link between biodiversity loss and the increasing spread of zoonotic diseases. https://www.europarl.europa.eu/thinktank/en/document/IPOL_IDA(2020)658217

201. WHO. Antimicrobial resistance. https://www.who.int/health-topics/antimicrobial-resistance