PhD DISSERTATION

UNIVERSITY OF TRENTO

DEPARTMENT OF MATHEMATICS

Doctoral School in Mathematics

CYCLE XXXVIII

MATHEMATICAL MODELLING FOR REAL-TIME AND PROSPECTIVE ASSESSMENT OF INFECTIOUS DISEASES

Advisors

Dr. Giorgio Guzzetta

Prof. Andrea Pugliese

Prof. Ilaria Dorigatti

Author

Andrea Bizzotto

22nd December 2025

ACKNOWLEDGEMENTS

I believe the best way to learn is to be surrounded by people you can learn each day, and that's exactly what I experienced over these years. I'm grateful to the Health Emergencies group at the Bruno Kessler Foundation, and I thank each member for their generous sharing of expertise.

Special thanks go to my scientific pillar and supervisor, Dr. Giorgio Guzzetta, who helped me refine my work with the same care that nowcasting applies to data. You taught me to look beyond the present toward future projections, like a mathematical model that goes deep to reveal what the data hide rather than stopping at appearances. I never felt alone; your guidance, like that of a cruise-ship captain directing and controlling the ship, always made me feel `in una botte di ferro`.

A profound thanks also goes to Prof. Andrea Pugliese, who brought me into this field and, with his immense scientific expertise, has always been a source of inspiration. Another sincere thanks to Prof. Ilaria Dorigatti, whose professionalism and contagious passion led me to embark on this PhD journey.

I would also like to thank Prof. Tom Britton, who hosted me at Stockholm University in last months, where I felt truly welcome. The long conversations and collaboration with you and Prof. Frank Ball, whom I also thank as another example of immense expertise and genuine passion for his work, have been wonderful.

Finally, I wish to acknowledge all the colleagues I have collaborated with for their generosity and for sharing different points of view, which is precisely what made this journey so stimulating.

CONTENTS

LIST OF FIGURES

which either estimate is the closest one to the reference value, for different values of completeness.

Figure 2.4. *COVID-19 epidemic periods in Italy (2020-2021).* a) Reference reproduction number by date of symptom onset, as computed from consolidated data reported on April 5, 2023 (solid black line). The gray shaded area around the reference reproduction number (visible only in the period May-August 2020, due to the low number of cases contributing to R estimates) represents the 95%CI in its estimate. Dark gray bars above the graph highlight the days where the reference reproduction number was above 1, and dark gray vertical dashed lines delimit epidemic periods (see Table 2.1), labeled with a progressive number just above the x axis. Blue and red bars identify the dates of reporting for which the net and nowcasted reproduction numbers, respectively, estimated at the nearest date afforded by a completeness of 90%, were above 1. b) Dates where the reference reproduction number was above 1 (dark gray), and dates of reporting for which the net and nowcasted reproduction numbers, estimated at the nearest date afforded by a completeness between 50% and 90%, were above 1 (red and blue).

Figure 2.5. *Example of application of the proposed nowcasting technique using data from the Italian COVID-19 integrated surveillance system for a selected reporting date (Δ = April 1, 2021).* a) Approximation of the consolidation distribution. Light blue lines represent consolidation distributions for symptom onset dates between February 1 ($p_{\Delta,N-1}$) and March 2 ($p_{\Delta,0}$), 2021; $p_{\Delta,0}, p_{\Delta,10}, p_{\Delta,20}$ and $p_{\Delta,N-1}$ are reported in thicker and darker lines to highlight the variability across consolidation distributions. Dotted blue lines connect the values of different consolidation distributions for the same number z of days elapsed since the corresponding symptom onset, $p_{\Delta,i}(z)$, for z=0, 10, 20 and θ (gray lines show $p_{\Delta,i}(z)$ for the other values of z). The $p_{\Delta,i}(z)$ distributions for the selected values of z are reported on the right-hand side of the panel as blue dots, and their mean, corresponding to the estimated value of $\hat{\pi}_{\Delta}(z)$, is reported further to the right as a dark red dot. Pink dots represent the mean of $p_{\Delta,i}(z)$ distributions, corresponding to the estimated value of $\hat{\pi}_{\Delta}(z)$, for the remaining values of z. b) Estimated consolidation distribution and corresponding consolidation lags. Light blue dots represent the distributions of $p_{\Delta,i}(z)$ for all values of z; dark red dots represent the estimated consolidation distribution at Δ, $\hat{\pi}_{\Delta}(z)$. Horizontal lines define selected completeness thresholds (50%, 70% and 90%) and corresponding vertical lines define the corresponding consolidation lags.

Figure 2.6. *Error metrics for assessing the choice of θ.* a) Mean absolute percentage error, $\eta(\theta)$; b) Root mean square error, $\varepsilon(\theta)$. The error value for the adopted value of θ=30 days is highlighted in red. To better appreciate differences in errors for values of θ>20 days, insets show the same graph with a zoom on the y-axis.

Figure 2.7. D*istribution of consolidation lags for different combinations of the parameters θ and N.* a) 50% completeness. B) 70% completeness. C) 90% completeness.

Figure 2.8. *Distribution of the absolute errors between the reference reproduction number and the net and nowcasted reproduction numbers, respectively, for different combinations of parameters θ and N and for selected levels of completeness.* a) 50% completeness. B) 70% completeness. C) 90% completeness.

Figure 2.9. *Comparison of performance between the net and nowcasted estimates using alternative error functions.* a) Bias error; positive values represent underestimations of the reproduction number. b) Percentage error. c) Error with tolerance on the statistical variability of the reference estimate.

Figure 2.10. *Performance of the net and nowcasted reproduction numbers when considering the stochastic variability of the reference estimate, for different data completeness levels.*

Figure 2.11. *Example of alternative definitions for the data consolidation lag.* Blue and red dots represent the consolidation distributions associated to reporting dates July 15, 2020, and July 15, 2021, respectively. Triangles on the x-axis represent the estimated data consolidation lags according to the two definitions. In the case of a monotonic curve (Δ_2 = July 15, 2021), the two definitions are equivalent, while when the curves are non-monotonic (Δ_1 = July 15, 2020), the two definitions may give different results.

Figure 2.12. *Impact of alternative definitions for the data consolidation lag.* The figure shows the data consolidation lag distributions for different levels of data completeness, according to the baseline and alternative definition.

Figure 3.1. *Example of forecasts for a given outbreak and time of analysis.* Top row: constant incidence model (M0, baseline, light blue); upper-middle row: renewal equation model (M1, dark blue); lower-middle row: renewal equation model with isolation of cases (M2, lilac); bottom row: renewal equation model with isolation of cases and overdispersion in the offspring distribution (M3, purple). Left column: cases by symptom onset time; right column: cases by diagnosis date. Orange points: data available by the time of analysis; dark grey points: data available at the end of the outbreak; solid lines: median model predictions; shaded areas: 50% (darker color) and 95% (lighter color) prediction interval (PrI). The gold shaded rectangle identifies a 24-hour time window before the time of analysis.

Figure 3.2. *Forecasting scores across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; center column: RPS score; right column: 95% coverage.

Figure 3.3. *Model performance against the baseline across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. The 100% stacked barcharts show the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (yellow).

Figure 3.4. *Standardized rank of forecasting models across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by diagnosis date; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. Points: mean standardized rank; whiskers: 95% CI of the mean standardized rank.

Figure 3.5. *Forecasting scores across all outbreaks and times of analysis, disaggregated by forecasting horizon.* Top row: forecasts by time of symptom onset; bottom row: forecasts by diagnosis date. Left column: logS score; center column: RPS score; right column: 95% coverage. Points: median of the distribution; vertical bars: interquartile range (IQR).

Figure 3.6. *Model performance against the baseline across all outbreaks and times of analysis, disaggregated by forecast horizon.* Top row: forecasts by time of symptom onset; bottom row: forecasts by date of diagnosis. Left column: logS score; right column: RPS score. The graph shows the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (darker shaded areas in between).

Figure 3.7. *Workflow diagram for the forecasting methodology.* Blue boxes represent available data in the form of line lists; purple boxes represent one- or two-dimensional time-series (by time of symptom onset t, by time of diagnosis τ or by date of diagnosis d; gray boxes represent total number of cases; green boxes represent parameters estimated within this workflow; light blue boxes represent operations on inputs; the larger square with gray boundaries and light blue background represents the iterative relationship between stochastic forecasting through models M1-M3 and selection through particle filtering. Variables that are inputs for operations represent parameters, which were either obtained from the literature (printed in black) or within the workflow (printed in green); yellow boxes represent the scores for the forecasting endpoints which are used to evaluate the performance of the model.

Figure 3.8. *Distributions of the diagnostic delay by cruise ships.* x-axis: time of symptoms onset by day, y-axis: distribution of the diagnostic delay, aggregating cases by time steps of 4 hours. Horizontal black line: mean; light blue rectangles: interquartile range; whiskers: 95% percentile interval of the distribution. First row: cruise ships A, B, C and D, second row: cruises ship E, F, G and aggregation across all cruise ships.

Figure 3.9. *Estimated distributions of diagnostic delay, aggregated by day of symptom onset and across all cruises.* First row: symptoms onset on day of travel 1, 2, 3 and 4, second row: symptoms onset on day of travel 5, 6, 7, and pooled distribution aggregating cases with symptom onset on days 1-5.

Figure 3.10. *Cumulative distribution of the diagnostic delay, π*

Figure 3.11. *Effect of isolation in cabin.* Left: boxplots of the posterior distributions for the transmissibility reduction provided by cabin isolation α estimated in the different cruises. Horizontal black lines: mean; light blue rectangles: 50% credible intervals, CrI; light green rectangles: 95% CrI. Right: pooled probability distribution of α across all cruises.

Figure 3.12. *Forecasting scores across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; center column: RPS score; right column: 95% coverage.

Figure 3.13. *Model performance against the baseline across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. The 100% stacked barcharts show the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (yellow).

Figure 3.14. *Standardized rank of forecasting models across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by diagnosis date; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. Points: mean standardized rank; whiskers: 95% CI of the mean standardized rank.

Figure 3.15. *Forecasting scores across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; center column: RPS score; right column: 95% coverage.

Figure 3.16. *Model performance against the baseline across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. The 100% stacked barcharts show the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (yellow).

Figure 3.17. *Standardized rank of forecasting models across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by diagnosis date; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. Points: mean standardized rank; whiskers: 95% CI of the mean standardized rank.

Figure 4.1. *Chlamydia Transmission Model.* a) Compartmental scheme. S: susceptible; P: exposed infectious; IA: asymptomatic infectious; IS: symptomatic infectious; R: temporarily immune. Superscript P indicates the population of PrEP users; subscript l represents the SAL. PrEP enrolments and dropouts are shown in dashed colored arrows. State transitions due to PrEP screening are shown in green arrows. Parameter descriptions and values are reported in Table 4.1. b) Distribution of SALs in the general MSM population [24] and in the PrEP population (from partial cohort data). Bars: mean value over time; vertical lines: 95 percentile of the temporal variation. c) Yearly number of new PrEP enrolments; data for 2018-2024 from the full cohort. Color shades indicate SALs. d) Modelled yearly proportion of MSM population actively enrolled in the PrEP program (%).

Figure 4.2. *Model calibration to partial cohort data from the PrEP program at the Verona University Hospital.* a) number of diagnosed CT cases per year; b) distribution of diagnosed cases by SAL; c) distribution of diagnosed cases by symptoms. Boxplots represent the model-estimated mean (horizontal line), 50% CrI (boxes) and 95%CrI (whiskers).

Figure 4.3. *Estimated incidence of CT infections in the MSM population (both in PrEP and not in PrEP) and among PrEP users only, assuming the maintenance until the end of 2035 of the current PrEP screening protocol (scenario 6m) introduced at the start of 2018.* a) Temporal evolution. Colored vertical bands denote periods at which the endpoints are evaluated: just before PrEP introduction, i.e. in the last quarter (Q4) of 2017 (purple); before potential changes in the screening protocol, Q4 2025 (blue); and at the end of the projection period, Q4 2035 (red). Solid lines represent mean values, shaded areas represent 95% CrI. b) Quarterly average of estimates in correspondence of the key periods. c) Average prevalence of temporary immune individuals in the high-risk SAL. For panels b) and c), boxplots represent the model-estimated mean (horizontal line), 50% PrI (boxes) and 95%PrI (whiskers).

Figure 4.4. *Impact of alternative scenarios for CT control.* a) CT incidence; b) prevalence of symptomatic infections; c) CT-specific antibiotic consumption. Boxplots represent the model-estimated mean (horizontal line), 50% PrI (boxes) and 95%PrI (whiskers). Horizontal dashed lines represent mean estimates for 2035 under the current protocol (scenario 6m), reported as a reference for comparisons with past values and with alternative scenarios.

Figure 4.5. *Estimated effective reproduction number Re for CT in the province of Verona under alternative scenarios.* a) Temporal evolution; lines represent mean values; b) distributions after potential changes in CT interventions (2026) and at the end of the study period (end of 2035). Boxplots represent the model-estimated mean (horizontal line), 50% CrI (boxes) and 95%CrI (whiskers). In both panels, black horizontal lines represent the epidemic threshold of 1.

Figure 4.6. *Probability of condom use.* The brown line represents the probability of condom use during a sexual intercourse in the cohort of MSM that are not enrolled in the PrEP program, green bars represent the observed probability of condom use in the cohort of PrEP users for each year, computed as the average reported in questionnaires from the partial cohort.

Figure 4.7. *Estimates of the calibrated parameters.* The top row shows the estimated parameter distributions, the bottom row the accepted values during the MCMC algorithm. Lighter colors represent the burn-in phase, while darker colors indicate the selected parameter estimates. Left, in blue: basic reproduction number; middle, in green: assortativity parameter; right, in red: probability of remaining asymptomatic.

Figure 4.8. *Flowchart of scenario doxyPEP-R.* Compartmental model schematic representing the dynamics of infection and treatment stratified by PrEP status and SAL (l). The model includes two main blocks: the PrEP group (upper panel, green background) and the No PrEP group (lower panel, beige background). Each compartment denotes a different epidemiological state: susceptible (S), exposed (P), asymptomatic infected (I^A), symptomatic infected (I^S), and immunized (R), with superscripts and subscripts indicating group (l) and PrEP status (P for PrEP). Compartments denoted with a bar (e.g., $\bar{S}$) contain high SAL individuals who have been previously diagnosed with a CT infection.

LIST OF TABLES

INTRODUCTION

Infectious diseases have long been, are, and will likely remain recurring challenges for public health. Substantial progress, through improved hygiene, vaccination, and expanded access to care, has reduced the impact of many infections; at the same time, the risk landscape continues to evolve with changing mobility patterns, rapid urbanization, environmental pressures, antimicrobial resistance, and the recurrent emergence of novel pathogens. In such a setting, timely, data-informed, and forward-looking assessments are essential.

Mathematical modeling offers a quantitative framework to meet this need. First, by integrating surveillance streams with epidemiological knowledge, models provide an understanding of the current epidemiological situation, clarifying what is known and with what uncertainty. Second, they forecast potential outbreaks and burdens, providing short and medium projections that inform timely operational decisions. Third, they offer the possibility to simulate counterfactual scenarios of transmission dynamics under explicit assumptions, allowing exploration of behavioral, biological, and policy changes and their expected consequences, supporting strategy evaluation. Taken together, these capabilities help translate data into coherent insights and evidence-based public health decision-making.

BACKGROUND

The use of mathematical modelling in epidemiology dates back to well before the 20th century. A notable early contribution came from Daniel Bernoulli in 1760, who developed a model to estimate the benefits of smallpox inoculation on life expectancy [1]. In the early 1900s, Ronald Ross introduced key theoretical insights through his work on malaria, laying the foundations for quantitative approaches to disease transmission and the concept of a threshold for epidemic control [2]. These early efforts culminated in the work of William Kermack and Anderson McKendrick, who in 1927 formulated the first compartmental model, providing a systematic framework to describe the temporal dynamics of infectious disease spread [2].

Building on these foundations, the field of infectious disease modelling has progressively evolved, shifting from simplified theoretical constructs to increasingly realistic and application-oriented frameworks [3], [4]. This evolution has been driven not only by growing interest in infectious diseases, but also by a broader recognition of the analytical and predictive power of mathematical models [6], and by major advances in computational capacity, the expansion of epidemiological data, and continuous methodological innovation [6]-[12].

Today, a wide range of modelling approaches are used to study infectious diseases, each offering specific advantages depending on the context and research question. Among these, compartmental models remain a cornerstone of mathematical modelling in infectious disease epidemiology [12]. They divide the population into distinct categories based on disease status and describe the transitions between compartments over time using mathematical rules. The baseline example is the susceptible-infected-removed (SIR) model, which captures transmission with lifelong immunity and has inspired numerous extensions [13]. One common variant is the susceptible-exposed-infected-removed (SEIR) model, which introduces an exposed compartment to represent individuals who are infected but not yet infectious, thereby capturing the latency period typical of many diseases. Another important extension is the susceptible-exposed-infected-susceptible (SIRS) model, which incorporates the possibility of waning immunity by allowing recovered individuals to return to the susceptible compartment. Further refinements of compartmental models can account for additional complexity, such as age structure, vaccination

strategies, hospitalization pathways, or the coexistence of multiple pathogen strains [15], [16]. These adaptations account for additional complexity, but on the other hand they enhance the models' realism and expand their applicability to a wide range of epidemiological situations.

The deterministic implementation of compartmental models typically relies on systems of ordinary differential equations to describe the rate of change in each compartment. These models produce a single epidemic trajectory based on fixed parameters and initial conditions, under the assumption that the population's dimension is sufficiently large to be approximated with a continuous variable [17]. Deterministic models are computationally efficient and analytically more tractable. However, they may not fully capture the variability observed in real epidemics, such as in the early stages of an outbreak or in small populations, where the chance events can determine whether a pathogen successfully establishes or goes extinct [18], [19]. To address this limitation, stochastic versions of deterministic compartmental models have been developed [20].

Stochastic models, for a given set of parameters, can generate a distribution of possible epidemic outcomes and therefore they can reflect the inherent randomness of infectious diseases transmission.

More recently, agent-based models, which are inherently stochastic, have gained prominence in infectious disease modelling [21]–[23]. These approaches simulate the behavior, characteristics, and interactions of individual agents within a population, allowing for a highly detailed and granular representation of heterogeneity. By modelling each individual separately, agent-based models can capture complex patterns of contact, movement, and behavior that influence disease transmission, including variations in susceptibility, compliance with interventions, and social dynamics. This level of detail is especially valuable for understanding outbreaks in structured populations or when individual heterogeneities significantly affect epidemic outcomes. A closely related class are metapopulation models, which couple compartmental systems across subpopulations via a contact or mobility matrix [24].

In parallel, network-based models have emerged as powerful tools to represent the underlying structure of networks. These models explicitly describe how individuals are connected through various types of interactions,

capturing the heterogeneity, and clustering typical of real-world networks [25]–[27]. Network modelling is particularly relevant for diseases transmitted through close, repeated, or preferential contacts, such as sexually transmitted infections, respiratory diseases in households or workplaces, and infections spreading through social communities [28]. By accounting for the network topology, these models help identify key nodes or groups that drive transmission, optimize targeted interventions, and explore how changes in connectivity impact epidemic dynamics. Agent-based and network models complement population-level approaches by providing more granular insights into transmission dynamics that are often inaccessible through aggregate models.

The wide range of mathematical modelling approaches helps align methods with specific questions and data constraints, thereby strengthening their role in public health decision-making. Relevant applications where mathematical models play a central role are in the real-time monitoring, decision-oriented forecasting, and scenario evaluation to guide control interventions [29]–[31]. In routine disease surveillance, models are invaluable for interpreting incomplete, delayed, or biased data, estimating the degree of underreporting, and detecting early warning signals of disease resurgence. They facilitate continuous monitoring of real-time transmission trends and provide a framework for evaluating the effectiveness of ongoing interventions, allowing timely adjustments to public health responses [32]-[34]. Another critical application of modelling is forecasting across multiple horizons [35], [36]. Short-term forecasts (typically until a few days to a few weeks in the future), are often the most operationally necessary and the most feasible to generate since they rely on recent observations and simpler assumptions that limit structural uncertainty. They estimate the near-future course of transmission and burden with quantified uncertainty to support time-sensitive operational decisions and early warning systems. Furthermore, through detailed scenario analyses, models enable public health officials to explore "what-if" questions and counterfactual scenarios, offering insights into the possible outcomes and trade-offs of alternative strategies before they are implemented [37], [38].

RESEARCH QUESTIONS AND INNOVATIVE CONTRIBUTIONS

The relevance of mathematical modelling in infectious disease epidemiology has become increasingly evident in recent years. Their importance was particularly underscored during the COVID-19 pandemic, which highlighted both their potential and their limitations in guiding real-time public health responses [39], [40]. Ultimately, the overarching goal of mathematical modelling is to develop approaches that are not only methodologically robust, but also responsive to the practical needs and constraints of real-world public health practice, both in the context of routine surveillance and in emergency situations. This thesis focuses on three key areas in which mathematical modelling efforts play a central role: understanding, anticipating, and controlling the spread of infectious diseases. Each of these research domains involves specific methodological challenges related to model design, parameter estimation, validation, and interpretation.

In this thesis, I present a collection of modelling studies developed during my PhD that address different epidemiological questions through the application of mathematical and statistical tools. The work spans a variety of contexts, from real-time monitoring of emerging infectious diseases to scenario analysis supporting public health decision-making, passing through the evaluation of short-term forecasting models. Each study reflects not only the scientific challenges posed by specific epidemiological contexts but also the methodological adaptations required to address them.

NOWCASTING SURVEILLANCE DATA

Understanding the spread of an infectious disease is necessary to interpret surveillance data and to quantify transmission in real time. Epidemiological surveillance often suffers from reporting delays: the most recent notifications could arrive late, so recent data could be underestimated [41], [42], [43]. This undercount limits our ability to assess the current trajectory and distorts real-time estimates of key quantities such as the reproduction number [44]. This phenomenon is known as right truncation of epidemic curves [45]. To address this issue nowcasting methods have been developed but not yet formally integrated into surveillance-based modelling frameworks to support the estimation of key parameters [46]. Nowcasting algorithms are techniques that

use estimates of reporting delays to reconstruct the still-incomplete recent part of the epidemic curve and produce timely, bias-corrected indicators [47]. The first research question of this study investigates the practical advantages of incorporating a nowcasting algorithm to produce real-time estimates of the reproduction number.

I developed a flexible nowcasting framework to estimate the time-varying reproduction number from incomplete epidemic curves. The method uses recurrent Bayesian updating and explicitly accounts for uncertainty from reporting delays. I applied it to COVID-19 data from Italy and evaluated it retrospectively: using finalized datasets to mimic how the pipeline would have performed in real time during the epidemic. The work illustrates that deploying nowcasting in real time to correct recent underreporting, improves the estimation of the reproduction number for control and surveillance. Therefore, it results in offering timely, reliable support to policymakers during critical phases of a pandemic.
The first study resulted in the publication: Bizzotto A, Guzzetta G, Marziano V, et al. Increasing situational awareness through nowcasting of the reproduction number. Front Public Health. 2024 [48].

FORECASTING EPIDEMIC TRENDS

Short- and medium-term epidemic forecasts, which aim to anticipate future trends in the epidemic curve is a highly active and timely research area. Forecasting is critical for healthcare planning and risk communication, but it is subject to uncertainty and structural limitations [49], [50]. The second research question focuses on methodological approaches used for forecasting and on how to assess their reliability and usefulness in practice.
I conducted a systematic evaluation and comparison of multiple short-term forecasting models for norovirus cases to support outbreak management on board cruise ships. I calibrated the models with real outbreak data from multiple ships and evaluated performance using retrospective forecasts that reproduced real-time conditions. I then assessed reliability and uncertainty quantification to identify strengths, limitations, and conditions of robust performance, thereby enabling informed decisions that account for uncertainty. This work served as an applied case study to test forecasting strategies in a real-world operational context [51]-[53]. The resulting pipeline provides forecasts with uncertainty

giving timely, actionable guidance to public health officers operating on cruise ships.
The second work resulted in the publication: Bizzotto A, De Bellis A, Marziano V, et al. Forecasting norovirus cases on cruise ships to support outbreak management on board. Travel Med Infect Dis. 2025 [54].

IMPACT EVALUATION OF INTERVENTION SCENARIOS

Controlling the spread of an infectious disease raises the need for scenario exploration to evaluate alternative policy pathways, helping compare options and guide action under uncertainty [55], [56]. This counterfactual analysis does not aim to predict a single future; rather, it maps plausible outcomes across assumptions, helping epidemiologists and policymakers align actions with goals and constraints while maintaining a clear link between evidence, modelling assumptions, and decisions. The third research question thus investigates how counterfactual scenarios and intervention strategies can be modelled and compared maintaining a link between evidence, assumptions, and decisions.
In this research question I focused on exploring plausible long-term epidemiological outcomes under different intervention scenarios comparing intervention by simulating alternative policy choices [57], [58]. I developed a transmission model of Chlamydia trachomatis calibrated on data from men who have sex with men (MSM) in a Northern Italian province, where a subset of the population is engaged in a pre-exposure prophylaxis (PrEP) program for HIV. The model was designed to evaluate how changes in screening strategies could affect transmission dynamics [59]-[61]. This scenario-based analysis allowed to address "what-if" questions in a structured way, offering insights into the potential impact of public health interventions under uncertainty. The key contribution is a decision-focused workflow that links evidence, assumptions, and choices to rank strategies by benefits and trade-offs. An additional aspect is that I addressed scientific questions raised by clinicians through the application of our model and our mathematical expertise.
This article is currently in preparation.

Across these works, a common theme is the adaptation of modelling techniques to the nature of the available data and the specific public health questions at hand.

The results of these studies provide concrete examples of how mathematical modelling can contribute to infectious disease surveillance, preparedness, and response. They also highlight the importance of flexibility in the modelling approach, both in terms of mathematical structure and computational implementation, to reflect the complexity of real-world epidemiological systems.

FURTHER CONTRIBUTIONS

During my Ph.D., I also had the opportunity to contribute to three additional projects beyond those presented in the main body of this thesis.

The first supplementary contribution was developed within the Influcast forecasting hub, a collaborative initiative aiming to improve real-time forecasts of influenza-like illness (ILI) incidence in Italy. The project brought together several modeling teams using different approaches to forecast short-term national ILI trends. My contribution focused on the development of a model based on the renewal equation, calibrated on weekly surveillance data, to generate prospective forecasts of ILI incidence. This work was published as part of a multi-model comparison study that assessed the strengths and limitations of the various forecasting approaches employed: Fiandrino S, Bizzotto A, Guzzetta G, et al. Collaborative forecasting of influenza-like illness in Italy: The Influcast experience. Epidemics. 2025 [62].

The second additional work is related to the norovirus modeling work previously discussed. In this context, I developed a branching process model to simulate different outbreak scenarios aboard cruise ships. The aim was to quantify the impact of reporting delays and diagnostic timing on the effectiveness of isolation policies during norovirus outbreaks. The work resulted on the publication: De Bellis A, Bizzotto A, Anagnostopoulou L, et al. Mitigating norovirus spread on cruise ships: A model-based assessment of diagnostic timing and isolation. J Travel Med. 2025 [63].

In the third further project, we analyze a multitype epidemic model with a next-generation matrix (NGM) and type-specific community fractions from the theoretical point of view, focusing on two quantities: the basic reproduction number R_0 and the final epidemic size by type. We examine two NGM settings: the general case and the case derived from a contact matrix. When only partial NGM information is available (row and/or column sums), we derive sharp upper

and lower bounds for R_0 and final size in the general case. When the NGM is derived from a contact matrix, the problem is harder: our bounds for R_0 are tighter than in the general case but not sharp, and final-size bounds are obtained only for two types. My contribution includes providing most of the formal proofs of these bounds; the manuscript is ongoing.

STRUCTURE OF THE THESIS

- Chapter 1 describes the mainly mathematical framework and methodologies used in this thesis.

- Chapter 2 enhances real-time situational awareness by nowcasting the time-varying reproduction number from surveillance data, supporting timely public-health decisions.

- Chapter 3 develops and evaluates short-term forecasts of norovirus cases on cruise ships to support on-board outbreak operational management.

- Chapter 4 presents a mathematical model of Chlamydia trachomatis transmission among MSM in PrEP programs to evaluate how changes in screening strategies could affect transmission dynamics.

- Chapter 5 summarizes the main conclusions of the thesis and puts them in perspective with respect to the scientific collaborations they were developed in and with future work that may stem from these results.

1. THE MATHEMATICAL FRAMEWORK AND METHODOLOGIES

In this section, I outline the main mathematical methods used throughout the thesis; their problem-specific applications are then detailed in the respective chapters. The aim is to provide a common mathematical background. This will make it easier to understand and interpret the modelling choices, assumptions, and results in the chapters that follow.

1.1. THE REPRODUCTION NUMBER

The reproduction number is a central focus across all the studies. Broadly speaking, the reproduction number quantifies the average number of secondary infections generated by an infectious individual at a given time, accounting for the current state of the population, including levels of immunity and any public health interventions in place [6]. Because epidemics spread in transmission chain, the infectiousness profile of the individuals has a direct impact on the reproduction number [64], [65]. This profile is driven by time periods that include:

- *latent period*, time from infection to the onset of infectiousness;
- *infectious period*, the length of time an infected person remains capable of transmitting the pathogen to others;
- *incubation period*, time from infection to symptom onset;
- *symptomatic period*, time from symptom onset to symptom resolution.

Other important time intervals commonly used when inferring the reproduction number are:

- *generation time*, delays from infection of an infector to infection of their infectee;

- *serial interval*, the delay from symptom onset in an infector to symptom onset in the infectee.

See Figure 1.1 for a schematic overview.

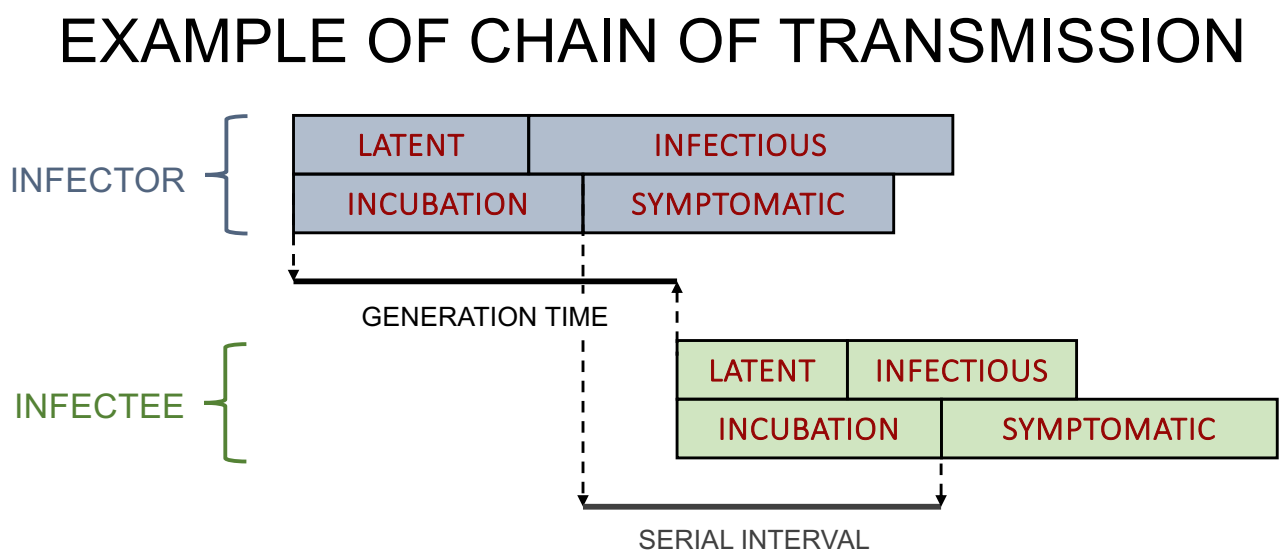


Figure 1.1. *Example of chain of transmission.* Timelines for infector (top) and infectee (bottom) showing latent period followed by the beginning of the infectivity and the incubation period followed by the time of symptoms' onset. Arrows mark generation time and serial interval.

The reproduction number is fundamentally important for surveillance: values above the critical threshold of 1 indicate epidemic growth of the incidence; values below 1 indicate epidemic decline of the incidence; values near 1 mark a critical regime in which small shifts in behavior, immunity, or policy may shift trends in either direction.
In the literature, various terms are used to refer to this concept, depending on slight differences in assumptions. When calculated at the beginning of an epidemic, under the assumption that the entire population is fully susceptible and no control measures (such as isolation, mask use, or contact reduction) are in place, it is referred to as the basic reproduction number R_0 [4], [13]. The basic reproduction number is generally not directly observable in practice; it must be inferred with a model. For instance, in the deterministic SIR model mean R_0 is given by the product between the infectivity rate β and the mean infectious period $1/\gamma$

$$R_0 = \beta \cdot \frac{1}{\gamma}.$$

Different modelling choices, both in structure and parameter values, can yield different formulations and estimates. Moreover, R_0 reflects not only pathogen biology but also individuals' characteristics and contact mixing patterns, so it can vary across settings, regions, and populations [7], [25]. One common way to compute it is via the next-generation matrix approach, see Section 1.3.

Complementarily, when the reproduction number is computed at a specific time t, reflecting the real-time context of the epidemic (e.g. current immunity levels, interventions, and behavioral changes), it is typically called the time-varying reproduction number $R(t)$ [4], [13]. The time-varying reproduction number is likewise treated as a model-based quantity: it is rarely directly observed and must be inferred under explicit assumptions. In the deterministic SIR model, the analytic formula for the time-varying reproduction number is:

$$R(t) = R_0 \cdot s(t),$$

where $s(t)$ is the proportion of susceptible individuals at time t.
Thus, a simulation of this SIR model with both R_0 and $1/\gamma$ fixed, produces a specific trajectory for $R(t)$. See Figure 1.2.

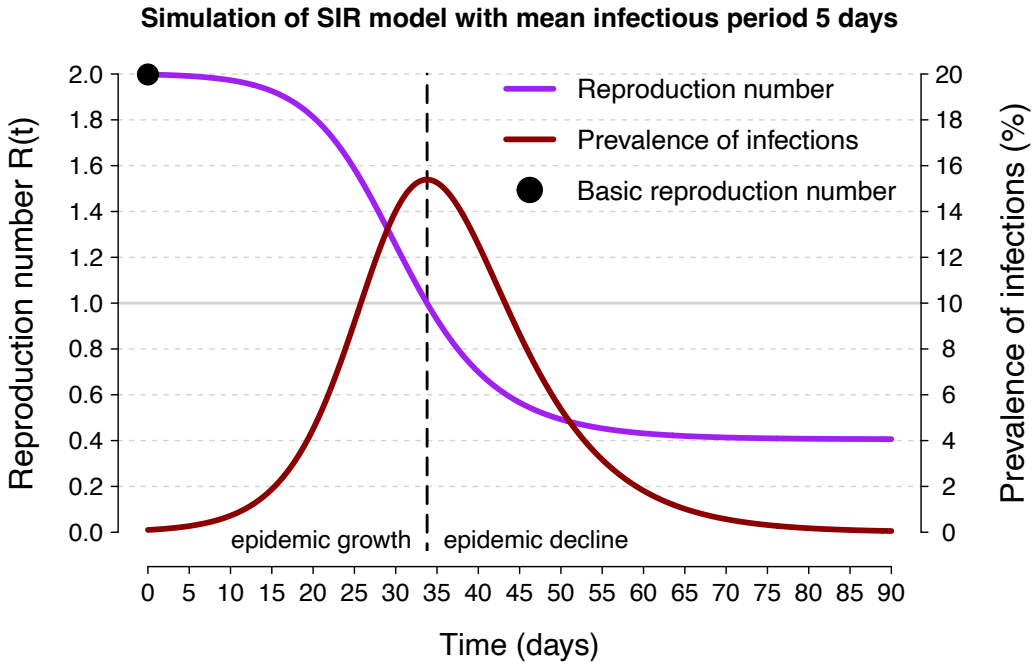


Figure 1.2. *Deterministic simulation of SIR model.* In purple the trajectory of the time-varying reproduction number with respect to left y-axis and in red the prevalence of infectious individuals in percentage, with respect right y-axis. The grey horizontal line marks the threshold $R = 1$: above it the epidemic grows; below it the epidemic declines toward extinction. The x-axis shows time in days.

A more specific form of the time-varying reproduction number is the instantaneous reproduction number, proposed by Cori et al. [29] and denoted $R_i(t)$. It measures transmissibility at time t under the assumption that the conditions at time t (contact patterns, interventions, behavior) remain constant over an individual's entire infectious period. It is a real-time, forward-looking measure: it reflects how many secondary cases an average infected individual would generate if they were infected at time t and experienced their entire infectious period under the conditions prevailing at that time. Operationally, it can be estimated via a renewal model, see next section:

$$I_t = R_i(t) \int_0^{+\infty} I_{t-s}\, \omega_s\, ds$$

where I_t is the incidence at time t and $\omega_.$ is a distribution that weights how past cases contribute to infectiousness. In Cori's framework, ω ideally denotes the generation-time distribution. Because infection times are rarely observed, analysts typically use the serial-interval distribution as the closest observable proxy. When the incubation period is similar across cases and symptom onset aligns with infectiousness, the serial interval can approximate the generation time reasonably well [43], [64]-[66]. Cori's approach is well suited to real-time estimation because it relies only on past incidence data and an assumed generation-time (often serial-interval) distribution, without requiring forecasts [29].

The time-varying reproduction number is used in Chapters 2,3,4, and most prominently in Chapter 2, where it is central to the analysis. In Chapter 4 it is also estimated the basic reproduction number R_0.

1.2. THE RENEWAL EQUATION

In the previous section I introduced the renewal equation for estimating the instantaneous reproduction number. In practice, common implementations of the renewal equation [34] stabilize estimates quantifying the uncertainty within a Bayesian framework, especially at low incidence or between epidemic waves where noise can dominate signal [67]. In discrete time (e.g., daily or weekly counts), the renewal model is

$$I_{t+1} = \mathbb{E}_p \left[R_i(t) \cdot \sum_{s=0}^{t} I_{t-s} \cdot \omega_s \right].$$

Whereas it explicitly describes the number of new infections I_{t+1} at time $t+1$ as a function of past incidence I_{t-s} and the probability mass function of the generation time ω_s. Formally, it assumes that the expected number of new cases at time t, with respect to a probability distribution p, is given by the product of the instantaneous reproduction number $R_i(t)$ and the weighted sum of past incidences, where the weights correspond to the generation interval distribution [64]. Because the infections are represented by case counts, the typical choice for the probability distribution p is either the Poisson distribution or the Negative Binomial distribution, the latter being preferred when accounting for heterogeneity in the number of secondary infections [43]. In Bayesian estimation, prior probabilities on $R_i(t)$ are combined with the likelihood of observing the incidence data to obtain a posterior for $R_i(t)$. Inference can be done via MCMC (see Section 1.4), yielding credible intervals that quantify uncertainty from observation noise and other uncertainties [29], [68].

The same renewal equation can also be used for incidence forecasting once a path for $R_i(t)$ is specified [35]. A one-step-ahead prediction can be obtained by plugging the current estimate of $R_i(t)$ into the renewal equation, and a multi-step forecast is produced by iterating this update, replacing unknown future incidences with their predicted values at each step:

$$I_{t+1} \sim p \left[R_i(t) \cdot \sum_{s=0}^{t} I_{t-s} \cdot \omega_s \right], \qquad I_{t+2} \sim p \left[R_i(t) \cdot \sum_{s=0}^{t+1} I_{t-s} \cdot \omega_s \right], \qquad \dots$$

I applied the renewal equation in Chapters 2 and 3 to estimate the time-varying reproduction number. In Chapter 3 it further serves as the core tool for forecasting infectious cases.

1.3. THE NEXT GENERATION METHOD

As anticipated in Section 1.1, one common way to compute the basic reproduction number is via the Next Generation Matrix approach proposed by Diekmann [7], [69]. The NGM method linearizes the infection subsystem around a state of interest. Linearizing at the disease-free state (i.e. no infectious

individuals in the population) gives the basic reproduction number R_0, linearizing at a different state reflecting immunity and interventions gives the time-varying reproduction number $R(t)$ [70]. The NGM method is describes below.

Let $y = y(t) = [y_1(t), y_2(t), \dots, y_n(t)]$ the vector that describes the state of all compartments $y_i(t)$ at time t. The evolution of the epidemiological dynamics is described by the ODEs

$$y_i' = g_i\big(y(t)\big) \quad i = 1, \dots, n.$$

Where $g(\cdot)$ is a deterministic function. Without loss of generality order the compartments so that the first k are the "infected" (i.e., carrying the pathogen) writing:

$$\hat{y} = \hat{y}(t) = [y_1(t), y_2(t), \dots, y_k(t)].$$

On this infected subsystem, decompose the vector field as

$$\hat{y}_i' = g_i(y) = f_i(y) - v_i(t) \qquad i = 1, \dots, k \leq n,$$

where:

- f collects new infection flows, the entries into the infected compartments caused by transmission,

- v collects all other flows affecting the infected compartments.

This decomposition is not unique [7], [70].

Let t_x the time and x be the state, in which we linearize. Define the Jacobian matrices (i.e. first order approximation of the vector field) $F = \left[F_{ij}\right]_{i,j}$ and $V = \left[V_{ij}\right]_{i,j}$, with respect to the infected variables and evaluate them at x:

$$F_{ij} = \frac{\partial}{\partial \hat{y}_j} f_i(y)_{|y=x} \qquad V_{ij} = \frac{\partial}{\partial \hat{y}_j} v_i(y)_{|y=x} \qquad i, j = 1, \dots, k$$

where:

- F_{ij} represents the instantaneous per-capita rate at which one individual in infected class j generates new infections entering class i when the system is fixed at the linearization state x.
- V_{ij} is the per-capita linear effect of an individual in infected class j on the net outflow rate from class i at state x.

The next generation matrix K is defined as the matrix product between F and the inverse of V:

$$K = F \cdot V^{-1}, \qquad \text{then} \qquad K_{ij} = \sum_{h=1}^{k} F_{ih} \cdot V_{hj}^{-1},$$

where:

- V_{hj}^{-1} is the expected total time an individual who starts in infected class j will spend in infected class h over their entire infected lifetime, with the environment held at x.

- K_{ij}, is the expected number of new infections entering class i that are caused, over the entire infected lifetime, by one individual who starts in infected class j.

In K_{ij}, the sum over h accounts for the time the index case spends in each infected stage h before removal; while in h, they create new infections into i at rate F_{ih} , for an expected duration V_{hj}^{-1}.

The reproduction number is defined as the spectral radius (i.e. dominant eigenvalue) of K

$$R(t_x) = \rho(K)$$

Note that the spectral radius of K is equivalent to the one of its transpose matrix. In this case, the roles of i and j swap in the interpretation of K_{ij}: it becomes the expected number of new infections entering class j that are caused, over the entire infected lifetime, by one individual who belongs in infected class i. Further details are in the classic references [7], [69], [71].
Overall, the NGM method allows quantifying the pathogen's transmission potential under specific population conditions by capturing how infections spread through different compartments or groups within the population. By incorporating factors such as immunity levels, behavioral changes, or public health interventions, the Next Generation Method provides a flexible framework to evaluate how these elements influence transmission dynamics. As a result, it estimates not only the initial outbreak potential but can also track changes in transmissibility over time or in response to control measures, offering critical insights for epidemic management and policy decisions.

In Chapter 4, I apply the NGM in two ways: to estimate R_0 at the disease-free equilibrium, and to estimate the time-varying $R(t)$ at states reflecting different epidemiological conditions.

1.4. BAYESIAN ESTIMATION OF PARAMETERS

Bayesian parameter estimation enables the explicit incorporation of uncertainty in a rigorous mathematical framework. The Bayesian approach is based on the eponymous Bayes' theorem [72]

$$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)}$$

where:

- θ is the parameter (or vector of parameters);
- D are the observed data;
- $p(\theta|D)$ is the posterior distribution (i.e. updated beliefs about θ after observing D);
- $p(\theta)$ is the prior distribution (it encodes beliefs/knowledge about parameters before seeing the data);
- $p(D|\theta)$ is the likelihood function (how probable the observed data are under a given parameter value; links the model to the data);
- $p(D) = \int p(D|\theta)\, p(\theta)\, d\theta$ is the evidence (normalizing constant).

Knowing the right-hand side of the Bayes' theorem one can infer parameter values that best reconcile the model with the data. However, in most realistic models, including ours, the posterior distribution is rarely available in closed form because the evidence is intractable, for instance due to the high dimensionality of the θ space. A classic approach is to approximate numerically the posterior distribution with mathematical methods called Markov Chain Monte Carlo (MCMC) methods [73]. The MCMC are a class of algorithms that use Monte Carlo approach, designed to generate samples from complex probability distributions by constructing a Markov Chain. At each step k, the

algorithm proposes a new value $\theta^{\star}$ from a proposal rule $q\left(\cdot \left| \theta^{(k-1)} \right.\right)$ which depends on the current sample parameter $\theta^{(k-1)}$, and then applies an acceptance criterion to decide whether to move to proposed $\theta^{\star}$ or stay at the current one $\theta^{(k-1)}$ so the chain iteratively updates the sample set by accepting or rejecting proposals.
It is proved that under regularity conditions, the MCMC algorithm with infinite number of iterations, converges to the analytical posterior [74]. However, in practice, with a finite run N, the posterior distribution can only be approximated and assessed by checking appropriate diagnostics [74]. In a specific MCMC called Metropolis–Hastings (MH) [75], [76], the move is accepted with probability:

$$\alpha\left(\theta^{(k-1)}, \theta^{\star}\right) = min\left\{1, \frac{p(D|\theta^{\star}) \cdot p(\theta^{\star}) \cdot q\left(\theta^{(k-1)} \left| \theta^{\star} \right.\right)}{p\left(D \left| \theta^{(k-1)} \right.\right) \cdot p(\theta^{(k-1)}) \cdot q\left(\theta^{\star} \left| \theta^{(k-1)} \right.\right)}\right\}$$

That account for:

- the fit to the data via the likelihood function $p(D|\theta)$;
- the prior preference via $p(\theta)$;
- any asymmetry in the proposal distribution $q(\cdot \mid \cdot)$.

A pseudo-code of the Metropolis–Hastings algorithm is shown in Figure 1.3.

Metropolis-Hastings (MH)

Require: initial parameter $\theta^{(0)}$, iterations N, proposal $q(\cdot \mid \cdot)$
for $n = 1, \ldots, N$ **do**
    propose a new value $\theta^{\star} \sim q(\cdot \mid \theta^{(n-1)})$
    compute acceptance probability

$$\alpha\left(\theta^{(n-1)}, \theta^{\star}\right) = \min\left\{1, \frac{p(\mathcal{D} \mid \theta^{\star})\, p(\theta^{\star})\, q(\theta^{(n-1)} \mid \theta^{\star})}{p(\mathcal{D} \mid \theta^{(n-1)})\, p(\theta^{(n-1)})\, q(\theta^{\star} \mid \theta^{(n-1)})}\right\}$$

    sample $u \sim \mathrm{Unif}(0, 1)$
    if $u < \alpha$ **then**
        accept: $\theta^{(n)} \leftarrow \theta^{\star}$
    else
        reject: $\theta^{(n)} \leftarrow \theta^{(n-1)}$
    end if
end for
Return: $\{\theta^{(n)}\}_{n=0}^{N}$

Figure 1.3. *Metropolis-Hasting (MH) algorithm.*

In the special case when no information is available to prefer a specific region of the parameter space, is natural to assume q symmetric (e.g. random-walk Gaussian process). In this case, $q\left(\theta^{(k-1)} \middle| \theta^{\star}\right) = q\left(\theta^{\star} \middle| \theta^{(k-1)}\right)$ and the algorithm is known as Metropolis [77].
I performed Bayesian inference with Metropolis algorithm to estimate model parameters in Chapter 2 for the estimation of the time-varying reproduction number and in Chapter 3 for the estimation of cabin isolation effect and the $R(t)$.

1.5. PARTICLE FILTERING

When working with small populations, stochastic effects can dominate and are not captured well by deterministic models; a stochastic formulation is required instead. To calibrate a stochastic model to observed time series $y_{1:T}$, alternative approaches are needed, as traditional MCMC algorithms become inefficient or inappropriate. Natural formalism is a state–space or hidden Markov model [78] where the principal components are:

- *latent (unobserved) states* x_t : quantities that evolve over time but are not directly observed (e.g. true number of infectious individuals, new infections, compartment counts or random effects capturing unobserved heterogeneity and process noise);

- *observation model* $p(y_t \mid x_t, \theta)$: links states to observed data y_t (e.g., reported cases via reporting delay distributions);

- *state transition model* $p(x_t \mid x_{t-1}, \theta)$: describes the epidemiological dynamics (e.g. deterministic ODEs with noise, stochastic compartmental models, branching processes or Renewal equation with process noise).

In hidden Markov models we want the posterior distribution over parameters and latent states:

$$p(\theta, x_{1:T} \mid y_{1:T}) \propto p(x_{1:T}, y_{1:T} \mid \theta) \cdot p(\theta).$$

However, the likelihood $p(y_{1:T} \mid \theta) = \int p(x_{1:T}, y_{1:T} \mid \theta)\, dx_{1:T}$ is typically intractable [78]. Particles filtering provides an approximation by propagating a multitude of weighted particles forward in the time using the state transition model, weighting each particle by the likelihood of the new data under the observation model $p(y_t \mid x_t, \theta)$, and resampling to focus computational effort on high-weight particles. As result, it yields an estimate $\hat{p}_\theta(y_{1:T})$ of the marginal likelihood $p(y_{1:T} \mid \theta)$, obtained as the product of normalized weight averages across time [78]. See Figure 1.4 for the pseudocode of particles filtering.

Particle filtering (PF)

Input: parameter θ, number of particles M, data $y_{1:T}$
Output: likelihood estimate $\widehat{p}_\theta(y_{1:T})$
for $m = 1, \dots, M$ **do**
    sample $x_1^{(m)} \sim p(x_1 \mid \theta)$
    compute weight $\tilde{w}_1^{(m)} \leftarrow p(y_1 \mid x_1^{(m)}, \theta)$
end for
normalize weights $w_1^{(m)} \leftarrow \tilde{w}_1^{(m)} / \sum_{j=1}^{M} \tilde{w}_1^{(j)}$
set $\ell_1 \leftarrow \log\left(\frac{1}{M} \sum_{m=1}^{M} \tilde{w}_1^{(m)}\right)$
set $z_1^{(m)} \leftarrow m$ for all m
resample indices $Z_1^{(m)} \sim \text{Categorical}\left(w_1^{(1)}, \dots, w_1^{(M)}\right)$
set $x_1^{(m)} \leftarrow x_1^{(A_1^{(m)})}$
set $w_1^{(m)} \leftarrow 1/M$
for $t = 2, \dots, T$ **do**
    for $m = 1, \dots, M$ **do**
        propagate $x_t^{(m)} \sim p(x_t \mid x_{t-1}^{(m)}, \theta)$
        compute weight $\tilde{w}_t^{(m)} \leftarrow p(y_t \mid x_t^{(m)}, \theta)$
    end for
    normalize weights $w_t^{(m)} \leftarrow \tilde{w}_t^{(m)} / \sum_{j=1}^{M} \tilde{w}_t^{(j)}$
    set $\ell_t \leftarrow \log\left(\frac{1}{M} \sum_{m=1}^{M} \tilde{w}_t^{(m)}\right)$
    if $t < T$ **then**
        resample $Z_t^{(m)} \sim \text{Categorical}\left(w_t^{(1)}, \dots, w_t^{(M)}\right)$
        set $z_t^{(m)} \leftarrow Z_t^{(m)}$
        set $x_t^{(m)} \leftarrow x_t^{(Z_t^{(m)})}$
        set $w_t^{(m)} \leftarrow 1/M$
    end if
end for
set $\widehat{p}_\theta(y_{1:T}) \leftarrow e^{\sum_{t=1}^{T} \ell_t}$
Return $\widehat{p}_\theta(y_{1:T})$

Figure 1.4. *Particle filtering (PF) algorithm.*

Once an estimate of the likelihood is provided by particle filtering, Bayesian inference for model parameter calibration can proceed. A typical approach is Particle Markov Chain Monte Carlo (pMCMC), which plugs the likelihood estimator from the particle filter into a Metropolis–Hastings step [79]. At each iteration k:

- propose a new parameter $\theta^\star$;

- run a particles filter under $\theta^{\star}$ and the observed data $y_{1:T}$ to obtain $\hat{p}_{\theta^{\star}}(y_{1:T})$;

- accept $\theta^{\star}$ with probability

$$\alpha\left(\theta^{(k-1)}, \theta^{\star}\right) = min\left\{1, \frac{\hat{p}_{\theta^{\star}}(y_{1:T}) \cdot p(\theta^{\star}) \cdot q\left(\theta^{(k-1)} \middle| \theta^{\star}\right)}{\hat{p}_{\theta^{(k-1)}}(y_{1:T}) \cdot p(\theta^{(k-1)}) \cdot q\left(\theta^{\star} \middle| \theta^{(k-1)}\right)}\right\}.$$

The advantage of this method is that it avoids simulating a large number of full epidemic curves, most of which would be rejected in a standard MCMC approach. Instead, computational effort is focused on the most plausible trajectories, improving efficiency, and allowing parameter estimation even in highly stochastic dynamics [80], [81]. See Figure 1.5 for the pseudocode.

Particle Markov Chain Monte Carlo (pMCMC)

Require: initial parameter $\theta^{(0)}$, proposal $q(\cdot \mid \cdot)$, iterations N, particles M, data $y_{1:T}$
run PF with $\theta^{(0)}$, M particles and $y_{1:T}$ data to obtain $\widehat{p}_{\theta^{(0)}}(y_{1:T})$
for $n = 1, 2, \ldots, N$ **do**
    propose a new value $\theta^{\star} \sim q(\cdot \mid \theta^{(n-1)})$
    run PF with $\theta^{\star}$, M particles and $y_{1:T}$ data to obtain $\widehat{p}_{\theta^{\star}}(y_{1:T})$
    compute acceptance probability

$$\alpha\left(\theta^{(n-1)}, \theta^{\star}\right) = \min\left\{1, \frac{\widehat{p}_{\theta^{\star}}(y_{1:T})\, p(\theta^{\star})\, q(\theta^{(n-1)} \mid \theta^{\star})}{\widehat{p}_{\theta^{(n-1)}}(y_{1:T})\, p(\theta^{(n-1)})\, q(\theta^{\star} \mid \theta^{(n-1)})}\right\}$$

    sample $u \sim \mathrm{Unif}(0, 1)$
    if $u < \alpha\left(\theta^{(n-1)}, \theta^{\star}\right)$ **then**
        accept: $\theta^{(n)} \leftarrow \theta^{\star}$; $\widehat{p}_{\theta^{(n)}}(y_{1:T}) \leftarrow \widehat{p}_{\theta^{\star}}(y_{1:T})$
    else
        reject: $\theta^{(n)} \leftarrow \theta^{(n-1)}$; $\widehat{p}_{\theta^{(n)}}(y_{1:T}) \leftarrow \widehat{p}_{\theta^{(n-1)}}(y_{1:T})$
    end if
end for
Return: $\{\theta^{(n)}\}_{n=0}^{N}$

Figure 1.5. *Particle Markov Chain Monte Carlo (pMCMC) algorithm.*

I used particles filtering in Chapter 3 for the selection of the most likely forecasting trajectories and pMCMC for the calibration of the model in Chapter 4.

1.6. MODEL PERFORMANCE EVALUATION METRICS

In the context of nowcasting and forecasting, performance evaluation is essential. If a ground truth (or trusted proxy) is available, we can quantify how close our predictions are to reality and we can determine if, when, and how well an algorithm or a model works [31], [82], [83]. I used different metrics divided into two complementary families, the point metrics used in the nowcasting context (Chapter 2) and the probabilistic metrics in context of forecasting (Chapter 3).
For each time $t = 1, \dots, T$ and sample $i = 1, \dots, N$:

- $y_i(t)$ is the i^{th} element of the predicted values vector $y(t)$ with mean

$$\bar{y}(t) = \frac{1}{N}\sum_{i=1}^{N} y_i(t) .$$

- $\hat{y}_i(t)$ is the i^{th} corresponding reference value with mean

$$\hat{y}(t) = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i(t) .$$

If a reference distribution is not available but only a point reference value, let

- $\tilde{y}(t)$ be the reference value at timestep t.

Moreover, $y^U(t)$ and $y^L(t)$ denote the 97.5th and 2.5th percentiles of the predicted values at time t. Respectively, $\hat{y}^U(t)$ and $\hat{y}^L(t)$ denote the 97.5th and 2.5th percentiles of the reference values.
The metrics used are:

POINT METRICS

These are metrics useful when uncertainty is small, and the central estimate is the main focus [84].

- *Absolute Error*, which measures the average magnitude of prediction errors:

$$AE(t) = |\hat{y}(t) - \bar{y}(t)|.$$

- *Bias Error*, which captures systematic under- or over-estimation:

$$BE(t) = \hat{y}(t) - \bar{y}(t).$$

- *Absolute Percentage Error*, which express the error as a percentage:

$$APE(t) = 100 \cdot \left| 1 - \frac{\bar{y}(t)}{\hat{y}(t)} \right|.$$

- *Error Within Tolerance*, which evaluates how far the prediction lies outside the acceptable range $[\hat{y}^L(t), \hat{y}^U(t)]$:

$$EWT(t) = max\{\bar{y}(t) - \hat{y}^U(t);\ \ \hat{y}^L(t) - \bar{y}(t);\ \ 0\}.$$

I used also aggregated metrics:

- *Mean Absolute Percentage error,* which summaries the APE errors

$$MAPE = \frac{100}{T} \cdot \sum_{t=1}^{T} \left| 1 - \frac{y(t)}{\hat{y}(t)} \right|.$$

- *Root Mean Squared Error*, which penalizes large errors

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} |\hat{y}(t) - y(t)|^2}.$$

PROBABILISTIC METRICS

The probabilistic metrics are crucial when uncertainty is substantial so full predictive distributions must be assessed [49].

- *Logarithmic Score*, is defined as the negative logarithm of the proportion of predicted values at time t that correspond to the reference value $\tilde{y}(t)$. This score penalizes predictions that assign low probability to the reference outcome.

$$logS(t) = -\,log\left(\frac{\left|\{\hat{P}_{t_A} = X\}\right|}{N}\right)$$

where $|\{y(t) = \tilde{y}(t)\}|$ the numerosity of the set $\{y(t) = \tilde{y}(t)\}$, i.e. the number of times that $y_i(t) = \tilde{y}(t)$:

$$|\{y(t) = \tilde{y}(t)\}| = \sum_{i=1}^{N} \delta_{\{y_i(t) = \tilde{y}(t)\}}$$

and δ is the Kronecker's delta function:

$$\delta_{\{y_i(t) = \tilde{y}(t)\}} = \begin{cases} 1 & if \quad y_i(t) = \tilde{y}(t) \\ 0 & otherwise. \end{cases}$$

- *Ranked Probability Score (RPS)*, is a generalization of the mean absolute error and it assesses the accuracy of probabilistic predictions

$$RPS(t) = \frac{1}{N} \cdot \sum_{i=1}^{N} |\tilde{y}(t) - y_i(t)| - \frac{1}{2 \cdot N^2} \cdot \sum_{j=1}^{N} \sum_{k=1}^{N} |y_j(t) - y_k(t)|.$$

The first term $\frac{1}{N} \cdot \sum_{i=1}^{N} |\tilde{y}(t) - y_i(t)|$ corresponds to the mean absolute error and the second term accounts for the shape of the predicted values distribution.

- *95% coverage*, which quantifies how often the reference values fall within the model's 95% predictive intervals.

$$Cov = \frac{1}{T} \, |\{t \in \{1, \dots, T\}: y^L(t) \leq \tilde{y}(t) \leq y^U(t)\}|.$$

These metrics provide complementary insights into the model's performance, both in terms of point predictions and uncertainty quantification.

1.7. THE BINOMIAL CHAIN MODEL

The binomial chain model is a stochastic version of the usual ODEs compartmental model [17]-[20]. In these model transitions between

compartments are represented as binomial random variables, allowing the model to reflect the inherent randomness in the system. The chain binomial model is especially suitable when population sizes are small and events are discrete, conditions under which deterministic models may fail to represent the dynamics realistically. At each time t, the number of individuals $f_{\mathrm{X}\to\mathrm{Y}}(t_i)$ moving from compartment X to another compartment Y is sampled from a binomial distribution,

$$f_{X\to Y}(t) \sim Binomial(n = X(t);\ \ p = p_{X\to Y}),$$

where $X(t)$ is the number of individuals in X at time t and the success probability $p_{X\to Y}$ is derived by converting a (possibly time-varying) the instantaneous transition rate $\lambda(t)$ from X to Y into a probability over a timestep Δt. The mathematical steps of the derivation are as follows.
Assuming:

- Temporal independence of infinitesimal events: the probability that an individual transitions in $[t,\ t+dt]$ is independent of the past;

- T is the random transition time of an individual (i.e. the time spent in X before going to Y).

Over an infinitesimal interval,

$$\mathbb{P}(T \in [t, t+dt]) = \lambda(t)\cdot dt \qquad \mathbb{P}(no\ transition\ in\ [t, t+dt]) = 1 - \lambda(t)\cdot dt$$

To obtain the probability of no transition up to time t, multiply the no-transition probabilities across n subintervals of length $dt = t/n$:

$$\mathbb{P}(T > t]) = \prod_{i=1}^{n}(1 - \lambda(t_i)\cdot dt)$$

Where t_i is the beginning of the i^{th} interval. Applying the logarithm and using first order approximation assuming $n \gg 1$ then $dt \approx 0$

$$\log \mathbb{P}(T > t]) = \sum_{i=1}^{n} \log(1 - \lambda(t_i)\cdot dt) \approx -\sum_{i=1}^{n} \lambda(t_i)\cdot dt \ \longrightarrow\ -\int_0^t \lambda(s)\, ds\,,$$

applying the exponential function

$$\mathbb{P}(T > t]) = e^{-\int_0^t \lambda(s)\, ds}$$

so the cumulative probability that the transition occurs by time t is

$$\mathbb{P}(T \leq\ t]) = 1 - e^{-\int_0^t \lambda(s)\, ds}$$

Therefore, over a discrete step $[t,\ t + \Delta t]$, the per-step transition probability for one individual is:

$$p = \mathbb{P}(T \leq t + \Delta t] \mid T > t) = 1 - e^{-\int_t^{t+\Delta t} \lambda(s)\, ds}.$$

In particular, if $\lambda(t)$ is approximately constant on $[t,\ t + \Delta t]$ with value λ, then

$$p = 1 - e^{-\lambda\, \Delta t}$$

and moreover, if also $\Delta t \approx 0$ applying the first order approximation

$$p \approx \lambda \cdot \Delta t.$$

I used the chain binomial model in Chapter 4 where I focused on a specific cohort of individuals enrolled in a PrEP program. Within this cohort, the number of individuals in certain compartments and groups was quite low, making stochastic effects particularly important to capture accurately.

2. INCREASING SITUATIONAL AWARENESS THROUGH NOWCASTING OF THE REPRODUCTION NUMBER

2.1. ABSTRACT

The time-varying reproduction number R is a critical variable for situational awareness during infectious disease outbreaks; however, delays between infection and reporting of cases hinder its accurate estimation in real-time. A number of nowcasting methods, leveraging available information on data consolidation delays, have been proposed to mitigate this problem. In this work, I retrospectively validate the use of a nowcasting algorithm during 18 months of the COVID-19 pandemic in Italy by quantitatively assessing its performance against standard methods for the estimation of R. Nowcasting significantly reduced the median lag in the estimation of R from 13 to 8 days, while concurrently enhancing accuracy. Furthermore, it allowed the detection of periods of epidemic growth with a lead of between 6 and 23 days. Nowcasting augments epidemic awareness, empowering better informed public health responses.

2.2. INTRODUCTION

Epidemiological surveillance is a critical tool for policy making, allowing public health professionals to monitor epidemic trends and the effectiveness of the adopted interventions. One important quantity that can be monitored during an epidemic outbreak by relying on surveillance system data is the time-varying reproduction number (R) [85]. The reproduction number is defined as the average number of secondary infections caused by an average infectious individual and represents a summary metric that measures changes in transmissibility over time, indicating whether and how fast an epidemic is growing (when $R > 1$) or declining (when $R < 1$). R can be estimated with established statistical methods [29], [34], [86] [87] from the time series of the number of cases occurring in a given geographic unit (also known as "epidemic curve"), provided that the generation time distribution of the considered infection is known.

Surveillance systems are in most cases unable to trace the date at which cases were infected, and the temporally closest proxy event that can be measured is the onset of symptoms. Therefore, estimating R from epidemic curves organized as number of cases by date of symptom onset provides the closest estimate in time to the actual transmission events, even when symptomatic cases represent a small subset of all identified cases [44]. Indeed, it has been shown that estimates of the reproduction number are robust as long as the proportion between symptomatic cases and total infections remains stable or even drifts slowly over time [87], [88].

Between the onset of symptoms for a patient and the insertion of their record in the surveillance database, temporal delays occur depending on the medical-seeking behavior of the individual, the logistics of case ascertainment (e.g., testing), the organization of healthcare response systems (e.g., administration of epidemiological questionnaires, contact tracing), and the socio-technological infrastructure for data collection, quality control, data upload and integration. These delays may change over time not only as a function of the progressive improvement of organizational aspects as the outbreak develops but also depending on the saturation of resources for diagnosis, contact tracing, data collection and transmission, and on other factors such as the population's perceived importance of medical seeking at different stages of the epidemic. As much as surveillance systems can be optimized to minimize some components of these delays, their data provide information that is always somewhat lagged with respect to the current epidemiological situation. The above-mentioned delays result in an underestimation (right-truncation) of epidemic curves for symptom onset dates close to the date of reporting, and data relative to these dates will consolidate only in successive updates of the surveillance system. This eventually limits the ability of public health officers to promptly assess the current situation or the effectiveness of recently implemented interventions.

If information on data consolidation delays is available, it can be exploited to "nowcast" epidemic curves, i.e. to adjust for right-truncation. The first methods were proposed in the wake of the AIDS pandemic [89–93], and further approaches were proposed in later years [41], [42], [94]-[101]. The COVID-19 pandemic provided further momentum to this research topic [32], [67], [102], [103] which also found application during the 2022 mpox epidemic [104], [105]. Even though nowcasting may partially compensate the lack of knowledge on recent "occurred but not reported events" [42], in practical terms the right-truncation of incidence curves makes the following two questions of critical

importance for real-time monitoring purposes: i) until what date in the past can the epidemic curve (and therefore R estimates) be trusted, and ii) how the incompleteness in recent data affects the accuracy of R estimates. In this chapter, I retrospectively assess the application of a simple nowcasting algorithm from a point of view of the improvement in situational awareness during an actual health emergency, using extensive surveillance data from the COVID-19 epidemic in Italy collected over more than 18 months.

2.3. METHODS

NOWCASTING ALGORITHM

The nowcasting algorithm implemented here is a variant of the simple non-parametric method proposed by Lawless [42] and was independently developed for surveillance purposes during the early stages of the COVID-19 pandemic. If $C_D(t)$ is the epidemic curve as reported in the surveillance database at a given reporting date D, and $C^*(t)$ is the consolidated epidemic curve that will be reported at the end of the outbreak, the following relationship holds:

$$C_D(t = D - z) = \pi_D(z) \cdot C^*(t = D - z)$$

where $\pi_D(z)$ is a "consolidation distribution" representing the degree of completeness in reported cases within z days from symptom onset estimated at time D. Equivalently, $\pi_D(z)$ can be interpreted as the cumulated probability distribution of a symptomatic case to be reported z days after symptom onset at time D. From this relationship, an estimate of the consolidated epidemic curve can be obtained from the reported epidemic curve if $\pi_D(z)$ is known. An approximation of $\pi_D(z)$ can be obtained by comparing successive updates of epidemic curves [42] over a window of recent epidemic curves with respect to D; in particular, we propose to approximate $\pi_D(z)$ with the average over observed consolidation distributions relative to N symptom onset dates closest to D and that can be considered consolidated at D (see Section 2.6 for a formal and full description of the method). From $\pi_D(z)$ we can also define a "consolidation lag" $T_{D,F}$ as the minimum number of days that will elapse before the completeness in reported cases exceeds a given fraction F of the final count. Note that $\pi_D(z)$ and $T_{D,F}$ are continuously re-updated at every reporting date D, thereby implicitly keeping track of possible temporal changes in consolidation delays due to changes in the surveillance system or in the epidemiology of the pathogen under scrutiny.

We applied the algorithm to data on confirmed symptomatic SARS-CoV-2 infections collected by regional health authorities in Italy and collated by the Istituto Superiore di Sanità (Italian National Institute of Health) within the Italian COVID-19 integrated surveillance system [106] (a description of the system is written in the Section 2.6). Here, we used the national-level epidemic curves by date of symptom onset as reported daily between May 1, 2020, and December 31, 2021. Epidemic curves between May 1, 2020 and June 28, 2020 were used to obtain the first stable estimates of the consolidation distribution; therefore, the algorithm was applied to n=551 epidemic curves reported dates between June 29, 2020 and December 31, 2021.

For each of the n reporting dates D, we estimated the mean values of the time-varying reproduction number computed from non-adjusted epidemic curves ($R_D(t)$, or "net reproduction number"), and from nowcasted ones ($\hat{R}_D(t)$, or "nowcasted reproduction number"), evaluated at dates $D - T_{D,F}$ (with F ranging between 10% and 90% at intervals of 10%). The two estimates were then compared with corresponding estimates obtained from a consolidated epidemic curve reported several months after the end of the study period ($R^*(t)$, or "reference reproduction number"), evaluated at the same date. For each level of completeness F, we denote with $\boldsymbol{R}_F$, $\widehat{\boldsymbol{R}}_F$ and $\boldsymbol{R}^*_F$, respectively the vectors of estimates $R_D\big(D - T_{D,F}\big)$, $\hat{R}_D\big(D - T_{D,F}\big)$ and $R^*_D\big(D - T_{D,F}\big)$, obtained for different values of D.

A PRACTICAL EXAMPLE: APPLICATION OF THE NOWCASTING ALGORITHM

Figure 2.1 reports a practical example of the proposed method for a specific reporting date Δ = April 1, 2021, of the COVID-19 dataset. After estimating the consolidation distribution $\pi_\Delta(z)$ relative to the date of reporting Δ (Figure 2.1a), we compute the corresponding consolidation lags $T_{\Delta,F}$ for different levels of completeness F, the nowcasted epidemic curve (Figure 2.1b), and the estimates of the reproduction numbers $R_\Delta(t)$, $\hat{R}_\Delta(t)$, $R^*(t)$ (Figure 2.1c). This example shows that the net reproduction number significantly underestimated the reference value when the completeness value was below 90%. Thus, in this example the most recent reliable estimate obtainable on Δ = April 1, 2021 without nowcasting is relative to March 16, 2021, corresponding to a consolidation lag of $T_{\Delta,90}$= 16 days before the date of reporting, whereas the nowcasted estimate would be reliable even for the day before (March 31, 2021). In the overall analysis, this qualitative assessment is done quantitatively and for

all reporting dates in the study period, with the aim of assessing the added value of applying nowcasting for situational awareness.

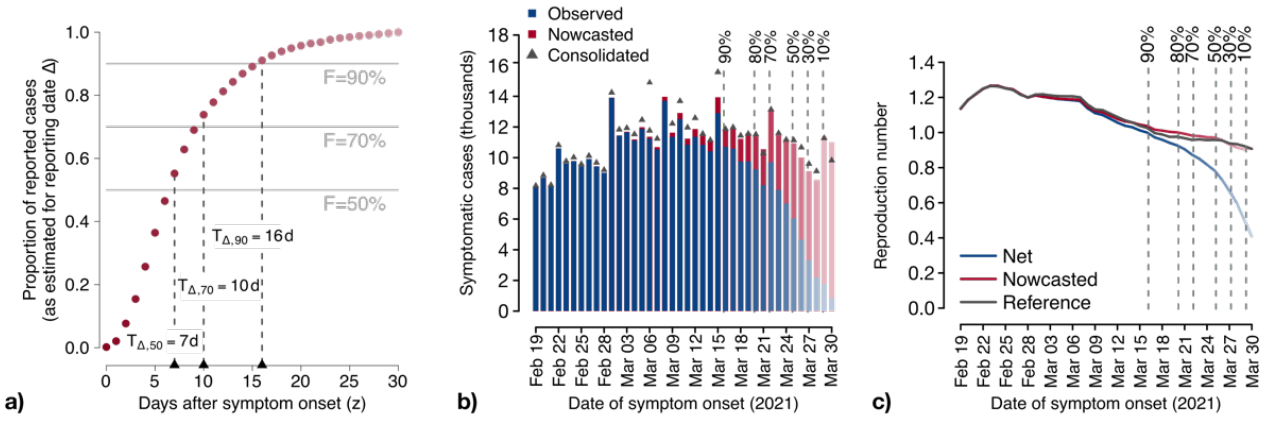


Figure 2.1. *Example of application of the proposed nowcasting technique using data from the Italian COVID-19 integrated surveillance system for a selected reporting date (Δ = April 1, 2021).* a) Estimated consolidation distribution and corresponding consolidation lags. Dark red dots represent the estimated consolidation distribution at Δ. Horizontal lines define selected completeness thresholds (50%, 70% and 90%) and vertical lines define the corresponding consolidation lags. b) Observed and nowcasted epidemic curves by date of symptom onset. The consolidated epidemic curve is shown as dark gray triangles. Vertical dashed lines show the dates at which the observed number of cases is estimated to have reached a given completeness value. Bars in the epidemic curve are reported in fading colors with a level of darkness proportional to the estimated completeness. c) Mean estimates of the net, nowcasted, and reference reproduction numbers over time. Vertical dashed lines show the dates at which the observed number of cases is estimated to have reached a given completeness value. The level of darkness in line colors is proportional to the estimated completeness.

PERFORMANCE METRICS

For each value of completeness F and each considered reporting date D, we calculated: the absolute error of the net and nowcasted estimates of the reproduction number against the corresponding reference value; the proportion of all reporting dates in which each estimate underestimated the reference value; and the proportion of all reporting dates for which either estimate was closer than the other to the reference value.

Additionally, we defined an "epidemic period" as a sustained period of time during which the reproduction number remains above the epidemic threshold of 1. Specifically, we defined the start of an epidemic period as the first of at least 3 consecutive reporting days where the estimate of the reproduction number was above 1; and the end of an epidemic period as the day before the first of at least 7 consecutive reporting days where the estimate of the reproduction number was below 1. We compared the lags with which the net and nowcasted reproduction numbers were able to identify epidemic periods

of duration longer than 15 days, as defined by the reference reproduction number.

2.4. RESULTS

During the study period, the Italian COVID-19 integrated surveillance system took about 6 days (median over the study period; 95% quantile: 5-8 days) from symptom onset to record at least 50% of all cases, and about 13 days (95% quantile: 7-17 days) to record at least 90% (Figure 2.2).

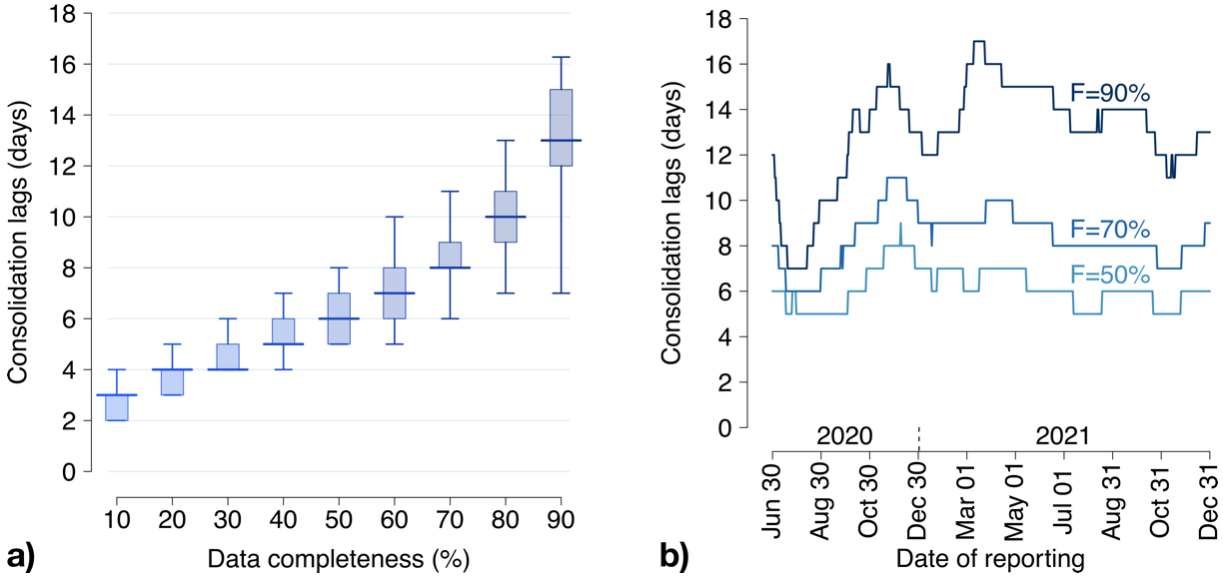


Figure 2.2. *Consolidation lags for the Italian COVID-19 integrated surveillance system.* a) Distribution of the consolidation lag across the estimation period (June 29, 2020 – December 31, 2021) for different values of completeness. Boxplots show the median (horizontal line), interquartile range (rectangle) and 95% quantiles (whiskers) over the n=551 reporting dates. b) Consolidation lags at different reporting dates as estimated for three selected values of completeness F.

Overall, the net reproduction number was a good approximation of the reference reproduction number R^* when evaluated at the date of 90% completeness, with a median absolute error for R_{90} over the entire study period of 0.042 (interquartile range, IQR: 0.025-0.066, Figure 2.3a). The corresponding nowcasted estimate $\hat{R}_{90}$ had a median absolute error of 0.017 (IQR: 0.007-0.036, significantly smaller than the error for R_{90}: p-value of paired t-test between the errors of the net and nowcasted estimates $\ll 0.001$).
When accepting lower thresholds for data completeness, the accuracy of the net reproduction number degraded rapidly, but not the one for nowcasted estimates. For example, with a completeness of 70% (corresponding to a median lag of 8 days), the median error was 0.116 (IQR: 0.078-0.171) for R_{70} but 0.032 (IQR: 0.016-0.068) for $\hat{R}_{70}$, i.e., still significantly smaller than the error

for R_{90} (paired t-test p-value << 0.001). The median error for $\hat{R}_{10}$ (0.115; IQR: 0.057-0.191), corresponding to a median lag of 3 days, was comparable to the median error for R_{70} (paired t-test p-value: 0.42). The net estimate systematically underestimated the reference value, while the nowcasted did so in about half of the reporting dates; the latter result was roughly independent on the considered completeness (Figure 2.3b). Furthermore, the nowcasted estimates were closer to the reference value, compared to net estimates obtained with the same completeness. This occurred for more than 90% of reporting dates when completeness values of 80% or lower were considered (Figure 2.3c).

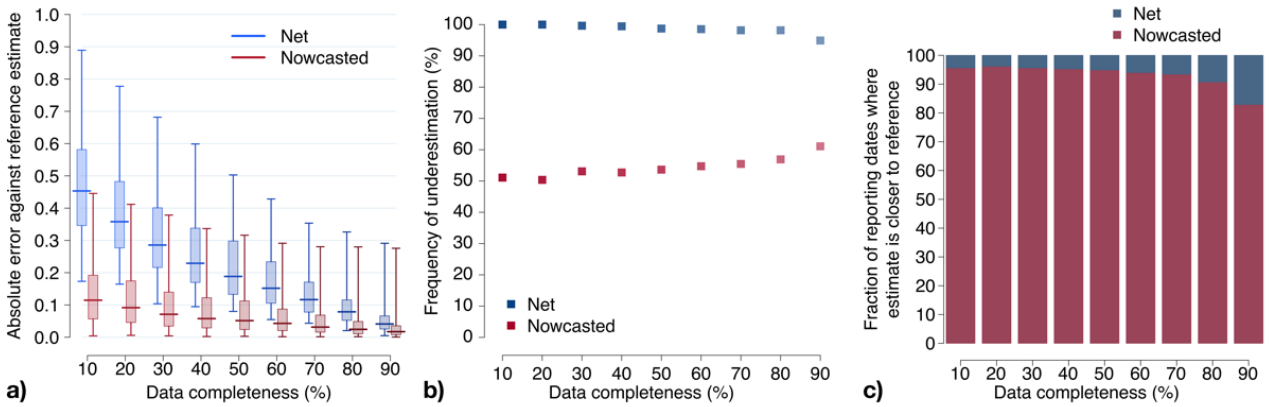


Figure 2.3. *Accuracy of the net and nowcasted reproduction numbers.* a) Distributions of the absolute error between the reference reproduction number and the net and nowcasted reproduction numbers, computed at different reporting dates (daily between June 29, 2020, and December 31, 2021), and evaluated at the date corresponding to a specified level of completeness. Boxplots show the median (horizontal line), interquartile range (rectangle) and 95% quantiles (whiskers). b) Fraction of reporting dates for which the net and nowcasted estimates of the reproduction number underestimate the reference value. c) Fraction of reporting dates for which either estimate is the closest one to the reference value, for different values of completeness.

We identified 5 epidemic periods where the reference reproduction number was above the epidemic threshold for more than 15 days between June 29, 2020, and December 31, 2021 (Figure 2.4 and Table 2.1). The net estimate at 90% completeness detected the epidemic periods with a lag that ranged between 13 and 30 days (Table 2.1). The net estimate at 70% completeness reduced by 2-3 days the detection lag for epidemic periods 4-5 but increased the lag by 8 days for period 1 and missed the detection of period 2. The net estimate at 50% completeness missed the detection of 3 periods out of 5. The nowcasted estimate allowed to anticipate the detection by 1 to 15 days (lags range: 12-20 days) at 90% completeness, and by 6 to 23 days (lags range: 7-17

days) at 70% completeness, compared to the net estimate at 90% completeness. In the case of 50% completeness, the nowcasted estimate performed worse than the net estimate at 90% completeness during period 1 (lag of 27 days vs. 23) but reduced the detection lag by 8 days for periods 4 and 5; for periods 2 and 3, the estimate provided an early warning 6 and 4 days compared to the actual start of the periods; however, it also falsely flagged an additional epidemic period between July 2 and August 13, 2020.

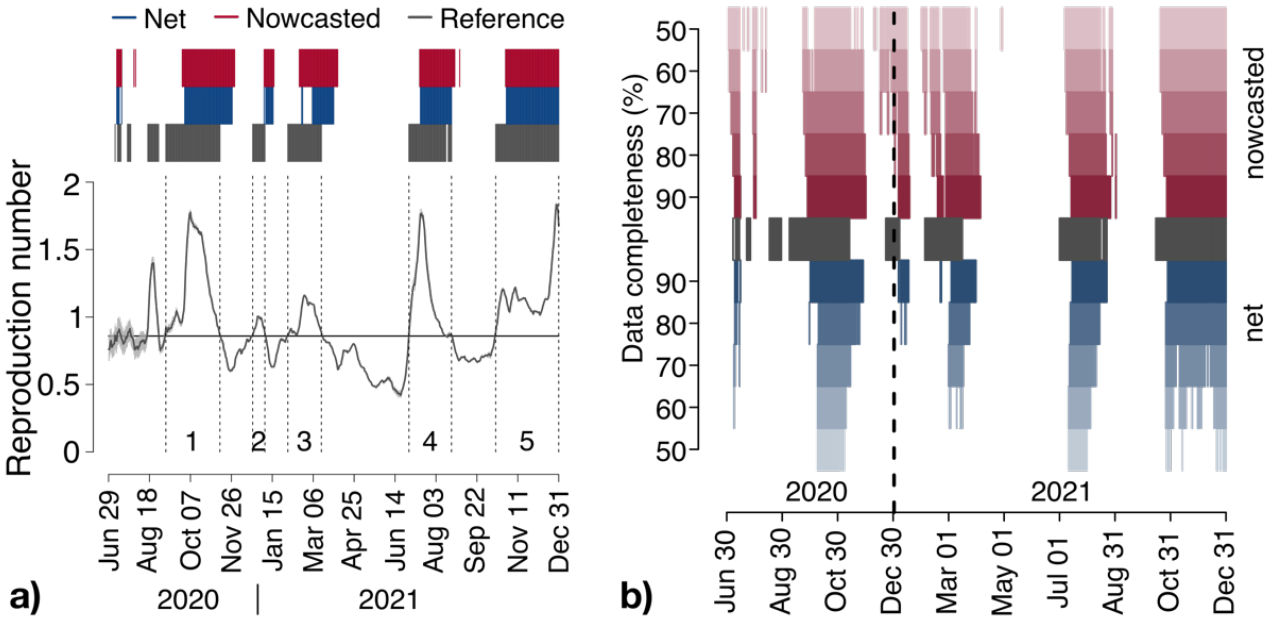


Figure 2.4. *COVID-19 epidemic periods in Italy (2020-2021).* a) Reference reproduction number by date of symptom onset, as computed from consolidated data reported on April 5, 2023 (solid black line). The gray shaded area around the reference reproduction number (visible only in the period May-August 2020, due to the low number of cases contributing to R estimates) represents the 95%CI in its estimate. Dark gray bars above the graph highlight the days where the reference reproduction number was above 1, and dark gray vertical dashed lines delimit epidemic periods (see Table 2.1), labeled with a progressive number just above the x axis. Blue and red bars identify the dates of reporting for which the net and nowcasted reproduction numbers, respectively, estimated at the nearest date afforded by a completeness of 90%, were above 1. b) Dates where the reference reproduction number was above 1 (dark gray), and dates of reporting for which the net and nowcasted reproduction numbers, estimated at the nearest date afforded by a completeness between 50% and 90%, were above 1 (red and blue).

Table 2.1. *Characteristics of COVID-19 epidemic periods in Italy (2020-2021) and detection lags due to consolidation of epidemic curves.* The dates of start and end of epidemic periods refer to dates in which the reference reproduction number (computed from consolidated data reported on April 5, 2023) was above 1 (see methods for the algorithmic definition). The dates of detection correspond to the first date of analysis for which the reproduction number, estimated at the date of the given completeness, was above 1.

Epidemic period		1	2	3	4	5
Date of start		Sep 07, 2020	Dec 22, 2020	Feb 03, 2021	Jul 01, 2021	Oct 15, 2021
Date of end		Nov 12, 2020	Jan 06, 2021	Mar 16, 2021	Aug 22, 2021	After Dec 31, 2021
Duration (days)		67	16	42	53	>78
Date of detection	Net (90% completeness)	Sep 30, 2020	Jan 07, 2021	Mar 05, 2021	Jul 15, 2021	Oct 28, 2021
	Net (70% completeness)	Oct 8, 2020	-	Mar 01, 2021	Jul 12, 2021	Oct 26, 2021
	Net (50% completeness)	Oct 10, 2020	-	-	Jul 11, 2021	-
	Nowcasted (90% completeness)	Sep 27, 2020	Jan 05, 2021	Feb 17, 2021	Jul 14, 2021	Oct 27, 2021
	Nowcasted (70% completeness)	Sep 24, 2020	Dec 30, 2020	Feb 10, 2021	Jul 09, 2021	Oct 22, 2021
	Nowcasted (50% completeness)	Oct 4, 2020	Dec 16, 2020	Jan 30, 2021	Jul 07, 2021	Oct 20, 2021
Detection lag	Net (90% completeness)	23	16	30	14	13
	Net (70% completeness)	31	-	26	11	11
	Net (50% completeness)	33	-	-	10	-
	Nowcasted (90% completeness)	20	14	14	13	12
	Nowcasted (70% completeness)	17	8	7	8	7
	Nowcasted (50% completeness)*	27	-6	-4	6	5

* The nowcasted estimate at 50% completeness falsely flagged an additional epidemic period (July 2 - August 13, 2020).

2.5. DISCUSSION

We performed a quantitative assessment of the use of nowcasting to improve situational awareness during an epidemic outbreak, based on extensive data from over 18 months of the COVID-19 epidemics in Italy. Nowcasting systematically outperformed the non-adjusted estimation of reproduction numbers, improving both the accuracy and timeliness of the estimates. In particular, the adopted nowcasting algorithm was able to more than halve the mean absolute error compared to the net estimate (i.e., from non-adjusted epidemic curves) when evaluated at the same median lag of 13 days before the date of analysis, Nowcasting maintained a better accuracy even for estimates

relative to a median of only 8 days before the date of analysis. More importantly, the nowcasted estimates markedly reduced by between 6 and 23 days the lag in detecting the beginning of epochs of sustained epidemic circulation, a notoriously difficult task for forecasting approaches [107].

In Italy, official estimates of the COVID-19 reproduction number were based on net estimates with reference to 14 days before the date of reporting, which we found to approximately correspond to a data completeness of 90%. The analysis proposed here retrospectively validates the choice of a 14-days lag, given the remarkable worsening of accuracy that would be obtained with shorter lags. Official estimates were made available to the public through a weekly bulletin and used for decisions on non-pharmaceutical interventions [108]. Nowcasted estimates of the reproduction number were additionally used by national and regional health authorities throughout the course of the emergency to gather further insights during weekly situation assessments, thus improving situational awareness. The analysis proposed here represents a retrospective validation of this additional estimate.

There are a few limitations that need to be considered when interpreting our results. First, the estimates of the reproduction numbers were obtained by assuming a fixed distribution of the generation time throughout the study period, corresponding to the one estimated for SARS-CoV-2 ancestral lineages. This choice was taken for simplicity and based on estimates for the Alpha [109], Delta [109] and Omicron variants [110] suggesting limited changes in the distribution of the SARS-CoV-2 generation time in Italy in the study period. Second, the main analysis compared different reproduction numbers only in terms of their mean estimate, without considering their variability. However, results remained robust when considering alternative error functions that considered the variability in estimates (see Section 2.6). For what concerns the definition of epidemic periods, we used a heuristic definition to compare the potential in early warning of the net and nowcasted estimates at different levels of completeness. This definition does not distinguish situations of moderate (R slightly above 1) vs. catastrophic (R much above 1) epidemic growth and therefore does not necessarily correspond to the need for public health decision makers to take urgent action. Still, the identified epidemic periods corresponded to the main periods of expansion of the COVID-19 epidemics in Italy during the study period (Figure 2.4a), including the second wave in the fall of 2020 (period 1), the short resurgence during Christmas holidays of 2020 (period 2), the wave related to the expansion of the Alpha variant in spring 2021 (period 3), the increase of cases in summer 2021, partially related to the

celebrations for the Italian victory in Euro2020 soccer championship [111] (period 4), and the wave due to Delta in the fall of 2021, replaced by Omicron in the last week of the year [112] (period 5). Although daily updates of epidemic curves for SARS-CoV-2 were available until April 15, 2022 (after which updating over the weekends has been suspended), we decided to stop the benchmarking exercise on December 31, 2021, given the lower severity of the pandemic in 2022 [113] – [115], the broad diffusion of self-tests to be performed at home, and the progressive shift towards hospital surveillance. Finally, we were not able to assess the method on the first wave of the COVID-19 pandemic, due to the lack of harmonization in data collection across different regional health systems (especially in the earliest weeks after the first viral detection) and to the rapid temporal variations in ascertainment rates, organizational set-up in epidemiological investigations and data collection.

The overall purpose of this study was to demonstrate and quantitatively assess the usefulness of a simple non-parametric nowcasting method from the perspective of situational awareness in real-time during epidemic outbreaks. Nowcasting can empower better informed public health responses through improved accuracy and timeliness of the estimates of the reproduction number and an earlier identification of periods of sustained epidemic growth. We suggest that nowcasting should become standard practice in surveillance activities, especially in situations of public health emergencies. Several methods have been proposed for nowcasting and some have been made publicly available as packages for statistical software [116]-[119]. A systematic comparison of requirements, advantages and disadvantages, and performances against standardized might support epidemiological data analysts in choosing the most appropriate nowcasting tool for different situations. Although an assessment of the reliability of nowcasting in the early phase of the COVID-19 pandemic was not possible in our study, we argue that pandemic preparedness towards harmonized data collection can minimize the time until the condition is met (i.e., a stable surveillance system providing regular updates of observed epidemic curves) for a reliable application of nowcasting.

2.6. APPENDIX

A SIMPLE METHOD FOR NOWCASTING EPIDEMIC CURVES

Here we provide the details of the nowcasting method used in this chapter, which was developed within surveillance activities in the early months of the

COVID-19 epidemic in Italy and is based on the same principles of the seminal paper by Lawless [42].

DEFINITIONS

Given a generic infectious disease, we define D as a generic date of update of the epidemic curve by the epidemiological surveillance system. Therefore, D implicitly identifies different versions of the surveillance database. We define the "observed epidemic curve" $C_D(t)$ as the number of symptomatic cases C with symptom onset at date t reported by the surveillance system at a date of reporting D. We define the "consolidated epidemic curve" $C^*(t)$ as the epidemic curve that is observed at the end of the outbreak, when all cases have been inserted in the database. For dates of symptom onset sufficiently distant in the past from D, we can assume that the observed epidemic curve is consolidated, i.e., that there is a time interval θ such that for $t \leq D - \theta$, $C_D(t)$ equals $C^*(t)$. We thus define for each reporting date D a "consolidation distribution", π_D, representing the proportion of cases recorded in the consolidated curve C^* and with symptom onset dates between $D - \theta$ and D that were already reported at date D:

$$\pi_D(z) = \frac{C_D(D-z)}{C^*(D-z)}, \qquad (0 \leq z < \theta).$$

The consolidation distribution π_D can be interpreted also as the proportion of cases having symptom onset at date $t = D - z$ that are reported in the surveillance database within z days after symptom onset (i.e., at date D). By definition, π_D cannot be known exactly at the date of reporting D; however, if we can obtain an approximation $\hat{\pi}_D$ using data available until D, an approximation $\hat{C}_D(t)$ of $C^*(t)$ can be inferred at date D:

$$C^*(t = D - z) \cong \hat{C}_D(t = D - z) = \frac{1}{\hat{\pi}_D(z)}\, C_D(t = D - z).$$

We refer to $\hat{C}_D(t)$ as the "nowcasted epidemic curve" estimated at time D.
From the approximated $\hat{\pi}_D$, we can additionally define a "consolidation lag" $T_{D,F}$, representing the minimum number of days that need to elapse before the number of cases of a given symptom onset date exceeds a given fraction F of all cases that will be recorded at the end of the outbreak:

$$T_{D,F} = \min\{z : \hat{\pi}_D(z) \geq F\}.$$

The consolidation lag gives an indication on until which symptom onset date the number of cases observed at time D can be considered sufficiently complete; for example, we can consider the epidemic curve observed at time D to be at least 90% complete until symptom onset date $t \leq D - T_{D,90}$.

ESTIMATION OF CONSOLIDATION DISTRIBUTIONS AND LAGS

One way to estimate $\hat{\pi}_D(z)$ is to rely on consolidation distributions relative to symptom onset dates $t \leq D - \theta$ that are consolidated at the date of reporting D. These consolidation distributions can be obtained by comparing the number of cases for a given symptom onset date across successive reporting updates. For example, for $t = D - \theta$ we will have:

$$p_{D,0}(z) = \frac{C_{D-\theta+z}(D-\theta)}{C^*(D-\theta)} = \frac{C_{D-\theta+z}(D-\theta)}{C_D(D-\theta)}$$

since we have assumed that $C_D(t = D - \theta) = C^*(t = D - \theta)$. Equivalently, for any symptom onset date $t = D - \theta - i$, $i \geq 0$:

$$p_{D,i}(z) = \frac{C_{D-\theta-i+z}(D-\theta-i)}{C_{D-i}(D-\theta-i)}$$

We propose to approximate the unknown consolidation distribution $\pi_D(z)$ with the average over consolidation distributions $p_{D,i}$ relative to the N symptom onset dates closest to D that are consolidated at D:

$$\hat{\pi}_D(z) = \frac{1}{N}\sum_{i=0}^{N-1} p_{D,i}(z)$$

A PRACTICAL EXAMPLE: ESTIMATING THE CONSOLIDATION DISTRIBUTION

Figure 2.5 reports a practical example of the proposed method to approximate the consolidation distribution for a specific reporting date Δ = April 1, 2021, of the COVID-19 dataset. The closest date of symptom onset for which the number of cases can be assumed to be consolidated (i.e., numbers for this date will not change in reported epidemic curves successive to April 1, 2021) is $\Delta - \theta$ = March 2, 2021, given our choice of θ=30 days. The estimate of $p_{\Delta,0}$ will be given by the number of cases with symptom onset on March 2 reported in epidemic curves produced between March 2 and April 1 and normalized by the

consolidated number reported on April 1 (Figure 2.5a). The estimate of $p_{\Delta,N-1}$, for N=30, will be given by the number of cases with symptom onset on $\Delta - \theta - (N-1)$ = February 1, reported between February 1 and March 3, normalized by the number reported on March 3. After averaging the N observed consolidation distributions, we obtain the consolidation distribution $\hat{\pi}_\Delta(z)$ relative to the date of reporting Δ (Figure 2.5b). Figure 2.1 in the methods section 2.3 shows an example of using the consolidation distribution to compute the nowcasted epidemic curve and reproduction numbers for the same date of reporting.

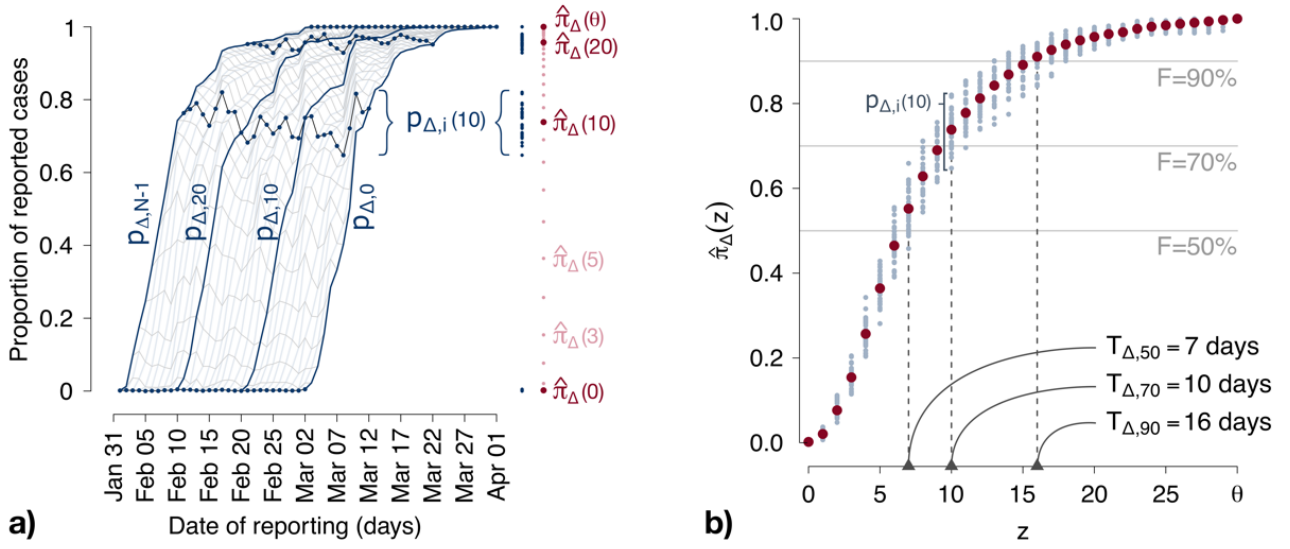


Figure 2.5. *Example of application of the proposed nowcasting technique using data from the Italian COVID-19 integrated surveillance system for a selected reporting date (Δ = April 1, 2021).* a) Approximation of the consolidation distribution. Light blue lines represent consolidation distributions for symptom onset dates between February 1 ($p_{\Delta,N-1}$) and March 2 ($p_{\Delta,0}$), 2021; $p_{\Delta,0}, p_{\Delta,10}, p_{\Delta,20}$ and $p_{\Delta,N-1}$ are reported in thicker and darker lines to highlight the variability across consolidation distributions. Dotted blue lines connect the values of different consolidation distributions for the same number z of days elapsed since the corresponding symptom onset, $p_{\Delta,i}(z)$, for z=0, 10, 20 and θ (gray lines show $p_{\Delta,i}(z)$ for the other values of z). The $p_{\Delta,i}(z)$ distributions for the selected values of z are reported on the right-hand side of the panel as blue dots, and their mean, corresponding to the estimated value of $\hat{\pi}_\Delta(z)$, is reported further to the right as a dark red dot. Pink dots represent the mean of $p_{\Delta,i}(z)$ distributions, corresponding to the estimated value of $\hat{\pi}_\Delta(z)$, for the remaining values of z. b) Estimated consolidation distribution and corresponding consolidation lags. Light blue dots represent the distributions of $p_{\Delta,i}(z)$ for all values of z; dark red dots represent the estimated consolidation distribution at Δ, $\hat{\pi}_\Delta(z)$. Horizontal lines define selected completeness thresholds (50%, 70% and 90%) and corresponding vertical lines define the corresponding consolidation lags.

INTEGRATED SARS-CoV-2 SURVEILLANCE SYSTEM

Since February 27, 2020, the Istituto Superiore di Sanità has coordinated a national case-based surveillance system reporting on all human cases with laboratory-confirmed SARS-CoV-2 infections [106] defined as per the concurrent European Case Definition [120]. Until January 2021, all cases were confirmed by RT-PCR; after that date, confirmation with rapid antigen tests was also accepted [121]. The system was mandated by national law, regulated by dedicated technical documents that described data and data quality requirements and provided weekly data quality verification reports and ad hoc verification of incomplete/inconsistent items to each of the 21 Italian Regions and Autonomous Provinces [122].

The system was based on an online dedicated and secure platform initially compiled manually by regional public health officers. As the epidemic progressed, with high case-loads, standardization for the upload of regional datasets was defined and implemented. The system allowed the collection of demographic data, geographic location data, date of symptom onset, date of diagnoses, date of hospitalization as well data on clinical severity and outcome. The definition of a case as symptomatic and of its date of symptom onset included any SARS-CoV-2-related symptoms, independently of its severity, as defined by clinical evaluation of the case. The definition of a case as imported was based on exposure outside the Italian territory ascertained through epidemiological investigations.

Results of the surveillance system were summarized in a daily dashboard, epidemiological bulletins and used as one of the sources of a mixed method risk assessment protocol that supported pandemic response in Italy [108].

ESTIMATION OF THE TIME-VARYNG REPRODUCTION NUMBER

The estimation of the time-varying reproduction number was performed by applying a standard method [29], [34], [87] which requires as input the number of autochthonous (locally transmitted) cases by date of symptom onset $A(t)$, the number of imported cases (infections acquired outside the geographical setting of interest) $I(t)$, and an estimate of the distribution of the generation time $\varphi(t)$ for the infection under study. The posterior distribution of $\tilde{R}(t)$ was obtained by applying Markov Chain Monte Carlo with Metropolis-Hasting sampling ($n = 50{,}000$ iterations) to the likelihood function defined below:

$$\mathcal{L} = \prod_{t=1}^{D} P\left(A(t);\ \tilde{R}(t)\sum_{s=1}^{t}\varphi(s)C(t-s)\right)$$

where:

- $P(k;\ \lambda)$ is the probability mass function of a Poisson distribution (i.e., the probability of observing k events if these events occur with rate λ).
- $C(t) = A(t) + I(t)$is the total epidemic curve (total number of cases with symptom onset at time t).
- $\varphi(s)$ is the integral of the probability density function of the generation time evaluated between day $s-1$ and s.
- $\tilde{R}(t)$ is the daily reproduction number at time t.

The posterior distribution of the time-varying reproduction number $R(t)$ was computed by applying a weekly moving average to the posterior distribution of $\tilde{R}(t)$.

The same procedure reported above was applied to the consolidated epidemic curves $C^*(t)$ in order to estimate the reference reproduction number $R^*(t)$, and, at each reporting date D, to the net and nowcasted epidemic curves ($C_D(t)$ and $\hat{C}_D(t)$ respectively), in order to estimate the net ($R_D(t)$) and nowcasted ($\hat{R}_D(t)$) reproduction numbers over time.

For application to the Italian COVID-19 data, nowcasting was applied only to the autochthonous component of the epidemic curve, based on the observation that imported cases represented in most cases a negligible fraction of the total (median 0.4%, IQR 0.20-1.88% across symptom onset dates between January 28, 2020, and December 31, 2021). For the distribution of the generation time $\varphi(s)$, we used the distribution of the serial interval estimated in Italy for ancestral lineages [123], given by a gamma function with shape 1.87 and rate 0.28, for a mean of 6.68 days. Further estimates of the distribution of the generation time in Italy for successively emerged variants showed minimal changes with respect to these values [109], [110].

RETROSPECTIVE ASSESSMENT ON THE CHOICE OF θ

A critical assumption in the method for nowcasting epidemic curves is that, for $t \leq D - \theta$, the reported epidemic curve at day D, $C_D(t)$, approximates the consolidated epidemic curve $C^*(t)$. In order to assess the goodness of the chosen value for θ (equal to 30 days in the baseline analysis) when nowcasting COVID-19 epidemic curves in Italy, we computed the mean absolute percentage error $\eta(\theta)$ and the root mean squared error $\varepsilon(\theta)$ between the last

values considered stable for the reported epidemic curve, given by $\tilde{C}(t;\theta) = C_{t+\theta}(t)$ at a certain value of θ, and the corresponding reference values $C^*(t)$:

$$\eta(\theta) = \frac{1}{n}\sum_{t=t_0}^{T}\frac{\left|\tilde{C}(t;\theta) - C^*(t)\right|}{C^*(t)} \qquad \varepsilon(\theta) = \sqrt{\frac{1}{n}\sum_{t=t_0}^{T}\left|\tilde{C}(t;\theta) - C^*(t)\right|^2}$$

Where t_0 and T are the first and last symptom onset dates over which the error is computed and $n = T - t_0 + 1$ is the number of data points in $\tilde{C}(t)$. We evaluated $\eta(\theta)$ and $\varepsilon(\theta)$ for θ between 1 and 50, over the period comprised between t_0 = May 1, 2020 (the first reporting date that was available in our dataset) and T = November 11, 2021 (the last date for which $\tilde{C}$ could be computed when θ=50), resulting in n=560 data points. Figure 2.6 shows that both error metrics stabilize when θ is above 20 days, confirming that θ=30 days was an appropriate choice for Italian COVID-19 data.

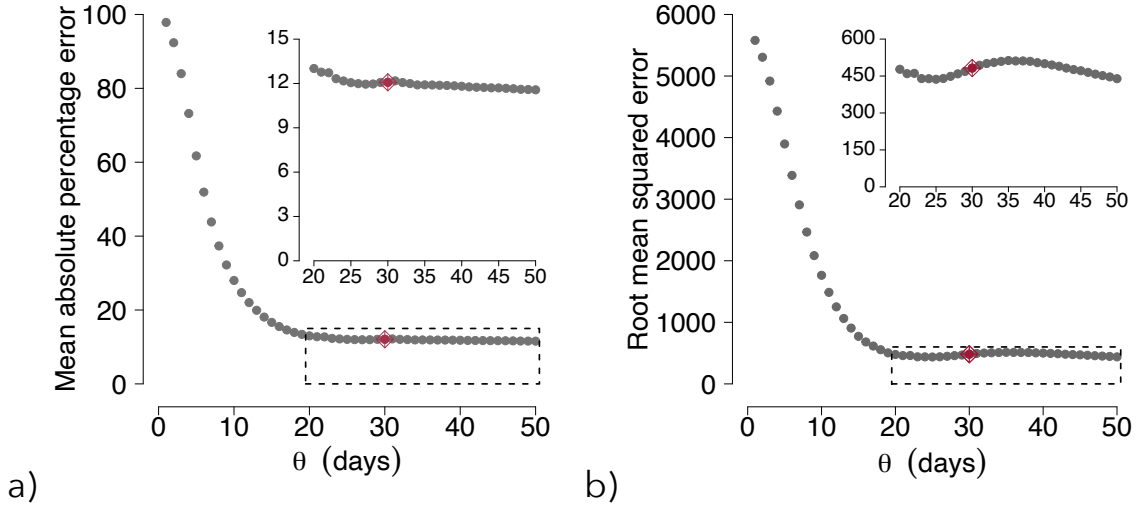


Figure 2.6. *Error metrics for assessing the choice of θ.* a) Mean absolute percentage error, $\eta(\theta)$; b) Root mean square error, $\varepsilon(\theta)$. The error value for the adopted value of θ=30 days is highlighted in red. To better appreciate differences in errors for values of θ>20 days, insets show the same graph with a zoom on the y-axis.

SENSITIVITY ANALYSIS ON θ AND N

The performance of the proposed method may depend on the choice of its two parameters: the number of days after which a symptom onset date can be considered consolidated (θ) and the number of observed consolidation distributions used to nowcast the epidemic curves (N). In order to evaluate the sensitivity of model results with respect to the chosen values, we repeated the whole analysis for 6 additional combinations of N and θ, where both parameters

can assume values of 20, 30 and 40. Since the first nowcasted estimates are available $\theta + N - 1$ days after the date of the first dataset, we evaluated results for reporting dates that were common to all combinations of θ and N, i.e. between July 19, 2020 (first date of availability of nowcasted estimates for $N = \theta = 40$) and December 31, 2021. Figures 2.7 and 2.8 show minimal changes in the estimation of consolidation lags and in the performance of the method when choosing alternative values of the two parameters for selected data completeness levels (F = 50%, 70% and 90%).

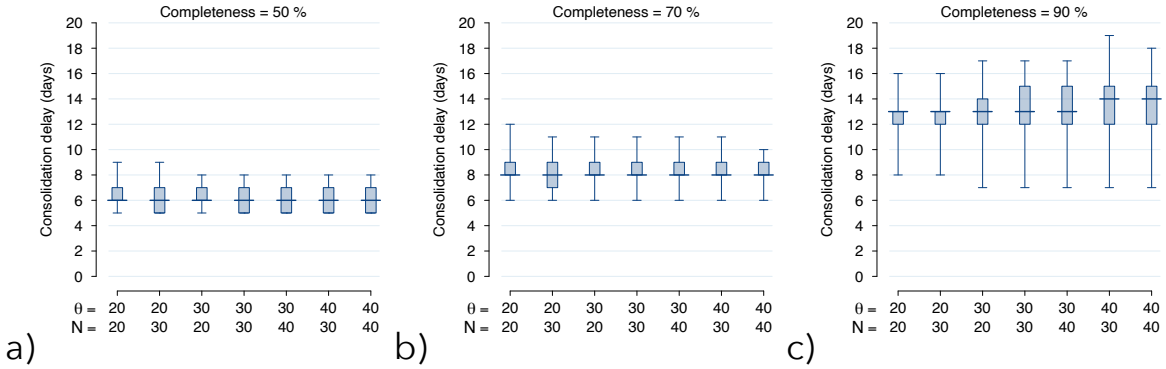


Figure 2.7. *Distribution of consolidation lags for different combinations of the parameters θ and N.* a) 50% completeness. B) 70% completeness. C) 90% completeness.

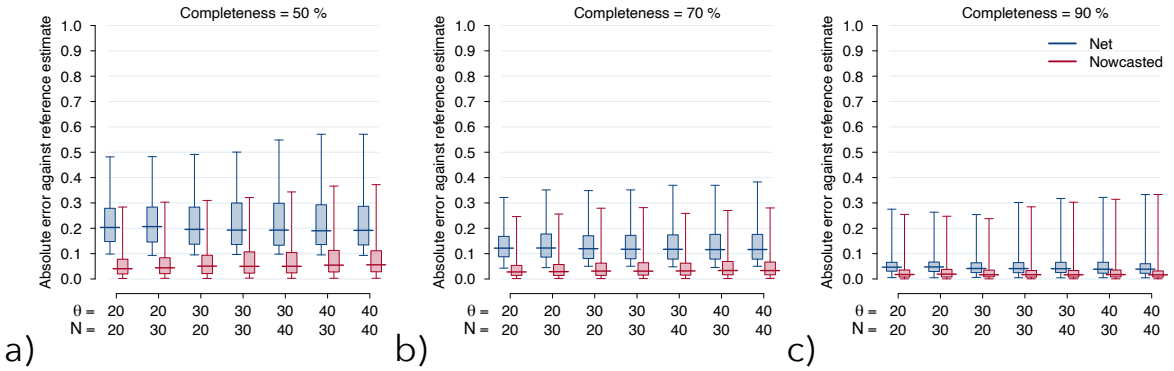


Figure 2.8. *Distribution of the absolute errors between the reference reproduction number and the net and nowcasted reproduction numbers, respectively, for different combinations of parameters θ and N and for selected levels of completeness.* a) 50% completeness. B) 70% completeness. C) 90% completeness.

FUTHER RESULTS WITH ALTERNATIVE ERROR FUNCTIONS

In this section, we report results on further performance metrics for the net and nowcasted reproduction numbers in the baseline analysis (θ=30 days, N=30). Given the net reproduction numbers $R_F(D) = R\left(D - T_{D,F}\right)$ estimated for a given data completeness F and at each reporting date D and consolidation lag $\mathrm{T_{D,F}}$, we considered:

i) the bias error:

$$E_F^B(D) = R^*(D) - R_F(D),$$

ii) the percentage error:

$$E_F^P(D) = 100 \cdot \frac{|R_F(D) - R^*(D)|}{R^*(D)},$$

iii) the error with tolerance:

$$E_F^T(D) = max\{R_F(D) - R^*_{97.5}(D);\ R^*_{2.5}(D) - R_F(D);\ 0\},$$

where $R^*_{2.5}$ and $R^*_{97.5}$ are respectively the lower and upper bounds of the 95%CI for the posterior distribution of R^*. The error with tolerance is defined in such a way to be zero when the average value of the net reproduction is within the 95% confidence interval of the reference estimate, and to be equal to the distance between the estimate and the closest boundary of the confidence interval otherwise. The corresponding errors were computed for the nowcasted reproduction numbers by substituting R_F with $\hat{R}_F$.

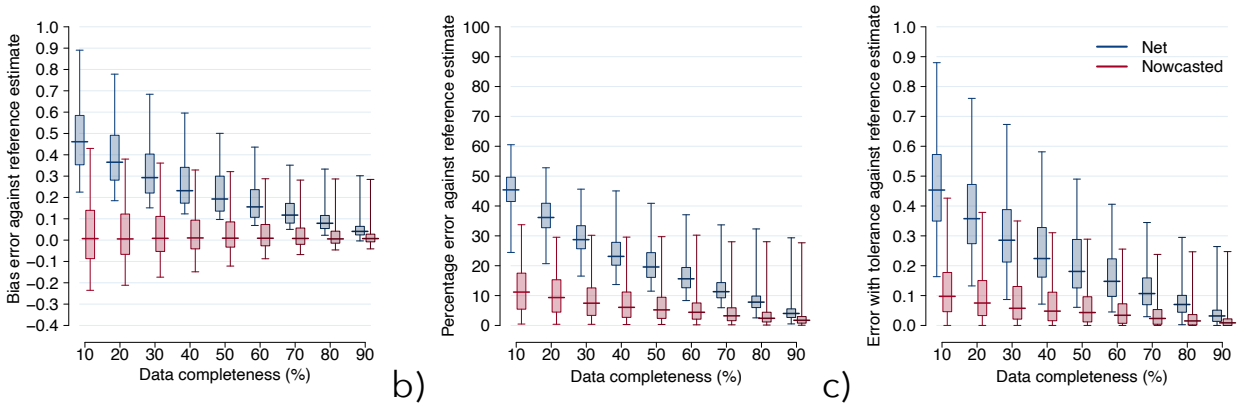
Figure 2.9 and Table 2.2 compare the overall performance of the net and nowcasted estimates for different values of data completeness. The bias error shows a significant downward bias (underestimation) for the net reproduction number and a substantial lack of bias for the nowcasted reproduction number at all levels of data completeness (see also Figure 2.7b). The percentage error shows that the net estimate at 90% completeness and the nowcasted estimate at 70% completeness approximate the reference value with an accuracy higher than 95% in a majority of reporting dates. The results of the error with tolerance may be better appreciated in Figure 2.10 and Table 2.3, where we considered the fraction of times in which the net and nowcasted estimates are:

i) correct (when the mean estimate falls within the 95%CI of the reference);

ii) underestimated (when the mean estimate falls below the 95%CI of the reference);

iii) overestimated (when the mean estimate falls above the 95%CI of the reference).

Net estimates of the reproduction number are correct only for 9% of reporting dates at a data completeness level of 90%, against 26% for the corresponding nowcasted estimate. Even with a data completeness of 50%, the nowcasted estimates are correct in a higher proportion of reporting dates (14%) compared to the net estimate at 90%.



a) b) c)

Figure 2.9. *Comparison of performance between the net and nowcasted estimates using alternative error functions.* a) Bias error; positive values represent underestimations of the reproduction number. b) Percentage error. c) Error with tolerance on the statistical variability of the reference estimate.

Estimate	F (%)	Absolute error (median and IQR)	Bias error (median and IQR)	Percentage error (median and IQR)	Error with tolerance (median and IQR)
Nowcasted	50	0.051 [0.023, 0.109]	0.008 [-0.035, 0.084]	5.3% [2.7, 9.8]	0.044 [0.012, 0.097]
	70	0.032 [0.016, 0.068]	0.006 [-0.021, 0.055]	3.4% [1.6, 6.2]	0.023 [0.003, 0.054]
	90	0.017 [0.007, 0.036]	0.007 [-0.009, 0.028]	1.8% [0.8, 3.1]	0.009 [0, 0.022]
Net	70	0.116 [0.078, 0.171]	0.116 [0.078, 0.171]	11.2% [9.1,14.3]	0.105 [0.067, 0.158]
	90	0.042 [0.025, 0.066]	0.039 [0.023, 0.064]	4.0% [2.7, 5.6]	0.029 [0.012, 0.049]

Table 2.2. *Comparison of performance metrics for the net and nowcasted estimates for selected levels of data completeness.*

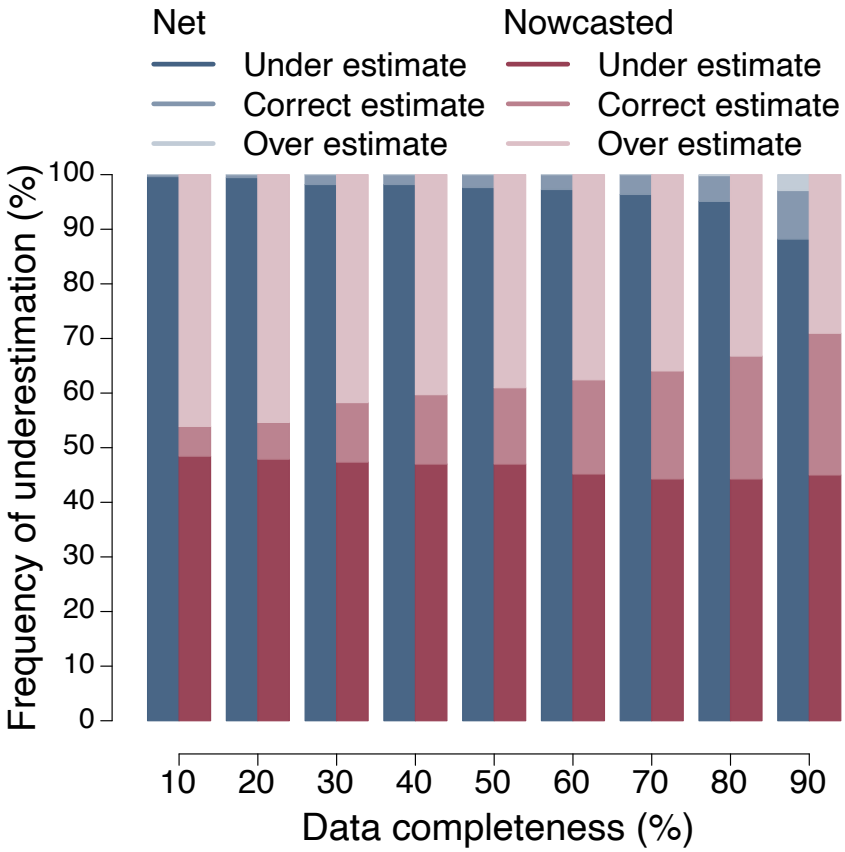


Figure 2.10. *Performance of the net and nowcasted reproduction numbers when considering the stochastic variability of the reference estimate, for different data completeness levels.*

Method	Data completeness (%)	Under-estimate %	Over-estimate %	Correct estimate %
Nowcasted	50	47	39	14
	70	44	36	20
	90	45	29	26
Net	90	88	3	9

Table 2.3. *Performance of the net and nowcasted reproduction numbers when considering the stochastic variability of the reference estimate, for selected data completeness levels.*

ALTERNATIVE DEFINITION OF THE DATA CONSOLIDATION LAG

In the main analysis, we considered explicitly the case in which consolidation distributions $\hat{\pi}_\Delta(z)$ are strictly monotonic with z (Figure 2.1 and 2.5). However, consolidation distributions in actual data may be non-monotonic when the process of consolidation due to reporting delays is counterbalanced by the retrospective removal of cases associated to a date of symptom onset. This may happen for a number of reasons, such as the re-evaluation of the date of symptom onset for a case after additional epidemiological investigations or

after identification of data entry errors, the reclassification of cases as non-cases, or as asymptomatic cases, the identification of case duplications in the dataset, and so on. In these situations, the functions $\hat{\pi}_{\Delta}(z)$ may exceed the value of 1 at some $\mathbf{z}$ between $\mathbf{0}$ and θ. Figure 2.11 shows a comparison of the consolidation distribution for two reporting dates: July 15, 2020, and July 15, 2021. We evaluated the effect of an alternative definition of the data consolidation lag $T_{\Delta,F}$ where we considered, instead of the earliest z for which $\hat{\pi}_{\Delta}(z)$ reaches the desired completeness F, the latest z for which $\hat{\pi}_{\Delta}(z)$ is less than $1-F$ away from its ideal value of 1:

$$T'_{\Delta,F} = \max_{z}\{|1-\hat{\pi}_{\Delta}(z)| \leq 1-F \ \wedge \ |1-\hat{\pi}_{\Delta}(z-1)| > 1-F\}.$$

An example of this second definition is shown in 2.11 for July 15, 2020, and F=90%. In Figure 2.12, differences in the distribution of the data consolidation lag when considering the alternative definition are shown for different values of F and limited to the 210 reporting dates with a non-monotonic consolidation distribution. For F<70%, the data consolidation lags are identical; for F=80% and 90% the alternative definition produces longer data consolidation lags, but with substantially similar medians. In particular, with a data completeness of 80%, the alternative definition yields the same median lag of 10 days but with a 95% confidence interval shifted upward (7-22 days versus 6-13 days in the baseline); with a data completeness of 90% the median lag grows from 12 days (95%CI 7-15) in the baseline to 13 days (95%CI 8-26) in the alternative definition. Table 2.4 shows that the alternative definition improves the performance of both the net and nowcasted estimates at 80% and 90% completeness, although at the cost of higher lags.

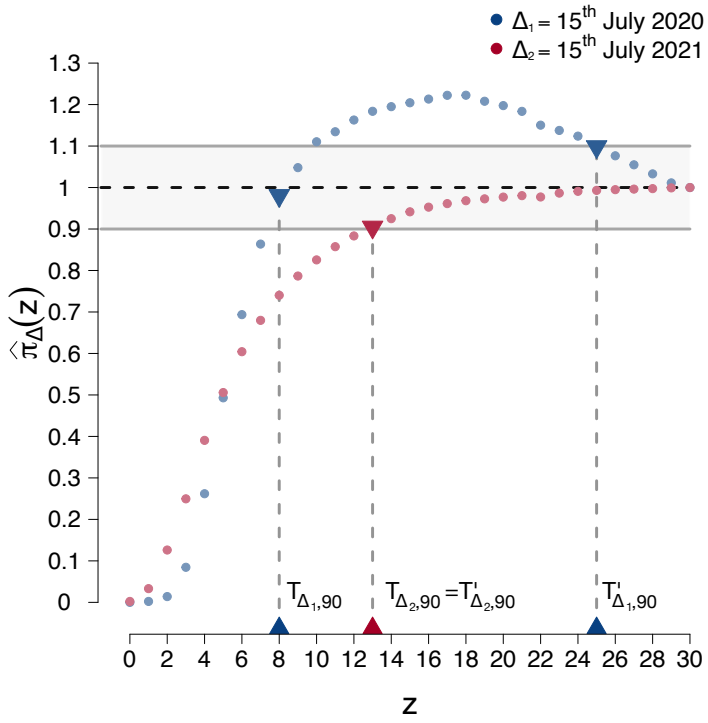


Figure 2.11. *Example of alternative definitions for the data consolidation lag.* Blue and red dots represent the consolidation distributions associated to reporting dates July 15, 2020, and July 15, 2021, respectively. Triangles on the x-axis represent the estimated data consolidation lags according to the two definitions. In the case of a monotonic curve (Δ_2 = July 15, 2021), the two definitions are equivalent, while when the curves are non-monotonic (Δ_1 = July 15, 2020), the two definitions may give different results.

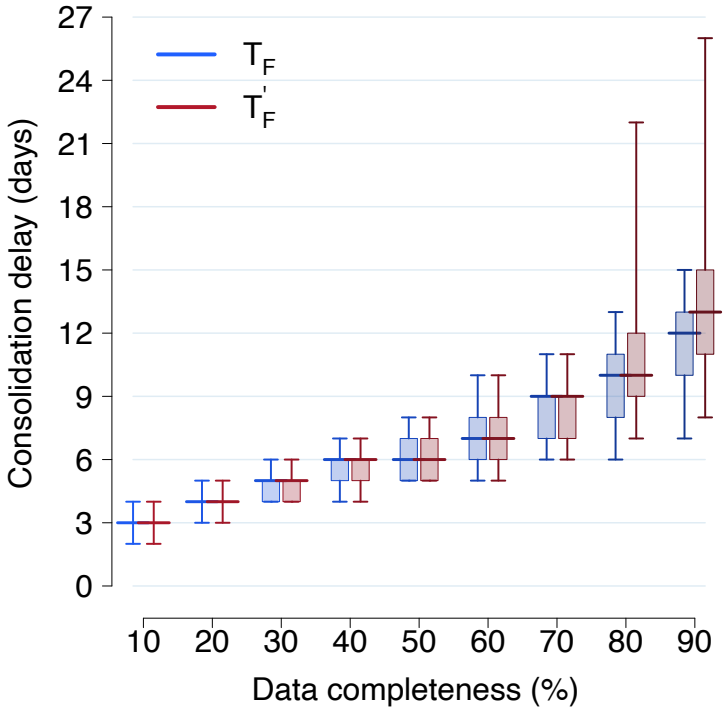


Figure 2.12. *Impact of alternative definitions for the data consolidation lag.* The figure shows the data consolidation lag distributions for different levels of data completeness, according to the baseline and alternative definition.

Method	Data completeness (%)	Definition of data consolidation lag	Absoluteerror (median and IQR)
Nowcasted	80	Baseline	0.048 [0.02, 0.106]
	80	Alternative	0.044 [0.019, 0.102]
	90	Baseline	0.041 [0.018, 0.091]
	90	Alternative	0.031 [0.016, 0.07]
Net	80	Baseline	0.083 [0.055, 0.168]
	80	Alternative	0.074 [0.041, 0.148]
	90	Baseline	0.054 [0.027, 0.099]
	90	Alternative	0.041 [0.018, 0.093]

Table 2.4. *Absolute errors of the nowcasted and net reproduction numbers against the reference estimate, using alternative definitions of the data consolidation lag, for data completeness levels of 80% and 90%.*

3. FORECASTING NOROVIRUS CASES ON CRUISE SHIPS TO SUPPORT OUTBREAK MANAGEMENT ON BOARD

3.1. ABSTRACT

Norovirus outbreaks on cruise ships are a significant threat to the cruising industry. Mathematical models have the potential to leverage routinely collected syndromic surveillance data on board to provide insight into outbreak evolution. We used historical data from seven norovirus outbreaks occurred in 2011-2013, totalling 359 diagnosed cases, to assess the performance of automated forecasts in real-time. We compared the performance of a set of alternative models on three endpoints (the number of cases by symptom onset time, by diagnosis date, and the total number of cases until the end of the cruise), using the logarithmic score (logS), the ranked probability score (RPS), and the 95% coverage. We found that the best forecasting performance was given by a model that includes both superspreading and the effect of case isolation. This model had in most cases a better score than that of a baseline model assuming constant incidence; this happened in 59-70% of data points when assessed using the logS and 53-57% with the RPS (depending on the considered endpoint). The best model also had the highest coverage over all endpoints. Its added value was especially evident for longer forecasting horizons, with an improvement in performance for up to 78% of data points, both according to the logS and the RPS. Simple mathematical models integrating key mechanisms of norovirus transmission can help predict the number of cases on board. This knowledge can be automatized in syndromic surveillance systems to support decision making for the management of outbreaks.

3.2. INTRODUCTION

Acute gastroenteritis outbreaks on cruise ships are most often caused by norovirus [51], with harmful consequences for passengers and crew members' health, disruption of holidays, and high operation costs for cruise companies.

The public health impact and economic burden of norovirus outbreaks for the tourism industry has been broadly recognised [124]. The steady growth of cruising industry since 2004 [125], only temporarily halted by the effects of the COVID-19 pandemic [126], suggests that norovirus outbreaks may be a threat for tourism industry in the future.

Norovirus transmission on board occurs through the ingestion of contaminated food or water, contact with contaminated surfaces, or from infected individuals [127], directly or indirectly, including also the ingestion of aerosolised virus particles emitted through vomitus or feces [127], [128]. The prevention and management of acute gastroenteritis outbreaks on cruise ships follows internationally agreed hygiene standards and plans [129]-[131]. Among the main routine preventative measures are pre-embarkation screening, syndromic surveillance on board, the isolation of infected individuals, the application of environmental hygiene measures and education of crew and passengers on hand washing and symptoms reporting. Management measures when transmission is ongoing may include active surveillance, enhanced disinfection of public surfaces and discontinuation of self-service restaurants [129]-[131].

Syndromic surveillance of acute gastroenteritis cases, implemented using standard definitions [129], [130], allows the modulation of prevention and management measures, thereby supporting the timely activation of outbreak management plans, helping to prevent further person-to-person transmission and eventually contributing to limiting the outbreak size. Mathematical models may constitute a powerful tool to complement epidemic intelligence on board. In recent years, the task of forecasting infectious diseases has been a growing field of investigation, with applications to endemic diseases such as seasonal influenza [132] or emerging infections such as Ebola [49] or COVID-19 [133]. Research efforts have been mostly focused on open populations, with the aim of supporting national or regional public health authorities. 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. To do this, we leveraged data on seven norovirus outbreaks (Table 3.1), collected through existing onboard syndromic surveillance systems, and compare the predictive performance of a set of models against endpoints that may be useful for assisting decision making for public health ship operators.

3.3. METHODS

We considered data from seven outbreaks that occurred between 2011 and 2013 on non-consecutive cruises on two ships belonging to the same cruise line, calling at ports in Mediterranean countries (Table 3.1). We considered as outbreaks voyages where the total number of cases was equal or larger than 20, irrespectively of whether these happened in crew or passengers. Cases were defined by the presence of gastrointestinal symptoms (vomiting, diarrhea, or both). Cases in crew were always less than 4 and less than 10% of all diagnosed cases (Table 3.1). A detailed description of the data is published elsewhere [134]. Because at least one third of cases in the outbreaks included in this study had a positive PCR confirmation for norovirus, we assumed them to be norovirus outbreaks [134]. We repeated forecasts at intervals of four hours starting from the 24th hour of each cruise, using only available data until that time. The four-hour interval was selected during preliminary analyses as a compromise between temporal resolution and having a proper numerosity of cases in the resulting temporal bins. Cases by time of symptom onset were forecasted according to a set of stochastic models until the end of the cruise. Cases by date of diagnosis were obtained by sampling, for each forecasted symptomatic case, a diagnostic delay from a discrete distribution estimated from available data (mean 17 hours, 95% percentile interval: 0-64 hours; Section 3.6). Cases that would be diagnosed after the end of the cruise (due to their sampled diagnostic delay) were discarded from the forecasts. The precise time of occurrence was available only for symptom onset, while for the diagnosis only the day of cruise was known; for this reason, cases by time of diagnosis were aggregated in one-day intervals according to their date. Forecasted endpoints at any time of analysis were compared against the corresponding endpoints observed at the end of each cruise, which were considered as the true value. All analyses were run using R version 4.4.0 (2024-04-24).

Table 3.1. *Cruise characteristics and cumulative incidence among passengers and crew.*

Cruise ID	Length of cruise [days]	Number of diagnosed cases		Population onboard		Cumulative incidence (%)	
		Passengers	Crew	Passengers	Crew	Passengers	Crew
A	7	118	2	1229	487	9.6	0.4
B	7	19	1	1222	476	1.6	0.2
C	7	23	2	1086	474	2.1	0.4
D	7	81	1	1245	475	6.5	0.2
E	14	27	1	1153	480	2.3	0.2
F	7	36	4	1300	510	2.8	0.8
G	7	40	4	981	381	4.1	1.0

MODELS

The baseline model (M0) forecasted cases at each time-step in the future by sampling from a Poisson distribution with rate equal to the mean incidence of cases recorded in the 24-hour window before the time of analysis. The other three models (M1-M3) are based on variations of the well-established renewal equation [64].

In its standard formulation, which was adopted for model M1, the number of cases by date of symptom onset $C^f_{t_A}(t)$ forecasted at time of analysis t_A for time-step $t+1$ is given by:

$$C^f_{t_A}(t+1) \sim Poisson\left(\rho_t \tilde{R}_{t_A}\left(\sum_{s=1}^{t_F} C_{t_A}(s)\cdot\varphi(t-s) + \sum_{\sigma=t_F+1}^{t} C^f_{t_A}(\sigma)\cdot\varphi(t-\sigma)\right)\right)$$

Where ρ_t represents the proportion of susceptible individuals on board at time t, $\tilde{R}_{t_A}$ is an estimate of the net reproduction number obtained at time t_A, assumed to be maintained unchanged throughout the forecast until the end of the cruise; C_{t_A} is the number of cases by time of symptom onset recorded in the cruise surveillance system from the start of the cruise (corresponding to time-

step 1), φ is the discretized distribution of the generation time (taken from [135]), and $t_F = t_A - D$ is the time at which the number of observed cases is considered sufficiently stable considering the effect of diagnostic delays.

For model M2, we introduce the effect of reduced transmission due to isolation by substituting $C_{t_A}(t)$ (and, accordingly, $C^f_{t_A}(t)$) with the expression $N_{t_A}(t) + (1-\alpha) \cdot D_{t_A}(t)$, where $N_{t_A}(t)$ are cases that have not been diagnosed and isolated at time t and generate new cases with transmissibility equal to $\tilde{R}_{t_A}$ and $D_{t_A}(t)$ are cases that have been diagnosed at time t and therefore their transmissibility is reduced by α. Using consolidated outbreak data, we estimated α to be 94.4% (95 CrI: 75.0-100.0%; Section 3.6).

Model M3 modifies model M2 by sampling the number of new cases from a negative binomial distribution, with rate equal to the Poisson rate for M2 and overdispersion parameter set from a preliminary analysis of transmission chain in the largest cruise (mean 0.4, 95% percentile: 0.15-0.65) [63]. Alternative distributions for the overdispersion parameters were considered in the sensitivity analysis (Section 3.6).

A number of Z=1000 forecasting trajectories were produced for each model, where in addition to the stochastic uncertainty we considered the uncertainty in parameters (i.e., $\tilde{R}_{t_A}$, α and ϕ) by sampling each time a different value from the estimated distribution.

We estimated the distribution of $\tilde{R}_{t_A}$ as $R_{t_A}(t_F)$, where $R_{t_A}(t)$ is the net-reproduction number, computed by applying well-established techniques [29, 34] to the epidemic curve by time of symptom onset available at t_A. Cases are projected starting with time step $t_F + 1$, and projected cases are iteratively used in the renewal equation to estimate the successive time steps. In the main analysis we assumed D = 24 hours, and we ran sensitivity analyses for other values of D.

For cases by time of symptom onset, $C^f_{t_A}(t)$, projected between t_{F+1} and t_A, we apply particle filtering to select stochastic trajectories that reproduce as much as possible observed data within this interval. Full specifications of the different procedures involved in the forecasts are reported in the Section 3.6.

SCORING

We used different metrics to evaluate forecasts that penalize differently the discrepancy between the forecast distribution and the observed value.

The logarithmic score, logS, is defined as the negative logarithm of the proportion r_{X,t_A} of forecasts produced at time t_A that correspond to the observed value X:

$$logS = -log\left(r_{X,t_A}\right) = -log\left(\frac{\left|\{\hat{P}_{t_A} = X\}\right|}{Z}\right)$$

Where $\hat{P}_{t_A}$ is the set of forecasts for the data point corresponding to X, $|\{x\}|$ is the numerosity of set $\{x\}$, and Z is the number of forecasts produced.
The Ranked Probability Score (RPS) is a generalization of the mean absolute error, given by:

$$RPS = \frac{1}{Z} \cdot \sum_{i=1}^{Z} \left|X - \hat{P}_{t_A,i}\right| - \frac{1}{2 \cdot Z^2} \cdot \sum_{j=1}^{Z} \sum_{k=1}^{Z} \left|\hat{P}_{t_A,j} - \hat{P}_{t_A,k}\right|$$

Where $\hat{P}_{t_A,i}$ represents the i-th forecast in the set $\hat{P}_{t_A}$. The first term $\frac{1}{Z} \cdot \sum_{i=1}^{Z} \left|X - \hat{P}_{t_A,i}\right|$ corresponds to the mean absolute error and the second term accounts for the shape of the forecast distribution.
Finally, the coverage for a given percentile w was defined as:

$$Cov_w = \frac{1}{N_p} \left|\{i \in \{1, \dots N_p\}: l_{w,t_A,i} \leq X_i \leq u_{w,t_A,i}\}\right|$$

Where $l_{w,t_A,i}$ and $u_{w,t_A,i}$ represent the lower and upper bounds of the Prediction Interval forecasted at t_A for the i-th data point X_i, and N_p is the total number of data points.
When comparing the ranking of model performance over the logS and RPS, we computed the distributions of the standardized rank, given by:

$$SR_{m,i} = 1 - \frac{q_{m,i} - 1}{N_m - 1}$$

Where $q_{m,i}$ is the rank of model m for the i-th forecast with the considered score (from 1 for the best model to N_m=4 for the worst model). Therefore, $SR_{m,i}$ is bounded to have values between 0 and 1.

3.4. RESULTS

In order to assess the performance of different models, we simulated real-time forecasts by feeding models with case updates every four hours, for each of the seven outbreak cruises. At any given time of analysis, the input for the forecasts was given by the total number of cases (including both passengers and crew) recorded until that time by the surveillance system. Cases were aggregated by

time of symptom onset and by date of diagnosis. The forecast output was always provided until the end of the cruise and was composed of three endpoints: the number of cases by time of symptom onset, the number of cases by date of diagnosis, and the total number of cases between the time of analysis and the end of the cruise.

In the main analysis, we considered four alternative forecasting models: i) a trivial model that forecasts, for every four-hour time step until the end of the cruise, a Poisson-distributed number of cases with constant rate equal to the average rate observed in the last 24 hours (model M0, used as a baseline to assess the other models' performances); ii) a model using the renewal equation to project cases under the assumption that the most recent estimate of the reproduction number at the time of analysis will remain constant until the end of the cruise (model M1); iii) a model that builds on M1 by additionally including the effect of isolating individuals immediately after diagnosis (model M2); the reduction in transmissibility for isolated cases was robustly estimated at 94% (95%CI: 75-100%) from outbreak data (see Section 3.6); iv) a model that builds on M2 by considering overdispersion in the offspring distribution of infectious cases (model M3); we considered an overdispersion parameter of 0.4, corresponding to an intermediate level of variance respect to alternative values considered in sensitivity analyses. Full model specifications are reported in the above Section 3.3 and Section 3.6.

To illustrate the forecasting procedure, Figure 3.1 shows one example for cruise D (Table 3.1) at the time of analysis corresponding to the 76th hour of navigation. Forecasts are overlayed with cases reported by the time of analysis (i.e., those available for forecasting, orange squares) and with consolidated data available at the end of the cruise (dark gray dots). Note that cases by symptom onset reported at the time of analysis may be underestimated with respect to consolidated ones due to delays between symptoms and diagnosis (termed hereafter "diagnostic delays"), since some of the cases who had symptom onset before the time of analysis will report symptoms and be diagnosed only after the time of analysis. For this reason, models based on the renewal equation (M1-M3) also attempt an estimation of cases by symptom onset in the last 24 hours before the time of analysis (Figure 3.1). This estimation is necessary to provide forecasts for these models but is not evaluated as a forecast in this study.

Each forecast is scored against the corresponding consolidated number of cases (observed at the end of the cruise), according to three standard evaluation metrics for forecasts: the logarithmic score (logS) [136], [137],

representing the likelihood of observing a data point given model predictions; the Ranked Probability Score (RPS), representing a measure of accuracy of the forecast that also takes into account its variability [RefF14,RefF15]; and the 95% coverage, representing the proportion of data points that fall within the 95% Prediction Interval (PrI) [137]. Lower values of the logS and RPS scores indicate better forecasting performance, while the optimal value for the coverage is 95%; full specifications of the scores are reported in the Methods.

Figure 3.2 shows the cumulative distribution functions (CDF) of the logS and RPS scores, obtained by pooling together all outbreaks, times of analysis, and forecasted data points. The CDF representation allows the display of the scores variability when their distributions have heavy tails for an easier comparison of model performances. In a perfectly predictive model, all forecasts would have logS and RPS scores equal to zero, and the corresponding CDF would be a Heaveside step-function; thus, the best performing model is the one whose CDF is higher than the others. Models M0, M2 and M3 tended to have similar performances over the different endpoints, while M1 fared consistently worse (Figure 3.2). M3 tended to be superior for lower and higher scores, while M0 never had excellent scores, but frequently had intermediate scores. M3 also had the highest coverage over the three endpoints, with values close to 95% for the number of cases by symptom onset time and by diagnosis date. Unsurprisingly, the most difficult endpoint to predict was the total number of cases until the end of the cruise, which resulted in generally worse scores and a massively lower coverage; model M3 was substantially better than others in estimating this endpoint, and 56% of all M3 forecasts contained the actually observed value within its 95% PrI.

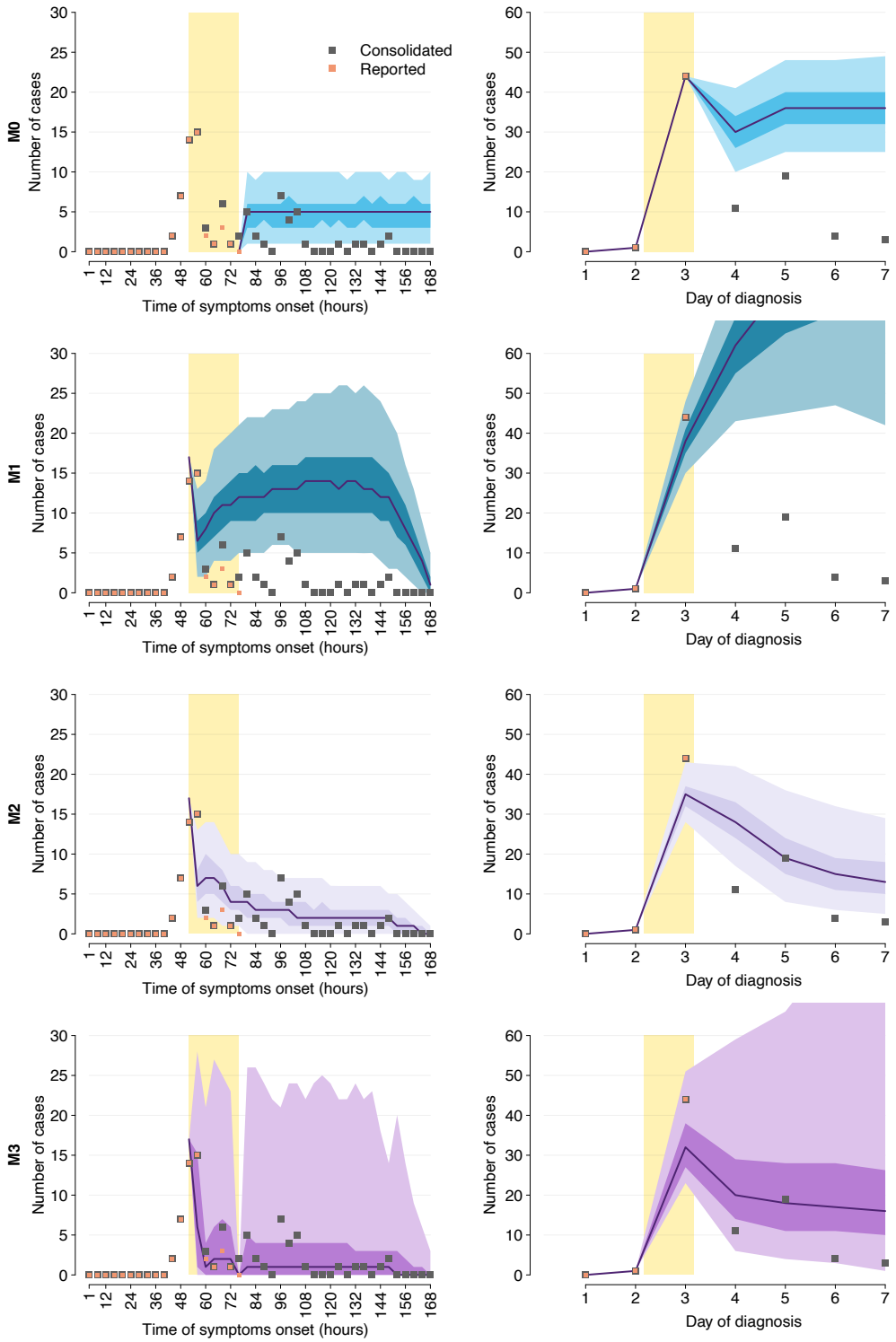


Figure 3.1. *Example of forecasts for a given outbreak and time of analysis.* Top row: constant incidence model (M0, baseline, light blue); upper-middle row: renewal equation model (M1, dark blue); lower-middle row: renewal equation model with isolation of cases (M2, lilac); bottom row: renewal equation model with isolation of cases and overdispersion in the offspring distribution (M3, purple). Left column: cases by symptom onset time; right column: cases by diagnosis date. Orange points: data available by the time of analysis; dark grey points: data available at the end of the outbreak; solid lines: median model predictions; shaded areas: 50% (darker color) and 95% (lighter color) prediction interval (PrI). The gold shaded rectangle identifies a 24-hour time window before the time of analysis.

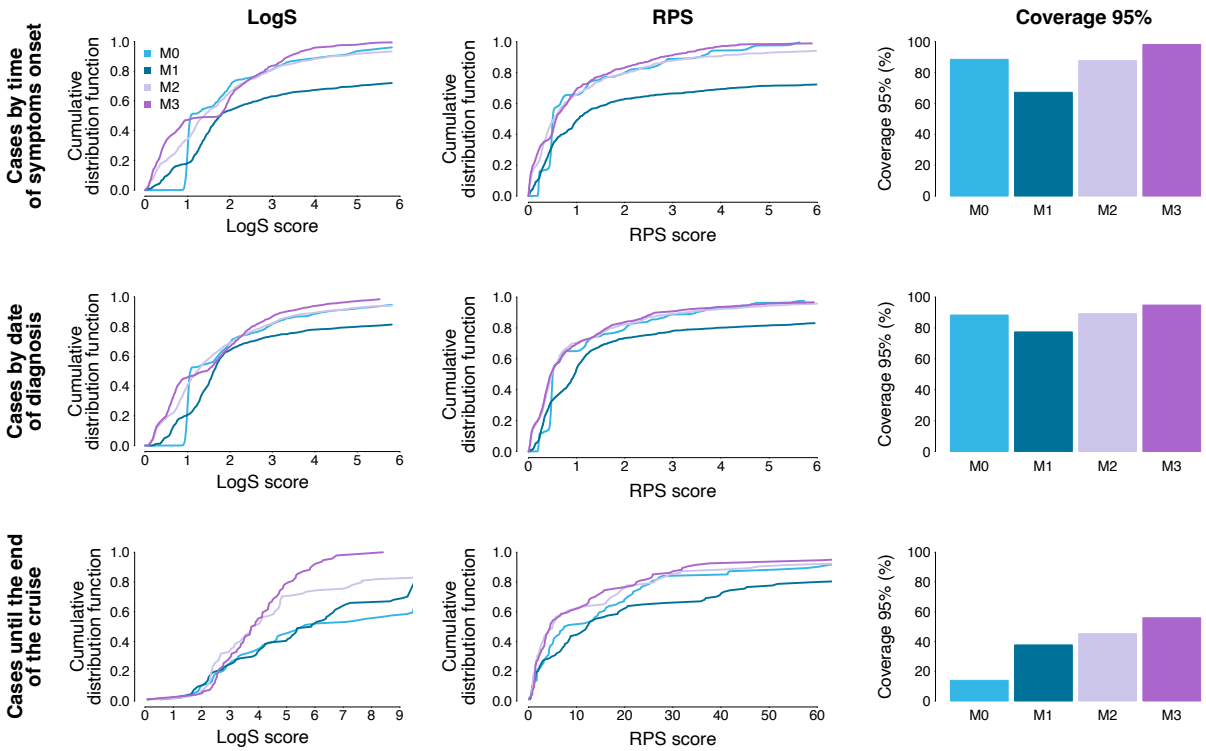


Figure 3.2. *Forecasting scores across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; center column: RPS score; right column: 95% coverage.

To quantify the improvement of each model from the baseline, we computed the proportion of data points where the score was better, equal or worse than the one obtained with the baseline (Figure 3.3). M1 performed significantly worse than the baseline over all scores and endpoints; M2 was substantially equivalent to the baseline on all measures for forecasts by symptom onset, but performed better on forecasts by date of diagnosis and on the total cases until the end of the cruise; M3 performed better than the baseline on all scores and endpoints, having a better score on forecasts by symptom onset (59%with the logS and 53% with the RPS), by diagnosis date (66% with the logS and 61% with the RPS) and by total cases until the end of the outbreak (70% with the logS and 57% with the RPS). We also evaluated the performance of all models compared with each other by computing the standardized rank, a score that takes into account the ranking of each model's score for a given forecast. The standardized rank score is zero for the worst-ranked model and one for the best-ranked one. In Figure 3.4, we show that M3 has the highest mean of the standardized ranks across all outbreaks, times of analysis and data points, for all endpoints and both scores. M2 is consistently the second-best model according to this measure (although very close to M0 for forecasts by symptom

onset), with the advantage of having a generally smaller variability than both M3 and M0.

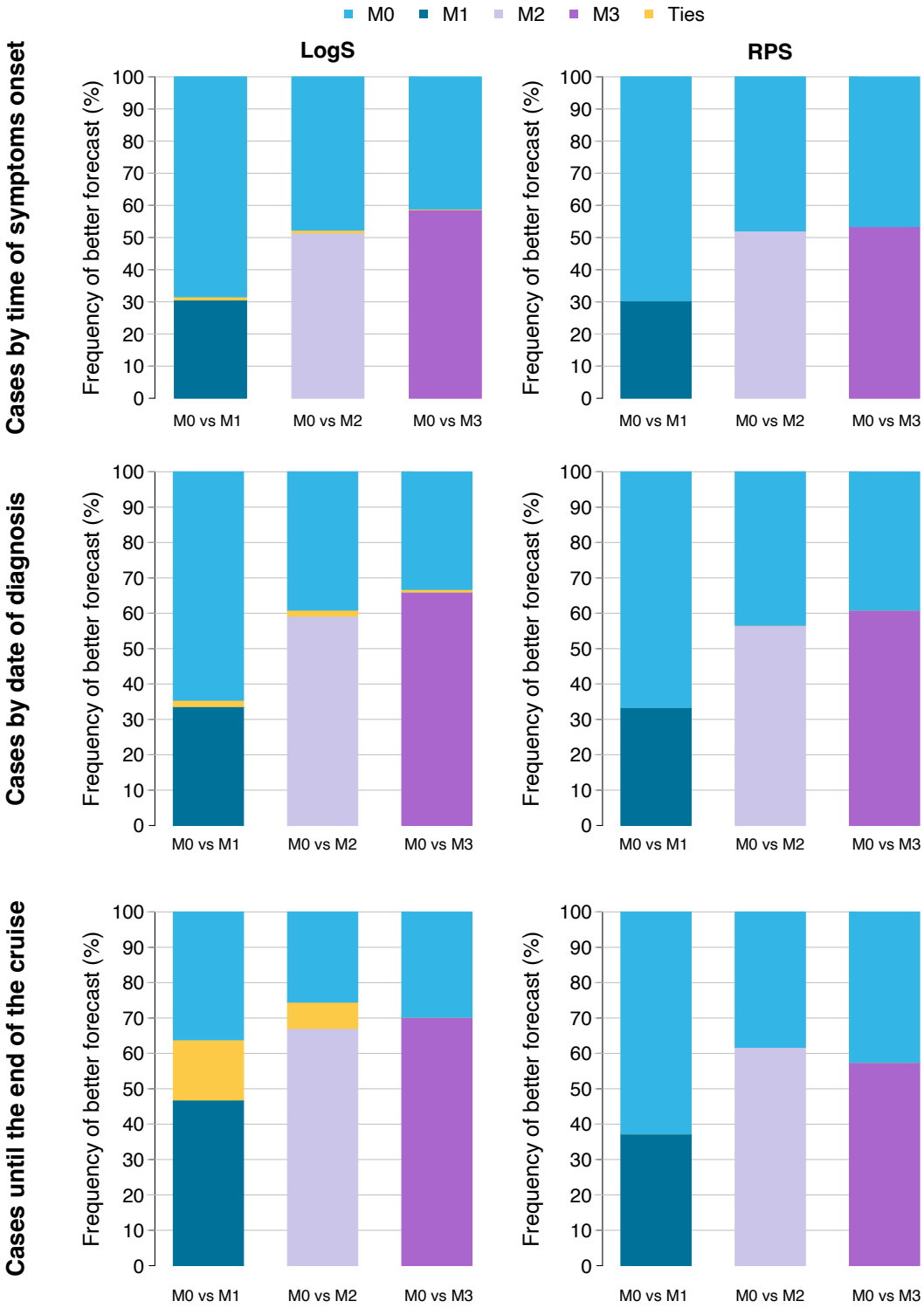


Figure 3.3. *Model performance against the baseline across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. The 100% stacked barcharts show the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (yellow).

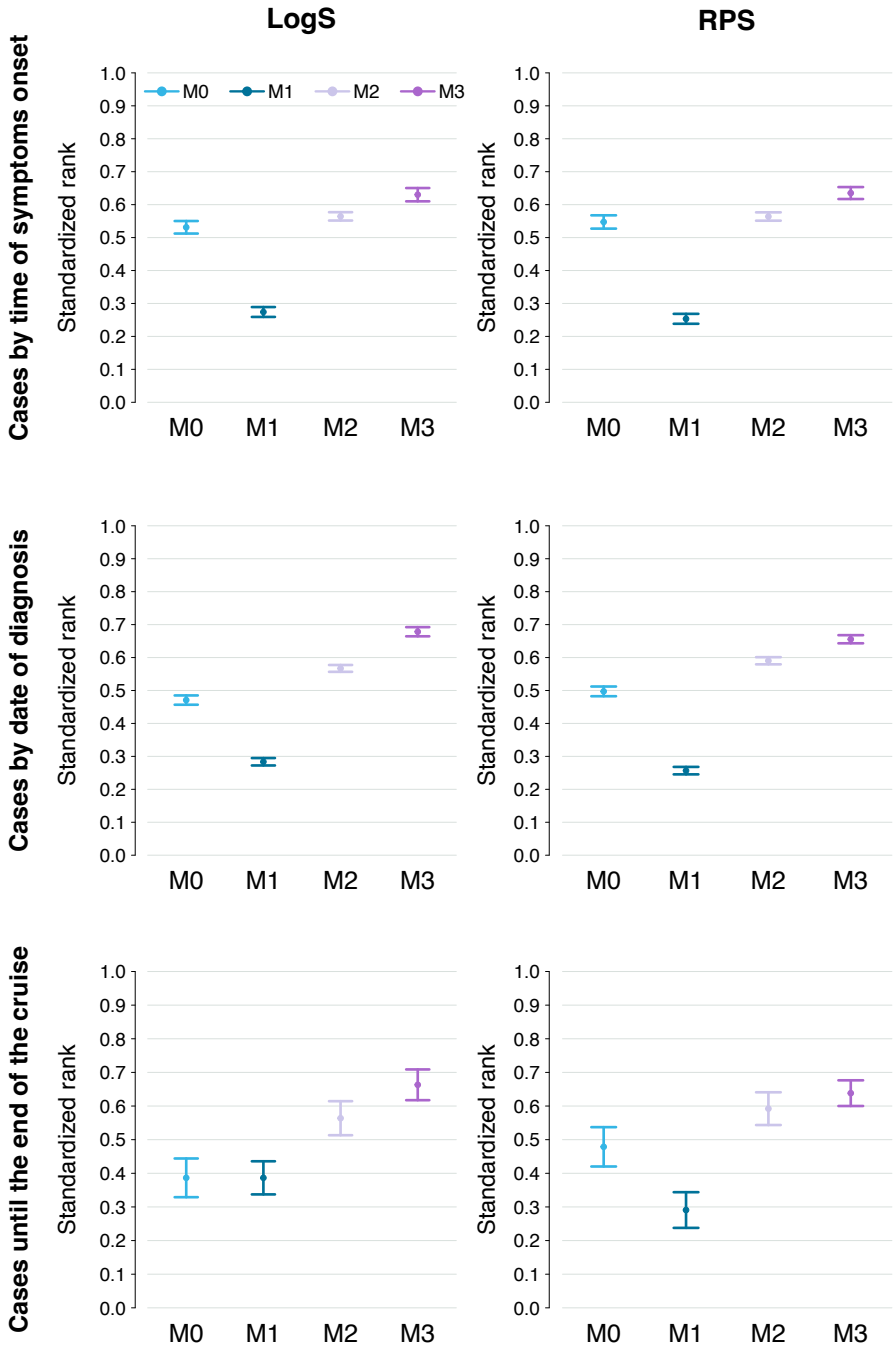


Figure 3.4. *Standardized rank of forecasting models across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by diagnosis date; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. Points: mean standardized rank; whiskers: 95% CI of the mean standardized rank.

Figures 3.5 and 3.6 disaggregate model performances by forecast horizon, i.e. by the distance in the future from the time of analysis at which data points are assessed. All models tended to perform better for longer horizons, with the strongest improvements occurring for M3. Figure 3.6 confirms that the added value of M2 and M3 arises mainly in projections for longer horizons. When

considering cases by symptom onset within a horizon of 24 hours, M2 and M3 tend to be slightly outperformed by the baseline since less than 50% of data points have a better score than M0; however, they are consistently superior for cases occurring later than 48 hours after the date of forecasting, with up to 78% predictions being more accurate (both according to the logS and the RPS) for a forecast horizon occurring later than 72 hours after the analysis. In the intermediate horizon (24-48h), M3 has a similar performance to the baseline, depending on the score measure, while M2 is consistently inferior. When considering the cases by diagnosis date, model M3 is consistently superior to the baseline at all forecasting horizons, again with improved performance for longer-term predictions.

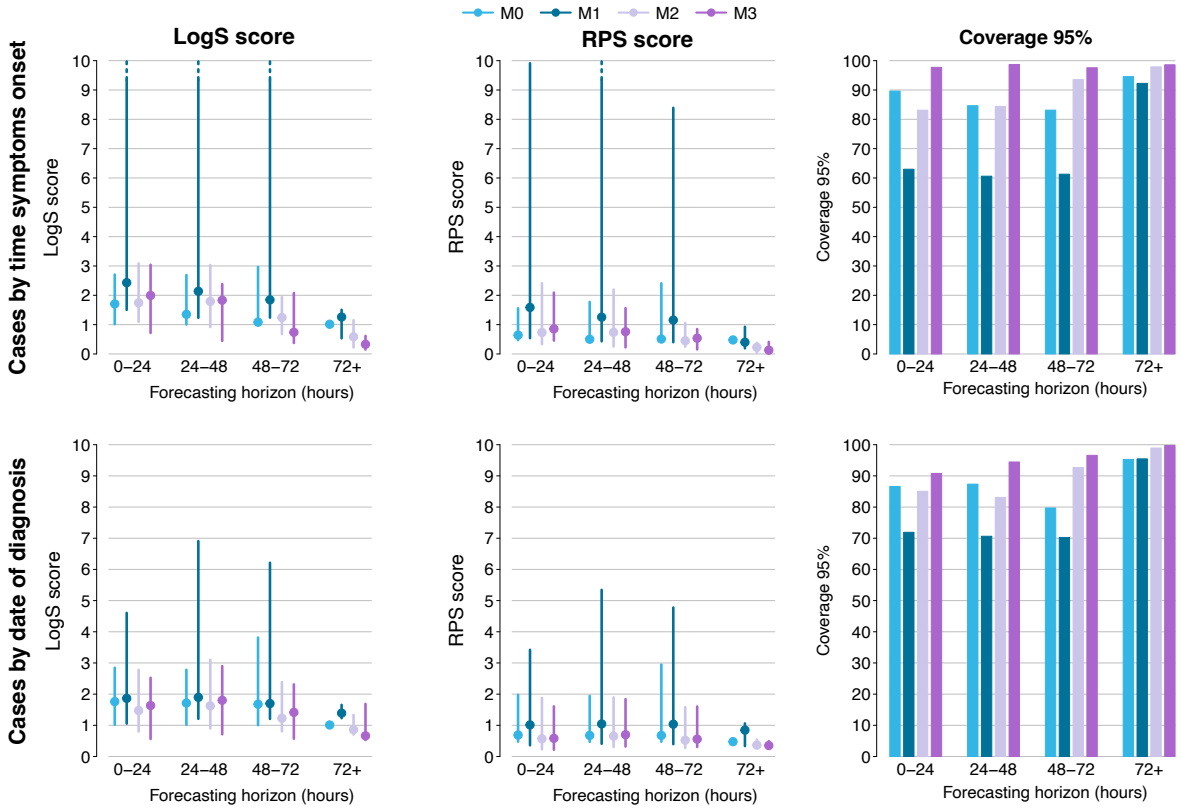


Figure 3.5. *Forecasting scores across all outbreaks and times of analysis, disaggregated by forecasting horizon.* Top row: forecasts by time of symptom onset; bottom row: forecasts by diagnosis date. Left column: logS score; center column: RPS score; right column: 95% coverage. Points: median of the distribution; vertical bars: interquartile range (IQR).

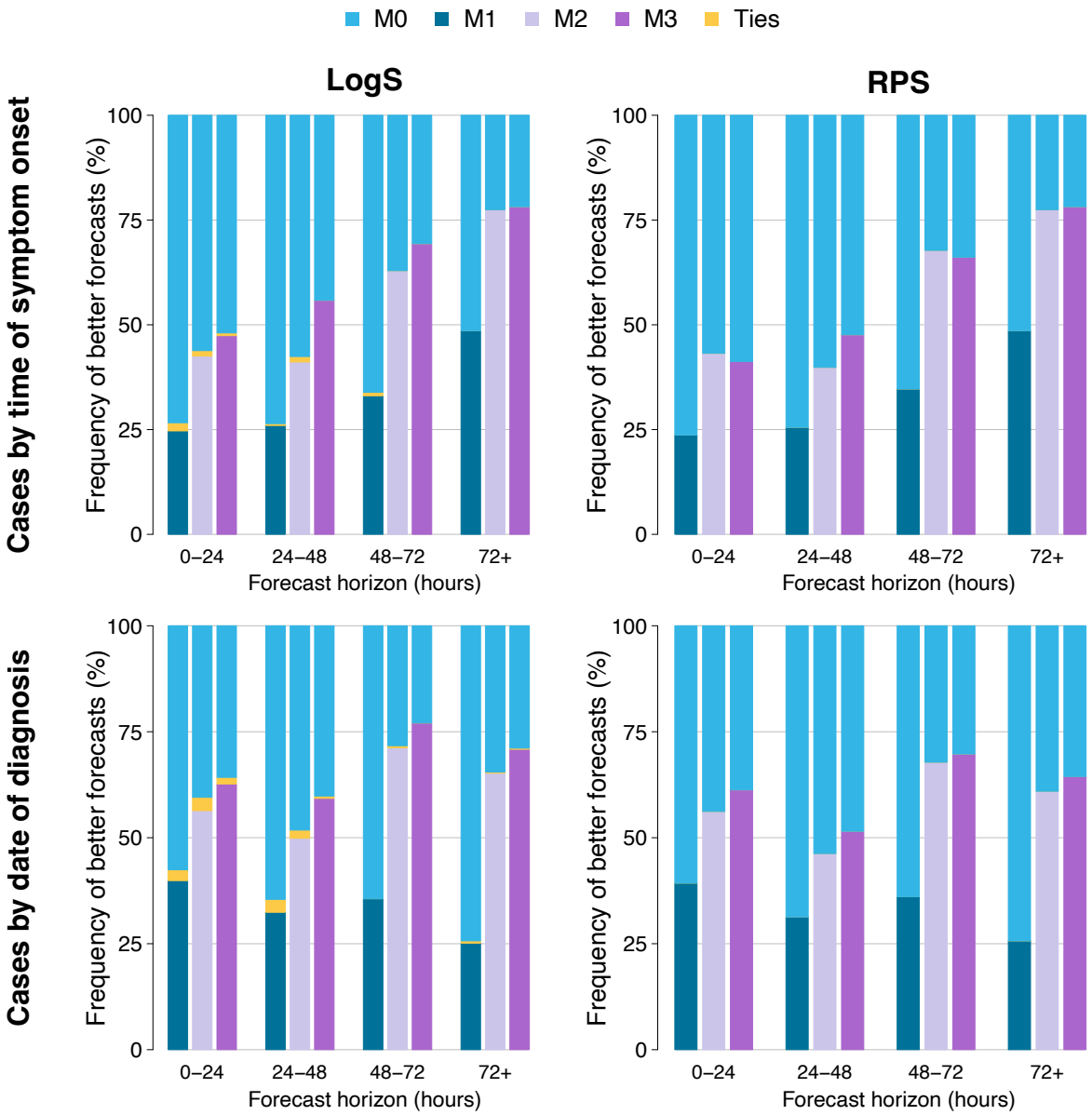


Figure 3.6. *Model performance against the baseline across all outbreaks and times of analysis, disaggregated by forecast horizon.* Top row: forecasts by time of symptom onset; bottom row: forecasts by date of diagnosis. Left column: logS score; right column: RPS score. The graph shows the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (darker shaded areas in between).

3.5. DISCUSSION

In this work, we assessed the real-time forecasting performance of a set of mathematical models for norovirus transmission on cruise ships. The objective of the study was to evaluate the feasibility of introducing automated forecasts within syndromic surveillance systems on board, in order to support decision-making for public health decisions.

We found that simple semi-mechanistic models based on the renewal equation [64] could outperform a baseline model predicting a constant incidence of

future cases. However, this could happen only for models (M2 and M3) that took into consideration the effect of isolation of cases after diagnosis on the individual transmissibility. Indeed, model M1, which was based on the application of the renewal equation under the assumption of a constant transmissibility until the end of the cruise, generally resulted in a massive overestimation of cases (see Figure 3.1 for an example). The best forecasting performance was given by a model that also included in forecasts the heterogeneity in transmission due to superspreading phenomena [138].

The assessment of the models' performance on three forecasting endpoints measured the model's ability to: i) capture the epidemiological dynamics (cases by symptom onset time); ii) provide an indication of the future number of individuals presenting with norovirus to the ship's medical center (cases by diagnosis date); iii) offer an assessment of the future residual outbreak size (cases until the end of the cruise), which may be useful for public health operators to decide whether to further enforce or relax infection control measures. The best performing model was superior to the baseline and to alternative models on all endpoints, according to a set of alternative scores commonly used to evaluate forecasts. The model provided the highest improvements in performance for the number of cases by date of diagnosis and the total number of cases until the end of the cruise. The latter quantity turned out to be much more difficult to gauge correctly for the models: for example, the best model forecasted the correct value within the 95% PrI in only 56% of forecasts (compared to 14% of the baseline model). The difficulty in estimating the total number of cases at the end of the outbreak can be attributable to the propagation of forecasting errors over time, since an over- or under-estimation of the number of cases at a given time point is more likely to bring to further errors in the same direction for successive time points.

The added value of the best model was especially evident for longer forecasting horizons: while forecasts within 48 hours after the time of analysis were substantially equivalent to the baseline model, up to 78% of those for data points beyond 72 hours were more accurate than the baseline. It must be noted, somewhat counterintuitively, that all models improved their performance at longer time horizons. Due to the short duration of the considered cruises (7 days for six out of the seven under study), time horizons longer than 72 hours corresponded in many cases to time points closer to the end of the cruise, when lower numbers of cases were recorded in all outbreaks; this may be due partly to the implemented control measures successfully mitigating the outbreak, and partly to some cases differing healthcare-seeking to after the end of the cruise.

Lower case counts generally result in forecasts with lower variance (resulting in lower logS scores) and lower absolute errors (resulting in lower RPS scores), explaining the improved performance of models for longer time horizons.

We estimated that an isolated case transmitted about 94% less than non-isolated ones. In the considered outbreaks, all cases were isolated in their cabin for 72 hours immediately after diagnosis. This protocol is potentially stricter than the standards followed by the cruise ship, according to which the recommended isolation period is 24 hours after the resolution of the symptoms [129], [130]. Considering an average incubation period of 35 hours [139], an average diagnostic delay of 17 hours (estimated from outbreak data), and an average generation time of 89 hours [135], an isolation of 72 hours corresponds in practice to isolation until the end of the infectious period, explaining the high estimated efficacy of the isolation protocol. The residual transmissibility of isolated cases may be mostly attributed to transmission to cabin mates or to a limited lack of compliance to isolation. The estimated value was remarkably robust across the seven outbreaks (see Section 3.6). Different cruise ships may have different isolation protocols (e.g., isolation in separate cabins rather than with cabinmates, and/or for different durations) or compliance, requiring a specific estimate of its effectiveness and limiting the generalizability of these forecasting models. The overdispersion parameter for the best performing model has not been thoroughly estimated for norovirus transmission on cruise ships; however, a sensitivity analysis on this value revealed that models considering different degrees of superspreading were still robustly superior to the baseline (Section 3.6). The distribution of diagnostic delays may also differ from cruise to cruise depending on passenger and crew education, communication campaigns, or potential logistic or economic barriers in the access to medical facilities on board [140]. However, in an automated forecasting system on board, the distribution of diagnostic delays can be easily estimated and continuously updated using routinely collected data within the syndromic surveillance system of the ship. In the data used in this work, we did not find evidence of changes over time in diagnostic delays, justifying the use of a constant distribution throughout the study.

Although syndromic surveillance of infectious diseases on board cruises considers the passengers and crew members as different populations, in this analysis we considered cases from the two populations together. A separated analysis by population was not possible with available data because the number of cases in crew was always very small (Table 3.1). We were also not able to include in the forecasts the effect of other interventions such as the disinfection

of common surfaces or the suspension of self-service restaurants. Quantitative estimates of the effectiveness of these interventions are not available and are an open field of investigation. Finally, we did not comprehensively address underreporting of cases, which was estimated at 23% and 40% in previous epidemiological studies [82], [141]. Underreported cases due to postponed health-seeking to after the end of the cruise was included in the model forecasts as they participated to transmission onboard but were not included in the forecasted endpoints. However, a residual fraction of cases that may not seek for medical care at all can alter the transmission dynamics and potentially reduce the predictive performance.

Real-time forecasting of infectious diseases is an intrinsically difficult and imperfect task, since models cannot integrate all possible sources of heterogeneity in transmission and control interventions. Similarly to many other forecasting efforts conducted on open populations [31], [50], [82], [132]-[133], [142]-[144], we found that forecasts provided under extremely simplistic assumptions (e.g., constant incidence until the end of the cruise) are not so easily outperformed even in small, semi-closed populations. However, in this work we have shown that simple semi-mechanistic models including the appropriate epidemiological hypotheses can provide useful estimates on the future number of norovirus cases on cruise ships with minimal computational requirements for calibration and execution. These results represent a proof-of-concept for building automated real-time forecasting systems that can enhance epidemic intelligence on board, providing public health officers with better information to support decision making during outbreaks.

3.6. APPENDIX

DETAILS ON FORECASTING METHODOLOGY

OVERVIEW

The main goal of this study was to retrospectively assess the performance of real-time forecasts for norovirus outbreaks on board of cruise ships, considering a set of alternative models. Figure 3.7 gives an overview of the pipeline of analysis, and Table 3.2 summarizes the mathematical notation and description of variables and parameters used in the manuscript. Starting from the line list of cases Z, we first perform preliminary analyses (top part of Figure 3.7) to estimate quantities that will be used for real-time forecasting and for its assessment. In particular, we compute the ground truth for the three endpoints on which forecasts will be assessed, i.e., the total number of cases at the end of each cruise, S, the final epidemic curve by date d of diagnosis $Q(d)$ and by time t of symptom onset, $C(t)$. We also estimate two parameters (green squares) which are critical for the forecasting pipeline: the distribution of diagnostic delays π, and the reduction α of transmissibility among diagnosed and isolated cases. The forecasting pipeline is executed at each time of analysis t_A (bottom part of Figure 3.7), starting from X_{t_A}, the subset of the line list Z which was available at time t_A. The below forecasting pipeline section describes methodological details of each element of the pipeline, which are briefly outlined here. From X_{t_A} we obtain a two-dimensional time-series of the number of cases that have had symptom onset at time t and have been diagnosed before time τ, $D_{t_A}(t, \tau)$, for $t, \tau \leq t_A$. The presence of diagnostic delays implies that there may be an unknown number of cases who have already had symptom onset before t_A but will be diagnosed after t_A; the missing information on these cases may lead to an underestimate of individuals who can further transmit the disease. Therefore, we use a nowcasting technique to adjust for this effect, obtaining a nowcasted time series of cases by time of symptom onset, $C^n_{t_A}(t)$, and an estimate of the number of cases $N_{t_A}(t, \tau)$ that have developed symptoms at time t but have not been diagnosed before time $\tau > t_A$. Note that, due to intrinsic limitations in the nowcasting, $C^n_{t_A}(t)$ and $N_{t_A}(t, \tau)$ are considered reliable only until a time of symptom onset $t_F < t_A$. We use $C^n_{t_A}(t \leq t_F)$ to compute an estimate of the reproduction number $\tilde{R}$ that will be used as an input parameter for all forecasting models M1-M3, excluding the baseline. Then, $C^n_{t_A}(t \leq t_F)$ is used as an input for model M1 and $D_{t_A}(t \leq t_F, \tau)$ and $N_{t_A}(t \leq t_F, \tau)$ are used as inputs for models M2, and M3, to start the forecasting. At each forecasting time-step comprised between t_F and t_A, the stochastic forecasts

$C^f_{t_A}(t,\tau)$ produced by each model are iteratively compared against corresponding observations available in $D_{t_A}(t,\tau)$ and the realizations who reproduce most closely available data are selected using a particle filtering algorithm. The final $C^f_{t_A}(t,\tau)$ resulting from this iterative procedure are aggregated by time of symptom onset and date of diagnosis to obtain the final forecasts $\hat{C}^f_{t_A}(t > t_A)$, $\hat{Q}^f_{t_A}(d > d_A)$ (where d_A is the day of analysis), and the total number of cases forecasted between t_A and the end of the cruise, $\hat{S}^f_{t_A}$. Finally, the three forecasting endpoints are scored with the corresponding ground truths estimated in the preliminary analyses, according to three metrics (the log score, logS, the Ranked Probability Score, RPS, and the 95% coverage, Cov95%).

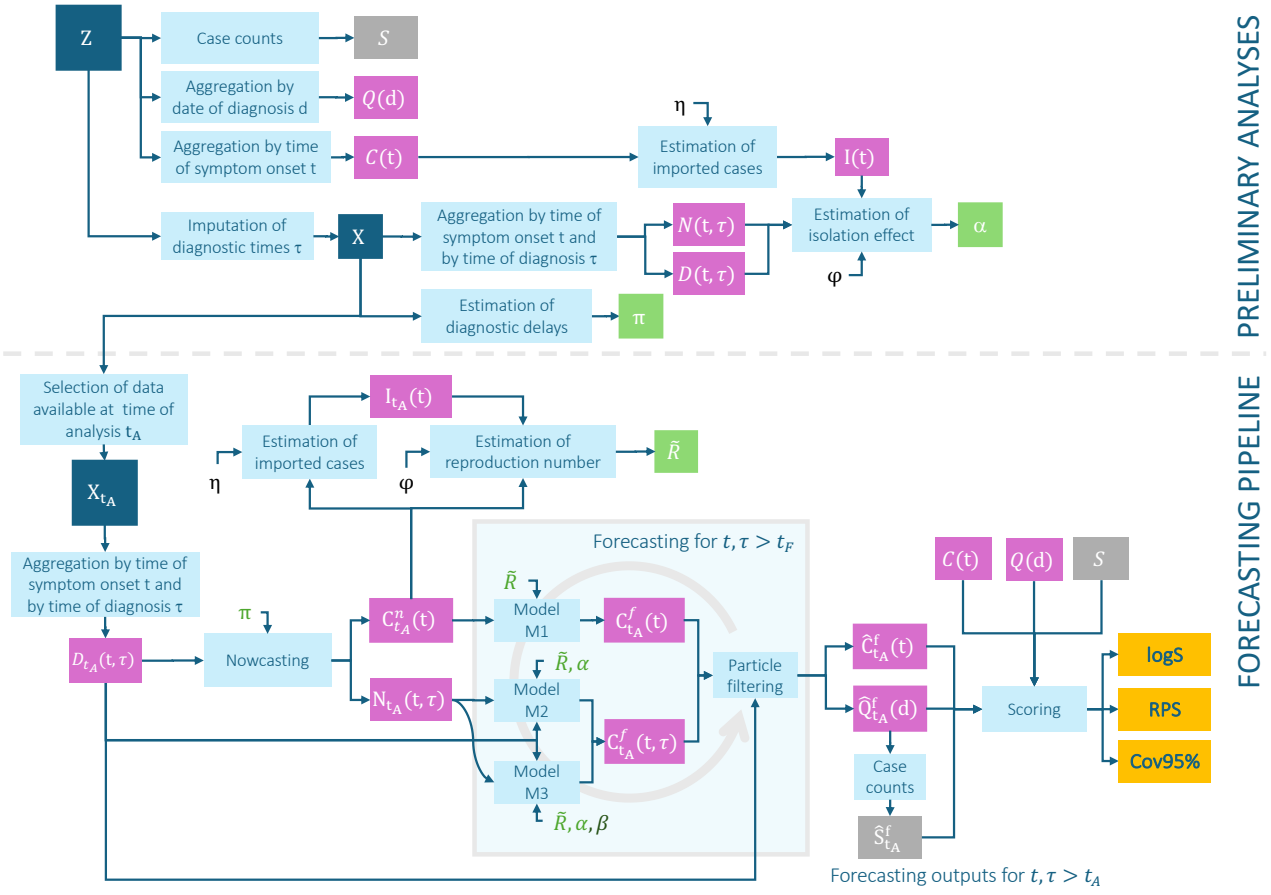


Figure 3.7. *Workflow diagram for the forecasting methodology.* Blue boxes represent available data in the form of line lists; purple boxes represent one- or two-dimensional time-series (by time of symptom onset t, by time of diagnosis τ or by date of diagnosis d; gray boxes represent total number of cases; green boxes represent parameters estimated within this workflow; light blue boxes represent operations on inputs; the larger square with gray boundaries and light blue background represents the iterative relationship between stochastic forecasting through models M1-M3 and selection through particle filtering. Variables that are inputs for operations represent parameters, which were either obtained from the literature (printed in black) or within the workflow (printed in green); yellow boxes represent the scores for the forecasting endpoints which are used to evaluate the performance of the model.

Table 3.2. *Notation and description of variables and parameters used in the manuscript.*

NOTATION	DESCRIPTION OF VARIABLE
t_A	Time of analysis at which forecasts are produced
t_F	Time until the number of reported cases is considered incomplete
d_F	Width of the "incompleteness window" (where reported cases are considered incomplete)
Z	Line list of cases by time of symptoms onset and by day of diagnosis
X	Line list of cases by time of symptoms onset and by time of diagnosis
X_{t_A}	Line list of cases by time of symptoms onset and by time of diagnosis, as available at time of analysis t_A
S	Total number of cases at the end of the cruise
$Q(d)$	Epidemic curve by date of diagnosis d at the end of the cruise
$C(t)$	Epidemic curve of cases by time of symptoms onset t at the end of the cruise
$I(t)$	Epidemic curve of imported cases by time of symptoms onset t estimated at the end of the cruise
$N(t,\tau)$	Number of cases with time of symptoms onset t that are not yet diagnosed at time τ
$D(t,\tau)$	Number of cases with time of symptoms onset t that were diagnosed before time τ
$N_{t_A}(t,\tau)$	Number of cases with time of symptoms onset t that are not yet diagnosed at time τ, available at time of analysis t_A
$D_{t_A}(t,\tau)$	Number of cases with time of symptoms onset t that were diagnosed before time τ, available at time of analysis t_A
$I_{t_A}(t)$	Epidemic curve of imported cases by time of symptoms onset t, available at time of analysis t_A
$C^n_{t_A}(t)$	Nowcasted time series of cases by time of symptom onset t, available at time of analysis t_A
$C^f_{t_A}(t)$	Provisional forecasted time series of cases by time of symptom onset t, before particle filtering
$C^f_{t_A}(t,\tau)$	Provisional forecasted time series of cases by time of symptom onset t that were diagnosed before time τ, before particle filtering
$\widehat{C}^f_{t_A}(t)$	Final forecasted time series of cases by time of symptom onset t at time of analysis t_A
$\widehat{Q}^f_{t_A}(d)$	Final forecasted time series of cases by date of diagnosis d, at time of analysis t_A
$\widehat{S}^f_{t_A}$	Final forecasted total number of cases at the end of the cruise, at time of analysis t_A

α	Reduction in transmissibility of norovirus for cases isolated in cabins
π	Cumulative probability distribution of the diagnostic delay
η	Cumulative probability distribution of the incubation time
φ	Discretized probability distribution function of the generation time
$\widetilde{R}$	Estimate of the reproduction number
ν	Overdispersion parameter in the Negative Binomial distribution
$logS$	Logarithmic Score
RPS	Ranked Probability Score
$Cov95\%$	95% Coverage

PRELIMINARY ANALYSIS

DIAGNOSTIC DELAY DISTRIBUTION

We define the diagnostic delay as the interval between the onset of symptoms and diagnosis of a case. Since available data only included the day of diagnosis, not the specific time, we imputed for each case a diagnostic time, assuming that diagnosis occurred uniformly during daytime (between 8:00 and 23:59 of each day) and only after the appearance of symptoms (given that no norovirus screening was conducted aboard the cruise ship). The imputation process was repeated NA=50 times, and the distributions of diagnostic delays were pooled together across the NA samples. Except for cruise A, where the first symptomatic cases had clearly longer diagnostic delays, in all other cruises and in the pooled distribution across all cruises, diagnostic delays were homogeneously distributed over time (Figure 3.8).

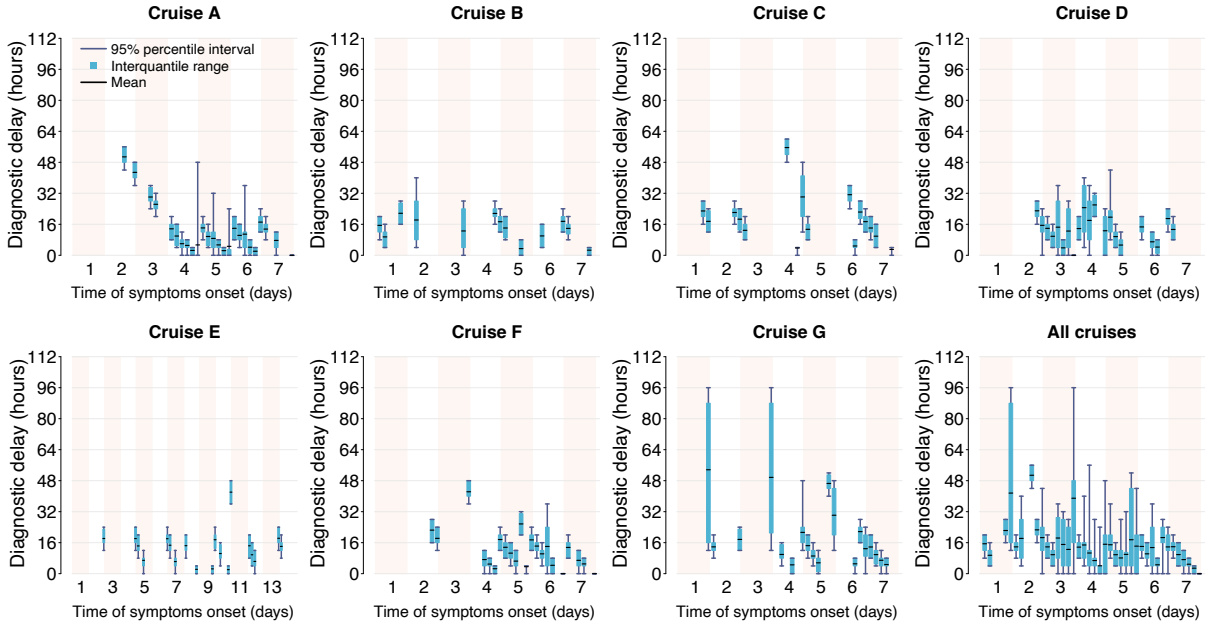


Figure 3.8. *Distributions of the diagnostic delay by cruise ships.* x-axis: time of symptoms onset by day, y-axis: distribution of the diagnostic delay, aggregating cases by time steps of 4 hours. Horizontal black line: mean; light blue rectangles: interquartile range; whiskers: 95% percentile interval of the distribution. First row: cruise ships A, B, C and D, second row: cruises ship E, F, G and aggregation across all cruise ships.

The final diagnostic delay distribution was obtained by pooling all cases with time of symptom onset before or on day 5, since diagnostic delays for cases with time of symptom onset on days 6 and 7 were strongly affected by right truncation (Figure 3.9). The resulting cumulative distribution π is shown in Figure 3.10; the mean of the diagnostic delay was 17 hours with 95% percentile interval 0-64 hours.

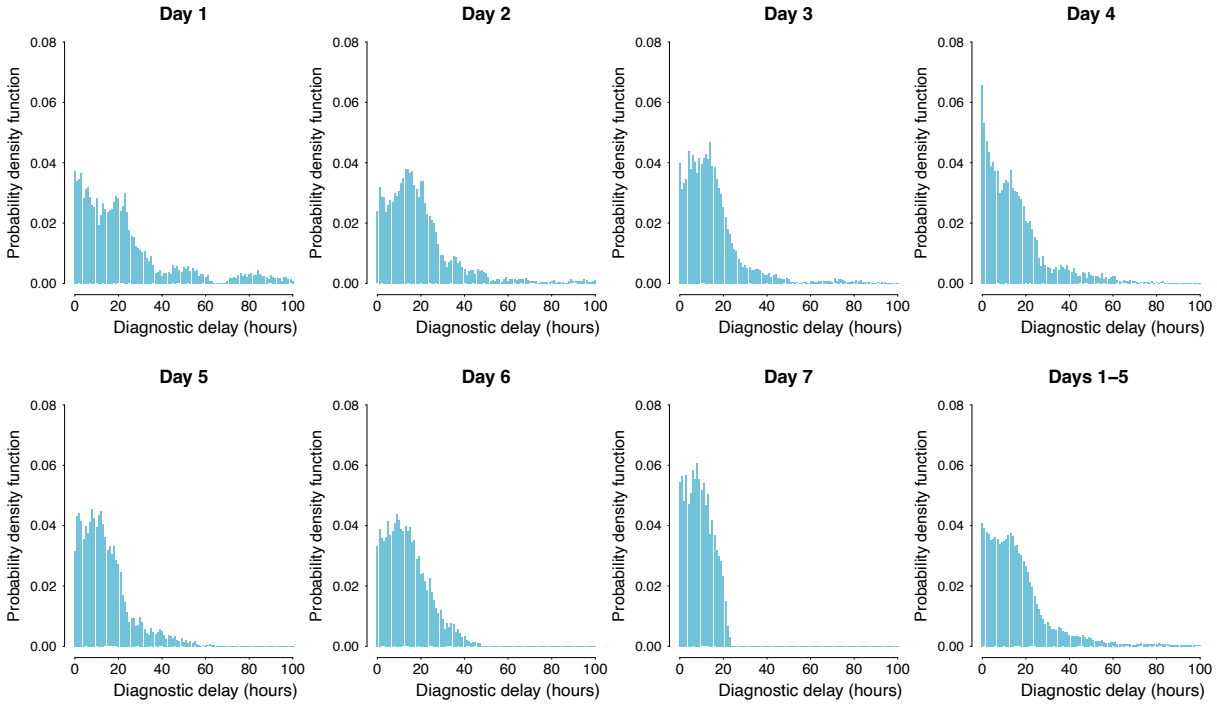


Figure 3.9. *Estimated distributions of diagnostic delay, aggregated by day of symptom onset and across all cruises.* First row: symptoms onset on day of travel 1, 2, 3 and 4, second row: symptoms onset on day of travel 5, 6, 7, and pooled distribution aggregating cases with symptom onset on days 1-5.

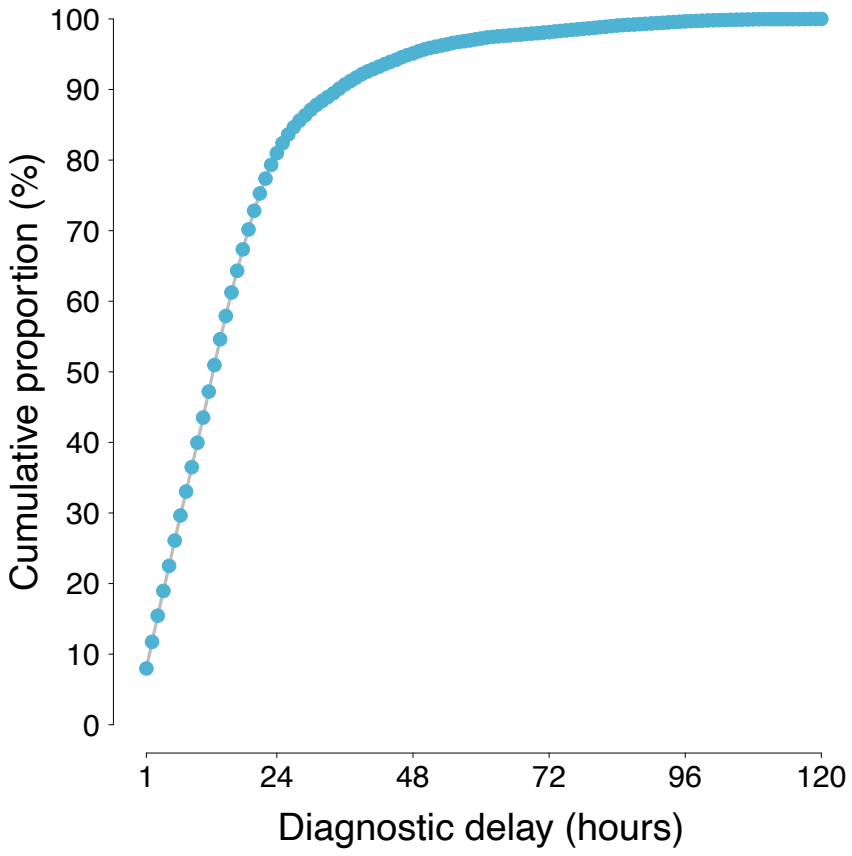


Figure 3.10. *Cumulative distribution of the diagnostic delay, π*

TRANSMISSIBILITY REDUCTION AFTER ISOLATION IN CABINS

In the considered cruises, whenever an infected individual was diagnosed with gastrointestinal illness (GI), they were immediately isolated in their cabin for 72 hours. We estimated the reduction in transmissibility associated to isolation by considering a modified version of the renewal equation [64] that considers isolated and non-isolated infectors separately. Let $D(t,\tau)$ represent the number of cases with symptoms onset t diagnosed and isolated before time τ and $N(t,\tau)$ the number of cases with symptoms onset t still undiagnosed at time τ. Then, the number of observed cases over time is modeled as:

$$X(t;R(t),\alpha) = R(t)\cdot\sum_{s=1}^{t-1}[N(t-s,t)+(1-\alpha)D(t-s,t)]\,\varphi(s)$$

Where $R(t)$ represents the time-varying reproduction number, α is the reduction in transmissibility of norovirus for patients isolated in cabins and $\varphi(s)$ is the probability distribution function of the generation time, discretized by 4-hours timesteps, at timestep s. $\varphi(s)$ was estimated in previous studies to be a gamma distribution with mean 88 hours and 95% percentile interval 22-205 hours [136].

For each cruise, we estimated $R(t)$ and the value of α, using the MCMC Metropolis-Hastings method with N_{MH} = 100,000 iterations and a likelihood L given by:

$$L = \prod_t Pois(C(t)-I(t)\,;\,X(t;R(t),\alpha))$$

where $Pois(x;\lambda)$ represents the Poisson probability of observing x cases when the rate is λ, $C(t)$ is the total number of cases with symptom onset t observed at the end of a given cruise and $I(t)$ is the number of cases with symptom onset t that were not transmitted onboard. Neglecting cases that are infected off board during visits at ports [63], we consider that imported cases are those who were infectious before first embarkation on the ship. Therefore, $I(t)$ can be estimated as a time-varying proportion $\xi(t)$ of the observed cases:

$$I(t) = \lfloor \xi(t)\cdot C(t) \rfloor = \lfloor (1-\eta(t))\cdot C(t) \rfloor$$

where $\xi(t) = 1-\eta(t)$ is the probability of having had an incubation period longer than the time of symptom onset t, $\eta(k)$ is the cumulative probability distribution of the incubation time k and $\lfloor y \rfloor$ represents the rounding to the

closest integer of y. If $I(t)$ was identically zero in a given cruise, we assumed that only one case with earliest symptom onset was an imported case.
The posterior distributions of α estimated for each cruise ship (Figure 3.11) were roughly similar, with average values always exceeding 90%. For the rest of the manuscript, we considered a distribution of α given by the pooled posterior distributions of the cruise-specific estimates, obtaining a mean of 94.3% with 95% credible interval (CrI): 74.9-99.9%.

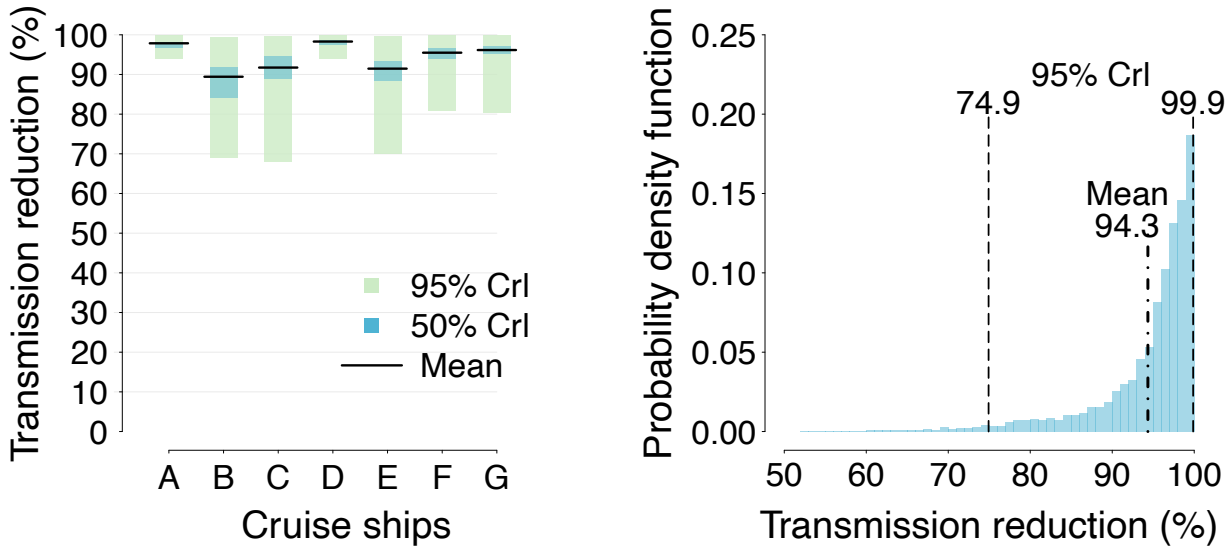


Figure 3.11. *Effect of isolation in cabin.* Left: boxplots of the posterior distributions for the transmissibility reduction provided by cabin isolation α estimated in the different cruises. Horizontal black lines: mean; light blue rectangles: 50% credible intervals, CrI; light green rectangles: 95% CrI. Right: pooled probability distribution of α across all cruises.

FORECASTING PIPELINE

In the perspective of implementing forecasts on electronic surveillance systems on board, we compared the predictive performance of a set of mathematical models that do not require a complex calibration to data.
We use a baseline model M0, relying on syndromic surveillance data with minimal preprocessing, to benchmark alternative models. The baseline model assumes that the trend in the number of cases observed over the last $W = 24$ hours will remain stable in the future. Let $C_{t_A}(t)$ be the observed time series of cases aggregated by time of symptom onset t, as available at time of analysis t_A, and $Q_{t_A}(d)$ the corresponding time series aggregated by date of diagnosis d. Then, the forecasted time series $\hat{C}_{t_A}(t > t_A)$ and $\hat{Q}_{t_A}(t > t_A)$ are obtained as:

$$\hat{X}_{t_A}(t > t_A) \sim Poisson\left(\frac{1}{W}\sum_{s=t_A-W}^{t_A} X_{t_A}(s)\right)$$

Where $Poisson(\lambda)$ is a random sample from a Poisson distribution of rate λ and $X = \{C, Q\}$ determines the type of output to be forecasted. As for all other forecasting models, Z=1,000 stochastic repetitions are run and the results from the Z simulations are pooled together to provide the forecasted distribution at each time of forecast.
Alternative models were all based on variations of the renewal equation [64], to keep low the computational demand, but leverage a complex pipeline of automated pre- and post-processing analyses that is outlined in Figure 3.7 and for which we describe the building blocks in detail below.

NOWCASTING EPIDEMIC CURVES

Forecasts for models alternative to the baseline are based on epidemic curves computed by time of symptom onset, as the symptom onset is the measurable event that is closest to the actual infectious episode. When considering real-time data, it is important to consider that diagnostic delays may bias downwards the number of cases reported to have symptoms close to t_A. Indeed, some of the cases who had symptom onset before t_A may be reported after this date. Nowcasting allows to partially compensate for this bias by leveraging information on the distribution of diagnostic delays [42], [48]. In this case, we adopted the technique proposed in (4) to $C_{t_A}(t) = D_{t_A}(t, \tau)$, in order to estimate a nowcasted time series $C^n_{t_A}(t)$ as follows:

$$C^n_{t_A}(t) = \left\lfloor \frac{1}{\pi(t_A - t)} \cdot C_{t_A}(t) \right\rfloor$$

Cases with symptom onset before t_A but diagnosed after t_A are given by $U_{t_A}(t) = C^n_{t_A}(t) - C_{t_A}(t)$ and were stochastically assigned a time of reporting by drawing a sample δ of diagnostic delay from π after constraining $t + \delta > t_A$, obtaining the two-dimensional time series of undiagnosed cases $N_{t_A}(t, \tau)$.
For values of t close to t_A, the number of reported cases $C_{t_A}(t)$ may become too small and the factor $\frac{1}{\pi(t_A - t)}$ too large; therefore the values of $C^n_{t_A}(t)$ and consequently of $N_{t_A}(t, \tau)$ may be poorly compensated. Therefore, we consider the nowcasted estimates as stable only until a time $t_F = t_A - d_F$. In the main analysis we choose d_F =24 hours and we run sensitivity analysis on the choice of this value.

ESTIMATION OF THE REPRODUCTION NUMBER

The time-varying reproduction number $R_{t_A}(t)$ at time of analysis t_A can be computed using the method proposed by [34], knowing the distribution of the generation time φ, the nowcasted number of cases by time of symptom onset $C^n_{t_A}(t)$, and the number of cases transmitted offboard and imported on the ship $I_{t_A}(t) = \lfloor (1 - \eta(t)) \cdot C^n_{t_A}(t) \rfloor$

We thus estimated $R_{t_A}(t)$ at each time of analysis t_A, by applying Markov Chain Monte Carlo with Metropolis- Hasting sampling (NMH = 100,000 iterations) to the likelihood function L^R below:

$$L^R = \prod_{t=1}^{t_F} Pois\left(C^n_{t_A}(t) - I_{t_A}(t); R_{t_A}(t) \sum_{s=1}^{t-1} \varphi(s) C^n_{t_A}(t-s) \right)$$

We used the estimate of $\tilde{R}_{t_A} = R_{t_A}(t = t_F)$, i.e., the most recent stable estimate with respect to t_A, as an input to the different forecasting models, which are described below.

FORECASTING MODELS

We considered three alternative models, based on variations of the renewal equation [64]. The renewal equation takes as input a time-series of cases by date of symptom onset and the values that the reproduction numbers is expected to assume in the future.

Model M1 projects cases as follows:

$$C^f_{t_A}(t) \sim Poisson\left(\rho_{t-1} \tilde{R}_{t_A} \sum_{s=1}^{t-1} C^f_{t_A}(s)\, \varphi(t-s) \right)$$

Where, for $t \leq t_F$, $C^f_{t_A}(t) = C^n_{t_A}(t)$ corresponds to the time series by time of symptom onset nowcasted at time t_A, and, for $t > t_F$, $C^f_{t_A}(t)$ corresponds to the time series of forecasted cases; in other words, forecasts are provided iteratively for the successive forecasting step based in part on forecasts from the previous step. ρ_t represents the proportion of susceptible individuals on board at time t, given by

$$\rho_{t-1} = 1 - \frac{\sum_{s=1}^{t-1} C^f_{t_A}(s)}{N_c}$$

and N_c is the total population of cruise ship c. In practice, this model assumes that the underlying transmissibility will change over time only due to the depletion of susceptible individuals.
In model M2, we consider the heterogeneous transmissibility of undiagnosed and isolated individuals, under the assumption that diagnosed individuals become immediately isolated in their cabins (as per protocols of the considered cruise ships):

$$C^f_{t_A}(t) \sim Poisson\left(\rho_{t-1}\tilde{R}_{t_A}\sum_{s=1}^{t-1}\left[N^f_{t_A}(s,t-1)+(1-\alpha)\,D^f_{t_A}(s,t-1)\right]\varphi(t-s)\right)$$

where $N^f_{t_A}(t,\tau)$ is the two-dimensional time-series of cases with symptom onset at time t that have not yet been diagnosed by time τ and therefore transmit at the full rate, given by $\rho_{t-1}\tilde{R}_{t_A}$, as in model M1; and $D^f{}_{t_A}(t,\tau)$ is the two-dimensional time-series of cases with symptom onset at time t that have been diagnosed before time τ and therefore transmit at a rate reduced by parameter α. Similarly to the previous case of model M1, $N^f_{t_A}(t,\tau)$ and $D^f_{t_A}(t,\tau)$ are given, for $t \le t_F$, by $N_{t_A}(t,\tau)$ and $D_{t_A}(t,\tau)$. For $t \ge t_F$, every case in $C^f_{t_A}(t)$ is stochastically assigned a time of reporting by drawing a sample δ of diagnostic delay from π, obtaining an update for the two-dimensional time-series of cases undiagnosed and diagnosed by time τ, $N^f_{t_A}(t,\tau)$ and $D^f_{t_A}(t,\tau)$.
In model M3, we consider similar transmission dynamics as in model M2, except for the assumption that secondary cases are distributed according to a negative binomial with overdispersion ν [63]:

$$C^f_{t_A}(t) \sim NegBin\left(\rho_{t-1}\tilde{R}_{t_A}\sum_{s=1}^{t-1}\left[N^f_{t_A}(s,t-1)+(1-\alpha)\,D^f_{t_A}(s,t-1)\right]\varphi(t-s)\,,\nu\right)$$

At each stochastic iteration, the values of $\tilde{R}_{t_A}$ and α are sampled from the posterior distribution estimated with the corresponding MCMC procedures; for model M3, the values of ν were sampled from a Gaussian distribution with mean 0.4 and standard deviation 0.125, following estimates on cruise A from a recent study [63]. However, other values for the overdispersion parameters were assessed in sensitivity analyses.

PARTICLE FILTERING

Because we consider the nowcasted epidemic curve to be reliable until t_F, we project cases not only for $t > t_A$, but also for the period $t_F < t \le t_A$. However, during this period the observations that we have, although partial, can be used

to filter stochastic projections which more closely reproduce the observations. For any given forecasting time step $t_F < \theta \leq t_A$, M = 10,000 stochastic forecasts ("particles") are actually drawn from models M1-M3; for each particle m, we compute the total absolute error $E_{t_A}^m$ between forecasts and the corresponding observation:

$$E_{t_A}^m = \begin{cases} \sum_{s=t_F}^{\theta} \left| C_{t_A}(s) - C_{t_A}^{f,m}(s) \right| & for \quad model \quad M1 \\ \sum_{s=t_F}^{\theta} \sum_{u=t_F}^{s} \left| D_{t_A}(u,s) - D_{t_A}^{f,m}(u,s) \right| & for \quad models \quad M2, M3 \end{cases}$$

Z particles are then selected from the M forecasts by sampling with probability directly proportional to the inverse of $E_{t_A}^m$, while the remaining forecasts are filtered out. These are then re-sampled with replacement M times with the same renormalized probabilities, to use as a starting point for the next forecasting step. By iteratively filtering particles and forecasting the next time step, we eventually select Z trajectories at forecasting step $\theta = t_A$, after which forecasts proceed normally according to the respective renewal equation model.

ENDPOINTS COMPUTATION

All models forecasted directly the first endpoint considered in this study, i.e. the time series of cases by time of symptom onset, $\hat{C}_{t_A}^f(t > t_A) = C_{t_A}^f(t > t_A)$. The total number of cases until the end of the cruise $\hat{S}_{t_A}^f$ was computed as $\hat{S}_{t_A}^f = \sum_{s=t_A}^{t_{end}} C_{t_A}^f(s)$, where t_{end} is the ending time of the cruise. The time-series of cases by date of diagnosis d, $\hat{Q}_{t_A}^f(d \geq d_A)$, was obtained for model M1 by stochastically assigning, for each forecasted case in $C_{t_A}^f(t)$, a diagnostic delay δ drawn from π and then aggregating diagnoses over days. For models M2 and M3, it was computed as $\hat{Q}_{t_A}^f(d) = \sum_s D_{t_A}^f(s,d)$. The three endpoints computed by the baseline model M0 and by models M1-M3 were each assessed against observed values at the end of the cruise, according to three score metrics (the logarithmic score, the Ranked Probability Score and the 95% Coverage), as defined in Section 3.3.

SENSITIVITY ANALYSIS

ADDITIONAL MODELS

In addition to the three models presented in the baseline analysis, we considered five additional model variants. M4 and M5 are equivalent to M3 but

assume a different distribution for the overdispersion parameter ν. In particular, for model M4 we assumed a more overdispersed distribution of secondary cases (with a mean value for ν of 0.1 and a standard deviation of 0.0375), and for model M5 a less overdispersed one (mean value for ν of 1 and standard deviation of 0.125). Models M6, M7 and M8 were obtained as the ensemble of projections of model M2 with M3, M4 and M5, respectively.

Overall, the sensitivity analysis shows that all additional models produced similar results to M3, with slight variations on the best performing model depending on the considered endpoint and score metrics (Figures 3.12-14; note that values of the standardized rank scores in Figure 3.14 are not comparable with analogous values in Figure 3.4, since they are computed over a different number of competitor models).

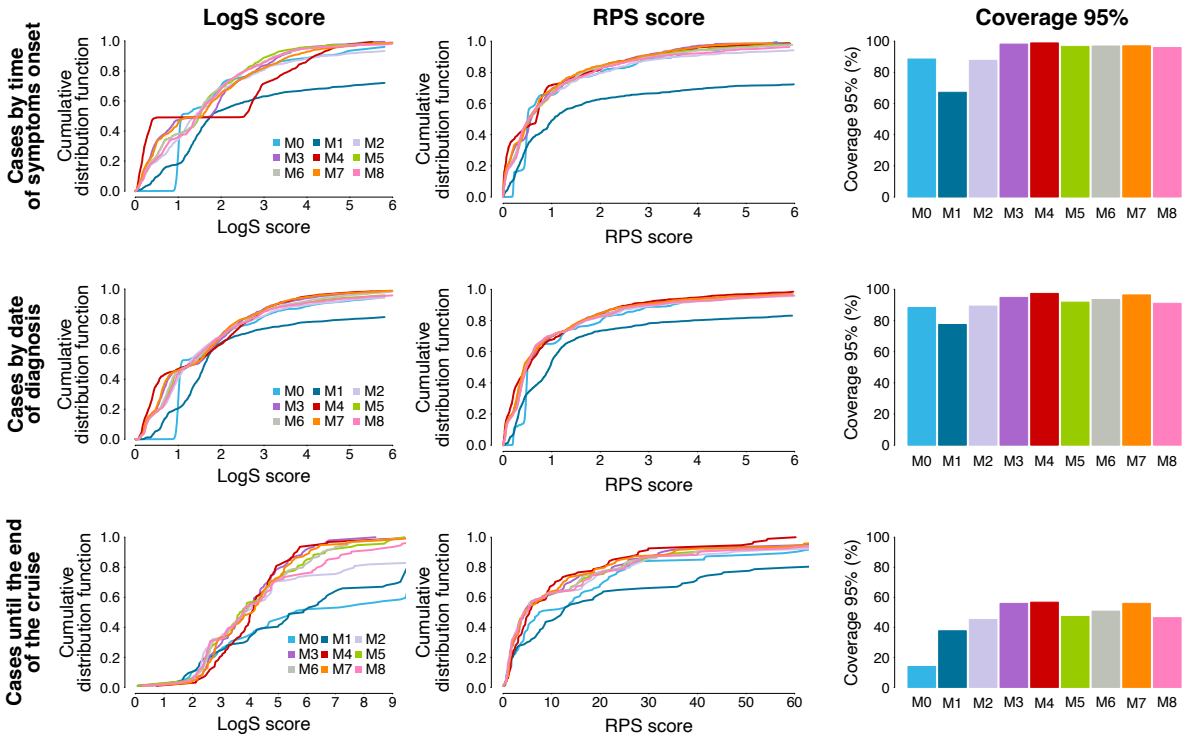


Figure 3.12. *Forecasting scores across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; center column: RPS score; right column: 95% coverage.

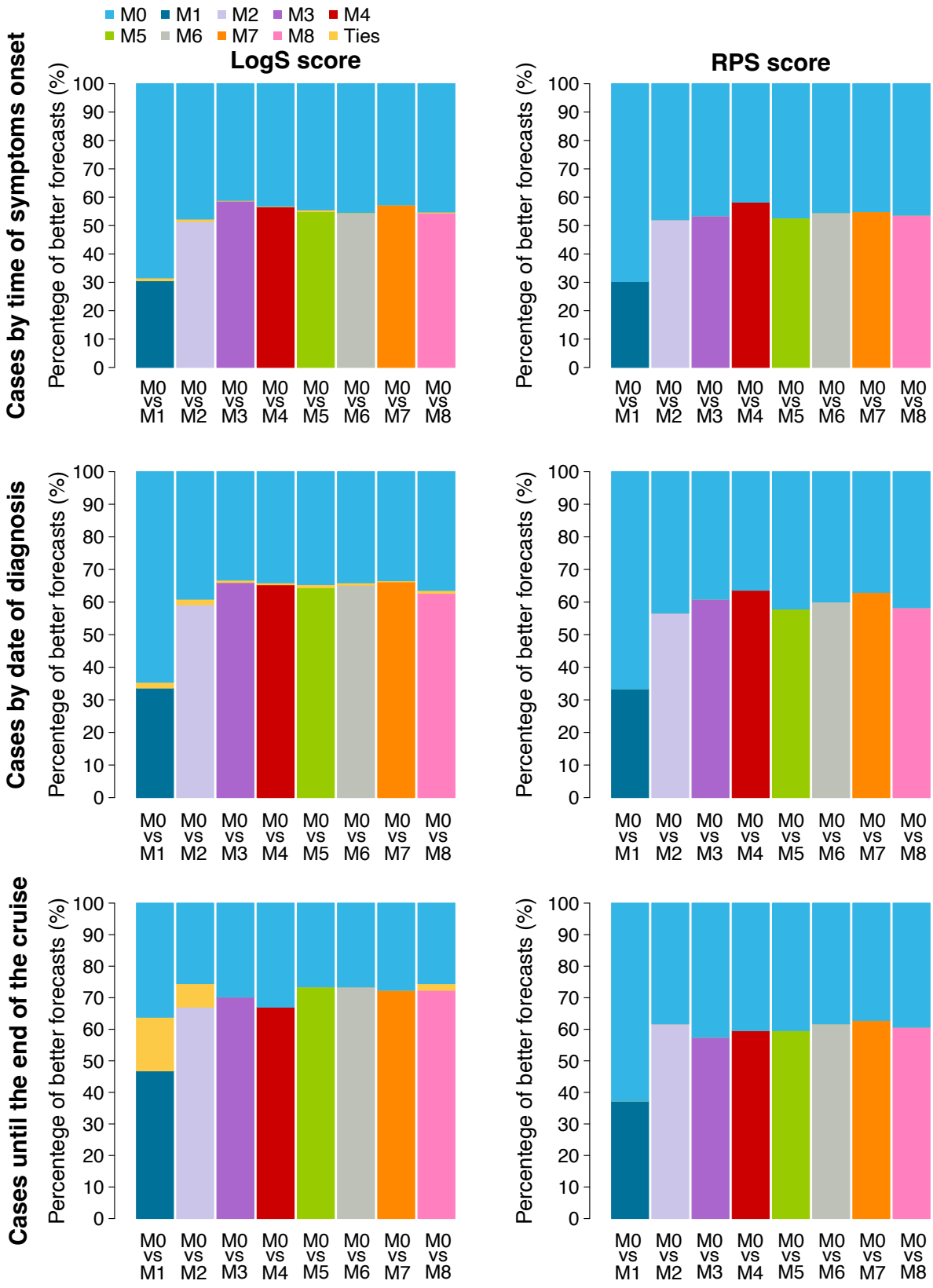


Figure 3.13. *Model performance against the baseline across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. The 100% stacked barcharts show the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (yellow).

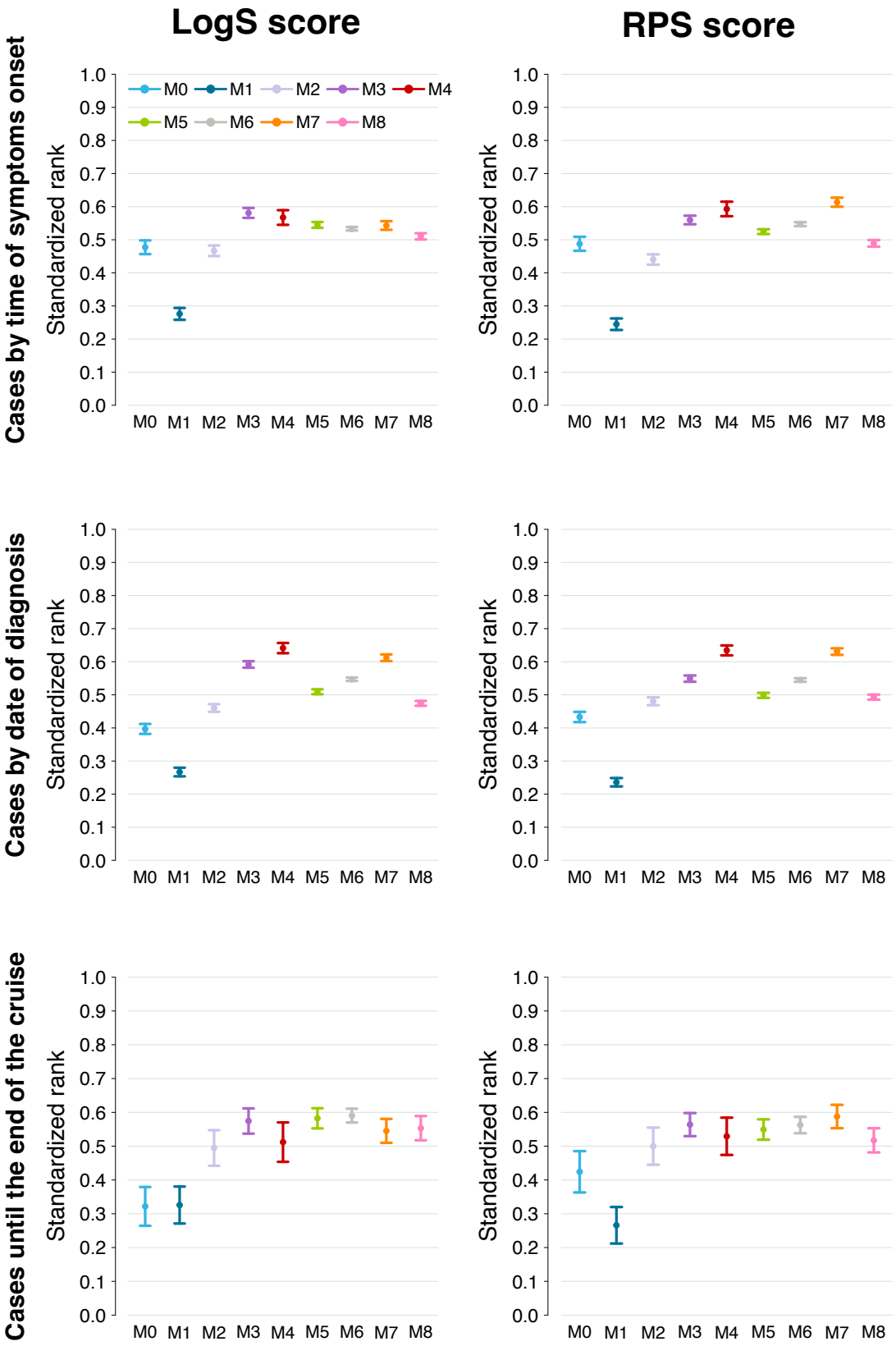


Figure 3.14. *Standardized rank of forecasting models across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by diagnosis date; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. Points: mean standardized rank; whiskers: 95% CI of the mean standardized rank.

ALTERNATIVE CHOICES OF THE PARAMETER d_F

We evaluated how the performance of the forecasts change when choosing different values for parameter d_F from that adopted in the baseline (24 hours). d_F represents the time window over which the number of reported cases by date of symptom onset is considered incomplete due to reporting delays and cannot be compensated by nowcasting, i.e. the width of the "incompleteness window". Because the first forecasts can be given after a time of at least d_F since the start of the cruise, considering values larger than 24 hours would make forecasting less relevant, given that most of the considered cruises had a duration of 7 days. Therefore, we assessed values of 20, 16, 12, 8 and 4 hours as alternative values for d_F, and we show results on the performance of model M3, the best performing model in the main analysis (results for all other models were similar). Although lower values of d_F have a limited impact on predictions (Figure 3.15-16), the standardized rank for forecasts produced with different values of d_F shows a general tendency of worsening performances as d_F decreases (Figure 3.17).

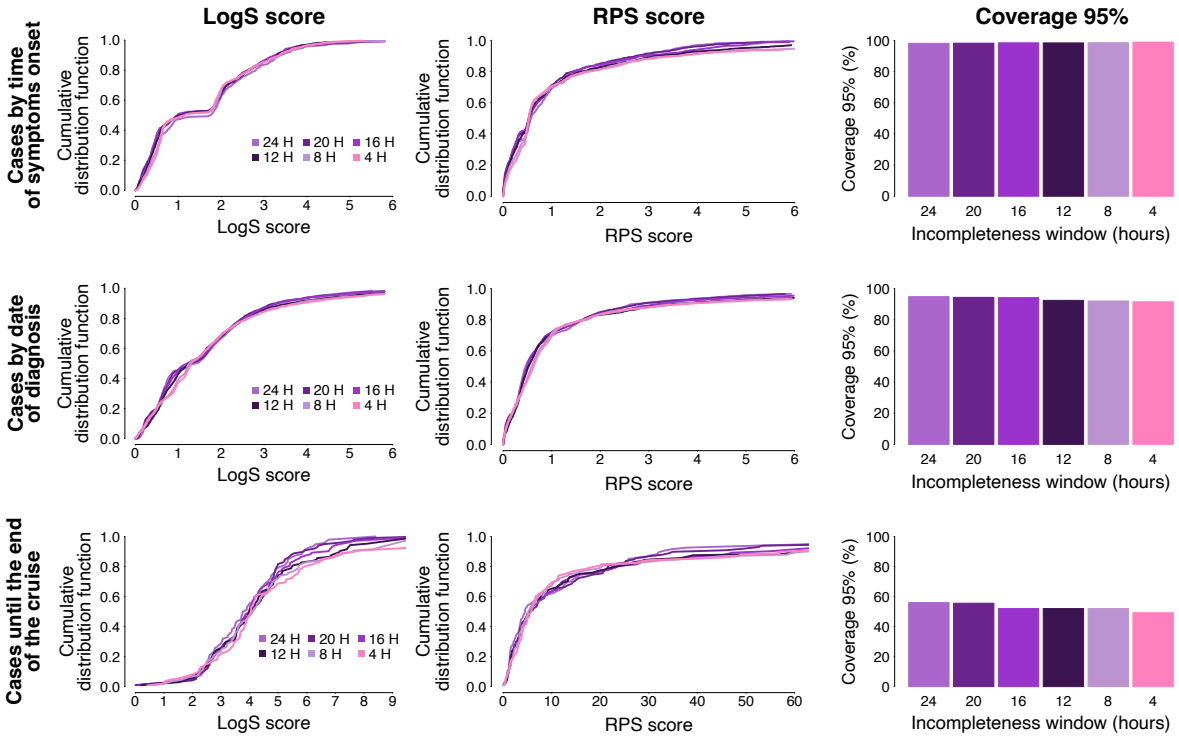


Figure 3.15. *Forecasting scores across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; center column: RPS score; right column: 95% coverage.

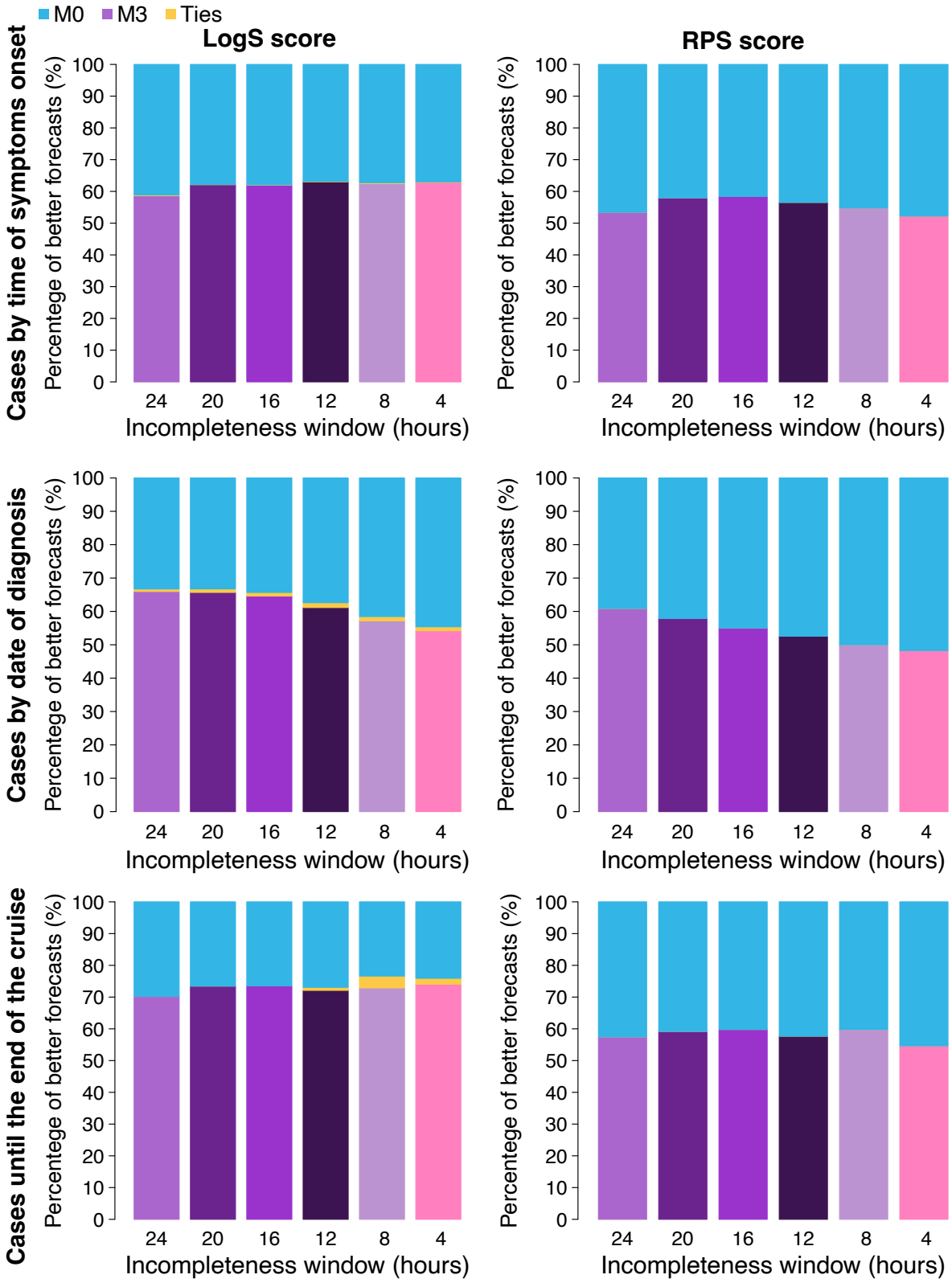


Figure 3.16. *Model performance against the baseline across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by date of diagnosis; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. The 100% stacked barcharts show the proportion of forecast points where: each model has a better score than the baseline M0 (color-coded by model); M0 has a better score (light blue); M0 and the model have the same score (yellow).

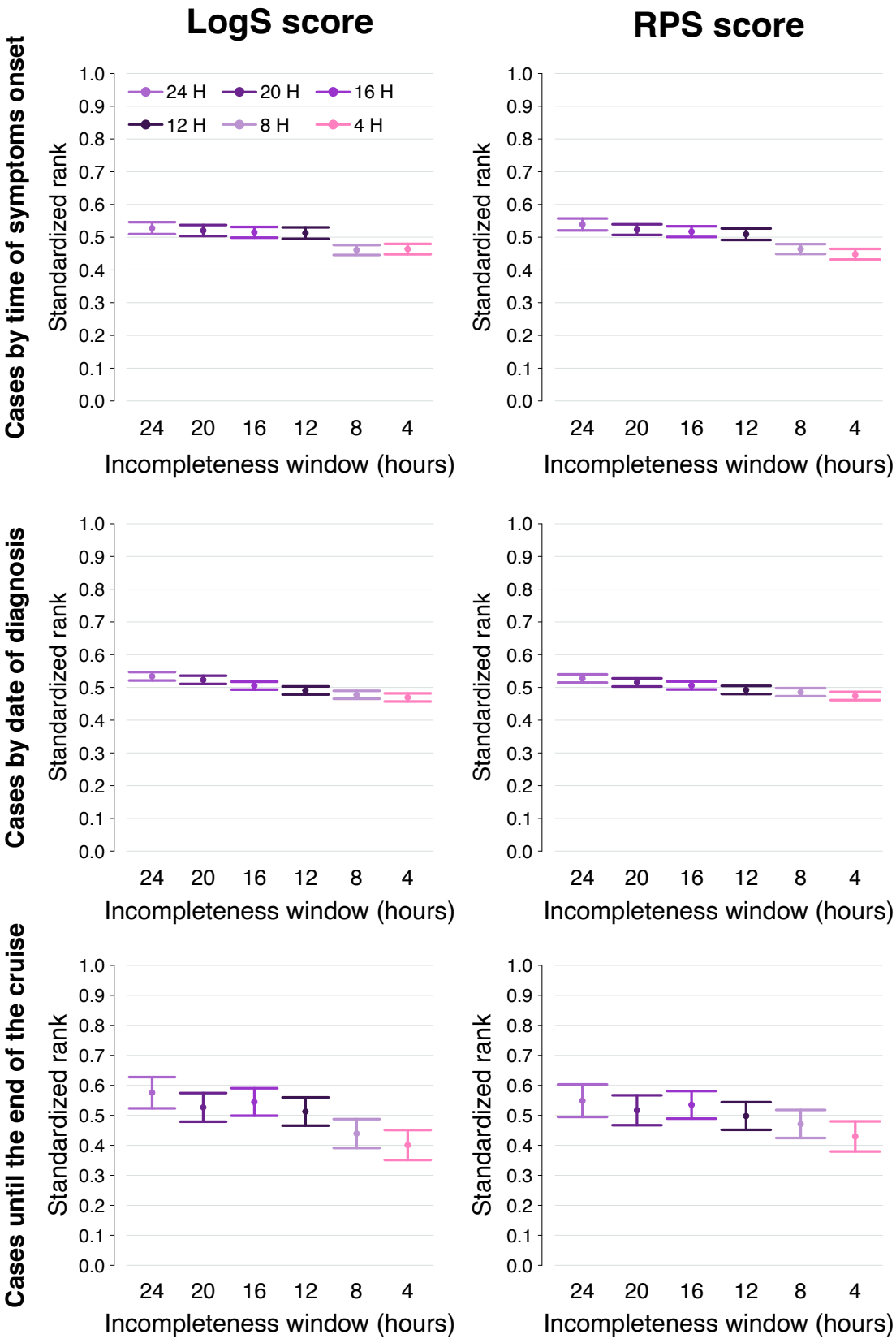


Figure 3.17. *Standardized rank of forecasting models across all outbreaks, times of analysis and forecasted data points.* Top row: forecasts by time of symptom onset; middle row: forecasts by diagnosis date; bottom row: forecasts of total cases until the end of the cruise. Left column: logS score; right column: RPS score. Points: mean standardized rank; whiskers: 95% CI of the mean standardized rank.

4. RETROSPECTIVE AND PROSPECTIVE ASSESSMENT OF CHLAMYDIA CONTROL INTERVENTIONS AMONG MEN WHO HAVE SEX WITH MEN IN HIV PRE-EXPOSURE PROPHYLAXIS PROGRAM

4.1. ABSTRACT

Screening protocols associated with HIV pre-exposure prophylaxis (PrEP) programs have a complex impact on the epidemiology of Chlamydia trachomatis (CT) by reducing infectious periods but also possibly favoring risky behaviors among users. Questions have been raised on screening asymptomatic individuals due to concern that treating clinically irrelevant infections may reduce the development of natural immunity and select for antibiotic resistance. Finally, novel control interventions such as doxycycline post-exposure prophylaxis (DoxyPEP) require a quantitative assessment.

We calibrated a mathematical model of CT transmission to data from men who have sex with men (MSM) enrolled in a PrEP program between January 2018 and December 2024 in the province of Verona, Italy. The model population is stratified by sexual activity level and participation to the PrEP program. We projected the model-estimated CT incidence, prevalence of symptomatic disease and CT-associated antibiotic consumption until 2035, under alternative control interventions.

By the end of 2017, infection incidence was estimated at 6.8 (95% Prediction Interval, PrI: 3.5-9.3) per 1,000 person-months and symptomatic prevalence at 0.23% (95%PrI: 0.12-0.35). We estimate a marginal impact of the current PrEP program (symptom-agnostic screening every six months) in the overall MSM population, with epidemiological indicators only slightly increasing (by less than 10% until 2035). Halving the screening interval would imply a 25.9% reduction from current levels (95%PrI: 22.5-32.6%) in the overall symptomatic prevalence among MSM by 2035, while doubling it would result in a 18.7%

increase (95%PrI: 14.6-30.4%). Maintaining the six-months frequency with symptom-based screening would not only increase the symptomatic prevalence by 40.3% (95%PrI: 30.9-62.9%) but also the overall CT-associated antibiotic consumption rate by 34.6% (95%PrI: 27.8-56.1%). DoxyPEP prescribed to high-risk PrEP users (15 or more sexual partners per year) may reduce the burden of disease by 62.7% (95% PrI: 55.9-88.6%) in 10 years.
The impact of PReP-associated screening on CT in the MSM population strictly depends on the screening protocol. In settings with limited circulation, screening asymptomatic individuals reduces both the burden of disease and the population-level antibiotic consumption associated with CT infections. DoxyPEP may provide sustained reductions in CT circulation, although the risks for emerging doxycycline resistance needs to be carefully evaluated.

4.2. INTRODUCTION

Sexually transmitted infections (STIs) represent a major global public health challenge. According to the World Health Organization, more than one million curable STIs are acquired every day, amounting to approximately 374 million annual cases of chlamydia, gonorrhea, syphilis, and trichomoniasis combined [145]. Men who have sex with men (MSM) are disproportionately affected due to a combination of behavioral factors and biological susceptibility associated with the types of sexual contacts. Among bacterial STIs, Chlamydia trachomatis (CT) is the most prevalent, with an estimated 129 million new cases reported globally in 2020 [146]. CT infections are often asymptomatic and underreported [147], which facilitates onward transmission. If left untreated, chlamydial infections in men can lead to serious complications, including chronic urethritis, epididymitis, prostatitis, and reactive arthritis [148]-[150]. Furthermore, ongoing CT infection can facilitate the transmission of HIV [151], [152]. The current standard of care for CT infections involves antibiotic treatment, most commonly with azythromycin and doxycycline, which are highly effective in clearing the infection [153].

Recently, the introduction of pre-exposure prophylaxis (PrEP) for the prevention of HIV transmission has perturbed the transmission dynamics of CT and other STIs. PrEP is highly effective in preventing HIV infection, and users tend to reduce condom use as a form of behavioral risk compensation [154]. Since condoms have some efficacy in reducing the risk of transmission for other STIs, access to PrEP has been conditioned to the administration of periodic STI screening tests [155]. Screening programs for bacterial STIs enhance the detection and antibiotic treatment of asymptomatic cases, reducing the length

of their infectious period and therefore contributing to reducing bacterial circulation. However, growing evidence suggests that antibiotic therapy may interfere with the development of natural immunity, potentially increasing susceptibility to reinfection, a phenomenon referred to as "arrested immunity" [156], [157]. Furthermore, there is concern [158] that the administration of antibiotics to asymptomatic carriers may favor the emergence and spread of antimicrobial resistance while being unjustified from the perspective of clinical risks. Together, these factors create a complex interplay in which PrEP affects not only HIV prevention but also the broader transmission patterns of STIs [159]-[162]. The picture is further complicated by the recent interest for novel control strategies for bacterial STI, such as doxycycline post-exposure prophylaxis (DoxyPEP). DoxyPEP involves taking a short course of doxycycline immediately after potential exposure to an STI and has shown promise in reducing the risk of acquiring CT and other bacterial STIs in high-risk populations [163].

A quantitative understanding of these interacting effects is needed for designing effective control strategies [59], [60], [164]. In this study, we develop a mathematical model calibrated to epidemiological and behavioral data from a cohort of MSM enrolled in a PrEP program at the Verona University Hospital, Italy. Our aim is to assess how PrEP-associated screening protocols and behavior changes have influenced the transmission dynamics and burden of CT, and to explore the potential impact of prospective changes to the screening protocols and the additional introduction of DoxyPEP.

4.3. METHODS

DATA

We considered data from all 542 individuals who were enrolled in the PrEP program at the STI clinic of the Verona University Hospital between January 1, 2018 (date of the start of the program) and December 31, 2024. The hospital serves as a reference STI clinic for the whole province (total population approximately 900,000). We had access to information on the date of enrollment (first medical visit) for all individuals enrolled in the study period ("full cohort"). At each medical visit, information was collected on the number of unique sexual partners in the last three months, the frequency of condom use during sex acts, the presence of CT-compatible symptoms, as well as CT tests performed and corresponding results. Such information was available for a subset of 123 individuals enrolled between January 1, 2018, and February 28, 2021 ("partial cohort").

TRANSMISSION MODEL

We developed a stochastic model for CT transmission in the MSM population of the Italian province of Verona (assuming a constant MSM population of approximately 33,000; see Section 4.6). The model was run until December 31, 2035, considering changes occurring at two milestones: the introduction of the PrEP program on January 1, 2018, until which the system is assumed at endemic equilibrium; and potential prospective changes in PrEP-associated screening and other CT control strategies occurring on January 1, 2026.

Susceptible individuals (S) may acquire CT infection through sexual contacts with infectious partners; infected individuals undergo an incubation period within an exposed state (P), and may successively develop symptoms (IS) or remain asymptomatic (IA) with a certain probability (Figure 4.1a). Symptomatic individuals are assumed to seek for medical help from their general practitioner, where they will be typically be prescribed antibiotics on the basis of symptoms without a laboratory confirmation of CT, heal the infection and revert to S. Asymptomatic individuals are assumed to clear the infection spontaneously and develop temporary immunity to reinfection (R) [156], [157] before reverting to S after some time [165]. The population is stratified in three sexual activity levels (SAL), indicated with subscript l, according to the number of unique sexual partners reported in the previous 12 months within the 2017 European MSM Internet Survey (EMIS) study for Italy [166]. We considered the SAL as low (for individuals with 4 or less partners per year), middle (between 5 and 14) and high (more than 15) (see distribution in Figure 4.1b). Sexual contacts across SALs were modelled by considering a contact matrix resulting from the linear combination of a fully assortative matrix (individuals have contacts only with members of the same SAL) and a fully homogeneous matrix (individuals have contacts irrespectively of their SAL) [167]. All infected individuals (P, IS, IA) are assumed to be equally infectious. We assume that the use of condom during sexual intercourse (with probability estimated from available data [166]) reduces the transmission rate by a previously estimated efficacy against CT [168].

Starting from January 1, 2018, individuals are enrolled to the PrEP program according to the yearly number of first visits recorded in the full cohort data (Figure 4.1c). Newly enrolled individuals were moved to the PrEP program from the population not in PrEP, proportionally to the SAL distribution recorded from partial cohort data (Figure 4.1b-c). Starting from 2025, we assumed a constant yearly number of new PrEP enrolments, equal to the average observed in 2021-2024. Individuals may drop out from the PrEP program with an annual rate

estimated at about 6% per year from partial cohort data (see Section 4.6). We considered a dropout as an individual in the partial cohort whose last screening visit occurred more than 9 months before the end of data, i.e., December 31, 2024. These assumptions result in a projected PrEP population that would be still growing by the end of our simulation period (Figure 4.1d, reported as a proportion of the total MSM population).

Individuals in the PrEP program follow the same epidemiologic dynamics of the rest of the MSM population, except that infected individuals (P, IA, IS) will revert to S upon positive screening results and the ensuing administration of antibiotic treatment. In addition, PrEP users are assigned a lower probability of condom use, estimated from partial cohort data.

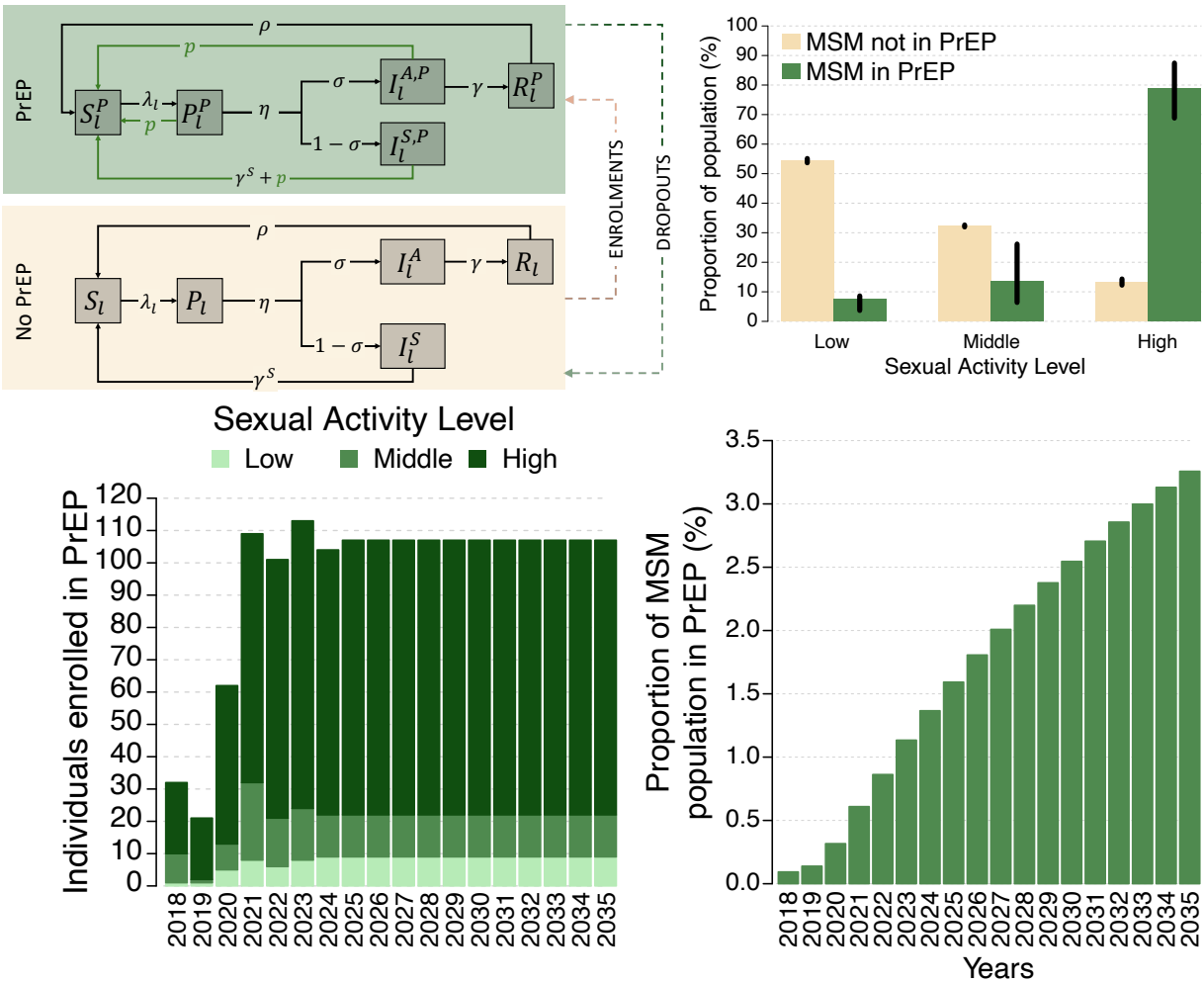


Figure 4.1. *Chlamydia Transmission Model.* a) Compartmental scheme. S: susceptible; P: exposed infectious; IA: asymptomatic infectious; IS: symptomatic infectious; R: temporarily immune. Superscript P indicates the population of PrEP users; subscript l represents the SAL. PrEP enrolments and dropouts are shown in dashed colored arrows. State transitions due to PrEP screening are shown in green arrows. Parameter descriptions and values are reported in Table 4.1. b) Distribution of SALs in the general MSM population [24] and in the PrEP population (from partial cohort data). Bars: mean value over time; vertical lines: 95 percentile of the temporal variation. c) Yearly number of new PrEP enrolments; data for 2018-2024 from the full

cohort. Color shades indicate SALs. d) Modelled yearly proportion of MSM population actively enrolled in the PrEP program (%).

MODEL CALIBRATION

Three model parameters (the basic reproduction number, the proportion of asymptomatic infections and the degree of assortative mixing) were calibrated using CT diagnosis data from the partial cohort between January 2018 and December 2024 (Figure 4.2). We applied a Markov Chain Monte Carlo with a Metropolis-Hastings sampling algorithm to a likelihood defined as the product of three components: (i) the Poisson likelihood of the yearly number of newly diagnosed CT cases; (ii) the multinomial likelihood of newly diagnosed CT cases across the three SALs; (iii) the binomial likelihood of the observed proportion of symptomatic cases. The calibrated parameters were used to simulate CT transmission dynamics in the MSM population until 2035 under the current screening protocol (symptom-agnostic screening every six months) and alternative scenarios. All other model parameters are reported in Table 4.1 and full methodological details are provided in Section 4.6

SCENARIO ANALYSIS

We considered 5 scenarios (Table 4.2), the first represent the current protocol, three scenarios assessed alternative screening protocols and one considered the additional prescription of DoxyPEP to high-risk PrEP users. We modeled the effect of DoxyPEP as a reduction of the transmission rate to PrEP users in the high SAL, using literature estimates for the potential coverage, adherence and effectiveness [163]. We assessed the impact of each scenario using three epidemiological endpoints calculated for the overall MSM population: CT incidence (as an indicator of the transmission dynamics), prevalence of symptomatic infections (for disease burden) and CT-specific antibiotic consumption. We chose the prevalence as an indicator of burden because it summarizes well the combination of disease incidence (which may be inflated by frequent reinfections due to rapid case identification) and symptom duration (which may be shortened by screening protocols). To quantify antibiotic consumption, we considered all symptomatic infections which get treated in the population not in PrEP, summed with all infections identified in the PrEP population by spontaneous presentation or at the screening visit independently of symptoms. We defined as one dose the average amount of antibiotic that is used in a typical course for CT infection. The endpoints were evaluated as averages over the last quarter of 2017, 2025 and 2035,

corresponding to the end of the three simulation periods. To provide insight in CT transmissibility under each scenario, we also evaluate the effective reproduction number (R) over time, estimated with the Next Generation Matrix approach [7] (see Section 4.6 for methodological details).

Table 4.1. *Model parameters.*

PARAMETER	DESCRIPTION	UNIT	VALUE	REFERENCE
$\boldsymbol{N}$	MSM population	-	33,000*	Assumption
$\frac{\mathbf{1}}{\boldsymbol{\eta}}$	Mean length of incubation period	Days	16	[169]
$\frac{\mathbf{1}}{\boldsymbol{\gamma}}$	Mean length of asymptomatic infection (excluding the incubation period)	Days	116	[169]
$\frac{\mathbf{1}}{\boldsymbol{\gamma^S}}$	Mean time to treatment for symptomatic infection after symptom onset	Days	17	[170]
$\frac{\mathbf{1}}{\boldsymbol{\rho}}$	Mean length of temporary immunity	Days	218*	[165]
$\boldsymbol{\alpha}$	Probability of condom use in the general MSM population	%	70	[166]
$\boldsymbol{\alpha^P}$	Average probability of condom use among PrEP users (2018-2024)	%	36	estimate from partial cohort data
$\boldsymbol{\phi}$	Reduction of transmissibility associated to condom use	%	60	[168]
$\frac{\mathbf{1}}{\boldsymbol{p}}$	Screening interval (current protocol)	Months	6	-
$\boldsymbol{\chi}$	Coverage of DoxyPEP in high-risk PrEP users	%	89	[163]
$\boldsymbol{\psi}$	DoxyPEP adherence	%	86	[163]
$\boldsymbol{\xi}$	Efficacy of DoxyPEP	%	88	[163]
$\boldsymbol{R_0}$	Basic reproduction number	-	1.14 (95%CI: 1.08-1.21)*	Calibrated
$\boldsymbol{\sigma}$	Probability of remaining asymptomatic	%	39.7 (95%CI: 22.3-54.7)*	Calibrated
$\boldsymbol{\omega}$	Degree of assortative mixing among SALs	%	77.0 (95%CI: 45.0-99.1)*	Calibrated

* parameters subject to a sensitivity analysis.

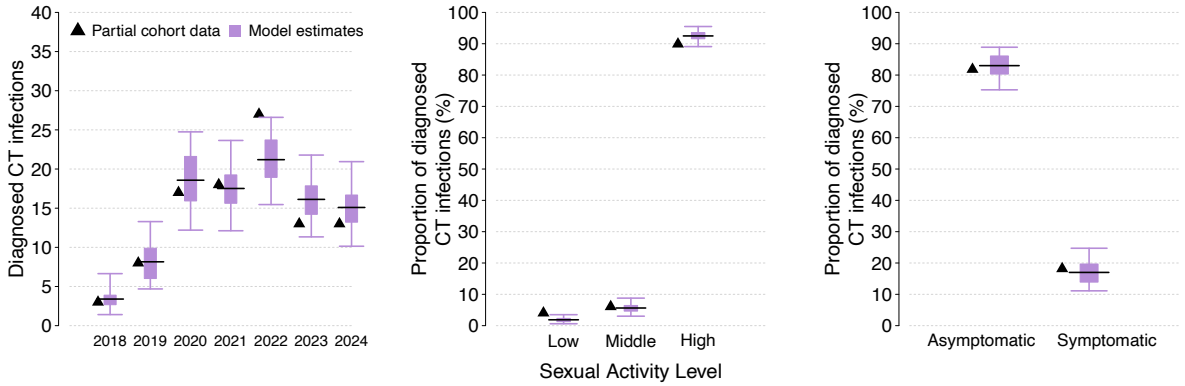


Figure 4.2. *Model calibration to partial cohort data from the PrEP program at the Verona University Hospital.* a) number of diagnosed CT cases per year; b) distribution of diagnosed cases by SAL; c) distribution of diagnosed cases by symptoms. Boxplots represent the model-estimated mean (horizontal line), 50% CrI (boxes) and 95%CrI (whiskers).

Table 4.2. *Description of the evaluated scenarios.*

Scenario	Description
6m	All PrEP users are screened every 6 months (current protocol)
3m	All PrEP users are screened every 3 months.
12m	All PrEP users are screened every 12 months.
6m-sym	PrEP users are screened every 6 months if they present with CT symptoms
6m-DoxyPEP	Prescription of DoxyPEP to high SAL PrEP users in addition to the current protocol

4.4. RESULTS

The model calibration identified a tendency to assortative mixing across SALs in our study population, with a mean degree of assortativity of 77.0 (95%CI: 45.0-99.1%). Asymptomatic infections were found to be a minoritary but significant proportion of all infections (mean 39.7%, 95%CI: 22.3-54.7). The basic reproduction number was estimated at 1.14 (95%CI: 1.08-1.21), compatible with the infection being at endemic equilibrium at a low prevalence (Table 4.1). We estimate that CT incidence in the overall MSM population increased slightly after the introduction of the current semestral PrEP screening protocol ("scenario 6m"), from approximately 6.8 (95% Prediction Interval, PrI: 3.5-9.3) infections per 1,000 person-months at the end of 2017 (Figure 4.3), to 7.2 (95% PrI: 3.6-10.2) in 2025, with a projection of 7.7 (95%PrI: 4.2-10.5) by the end of 2035. By comparison, the corresponding figure at the end of 2025 for the PrEP subgroup, characterized by higher sexual activity and lower condom use, was estimated at 23.5 (95%PrI: 12.2-35.6-33.9) per 1,000 person-months, over three times higher, compatibly with the higher SAL distribution (Figure

4.1b) and lower condom use (Table 4.1) in this subpopulation. The estimated prevalence of immune individuals was below 11% before the introduction of PrEP even in the high SAL group (Figure 4.3c) and decreased to less than 2% in the high SAL among PrEP users, in agreement with the arrested immunity hypothesis. We also estimated a steady prevalence of symptomatic infections in the overall population, from 0.23% (95%PrI: 0.12-0.35) at the end of 2017 to 0.26% (95%PrI: 0.13-0.40) at the end of 2035, and a slight increase in the incidence of administered antibiotic treatment for CT from about 13.5 (95%PrI: 7.0-20.5) doses per 100,000 person-months in 2017 to about 15.8 (95%PrI: 8.2-24.8) per 100,000 person-months in 2035 (Figure 4.4).

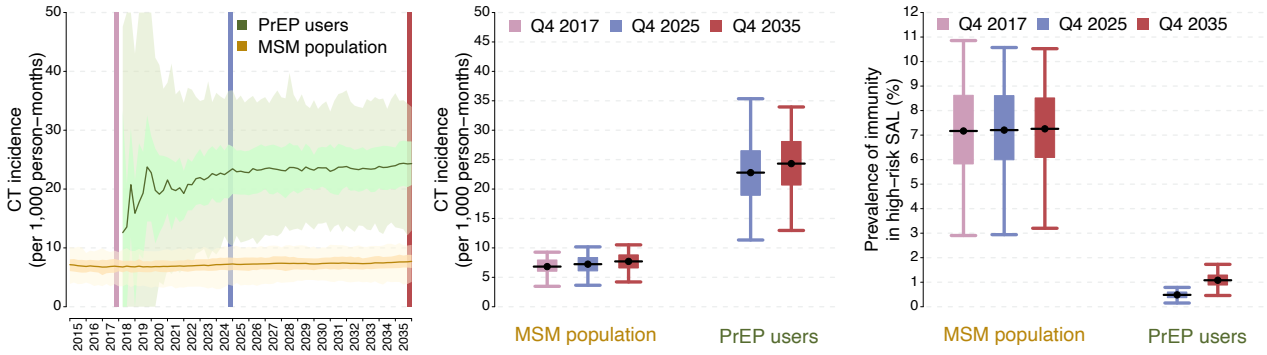


Figure 4.3. *Estimated incidence of CT infections in the MSM population (both in PrEP and not in PrEP) and among PrEP users only, assuming the maintenance until the end of 2035 of the current PrEP screening protocol (scenario 6m) introduced at the start of 2018.* a) Temporal evolution. Colored vertical bands denote periods at which the endpoints are evaluated: just before PrEP introduction, i.e. in the last quarter (Q4) of 2017 (purple); before potential changes in the screening protocol, Q4 2025 (blue); and at the end of the projection period, Q4 2035 (red). Solid lines represent mean values, shaded areas represent 95% CrI. b) Quarterly average of estimates in correspondence of the key periods. c) Average prevalence of temporary immune individuals in the high-risk SAL. For panels b) and c), boxplots represent the model-estimated mean (horizontal line), 50% PrI (boxes) and 95%PrI (whiskers).

If the screening interval were shortened to 3 months (scenario 3m), the CT incidence would be reduced by 24.5% (95%PrI: 18.8-31.5%) and the symptomatic prevalence by 25.9% (95%PrI: 22.5-32.6%) by the end of 2035. Conversely, these quantities would grow by 17.7% (95%PrI: 13.1-33.0%) and 18.7% (95%PrI: 14.6-30.4%) respectively with a screening delayed to every 12 months (scenario 12m, Figure 4.4). A scenario with a semestral screening where a CT test is administered only upon the presence of symptoms (scenario 6m-sym) would entail not only an increase by 40.1% (95%PrI: 30.1-72.2%) in CT incidence and 40.3% (95%PrI: 30.9-62.9%) in symptomatic prevalence, but, counterintuitively, also by 34.6% (95%PrI: 27.8-56.1%) in the overall antibiotic consumption. Finally, we estimate that the frequent use of DoxyPEP among

high-risk individuals (scenario 6m-DoxyPEP) has potential to substantially reduce the incidence, symptomatic prevalence and rate of CT-specific antibiotics over the next 10 years projection period considered, with estimated reductions of 61.2% (95% PrI: 50.6-88.9%), 62.7% (95% PrI: 55.9-88.6%) and 62.7% (95%: 56.0-88.9%) respectively, compared to the current (6m) interventions.

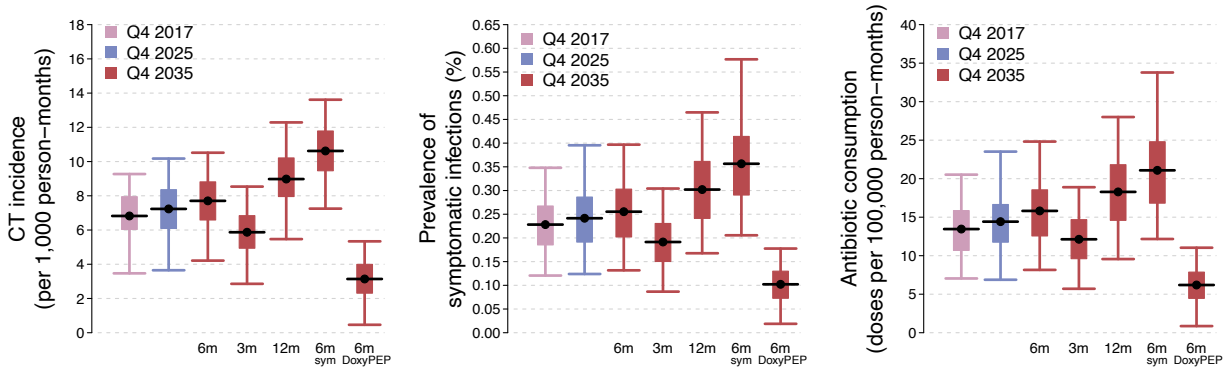


Figure 4.4. *Impact of alternative scenarios for CT control.* a) CT incidence; b) prevalence of symptomatic infections; c) CT-specific antibiotic consumption. Boxplots represent the model-estimated mean (horizontal line), 50% PrI (boxes) and 95%PrI (whiskers). Horizontal dashed lines represent mean estimates for 2035 under the current protocol (scenario 6m), reported as a reference for comparisons with past values and with alternative scenarios.

The results reported above for the considered scenarios are well reflected in the temporal evolution of the estimated effective reproduction number over time, R_e (Figure 4.5). As expected for a disease at endemic equilibrium, R_e was hovering over the epidemic threshold of 1 at the end of 2017; the introduction and progressive expansion of the PrEP program with semestral screening interval did not significantly influence the value of R_e. However, different screening intervals would result in a perturbation of R_e in a direction that is coherent with previously reported changes in epidemiological endpoints. When only changes in the screening protocol are considered, the dynamic system slowly tends to revert to an equilibrium R_e of 1: note however that by the end of 2035 this equilibrium is not fully reached in most scenarios, due to a still growing PrEP population (Figure 4.1d). The scenario with the highest upward perturbation in the reproduction number was the one with symptom-based screening.

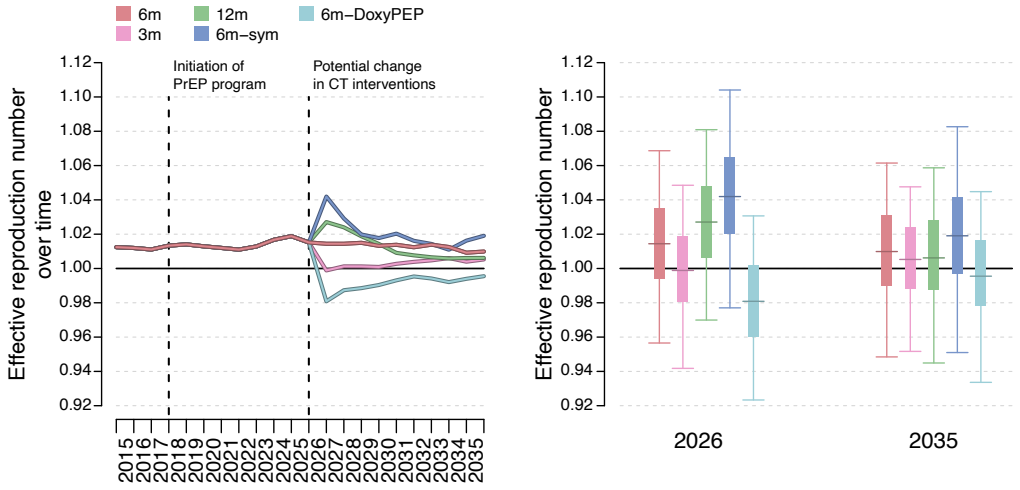


Figure 4.5. *Estimated effective reproduction number Re for CT in the province of Verona under alternative scenarios.* a) Temporal evolution; lines represent mean values; b) distributions after potential changes in CT interventions (2026) and at the end of the study period (end of 2035). Boxplots represent the model-estimated mean (horizontal line), 50% CrI (boxes) and 95%CrI (whiskers). In both panels, black horizontal lines represent the epidemic threshold of 1.

We run three sensitivity analyses on model results: one in which we considered an MSM population of half the value used in the main analysis; and two where we consider extreme alternative values on the duration of temporary immunity - a parameter characterized by high uncertainty - namely 11 days and 4.6 years, rather than 7 months in the main analyses. We found that, for smaller MSM populations and shorter immunities, qualitative results would be the same but with exacerbated quantitative effects; however, a very long immunity would imply negligible variations across screening protocols, thereby favoring the adoption of symptom-based screening to reduce the screening workload.

4.5.DISCUSSION

In this study, we evaluated the impact of an HIV PrEP screening program, introduced since January 2018 in the province of Verona, Italy, on the transmission dynamics of Chlamydia trachomatis (CT) among MSM. We used a compartmental stochastic model informed with and calibrated to data collected during periodic screening visits of PrEP users. We found that, under the current semestral screening protocol, the program did not significantly alter the epidemiology of CT, due to an almost balanced compensation between different processes: on one hand, the reduction in the length of infectious period of infected PrEP users, allowed by periodic screening and antibiotic treatment of diagnosed CT cases; and on the other hand, the reduced condom use observed among PrEP users. We estimated a marginal prevalence of

temporarily immune individuals (less than 11% even in high sexual activity groups), providing little support to a significant impact of arrested immunity on the overall epidemiology of CT [156]. When considering shorter or longer CT screening intervals among PrEP users, the above mentioned compensation became unbalanced, leading to meaningful future changes in estimated transmissibility and burden: in particular, we estimated an increase within 10 years in CT incidence, disease burden and antibiotic consumption in the overall MSM population by about 20%, if screening were loosened to once every 12 months, and a decrease of about 25% if it were intensified to once every 3 months. Thus, even though PrEP users were a small percentage of the overall MSM population (projected to be less than 3.5% in 2035), they represented a non-negligible proportion of the group with highest sexual activity level (up to one quarter in 2035) and were characterized by a higher incidence of infections (Figure 4.3), explaining the sensitivity of projections to the screening intensity.
We also show the critical importance of limiting cryptic transmission by asymptomatic individuals by maintaining screening for asymptomatic individuals and treatment for diagnosed asymptomatic infections. In absence of this component of the PrEP screening protocol, we estimate an overall 35% increase in antibiotic use in the general MSM population and 40% in the burden of symptomatic infections. These results can be explained with the long infectious period of asymptomatic individuals and the limited estimated prevalence of immune individuals across all population subgroups. Nonetheless, we note that this result is sensitive on the assumed duration of temporary immunity. When considering a duration of about 5 years (rather than 7 months in the main analysis), the impact of screening protocols on CT epidemiology would be negligible, so that switching to a symptom-based protocol to reduce the screening workload would bear no counterindications. In our study, the duration of immunity was estimated from an epidemiological study on vaginal CT infections in women [165]. Further research on the duration of immunity in CT infection in males is therefore warranted.
Finally, we have shown that the prescription of DoxyPEP to high-risk PrEP users (here defined as reporting 15 or more unique sexual partners per year) has the potential to result in a marked reduction of CT transmissibility below the epidemic threshold, providing declines in CT burden and overall CT-specific antibiotic use. However, this benefit needs to be carefully evaluated against potential costs (both economic and in terms of potential emergence of resistance) associated with doxycycline consumption. In this study, this quantity could not be evaluated, since it would have required reliable data on the

number of sexual encounters (rather than on the number of sexual partners) within the different SALs.

Further limitations may apply to this study. Unknowns on the future rollout of the PrEP program may alter the scale of projected epidemiological effects. We assumed that enrollments and dropouts will continue with the stable annual rates that have been observed over the last four years, and that the distribution of sexual activity levels in newly enrolled individuals will follow the one observed among enrolled PrEP users between the start of the program in January 2018 and the end of available data in December 2024. If the scale-up of PrEP roll-out is more rapid, e.g. through the creation of further PrEP clinics in the province, the estimated effects of changes in screening protocols might be amplified. Our results may also be impacted by a number of uncertainties on sexual behavior. Persisting social stigma does not allow the accurate estimation of the proportion of MSM over the total population. We assumed 10% of the male population in ages of typical sexual activity (resulting in a 3.6% of the total population of the province of Verona) [171], a proportion that is generally higher than commonly assumed [164], [167] to compensate for stigma-induced underestimation. However, results were analogous in direction and enhanced in effect size within a sensitivity analysis considering a halved MSM population. Due to lack of data, we could not consider potential age heterogeneities in sexual activity levels, age-assortativity in sexual contacts and possible preferential mixing between PrEP users. Additional uncertainties may come from geographical heterogeneities (e.g., a higher concentration of MSM and sexual activity in the city of Verona as opposed to smaller and less densely populated rural towns). We did not consider epidemiologically relevant differences between genital, rectal, or oropharyngeal infections, such as the duration of carriage or the probability of transmission and symptom development. Explicitly considering different anatomical sites would require data on the frequency of different sexual practices, both in the general population and in the PrEP cohort, which were not available for this study. Finally, quantitative results on the impact of screening in different contexts will be highly influenced by specific sexual behavior (partner distribution and condom use of the PrEP and general MSM population), PrEP coverage and local CT epidemiology.

We conclude that screening protocols within PrEP programs can have a relevant impact on CT transmission dynamics, and evaluations on potential protocol changes need to balance acceptable levels of disease burden, laboratory effort and antibiotic consumption, also taking into account cost-

benefits associated with other STIs (gonorrhea, syphilis) on which the screening protocol may impact. Our study robustly highlights the importance of maintaining CT screening and antibiotic treatment for asymptomatic individuals in PrEP as a strategy to reduce the overall antibiotic consumption for CT in the MSM population, provided the temporary immunity associated to natural clearance is confirmed to be short-lived (less than one year). We also demonstrate the potential role of DoxyPEP as a complementary prevention strategy which could bring down the CT burden significantly and/or allow for relaxations in the screening workload. The exploration of such an approach should assess its feasibility, acceptability and cost-effectiveness, including potential risks towards the emergence of doxycycline resistance.

4.6. APPENDIX

MODEL DETAILS

Model equations

We described the dynamics of CT transmission with the following $10 \cdot N_L$ ordinary differential equations (ODEs), where $N_L = 3$ is the number of sexual activity levels (SALs) represented by subscript $l \in \{1, 2, 3\}$ and 10 is the number of epidemiological states, 5 for the general MSM population (excluding PrEP users) and 5 for PrEP users (indicated with superscript P).

$$
\begin{aligned}
S_l' &= -\lambda_l S_l + \gamma^S I_l^S + \rho R_l - E_l^S + D_l^S \\
P_l' &= \lambda_l S_l - \eta P_l - E_l^P + D_l^P \\
I_l^{A\prime} &= \eta\sigma P_l - \gamma I_l^A - E_l^{I^A} + D_l^{I^A} \\
I_l^{S\prime} &= \eta(1-\sigma) P_l - \gamma^S I_l^S - E_l^{I^S} + D_l^{I^S} \\
R_l' &= \gamma I_l^A - \rho R_l - E_l^R + D_l^R \\
S_l^{P\prime} &= -\lambda_l S_l^P + \gamma^S I_l^{S,P} + \rho R_l^P + p_l^A P_l^P + p_l^A I_l^{A,P} + p_l^S I_l^{S,P} + E_l^S - D_l^S \\
P_l^{P\prime} &= \lambda_l S_l - \eta P_l^P - p_l^A P_l^P + E_l^P - D_l^P \\
I_l^{A,P\prime} &= \eta\sigma P_l^P - \gamma I_l^{A,P} - p_l^A I_l^{A,P} + E_l^{I^A} - D_l^{I^A} \\
I_l^{S,P\prime} &= \eta(1-\sigma) P_l^P - \gamma^S I_l^{S,P} - p_l^S I_l^{S,P} + E_l^{I^S} - D_l^{I^S} \\
R_l^{P\prime} &= \gamma I_l^{A,P} - \rho R_l^P + E_l^R - D_l^R
\end{aligned}
$$

S represents susceptible individuals, P exposed infectious individuals, I^A infectious individuals who remain asymptomatic, I^S infectious individuals who develop symptoms, R temporarily immune individuals. The total numbers of individuals in the general MSM population and in the PrEP program are respectively given by:

$$N_l = S_l + P_l + I_l^A + I_l^S + R_l$$

and

$$N_l^P = S_l^P + P_l^P + I_l^{A,P} + I_l^{S,P} + R_l^P$$

Although the total population $N = \sum_{l=1}^{N_L} N_l + N_l^P$ is assumed constant, N_l and N_l^P are time-varying depending on movements between the non-PrEP and PrEP populations. These consist in the time-varying flows of individuals newly enrolled in PrEP E_l^X, and of individuals dropping out of PrEP, D_l^X, for each epidemiological state X and SAL l . The computation of these terms from available data is described in the below subsection.

Note that before January 1, 2018, the PrEP population is zero, $N_l^P = S_l^P = P_l^P = I_l^{A,P} = I_l^{S,P} = R_l^{S,P} = 0$, for every state X and SAL l, so that the model can be simply described by the first 5 equations with all enrollment and dropout rates also set to zero $E_l^X = D_l^X = 0$.

The dynamics is the same in every SAL l. In both populations, susceptible individuals become infected with a force of infection λ_l (described in detail below), enter the incubating state P_l and leave it at rate η, so that $1/\eta$ represents the mean incubation period. After incubation, individuals remain asymptomatic with probability σ or become symptomatic with probability $1-\sigma$. Asymptomatic infections recover at rate γ (mean duration $1/\gamma$), moving to the temporarily immune class R_l. Symptomatic infections seek for medical assistance and are healed after antibiotic treatment at rate γ^S (mean time to treatment $1/\gamma^S$) and, upon recovery, return directly to the susceptible compartment S_l. Temporary immunity wanes at rate ρ, with mean duration $1/\rho$, after which individuals re-enter the susceptible state.

PrEP users additionally undergo routine screening with rates p_l^A (if asymptomatic) and p_l^S (if symptomatic); this rate is applied to infected PrEP users (either incubating, symptomatic, or asymptomatic) to represent detection and treatment, returning them to the susceptible state. For the purpose of this study, we assume p_l^A and p_l^S to be independent of l. Both values are set to 6 months in the baseline scenario, coherently with the actual screening interval adopted in 2018-2025, and may change thereafter depending on the considered scenario.

Enrolments, dropouts and patient population in PrEP program

The size of the PrEP cohort varies over time as people enroll and drop out of the program (see Figure 4.1b). To realistically model these dynamics, we used two complementary data sources. The full cohort data provide the total number of new enrollments per year, E_Y from 2018 to 2024, totalling 542 individuals over the period. The partial cohort data, including 123 individuals enrolled between 2018 and 2021, provides yearly enrollments $\tilde{E}_Y$ and allows to estimate yearly dropouts $\widetilde{D}_Y$ based on the date of the individual's last visit. A dropout is defined as a person whose last screening visit occurred more than nine months before December 31, 2024. The partial cohort dataset also reports information on the individual number of reported sexual partners, allowing to estimate the proportion of individuals in each SAL and year, $z_{Y,g}$. Because the questionnaire administered during PrEP visits asks for the number of unique sexual partners in the last three months, while the EMIS questionnaire adopted for the overall MSM population reports partners over the last 12 months, we aligned the two by scaling quarterly counts to an annual scale according to Table 4.3 below.

Table 4.3. *Sexual Activity Levels.* Ono-to-one correspondence between EMIS SAL and data SAL.

SAL ($\boldsymbol{l}$)	EMIS partners (last year)	PrEP partners (last 3 months)
Low - 1	$0-4$	$0-1$
Middle - 2	$5-14$	$2-3$
High - 3	≥ 15	≥ 4

The yearly dropout rate r_Y was estimated from the partial cohort dataset as:

$$r_Y = \frac{\widetilde{D}_Y}{\sum_{Z=2018}^{Y} \tilde{E}_Z - \sum_{Z=2018}^{Y-1} \widetilde{D}_Z}$$

i.e., dropouts in year Y divided by the number of patients currently in PrEP, explicitly including the new enrollees of year Y. For years equal and successive to 2024, we assume a constant dropout rate equal to the average of the annual dropout rates estimated from 2018 to 2023 (see Table 4.4; we excluded data for 2024 due possible censorships in the estimates of dropped out individuals.

Table 4.4. *Estimates of dropout rate.* Left column contains the year, central column contains the estimates of the dropout rates in percentage and right column contains the number of PrEP users with respect to the year.

Year	Dropout rate (%)	Patients in partial cohort
2018	3.12	31
2019	11.5	46
2020	2.80	104
2021	5.31	107
2022	8.41	98
2023	3.06	95

The average dropout rate was therefore computed as:

$$r = \frac{1}{6}\sum_{Z=2018}^{2023} r_Z = 5.71\%$$

From 2025 onward, we assume that the yearly enrollment rate is stable and equal to the four-year average of annual enrollments over 2021-2024 in the aggregated dataset. Annual dropouts in the full cohort are then computed recursively as:

$$D_Y = r\left(\sum_{Z=2018}^{Y} E_Z - \sum_{Z=2018}^{Y-1} D_Z\right)$$

and the number of PrEP users at the end of year Y is

$$N_Y^P = \sum_{Z=2018}^{Y} E_Z - \sum_{Z=2018}^{Y} D_Z$$

We allocate annual enrollments across SALs using proportions estimated from screening questionnaires in the partial cohort data.

$$E_{Y,g} = z_{Y,l} \cdot E_Y$$

where $v_{Y,l}$ is the proportion of individuals in PrEP reporting to be in SAL l in year Y. We assume a SAL distribution fixed at its 2024 values for the following years. This hypothesis is supported by a chi-square test of homogeneity for the SAL

composition across the last three years, finding no differences (χ^2=0.78, p=0.94)).
After assuming that the population also follows the SAL distribution from the screening questionnaires ($N_{Y,l}^P = z_{Y,l} \cdot N_Y^P$), we estimated the number of dropouts per year and SAL recursively as the number of individuals present at the beginning of the year, updated by the difference between new enrolments and the number observed at the end of the year:

$$D_{Y,l} = N_{Y-1,l}^P + E_{Y,l} - N_{Y,l}^P$$

The estimated annual enrollments and dropouts for each SAL were spread uniformly within each year and assigned across epidemiological states proportionally to the number of individuals in each epidemiological state in the population of origin. For example, the average number of susceptible individuals of low SAL enrolling to the PrEP program in a given year was obtained by the number of enrolled individuals of low SAL multiplied by the proportion of susceptible in the general population of low SAL.

Force of infection

The force of infection (FOI) λ_l represents the rate at which susceptible individuals in SAL l acquire the infection and takes into account sexual partnerships across different SALs as well as protective behaviors such as condom use. In our case, it is defined as:

$$\lambda_l = \beta \cdot \sum_{g=1}^{N_G} M_{l,g} \left[\frac{\left(1 - \alpha_y^P + \alpha_y^P \phi\right)\left(P_g^P + I_g^{A,P} + I_g^{S,P}\right) + (1 - \alpha + \alpha\phi)\left(P_g + I_g^A + I_g^S\right)}{N_g^P + N_g} \right]$$

Where:

- β is an unknown transmissibility parameter estimated through calibration;
- M is a sexual contact matrix with entries M_{gl} defined as the (per-capita) contact rate from SAL g to SAL l;
- The terms $\left(1 - \alpha_y^P + \alpha_y^P \phi\right)$ and $(1 - \alpha + \alpha\phi)$ modulate the effect of condom use in the PrEP and non PrEP populations; in particular,
    - ϕ represent the relative infectiousness of infectious individuals using condom during sexual intercourse;
    - α is the probability of using a condom in the non-PrEP cohort;
    - α_y^P is the probability of using a condom in the PrEP cohort, depending on calendar year y, estimated for each year from partial

cohort data up to 2024, and assumed constant at the 2024 level, thereafter (Figure 4.6)

Thus, the terms $1-\alpha$ and $1-\alpha_y^P$ represent the proportion of condomless sexual intercourses, which retain full infectivity, whereas the remaining proportions of sexual intercourses contribute with a reduced infectivity ϕ.

- Individuals in an infectious state are assumed to have equal infectivity, independently of whether they are incubating or showing symptoms and of participation to the PrEP program.

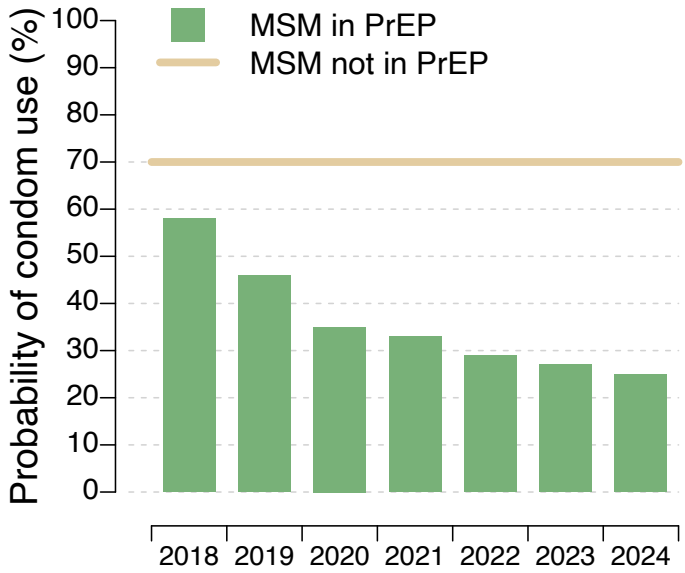


Figure 4.6. *Probability of condom use.* The brown line represents the probability of condom use during a sexual intercourse in the cohort of MSM that are not enrolled in the PrEP program, green bars represent the observed probability of condom use in the cohort of PrEP users for each year, computed as the average reported in questionnaires from the partial cohort.

The sexual contact matrix

We define a sexual contact matrix $M = \left[M_{g,l}\right]_{g,l=1}^{N_L}$, where each element $M_{g,l}$ represents the average number of sexual partners per unit time that an individual in SAL g has with individuals in risk group l. Following [172], each element is defined as:

$$M_{g,l} = c_g \cdot p_{g,l}$$

where c_g denotes the average number of sexual partners per unit time for individuals in SAL g and $p_{g,l}$ indicates the proportion of sexual partners that an individual from SAL g has with individuals in SAL l. c_g is defined by considering the average number of yearly sexual partners reported in the EMIS study within

the three SALs [166], obtaining $c_1 = 1.7$ / year, $c_1 = 8.6$ / year and $c_3 = 23.1$ / year.

$$p_{g,l} = (1 - \omega) \cdot \delta_{g,l} + \omega \cdot \frac{c_l \cdot N_l}{\sum_{k=1}^{N_G} c_k \cdot N_k}$$

Where N_k denotes the number of individuals in SAL k, and $\delta_{g,l}$ is the Kronecker delta, equal to 1 if $g = l$ and 0 otherwise. Parameter ω is generally unknown and it is estimated here during calibration.

Duration of immunity

To estimate the duration of immunity, we consider an empirical study where subjects with vaginal chlamydia who had been treated were found to have an Odds Ratio of reinfection of 4.0 (95% CI 1.1–25.6) compared to those with spontaneous resolution over a follow-up period of 1 to 12 months [165]
We modelled the empirical cohort as follows. Let q_S denote the probability that a susceptible individual becomes reinfected before time t, and q_R the corresponding probability for a temporary immune individual, assuming that both are subject to a constant force of infection λ. We can compute q_S as

$$q_S = \int_0^t \lambda \, e^{-\lambda s} \, ds = 1 - e^{-\lambda t}$$

where λ is the rate of being infected at any time s and $e^{-\lambda t}$ is the probability of escaping infection until time t.
In order to compute q_R, we need to additionally consider the probability of having lost the immunity by time $s < t$, obtaining:

$$q_R = \int_0^t \rho \, e^{-\rho s} \int_s^t \lambda \, e^{-\lambda(u-s)} \, du \, ds = \frac{\rho}{\rho - \lambda}\left(1 - e^{-\lambda t}\right) - \frac{\lambda}{\rho - \lambda}(1 - e^{-\rho t})$$

Where ρ is the rate of losing the immunity at any time s, $e^{-\rho s}$ is the probability of retaining it until time s and $e^{-\lambda(u-s)}$ is the probability of escaping infection between time s and time $u > s$.
The odds ratio between susceptible and temporary immune individuals of developing infection before time t can be approximated under the rare events assumption ($\lambda \approx 0$) as:

$$OR^{S,R} = \frac{\frac{q_S}{(1-q_S)}}{\frac{q_R}{(1-q_R)}} \approx \frac{q_S}{q_R} \approx \frac{\rho t}{\rho t + e^{-\rho t} - 1}$$

Assuming that the time at which follow-up occurs is uniformly distributed during the time interval $[t_1 = 1\,month,\ t_2 = 12\,months]$, the odds ratio is obtained by

$$OR^{S,R} \approx \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{\rho t}{\rho t + e^{-\rho t} - 1} dt$$

Applying this formula to OR values estimated in [A3], we obtain a mean value of $\rho = 4.58 \cdot 10^{-3}$ days-1 (i.e., an average duration of immunity of 218 days), with a range of $6.04 \cdot 10^{-4}$ (corresponding to 1655 days) to $8.75 \cdot 10^{-2}$ (corresponding to 11 days) when considering the OR confidence intervals of the experimental study.

Scenario implementations

To implement scenarios 6m, 3m and 12m in Section 4.3, we simply change the values of $p_l^A = p_l^S$ in such a way to reflect the screening interval. For the scenario 6m-sym, with symptom-based semestral screening, we set $p_l^A = 0$ and $p_l^S = \frac{1}{6\,months}$. Scenario 6m + DoxyPEP required a redefinition of the force of infection, where high SAL individuals in PrEP may reduce their probability of contracting infection when not using condoms:

$$\lambda_l = \beta \cdot \sum_{l=1}^{N_G} M_{l,g} \left[\frac{\left[\alpha_y^P \phi + (1-\alpha_y^P)(1-\chi_g \psi) + (1-\alpha_y^P)\chi_g \psi (1-\xi)\right](P_g^P + I_g^{A,P} + I_g^{S,P}) + (1 + \alpha\phi - \alpha)(P_g + I_g^A + I_g^S)}{N_g^P + N_g} \right]$$

where:

- χ_g is the coverage of DoxyPEP by SAL, set to zero for low and middle SAL and to 89% for the upper SAL based on an open-label randomized trial in MSM and transgender women [163]
- ψ is the adherence to DoxyPEP (i.e., the probability that an individual under DoxyPEP prescription will actually take doxycycline after a risky sexual contact), set to 86% [163])
- ξ is the efficacy of DoxyPEP in preventing chlamydia, set to 88% [163].

Estimation od reproduction numbers

We adopted the Next Generation Matrix (NGM) approach [7] to estimate reproduction numbers at different time points. The NGM K is defined as:
$K = F \cdot V^{-1}$,
where Jacobian matrices $F = \left[F_{g,l}\right]_{g,l}$ (the flux of new infected individuals in a specific time) and $V = \left[V_{g,l}\right]_{g,l}$ (transitions of the infected individuals), are given by

$$F_{g,l} = S_g \beta M_{gl} \frac{1+\alpha\phi-\alpha}{N_l} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad V_{g,l} = \delta_{gl} \cdot \begin{bmatrix} \eta & 0 & 0 \\ -\eta\sigma & -\gamma & 0 \\ -\eta(1-\sigma) & 0 & -\gamma^S \end{bmatrix} \quad g, l \in 1,2,3 = N_L$$

and δ_{gl} is the Kronecker's delta.
When all the population is susceptible and no intervention is implemented (i.e., at the beginning of our simulations), the spectral radius of this next generation matrix corresponds to the basic reproduction number:

$$R_0 = Sp(K)$$

In all other cases, we estimate the effective reproduction number R_{eff} as the spectral radius of K at different time points.
The basic reproduction number is directly proportional to the value of the transmission rate β, estimated during calibration, and is more easily interpretable: for this reason, we reported the value of the reproduction number in Table 4-1.

Stochastic model framework

The compartmental model has been implemented within a stochastic framework, specifically using both discrete time steps and discrete system states. For each of N_S numerical simulations, the stochastic version of the model was run over the time interval from T_{ini} to T_{end}, discretized in N_τ fixed timesteps of size $\tau = \frac{T_{end}-T_{ini}}{N_\tau}$. We define the discrete time points as $t_i = T_{ini} + i \cdot \tau$, where $i = 0, \dots, N_\tau$. The transitions of the individuals in the N_C different compartments X_n $(n = 1, \dots, N_C)$ are described in the stochastic model by the following equation:

$$X_n(t_{i+1}) = X_n(t_i) + F_{X_n}(t_i)$$

Where $X_n(t_i)$ is the number of individuals in the compartment X_n at the time t_i and $F_{X_n}(t_i)$ is the total number of individuals that at the time t_i enter from other compartments less the total number of individuals that exit, as described below

$$F_{X_n}(t_i) = \sum_{k=1}^{N_C} \Big(f_{X_k \to X_n}(t_i) - f_{X_n \to X_k}(t_i) \Big).$$

Where $f_{X_k \to X_n}(t_i)$ is the flux of individuals that enter from the compartment X_k to the compartment X_n at time t_i. Assuming a time step enough small we write the flux as

$$f_{X_k \to X_n}(t_i) \sim Binomial\big(n = X_k(t_i);\ \ prob = \tau \cdot r_{X_k \to X_n}\big)$$

Model calibration

The model was calibrated using Markov Chain Monte Carlo (MCMC) with a Metropolis-Hastings sampling to estimate the posterior distributions of key parameters: the transmissibility β (directly proportional to the basic reproduction number R_0, given all other parameters), the assortativity parameter ω, and the probability of remaining asymptomatic σ. Calibration was performed using data on new positive cases identified within the PrEP program between 2018 and 2024, including their distribution by SAL and symptom presentation. The overall likelihood used in the algorithm was defined as the product of the likelihoods corresponding to each distinct type of data.

$$L = P_B(x = D_V, size = D, p = p_V^M) \cdot P_M(q = D_G, size = D, p = p_G^M) \cdot \prod_{Y=2018}^{2024} P_P(x = D_Y, \lambda = D_Y^M)$$

Where:

- P_B is the density probability function of the Binomial distribution;
- P_M is the density probability function of the Multinomial distribution;
- P_P is the density probability function of the Poisson distribution;
- D is the observed total number of diagnoses between 2018 and 2024;
- D_V is the observed number of symptomatic infections;
- D_G is a vector containing the observed numbers of diagnoses by SAL;
- D_Y is a vector containing the observed number of yearly diagnoses;
- p_V^M is the model-estimated proportion of symptomatic infections
- p_G^M is a vector containing the model-estimated proportions of diagnoses by SAL;
- D_Y^M is a vector containing the model-estimated number of yearly diagnoses.

Figure 4.7 shows the estimated posterior distributions of parameters and the traceplots of the MCMC chain.

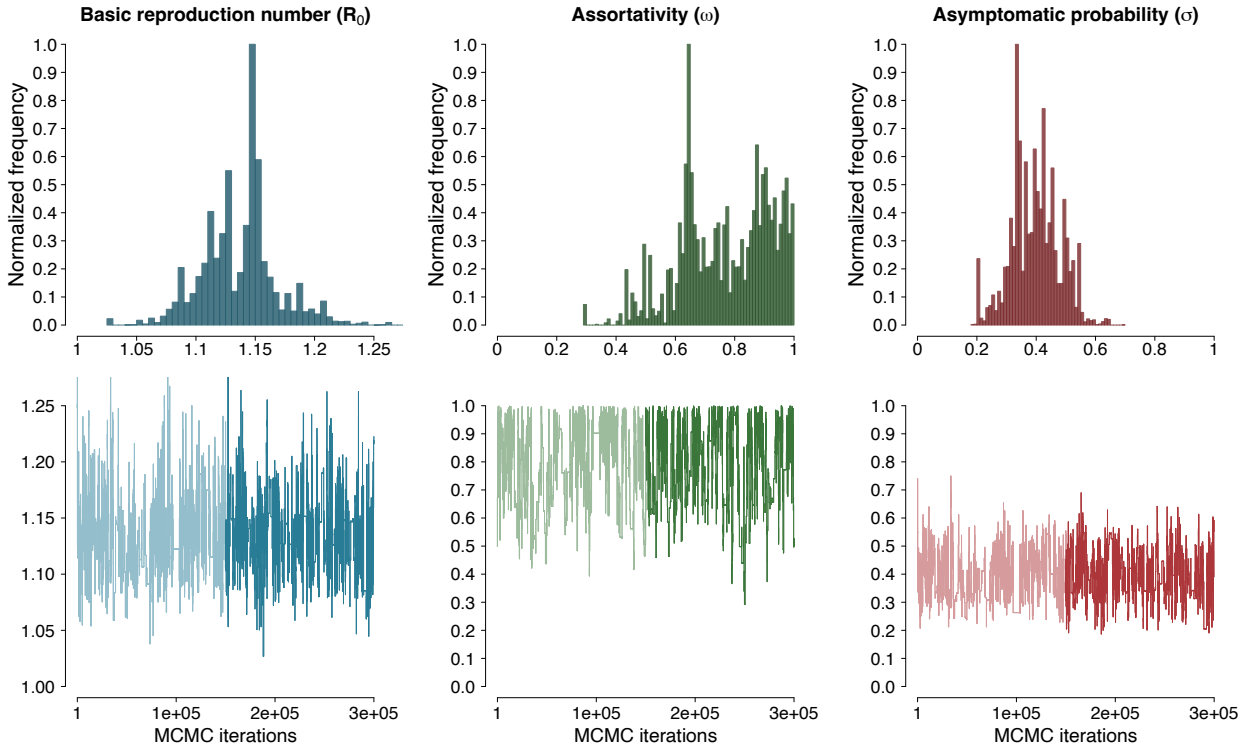


Figure 4.7. *Estimates of the calibrated parameters.* The top row shows the estimated parameter distributions, the bottom row the accepted values during the MCMC algorithm. Lighter colors represent the burn-in phase, while darker colors indicate the selected parameter estimates. Left, in blue: basic reproduction number; middle, in green: assortativity parameter; right, in red: probability of remaining asymptomatic.

ADDITIONAL RESULTS

DoxyPEP to prevent chlamydia reinfection

Here, we evaluate an additional scenario in which DoxyPEP is prescribed only to PrEP users who have a documented CT infection since PrEP enrolment (scenario 6m-DoxyPEP-R, where R stands for recurrent chlamydia). To model this intervention, we extend the PrEP dynamics with additional compartments to track individuals who have already received a prior positive diagnosis of CT. The flowchart is shown in Figure 4.8.

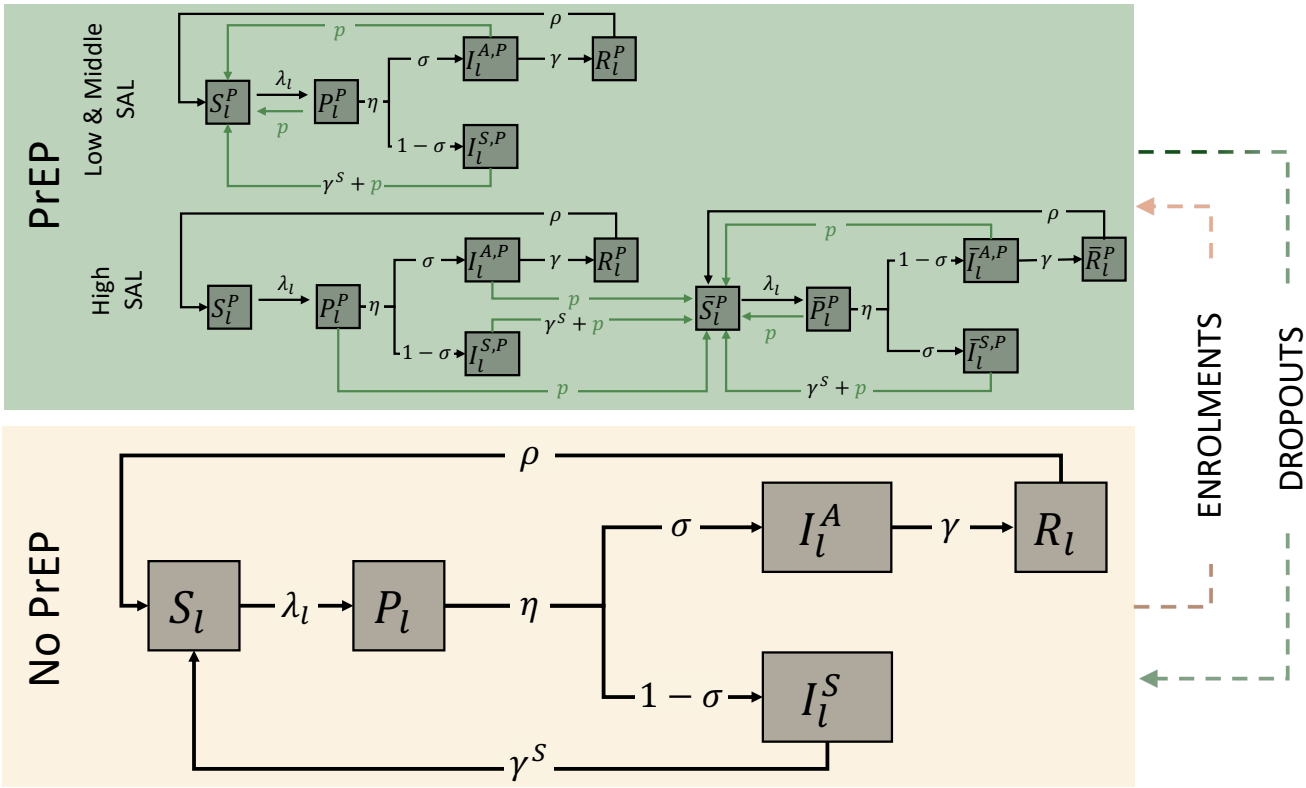


Figure 4.8. *Flowchart of scenario doxyPEP-R.* Compartmental model schematic representing the dynamics of infection and treatment stratified by PrEP status and SAL (l). The model includes two main blocks: the PrEP group (upper panel, green background) and the No PrEP group (lower panel, beige background). Each compartment denotes a different epidemiological state: susceptible (S), exposed (P), asymptomatic infected (I^A), symptomatic infected (I^S), and immunized (R), with superscripts and subscripts indicating group (l) and PrEP status (P for PrEP). Compartments denoted with a bar (e.g., $\bar{S}$) contain high SAL individuals who have been previously diagnosed with a CT infection.

The FOI was modified accordingly:

$$\lambda_l = \beta \cdot \sum_{l=1}^{N_L} M_{l,g} \left[\frac{(1+\alpha^P\phi-\alpha^P)\left(P_g^P + I_g^{A,P} + I_g^{S,P}\right) + (1+\alpha\phi-\alpha)\left(P_g + I_g^A + I_g^S\right)}{N_g^P + N_g} \right] +$$
$$\beta \cdot \frac{\left[\alpha_y^P\phi + \left(1-\alpha_y^P\right)\left(1-\chi_{N_L}\psi\right) + \left(1-\alpha_y^P\right)\chi_{N_L}\psi(1-\xi)\right]\left(\bar{P}_{N_L}^P + \bar{I}_{N_L}^{A,P} + \bar{I}_{N_L}^{S,P}\right)}{N_{N_L}^P + N_{N_L}}$$

since, in this scenario, doxy-PEP is prescribed only to PrEP users who experience a recurrent CT episode, coverage is zero at 2025 and rises as the PrEP cohort accumulates recurrent CT diagnoses. We show that, under calibrated parameters, DoxyPEP coverage among high SAL individuals reaches 64.3% (95% Prl: 53.9-71.6%) by year 10 of implementation (Figure 4.9), which is below the 89% achieved in the scenario where doxy-PEP prescription

is not conditioned on prior CT. The averaged coverage over the period 2025-2035 is 45.7% (95% PrI: 2.3-68.2%).

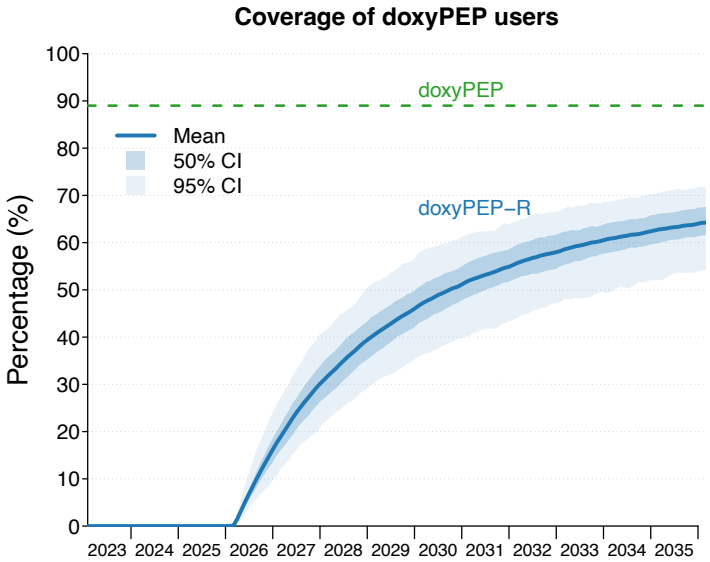


Figure 4.9. *Doxy-PEP coverage when prescribed after a diagnosed chlamydia infection.* Left: coverage among high SAL PrEP users. The dashed green line denotes scenario 6m-DoxyPEP in the main analysis, where doxy-PEP prescription is independent of prior CT infection. For the 6m-DoxyPEP-R scenario, the solid blue line shows the mean coverage, with blue bands indicating the 50% (darker) and 95% (lighter) confidence intervals.

Restricting doxy-PEP to high-SAL PrEP users with prior CT yields smaller reductions than prescribing irrespective of prior CT. Over 10 years, Doxy-PEP-R reduces incidence by 44.9% (95% PrI: 35.6-69.8%), symptomatic prevalence by 45.1% (95% PrI: 41.2-59.9%). In contrast, the unrestricted prescription scenario achieves 61.2% (95% PrI: 50.6-88.9%), 62.7% (95% PrI: 55.9-88.6%), and 62.7% (95% PrI: 56.0-88.9%) reductions for the same endpoints, respectively (see Figure 4.10).

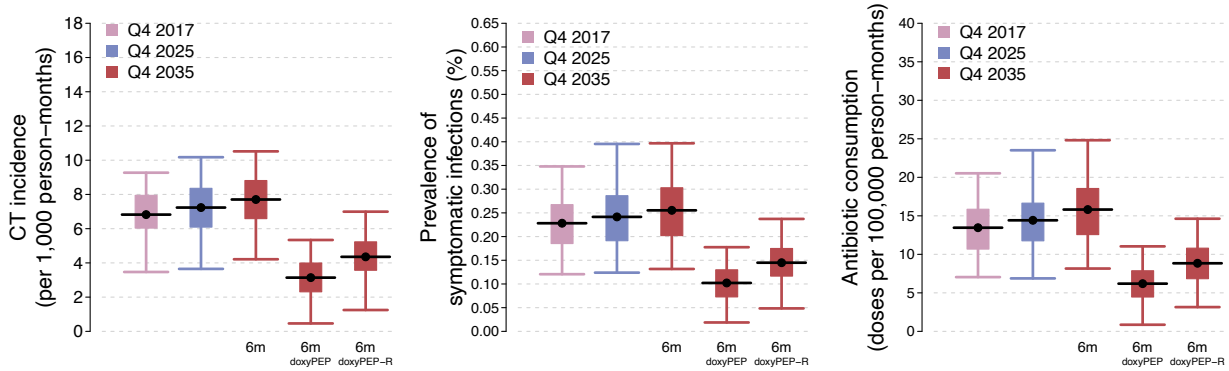


Figure 4.10. *Impact of alternative scenarios for CT control.* a) CT incidence; b) prevalence of symptomatic infections; c) CT-specific antibiotic consumption. Boxplots represent the model-estimated mean (horizontal line), 50% PrI (boxes) and 95%PrI (whiskers). Horizontal dashed lines represent mean estimates for 2035 under the current protocol (scenario 6m), reported as a reference for comparisons with past values and with doxy-PEP alternative scenarios.

Sensitivity analyses

We investigated alternative assumptions concerning the MSM population in the province of Verona and the duration of immunity. Specifically, we explored the impact of reducing the assumed proportion of MSM by half, considering 5% of the total male population in sexually active ages instead of the baseline value of 10%. Then, we considered the lower and upper estimates for the values of the average duration of immunity, i.e. 11 days and 1655 (about 4 years and half). Figure 4.11 shows a comparison of the posterior distribution of parameters, which remain roughly stable, except for a substantially higher basic reproduction number estimated for the case of longer duration of immunity. Figures 4.12–14 show the result over all scenarios and endpoints, showing that a smaller proportion of MSM or a shorter duration of immunity would imply a higher effect of PrEP, exacerbating the size of estimated effects. A longer duration of immunity, however, results in milder differences among scenarios, given the fact that a significant part of individuals with high SAL will be removed from transmission dynamics at any time.

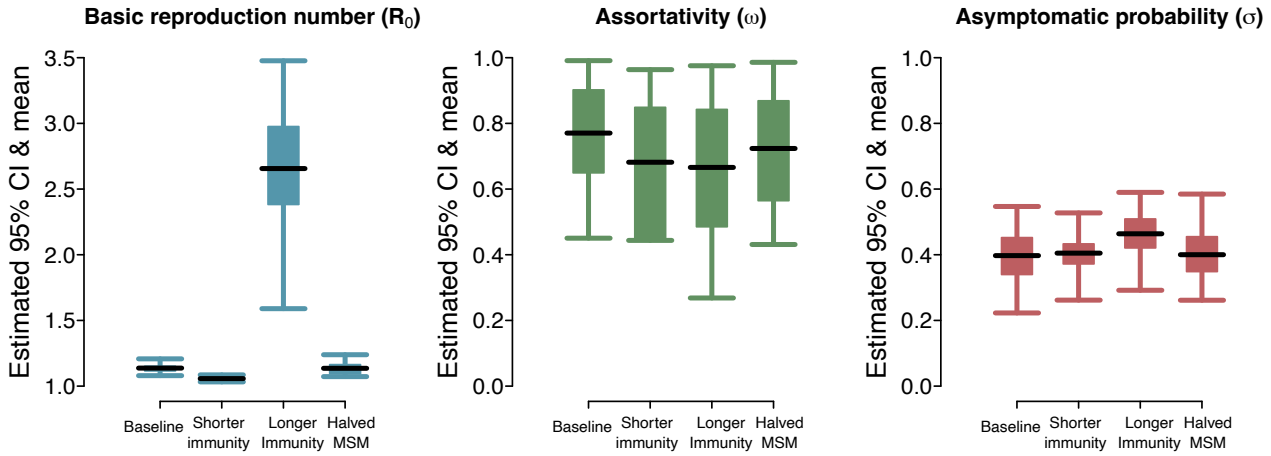


Figure 4.11. *Posterior distribution of the parameters in the calibration.* The left panels show the estimates of the basic reproduction number. The middle panel presents the estimates of the assortativity parameter. The right panel displays the posterior distribution of the probability of remaining asymptomatic.

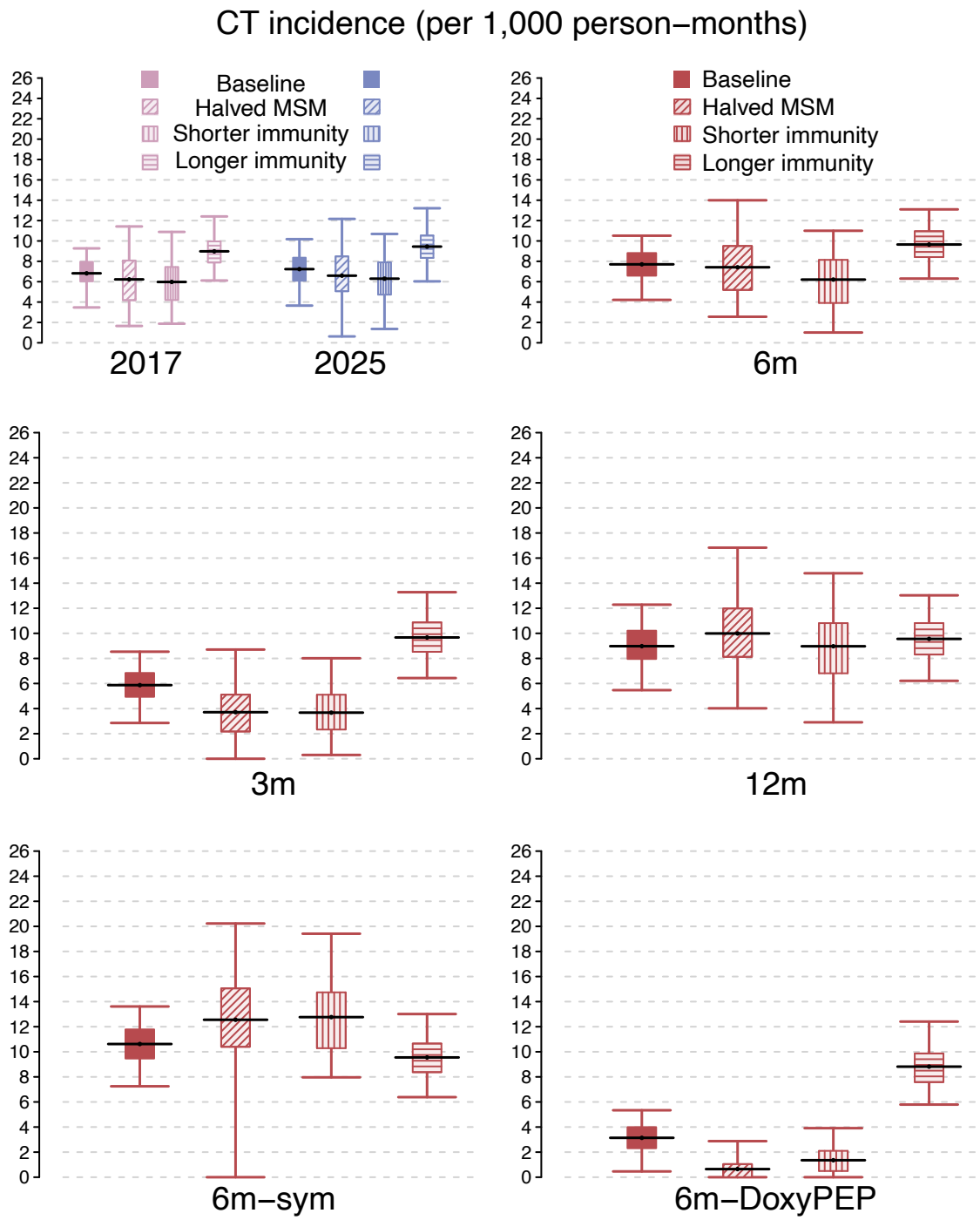


Figure 4.12. *Chlamydia Trachomatis incidence.* Comparative boxplots of the CT incidence under five scenarios and four sensitivity analysis setting. Boxes show mean and IQR; vertical lines denote the 95% CI.

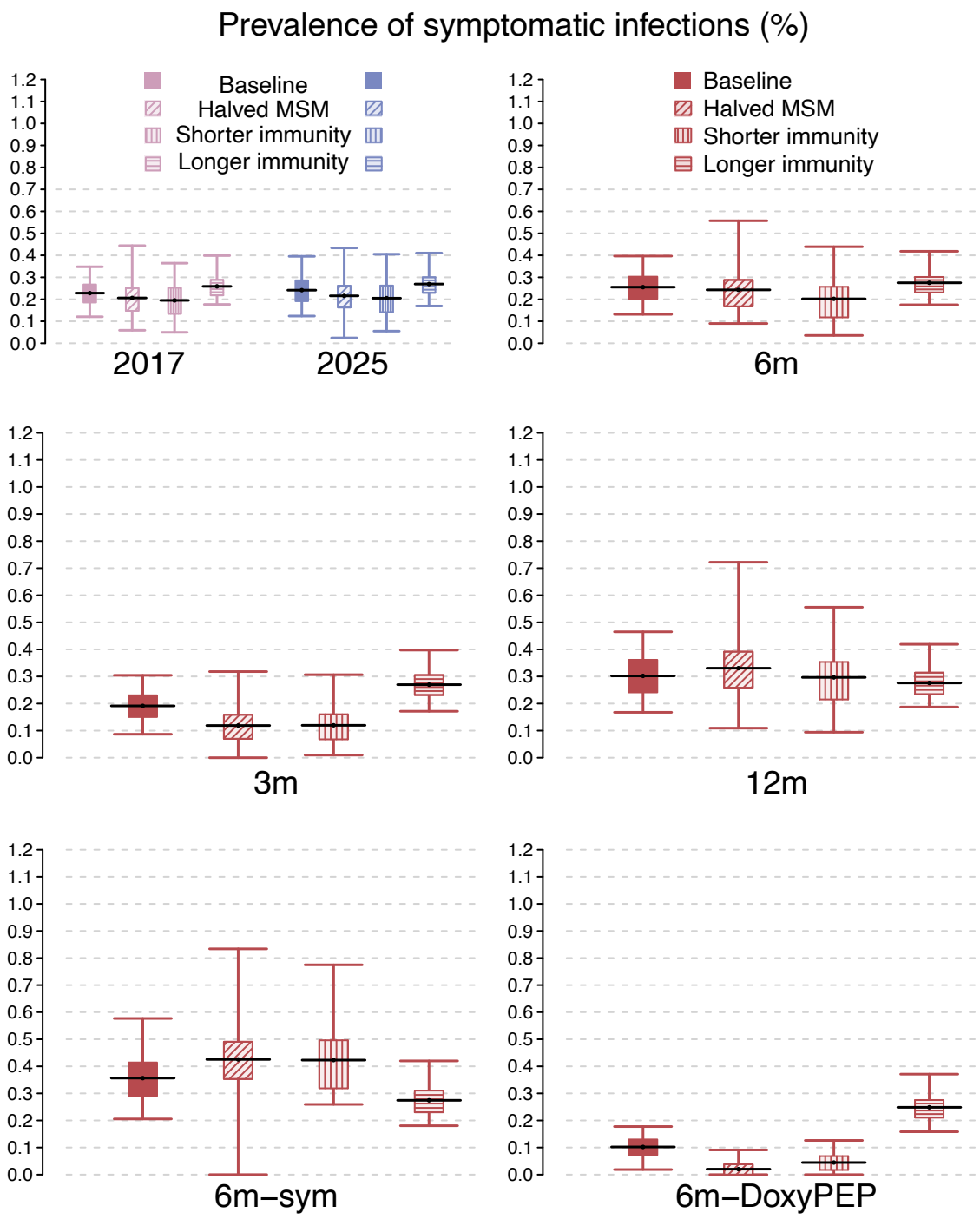


Figure 4.13. *Prevalence of symptomatic infections.* Comparative boxplots of the symptomatic prevalence under five scenarios and four sensitivity analysis setting. Boxes show mean and IQR; vertical lines denote the 95% CI.

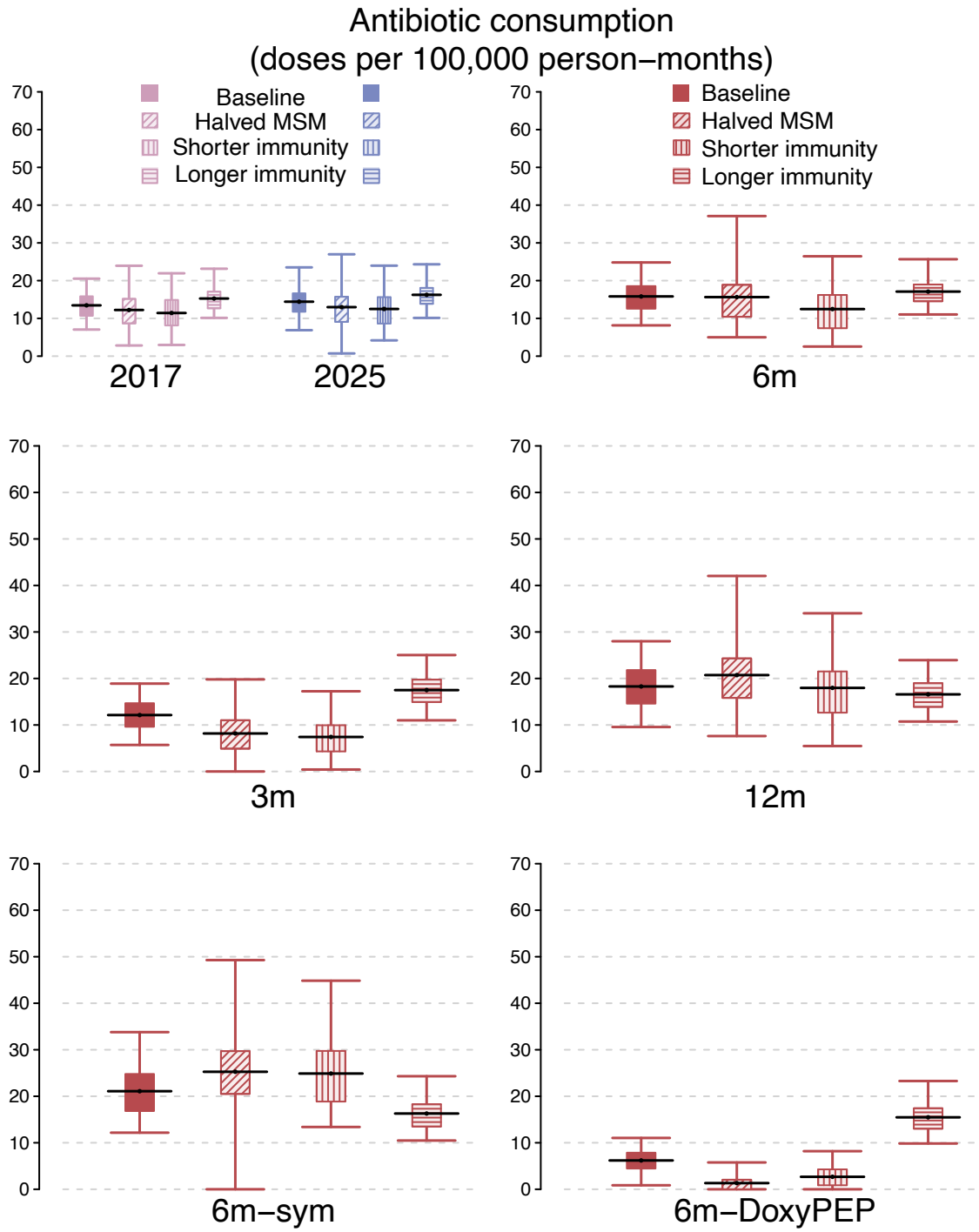


Figure 4.14. *Antibiotic consumption.* Comparative boxplots of the antibiotic consumption under five scenarios and four sensitivity analysis setting. Boxes show mean and IQR; vertical lines denote the 95% CI.

5. CONCLUSIONS

This thesis examines different problems of infectious-disease transmission, spanning different pathogen types, transmission mechanisms, populations, geographical contexts, and aims, through a variety of modelling methodologies. It focuses on nowcasting for surveillance, short-term forecasting, and policy evaluation for sexually transmitted infections control, using simple, mechanism-aware models fitted to routinely collected data. Although each chapter tackles a different problem with different tools, a common thread links them all: using mathematical models to turn surveillance data into coherent and robust inference and to inform evidence-based public health decisions for both real-time and prospective assessment of infectious diseases.
Here, I synthesize the core results and assess their implications for science and practice.

In the first chapter, I validated an operational nowcasting pipeline by measuring its impact on the accuracy of the reproduction number estimate within a retrospective setup that emulated real-time use. Correcting right-truncation in surveillance data yielded timelier and more accurate estimates when benchmarked against a consolidated reference over 18 months of Italy's COVID-19 data, provided by the national public-health authority (Istituto Superiore di Sanità). I quantified consolidation lags and compared standard versus nowcasted reproduction number. At the same 13-day median lag, nowcasting more than halved the mean absolute error relative to net (non-adjusted) estimates, and it maintained superior accuracy even at an 8-day median lag. Crucially, detection of sustained epidemic growth was advanced by 6–23 days, strengthening real-time situational awareness. The analysis also supports the operational choice of about 14-day reporting lag for the reproduction number, adopted by the Istituto Superiore di Sanità during weekly reporting for COVID-19. The work also shows the practical value of adopting nowcasted reproduction number alongside standard estimates for monitoring infectious diseases. These results have clear implications for pandemic preparedness: validated real-time nowcasting can make existing surveillance more actionable during future emergencies. However, its performance is good as long as reporting delays are estimated accurately and updated regularly. Because delays are time-varying and estimated from recent data, a key next step is to model and predict reporting delays in real time,

potentially using external covariates. Finally, this chapter benchmarks only one simple method; more sophisticated algorithms have been recently proposed and should be comparatively evaluated on common ground truths, then operationalized for use under shifting delay distribution.

In the second chapter I tested and validated an automated real-time forecasting system for norovirus outbreaks on cruise ships using routinely collected onboard data provided by a cruise line company. I estimated that the cabin isolation was very effective: an isolated case transmitted about 94% less than a non-isolated one. I compared simple renewal-based, semi-mechanistic models with a naive baseline across three operational endpoints: cases by symptoms onset, cases by diagnosis, and residual outbreak size. Models that included the effects of case isolation and transmission heterogeneity (superspreading) performed better than the baseline, with the biggest gains for longer horizons: beyond 72 hours, up to 78% of forecasts were more accurate than the baseline. Forecasts of the final outbreak size remained challenging, but the best model hit the true value within the 95% predictive interval in 56% of times (vs 14% for the baseline). This work was performed in collaboration with University of Thessaly, and the forecasting pipeline is now being implemented by the National and Kapodistrian University of Athens within the Healthy Sailing project for prospective use within a prototype syndromic surveillance software. The performance of the implemented prototype will likely benefit from the semi-closed nature of cruise ships and clear isolation rules; generalization depends on data quality and local practices, so endpoints and calibration should be adapted. A natural next step might be to extend this approach to other semi-closed environments, such as schools, prisons, or care homes, and adapt it to the specific settings.

In the third chapter I analyzed Chlamydia Trachomatis (CT) testing policies using a stochastic transmission model calibrated to real data from men who have sex with men, in Pre-Exposure-Prophylaxis (PrEP) program, in a northern province of Italy. The work has been done in collaboration with clinicians that provided data and supported the definition of alternative screening scenarios and doxycycline post-exposure prophylaxis (DoxyPEP) prescriptions. The current semestral screening keeps transmission roughly stable because of the balance between two effects, i.e., the shortening of infectious periods due to screening and treatment and the reduced condom use among PrEP users. Changing the frequency shifts this balance: moving to annual screening is projected to increase CT incidence and disease burden use by 18% within 10 years, whereas quarterly screening would decrease them by 25%. Screening

and treating asymptomatic infections are crucial: removing this component leads to 35% more antibiotic use overall and 40% more symptomatic burden. We also estimate a low prevalence of temporary immunity (≤ 11%) even in high-activity groups (≥ 15 partners/year). Results are sensitive to the assumed duration of immunity: if immunity lasted about 5 years (vs about 7 months in the main analysis), the impact of screening protocols would be minimal, and a symptom-based approach might be acceptable. Targeted DoxyPEP for high-activity PrEP users can push transmission below threshold and reduce burden and CT-specific antibiotic use, but feasibility and resistance risks must be weighed. Although PrEP users remain a small share of the total population, they remain a large fraction of the highest-activity group, which explains why policy changes focused on PrEP can shift population-level outcomes. Limitations include uncertainties in sexual behaviour (e.g., stigma-related underreporting and partner distributions) and key biology parameters (e.g., immunity duration). In addition, a model that explicitly represents oral and anal sex could reproduce the observed rates of single- and multi-site infection at the oropharynx, urethra, and rectum, providing clearer insight into site-specific transmission and coinfection dynamics. More broadly, closing knowledge of surveillance gaps on contact patterns and immunity in STIs is essential to avoid blind spots against emerging sexually transmitted epidemics and pandemics.

This thesis addresses diverse problems in infectious disease transmission with a single aim: to use simple, clear models, built on surveillance data and reflecting key transmission mechanisms, to improve real-time understanding and support future policy. These contributions also matter for pandemic preparedness: the first shows how existing tools can make real-time surveillance more informative; the second advances the intrinsically hard but strategic area of short-term forecasts; and the third improves knowledge on sexually transmitted infections, which constitute a weak spot in pandemic preparedness, as recently demonstrated by the global spread of mpox in 2022. Together, these studies demonstrate how basic mathematical principles, applied carefully to real surveillance data, can translate complexity into action, improving public health now and strengthening preparedness for future epidemics and pandemics.

6. BIBLIOGRAPHY

[1] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole.

[2] Ross R. The Prevention of Malaria. 2nd ed. London: John Murray; 1911.

[3] Kermack W. O. and McKendrick A. G. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London, series A, vol. 115, no. 772, 1927

[4] Anderson, R. M., & May, R. M. (1991). Infectious Diseases of Humans: Dynamics and Control. Oxford University Press.

[5] Keeling, M. J., & Rohani, P. (2008). Modeling Infectious Diseases in Humans and Animals. Princeton University Press.

[6] Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., et al. (2015). Modeling infectious disease dynamics in the complex landscape of global health. Science, 347(6227), aaa4339. https://doi.org/10.1126/science.aaa4339

[7] Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. Journal of the Royal Society Interface, 7(47), 873–885. https://doi.org/10.1098/rsif.2009.0386

[8] Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. American Journal of Epidemiology, 160(6), 509–516. https://doi.org/10.1093/aje/kwh255

[9] Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006).The role of the airline transportation network in the prediction and predictability of global epidemics. Proceedings of the National Academy of Sciences, 103(7), 2015–2020.

[10] Colizza, V., & Vespignani, A. (2007). Invasion threshold in heterogeneous metapopulation networks. Physical Review Letters, 99(14), 148701. https://doi.org/10.1103/PhysRevLett.99.148701

[11] Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. Science, 303(5656), 327–

332. https://doi.org/10.1126/science.1090727

[12] Germann, T. C., Kadau, K., Longini, I. M. Jr., & Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. Proceedings of the National Academy of Sciences, 103(15), 5935–5940. https://doi.org/10.1073/pnas.0601266103

[13] Hethcote, H. W. (2000). The Mathematics of Infectious Diseases. SIAM Review, 42(4), 599–653. https://doi.org/10.1137/S0036144500371907

[14] Brauer, F., & Castillo-Chavez, C. (2012). Mathematical Models in Population Biology and Epidemiology (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-1686-9

[15] Keeling, M. J., & Danon, L. (2011). Mathematical Modelling of Infectious Diseases. British Medical Bulletin, 92, 33–42. https://doi.org/10.1093/bmb/ldr010

[16] Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., & Burke, D. S. (2006). Strategies for Mitigating an Influenza Pandemic. Nature, 442(7101), 448–452. https://doi.org/10.1038/nature04795

[17] Bailey, N. T. J. (1975). The Mathematical Theory of Infectious Diseases (2nd ed.). Charles Griffin.

[18] Andersson, H., & Britton, T. (2000). Stochastic Epidemic Models and Their Statistical Analysis. Springer.

[19] Allen, L. J. S. (2008). An Introduction to Stochastic Epidemic Models. In: F. Brauer et al. (eds) Mathematical Epidemiology. Lecture Notes in Mathematics, vol 1945. Springer.

[20] Daley, D. J., & Gani, J. (1999). Epidemic Modelling: An Introduction. Cambridge University Press.

[21] Eubank S, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004;429:180–184.

[22] Halloran ME, Ferguson NM, Eubank S, et al. Modeling targeted layered containment of an influenza pandemic in the United States. PNAS. 2008;105(12):4639–4644.

[23] Longini IM Jr., Nizam A, Xu S, et al. Containing pandemic influenza at the source. Science. 2005;309(5737):1083–1087.

[24] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. PNAS. 2009;106(51):21484–21489.

[25] Keeling MJ, Eames KTD. Networks and epidemic models. J R Soc Interface. 2005;2(4):295–307.

[26] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett. 2001;86(14):3200–3203.

[27] Newman MEJ. Spread of epidemic disease on networks. Phys Rev E. 2002;66:016128.

[28] Salathé M, Jones JH. Dynamics and control of diseases in networks with community structure. PLoS Comput Biol. 2010;6(4):e1000736.

[29] Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers. Am J Epidemiol. 2013;178(9):1505–1512.

[30] Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. PNAS. 2012;109(50):20425–20430.

[31] Reich NG, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PNAS. 2019;116(8):3146–3154.

[32] Bergström F, Günther F, Höhle M, Britton T (2022) Bayesian nowcasting with leading indicators applied to COVID-19 fatalities in Sweden. PLOS Computational Biology 18(12): e1010767. https://doi.org/10.1371/journal.pcbi.1010767

[33] Höhle M, an der Heiden M. Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. BMC Med Res Methodol. 2014;14:38.

[34] R.N. Thompson, J.E. Stockwin, R.D. van Gaalen, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks, Epidemics, Volume 29, 2019, 100356, ISSN 1755-4365, https://doi.org/10.1016/j.epidem.2019.100356.

[35] Sebastian Funk, Anton Camacho, Adam J. Kucharski, et al. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model, Epidemics, Volume 22, 2018, Pages 56-61, ISSN 1755-4365, https://doi.org/10.1016/j.epidem.2016.11.003.

[36] Paireau J, Andronico A, Hozé N, et al. An ensemble model based on early predictors to forecast COVID-19 health care demand in France. Proc Natl Acad Sci U S A. 2022 May 3;119(18):e2103302119. doi: 10.1073/pnas.2103302119. Epub 2022 Apr 27. PMID: 35476520; PMCID: PMC9170016.

[37] Marziano V, Guzzetta G, Rondinone BM, et al. Retrospective analysis of the Italian exit strategy from COVID-19 lockdown. Proc Natl Acad Sci U S A. 2021 Jan 26;118(4):e2019617118. doi: 10.1073/pnas.2019617118. PMID: 33414277; PMCID: PMC7848712.

[38] Vollset, Stein Emil et al. Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting analysis for the Global Burden of Disease Study 2021. The Lancet, Volume 403, Issue 10440, 2204 - 2256

[39] Holmdahl I, Buckee C. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. New England Journal of Medicine. 2020;383(4):303–305. doi:10.1056/NEJMp2016822.

[40] Jewell NP, Lewnard JA, Jewell BL. Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. JAMA. 2020;323(19):1893–1894. doi:10.1001/jama.2020.6585.

[41] McGough SF, Johansson MA, Lipsitch M, Menzies NA. Nowcasting by Bayesian Smoothing: A Flexible, Generalizable Model for Real-Time Epidemic Tracking. PLoS Computational Biology. 2020;16(4):e1007735. doi:10.1371/journal.pcbi.1007735.

[42] Lawless JF. Adjustments for reporting delays and the prediction of occurred but not reported events. Canadian Journal of Statistics. 1994;22(1):15–31. doi:10.2307/3315826.

[43] Bastos LS, Economou T, Gomes MFC, et al. A modelling approach for correcting reporting delays in disease surveillance data. Statistics in Medicine. 2019;38(22):4363–4377. doi:10.1002/sim.8303.

[44] Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the effective reproductive number, Rt. PLoS Computational Biology. 2020;16(12):e1008409. doi:10.1371/journal.pcbi.1008409.

[45] Seaman SR, Samartsidis P, Kalli M, et al. Estimating a time-to-event distribution from right-truncated data, with application to COVID-19. Statistics in Medicine. 2021;40(24):5329–5345. doi:10.1002/sim.9071.

[46] Li T, White LF, Fitzpatrick MC, et al. Bayesian back-calculation and nowcasting for line list data. PLoS Computational Biology. 2021;17(7):e1009210. doi:10.1371/journal.pcbi.1009210.

[47] Beesley LJ, Tegunov D, Fox SJ, Meyers LA. Addressing delayed case reporting in infectious disease forecast modeling. Epidemics. 2022;40:100614. doi:10.1016/j.epidem.2022.100614.

[48] Bizzotto A, Guzzetta G, Marziano V, et al. Increasing situational awareness through nowcasting of the reproduction number. Front Public Health. 2024.

[49] Viboud C, Sun K, Gaffey R, et al. The RAPIDD Ebola forecasting challenge: Synthesis and lessons learnt.Epidemics. 2018;22:13–21. doi:10.1016/j.epidem.2017.08.003

[50] Johansson MA, Apfeldorf KM, Dobson S, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. PNAS. 2019;116(48):24268–24274. doi:10.1073/pnas.1909865116.

[51] [Jenkins KA, Mintz ED, Newton AE, et al. Acute Gastroenteritis on Cruise Ships—Maritime Illness and Death Reporting System, United States, 2006–2019. MMWR Surveill Summ. 2021;70(6):1–12. doi:10.15585/mmwr.ss7006a1.

[52] Isakbaeva ET, Widdowson M-A, Beard RS, et al. Norovirus Transmission on Cruise Ship. Emerg Infect Dis.2005;11(1):154–158. doi:10.3201/eid1101.040434.

[53] Towers S, Chen J, Cruz C, et al. Quantifying the relative effects of environmental and direct transmission of norovirus. R. Soc. Open Sci. 2018;5(3):170602. doi:10.1098/rsos.170602.

[54] Bizzotto A, De Bellis A, Marziano V, et al. Forecasting norovirus cases on cruise ships to support outbreak management on board. Travel Med Infect Dis. 2025

[55] Runge MC, Shea K, Brook CE, et al. Scenario design for infectious disease projections. Epidemics.2024;47:100756. doi:10.1016/j.epidem.2024.100756.

[56] Howerton E, Ray EL, Kirienko N, et al. Evaluation of the US COVID-19 Scenario Modeling Hub for multiple waves of transmission. Nature Communications. 2023;14:7070. doi:10.1038/s41467-023-42680-x.

[57] [Flaxman S, Mishra S, Gandy A, et al. Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe. Nature. 2020;584:257–261. doi:10.1038/s41586-020-2405-7.

[58] Davies NG, Barnard RC, Jarvis CI, et al. Association of Tiered Restrictions and a Second Lockdown with COVID-19 Deaths and Hospital Admissions in England: A Modelling Study. The Lancet Infectious Diseases. 2021;21(4):482–492. doi:10.1016/S1473-3099(20)30984-1.

[59] Jenness SM, Weiss KM, Goodreau SM, et al. Incidence of Gonorrhea and Chlamydia Following HIV PrEP Among MSM: A Modeling Study. Clinical Infectious Diseases. 2017;65(5):712–718. doi:10.1093/cid/cix439.

[60] Xiridou M, Vriend HJ, Lugner A, et al. Modelling the impact of chlamydia screening on the transmission of HIV among MSM. BMC Infectious Diseases. 2013;13:436. doi:10.1186/1471-2334-13-436.

[61] Rönn MM, Althaus CL, Heijne JCM, et al. The use of mathematical models of chlamydia transmission to address public health policy questions: a systematic review. Sexually Transmitted Diseases. 2017;44(5):278–283.

[62] Fiandrino S, Bizzotto A, Guzzetta G, et al. Collaborative forecasting of influenza-like illness in Italy: The Influcast experience. Epidemics. 2025.

[63] De Bellis A, Bizzotto A, Anagnostopoulou L, et al. Mitigating norovirus spread on cruise ships: A model-based assessment of diagnostic timing and isolation. J Travel Med. 2025.

[64] Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B. 2007;274(1609):599–604.

[65] Svensson Å. A note on generation times in epidemic models. Mathematical Biosciences. 2007;208(1):300–311.

[66] Britton T, Scalia Tomba G. Estimation in emerging epidemics: a review. J R Soc Interface.2019;16(150):20180630.

[67] Abbott S, Hellewell J, Thompson RN, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case

counts (EpiNow2). Wellcome Open Research. 2020;5:112. doi:10.12688/wellcomeopenres.16006.2.

[68] Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. Journal of the Royal Society Interface. 2021;18(181):20210079. doi:10.1098/rsif.2021.0079.

[69] Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R0R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology.1990;28(4):365–382.

[70] van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences. 2002;180:29–48.

[71] Diekmann O, Heesterbeek JAP. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley, 2000.

[72] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. Chapman & Hall/CRC, 2013.

[73] Robert CP, Casella G. Monte Carlo Statistical Methods. 2nd ed. Springer, 2004.

[74] Tierney L. Markov Chains for Exploring Posterior Distributions. Annals of Statistics. 1994;22(4):1701–1762.

[75] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics. 1953;21(6):1087–1092.

[76] Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika. 1970;57(1):97–109.

[77] Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science. 1992;7(4):457–472.

[78] Doucet A, de Freitas N, Gordon N (eds.). Sequential Monte Carlo Methods in Practice. Springer, 2001.

[79] Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2010;72(3):269–342.

[80] Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boëlle PY. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Statistics in Medicine. 2004;23(22):3469–3487.

[81] Endo A, van Leeuwen E, Baguelin M. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. Epidemics. 2019;29:100363.

[82] Cramer EY, et al. Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the United States. Proceedings of the National Academy of Sciences (PNAS). 2022;119(15):e2113561119. doi:10.1073/pnas.2113561119.

[83] Bracher J, et al. A Pre-Registered Short-Term Forecasting Study of COVID-19 in Germany and Poland. Nature Communications. 2021;12:5173. doi:10.1038/s41467-021-25207-0.

[84] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006;22(4):679–688.

[85] Nash, R. K., Nouvellet, P. & Cori, A. Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. PLOS Digit. Health 1, e0000052 (2022).

[86] O'Driscoll, M., Harry, C., Donnelly, C. A., Cori, A. & Dorigatti, I. A Comparative Analysis of Statistical Methods to Estimate the Reproduction Number in Emerging Epidemics, With Implications for the Current Coronavirus Disease 2019 (COVID-19) Pandemic. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 73, e215–e223 (2021).

[87] Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect. Dis. 20, 793–802 (2020).

[88] Liu, Q.-H. et al. The COVID-19 outbreak in Sichuan, China: Epidemiology and impact of interventions. PLOS Comput. Biol. 16, e1008467 (2020).

[89] Heisterkamp, S. H., Jager, J. C., Ruitenberg, E. J., Van Druten, J. a. M. & Downs, A. M. Correcting reported aids incidence: A statistical approach. Stat. Med. 8, 963–976 (1989).

[90] Brookmeyer R, Liao JG. The analysis of delays in disease reporting: methods and results for the acquired immunodeficiency syndrome. Am J Epidemiol. 1990 Aug;132(2):355-65. doi: 10.1093/oxfordjournals.aje.a115665. PMID: 2372012.

[91] Harris, J. E. Reporting Delays and the Incidence of AIDS. J. Am. Stat. Assoc. 85, 915–924 (1990).

[92] Rosenberg, P. S. A simple correction of AIDS surveillance data for reporting delays. J. Acquir. Immune Defic. Syndr. 3, 49–54 (1990).

[93] Kalbfleisch, J. D. & Lawless, J. F. Regression Models for Right Truncated Data with Applications to Aids Incubation Times and Reporting Lags. Stat. Sin. 1, 19–32 (1991).

[94] Cauchemez, S. et al. Real-time Estimates in Early Detection of SARS. Emerg. Infect. Dis. 12, 110–113 (2006).

[95] Azmon, A., Faes, C. & Hens, N. On the estimation of the reproduction number based on misreported epidemic data. Stat. Med. 33, 1176–1192 (2014).

[96] Noufaily, A. et al. Modelling Reporting Delays for Outbreak Detection in Infectious Disease Data. J. R. Stat. Soc. Ser. A Stat. Soc. 178, 205–222 (2015).

[97] van de Kassteele, J., Eilers, P. H. C. & Wallinga, J. Nowcasting the Number of New Symptomatic Cases During Infectious Disease Outbreaks Using Constrained P-spline Smoothing. Epidemiol. Camb. Mass 30, 737–745 (2019).

[98] [98] Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science, 369(6500), eabb9789.

[99] Dehning, J., Mohr, S. B., Contreras, S., Dönges, P., Iftekhar, E. N., Schulz, O., ... & Priesemann, V. (2023). Impact of the Euro 2020 championship on the spread of COVID-19. Nature Communications, 14(1), 122.

[100] Brauner, J. M., Mindermann, S., Sharma, M., et al. (2021). Inferring the effectiveness of government interventions against COVID-19. Science, 371(6531), eabd9338.

[101] Sharma, M., Mindermann, S., Rogers-Smith, C., Leech, G., Snodin, B., Ahuja, J., ... & Brauner, J. M. (2021). Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. Nature communications, 12(1), 5820.

[102] Günther, F., Bender, A., Katz, K., Küchenhoff, H. & Höhle, M. Nowcasting the COVID-19 pandemic in Bavaria. Biom. J. 63, 490–502 (2021).

[103] Lison, A., Abbott, S., Huisman, J. & Stadler, T. Generative Bayesian modeling to nowcast the effective reproduction number from line list data with missing symptom onset dates. Preprint at https://doi.org/10.48550/arXiv.2308.13262 (2023).

[104] Overton, C. E. et al. Nowcasting the 2022 mpox outbreak in England. PLOS Comput. Biol. 19, e1011463 (2023).

[105] Charniga, K. et al. Nowcasting and Forecasting the 2022 U.S. Mpox Outbreak: Support for Public Health Decision Making and Lessons Learned. http://medrxiv.org/lookup/doi/10.1101/2023.04.14.23288570 (2023) doi:10.1101/2023.04.14.23288570.

[106] Riccardo, F. et al. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull. 25, 2000790 (2020).

[107] Lopez, V. K. et al. Challenges of COVID-19 Case Forecasting in the US, 2020-2021. 2023.05.30.23290732 Preprint at https://doi.org/10.1101/2023.05.30.23290732 (2023).

[108] Riccardo, F. et al. COVID-19 response: effectiveness of weekly rapid risk assessments, Italy. Bull. World Health Organ. 100, 161–167 (2022).

[109] Manica, M. et al. Estimation of the incubation period and generation time of SARS-CoV-2 Alpha and Delta variants from contact tracing data. Epidemiol. Infect. 151, e5 (2022).

[110] Manica, M. et al. Intrinsic generation time of the SARS-CoV-2 Omicron variant: An observational study of household transmission. Lancet Reg. Health Eur. 19, 100446 (2022).

[111] Riccardo, F. et al. Winning during a pandemic: epidemiology of SARS-CoV-2 during EURO2020 in Italy. Epidemiol. Infect. 150, e166 (2022).

[112] EpiCentro. Monitoraggio delle varianti del virus SARS-CoV-2 di interesse in sanità pubblica in Italia. https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti-rapporti-periodici.

[113] Pulliam, J. R. C. et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. Science 376, eabn4947 (2022).

[114] Wolter, N. et al. Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study. The Lancet 399, 437–446 (2022).

[115] Petrone, D. et al. Reduction of the risk of severe COVID-19 due to Omicron compared to Delta variant in Italy (November 2021 – February 2022). Int. J. Infect. Dis. 129, 135 (2023).

[116] Abbott, S. et al. epinowcast/epinowcast: Epinowcast 0.2.2. (2023) doi:10.5281/ZENODO.5637165.

[117] Abbott [aut, S. et al. EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters. (2023).

[118] Lytras, T. thlytras/bayEStim. (2020). McGough, S. NobBS: Nowcasting by Bayesian Smoothing. (2023).

[119] McGough S, Menzies N, Lipsitch M, Johansson M. Nowcasting by Bayesian Smoothing. R package NobBS (2024). Available at: https://CRAN.R-project.org/package=NobBS (accessed April 1, 2024).

[120] Case definition for COVID-19. https://www.ecdc.europa.eu/en/infectious-disease-topics/z-disease-list/covid-19/facts/case-definition-covid-19 (2023).

[121] Italian Ministry of Health. Circolare n. 0705 "Aggiornamento della definizione di caso COVID-19 e strategie di testing" (Update on the case definition of COVID-19 and testing strategies), January 8, 2021.

[122] COVID-19, M. D. M., Xanthi Andrianou, Alberto Mateo Urdiales, Maria Fenicia Vescio, Maria Cristina Rota, Massimo Fabiani, Stefano Boros, Stefania Bellino, Paola Stefanelli, Alessandra Ciervo, Ornella Punzo, Antonietta Filia, Marco Tallon, Corrado Di Benedetto, Matteo Spuri, Patrizio Pezzotti, Flavia Riccardo, Antonella Bella, Gruppo Referenti regionali della Sorveglianza Integrata. La sorveglianza integrata Covid-19 in Italia: output e attività correlate. www.epiprev.it https://epiprev.it/5088.

[123] Cereda, D. et al. The early phase of the COVID-19 epidemic in Lombardy, Italy. Epidemics 37, 100528 (2021).

[124] Lopman BA, Steele D, Kirkwood CD, Parashar UD. The Vast and Varied Global Burden of Norovirus: Prospects for Prevention and Control. PloS Med.

2016;13(4):e1001999. https://doi.org/10.1371/journal.pmed.1001999 PMID: 27115709

[125] 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

[126] 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

[127] Mouchtouri VA, Simou E, Soteriades S, Rousou X, Kontouli KM, Kafetsouli D, Kourentis L, Anagnostopoulos L, Hadjichristodoulou C. Systematic literature review and meta- analysis on preventing and controlling norovirus outbreaks on cruise ships, 1990 to 2020: calling for behaviour change strategies of travellers. Eurosurveillance. 2024 Mar 7;29(10):2300345.

[128] de Graaf M, Villabruna N, Koopmans MP. Capturing norovirus transmission. Curr Opin Virol. 2017;22:64-70. https://doi.org/10.1016/j.coviro.2016.11.008 PMID: 28056379

[129] 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

[130] 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

[131] World Health Organization (WHO). Guide to ship sanitation. 3rd ed. Geneva: WHO; 2011. Available from: http://www.who.int/water_sanitation_health/publications/2011/ship_sanitation_guide/en/

[132] Matthew Biggerstaff, David Alper, Mark Dredze, Spencer Fox, Isaac Chun-Hai Fung, Kyle S Hick- mann, Bryan Lewis, Roni Rosenfeld, Jeffrey Shaman, Ming-Hsiang Tsou, et al. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. BMC Infectious Diseases, 16:1–10, 2016.

[133] Katharine Sherratt, Hugo Gruson, Helen Johnson, Rene Niehus, Bastian Prasse, Frank Sandmann, Jannik Deuschel, Daniel Wolffram, Sam Abbott, Alexander Ullrich, et al. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. Elife, 12:e81916, 2023.

[134] Mouchtouri VA, Verykouki E, Zamfir D, Hadjipetris C, Lewis HC, Hadjichristodoulou C. Gastroenteritis outbreaks on cruise ships: contributing factors and thresholds for early outbreak detection. Eurosurveillance. 2017 Nov 9;22(45):16-00576.

[135] Heijne JC, Teunis P, Morroy G, Wijkmans C, Oostveen S, Duizer E, Kretzschmar M, Wallinga J. Enhanced hygiene measures and norovirus transmission during an outbreak. Emerging infectious diseases. 2009 Jan;15(1):24.

[136] Good IJ. Rational Decisions. Journal of the Royal Statistical Society B, 1952; 14(1), 107–114. doi:10.1111/j.2517-6161.1952.tb00104.x.

[137] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association. 2007 Mar 1;102(477):359-78.

[138] Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005 Nov 17;438(7066):355-9.

[139] Götz H, 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 Sep 1;33(5):622-8.

[140] HEALTHY SAILING project. Prevention, mitigation, management of infectious disease on cruise ships and passenger ferries. Foundation report. 30/11/2023.

[141] 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.

[142] Holcomb KM, Mathis S, Staples JE, et al. Evaluation of an open forecasting challenge to assess skill of West Nile virus neuroinvasive disease prediction. Parasites & Vectors, 16(1):1–13, 2023.

[143] Del Valle SY, McMahon BH, Asher J, et al. Summary results of the 2014-2015 DARPA Chikungunya challenge. BMC Infectious Diseases, 18:1–14, 2018.

[144] McGowan CJ, Biggerstaff M, Johansson M, et al. Collaborative efforts to forecast seasonal influenza in the United States, 2015–2016. Scientific reports, 9(1):683, 2019.

[145] Sinka K, et al. The global burden of sexually transmitted infections. Clinics in Dermatology. 2024;42(2):110-118. https://doi.org/10.1016/j.clindermatol.2023.12.002.

[146] Rodrigues R, Sousa C, Vale N. Chlamydia trachomatis as a Current Health Problem: Challenges and Opportunities. Diagnostics (Basel). 2022;12(8):1795. doi: 10.3390/diagnostics12081795.

[147] Farley TA, Cohen DA, Elkins W, et al. Asymptomatic sexually transmitted diseases: the case for screening. Preventive Medicine. 2003;36(4):502-509. https://doi.org/10.1016/S0091-7435(02)00058-0.

[148] Workowski KA, Bolan GA, Center for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recomm Rep. 2015;64(RR-03):1-137.

[149] Carter JD, Espinoza LR, Inman RD, et al. Combination antibiotics as a treatment for chronic Chlamydia-induced reactive arthritis: a double-blind, placebo-controlled, prospective trial. Arthritis Rheum. 2010;62(5):1298-1307. doi: 10.1002/art.27394.

[150] Weidner W, Diemer T, Huwe P, et al. The role of Chlamydia trachomatis in prostatitis. Int J Antimicrob Agents. 2002;19(6):466-470. https://doi.org/10.1016/S0924-8579(02)00094-8.

[151] Wasserheit J: Epidemiological synergy. Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. Sex Transm Dis 1992, 19:61–77.

[152] Rottingen J, Cameron D, Garnett G: A systematic review of the epidemiologic interactions between sexually transmitted diseases and HIV: how much really is known? Sex Transm Dis 2001, 28:579–597.

[153] Geisler WM, Uniyal A, Lee JY, et al. Azithromycin versus Doxycycline for Urogenital Chlamydia trachomatis Infection. N Engl J Med. 2015 Dec 24;373(26):2512-21. doi: 10.1056/NEJMoa1502599. PMID: 26699167; PMCID: PMC4708266.

[154] Eaton LA, Kalichman S. Risk compensation in HIV prevention: implications for vaccines, microbicides, and other biomedical HIV prevention technologies. Curr HIV/AIDS Rep 2007; 4:165-72.

[155] Landovitz RJ, Molina J, Buchbinder SP, International Antiviral Society–USA (IAS-USA) Panel. Preexposure Prophylaxis for HIV: Updated Recommendations From the 2024 International Antiviral Society–USA Panel. JAMA. 2025;334(7):638–639. doi:10.1001/jama.2025.11410

[156] Brunham RC, Pourbohloul B, Mak S, et al. The unexpected impact of a Chlamydia trachomatis infection control program on susceptibility to reinfection. J Infect Dis. 2005;192(10):1836-1844. doi: 10.1086/497341.

[157] Brunham RC, Rekart ML. The arrested immunity hypothesis and the epidemiology of chlamydia control. Sex Transm Dis. 2008;35(1):53-54. doi: 10.1097/OLQ.0b013e31815e41a3.

[158] Williams E, et al. Frequent screening for asymptomatic chlamydia and gonorrhoea infections in men who have sex with men: time to re-evaluate? The Lancet Infectious Diseases. 2023;23(12):e558-e566.

[159] Stewart J, Baeten JM, et al. HIV pre-exposure prophylaxis and sexually transmitted infections: intersection and opportunity. Nat Rev Urol. 2022;19(1):7-15. doi: 10.1038/s41585-021-00527-4.

[160] Kojima N, Davey DJ, Klausner JD. Pre-exposure prophylaxis for HIV infection and new sexually transmitted infections among men who have sex with men. AIDS. 2016;30(14):2251-2252. doi: 10.1097/QAD.0000000000001185.

[161] Beymer MR, DeVost MA, Weiss RE, et al. Does HIV pre-exposure prophylaxis use lead to a higher incidence of sexually transmitted infections? Sex Transm Infect. 2018;94(6):457-462. doi: 10.1136/sextrans-2017-053377.

[162] Traeger MW, Cornelisse VJ, Asselin J, et al. Association of HIV Preexposure Prophylaxis With Incidence of Sexually Transmitted Infections Among Individuals at High Risk of HIV Infection. JAMA. 2019 Apr 9;321(14):1380-1390. doi: 10.1001/jama.2019.2947. PMID: 30964528; PMCID: PMC6459111.

[163] Luetkemeyer AF, Donnell D, Dombrowski JC, et al. DoxyPEP Study Team. Postexposure Doxycycline to Prevent Bacterial Sexually Transmitted Infections. N Engl J Med. 2023 Apr 6;388(14):1296-1306. doi: 10.1056/NEJMoa2211934. PMID: 37018493; PMCID: PMC10140182.

[164] Althaus CL, Heijne JCM, Roellin A, et al. Transmission dynamics of Chlamydia trachomatis affect the impact of screening programmes. Epidemics. 2010;2(3):123-131. doi: 10.1016/j.epidem.2010.04.002.

[165] Geisler WM, Lensing SY, Press CG, Hook EW 3rd. Spontaneous resolution of genital Chlamydia trachomatis infection in women and protection from reinfection. J Infect Dis. 2013 Jun 15;207(12):1850-6. doi: 10.1093/infdis/jit094. Epub 2013 Mar 7. PMID: 23470847; PMCID: PMC3654745.

[166] The EMIS Network. EMIS-2017 – The European Men-Who-Have-Sex-With-Men Internet Survey. Key findings from 50 countries. Stockholm: European Centre for Disease Prevention and Control; 2019.

[167] De Bellis, A., Willemsen, M.S., Guzzetta, G. et al. Model-based evaluation of the impact of a potential HIV cure on HIV transmission dynamics. Nat Commun 16, 3527 (2025). https://doi.org/10.1038/s41467-025-58657-x

[168] Hahn A, Frickmann H, Loderstädt U. Modelling of doxycycline-based prevention of bacterial sexually transmitted infections in comparison to condom-based and test-based prevention. Eur J Microbiol Immunol (Bp). 2024 Jan 30;14(1):50-58. doi: 10.1556/1886.2023.00059. PMID: 38289395; PMCID: PMC10895362.

[169] Korenromp EL, Sudaryo MK, de Vlas SJ, Gray RH, Sewankambo NK, Serwadda D, Wawer MJ, Habbema JD. What proportion of episodes of gonorrhoea and chlamydia becomes symptomatic? Int J STD AIDS. 2002 Feb;13(2):91-101. doi: 10.1258/0956462021924712. PMID: 11839163.

[170] Kretzschmar M, Turner KME, Barton PM, et al. Predicting the population impact of chlamydia screening programmes: comparative mathematical modelling study. Sexually Transmitted Infections 2009;85:359-366.

[171] Guzzetta, G., Marziano, V., Mammone, A. et al. The decline of the 2022 Italian mpox epidemic: Role of behavior changes and control strategies. Nat Commun 15, 2283 (2024). https://doi.org/10.1038/s41467-024-46590-4

[172] John A. Jacquez, Carl P. Simon, James Koopman, Lisa Sattenspiel, Timothy Perry, Modeling and analyzing HIV transmission: the effect of contact patterns, Mathematical Biosciences, Volume 92, Issue 2, 1988, Pages 119-199, ISSN 0025-5564, https://doi.org/10.1016/0025-5564(88)90031-4.