Contagions under control

Mathematical models to inform infectious disease prevention and control

THI MUI PHAM

Contagions under control

Mathematical models to inform infectious disease prevention and control

THI MUI PHAM

Contagions under control

Mathematical models to inform infectious disease prevention and control

Besmettingen in toom
Wiskundige modellen om infectieziektepreventie en -bestrijding te informeren

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
woensdag 23 februari 2022 des middags te 4.15 uur

door

Thi Mui Pham
geboren op 12 juli 1991
te Hanoi, Vietnam

Promotor

Prof. dr. Mirjam E. Kretzschmar

Copromotoren

Dr. Martin C.J. Bootsma

Dr. Ganna Rozhnova

"I simply wish that in a matter which so closely concerns the well-being of the human race, no decision shall be made without all knowledge which a little analysis and calculation can provide."

Daniel Bernoulli, 1760

Beoordelingsccommissie:

prof. dr. ir. J.A.P. (Hans) Heesterbeek	Universiteit Utrecht
prof. dr. Niel Hens	Universiteit Antwerpen
prof. dr. J.A.J.W. (Jan) Kluijtmans	Universiteit Medisch Centrum Utrecht
prof. dr. Rafael Mikolajczyk	Martin Luther University Halle-Wittenberg
prof. dr. R.J.L. (Rob) Willems	Universiteit Medisch Centrum Utrecht

Paranimfen:

Remy Willems

Francisco Simoes

Contents

Appendices   **404**

Introduction

Thi Mui Pham

From the plague to the ongoing coronavirus disease 2019 (COVID-19) pandemic, infectious diseases have been a steady companion of humans throughout history. Despite a global decline in deaths from communicable diseases as reported by the World Health Organization (WHO) in 2019, lower respiratory tract infections and diarrhoeal diseases still ranked among the top 10 causes of death worldwide [1]. They also remain a leading cause of morbidity and mortality in low- and middle-income countries, with 6 of the top 10 causes of death in low-income countries assigned to communicable diseases [1]. In addition, infections caused by antibiotic-resistant bacteria are widely considered a major public health concern of the 21st century [2–4]. In 2018, the WHO published a priority list for research and development of new antibiotics for antibiotic-resistant bacteria [5]. Apart from multidrug-resistant and extensively-resistant Mycobacterium, priority should be given to multidrug resistance and extensively drug-resistant Gram-negative bacteria, with carbapenem resistant *Acinetobacter baumanni*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* in the highest priority category.

The importance of an organized response to these challenges has been highlighted most recently, with the emergence of the *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2), the virus causing the Coronavirus Disease 2019 (COVID-19) pandemic. Public health policy makers have been confronted with the urgent need to find the most adequate and effective intervention strategies. However, assessing their effectiveness requires evidence-based tools to support their decision. The field of mathematical modeling has become an essential part of epidemic response efforts for many recent infectious disease outbreaks and achieved greater attention and appreciation in particular during the ongoing COVID-19 pandemic.

In this thesis, we will illustrate the value and important role of mathematical models for infectious disease control with a focus on the transmission of SARS-CoV-2 in the community and in hospital settings, as well as the nosocomial transmission of *Pseudomonas aeruginosa*. We will begin with a general introduction by providing a summary of non-pharmaceutical interventions followed by a brief overview of mathematical models relevant for the content of this thesis. Building upon this theoretical background, we will elaborate on the research questions addressed in this thesis and on how they

build on the existing literature of mathematical modeling of infectious diseases. We will then give an outline of this thesis with a brief description of each individual chapter.

Mathematical modeling of infectious diseases in a nutshell

A mathematical model can be used as a conceptual tool to understand and to quantify dynamic behaviour of infectious diseases. More specifically, infectious disease models serve three aims, *understanding* the dynamics of disease spread, *predicting* the future course of an outbreak, and ultimately *devising and evaluating* measures and interventions for disease control. Using these models, we are able to translate epidemiological assumptions of biological processes into a mathematical framework in a transparent and systematic way. As such, mathematical models allow us to test our understanding of the epidemiology of the disease by comparing model results and observations from the real world. They complement traditional experimental approaches, in particular when experimental manipulation of the studied system is not feasible (as it is the case during infectious disease outbreaks) [6].

The field of mathematical modeling of infectious disease dynamics has a long history with the first known model presented by Daniel Bernoulli at the Royal Academy of Sciences in Paris in 1760 and later published in 1766. He demonstrated through a mathematical analysis that inoculation against smallpox (variolation) was beneficial for society as a whole despite the risks of infection to individuals. By calculating the gain in life expectancy that would be achieved if smallpox were eradicated, he argued for a universal inoculation of smallpox. Bernoulli showed how a simple mathematical model could assess the impact of an intervention on people's health on a population level without the need of performing an experiment.

"I simply wish that in a matter which so closely concerns the well-being of the human race, no decision shall be made without all knowledge which a little analysis and calculation can provide."

Daniel Bernoulli, 1760

How to control infectious disease transmission

Infectious disease control can be defined as practices and programs that aim at preventing the disease by reducing the transmission of infections [7]. One of the most important epidemiological parameters to measure transmissibility is the basic reproduction number R_0, defined as the average number of secondary cases that an index case can generate during its entire infectious period in a fully susceptible population. The effective reproduction number R_E applies to a population where some people have gained immunity and/or control measures are implemented. Control measures aim at reducing R_E to below 1 which will lead, in the long term, to the elimination of the disease. A plethora of such measures exist but the most appropriate may depend on the disease, the host, its routes of transmission as well as the setting.

Non-pharmaceutical interventions

In this thesis, we will focus on non-pharmaceutical interventions (NPIs), public health measures that aim at preventing disease transmission without requiring pharmaceutical drug treatments. Below, we provide a summary of those considered in this thesis.

Hand hygiene or specifically hand washing is a personal protective measure where hands are cleaned with soap and water, or with alcohol-based hand sanitizers. In the 19th century, Ignaz Philip Semmelweis, a Viennese obstetrician, discovered that infections could be passed to patients on the hands of health-care workers. He showed that hand washing could drastically reduce mortality rates due to puerperal fever in obstetrical clinics. Hand hygiene is now an essential measure in hospitals, and it has also been advocated for to be used to reduce community spread of respiratory tract infections (such as with SARS-CoV-2) or gastrointestinal diseases.

Face masks can be worn as a personal protective measure to reduce the airborne transmission of a pathogen. During the 1910 Manchurian Pneumonic Plague, Wu Lien Teh developed a mask made of cotton and gauze to filter the air people inhaled, and promoted their use as "the principal means of personal protection". He not only received international acclaim for his contribution to controlling the epidemic but was also recognized as a public health pioneer. In East Asia, masks are commonly used to control

the transmission of respiratory tract infections. While masks were infrequently used in other parts of the world, a drastic shift occurred during the COVID-19 pandemic, where community face mask use has been advocated for in nearly all regions of the world.

Physical distancing reduces the number of contacts in a population through which a disease can spread. Governments may impose physical distancing measures through interventions like stay-at-home orders, limiting the number of individuals at venues, or closing of schools and work places. Individuals themselves may also choose to *self-impose* distancing measures. During the COVID-19 pandemic, the term *lockdown* was colloquially used to describe the restriction policies imposed by governments for people to stay at home and reduce their contacts. Numerous countries and territories around the world have implemented physical distancing to limit the spread of SARS-CoV-2.

Quarantine and isolation prevent (potentially) infectious individuals from mixing with the rest of the population and hence prevent their contributions to onward transmission. *Quarantine* refers to the restriction of movements of individuals who were exposed to a contagious disease whereas *isolation* refers to the separation of already infected individuals from those who are not infected. These two forms of disease control are among the oldest known control measures. For example, in the 14th century, coastal cities (such as Venice) quarantined ships arriving at their port in order to contain diseases such as the Bubonic plague. During the COVID-19 pandemic, quarantine and isolation have been implemented across the world as key strategies to combat the spread of SARS-CoV-2.

Regular screening is a form of testing of individuals at certain time intervals independent of symptoms. This measure itself does not directly impact transmission, but if detected cases adhere to pharmaceutical or non-pharmaceutical measures, transmission chains may be interrupted. Regular screening is critical for identifying asymptomatic cases and is, therefore, particularly effective when asymptomatic individuals make a large contribution to transmission and transmission levels of the pathogen are high. In hospitals, screening programs have been widely implemented to prevent colonizations (and subsequently infections) by multi-drug resistant bacteria. During the COVID-19 pandemic, public health authorities have highlighted the importance of regular antigen self-testing to detect asymptomatic SARS-CoV-2 infections, especially in occupational and educational settings [8, 9].

Contact tracing operates by identifying individuals who have been exposed to a disease. Index cases (individuals identified as having the disease) represent a starting point for the process. They are interviewed by public health staff in order to establish the source of infection, i.e., who infected them, with whom they had contact, and who they might have infected. Self-isolation or self-quarantine is then either imposed or advised to the index case and the contacts, respectively. If conducted sufficiently quickly, contact tracing can prevent secondary transmission from individuals who are likely to be infected but are not identified otherwise (e.g., through symptoms). This intervention relies on sufficient resources and is particularly efficient if the case numbers are low. Throughout the history of infectious diseases, contact tracing (though often not named as such) has played a crucial part in infection control. Contrary to popular opinion, the eradication of smallpox was achieved not by universal vaccination alone, but in combination with contact tracing, quarantine, and the treatment of infected individuals. During the COVID-19 pandemic, contact tracing has regained importance in outbreak control, in particular through the deployment of digital apps.

Other non-pharmaceutical measures exist to prevent infectious disease spread, including travel-related measures, such as border closures or entry and exit screening, disinfection of equipment and the environment, or control of the vectors of infection. We do not consider those in this thesis.

Role of mathematical modeling in controlling infectious diseases

To effectively combat the spread of an infectious disease, we need to understand its dynamics as well as the effect of interventions that are already implemented or their potential impact if prospectively implemented. This is important for efficiently allocating current and future resources and for designing new interventions. In this thesis, we will illustrate how mathematical models can contribute to evaluating interventions and their value in informing infection control policies. We will focus on transmission dynamics and control measures in human populations and will consider only infections where the immune system reacts relatively fast and removes the pathogen after a short period of time (days or weeks) [10].

Types of models

In essence, mathematical modeling of infectious diseases is about studying real world phenomena by describing them, in a simplified form, in the mathematical language. In order to evaluate the effect on infectious disease dynamics, tools are necessary 1) to translate biological aspects and assumptions into the mathematical language, 2) to perform analyses and estimate relevant parameters, and subsequently 3) to translate the results back for interpretation of observations from the real world. The basic foundations of the mathematical theory of epidemics were established by Kermack and McKendrick in 1927 [11]. The fundamental principle can be described by the dependence of the per capita rate of infectivity and the rate of removal (as the sum of recovery and death rates) on the time since infection. An important special case of Kermack and McKendrick's general theory is the classification of individuals into compartments by their epidemiological/disease status, e.g., their ability to transmit the pathogen. Here, we assume that the pathogen causes an infection for a typically short period of time after which the host will develop immunity (often lifelong).

While there is an infinite number of possible ways to develop models for epidemic processes, it is possible to define some broad categories of infectious disease models based on their commonalities and differences. Almost all epidemic models have in common that they aim to describe the number of infected individuals as a function of time. What follows is an overview of types of models that serve as the basis for the articles in this thesis. While the presented list of models and their elaborations are not exhaustive and kept deliberately short, they address important aspects in infectious disease modeling. A general overview of mathematical tools in infectious disease modeling can be found in, e.g., [10, 12–14].

Compartmental models

A classical example of compartmental models is the *SIR* model, where for each time point t, the population is divided into the susceptible $S(t)$, infectious $I(t)$, and recovered $R(t)$ (individuals that have cleared the disease and are not transmitting anymore) compartments. One may remove or add compartments to the model, depending on the disease and the level of detail and realism one wishes to incorporate. The essential assumption of these models is that no difference is made between individuals within each compartment.

Transitions between compartments are characterised by rates at which individuals move from one compartment to another. To account for population demography (i.e., birth, death, and migration), rates for entering and leaving the population may be added. Since the dynamics of an epidemic are often much faster than the demographic dynamics of population, the latter are often neglected in simple compartmental models. In this case, the total population size, i.e., the sum of the numbers of individuals in each compartment, is constant.

The transition rates may be either constant or may vary with time. For a simple SIR model, it is assumed that the infectivity of an infected individual and the recovery rate are constant leading to exponentially distributed sojourn times. Although for some diseases this assumption is an acceptable approximation, it is unrealistic for other diseases as demonstrated for HIV [15] or measles (e.g., [16, 17]). While alternative formulations, such as gamma distributed infectious periods, exist, exponentially distributed sojourn times are often used for mathematical convenience.

The rate of progression from S to I is typically assumed to be time-varying and referred to as *force of infection*, and defined as the per capita rate at which susceptible individuals become infected [12]. Another critical assumption of many compartmental models is that contacts are made according to the mass-action principle: All individuals mix randomly, in analogy to chemical reactions, i.e., each pair of individuals are equally likely to come in contact and each individual can contact all other individuals. In the simplest case the force of infection is then determined by three distinct factors: 1) the contact rate (the number of contacts per unit of time), 2) the fraction of infectious individuals (also known as prevalence), and 3) the transmission probability per contact.

Deterministic compartmental models

The defining property of deterministic models is that the state of the system is uniquely determined by the values of the parameters and the initial state of the system. Usually deterministic continuous time models consist of sets of ordinary differential equations (ODEs), which describe the rate of change of the number of individuals in each compartment. If the basic reproduction number R_0 is less than 1, there is no (big) epidemic while if R_0 exceeds 1, the epidemic grows exponentially, infecting a substantial proportion of the population. ODE models may be good approximations to reality for infectious disease dynamics in large populations. For phenomena on a smaller scale, such as in hospitals or

in schools, one would expect that randomness may play an important role. Deterministic models usually cannot capture the full range of stochastic variability that may occur in epidemics in these smaller populations.

Stochastic compartmental models

For stochastic compartmental models, chance processes play a crucial role, and contrary to deterministic models, model results may differ for the same initial condition and parameter values. They provide answers to questions that cannot be addressed with deterministic methods such as the probability of an outbreak, the outbreak size distribution, or the probability that a disease has been eradicated. In addition, they enable the quantification of uncertainty of model parameter estimates from (epidemic) data [14, 18, 19].

Beyond compartmental models

While equation-based models as described above are simple and intuitive, they often rely on the critical assumption of homogeneous mixing, and on fixed or exponentially distributed infectious periods. However, heterogeneity may affect the contact patterns (and therefore infectious disease spread), and the infectiousness of infected individuals does not necessarily follow an exponential distribution for many diseases (e.g., [15–17]). In principle, it is possible to extend equation-based models to account for non-exponentially distributed infectious periods, for example, by incorporating additional disease compartments (also known as the linear chain trick [20, 21]) or by using a continuous time stochastic model represented by distributed delay equations (see e.g., [11, 22]). While these models provide more flexibility, they can also be more challenging to analyze mathematically, and to simulate [23].

Agent-based models

For understanding small-scale effects of epidemics as experienced in schools or hospitals, more detailed and complex models may be much more appropriate. Agent-based models have become a popular addition to the tool set of infectious disease modeling in recent years [24, 25]. A defining feature is to put the individual central to the model and keep track of their corresponding characteristics and individual interactions. This not only allows to capture behaviour at the individual level but also to incorporate more realistic

contact networks in the model. Usually agent-based models are stochastic and can take stochastic effects, which occur in smaller populations, like schools or hospitals, or network effects in large populations, into account. They provide a rich model framework to incorporate heterogeneity that affects the contact patterns and therefore infectious disease spread without running an experiment.

Parameterization

Results of a model analysis are only informative for interpretation in the real world if the model is appropriately parameterized. Where possible, model parameters are set to plausible values or ranges based on expert opinion, best available evidence from current data or literature to generate model results [18]. For example, the infectious period can be estimated by studying reported transmission events or by measuring the amount of pathogens excreted by an infected individual over time. If model results cannot directly be verified due to a lack of data, sensitivity analyses via parameter space exploration (e.g., using Latin Hypercube Sampling [26]) can aid in identifying influential parameters and may be used to verify model predictions qualitatively.

Model calibration and statistical inference

Model calibration is the process of identifying the set of model parameters that best explain the observed data retrieved from the underlying system that was modeled. For simple models, parameters can be estimated by fitting model results to observed data based on maximizing the likelihood of the model. One of the main difficulties in estimating parameters of infectious disease models is that the infection process is only partially observed, and observed quantities might be aggregated (e.g., weekly) [14]. Since likelihood methods typically rely on integrating over unobserved quantities, analytical evaluations quickly become unfeasible. Statistical inference techniques involving data imputation, such as the expectation-maximisation (EM) algorithm [27] and Markov chain Monte Carlo (MCMC) [28, 29] have been used to overcome this difficulty. Data-augmented MCMC methods [30] estimate the joint posterior distribution of the parameters by imputing unobserved data. This Bayesian approach allows a high flexibility in model choice and fitting but could result in large computational memory requirements and slower mixing (and therefore convergence issues) if a large amount of data is missing [31]. Moreover, the MCMC algorithm cannot be easily parallelized and is therefore poorly scalable [32]. Consequently,

this approach has been mainly applied to data from small populations such as households [33–35], schools [36], or hospital wards [37–39], for which the number of cases does not exceed a few thousands.

Due to their complexity, agent-based models usually contain a large number of parameters making efficient model calibration a key challenge. Often calibration of unknown parameters follow ad hoc and unstructured approaches [25] where parameters are adjusted sequentially to match observed epidemiological statistics or data. More advanced and systematic calibration methods exist [24, 40]. For example, the output of an ABM may be calibrated to data by the Nelder-Mead (NM) optimization algorithm. Alternatively, Bayesian calibration approaches [41] start with prior distributions over the parameter space (e.g., using Latin Hypercube Sampling [26]) and a Bayesian update algorithm to obtain the posterior distribution of the parameters. Due to their required extensive computation effort, these methods have been employed only occasionally for infectious disease agent-based models [42].

Mathematical models to tackle nosocomial transmission of multi-drug-resistant bacteria

Hospitals encompass a particularly vulnerable population including patients receiving treatments that may weaken their immune system or already immunocompromized patients. In this setting, even microorganisms that are usually not disease-causing in the general population may become pathogenic and cause nosocomial infections. The distinct characteristics of these infections relevant for this thesis are the following: Firstly, many of these microorganisms may be present on or within a body without yet causing an infection (colonization). Typically, only a small proportion of patients who are colonized will develop an infection, and the dynamics of disease transmission are therefore determined by individuals colonized with the microorganism rather than those who are infected. Thus, by preventing colonization, nosocomial infections may be effectively averted as well. Secondly, transmissions are observed in typically small hospital units where stochastic effects are likely to play an important role. In these settings, stochastic models are more appropriate to study the dynamics of disease transmission. Thirdly, patients in intensive-care units typically have a short length of stay (in the order of days) leading to a highly fluctuating hospital population with a constant risk of importations from the community.

The first published mathematical models for nosocomial infections used deterministic approaches [43, 44] to study the transmission dynamics in single hospital wards. However, due to the small population size in hospitals, the importance of using stochastic models was quickly recognized [45–47]. The first stochastic models [45, 47] quantified the effectiveness of infection control interventions and were based on the Ross-Macdonald model that described the transmission of malaria by mosquito vectors [12]. These models incorporated contaminated hands of healthcare workers (HCWs) in analogy to mosquitoes as vectors of transmission in hospitals. Cooper et al [45] studied the effect of improving hand hygiene and reducing the admission of colonized patients on the prevalence of methicillin-resistant *Staphylococcus aureus* (MRSA). The results of this model suggested an important role of hand hygiene on MRSA spread in hospital wards. Austin et al [47] similarly found that hand hygiene was the most effective intervention for reducing the prevalence of vancomycin-resistant *Enterococci* (VRE) (VRE). In addition, staff cohorting was found to be another powerful control measure.

Since the first employment of mathematical models as tools for assessing the effectiveness of controlling healthcare-associated infections, they have been increasingly used in the past years (e.g., Figure 2 in Kleef et al (2013) [48] or Assab et al (2017) [49]). Interventions to control nosocomial infections may differ between pathogens. Thus, pathogen specific transmission routes and their relative contributions to the overall number of transmissions are crucial to evaluate and design effective control strategies. Numerous models have been developed to estimate epidemiological parameters such as transmission rates and the relative contributions of transmission routes for various pathogens using hospital surveillance data [50–54]. To overcome the difficulties related to missing information on, e.g., timing of events (such as of colonizations and infections) and asymptomatic carriage, several studies employed Bayesian statistical methods [37–39].

According to a systematic review by Kleef et al [48], most studies modeling nosocomial transmission of bacteria between 1993 and 2011 focused on MRSA, followed by VRE. In 2020, the most common modeled pathogens were SARS-CoV-2, MRSA, and *Clostridioides difficile* (*C. difficile*). Although *Pseudomonas aeruginosa* (*P. aeruginosa*) is among the most frequently reported pathogens for healthcare-associated infections [55], less

than five studies modeled this pathogen in a nosocomial setting between 1993 and 2011 [48]. Even these studies have not been specifically developed for *P. aeruginosa* but for nosocomial pathogens in general. While observational and experimental studies [56–60] suggest an important role of environmmental contamination in the transmission process of (VIM-producing) *P. aeruginosa*, there have been no mathematical modeling studies that quantified this relationship. Due to its intrinsic resistance to multiple antibiotics, in particular the emerging resistance to carbapenems, *P. aeruginosa* is extremely difficult to treat. More efforts to understand the modes of nosocomial transmissions of *P. aeruginosa* are therefore needed to protect immunocompromized patients at risk. In this thesis, we present two studies that contribute to filling this gap, representing the first attempt to model the dynamics of nosocomial transmission of *P. aeruginosa* and estimate the role of environmental contamination for two different settings. Using longitudinal hospital surveillance data, we employed a Bayesian data-augmented MCMC method as presented in Cooper et al (2008) [39] to account for missing data.

Mathematical models to control COVID-19

The COVID-19 pandemic represents the most recent and prominent example of a public health problem for which mathematical modeling can be a powerful tools to inform disease control strategies. In the beginning of the outbreak, many (modeling) studies focused on estimating biological and epidemiological basic characteristics of SARS-CoV-2, such as the basic reproduction number [61], serial interval and incubation time distribution [62], and forecasting the trajectory of the epidemic on short but also longer term [63]. Concurrently, a myriad of models has been developed focusing on the evaluation of measures and interventions to control the pandemic [64]. These studies either focus on the impact of NPIs in the community as a whole, or in specific settings of interest, such as hospitals, long-term care facilities, work places, or schools. The extent to which NPIs were implemented in response to the rapid spread of SARS-CoV-2 varied across countries. Most governments implemented some sort of physical distancing measures, such as mass gathering cancellations, closure of public spaces (including restaurants, entertainment venues, non-essential shops, public transport etc.), closure of educational institutions (including daycare or nursery, primary schools, and secondary schools and higher education) [65]. In addition, self-imposed measures, such as social distancing, hand washing, and the use of face masks were promoted to further curb

transmission of SARS-Cov-2. This voluntary health-related behaviour is likely dependent on the progression of the epidemic with more measures adopted if the prevalence of the disease is high. However, most COVID-19 transmission models do not include this type of risk perception and reactive health-related behaviour. It is thus not known what the potential impact of these reactive self-imposed measures on the ongoing epidemic could be and how it compares to government-imposed physical distancing measures. We provide a theoretical basis to address these questions in Chapter 1 of this thesis. Since we were interested in possible implications for the general population and thus on a large-scale, we chose a deterministic compartmental model to address these questions.

While many uncertainties about SARS-CoV-2 remain to date, knowledge gaps and the lack of reliable data have made the policy response at the start of the pandemic especially difficult. In particular, little information was available on the predominant transmission routes of SARS-CoV-2. The WHO reported contact, droplet, airborne, fomite, fecal-oral, bloodborne, mother-to-child, and animal-to-human transmission as possible modes of transmission for SARS-CoV-2 in April 2020 [66]. Motivated by the COVID-19 pandemic, we were interested in how hand hygiene behavior could be optimized to effectively reduce virus transmission. We developed a statistical model in Chapter 2 to investigate the effect of different timings of hand washing and the duration of persistence of viable virus on hands on the risk of infection of an individual. Key results of this work were presented at the Scientific Advisory Group for Emergencies (SAGE) to inform the policy response at the beginning of the COVID-19 epidemic in the United Kingdom (UK). While it is now known that contaminated surfaces and thus hand hygiene are likely to play a minor role in SARS-CoV-2 transmission (in comparison with airborne transmission) [67–69], our work has far wider relevance to the control of respiratory tract infections in general.

The implementation of measures that reduce the spread of COVID-19 is vital for maintaining healthcare capacities and preventing hospitals to be overwhelmed. Nosocomial infections have been reported in healthcare settings in many countries during the COVID-19 pandemic [70, 71]. However, quantitative estimates of SARS-CoV-2 transmission in hospitals are lacking and its contribution to the overall COVID-19 epidemic is unknown. Quantification of nosocomial SARS-CoV-2 transmission is hindered by the fact that the

exact time of infection is rarely known. The definition of a hospital-acquired infection usually relies on the time of symptom onset [72], or the first positive test [73]. If the delay is above a pre-defined cutoff (usually the average incubation time), the infection would be defined as hospital-acquired. This definition might miss a substantial proportion of symptomatic hospital-acquired infections either because patients might be discharged before developing symptoms or because they developed symptoms before the defined threshold. Prepared to advise hospital policy decisions of SAGE in the UK, we used simulation modeling in Chapter 3 of this thesis to estimate the proportion of identified COVID-19 cases in English hospitals attributed to symptomatic hospital-acquired infections as well as the contribution of the latter to the COVID-19 epidemic in England in the first half of 2020. To help inform hospital infection control policies for subsequent COVID-19 waves, several models were developed to study the impact of interventions to control nosocomial transmission of SARS-CoV-2 [74–77]. Most studies investigated the use of personal protective equipment (PPE), isolation of COVID-19 patients, and asymptomatic screening of HCWs and patients. These studies were developed in the beginning of the pandemic and often assumed a time-invariant infectiousness of infected individuals and/or perfect test sensitivity. Huang et al (2021) [78] accounted for a time-varying infectiousness and imperfect test sensitivity but modeled SARS-CoV-2 transmission only in one hospital unit in Wuhan [77]. Following up on these studies, we were interested in comparing the impact of various hospital-based interventions targeted at HCWs on the SARS-CoV-2 transmission in a hospital consisting of COVID-19 as well as non-COVID-19 wards while accounting for time-varying infectiousness, and time-varying sensitivities of diagnostic tests. In particular, we were interested in the impact of contact-tracing of HCWs, an intervention supported by observational evidence [79] but not investigated in other modeling studies. To include information of contact networks and allow for time-varying infectiousness and test sensitivities, we developed an agent-based model and compared it to several other hospital-based interventions to investigate their effectiveness in reducing nosocomial transmission and health-care worker absenteeism. For evaluating the efficiency of testing-based interventions, we also calculated how many individuals that were tested positive would be detected among the overall number of tested.

Outline

The unifying theme of this thesis is the evaluation of the impact of preventive measures on infectious disease transmission by means of mathematical modeling. This is studied on two different levels dividing the thesis in two parts.

In **Part I** (Chapter 1 and 2) of the thesis, we focus on modeling types of infection prevention and control measures employed in the community, aiming to reduce infection disease spread in the general population. In **Part II** (Chapters 3-6), we focus on mathematical models developed for hospital settings where people from the community are blend together potentially increasing the risk of infection.

Mathematical models to control the spread of infectious diseases in the community

In **Chapter 1**, a deterministic compartmental model was developed linking biological disease progression with human health-related behaviour. We compared self-imposed prevention measures, such as social distancing, hand washing, and the use of face masks with a physical distancing measure that was imposed *once* by the government for a short-term in mitigating, delaying, or preventing a COVID-19 epidemic. We were mainly interested in how health-related behavior as a response to an ongoing epidemic may affect the epidemic itself.

Chapter 2 picks up hand hygiene as one self-imposed measure from Chapter 1 and applies a model-based statistical framework to influenza or similar respiratory tract infections. We were interested in how the effectiveness of this protective measure could be optimized and focused on different hand washing timings and frequencies as well as for different durations of persistence of viable virus on hands.

Mathematical models to control the spread of infectious diseases in hospital settings

Chapter 3 takes the reader from the general population to hospital settings where, without any control measures, the risk of infection may be greatly increased and the host population is highly vulnerable. For this population it is of great importance to investigate the routes of transmission of pathogens causing health-care associated infections and

to quantify their relative contribution to the overall number of transmissions. This will help in evaluating intervention strategies and in deciding which hospital infection control policies targeting the most important transmission routes need to be prioritized. We investigated the transmission routes of *P. aeruginosa*, known for its intrinsic resistance to many antibiotics, its omnipresence in moist environments, and included in the highest category of WHO's priority list of antibiotic-resistant bacteria. We developed a mathematical transmission model to quantify the contributions of different transmission routes using a Bayesian data-augmented MCMC estimation procedure. We distinguished background transmission, cross-transmission, and environmental contamination after the discharge of patients, and estimated their relative importance for the nosocomial transmission of *P. aeruginosa* in two ICUs in a French hospital in Besançon.

Chapter 4 is an application of a similar model and method discussed in Chapter 3 to a data set from two ICU wards of the Erasmus Medical Center in Rotterdam. Here, we focused on the role of persistent contamination in the environment in the transmission process of Verona integron-encoded metallo-β-lactamase (VIM) producing *P. aeruginosa*.

Since SARS-CoV-2 is known to spread efficiently in indoor environments, hospitals may be an important setting in the COVID-19 pandemic and understanding the extent of the problem is vital for hospital infection control policies. **Chapter 5** presents a quantification of symptomatic hospital-acquired infections and the contribution of hospital settings to the overall COVID-19 epidemic in England during the first wave.

To help inform hospital infection control policies for subsequent COVID-19 waves, we developed an agent-based model in **Chapter 6** to study the impact of hospital-based interventions aimed at HCWs to control nosocomial transmission of SARS-CoV-2. In contrast to chapters 3 and 4, the infectiousness of an infected individual, contacts between individuals as well as hospital-based interventions to limit the nosocomial spread of SARS-CoV-2 are explicitly modeled.

Finally, in the last part of this thesis, we summarize the most important results, discuss their applicability for informing infection control policies and their potential implications thereon.

References

[1] World Health Organization. *Global health estimates: Leading causes of death.* 2020. `https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death`.

[2] World Health Organization. Antimicrobial resistance: global report on surveillance. In: *World Health Organization* 61.3 (2014), pp. 383–94. `http://www.ncbi.nlm.nih.gov/pubmed/22247201`.

[3] Woolhouse, M and Farrar, J. Policy: An intergovernmental panel on antimicrobial resistance. In: *Nature* 509.7502 (May 2014), pp. 555–557. DOI: `10.1038/509555a`.

[4] Laxminarayan, R, Matsoso, P, Pant, S, et al. *Access to effective antimicrobials: A worldwide challenge.* Jan. 2016. DOI: `10.1016/S0140-6736(15)00474-2`.

[5] Tacconelli, E, Buhl, M, Humphreys, H, et al. Analysis of the challenges in implementing guidelines to prevent the spread of multidrug-resistant gram-negatives in Europe. In: *BMJ Open* 9.5 (May 2019), e027683. DOI: `10.1136/bmjopen-2018-027683`.

[6] Fitzpatrick, MC, Bauch, CT, Townsend, JP, et al. Modelling microbial infection to address global health challenges. In: *Nature Microbiology 2019 4:10* 4.10 (Sept. 2019), pp. 1612–1619. DOI: `10.1038/s41564-019-0565-8`.

[7] Detels, R, Karim, Q, Baum, F, et al. Oxford Textbook of Global Public Health. Vol. 38. 3. Oxford University Press, 2016, pp. 623–623. DOI: `10.1093/med/9780198816805.001.0001`.

[8] European Centre for Disease Prevention and Control. *Considerations on the use of self-tests for COVID-19 in the EU/EEA – 17 March 2021.* Stockholm, 2021. `https://www.ecdc.europa.eu/en/publications-data/considerations-use-self-tests-covid-19-eueea`.

[9] Centers for Disease Control and Prevention. *Testing strategies for SARS-CoV-2.* 2021. `https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/sars-cov2-testing-strategies.html`.

[10] Keeling, MJ and Danon, L. Mathematical modelling of infectious diseases. Vol. 92. 1. Dec. 2009, pp. 33–42. DOI: `10.1093/bmb/ldp038`.

[11] Kermack, WO and McKendrick, AG. Contributions to the mathematical theory of epidemics-I. 1927. eng. In: *Bulletin of Mathematical Biology* 53.1-2 (1991), pp. 33–55. DOI: `10.1007/BF02464423`.

[12] Anderson, RM and May, RM. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press (OUP), 1991.

[13] Diekmann, O, Heesterbeek, H, and Britton, T. Mathematical tools for understanding infectious diseases dynamics. In: (2013).

[14] Britton, T and Pardoux, E. Stochastic Epidemic Models. Vol. 2255. Springer, 2019, pp. 5–19. DOI: `10.1007/978-3-030-30900-8{\_}1`.

[15] Shiboski, SC and Jewell, NP. Statistical analysis of the time dependence of hiv infectivity based on partner study data. In: *Journal of the American Statistical Association* 87.418 (1992), pp. 360–372. DOI: `10.1080/01621459.1992.10475215`.

[16] Bailey, NTJ. On Estimating the Latent and Infectious Periods of Measles: I. Families with Two Susceptibles Only. In: *Biometrika* 43.1/2 (June 1956), p. 15. DOI: 10.2307/2333574.

[17] Bailey, NTJ. On Estimating the Latent and Infectious Periods of Measles: II. Families with Three or More Susceptibles. In: *Biometrika* 43.3/4 (Dec. 1956), p. 322. DOI: 10.2307/2332910.

[18] Keeling, MJ and Rohani, P. Modeling infectious diseases in humans and animals. Princeton University Press, 2011, pp. 1–368. DOI: 10.1016/s1473-3099(08)70147-6.

[19] Allen, LJ. An introduction to stochastic epidemic models. In: *Lecture Notes in Mathematics* 1945 (2008), pp. 81–130. DOI: 10.1007/978-3-540-78911-6{_}3.

[20] MacDonald, N. "Stability Analysis". In: *Time Lags in Biological Models. Lecture Notes in Biomathematics, vol 27*. Springer, Berlin, Heidelberg, 1978, pp. 13–38. DOI: 10.1007/978-3-642-93107-9{_}2.

[21] Fargue, D. Reductibilite des systemes hereditaires. In: *International Journal of Non-Linear Mechanics* 9.5 (Oct. 1974), pp. 331–338. DOI: 10.1016/0020-7462(74)90018-3.

[22] Hethcote, HW and Tudor, DW. Integral equation models for endemic infectious diseases. In: *Journal of Mathematical Biology* 9.1 (1980), pp. 37–47. DOI: 10.1007/BF00276034.

[23] Burton, TA. Volterra Integral and Differential Equations. 2nd edn. Amsterdam: Elsevier B.V., 2005.

[24] Bruch, E and Atwell, J. Agent-Based Models in Empirical Social Research. In: *Sociological Methods and Research* 44.2 (Oct. 2015), pp. 186–221. DOI: 10.1177/0049124113506405.

[25] Elizabeth Hunter, BMN, Kelleher, JD, Namee, BM, et al. A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology. In: *2016:84:3* 20.3 (June 2017). DOI: 10.18564/JASSS.3414.

[26] McKay, MD, Beckman, RJ, and Conover, WJ. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. In: *Technometrics* 21.2 (1979), pp. 239–245. DOI: 10.1080/00401706.1979.10489755.

[27] Dempster, AP, Laird, NM, and Rubin, DB. Maximum Likelihood from Incomplete Data Via the EM Algorithm. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22. DOI: 10.1111/j.2517-6161.1977.tb01600.x.

[28] Gelfand, AE and Smith, AF. Sampling-based approaches to calculating marginal densities. In: *Journal of the American Statistical Association* 85.410 (1990), pp. 398–409. DOI: 10.1080/01621459.1990.10476213.

[29] Gamerman, D and Lopes, HF. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition. In: *Texts in statistical science* 1 (2006), p. 245.

[30] Neal, P and Kypraios, T. Exact Bayesian inference via data augmentation. In: *Statistics and Computing* 25.2 (Dec. 2015), pp. 333–347. DOI: 10.1007/s11222-013-9435-z.

[31] McKinley, TJ, Ross, JV, Deardon, R, et al. Simulation-based Bayesian inference for epidemic models. In: *Computational Statistics and Data Analysis* 71 (Mar. 2014), pp. 434–447. DOI: 10.1016/j.csda.2012.12.012.

[32] Britton, T and Giardina, F. Introduction to statistical inference for infectious diseases. In: *Journal de la société*

fran\c{c}aise de statistique 157 (2014), pp. 53–70.

[33] Auranen, K, Arjas, E, Leino, T, et al. Transmission of Pneumococcal Carriage in Families: A Latent Markov Process Model for Binary Longitudinal Data. In: *Journal of the American Statistical Association* 95.452 (Dec. 2000), pp. 1044–1053. DOI: 10.1080/01621459.2000.10474301.

[34] O'Neill, PD, Balding, DJ, Becker, NG, et al. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. In: *Journal of the Royal Statistical Society. Series C: Applied Statistics* 49.4 (Jan. 2000), pp. 517–542. DOI: 10.1111/1467-9876.00210.

[35] Cauchemez, S, Carrat, F, Viboud, C, et al. A Bayesian MCMC approach to study transmission of influenza: Application to household longitudinal data. In: *Statistics in Medicine* 23.22 (Nov. 2004), pp. 3469–3487. DOI: 10.1002/sim.1912.

[36] Cauchemez, S, Temime, L, Guillemot, D, et al. *Investigating heterogeneity in pneumococcal transmission: A bayesian MCMC approach applied to a follow-up of schools.* Sept. 2006. DOI: 10.1198/016214506000000230.

[37] Cooper, B and Lipsitch, M. The analysis of hospital infection data using hidden Markov models. In: *Biostatistics* 5.2 (Apr. 2004), pp. 223–237. DOI: 10.1093/biostatistics/5.2.223.

[38] Drovandi, CC and Pettitt, AN. Multivariate Markov process models for the transmission of Methicillin-resistant Staphylococcus aureus in a hospital ward. In: *Biometrics* 64.3 (Sept. 2008), pp. 851–859. DOI: 10.1111/j.1541-0420.2007.00933.x.

[39] Cooper, BS, Medley, GF, Bradley, SJ, et al. An augmented data method for the analysis of nosocomial infection data. In: *American Journal of Epidemiology* 168.5 (Sept. 2008), pp. 548–557. DOI: 10.1093/aje/kwn176.

[40] Najmi, A, Rashidi, TH, Vaughan, J, et al. Calibration of large-scale transport planning models: a structured approach. In: *Transportation* 47.4 (June 2020), pp. 1867–1905. DOI: 10.1007/s11116-019-10018-6.

[41] Kennedy, MC and O'Hagan, A. Bayesian calibration of computer models. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 63.3 (Jan. 2001), pp. 425–464. DOI: 10.1111/1467-9868.00294.

[42] Najmi, A, Nazari, S, Safarighouzhdi, F, et al. Easing or tightening control strategies: determination of COVID-19 parameters for an agent-based model. In: *Transportation* (July 2021), pp. 1–29. DOI: 10.1007/s11116-021-10210-7.

[43] Massad, E, Lundberg, S, and Yang, HM. Modeling and simulating the evolution of resistance against antibiotics. In: *International Journal of Bio-Medical Computing* 33.1 (1993), pp. 65–81. DOI: 10.1016/0020-7101(93)90060-J.

[44] Sébille, V, Chevret, S, and Valleron, AJ. Modeling the Spread of Resistant Nosocomial Pathogens in an Intensive-Care Unit. In: *Infection Control and Hospital Epidemiology* 18.2 (Feb. 1997), pp. 84–92. DOI: 10.1086/647560.

[45] Cooper, BS, Medley, GF, and Scott, GM. Preliminary analysis of the transmission dynamics of nosocomial infections: Stochastic and management effects. In: *Journal of Hospital Infection* 43.2 (Oct. 1999), pp. 131–147. DOI: 10.1053/jhin.1998.0647.

[46] Austin, DJ and Anderson, RM. Studies of antibiotic resistance within the patient, hospitals and the community using simple mathematical models. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 354.1384 (Apr. 1999), pp. 721–738. DOI: `10.1098/rstb.1999.0425`.

[47] Austin, DJ, Bonten, MJ, Weinstein, RA, et al. Vancomycin-resistant enterococci in intensive-care hospital settings: Transmission dynamics, persistence, and the impact of infection control programs. In: *Proceedings of the National Academy of Sciences of the United States of America* 96.12 (June 1999), pp. 6908–6913. DOI: `10.1073/pnas.96.12.6908`.

[48] Kleef, E van, Robotham, JV, Jit, M, et al. Modelling the transmission of healthcare associated infections: A systematic review. In: *BMC Infectious Diseases* 13.1 (June 2013), pp. 1–13. DOI: `10.1186/1471-2334-13-294`.

[49] Assab, R, Nekkab, N, Crépey, P, et al. Mathematical models of infection transmission in healthcare settings: Recent advances from the use of network structured data. In: *Current Opinion in Infectious Diseases* 30.4 (Aug. 2017), pp. 410–418. DOI: `10.1097/QCO.0000000000000390`.

[50] Starr, JM and Campbell, A. Mathematical modeling of Clostridium difficile infection. In: *Clinical Microbiology and Infection* 7.8 (Aug. 2001), pp. 432–437. DOI: `10.1046/j.1198-743X.2001.00291.x`.

[51] Pelupessy, I, Bonten, MJ, and Diekmann, O. How to assess the relative importance of different colonization routes of pathogens within hospital settings. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.8 (Apr. 2002), pp. 5601–5605. DOI: `10.1073/pnas.082412899`.

[52] Bootsma, MC, Bonten, MJ, Nijssen, S, et al. An algorithm to estimate the importance of bacterial acquisition routes in hospital settings. In: *American Journal of Epidemiology* 166.7 (Oct. 2007), pp. 841–851. DOI: `10.1093/aje/kwm149`.

[53] Lanzas, C, Dubberke, ER, Lu, Z, et al. Epidemiological Model for Clostridium difficile Transmission in Healthcare Settings. In: *Infection Control & Hospital Epidemiology* 32.6 (June 2011), pp. 553–561. DOI: `10.1086/660013`.

[54] Kin, OK, Leung, GM, Wai, YL, et al. Using models to identify routes of nosocomial infection: A large hospital outbreak of SARS in Hong Kong. In: *Proceedings of the Royal Society B: Biological Sciences* 274.1610 (Mar. 2007), pp. 611–617. DOI: `10.1098/rspb.2006.0026`.

[55] Saleem, Z, Godman, B, Hassali, MA, et al. Point prevalence surveys of healthcare-associated infections: a systematic review. In: *Pathogens and Global Health* 113.4 (May 2019), pp. 191–205. DOI: `10.1080/20477724.2019.1632070`.

[56] Catho, G, Martischang, R, Boroli, F, et al. Outbreak of Pseudomonas aeruginosa producing VIM carbapenemase in an intensive care unit and its termination by implementation of waterless patient care. In: *Critical Care 2021 25:1* 25.1 (Aug. 2021), pp. 1–10. DOI: `10.1186/S13054-021-03726-Y`.

[57] Geyter, DD, Vanstokstraeten, R, Crombé, F, et al. Sink drains as reservoirs of VIM-2 metallo-β-lactamase-producing Pseudomonas aeruginosa in a Belgian intensive care unit: relation to patients investigated by whole-genome sequencing. In: *Journal of Hospital Infection* 115 (Sept. 2021), pp. 75–82. DOI: `10.1016/J.JHIN.2021.05.010`.

[58] Knoester, M, Boer, Md, Maarleveld, J, et al. An integrated approach to control a prolonged outbreak of multidrug-

resistant Pseudomonas aeruginosa in an intensive care unit. In: *Clinical Microbiology and Infection* 20.4 (Apr. 2014), O207–O215. DOI: `10.1111/1469-0691.12372`.

[59] Kizny Gordon, AE, Mathers, AJ, Cheong, EYL, et al. The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections—A Systematic Review of the Literature. In: *Clinical Infectious Diseases* 64.10 (May 2017), pp. 1435–1444. DOI: `10.1093/CID/CIX132`.

[60] Voor in 't holt, AF, Severin, JA, Lesaffre, EM, et al. A Systematic Review and Meta-Analyses Show that Carbapenem Use and Medical Devices Are the Leading Risk Factors for Carbapenem-Resistant Pseudomonas aeruginosa. In: *Antimicrobial Agents and Chemotherapy* 58.5 (2014), pp. 2626–2637. DOI: `10.1128/AAC.01758-13`.

[61] Ahammed, T, Anjum, A, Rahman, MM, et al. Estimation of novel coronavirus (COVID -19) reproduction number and case fatality rate: A systematic review and meta-analysis. In: *Health Science Reports* 4.2 (June 2021), e274. DOI: `10.1002/hsr2.274`.

[62] Alene, M, Yismaw, L, Assemie, MA, et al. Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. In: *BMC Infectious Diseases* 21.1 (Mar. 2021), pp. 1–9. DOI: `10.1186/s12879-021-05950-x`.

[63] Kissler, SM, Tedijanto, C, Goldstein, E, et al. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. In: *Science* 368.6493 (May 2020), pp. 860–868. DOI: `10.1126/science.abb5793`.

[64] Perra, N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. In: *Physics Reports* 913 (May 2021), pp. 1–52. DOI: `10.1016/J.PHYSREP.2021.02.001`.

[65] European Centre for Disease Prevention and Control. *Data on country response measures to COVID-19*. en. June 2021. `https://www.ecdc.europa.eu/en/publications-data/download-data-response-measures-covid-19`.

[66] World Health Organization. *Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations*. 2020. `https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations`.

[67] Greenhalgh, T, Jimenez, JL, Prather, KA, et al. *Ten scientific reasons in support of airborne transmission of SARS-CoV-2*. May 2021. DOI: `10.1016/S0140-6736(21)00869-2`.

[68] Heneghan, CJ, Spencer, EA, Brassey, J, et al. SARS-CoV-2 and the role of airborne transmission: a systematic review. In: *F1000Research* 10 (Mar. 2021), p. 232. DOI: `10.12688/f1000research.52091.1`.

[69] Goldman, E. Exaggerated risk of transmission of COVID-19 by fomites. In: *The Lancet Infectious Diseases* 20.8 (Aug. 2020), pp. 892–893. DOI: `10.1016/S1473-3099(20)30561-2`.

[70] Zhou, Q, Gao, Y, Wang, X, et al. Nosocomial infections among patients with COVID-19, SARS and MERS: a rapid review and meta-analysis. In: *Annals of Translational Medicine* 8.10 (May 2020), pp. 629–629. DOI: `10.21037/atm-20-3324`.

[71] Richterman, A, Meyerowitz, EA, and Cevik, M. *Hospital-Acquired SARS-CoV-2 Infection: Lessons for Public Health*. Dec. 2020. DOI: `10.1001/jama.2020.21399`.

[72] European Centre for Disease Prevention and Control. *Surveillance definitions for COVID-19*. 2021. `https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions`.

[73] Bhattacharya, A, Collin, SM, Stimson, J, et al. Healthcare-associated COVID-19 in England: a national data linkage study. In: *Journal of Infection* 0.0 (Aug. 2021). DOI: `10.1016/j.jinf.2021.08.039`.

[74] Evans, S, Agnew, E, Vynnycky, E, et al. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1829 (July 2021), p. 20200268. DOI: `10.1098/rstb.2020.0268`.

[75] Qiu, X, Miller, JC, MacFadden, DR, et al. Evaluating the contributions of strategies to prevent SARS-CoV-2 transmission in the healthcare setting: a modelling study. In: *BMJ Open* 11.3 (Mar. 2021), e044644. DOI: `10.1136/bmjopen-2020-044644`.

[76] Chin, ET, Huynh, BQ, Chapman, LAC, et al. Frequency of Routine Testing for Coronavirus Disease 2019 (COVID-19) in High-risk Healthcare Environments to Reduce Outbreaks. In: *Clinical Infectious Diseases* (Oct. 2020). DOI: `10.1093/cid/ciaa1383`.

[77] Huang, Q, Mondal, A, Jiang, X, et al. SARS-CoV-2 transmission and control in a hospital setting: An individual-based modelling study. In: *medRxiv* (Aug. 2020), p. 2020.08.22.20179929. DOI: `10.1101/2020.08.22.20179929`.

[78] Huang, Q, Mondal, A, Jiang, X, et al. SARS-CoV-2 transmission and control in a hospital setting: An individual-based modelling study. In: *Royal Society Open Science* 8.3 (Mar. 2021). DOI: `10.1098/rsos.201895`.

[79] Mandić-Rajčević, S, Masci, F, Crespi, E, et al. Source and symptoms of COVID-19 among hospital workers in Milan. In: *Occupational Medicine* 70.9 (Dec. 2020), pp. 672–679. DOI: `10.1093/occmed/kqaa201`.

Part I

Mathematical models to control the spread of infectious diseases in the community

Chapter 1

Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study

Alexandra Teslya[*], Thi Mui Pham[*], Noortje G. Godijk[*], Mirjam E. Kretzschmar,

Martin C. J. Bootsma, Ganna Rozhnova

These authors contributed equally to this work.

Abstract

Background: Many countries have implemented social distancing as a measure to *flatten the curve* of the ongoing COVID-19 epidemics. Evaluation of the impact of government-imposed social distancing and of other measures to control further spread of COVID-19 is urgent, especially because of the large societal and economic impact of the former. We aimed to compare the effectiveness of self-imposed prevention measures and of short-term government-imposed social distancing in mitigating, delaying, or preventing a COVID-19 epidemic.

Methods: We developed a deterministic compartmental transmission model of SARS-CoV-2 in a population stratified by disease status (susceptible, exposed, infectious with mild or severe disease, diagnosed and recovered) and disease awareness status (aware and unaware) due to the spread of COVID-19. Self-imposed measures were assumed to be taken by disease-aware individuals and included handwashing, mask-wearing, and social distancing. Government-imposed social distancing reduced the contact rate of individuals irrespective of their disease or awareness status.

Results: For fast awareness spread in the population, self-imposed measures can significantly reduce the attack rate, diminish and postpone the peak number of diagnoses. We estimate that a large epidemic can be prevented if the efficacy of these measures exceeds 50%. For slow awareness spread, self-imposed measures reduce the peak number of diagnoses and attack rate but do not affect the timing of the peak. Early implementation of short-term government-imposed social distancing alone is estimated to delay (by at most 7 months for a 3-month intervention) but not to reduce the peak. The delay can be even longer and the height of the peak can be additionally reduced if this intervention is combined with self-imposed measures that are continued after government-imposed social distancing has been lifted.

Conclusions: Our results suggest that information dissemination about COVID-19, which causes voluntary adoption of handwashing, mask-wearing and social distancing can be an effective strategy to mitigate and delay the epidemic. Early-initiated short-term government-imposed social distancing can buy time for healthcare systems to prepare for an increasing COVID-19 burden. We stress the importance of disease awareness in controlling the ongoing epidemic and recommend that, in addition to policies on social distancing, governments and public health institutions mobilize people to adopt self-imposed measures with proven efficacy in order to successfully tackle COVID-19.

Introduction

As of May 5, 2020, the novel coronavirus (SARS-CoV-2) has spread worldwide and only 13 countries have not reported any cases. It has caused over 3,640,835 confirmed cases of COVID-19 and nearly 255,100 deaths since the detection of its outbreak in China on December 31, 2019 [1]. On March 11, the World Health Organization officially declared the COVID-19 outbreak a pandemic [1]. Several approaches aimed at the containment of SARS-CoV-2 in China were unsuccessful. Airport screening of travelers was hampered by a potentially large number of asymptomatic cases and the possibility of pre-symptomatic transmission [2–4]. Quarantine of fourteen days combined with fever surveillance was insufficient in containing the virus due to the high variability of the incubation period [5].

Now that SARS-CoV-2 has extended its range of transmission in all parts of the world, it is evident that many countries face a large COVID-19 epidemic [6]. Initial policies regarding COVID-19 prevention were mainly limited to reporting cases, strict isolation of severe symptomatic cases, home isolation of mild cases, and contact tracing [7]. However, due to the potentially high contribution of asymptomatic and pre-symptomatic spread [8], these case-based interventions are likely insufficient in containing a COVID-19 epidemic unless they are highly effective [8–11]. Given the rapid rise in cases and the risk of exceeding critical care bed capacities, many countries have implemented social distancing as a short-term measure aiming at reducing the contact rate in the population and, subsequently, transmission [6, 12]. Several governments have imposed nationwide partial or complete lockdowns by closing schools, public places and non-essential businesses, canceling mass events, and issuing stay-at-home orders [6]. Previous studies on the 1918 influenza pandemic showed that such mandated interventions were effective in reducing transmission but their timing and magnitude had a profound influence on the course of the epidemic [13–18]. These short-term interventions were associated with a high risk of epidemic resurgence and their impact was limited if introduced too late or lifted too early [13–16].

Self-imposed prevention measures such as handwashing, mask-wearing, and social

distancing could also contribute to slowing down the epidemic [19, 20]. Alcohol-based sanitizers are effective in removing the SARS coronavirus from hands [21] and handwashing with soap may have a positive effect on reducing the transmission of respiratory infections [22]. Surgical masks, often worn for their perceived protection, are not designed nor certified to protect against respiratory hazards, but they can stop droplets being spread from infectious individuals [23–25]. Information dissemination and official recommendations about COVID-19 can create awareness and motivate individuals to adopt such measures. Previous studies emphasized the importance of disease awareness for changing the course of an epidemic [26–28]. Depending on the rate and mechanism of awareness spread, the awareness process can reduce the attack rate of an epidemic or prevent it completely [26], but it can also lead to undesirable outcomes such as the appearance of multiple epidemic peaks [27, 28]. The secondary epidemic waves may appear as the result of individuals relaxing adherence to self-imposed measures prematurely in a population where the susceptible pool following the first wave is still significantly large and disease has not been completely eliminated. It is essential to assess under which conditions, spread of disease aware-ness that instigates self-imposed measures can be a viable strategy for COVID-19 control.

The comparison of the effectiveness of early implemented short-term government-imposed social distancing and self-imposed prevention measures on reducing the transmission of SARS-CoV-2 are currently missing but are of crucial importance in the attempt to stop its spread. If a COVID-19 epidemic cannot be prevented, it is important to know how to effectively diminish and postpone the epidemic peak to give healthcare professionals more time to prepare and react effectively to an increasing health care burden. Moreover, given that several countries have peaked in cases, the importance of evaluating the effect of self-imposed measures after lifting lockdown measures is profound.

Using a transmission model we evaluated the impact of self-imposed measures (hand-washing, mask-wearing, and social distancing) due to awareness of COVID-19 and of a short-term government-imposed social distancing intervention on the peak number of diagnoses, attack rate, and time until the peak number of diagnoses since the first case. We provide a comparative analysis of these interventions as well as of their combinations

and assess the range of intervention efficacies for which a COVID-19 epidemic can be mitigated, delayed or even prevented completely. Qualitatively, these results will aid public health professionals to compare and select a combination of interventions for designing effective outbreak control policies.

Methods

Baseline transmission model

We developed a deterministic compartmental model describing SARS-CoV-2 transmission in a population stratified by disease status (Fig 1). In this baseline model, individuals are classified as susceptible (S), latently infected (E), infectious with mild disease (I_M), infectious with severe disease (I_S), diagnosed and isolated (I_D), and recovered after mild or severe disease (R_M and R_S, respectively). Susceptible individuals (S) can become latently infected (E) through contact with infectious individuals (I_M and I_S) with the force of infection dependent on the fractions of the population in I_M and I_S compartments. A proportion of the latently infected individuals (E) will go to the I_M compartment, and the remaining E individuals will go to the I_S compartment. We assume that infectious individuals with mild disease (I_M) do not require medical attention and recover (R_M) without being conscious of having contracted COVID-19. Infectious individuals with severe disease (I_S) are unable to recover without medical help, and subsequently get diagnosed and isolated (I_D) (in e.g. hospitals, long-term care facilities, nursing homes) and know or suspect they have COVID-19 when they are detected. Therefore, the diagnosed compartment I_D contains infectious individuals with severe disease who are both officially diagnosed and get treatment in healthcare institutions and those who are not officially diagnosed but have disease severe enough to suspect they have COVID-19 and require isolation. For simplicity, isolation of these individuals is assumed to be perfect until recovery (R_S), and, hence, they neither contribute to transmission nor to the contact process. Given the timescale of the epidemic and the lack of reliable reports on reinfections, we assume that recovered individuals (R_M and R_S) cannot be reinfected. The infectivity of infectious individuals with mild disease is lower than the infectivity of infectious individuals with severe disease [29]. Natural birth and death processes are neglected as the time scale of the epidemic is short compared to the mean life span of individuals. However, isolated infectious individuals with severe disease (I_D) may be removed from the population due
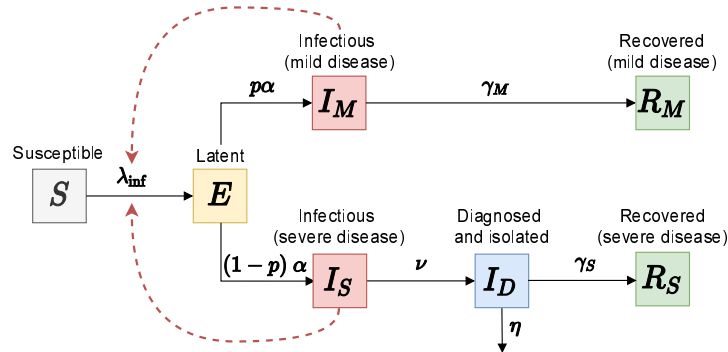
to disease-associated mortality.

Figure 1: **Schematic of the baseline transmission model.** Black arrows show epidemiological transitions. Red dashed arrows indicate the compartments contributing to the force of infection. Susceptible persons (S) become latently infected (E) with the force of infection λ_{inf} via contact with infectious individuals in two infectious classes (I_M and I_S). Individuals leave the E compartment at rate α. A proportion p of the latently infected individuals (E) will go to the I_M compartment, and the proportion $(1-p)$ of E individuals will go to the I_S compartment. Infectious individuals with mild disease (I_M) recover without being conscious of having contracted COVID-19 (R_M) at rate γ_M. Infectious individuals with severe disease (I_S) are diagnosed and kept in isolation (I_D) at rate ν until they recover (R_S) at rate γ_S or die at rate η. Table 1 provides the description and values of all parameters.

Transmission model with disease awareness

In the extended model with disease awareness, the population is stratified not only by the disease status but also by the awareness status into disease-aware (S^a, E^a, I_M^a, I_S^a, I_D^a, and R_M^a) and disease-unaware (S, E, I_M, I_S, I_D, and R_M) (Fig 2 A). Disease awareness is a state that can be acquired as well as lost. Disease-aware individuals are distinguished from unaware individuals in two essential ways. First, infectious individuals with severe disease who are disease-aware (I_S^a) get diagnosed and isolated faster (I_D^a), stay in isolation for a shorter period of time and have lower disease-associated mortality than the same category of unaware individuals. The assumption we make here is that disease-aware individuals (I_S^a) recognize they may have COVID-19 on average faster than disease-unaware individuals (I_S) and get medical help earlier which leads to a better prognosis of I_D^a individuals as compared to I_D individuals. Second, disease-aware individuals are assumed to use self-imposed measures such as handwashing, mask-wearing and self-imposed social distancing that can lower their susceptibility,

infectivity and/or contact rate. Individuals who know or suspect their disease status (I_D, I_D^a and R_S) do not adapt any such measures since they assume that they cannot contract the disease again. Hence, they are excluded from the awareness transition process and their behaviour in the contact process is identical to disease-unaware individuals.

Similarly to Perra et al [27], disease-unaware individuals acquire disease awareness at a rate proportional to the rate of awareness spread and to the current number of diagnosed individuals (I_D and I_D^a) in the population (Fig 2 B). We assume that awareness fades and individuals return to the unaware state at a constant rate. The latter means that they no longer use self-imposed measures. For simplicity, we assume that awareness acquisition and fading rates are the same for individuals of type S, E, I_M, and R_M. However, the rate of awareness acquisition is faster and the fading rate is slower for infectious individuals with severe disease (I_S) than for the remaining disease-aware population.

Table 1: Parameter values for the transmission model with and without awareness

		Value*	Source
Epidemiological parameters			
Basic reproduction number	R_0	2.5 (2–3)	Li et al [5], Park et al [30], sensitivity analyses
Probability of transmission per contact with I_S	ϵ	0.048	From $R_0 = \beta\,[p\sigma/\gamma_M + (1-p)/\nu]$
Transmission rate of infection via contact with I_S	β	0.66 per day	$\beta = c\epsilon$
Average contact rate (unique persons)	c	13.85 persons per day	Mossong et al [31]
Relative infectivity of infectious with mild disease (I_M)	σ	50% (25–75%)	Assumed, see e.g. Liu et al [29], sensitivity analyses
Proportion of infectious with mild disease (I_M)	p	82% (82–90%)	Wu et al [32], Anderson at al [20], sensitivity analyses
Delay between infection and onset of infectiousness (latent period)	$1/\alpha$	4 days	Shorter than incubation period [5, 30, 33]
Delay from onset of infectiousness to diagnosis for I_S	$1/\nu$	5 (3–7) days	Li et al [5], sensitivity analyses
Recovery period of infectious with mild disease (I_M)	$1/\gamma_M$	7 (5–9) days	Li Xingwang[†], sensitivity analyses
Delay from diagnosis to recovery for unaware diagnosed (I_D)	$1/\gamma_S$	14 days	WHO [34]
Relative infectivity of isolated (I_D)		0%	Assuming perfect isolation
Case fatality rate of unaware diagnosed (I_D)	f	1.6%	Althaus et al[35] Park et al[30]
Disease-associated death rate of unaware diagnosed (I_D)	η	0.0011 per day	$\eta = \gamma_S f/(1-f)$
Awareness parameters			
Rate of awareness spread (slow, fast and range)	δ	5×10^{-5}, 1 (10^{-6}–1) per year	Assumed, sensitivity analyses
Relative susceptibility to awareness acquisition for S, E, I_M, and R_M	k	50% (0–100%)	Assumed, sensitivity analyses
Duration of awareness for S^a, E^a, I_M^a, and R_M^a	$1/\mu$	30 (7–365) days	Assumed, sensitivity analyses
Duration of awareness for I_S^a	$1/\mu_S$	60 (7–365) days	Longer than $1/\mu$, sensitivity analyses
Delay from onset of infectiousness to diagnosis for I_S^a	$1/\nu^a$	3 (1–5) days	Shorter than $1/\nu$, sensitivity analyses
Delay from diagnosis to recovery of aware diagnosed (I_D^a)	$1/\gamma_S^a$	12 days	Shorter than $1/\gamma_S$
Case fatality rate of aware diagnosed (I_D^a)	f^a	1%	Smaller than f
Disease-associated death rate of aware diagnosed (I_D^a)	η^a	0.0008 per day	$\eta = \gamma_S^a f^a/(1-f^a)$
Prevention measure parameters			
Efficacy of mask-wearing (reduction in infectivity)		0–100%	Varied
Efficacy of handwashing (reduction in susceptibility)		0–100%	Varied
Efficacy of self-imposed contact rate reduction		0–100%	Varied
Efficacy of government-imposed contact rate reduction		0–100%	Varied
Duration of government-imposed social distancing		3 (1–13) months	Assumed, sensitivity analyses
Threshold for initiation of government-imposed social distancing		10 (1–1000) diagnoses	Assumed, sensitivity analyses

*Mean or median values were used from literature; range was used in the sensitivity analyses.
[†]Expert at China's National Health Commission

Prevention measures

We considered short-term government intervention aimed at fostering social distancing in the population and a suite of measures that may be self-imposed by disease-aware
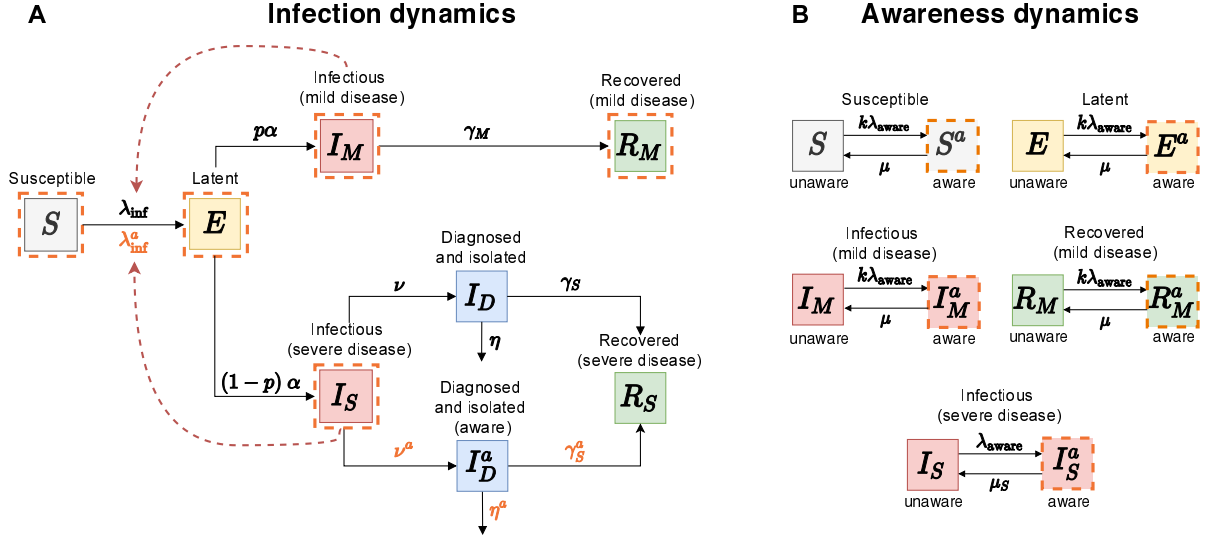
Figure 2: **Schematic of the transmission model with disease awareness.** (A) shows epidemiological transitions in the transmission model with awareness (black arrows). The orange dashed lines indicate the compartments that participate in the awareness dynamics. The red dashed arrows indicate the compartments contributing to the force of infection. Disease-aware susceptible individuals (S^a) become latently infected (E^a) through contact with infectious individuals (I_M, I_S, I_M^a, and I_S^a) with the force of infection λ_{inf}^a. Infectious individuals with severe disease who are disease-aware (I_S^a) get diagnosed and isolated (I_D^a) at rate ν^a, recover at rate γ_S^a and die from disease at rate η^a. (B) shows awareness dynamics. Infectious individuals with severe disease (I_S) acquire disease awareness (I_S^a) at rate λ_{aware} proportional to the rate of awareness spread and to the current number of diagnosed individuals (I_D and I_D^a) in the population. As awareness fades, these individuals return to the unaware state at rate μ_S. The acquisition rate of awareness ($k\lambda_{\text{aware}}$) and the rate of awareness fading (μ) are the same for individuals of type S, E, I_M, and R_M, where k is the reduction in susceptibility to the awareness acquisition compared to I_S individuals. Table 1 provides the description and values of all parameters.

individuals, i.e., mask-wearing, handwashing, and self-imposed social distancing.

Mask-wearing

Mask-wearing, while often adapted as a protective measure, may be ineffective in reducing the individual's susceptibility because laypersons, i.e., not medical professionals, are unfamiliar with correct procedures for its use (e.g. often engage in face-touching and mask adjustment) [36]. However, mask-wearing reduces infectious output [25] and, therefore, we assume that this measure lowers only the infectivity of disease-aware infectious individuals (I_M^a and I_S^a) with an efficacy ranging from 0% (zero efficacy) to 100% (full efficacy).

Handwashing

Since infectious individuals may transmit the virus to others without direct physical contact, we assume that handwashing only reduces one's susceptibility. The efficacy of handwashing is described by the reduction in susceptibility (i.e., probability of transmission per single contact) of susceptible disease-aware individuals (S^a) which ranges from 0% (zero efficacy) to 100% (full efficacy). Since transmission can possibly occur through routes other than physical contact, hand washing may not provide 100% protection to those who practice it.

Self-imposed social distancing

Disease-aware individuals, who consider themselves susceptible, may also practice social distancing, i.e., maintaining distance to others and avoiding congregate settings. As a consequence, this measure leads to a change in mixing patterns in the population. The efficacy of social distancing of disease-aware individuals is described by the reduction in their contact rate which is varied from 0% (no social distancing or zero efficacy) to 100% (complete self-isolation or full efficacy). Since contacts might not be eliminated entirely (e.g. household contacts remain), realistic values of the efficacy of self-imposed social distancing can be close to but may never reach 100%.

Short-term government-imposed social distancing

Governments may decide to promote social distancing policies through interventions such as school and workplace closures or by issuing stay-at-home orders and bans on large gatherings. These lockdown policies will cause a community-wide contact rate reduction, regardless of the awareness status. Here, we assume that the government-imposed social distancing is initiated if the number of diagnosed individuals exceeds a certain threshold (10–1000 persons) and terminates after a fixed period of time (1–3 months). As such, the intervention is implemented early into the epidemic. Government-imposed social distancing may be partial or complete depending on its efficacy, i.e., the reduction of the average contact rate in the population which ranges from 0% (no distancing) to 100% (complete lockdown). Since during a lockdown, some contacts in the population cannot be eliminated (e.g. household contacts), realistic values of the efficacy of government-imposed social distancing can be close to but never reach 100%. For example, a 73% reduction in the average daily number of contacts was observed during the lockdown in the UK [37] but the reduction could be different in countries with more or less stringent lockdown.

Model output

The model outputs are the peak number of diagnoses, attack rate (a proportion of the population that recovered or died after severe infection), the time to the peak number of diagnoses since the first case, and the probability of infection during the course of an epidemic (see S2 Text for a more detailed description of the latter). We compared the impact of different prevention measures and their combinations on these outputs by varying the reduction in infectivity of disease-aware infectious individuals (mask-wearing), the reduction in susceptibility of disease-aware susceptible individuals (handwashing), the reduction in contact rate of disease-aware individuals only (self-imposed social distancing) and of all individuals (government-imposed social distancing). We refer to these quantities as the efficacy of a prevention measure and vary it from 0% (zero efficacy) to 100% (full efficacy) (Table 1). The main analyses were performed for two values of the rate of awareness spread that corresponded to scenarios of slow and fast spread of awareness in the population (Table 1). For these scenarios, the proportion of the aware population at the peak of the epidemic was 40% and 90%, respectively. In the

main analyses, government-imposed social distancing was initiated when 10 individuals got diagnosed and was lifted after 3 months.

Estimates of epidemiological parameters were obtained from the most recent literature (Table 1). We used contact rates for the Netherlands, but the model is appropriate for other Western countries with similar contact patterns. A detailed mathematical description of the model can be found in the S1 Text. The model was implemented in Mathematica 10.0.2.0. The code reproducing the results of this study is available at `https://github.com/lynxgav/COVID19-mitigation`.

Sensitivity analyses

To allow for the uncertainty in the parameters of the baseline transmission model, we conducted sensitivity analyses with respect to the proportion of infectious individuals with mild disease, the relative infectivity of infectious individuals with mild disease, the recovery period of infectious individuals with mild disease, the delay from onset of infectiousness to diagnosis for infectious individuals with severe disease, and the basic reproduction number (see S3 Text). We also conducted sensitivity analyses for the model with disease awareness with respect to changes in the delay from the onset of infectiousness to diagnosis and isolation for disease-aware individuals, the rate of awareness spread, the relative susceptibility to awareness, and the duration of awareness (see S3 Fig). Parameter ranges used in these sensitivity analyses are specified in Table 1.

In addition, we present results for the impact on the model outcomes of all combinations of self-imposed prevention measures as their efficacy was varied from 0% to 100% and of the government-imposed social distancing with efficacy ranging from 0% to 100%, different thresholds for initiating the intervention (1 to 1000 diagnoses), and different durations of the intervention (3, 8 and 13 months) (see S1 Fig and S2 Fig for details).

Results

Our analyses show that disease awareness spread has a significant effect on the model predictions. We first considered the epidemic dynamics in a disease-aware population where handwashing is promoted, as an example of self-imposed measures (Fig 3). Then,
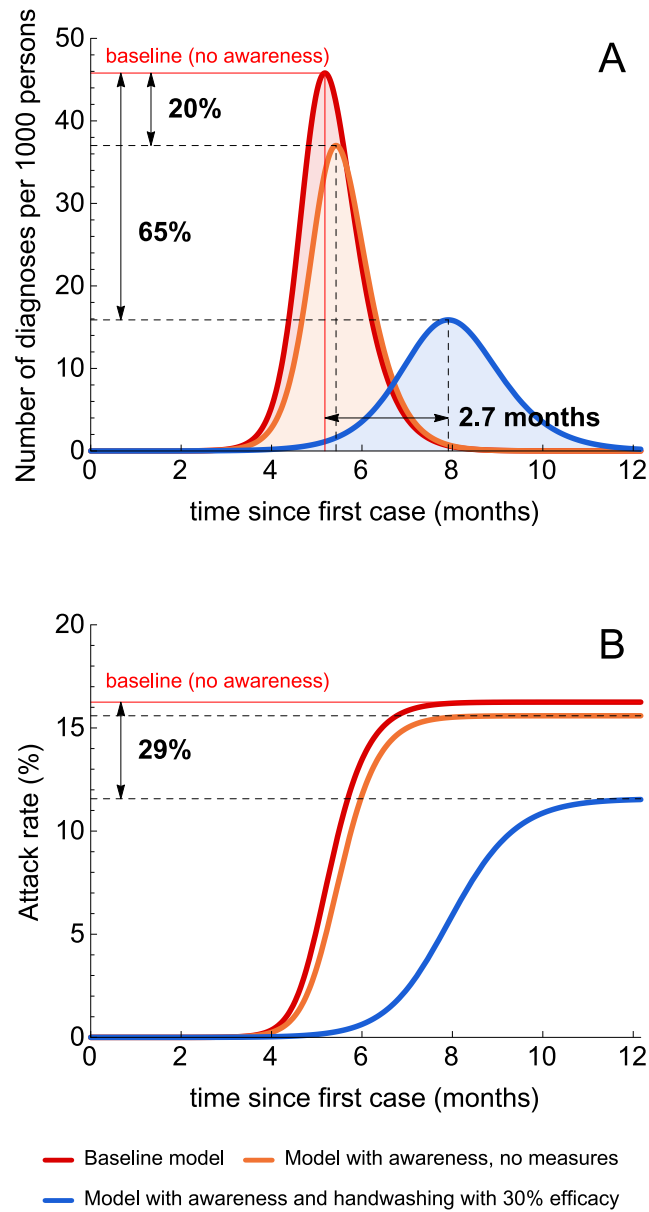
Figure 3: **Illustrative simulations of the transmission model.** (A) and (B) show the number of diagnoses and the attack rate during the first 12 months after the first case under three model scenarios. The red lines correspond to the baseline transmission model. The orange lines correspond to the model with a fast rate of awareness spread and no interventions. The blue lines correspond to the latter model where disease awareness induces the uptake of handwashing with an efficacy of 30%.

we performed a systematic comparison of the impact of different prevention measures on the model output for slow (Fig 4) and fast (Fig 5) rate of awareness spread.

Epidemic dynamics

All self-imposed measures and government-imposed social distancing have an effect on the COVID-19 epidemic dynamics. The qualitative and quantitative impact, however, depends strongly on the prevention measure and the rate of awareness spread. The baseline model predicts 46 diagnoses per 1000 individuals at the peak of the epidemic, an attack rate of about 16% and the time to the peak of about 5.2 months (red line, Fig 3 A and B). In the absence of prevention measures, a fast spread of disease awareness reduces the peak number of diagnoses by 20% but has only a minor effect on the attack rate and peak timing (orange line, Fig 3 A and B). This is expected, as disease-aware individuals with severe disease seek medical care sooner and therefore get diagnosed faster causing fewer new infections as compared to the baseline model. Awareness dynamics coupled with the use of self-imposed prevention measures has an even larger impact on the epidemic. The blue line in Fig 3 A shows the epidemic curve for the scenario when disease-aware individuals use handwashing as self-imposed prevention measure. Even if the efficacy of handwashing is modest (i.e., 30% as in Fig 3 A) the impact on the epidemic can be significant, namely we predicted a 65% reduction in the peak number of diagnoses, a 29% decrease in the attack rate, and a delay in peak timing of 2.7 months (Fig 3 A and B).

The effect of awareness on the disease dynamics can also be observed in the probability of infection during the course of the epidemic. In the model with awareness and no measures, the probability of infection is reduced by 4% for all individuals. Handwashing with an efficacy of 30% reduces the respective probability by 14% for unaware individuals and by 29% for aware individuals. Note that the probability of infection is highly dependent on the type of prevention measure. The detailed analysis is given in the S2 Text.
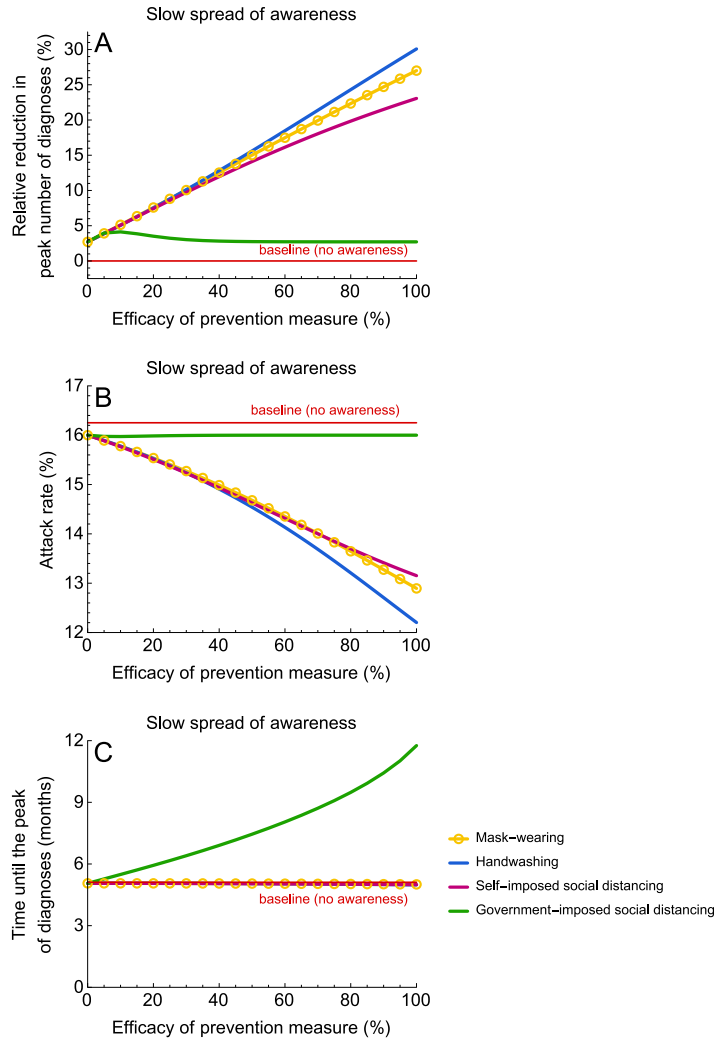
Figure 4: **Impact of prevention measures on the epidemic for a slow rate of awareness spread.**
(A), (B) and (C) show the relative reduction in the peak number of diagnoses, the attack rate (proportion of the population that recovered or died after severe infection) and the time until the peak number of diagnoses. The efficacy of prevention measures was varied between 0% and 100%. In the context of this study, the efficacy of social distancing denotes the reduction in the contact rate. The efficacy of handwashing and mask-wearing are given by the reduction in susceptibility and infectivity, respectively. The simulations were started with one case. Government-imposed social distancing was initiated after 10 diagnoses and lifted after 3 months. For parameter values, see Table 1. Please note that the blue line corresponding to handwashing is not visible in (C) since it almost completely overlaps with lines for mask-wearing and self-imposed social distancing.

A comparison of prevention measures

Slow spread of awareness

Fig 4 shows the impact of all considered self-imposed measures as well as of the government-imposed social distancing on the peak number of diagnoses, attack rate, and the time to the peak for a slow rate of awareness spread. In this scenario, the model predicts progressively larger reductions in the peak number of diagnoses and in the attack rate as the efficacy of the self-imposed measures increases. In the limit of 100% efficacy, the reduction in the peak number of diagnoses is 23% to 30% (Fig 4 A) and the attack rate decreases from 16% to 12-13% (Fig 4 B). The efficacy of the self-imposed measures has very little impact on the peak timing when compared to the baseline, i.e., no awareness in the population (Fig 4 C). Since the proportion of aware individuals who change their behavior is too small to make a significant impact on transmission, self-imposed measures can only mitigate but not prevent an epidemic.

When awareness spreads at a slow rate, a 3-month government intervention has a contrasting impact to the self-imposed measure scenario. The time to the peak number of diagnoses is longer for more stringent contact rate reductions. For example, a complete lockdown (government-imposed social distancing with 100% efficacy) can postpone the peak by almost 7 months but its magnitude and attack rate are unaffected (with respect to the baseline model without measures and awareness). Similar predictions are expected, as long as government-imposed social distancing starts early (e.g, after tens to hundreds of cases) and is lifted a few weeks to few months later. This type of intervention halts the epidemic for the duration of intervention, but, because of a large pool of susceptible individuals, epidemic resurgence is expected as soon as social distancing measures are lifted.

Fast spread of awareness

Since the government intervention reduces the contact rate of all individuals irrespective of their awareness status, it has a comparable impact on transmission for scenarios with fast and slow rate of awareness spread (compare Fig 4 and Fig 5). However, the impact of self-imposed measures is drastically different when awareness spreads fast. All self-imposed measures are more effective than the short-term government intervention.
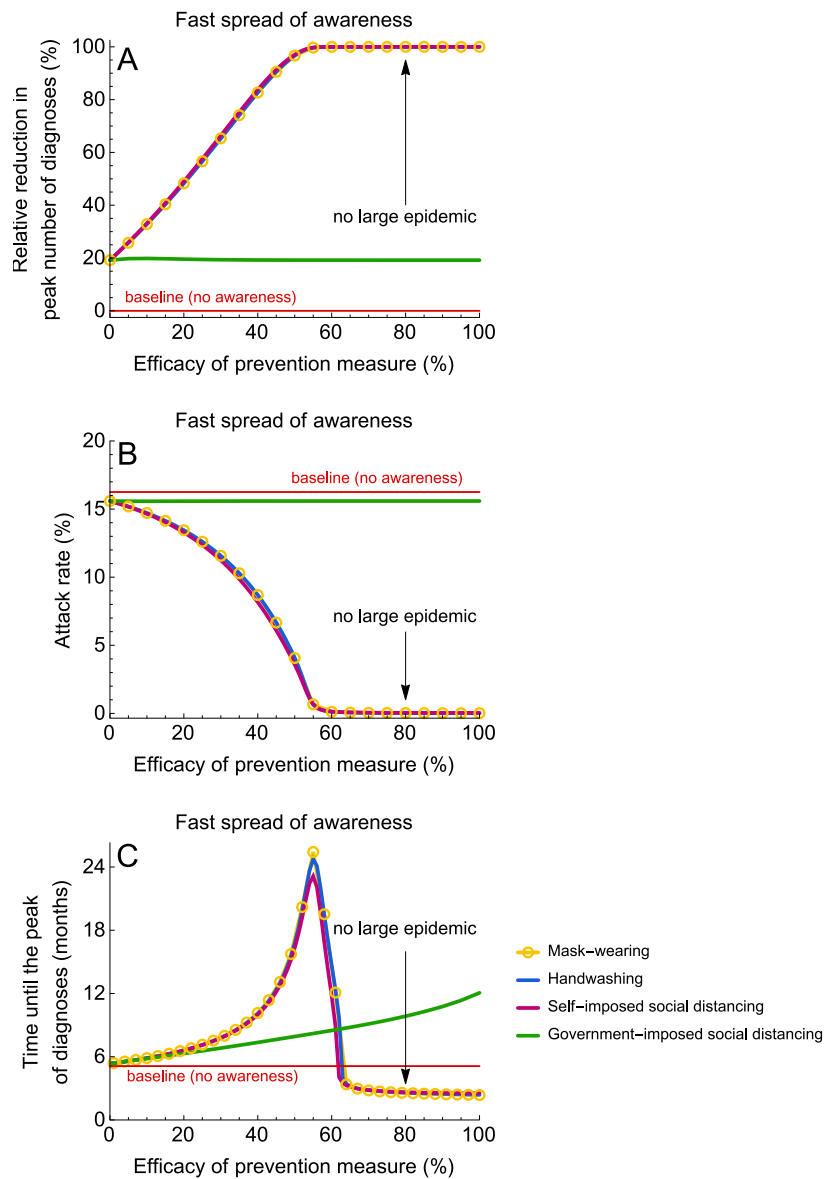
Figure 5: **Impact of prevention measures on the epidemic for a fast rate of awareness spread.** (A), (B) and (C) show the relative reduction in the peak number of diagnoses, the attack rate (proportion of the population that recovered or died after severe infection) and the time until the peak number of diagnoses. The efficacy of prevention measures was varied between 0% and 100%. In the context of this study, the efficacy of social distancing denotes the reduction in the contact rate. The efficacy of handwashing and mask-wearing are given by the reduction in susceptibility and infectivity, respectively. The simulations were started with one case. Government-imposed social distancing was initiated after 10 diagnoses and lifted after 3 months. For parameter values, see Table 1. Please note that the blue line corresponding to handwashing is not visible in (A) since it almost completely overlaps with lines for mask-wearing and self-imposed social distancing.

These measures not only reduce the attack rate (Fig 5 B), diminish and postpone the peak number of diagnoses (Fig 5 A and C), but they can also prevent a large epidemic altogether when their efficacy is sufficiently high (about 50%). Note that when the rate of awareness is fast, as the number of diagnoses grows, the population becomes almost homogeneous, with most individuals being disease-aware. It can be shown that in such populations prevention measures yield comparable results if they have the same efficacy.

Combinations of prevention measures

If government-imposed social distancing is combined with a self-imposed prevention measure, the model predicts that the relative reduction in the peak number of diagnoses and attack rate are determined by the efficacy of the self-imposed measure, while the timing of the peak is determined by the efficacies of both the self-imposed measure and the government intervention. This is demonstrated in Fig 6, where we used a combination of handwashing with efficacies of 30%, 45% and 60% and government-imposed social distancing with efficacy ranging from 0% to 100% for slow and fast spread of awareness. Our results show that the effect of the combined intervention highly depends on the rate of awareness spread. Fast awareness spread is crucial for a large reduction in the peak number of diagnoses (Fig 6 A) and in the attack rate (Fig 6 B). Note, that for fast spread of awareness, a combination of a complete lockdown and handwashing with an efficacy of 30% could postpone the time to the peak number of diagnoses by nearly 10 months (Fig 6 C). Thus, when combined with short-term government-imposed social distancing, handwashing can contribute to mitigating and delaying the epidemic in particular after the lockdown is relaxed. The second wave of the epidemic could be prevented completely if the efficacy of handwashing exceeds 50% (Fig 6 A). The results for the combination of mask-wearing and government-imposed social distancing are similar.

The effect of combinations of self-imposed measures (e.g. handwashing and mask-wearing) is additive (see S1 Fig). This means that, for fast spread of awareness, a large outbreak can be prevented by, for example, a combination of handwashing and self-imposed social distancing each with an efficacy of around 25% (or other efficacies adding up to 50%).
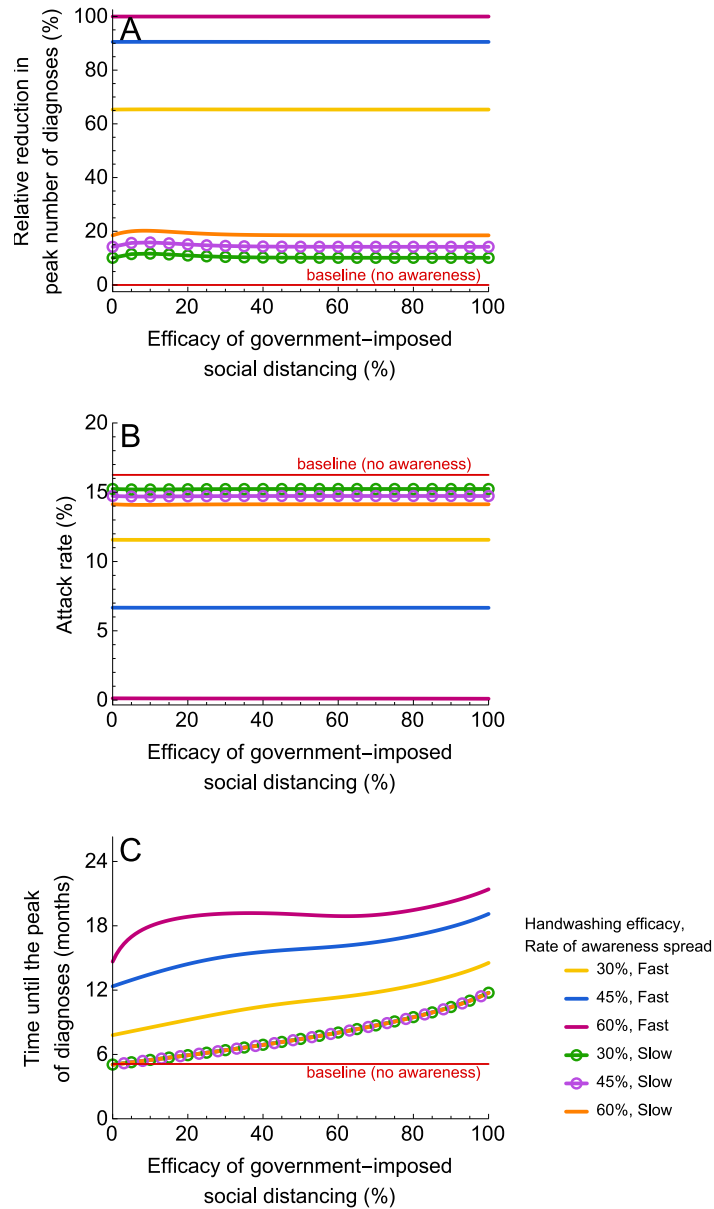
Figure 6: **Impact on the epidemic of a combination of government-imposed social distancing and handwashing** (A), (B) and (C) show the relative reduction in the peak number of diagnoses, the attack rate (proportion of the population that recovered or died after severe infection) and the time until the peak number of diagnoses. The efficacy of handwashing was 30%, 45% and 60%. In the context of this study, the efficacy of social distancing denotes the reduction in the contact rate. The efficacy of handwashing is given by the reduction in susceptibility. The simulations were started with one case. Government-imposed social distancing was initiated after 10 diagnoses and lifted after 3 months. For parameter values, see Table 1.

Discussion

For many countries around the world, the focus of public health officers in the context of COVID-19 epidemic has shifted from containment to mitigation and delay. Our study provides new insights for designing effective outbreak control strategies. Based on our results, we conclude that handwashing, mask-wearing, and social distancing adopted by disease-aware individuals can delay the epidemic peak, flatten the epidemic curve and reduce the attack rate. We show that the rate at which disease awareness spreads has a strong impact on how self-imposed measures affect the epidemic. For a slow rate of awareness spread, self-imposed measures have less impact on transmission, as not many individuals adopt them. However, for a fast rate of awareness spread, their impact on the magnitude and timing of the peak increases with increasing efficacy of the respective measure. For all measures, a large epidemic can be prevented when the efficacy exceeds 50%. Moreover, the effect of combinations of self-imposed measures is additive. In practical terms, it means that SARS-CoV-2 will not cause a large outbreak in a country where 90% of the population adopt handwashing and social distancing that are 25% efficacious (i.e., reduce susceptibility and contact rate by 25%, respectively).

Although our analyses indicate that the effects of self-imposed measures on mitigating and delaying the epidemic for the same efficacies are similar (see Fig 4 and Fig 5), not all explored efficacy values may be achieved for each measure. Wong et al [22] and Cowling et al [24] performed a systematic review and meta-analysis on the effect of handwashing and face masks on the risk of influenza virus infections in the community. While the authors highlight the potential importance of both hand hygiene and face masks, only modest effects could be ascertained with a pooled risk ratio of 0.73 (95% CI: [0.6, 0.89]) for a combination of these two measures. However, the authors also highlight the small number of randomized-controlled trials and the heterogeneity of the studies as notable limitations which may have led to these results. Given the high uncertainty around the efficacies of hand hygiene and mask-wearing on their own, the promotion of a combination of these measures might become preferable to recommending handwashing or mask-only measures. For self-imposed social distancing, contacts might not be eliminated entirely (e.g. household contacts remain) and therefore realistic values of the efficacy of self-imposed social distancing can be close to but may never reach 100%.

Thus, for a fair comparison between measures, realistic efficacy values of a specific measure should be taken into consideration.

We contrasted self-imposed measures stimulated by disease awareness with mandated social distancing. Our analyses show that short-term government-imposed social distancing that is implemented early into the epidemic, can delay the epidemic peak but does not affect its magnitude nor the attack rate. For example, a complete lockdown of 3 months imposing a community-wide contact rate reduction that starts after tens to thousands diagnoses in the country can postpone the peak by about 7 months. Such an intervention is highly desirable, when a vaccine is being developed or when healthcare systems require more time to treat cases or increase capacity. If this intervention is implemented in a population which exercises a self-imposed measure that is continued being practiced even after the lockdown is over, then the delay can be even longer (e.g. up to 10 months for handwashing with 30% efficacy). In the context of countries that implemented social distancing as a measure to 'flatten the curve' of the ongoing epidemics, peaked in cases and now are now planning or have already started gradual lifting of social distancing, it means that governments and public health institutions should intensify the promotion of self-imposed measures to diminish and postpone the peak of the potential second epidemic wave. The potential second wave could be prevented altogether if the coverage of a self-imposed measure in the population and its efficacy are sufficiently high (e.g. 90% and 50%, respectively). Our sensitivity analyses showed that lower or higher efficacies can be required to prevent a large epidemic for countries with smaller or larger basic reproduction numbers (see S3 Text).

Since for many countries the COVID-19 epidemic is still in its early stages, government-imposed social distancing was modeled as a short-term intervention initiated when the number of diagnosed individuals was relatively low. Our sensitivity analyses showed that government interventions introduced later into the epidemic (at 100—1000 diagnoses) and imposed for a longer period of time (3—13 months) not only delay the peak of the epidemic but also reduce it for intermediate efficacy values (see S2 Fig). Previous studies suggested that the timing of mandated social distancing is crucial for its viability in controlling a large disease outbreak [13, 14, 16, 38]. As discussed by Hollingsworth et al [16] and Anderson et al [20], a late introduction

of such interventions may have a significant impact on the epidemic peak and attack rate. However, the authors also showed that the optimal strategy is highly dependent on the desired outcome. A detailed analysis of government intervention with different timings and durations that also takes into account the economic and societal consequences, and the cost of SARS-CoV-2 transmission is a subject for future work.

To our knowledge, our study is the first to provide comparative analysis of a suite of self-imposed measures, government-imposed social distancing and their combinations as strategies for mitigating and delaying a COVID-19 epidemic. Several studies (e.g., [39–42]) looked at the effect of different forms of social distancing but they did not include self-imposed measures such as handwashing and mask-wearing. Some of these studies concluded that one-time social-distancing interventions will be insufficient to maintain COVID-19 prevalence within the critical care capacity [40, 42]. In our analyses, we explored the full efficacy range for all self-imposed prevention measures and different durations and thresholds for initiation of government intervention. Our results allow to draw conclusions on which combination of prevention measures can be most effective in diminishing and postponing the epidemic peak when realistic values for the measure's efficacy are taken into account. We showed that spreading disease awareness such that highly efficacious preventive measures are quickly adopted by individuals can be crucial in reducing SARS-CoV-2 transmission and preventing a large epidemics of COVID-19.

Our model has several limitations. It does not account for stochasticity, demographics, heterogeneities in contact patterns, spatial effects, inhomogeneous mixing, imperfect isolation of individuals with severe disease, and reinfection with COVID-19. Our conclusions can, therefore, be drawn on a qualitative level. Detailed models will have to be developed to design and tailor effective strategies in particular settings. The impact of the duration of immunity has been explored by Kissler et al [43]. The effect of non-permanent immunity on the results of our model would be an interesting subject for future work. To take into account the uncertainty in SARS-CoV-19 epidemiological parameters, we performed sensitivity analyses to test the robustness of the model predictions. As more data become available, our model can be easily updated. In addition, our study assumes that individuals become disease-aware with a rate of awareness acquisition proportional to the number of currently diagnosed individuals.

Other forms for the awareness acquisition rate that incorporate, e.g., the saturation of awareness, may be more realistic and would be interesting to explore in future studies. Furthermore, we assume that handwashing may reduce the susceptibility of an individual down to 0% and therefore neglect aerosol transmission of SARS-CoV-2. Thus, the impact of handwashing on the epidemic may be an overestimation. However, while there is preliminary evidence on SARS-CoV-2 RNA detection in aerosols [44], there is still uncertainty about the level of infectiousness of the detected aerosols and the significance of potential airborne transmission. Current recommendations by the WHO are still focused on droplet and contact precautions [45]. Our model may be adapted as more information on the relative contribution of the transmission routes of COVID-19 emerges.

In conclusion, we provide the first empirical basis of how stimulating the uptake of effective prevention measures, such as handwashing or mask-wearing, combined with government-imposed social distancing intervention, can be pivotal to achieve control over a COVID-19 epidemic. While information on the rising number of COVID-19 diagnoses reported by the media may fuel anxiety in the population, wide and intensive promotion of self-imposed measures with proven efficacy by governments or public health institutions may be a key ingredient to tackle COVID-19.

Declaration of interests

We declare that we have no conflicts of interest.

Author contributions

AT, TMP, NGG, MK, MCJB and GR developed the conceptual framework of the study. AT, TMP, NGG and GR developed the model. AT and GR performed the model analyses. GR produced the results for the main text and conducted sensitivity analyses. NGG conducted the literature search. NGG, AT, TMP and GR wrote the manuscript. AT wrote the appendix. MCJB and MK contributed to interpretation of the results and provided critical review of the manuscript. All authors approved its final version.

Funding

Aidsfonds project P-29704.

References

[1] World Health Organization. *Coronavirus disease 2019 (COVID-19). Situation Report − 51.* 2020. `https : / / www . who . int / docs / default - source / coronaviruse / situation - reports / 20200311-sitrep-51-covid-19.pdf`.

[2] Gostic, KM, Kucharski, AJ, and Lloyd-Smith, JO. Effectiveness of traveller screening for emerging pathogens is shaped by epidemiology and natural history of infection. In: *eLife* 4 (2015), e05564. DOI: `10.7554/eLife.05564`.

[3] Li, R, Pei, S, Chen, B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). In: *Science* (2020). DOI: `10.1126/science.abb3221`.

[4] Tindale, L, Coombe, M, Stockdale, JE, et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. In: *medRxiv* (2020). DOI: `10.1101/2020.03.03.20029983`.

[5] Li, Q, Guan, X, Wu, P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. In: *New England Journal of Medicine* 382.13 (Mar. 2020), pp. 1199–1207. DOI: `10.1056/nejmoa2001316`.

[6] European Centre for Disease Prevention and Control. *Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK − seventh update.* `https : / / www . ecdc . europa . eu / sites / default / files / documents / RRA - seventh - update - Outbreak-of-coronavirus-disease-COVID-19.pdf`.

[7] European Centre for Disease Prevention and Control. *Outbreak of acute respiratory syndrome associated with a novel coronavirus, China: first local transmission in the EU/EEA - third update.* `https : / / www . ecdc . europa . eu / sites / default / files / documents / novel - coronavirus - risk - assessment - china - 31 - january-2020_0.pdf`.

[8] Ferretti, L, Wymant, C, Kendall, M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. In: *Science* 368.6491 (Mar. 2020), eabb6936. DOI: `10.1126/science.abb6936`.

[9] Hellewell, J, Abbott, S, Gimma, A, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. In: *The Lancet Global Health* (2020). DOI: `10.1016/S2214-109X(20)30074-7`.

[10] Fraser, C, Riley, S, Anderson, RM, et al. Factors that make an infectious disease outbreak controllable. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.16 (2004), pp. 6146–6151. DOI: `10.1073/pnas.0307506101`.

[11] Klinkenberg, D, Fraser, C, and Heesterbeek, H. The Effectiveness of Contact Tracing in Emerging Epidemics. In: *PLoS ONE* 1.1 (2006). Ed. by Getz, W, e12. DOI: `10.1371/journal.pone.0000012`.

[12] Wilder-Smith, A and Freedman, DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. In: *Journal of Travel Medicine* (2020). DOI: `10.1093/jtm/taaa020`.

[13] Hatchett, RJ, Mecher, CE, and Lip-sitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. In: *Proceedings of the National Academy of Sciences* 104.18 (2007), pp. 7582–7587. DOI: `10.1073/pnas.0610941104`.

[14] Bootsma, MCJ and Ferguson, NM. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. In: *Proceedings of the National Academy of Sciences* 104.18 (2007), pp. 7588–7593. DOI: `10.1073/pnas.0611071104`.

[15] Markel, H, Lipman, HB, Navarro, JA, et al. Nonpharmaceutical Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic. In: *JAMA* 298.6 (2007), pp. 644–654. DOI: `10.1001/jama.298.6.644`.

[16] Hollingsworth, TD, Klinkenberg, D, Heesterbeek, H, et al. Mitigation Strategies for Pandemic Influenza A: Balancing Conflicting Policy Objectives. In: *PLoS Comput Biol* 7.2 (2011), pp. 1–11. DOI: `10.1371/journal.pcbi.1001076`.

[17] Cauchemez, S, Ferguson, NM, Wachtel, C, et al. Closure of schools during an influenza pandemic. In: *The Lancet Infectious Diseases* 9.8 (2009), pp. 473–481.

[18] Jackson, C, Vynnycky, E, Hawker, J, et al. School closures and influenza: systematic review of epidemiological studies. In: *BMJ Open* 3.2 (2013). DOI: `10.1136/bmjopen-2012-002149`. `https://bmjopen.bmj.com/content/3/2/e002149`.

[19] Control, ECfDPa. *Guidelines for the use of non-pharmaceutical measures to delay and mitigate the impact of 2019-nCoV*. `https://www.ecdc.europa.eu/en/publications-data/guidelines-use-non-pharmaceutical-measures-delay-and-mitigate-impact-2019-ncov`.

[20] Anderson, RM, Heesterbeek, H, Klinkenberg, D, et al. How will country-based mitigation measures influence the course of the COVID-19 epidemic? In: *The Lancet* (2020). DOI: `10.1016/S0140-6736(20)30567-5`.

[21] Kampf, G. Efficacy of ethanol against viruses in hand disinfection. In: *Journal of Hospital Infection* 98.4 (2018), pp. 331–338. DOI: `10.1016/j.jhin.2017.08.025`.

[22] Wong, V, Cowling, B, and Aiello, A. Hand hygiene and risk of influenza virus infections in the community: a systematic review and meta-analysis. In: *Epidemiology and infection* 142.5 (2014), pp. 922–932. DOI: `10.1017/S095026881400003X`.

[23] Institute Of Medicine. Preventing transmission of pandemic influenza and other viral respiratory diseases: Personal protective equipment for healthcare personnel. Update 2010. Washington, DC: The National Academies Press, 2011.

[24] Cowling, BJ, Zhou, Y, Ip, DKM, et al. Face masks to prevent transmission of influenza virus: a systematic review. In: *Epidemiology and Infection* 138.4 (2010), pp. 449–456. DOI: `10.1017/S0950268809991658`.

[25] Leung, NHL, Chu, DKW, Shiu, EYC, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. In: *Nature Medicine* (2020), pp. 1–5. DOI: `10.1038/s41591-020-0843-2`.

[26] Funk, S, Gilad, E, Watkins, C, et al. The spread of awareness and its impact on epidemic outbreaks. In: *Proceedings of the National Academy of Sciences* 106.16 (2009), pp. 6872–6877. DOI: `10.1073/pnas.0810762106`.

[27] Perra, N, Balcan, D, Gonçalves, B, et al. Towards a characterization of behavior-disease models. In: *PLoS ONE* 6.8 (2011), e23084. DOI: `10.1371/journal.pone.0023084`.

[28] He, D, Dushoff, J, Day, T, et al. Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. In: *Proceedings of the Royal Society B: Biological Sciences* 280.1766 (Sept. 2013), p. 20131345. DOI: `10.1098/rspb.2013.1345`.

[29] Liu, Y, Yan, LM, Wan, L, et al. Viral dynamics in mild and severe cases of COVID-19. In: *The Lancet Infectious Diseases* 0.0 (2020). DOI: `10.1016/S1473-3099(20)30232-2`.

[30] Park, M, Cook, AR, Lim, JT, et al. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. In: *J. Clin. Med.* 9.4 (2020), p. 967. DOI: `10.3390/jcm9040967`.

[31] Mossong, J, Hens, N, Jit, M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. In: *PLoS Medicine* 5.3 (2008), p. 1. DOI: `10.1371/journal.pmed.0050074`.

[32] Wu, Z and McGoogan, JM. *Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention.* Apr. 2020. DOI: `10.1001/jama.2020.2648`.

[33] Backer, JA, Klinkenberg, D, and Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. In: *Eurosurveillance* 25.5 (2020). DOI: `10.2807/1560-7917.ES.2020.25.5.2000062`.

[34] World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 — 24 February 2020.* 2020. `https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---24-february-2020`.

[35] Riou, J, Counotte, MJ, Hauser, A, et al. *Adjusted age-specific case fatality ratio during the COVID-19 epidemic in Hubei, China, January and February 2020.* 2020.

[36] World Health Organization. *Coronavirus disease (COVID-19) advice for the public: When and how to use masks.* 2020. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks`.

[37] Jarvis, CI, Van Zandvoort, K, Gimma, A, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. In: *BMC Medicine* 18.1 (May 2020), pp. 1–10. DOI: `10.1186/s12916-020-01597-8`.

[38] Lauro, FD, Kiss, IZ, and Miller, J. *The timing of one-shot interventions for epidemic control.* 2020.

[39] Prem, K, Liu, Y, Russell, TW, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. In: *The Lancet Public Health* 5.5 (2020), e261–e270. DOI: `10.1016/S2468-2667(20)30073-6`.

[40] Imperial College London. *Report 9 - Impact of non-pharmaceutical interventions ({NPI}s) to reduce {COVID}-19 mortality and healthcare demand.* 2020. `https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf`.

[41] Davies, NG, Kucharski, AJ, Eggo, RM, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. In: *The Lancet Public Health* 5.7 (July 2020), e375–e385. DOI: `10.1016/S2468-2667(20)30133-X`.

[42] Kissler, S, Tedijanto, C, Lipsitch, M, et al. Social distancing strategies for curbing the COVID-19 epidemic. In: *medRxiv* (2020). DOI: 10.1101/2020.03.22. 20041079.

[43] Kissler, SM, Tedijanto, C, Goldstein, E, et al. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. In: *Science* (2020). DOI: 10.1126/science.abb5793.

[44] Liu, Y, Ning, Z, Chen, Y, et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. In: *Nature* (2020). DOI: 10.1038/s41586-020-2271-3.

[45] World Health Organization. *Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations*. 2020. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations.

[46] Moberly, T. Covid-19: school closures and bans on mass gatherings will need to be considered, says England's CMO. In: *BMJ* 368 (Feb. 2020), p. m806. DOI: 10.1136/bmj.m806.

Supplementary material

Table of Contents

S1: Mathematical description of the model

Baseline transmission model

We developed a deterministic compartmental model describing SARS-CoV-2 transmission in a population stratified by disease status (see Figure 1 in the main text). In the baseline model, individuals are classified as susceptible (S), latently infected (E), infectious with mild disease (I_M), infectious with severe disease (I_S), diagnosed and isolated (I_D), and recovered (R_M and R_S after mild or severe disease, respectively). Susceptible individuals (S) become latently infected (E) through contact with infectious individuals (I_M and I_S) with the force of infection λ_{inf} dependent on the fractions of the population in I_M and I_S compartments. Latently infected individuals (E) become infectious at a rate α; a proportion p of the latently infected individuals will go to the I_M compartment, a proportion $(1-p)$ to the I_S compartment. We assume that infectious individuals with mild disease (I_M) do not require medical attention and recover (R_M) with rate γ_M without being conscious of having contracted COVID-19. Infectious individuals with severe disease (I_S) are unable to recover without medical help, and subsequently get diagnosed and isolated (I_D) with rate ν (in e.g. hospitals, long-term care facilities, nursing homes) and know or suspect they have COVID-19 when they are detected. Therefore, the diagnosed compartment I_D contains infectious individuals with severe disease who are both officially diagnosed and get treatment in healthcare institutions and are not officially diagnosed but have a disease severe enough to suspect they have COVID-19 and require treatment as well as isolation. For simplicity, isolation of these individuals is assumed to be perfect until recovery (R_S) which occurs at rate γ_S, and, hence, they neither contribute to transmission nor to the contact process. Given the timescale of the epidemic and the lack of reliable reports on reinfections, we assume that recovered individuals (R_M and R_S) cannot be reinfected. The infectivity of infectious individuals with mild disease is lower by a factor $0 \le \sigma \le 1$ than the infectivity of infectious individuals with severe disease [29]. Natural birth and death processes are neglected as the time scale of the epidemic is short compared to the mean life span of individuals. However, isolated infectious individuals with severe disease (I_D) may be removed from the population due to disease-associated mortality at rate η.

The transmission model without awareness is given by the following system of ordinary differential equations

$$
\begin{aligned}
\frac{\mathrm{d}S(t)}{\mathrm{d}t} &= -S(t)\lambda_{\mathrm{inf}}(t) \\
\frac{\mathrm{d}E(t)}{\mathrm{d}t} &= S(t)\lambda_{\mathrm{inf}}(t) - \alpha E(t) \\
\frac{\mathrm{d}I_M(t)}{\mathrm{d}t} &= p\alpha E(t) - \gamma_M I_M(t) \\
\frac{\mathrm{d}I_S(t)}{\mathrm{d}t} &= (1-p)\alpha E(t) - \nu I_S(t) \\
\frac{\mathrm{d}I_D(t)}{\mathrm{d}t} &= \nu I_S(t) - \gamma_S I_D(t) - \eta I_D(t) \\
\frac{\mathrm{d}R_M(t)}{\mathrm{d}t} &= \gamma_M I_M(t) \\
\frac{\mathrm{d}R_S(t)}{\mathrm{d}t} &= \gamma_S I_D(t),
\end{aligned}
\tag{1.1}
$$

where

$$
\lambda_{\mathrm{inf}}(t) = \frac{\beta}{N(t)}\left[\sigma I_M(t) + I_S(t)\right]
\tag{1.2}
$$

is the force of infection and $N(t) = S(t) + E(t) + I_M(t) + I_S(t) + R_M(t) + R_S(t)$ is the total number of individuals who participate in the contact process.

Transmission model with disease awareness

In the extended model with awareness, the population is stratified not only by the disease status but also by the awareness status into disease-aware (S^a, E^a, I_M^a, I_S^a, I_D^a, and R_M^a) and disease-unaware (S, E, I_M, I_S, and R_M) (Figure 2 A in the main text). Disease awareness is a state that can be acquires as well as lost. Disease-aware individuals are distinguished from unaware individuals in two essential ways. First, infectious individuals with severe disease who are disease-aware (I_S^a) get diagnosed and isolated faster (I_D^a) with rate ν^a, stay in isolation for a shorter period of time (recovery rate γ_S^a) and have lower disease-associated mortality (rate η^a) than the same category of unaware individuals. The assumption we make here is that disease-aware individuals (I_S^a) recognize they may have COVID-19 on average faster than disease-unaware individuals (I_S) and get medical help earlier which leads to a better prognosis of I_D^a individuals as compared to I_D individuals. Second, disease-

aware individuals are assumed to use self-imposed measures such as handwashing, mask-wearing and self-imposed social distancing that can lower their susceptibility, infectivity and/or contact rate. Individuals who know or suspect their disease status (I_D, I_D^a and R_S) do not adapt any such measures since they assume that they cannot contract the disease again. Hence, they are excluded from the awareness transition process and their behaviour in the contact process is identical to disease-unaware individuals.

A schematic representation of the awareness dynamics is given in Figure 2 B in the main text. Individuals of type S, E, I_M, I_S, and R_M become aware of the disease with the awareness acquisition rate $\lambda_{\text{aware}}(t)$ proportional to the current number of diagnosed individuals via information shared by the government or media

$$\lambda_{\text{aware}}(t) = \delta \cdot [I_D(t) + I_D^a(t)],$$

where δ is a constant which describes how fast unaware individuals become aware per unit of time. This formulation is based on Eq. (7) in Perra et al. [27].

We assume that awareness fades and individuals return to the unaware state at a constant rate. The latter means that they no longer use self-imposed measures. We propose that awareness acquisition and fading rates are the same for individuals who are susceptible (S), latently infected (E), infectious with mild disease (I_M) and recovered after mild disease (R_M). The rate of awareness acquisition for these individuals is a factor $0 \leq k \leq 1$ lower than the rate of awareness acquisition for infectious individuals with severe disease (I_S). Also, infectious individuals with severe disease are more cautious and, therefore, lose awareness at a slower rate than other individuals. Thus, we use μ to denote the decay rate in compartments S^a, E^a, I_M^a, and R_M^a and μ_S for compartment I_S^a, such that $\mu > \mu_S$.

The difference in disease severity and state of awareness affects the transmission rates and we define the following matrix to summarize transmission rates between different

types of susceptible and infectious individuals

$$M(t) = \begin{array}{c} \\ \text{unaware } S \\ \text{aware } S \end{array} \overset{\begin{array}{cccc} \text{unaware } I_M & \text{unaware } I_S & \text{aware } I_M & \text{aware } I_S \end{array}}{\left[\begin{array}{cccc} M_{11}(t) & M_{12}(t) & M_{13}(t) & M_{14}(t) \\ M_{21}(t) & M_{22}(t) & M_{23}(t) & M_{24}(t) \end{array} \right]}. \tag{1.3}$$

Here $[M]_{11}$ captures transmission of infection from unaware I_M to unaware S, $[M]_{12}$ from unaware I_S to unaware S, $[M]_{13}$ from aware I_M to unaware S, $[M]_{14}$ from from aware I_M to unaware S. Similarly, the second row of the matrix captures transmission of infection to susceptible individuals who are aware, S^a. To sum up,

$$S + I_M \xrightarrow{[M]_{11}} E + I_M \ , \ S + I_S \xrightarrow{[M]_{12}} E + I_S$$

$$S + I_M^a \xrightarrow{[M]_{13}} E + I_M^a \ , \ S + I_S^a \xrightarrow{[M]_{14}} E + I_S^a$$

$$S^a + I_M \xrightarrow{[M]_{21}} E^a + I_M \ , \ S^a + I_S \xrightarrow{[M]_{22}} E^a + I_S$$

$$S^a + I_M^a \xrightarrow{[M]_{23}} E^a + I_M^a \ , \ S^a + I_S^a \xrightarrow{[M]_{24}} E^a + I_S^a.$$

The transmission model with awareness is given by the following system of ordinary

differential equations

$$\frac{dS(t)}{dt} = -S(t)\lambda_{\text{inf}}(t) - kS(t)\lambda_{\text{aware}}(t) + \mu S^a(t)$$

$$\frac{dE(t)}{dt} = S(t)\lambda_{\text{inf}}(t) - \alpha E(t) - kE(t)\lambda_{\text{aware}}(t) + \mu E^a(t)$$

$$\frac{dI_M(t)}{dt} = p\alpha E(t) - \gamma_M I_M(t) - kI_M(t)\lambda_{\text{aware}}(t) + \mu I_M^a(t)$$

$$\frac{dI_S(t)}{dt} = (1-p)\alpha E(t) - \nu I_S(t) - I_S(t)\lambda_{\text{aware}}(t) + \mu_S I_S^a(t)$$

$$\frac{dI_D(t)}{dt} = \nu I_S(t) - \gamma_S I_D(t) - \eta I_D(t)$$

$$\frac{dS^a(t)}{dt} = -S^a(t)\lambda_{\text{inf}}^a(t) + kS(t)\lambda_{\text{aware}}(t) - \mu S^a(t)$$

$$\frac{dE^a(t)}{dt} = S^a(t)\lambda_{\text{inf}}^a(t) - \alpha E^a(t) + kE(t)\lambda_{\text{aware}}(t) - \mu E^a(t) \qquad (1.4)$$

$$\frac{dI_M^a(t)}{dt} = p\alpha E^a(t) - \gamma_M I_M^a(t) + kI_M(t)\lambda_{\text{aware}}(t) - \mu I_M^a(t)$$

$$\frac{dI_S^a(t)}{dt} = (1-p)\alpha E^a(t) - \nu^a I_S^a + I_S(t)\lambda_{\text{aware}}(t) - \mu_S I_S^a(t)$$

$$\frac{dI_D^a(t)}{dt} = \nu^a I_S^a(t) - \gamma_S^a I_D^a(t) - \eta^a I_D^a(t)$$

$$\frac{dR_M(t)}{dt} = \gamma_M I_M(t) - kR_M(t)\lambda_{\text{aware}}(t) + \mu R_M^a(t)$$

$$\frac{dR_M^a(t)}{dt} = \gamma_M I_M^a(t) + kR_M(t)\lambda_{\text{aware}}(t) - \mu R_M^a(t)$$

$$\frac{dR_S(t)}{dt} = \gamma_S I_D(t) + \gamma_S^a I_D^a(t),$$

where

$$\lambda_{\text{aware}}(t) = \delta \cdot [I_D(t) + I_D^a(t)] \qquad (1.5a)$$

$$\lambda_{\text{inf}}(t) = [M(t)]_{11}I_M(t) + [M(t)]_{12}I_S(t) + [M(t)]_{13}I_M^a(t) + [M(t)]_{14}I_S^a(t) \qquad (1.5b)$$

$$\lambda_{\text{inf}}^a(t) = [M(t)]_{21}I_M(t) + [M(t)]_{22}I_S(t) + [M(t)]_{23}I_M^a(t) + [M(t)]_{24}I_S^a(t). \qquad (1.5c)$$

For the population where disease-aware individuals do not use self-imposed measures matrix M takes the following form

$$M_0(t) = \frac{\beta}{N_T(t)} \begin{bmatrix} \sigma & 1 & \sigma & 1 \\ \sigma & 1 & \sigma & 1 \end{bmatrix} \qquad (1.6)$$

with $N_T(t) = S(t) + E(t) + I_M(t) + I_S(t) + S^a(t) + E^a(t) + I_M^a(t) + I_S^a(t) + R_M(t) + R_M^a(t) + R_S(t)$.

Estimates of epidemiological parameters were obtained from previous studies and are shown in Table 1 in the main text.

Prevention measures

We considered short-term government intervention aimed at fostering social distancing in the population and a suite of measures self-imposed by disease-aware individuals, i.e., mask-wearing, handwashing, and self-imposed social distancing.

Mask-wearing

Mask-wearing does not reduce the individual's susceptibility because laypersons, i.e., not medical professionals, are unfamiliar with correct procedures for its use and may often engage in face-touching and mask adjustment.[36] The efficacy of mask wearing is described by a reduction in infectivity of disease-aware infectious individuals (I_S^a and I_M^a) and is represented by a factor r_1, $0 \leq r_1 \leq 1$. The respective transmission matrix is given by

$$M_1 = \frac{\beta}{N_T(t)} \begin{bmatrix} \sigma & 1 & r_1\sigma & r_1 \\ \sigma & 1 & r_1\sigma & r_1 \end{bmatrix} \tag{1.7}$$

with $N_T(t) = S(t) + E(t) + I_M(t) + I_S(t) + R_M(t) + R_S(t) + S^a(t) + E^a(t) + I_M^a(t) + I_S^a(t) + R_M^a(t)$.

Handwashing

Since infectious individuals may transmit the virus to others without direct physical contact, we assume that handwashing only reduces one's susceptibility. The efficacy of handwashing is described by a reduction in susceptibility (i.e., probability of transmission of infection per single contact) of susceptible disease-aware individuals (S^a) and is represented by a factor r_2, $0 \leq r_2 \leq 1$. The respective transmission matrix is given by

$$M_2 = \frac{\beta}{N_T(t)} \begin{bmatrix} \sigma & 1 & \sigma & 1 \\ r_2\sigma & r_2 & r_2\sigma & r_2 \end{bmatrix} \tag{1.8}$$

with $N_T(t) = S(t) + E(t) + I_M(t) + I_S(t) + R_M(t) + R_S(t) + S^a(t) + E^a(t) + I_M^a(t) + I_S^a(t) + R_M^a(t)$.

Self-imposed social distancing

Disease awareness may also lead individuals to practice social distancing, i.e., maintaining distance to others and avoiding congregate settings. Social distancing of disease-aware individuals is modeled as a reduction in their contact rate. As a consequence, this measure leads to a change in mixing patterns in the population. We model the reduction in contact rate of aware individuals by using the parameter r_3, $0 \leq r_3 \leq 1$. Recall that individuals who recovered from a mild infection may still think of themselves as susceptible, which implies that they are affected by the awareness contagion process. They can, therefore, practice social distancing after they recover. The respective transmission matrix is given by

$$M_4 = \frac{\beta}{N(t) + r_3 N^a(t)} \begin{bmatrix} \sigma & 1 & r_3\sigma & r_3 \\ r_3\sigma & r_3 & r_3^2\sigma & r_3^2 \end{bmatrix}, \tag{1.9}$$

where $N(t) = S(t) + E(t) + I_M(t) + I_S(t) + R_M(t) + R_S(t)$ and $N^a(t) = S^a(t) + E^a(t) + I_M^a(t) + I_S^a(t) + R_M^a(t)$.

Short-term government-imposed social distancing

Governments may decide to promote social distancing policies through interventions such as school and workplace closures, or by issuing a ban on large gatherings and issuing stay-at-home orders [12, 17, 18, 46], if the number of diagnosed individuals exceeds a certain threshold. Such a policy will cause a community-wide contact rate reduction, regardless of the awareness status. We model government-imposed social distancing by reducing the average contact rate in the population by a factor r_4, $0 \leq r_4 < 1$. This intervention is initiated if the number of diagnosed individuals is above a certain threshold $\tilde{I}$ (e.g., $10 - 1000$ individuals) and terminates after a fixed period of time, denoted $t_{\text{intervention}}$ (e.g., $1 - 13$ months). As such, we assume that the intervention is implemented early in the epidemic. If t_{start} is the time for which $I_D(t) + I_D^a(t) \geq \tilde{I}$, then the transmission

matrix is given by

$$M_5(t) = \frac{\beta}{N_T(t)} \cdot \tilde{r} \cdot \begin{bmatrix} \sigma & 1 & \sigma & 1 \\ \sigma & 1 & \sigma & 1 \end{bmatrix}, \tag{1.10}$$

where

$$\tilde{r} = \begin{cases} r_4, & \text{if } I_D(t) + I_D^a(t) \geq \tilde{I} \text{ and } t \leq t_{\text{start}} + t_{\text{intervention}} \\ 1, & \text{otherwise} \end{cases}$$

and $N_T(t) = S(t) + E(t) + I_M(t) + I_S(t) + R_M(t) + R_S(t) + S^a(t) + E^a(t) + I_M^a(t) + I_S^a(t) + R_M^a(t)$.

S2: Impact of awareness process on the probability of infection

We compared changes in the probability of infection for individuals who are aware and who are unaware over the studied period of $T_{max} = 2.5$ years for various scenarios of self-imposed measures and government-imposed social distancing (Figure S1). The probabilities were calculated using the following equations

$$\text{Probability of infection of aware individuals} = 1 - \exp\left[-\int_0^{T_{max}} \lambda_{\text{inf}}(t)\,dt\right] \quad (1.11a)$$

$$\text{Probability of infection of unaware individuals} = 1 - \exp\left[-\int_0^{T_{max}} \lambda_{\text{inf}}^a(t)\,dt\right], \quad (1.11b)$$

where $\lambda_{\text{inf}}(t)$ and $\lambda_{\text{inf}}^a(t)$ are given by Eq. (5b) and (5c).

We observe that when aware individuals adapt mask wearing, the probability of infection is equally reduced for aware and unaware individuals, as it reduces the infectivity of a part of the population. This measure is most efficient when the rate of awareness spread is fast and infectivity reduction due to mask use is above 40%.

In the case of handwashing, the probability is reduced for both aware and unaware individuals. However, aware individuals experience a larger reduction. Handwashing yields direct protection to aware individuals, while unaware individuals benefit indirectly from the overall reduced infection level. Similar to mask-wearing, the infection probabilities for both aware and unaware individuals decrease drastically when the efficacy of handwashing exceeds 40% and the rate of awareness spread is fast.

Effects of self-imposed social distancing depend on the rate of awareness spread as well. While aware individuals have reduced probability of infection regardless of the rate of awareness spread, the unaware individuals will only benefit from it when the rate of awareness spread is fast. This is due to modified mixing patterns that emerge as a result of heterogeneous contact rates.

Finally, government-imposed short-term social distancing which lasts for 3 months has no effect on acquisition rates for aware and unaware individuals. The respective probability of infection is marked with dashed red line in Figure S1.
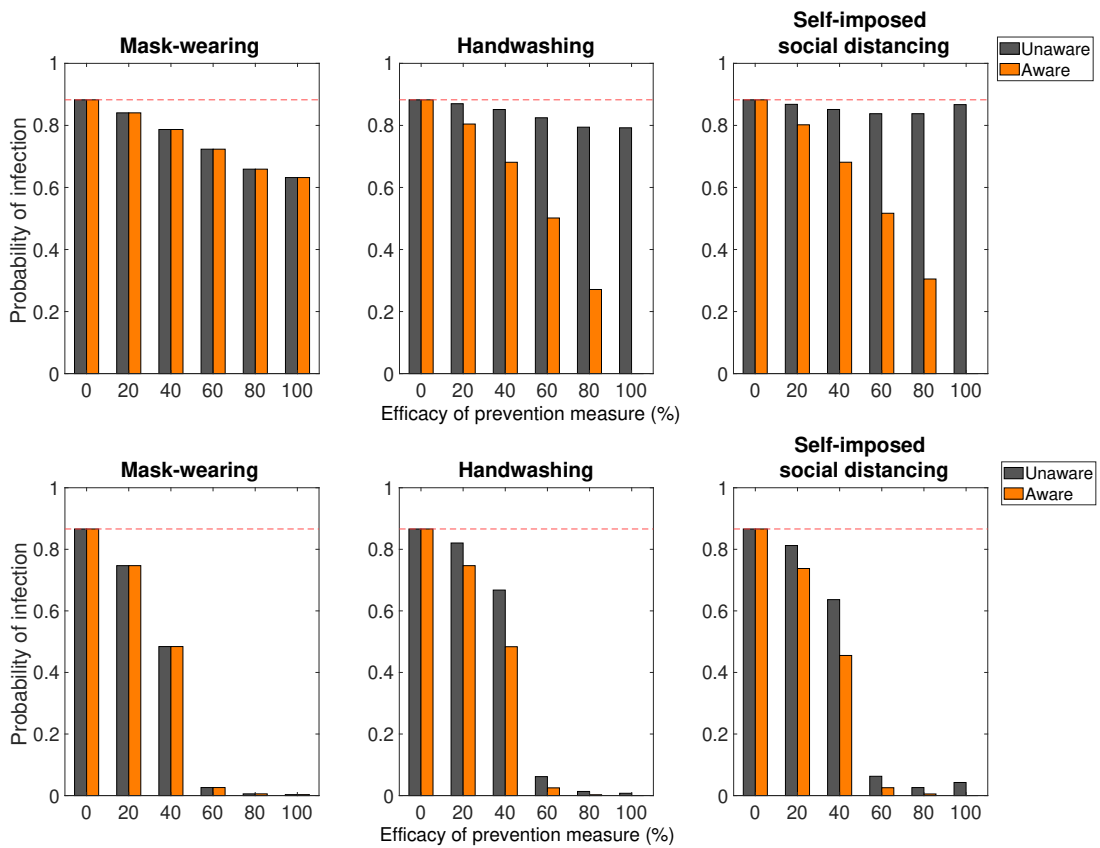
Impact of awareness process on the probability of infection. TOP panels were obtained for a slow rate of awareness spread. **BOTTOM** panels were obtained for a fast rate of awareness spread. The dashed red line indicates probability of infection in the model with awareness and no prevention measures.

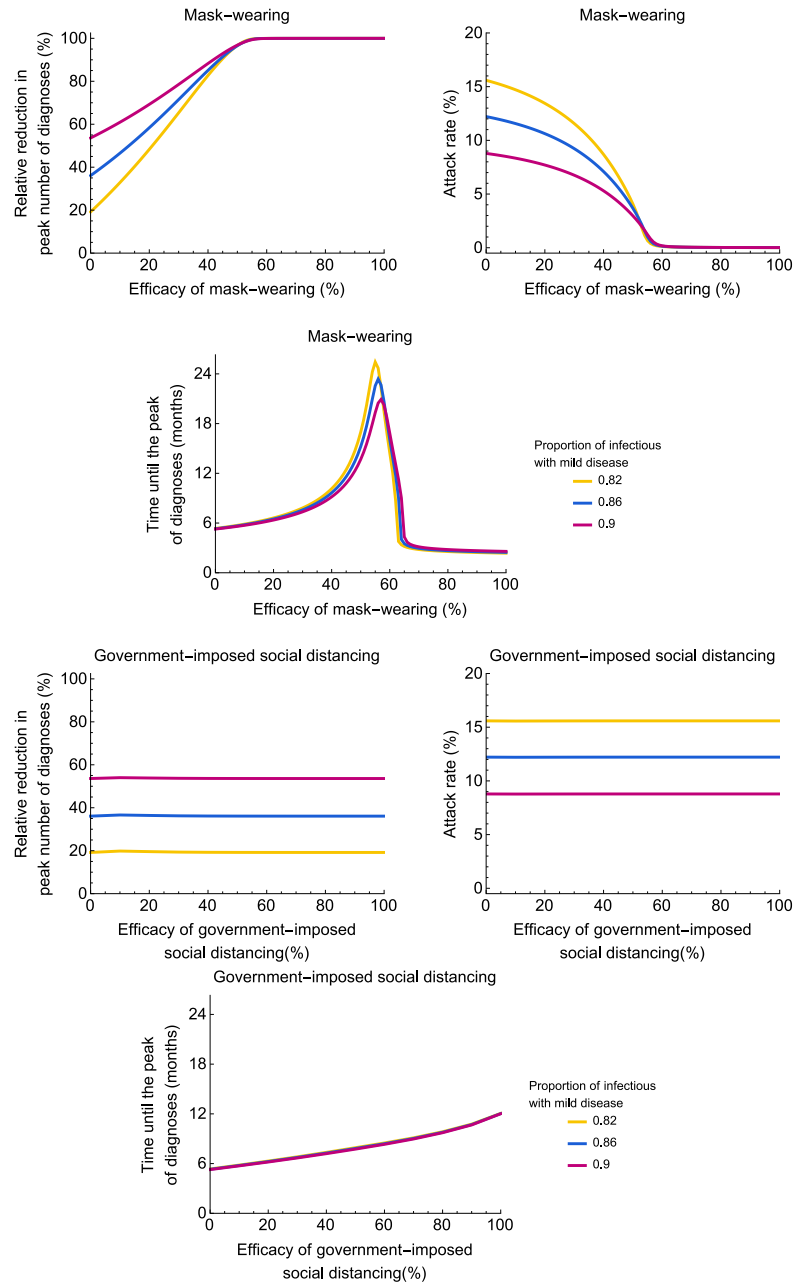# S3: Sensitivity analyses of the baseline transmission model

To allow for the uncertainty in the parameters of the baseline transmission model, we conducted sensitivity analyses with respect to the proportion of infectious individuals with mild disease (Figure S2), relative infectivity of infectious individuals with mild disease (Figure S3), recovery period of infectious individuals with mild disease (Figure S4), delay from onset of infectiousness to diagnosis for infectious individuals with severe disease (Figure S5), and basic reproduction number (Figure S6). Since our findings in the main text demonstrate that the impact of self-imposed measures is similar across all measures that we considered, we present here only sensitivity analyses for mask-wearing and government-imposed social distancing. All fig/chapter1/ were made for a fast rate of awareness spread. In each figure, the panels show the relative reduction in the peak number of diagnoses, the attack rate (proportion of the population that recovered or died after severe infection) and the time until the peak number of diagnoses. In the top row of panels, the efficacy of mask-wearing was varied between 0% and 100%. In the bottom row of panels, the efficacy of government-imposed social distancing was varied between 0% and 100%. In the context of this study, the efficacy of social distancing denotes the reduction in the contact rate. The efficacy of mask-wearing is given by the reduction in infectivity. The simulations were started with one case. Government-imposed social distancing was initiated after 10 diagnoses and lifted after 3 months. For fixed parameter values, see Table 1 in the main text. The parameter which was varied in the sensitivity analyses and the respective range is indicated in each figure.

The main findings of the sensitivity analyses for government-imposed social distancing are

- the time until the peak number of diagnoses does not depend much on any of the explored parameters, except for the basic reproduction number

- the relative reduction in the peak number of diagnoses increases and the attack rate decreases with increasing proportion of infectious individuals with mild disease, decreasing relative infectivity of infectious individuals with mild disease, shorter recovery period of individuals with mild disease and shorter delay from onset of infectiousness to diagnosis for individuals with severe symptoms

The main findings of the sensitivity analyses for mask-wearing are

- for all explored parameter ranges there is a value of efficacy of mask-wearing for which a large epidemic can be prevented

- this critical value of efficacy is very sensitive to the basic reproduction number; smaller value of efficacy is required to prevent a large epidemic for smaller R_0

- the critical value decreases with decreasing relative infectivity of infectious individuals with mild disease and shorter delay from onset of infectiousness to diagnosis for individuals with severe symptoms and is not sensitive to the remaining parameters
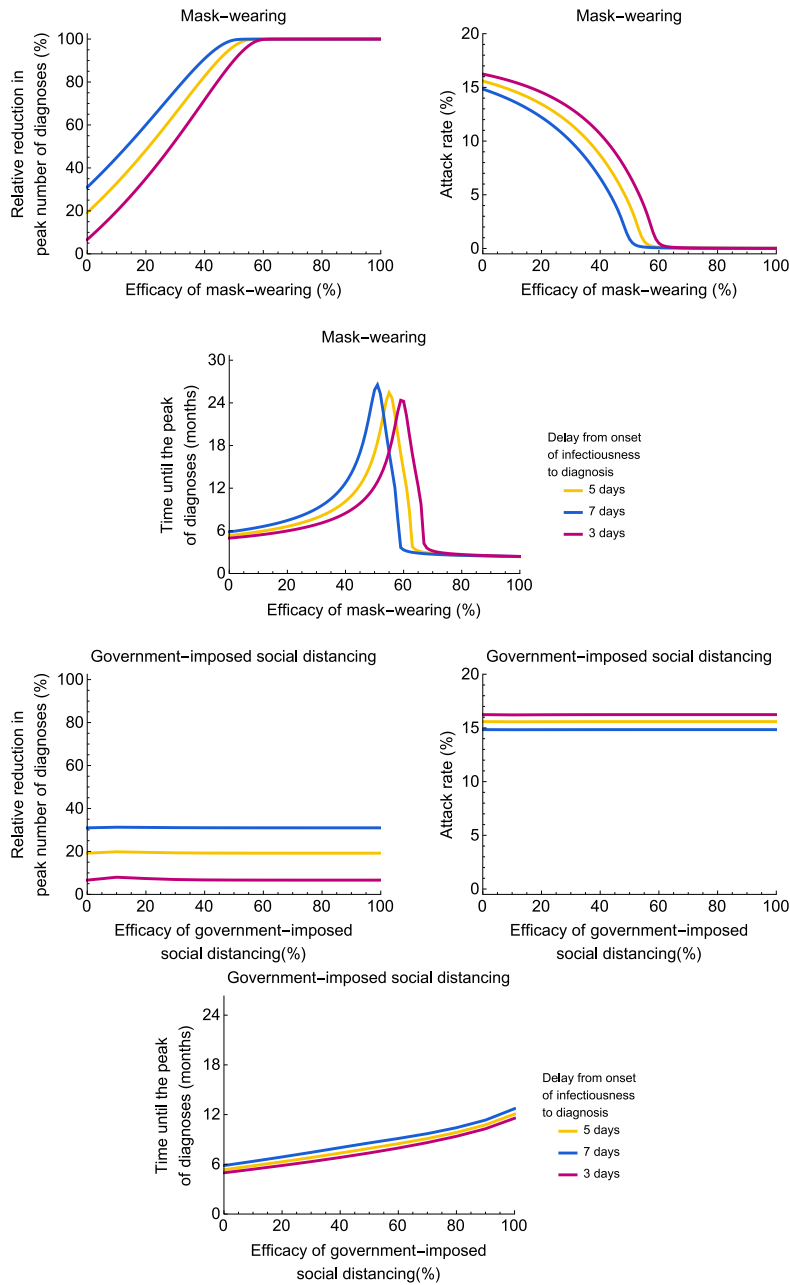
Sensitivity analyses of the baseline transmission model with respect to the proportion of infectious individuals with mild disease.
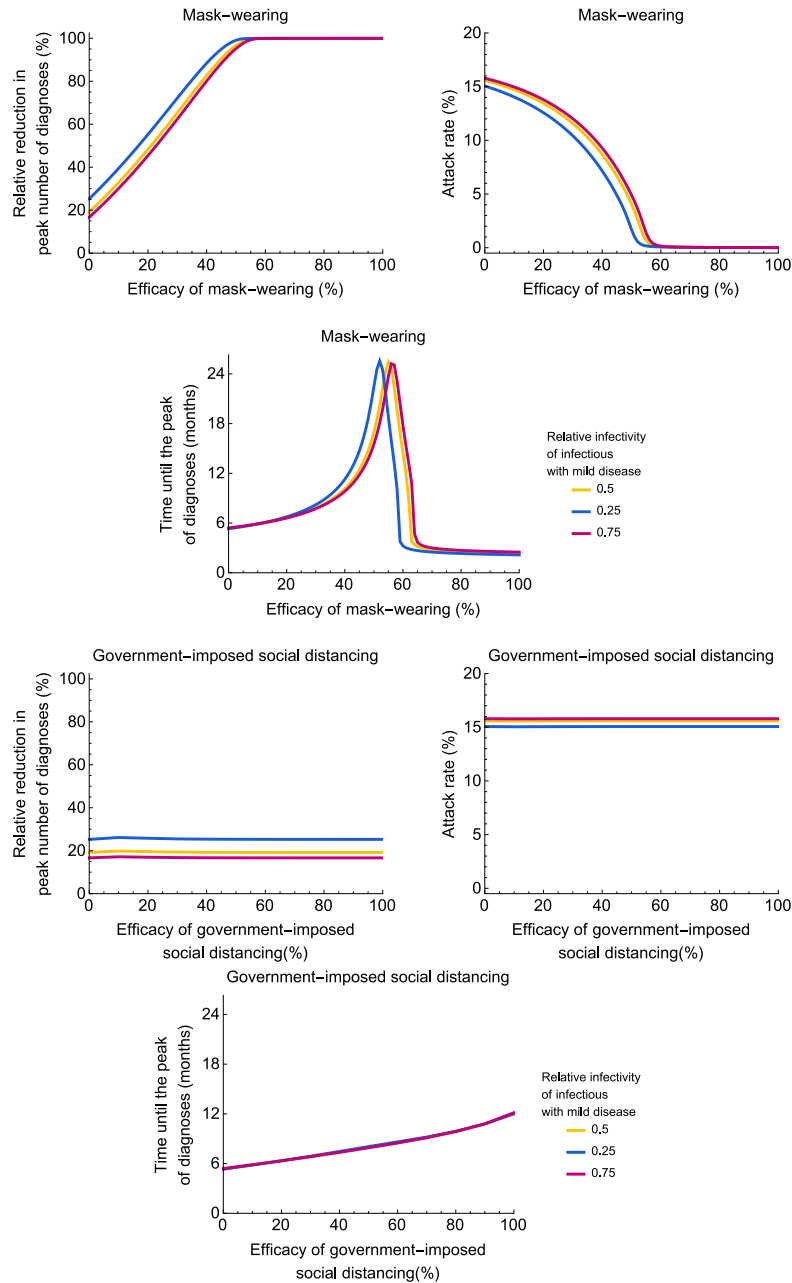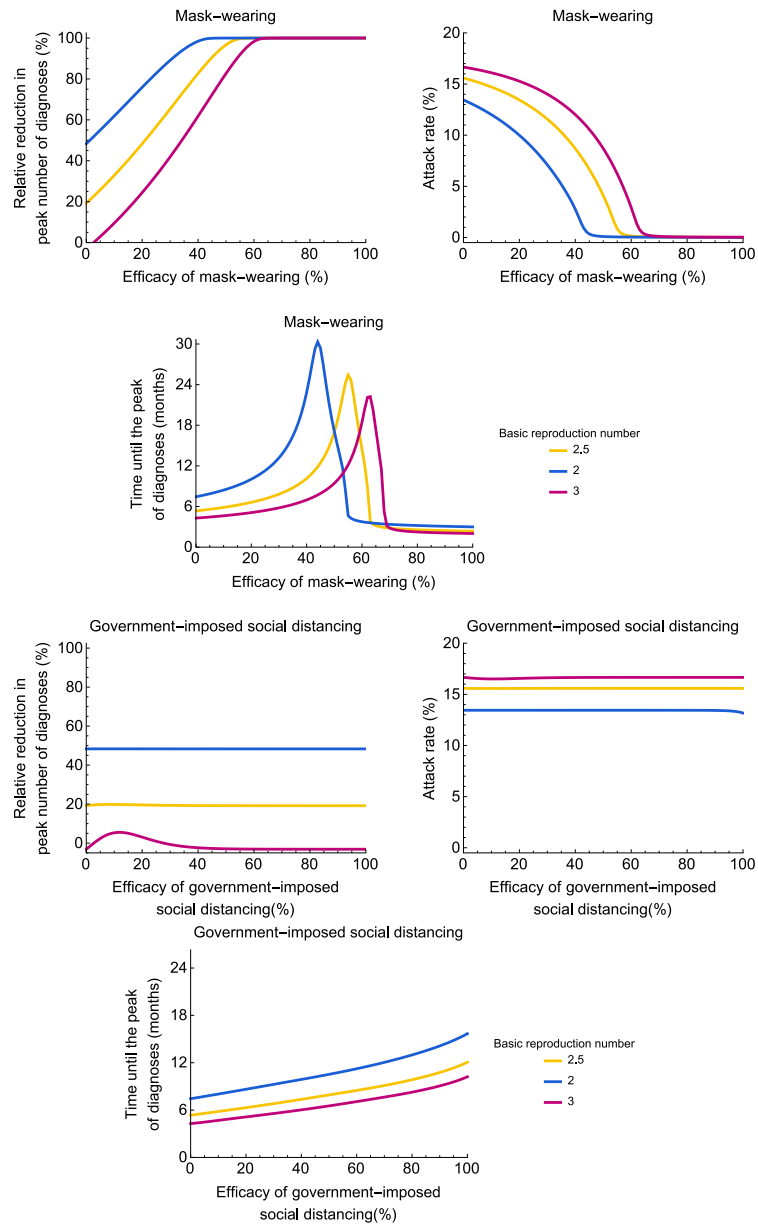
Sensitivity analyses of the baseline transmission model with respect to the relative infectivity of infectious individuals with mild disease.
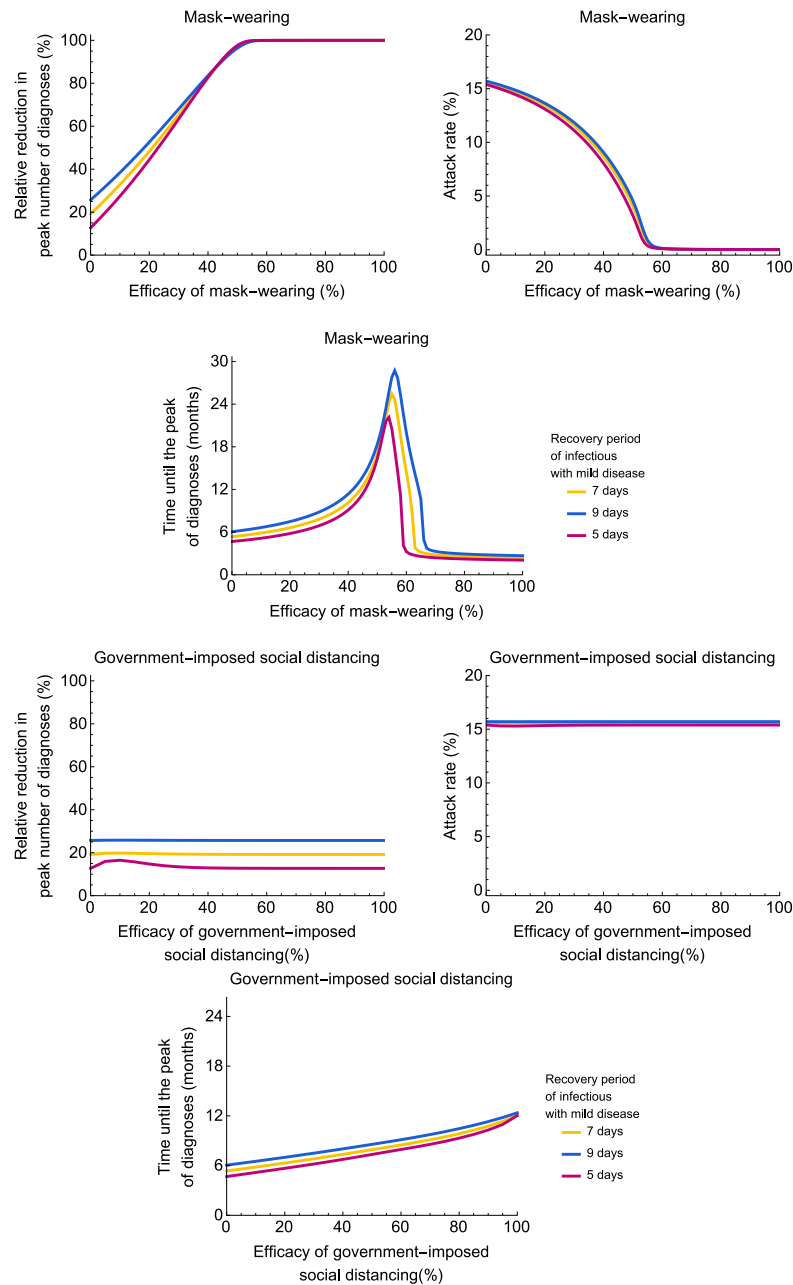
Sensitivity analyses of the baseline transmission model with respect to the recovery period of infectious individuals with mild disease.

Sensitivity analyses of the baseline transmission model with respect to the delay from onset of infectiousness to diagnosis for infectious individuals with severe disease.

Sensitivity analyses of the baseline transmission model with respect to the basic reproduction number.
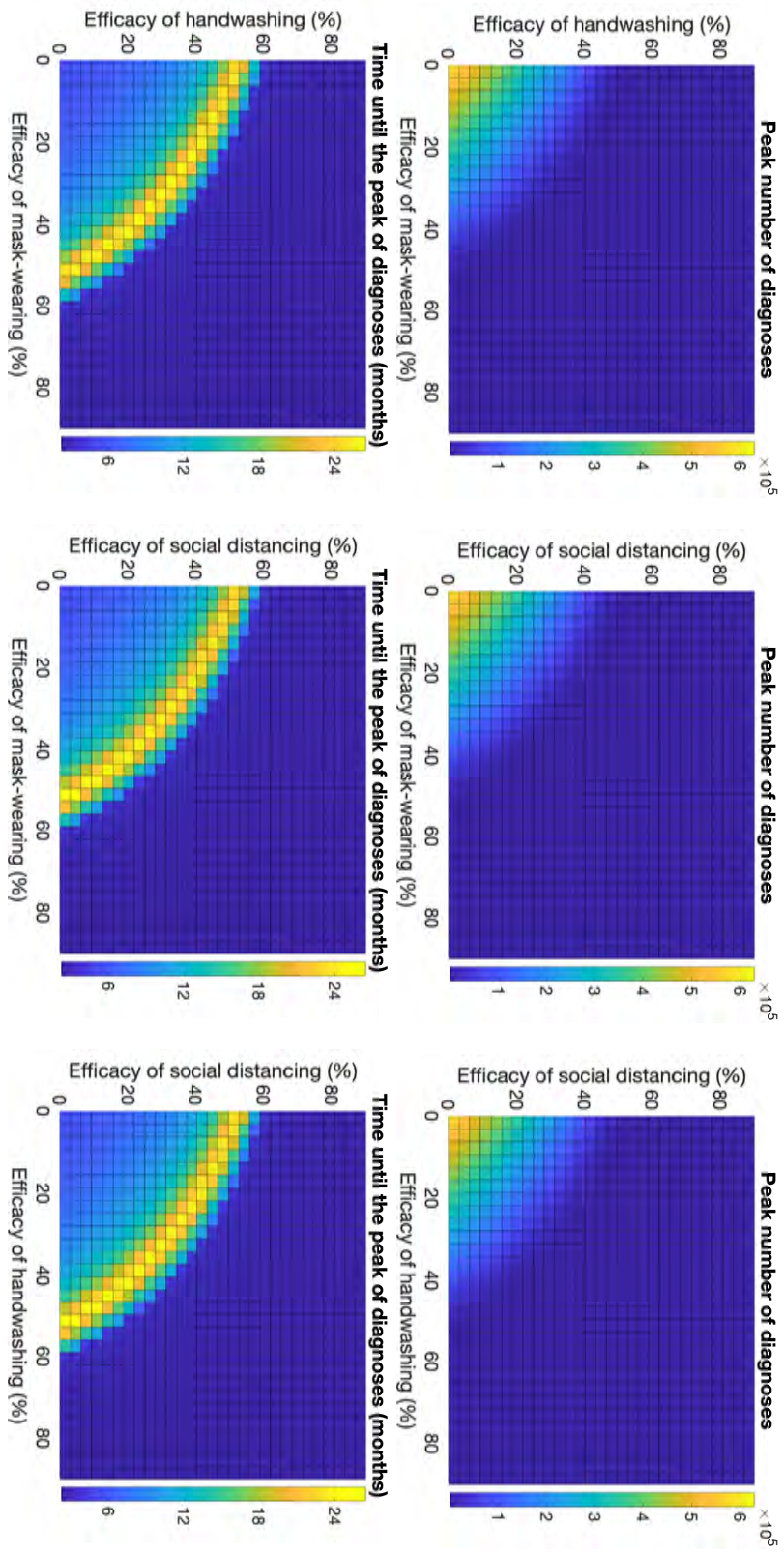
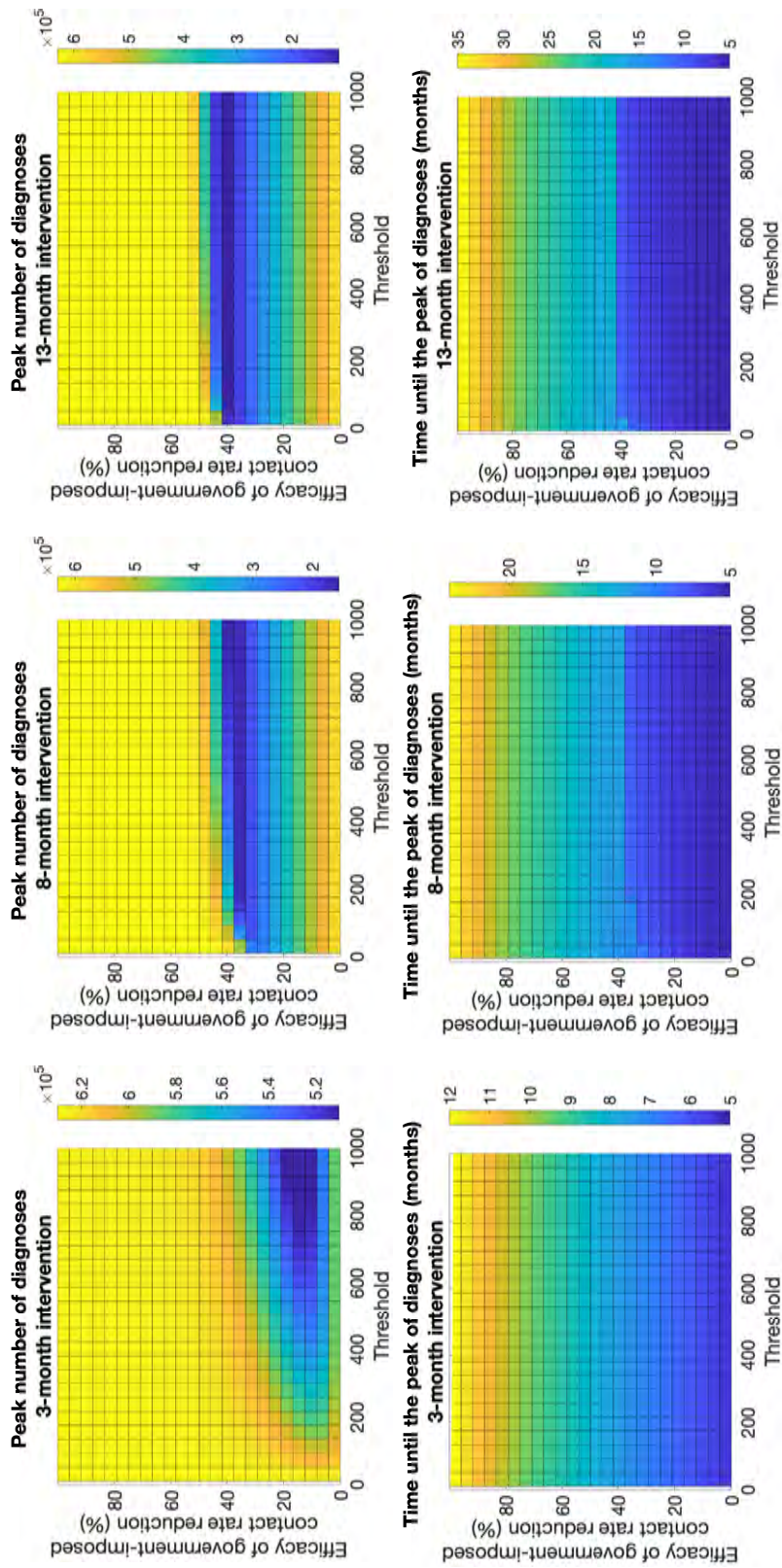**Figure S1: Impact of combinations of self-imposed prevention measures.**

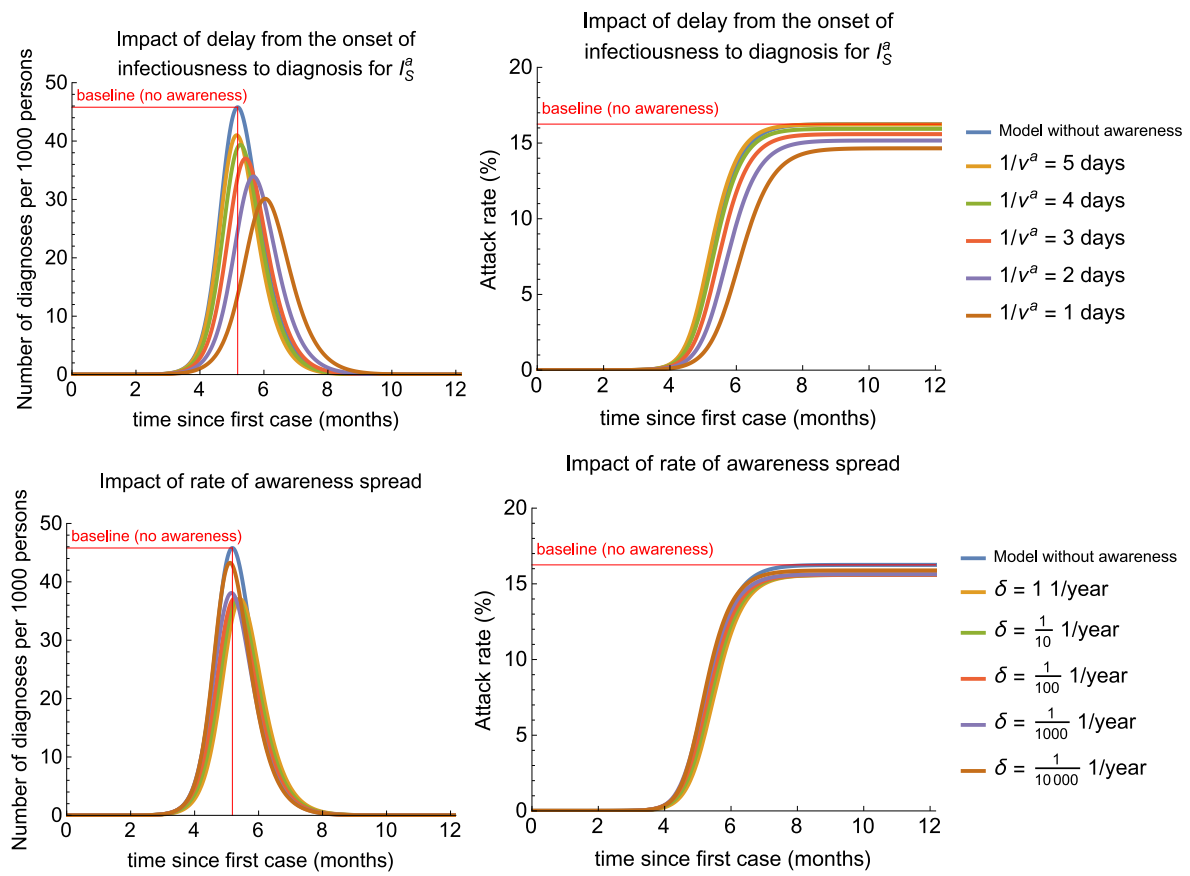**Figure S2**: Impact of government-imposed social distancing intervention.

Figure S3: **Sensitivity analyses of the transmission model with disease awareness. Part 1.**
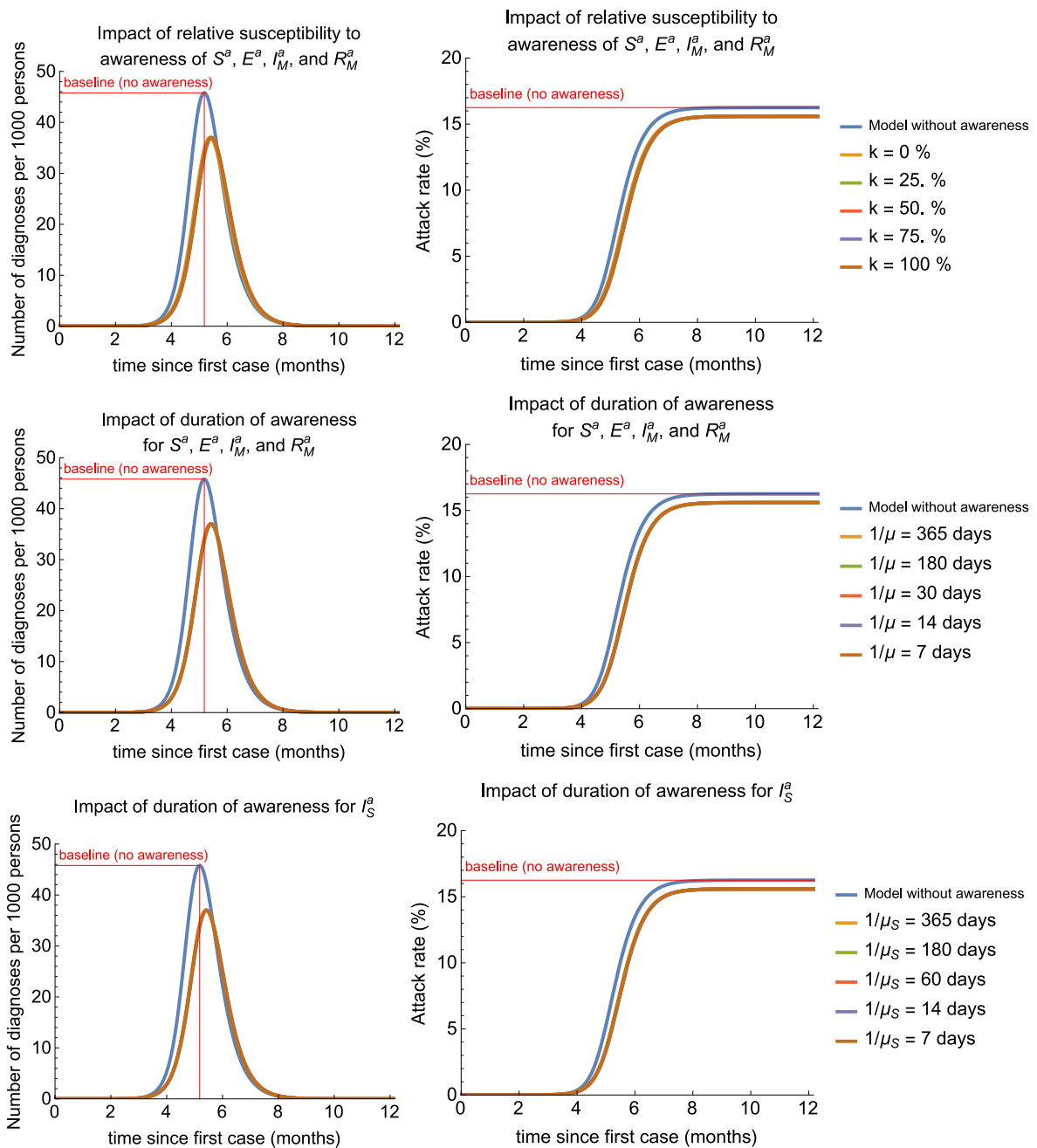
Figure S3: **Sensitivity analyses of the transmission model with disease awareness. Part 2.**

Chapter 2

The potential impact of intensified community hand hygiene interventions on respiratory tract infections: A modeling study

Thi Mui Pham, Yin Mo, Ben S. Cooper

Abstract

Background: Hand hygiene is amongst the most fundamental and widely-used behavioural measures to reduce the person-to-person spread of human pathogens and its effectiveness as a community intervention is supported by evidence from randomised trials. However, a theoretical understanding of the relationship between hand hygiene frequency and change in risk of infection is lacking.

Methods and Findings: Using a simple model-based framework for understanding the determinants of hand hygiene effectiveness in preventing viral respiratory tract infections we show that a crucial, but overlooked, determinant of the relationship between hand hygiene frequency and risk of infection via indirect transmission is persistence of viable virus on hands. If persistence is short, as has been reported for influenza, hand-washing needs to be performed very frequently or immediately after hand contamination to substantially reduce the probability of infection. When viable virus survival is longer (e.g., in the presence of mucus or for some enveloped viruses) less frequent hand washing can substantially reduce the infection probability. Immediate hand washing after contamination is consistently more effective than at fixed-time intervals.

Conclusions: Our study highlights that recommendations on hand hygiene should be tailored to persistence of viable virus on hands and that more detailed empirical investigations are needed to help optimise this key intervention.

Keywords: hand hygiene, respiratory infections, community, influenza, modelling

Introduction

Promotion of hand hygiene is a key public health intervention in preventing the spread of infectious diseases. Since the mid-1800s, when Ignaz Philip Semmelweis demonstrated that hand washing could dramatically reduce maternal mortality due to puerperal fever [1], hand hygiene has been the cornerstone of infection prevention and control policies. In hospital settings, hand hygiene has played a major role in successfully controlling hospital-acquired infections, especially those caused by methicillin-resistant *Staphylococcus aureus* [2]. In the community, there is evidence from randomized controlled trials that hand hygiene interventions can be effective in reducing both the risk of diarrhoeal disease [3] and respiratory tract infections [4–6].

Hand hygiene is simple, low-cost, minimally disruptive and, when widely adopted, may lead to substantial population-level effects [5, 7]. While randomized controlled trials of hand hygiene interventions in the community provide evidence that such interventions are effective in reducing the incidence of respiratory tract infections, reported effect sizes are highly variable [4, 6]. It is unclear to what extent this variability is explained by success in achieving substantial changes in hand hygiene behaviour in these trials. Understanding how the effectiveness of hand hygiene in reducing transmission scales with hand hygiene frequency is important for assessing the extent to which interventions that aim at achieving a large and sustained increase in community hand hygiene can contribute to infection suppression.

In this study, we took a theory-based approach and developed a simple mechanistic mathematical model to understand the relationships between the various components of respiratory tract infection transmission pathways involving hand contamination. We aimed to quantify the expected impact of different hand hygiene behaviours on risks of respiratory tract infections. Our work is motivated by published data on the survival of influenza A on human fingers. We therefore focus on viral respiratory tract infections but our model also applies to pathogens for which similar assumptions apply. Finally, we consider the implications of the outcomes of these analyses for the potential contribution of intensifying community hand hygiene to the suppression of respiratory tract infections.

Methods

Overview

We consider human pathogens where transmission is mediated by contaminated hands. We neglect direct droplet and aerosol transmission. Hands are assumed to become contaminated with infectious material via contact with contaminated surfaces or an infected person. In the absence of hand washing, hands do not remain contaminated indefinitely; instead, as has been shown experimentally, the probability of remaining contaminated and capable of transmitting infection declines over time (Figure 1, top panel) [8, 9]. If contaminated hands of a susceptible host make contact with the host's mucous membranes in the eyes, nose or mouth there is some probability of the host becoming infected. Effective hand washing interrupts this process by removing viable virus from the hands. An immediate consequence of this conceptualisation is that the time interval between
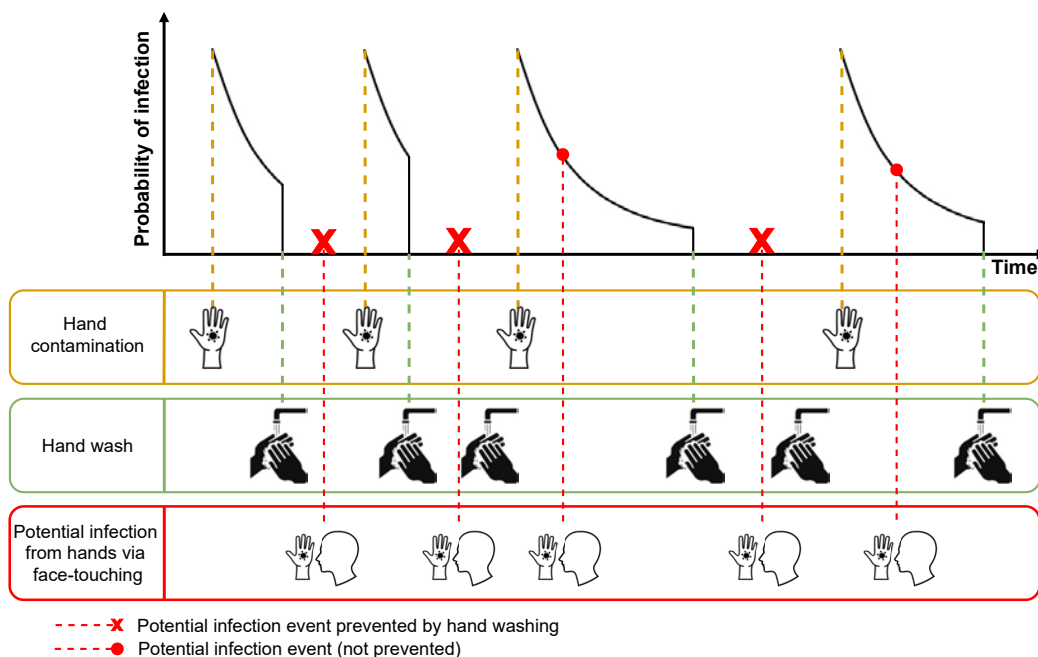

Figure 1: **Hand hygiene model.** Illustration of potential infection events from hands via face-touching, hand contamination events, and hand washing events. Hand contamination events cause a stepwise increase in the probability of infection resulting from face-touching events, which then decreases exponentially with time. Effective hand washing reduces the probability of infection to zero during subsequent face-touching if no further hand contamination events occur. An infection may occur between a hand contamination event and hand washing, depending on the probability of infection at the moment of face-touching.

the hands becoming contaminated and the potential transmission to the host can have a critical impact on how effective a given frequency of hand washing will be at interrupting transmission (Figure 2). Given a certain probability of infection, the time interval between hand contamination and transmission to the host's mucosa tends to be longer if pathogen persistence on hands is long and vice versa. If this time interval is relatively long, i.e., the virus survives on hands for a long time, regular effective hand hygiene will have a high chance of blocking potential transmission events (red squares in Figure 2 panel A) in the absence of hand hygiene. In contrast, if this time interval is short, i.e. the pathogen persists for only a short amount of time, much more frequent hand hygiene will be needed to block an appreciable proportion of transmission events (Figure 2 panel B).
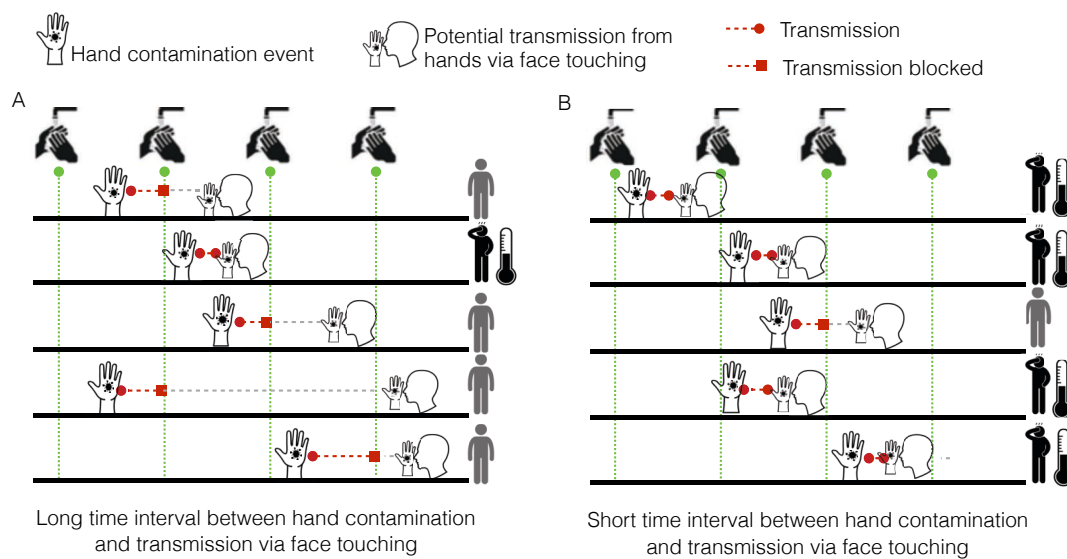
Figure 2: **Long versus short time interval between hand contamination and infection with regular hand washing.** A) When there are long time intervals between hand contamination and potential infection from hands via face-touching, hand washing can block many infection events and substantially reduce the risk of infection. B) When there are short time intervals between hand contamination and face-touching, it is likely that hand washing can disrupt only a few infections.

Hand hygiene scenarios

We explored the effect of hand hygiene on the probability of infection and considered two hand washing schemes that are distinguished by different timings of hand washing:

1. fixed-time hand washing (uniformly at fixed time intervals)

2. event-prompted hand washing (with a delay after hand contamination events).

Mathematical model

We assumed that hands of susceptible individuals become contaminated at random. These contamination events are assumed to occur independently of each other, and to follow a Poisson distribution with a mean of λ_c events per hour. Once contaminated, we assumed that in the absence of hand washing there is a constant rate at which hands get decontaminated. Thus, the probability of the virus persisting on hands at time t after contamination, $P(t)$, is assumed to decay exponentially with a half-life of $T_{1/2}$. This is consistent with experimental data for influenza A [9]. We further assumed that, in a given time interval $[0, T]$, individuals touch their face at random times $t_1, \ldots, t_F$ leading to potential self-infection events that are assumed to occur independently of each other, and follow a Poisson distribution with a mean of λ_f events per hour. The probability that a single face-touching contact with contaminated hands actually leads to transmission is denoted by ϵ. The force of infection that a susceptible individual at time t becomes infected is $\lambda_{\text{inf}}(t) = \epsilon P(t)$. The cumulative probability of infection over a given time period T is then given by: $1 - e^{-\sum_{i=1}^{F} \epsilon P(t_i)}$. We assumed that when hand washing is performed after the last hand contamination event and before a face-touching event at time t_i, the respective probability of pathogen persistence $P(t_i)$ is reduced to zero. Thus, hand washing is assumed to remove all virus on contaminated hands completely after one wash, regardless of the number of hand contamination events that took place between hand washing events. A more detailed mathematical description of the model is included in the supplementary material (pp. 15).

Parameters

When available, parameter estimates were obtained from the literature. Otherwise, we performed sensitivity analyses where parameters were varied within plausible ranges (see

Table 1).

The probability of transmission per face-touching event, ϵ, was constrained to meet a fixed probability of infection. In our main analysis, we assumed a cumulative probability of infection of 10% over a time period of 12 hours. This is roughly based on secondary attack rates for Influenza, influenza-like-illnesses and acute respiratory illness in household studies [10–13]. In sensitivity analyses, we examine the results for cumulative probabilities of infection of 30% and 50%.

In the fixed-time hand washing scheme, we varied time intervals between hand washing to be 5 min to 6 hours. For event-prompted hand washing, the delay of hand washing after hand contamination events was varied from 1 min to 6 hours.

There is little published data on the rate of hand contamination events susceptible individuals are exposed to when in contact with infected individuals who are shedding respiratory viruses. In a direct observation study conducted by Zhang et al [14], surface touching behaviour in a graduate student office was recorded. Approximately 112 surface touches per hour were registered. Another study by Boone et al [15] found that the influenza virus was detected on 53% of commonly touched surfaces in homes with infected children (using reverse transcriptase-polymerase chain reaction (RT-PCR)). Informed by these values, we took 60 events per hour as the upper bound for the rate of hand contamination events λ_c. We chose 1 event per hour as the lower bound. In our main analyses, we used a rate of 4 hand contamination events per hour.

In [9], the survival of influenza A on human fingers was experimentally investigated. We fitted exponential decay curves to these results in order to determine the half-life of the probability of persistence of H3N2 for two viral volumes, 2 μL and 30 μL (Table 1 and supplementary material). We use these values as examples for the half-life of the probability of pathogen persistence. In addition, we vary the half-life of the probability of persistence from 1 to 60 min in our analysis.

Model analyses and outcomes

The model output is the cumulative probability of a susceptible person becoming infected in twelve hours and we will refer to it subsequently as simply the probability of infection. We investigated the impact of hand washing on the probability of infection for different hand contamination rates. In addition, we compared the two hand washing schemes (fixed-time vs. event-prompted) to find the optimal hand washing strategy that will lead

Table 1: **Parameter values**

		Value	Source
Time period		12 hours	Assumed
Rate of infection events through face-touching (per hour)	λ_f	10	[16]*
Cumulative probability of infection (in 12 hours)		10 % (30 %, 50 %)†	Assumed
Probability of transmission per face-touching event	ϵ	Computed from cumulative probability of infection	
Rate of hand contamination events (per hour)	λ_c	4 hour⁻¹ (1–60 hour⁻¹)†	[14, 15]
Time between hand washing events (fixed-time)	t_F	5, 15, 30 min, 1 hour, 2, 6 hours	Assumed
Delay of hand washing after hand contamination events	t_D	1, 5, 15, 45 min, 1 hour, 2, 6 hours	Assumed
Half-life of virus persistence	$T_{1/2}$	1–60 min	Varied
Half-life of H3N2 persistence for 2 μL of viral inoculum		5.4 min	[9]
Half-life of H3N2 persistene for 30 μL of viral inoculum		36.1 min	[9]

* Mean face-touching frequency involving mucous membranes (eyes, mouth, nose)
† Sensitivity analyses

to the greatest reduction of the probability of infection. The model was implemented in R version 3.6.3 [17]. The code reproducing the results of this study is available at `https://github.com/tm-pham/covid-19_handhygiene`.

Results

Impact of half-life of pathogen persistence on probability of infection

Viral persistence on hands plays a key role on the effect of increasing hand hygiene frequency. The longer the virus survives on the hands, the larger the impact of increasing hand washing uptake on the probability of infection. For example, when the half-life of viral persistence is 1 min, hand washing every 15 min reduces the probability of infection from 10 % to 9.2 % (Figure 3 A). When the half-lives increase to 5.4 min and 36.1 min (equivalent to the half-lives of H3N2 persistence of 2 μL and 30 μL viral inoculum, respectively), the same hand washing frequency decreases the probability of infection to 6.9 % and to 4.6 %, respectively. Consequently, fewer hand washes are necessary to reduce the probability of infection by 50 % for long compared to short durations of viral persistence (see Figure S2). This observation can be explained by the fact that the shorter the virus persists on hands, the shorter the intervals between hand contamination and transmission events tend to be (with a higher transmission probability per contact needed for the same cumulative probability of infection, see Figure S4) and, therefore, the less likely hand washing is able to interrupt infection events. Figure S3 shows that delay between hand contamination and hand washing needed to prevent 50% of transmissions is shorter when the half-life of viral persistence on the hands is shorter, confirming the hypothesis that timely hand washing is especially crucial if the virus survives only a short time on

hands. Furthermore, the effect of hand washing on reducing the probability of infection plateaus with increasing duration of virus persistence (Figure 3). This can be attributed to the hand contamination rate, i.e. new events occur before the virus decays.
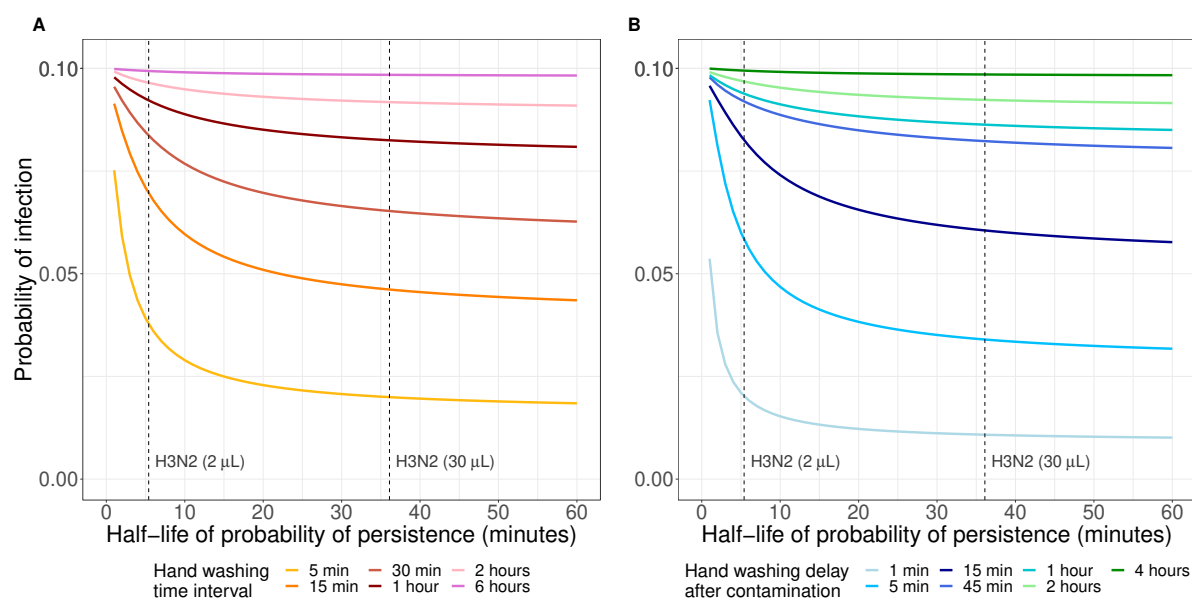

Figure 3: **Impact of half-life of viral persistence on probability of infection for different hand washing schemes and frequencies.** (A) Fixed-time hand washing (B) Event-prompted hand washing. In this graph, we assumed that a susceptible individual is exposed to a baseline probability of infection of 10% if no hand washing is performed within the time period of twelve hours. The dashed lines represent the half-life of viral persistence for H3N2 inoculum volumes of 2 μL and 30 μL (calculated from [9]). For each half-life value, the probability of transmission per face-touching event ϵ was determined for a probability of infection of 10% in the case of no hand washing. The probability of infection for the different hand washing frequencies/delays was then computed using this ϵ value. Hand contamination events are assumed to occur on average 4 times per hour. Sensitivity analyses with different values for baseline probabilities of infection as well as the half-life calculations are presented in the supplementary material.

Comparison of hand washing schemes

The second notable finding from the model is that event-prompted hand washing is more effective than fixed-time hand washing in reducing the probability of infection. We illustrate this in Figure 4 by comparing both schemes using four different hand washing frequencies/delays, each with approximately the same average number of hand washing events performed per hour. For example, hand washing regularly every fifteen minutes is compared to event-prompted hand washing one minute after each hand contamination

event (set at four per hour). If the half-life of viral persistence is similar to 2 μL of H3N2 inoculum ($T_{1/2} = 5.4\,\text{min}$), the baseline probability of infection of 10%(no hand washing) is reduced to about 6% and 2% when hand washing is performed every 15 min and one minute after hand contamination events, respectively. The differences between the two hand washing schemes are less pronounced if hand washing is performed less frequently or with a longer delay after hand contamination events since the two hand washing schemes become more similar. It follows that delays between hand contamination and hand washing decrease the effect of hand washing on reducing the probability of infection.

Comparison of hand washing schemes

The second notable finding from the model is that event-prompted hand washing is more effective than fixed-time hand washing in reducing the probability of infection. We illustrate this in Figure 4 by comparing both schemes using four different hand washing frequencies/delays, each with approximately the same average number of hand washing events performed per hour. For example, hand washing regularly every fifteen minutes is compared to event-prompted hand washing one minute after each hand contamination event (set at four per hour). If the half-life of viral persistence is similar to 2 μL of H3N2 inoculum ($T_{1/2} = 5.4\,\text{min}$), the baseline probability of infection of 10%(no hand washing) is reduced to about 6% and 2% when hand washing is performed every 15 min and one minute after hand contamination events, respectively. The differences between the two hand washing schemes are less pronounced if hand washing is performed less frequently or with a longer delay after hand contamination events since the two hand washing schemes become more similar. It follows that delays between hand contamination and hand washing decrease the effect of hand washing on reducing the probability of infection.

Another important parameter that affects the effect of hand hygiene is the hand contamination rate. Figure 5 shows the increase in hand hygiene frequency required to halve the probability of infection from 10% (no hand wash) to 5%. When the hand contamination rate is relatively low (i.e., less than 10 contamination events per hour), fewer hand washes are needed to reduce the probability of infection if hand washing is event-prompted. In addition, the longer the virus persists on hands, the smaller the number of hand washing events are necessary to reduce the probability of infection.
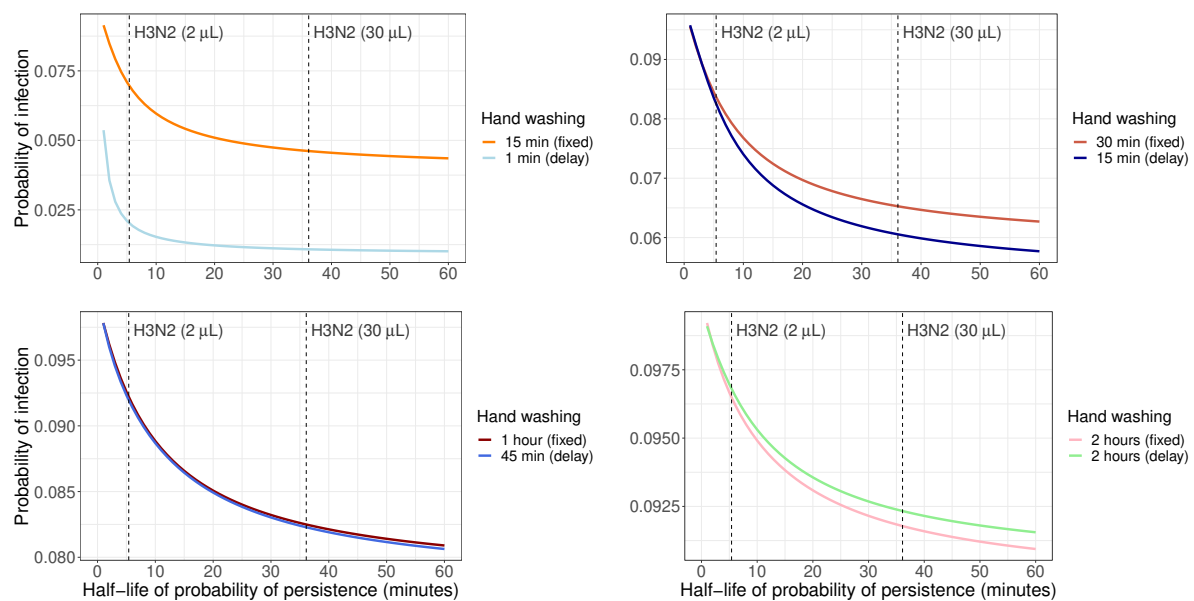
Figure 4: **Comparison of the impact of the two hand washing schemes on the cumulative probability of infection.** Hand washing at fixed time intervals and event-prompted hand washing (with a time delay) with similar average number of hand washing events per hour are compared for a hand contamination rate of $\lambda_c = 4$ hour^{-1}. A baseline probability of infection of 10% is assumed when there is no hand washing. The dashed lines represent the half-life values of H3N2 persistence for 2 μL and 30 μL inoculum volumes [9].

This effect is less pronounced for event-prompted than for time-fixed hand washing, re-emphasizing the finding that when hand contamination events occur very frequently, hand washing would need to be very frequent to have a substantial impact on reducing the probability of infection (e.g., at least five times per hour to prevent 50% of transmission in the case of a half-life of 36.1 min). In this case, susceptible individuals are exposed to a continuous risk of hand contamination and hand washing has only a limited impact on reducing the risk of infection.

Our qualitative conclusions do not change with respect to different baseline probabilities of infection and hand contamination rates (see supplementary material for sensitivity analyses with respect to these parameters).
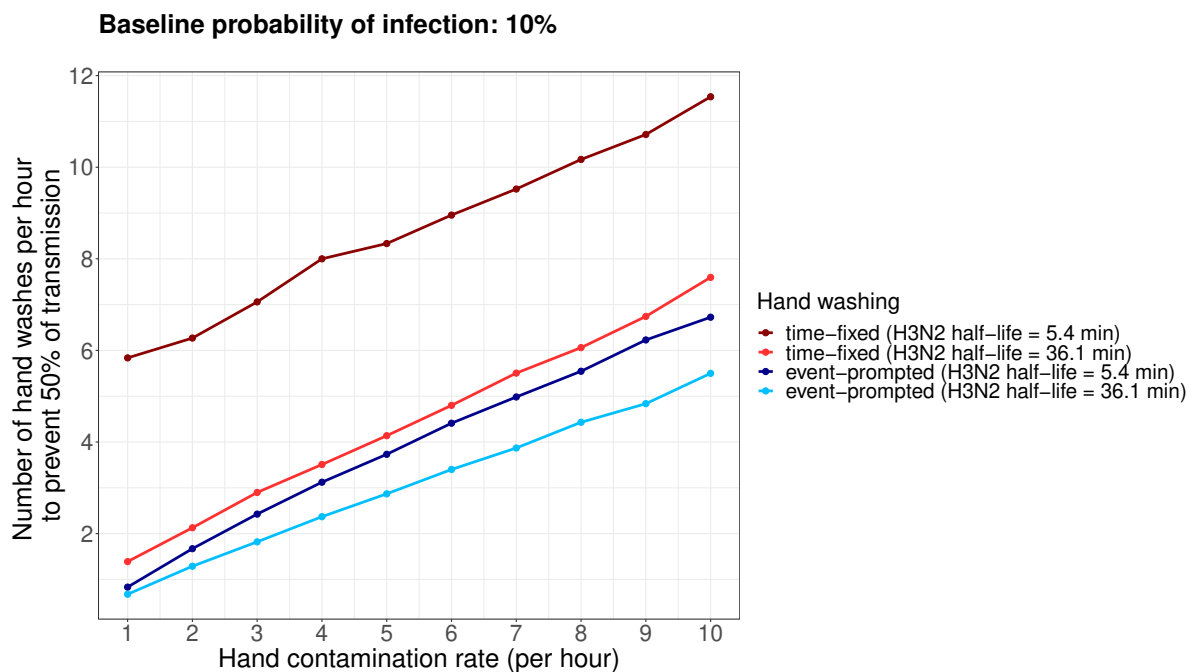

Baseline probability of infection: 10%

Figure 5: **Number of hand washes necessary to prevent 50% of transmissions.** For a baseline probability of infection of 10%, the number of hand washing events necessary to reduce the probability of infection to 5% was computed for time-fixed and event-prompted hand washing and a range of hand contamination rates. We used the half-life of H3N2 persistence for viral inoculum volumes of 2 μL and 30 μL (calculated from [9]).

Discussion

Our study provides new insights into factors that affect the effectiveness of hand hygiene behaviour in reducing the probability of infection. Firstly, we found that the shorter the virus survives on hands, the less effective increasing hand washing frequency is in reducing infection. The logic behind this is that when the virus dies off quickly before hand washing is performed, the time intervals between hand contamination and transmission tend to be shorter and the respective transmission probability per contact needs to be higher for the same cumulative probability of infection. Secondly, the contact frequency with contaminated surfaces is crucial for the effect of hand washing. The more often hands become contaminated, the more frequently hands need to be washed to reduce infection risk. Lastly, when hands are not constantly contaminated, event-prompted hand washing is more efficient than fixed-time hand washing given the same hand washing frequency. This is because delays in hand washing after contamination of hands in fixed-time compared to event-prompted hand washing tend to be longer, and, during this delay, susceptible hosts may become infected through face-touching.

These findings provide additional insights into the modest and heterogeneous effects of hand hygiene reported by hand hygiene trials aimed at reducing respiratory tract infections in the community [4, 6, 18], and also provide pointers to potentially more effective hand hygiene interventions. These trials are challenging to conduct due to the difficulties in implementing behaviour change, including poor adherence to hand washing recommendations [19], and loss to follow-up [20, 21]. However, given the low-cost and minimally-disruptive nature of the intervention we believe there would be considerable value in building on this experimental work and the theory outlined above to develop improved hand-hygiene interventions. This could offer considerable public health benefit both in interpandemic and pandemic periods.

Since the hand contamination rate directly impacts the effect of hand hygiene, specific hand hygiene advice should cater for different situations where surface contamination differs markedly. For example, contacts in the community and in a household with an infectious person would likely result in very different hand contamination rates. In the first case, where hand contamination events occur at a moderate rate, hand washing needs to be performed frequently or immediately after hand contamination events in order to substantially reduce the probability of infection. While individuals may not always be

aware of all hand contamination events, event-prompted hand washing can be facilitated by installing or providing hand sanitisers in public areas with high-touch surface areas, such as public transportation and supermarkets, to reduce the delay in hand cleansing. Furthermore, in the second case, where hands become contaminated very frequently, a substantial reduction in the probability of infection is unlikely to be attained unless hand washing frequency is increased drastically, i.e., every one to five minutes. Because hand washing at such a high rate is not practical, the recommendation in this scenario is to regularly clean the environment and/or isolate infected individuals to reduce hand contamination events.

We performed sensitivity analyses with varying parameter values and distributions to ensure our conclusions are robust on a qualitative level. Nevertheless, our results have several limitations. Firstly, we specifically modelled indirect transmission routes via hands and did not consider direct droplet and aerosol transmission. To date, there is little known about the relative importance of the various transmission routes of respiratory pathogens [22]. When other routes are considered, the effect of hand hygiene will be reduced. Secondly, there is limited literature on many parameters used in the model, which prevents us from making more precise quantitative conclusions. These include the probability of infection with contaminated hands, the survival of pathogens on contaminated hands and infective dose. Furthermore, we modelled all infection events with the same rate of decay, i.e., the same probability of pathogen persistence on the hands. In reality, hand contamination events are likely to be heterogeneous with small droplets persisting only a short amount of time and heavy contamination with mucus decaying at a slower rate. In addition, we specifically focus on viral respiratory infections and assumed an exponential decay for the probability of viral persistence. While our model can be applied to all pathogens where hand hygiene is relevant for reducing respiratory tract infections, our results are only applicable for pathogens with a similar persistence behaviour. However, our model can be easily adapted if information on the persistence behaviour of specific pathogens is available.

Conclusion

To conclude, in this study we highlight the important considerations in hand hygiene behaviour to improve its effect in stopping the community spread of respiratory tract

infections. Recommendations on hand hygiene should be tailored to the expected hand contamination rate and the half-life of virus persistence on hands.

Data accessibility

Data and code used in the analysis are publicly available:

`https://github.com/tm-pham/covid-19_handhygiene`

Author contributions

TMP, BC, and MY conceptualised and developed the model. TMP wrote the model code and performed the model analysis. TMP, MY, and BC interpreted the results. MY, TMP, and BC performed the literature search. TMP and MY drafted the manuscript. All authors read, reviewed, and approved the final manuscript, and all authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Funding

Acknowledgements

References

[1] Best, M and Neuhauser, D. Ignaz Semmelweis and the birth of infection control. In: *BMJ Quality \& Safety* 13.3 (2004), pp. 233–234.

[2] Pittet, D and Boyce, JM. Hand hygiene and patient care: pursuing the Semmelweis legacy. In: *The Lancet Infectious Diseases* 1 (2001), pp. 9–20.

[3] Curtis, V and Cairncross, S. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. In: *The Lancet infectious diseases* 3.5 (2003), pp. 275–281.

[4] Wong, V, Cowling, B, and Aiello, A. Hand hygiene and risk of influenza virus infections in the community: a systematic review and meta-analysis. In: *Epidemiology and infection* 142.5 (2014), pp. 922–932. DOI: `10.1017/S095026881400003X`.

[5] Aiello, AE, Coulborn, RM, Perez, V, et al. Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis. In: *American journal of public health* 98.8 (2008), pp. 1372–1381.

[6] Warren-Gash, C, Fragaszy, E, and Hayward, AC. Hand hygiene to reduce community transmission of influenza and acute respiratory tract infection: a systematic review. In: *Influenza and other respiratory viruses* 7.5 (2013), pp. 738–749.

[7] Cowling, BJ, Chan, KH, Fang, VJ, et al. Facemasks and hand hygiene to prevent influenza transmission in households: a cluster randomized trial. In: *Annals of internal medicine* 151.7 (2009), pp. 437–446.

[8] Bean, B, Moore, BM, Sterner, B, et al. Survival of influenza viruses on environmental surfaces. In: *Journal of Infectious Diseases* 146.1 (1982), pp. 47–51.

[9] Thomas, Y, Boquete-Suter, P, Koch, D, et al. Survival of influenza virus on human fingers. In: *Clinical Microbiology and Infection* 20.1 (2014), O58–O64. DOI: `10.1111/1469-0691.12324`.

[10] Azman, AS, Stark, JH, Althouse, BM, et al. Household transmission of influenza A and B in a school-based study of non-pharmaceutical interventions. In: *Epidemics* 5.4 (Dec. 2013), pp. 181–186. DOI: `10.1016/j.epidem.2013.09.001`.

[11] Carcione, D, Giele, CM, Goggin, LS, et al. Secondary attack rate of pandemic influenza A(H1N1)2009 in Western Australian households, 29 May-7 August 2009. In: *Eurosurveillance* 16.3 (Jan. 2011), p. 19765. DOI: `10.2807/ese.16.03.19765-en`.

[12] Cauchemez, S, Ferguson, NM, Fox, A, et al. Determinants of Influenza Transmission in South East Asia: Insights from a Household Cohort Study in Vietnam. In: *PLoS Pathogens* 10.8 (Aug. 2014), e1004310. DOI: `10.1371/journal.ppat.1004310`.

[13] Savage, R, Whelan, M, Johnson, I, et al. Assessing secondary attack rates among household contacts at the beginning of the influenza A (H1N1) pandemic in Ontario, Canada, April-June 2009: A prospective, observational study. In: *BMC Public Health* 11.1 (Apr. 2011), pp. 1–8. DOI: `10.1186/1471-2458-11-234`.

[14] Zhang, N and Li, Y. Transmission of Influenza A in a Student Office Based on Realistic Person-to-Person Contact

and Surface Touch Behaviour. In: *International journal of environmental research and public health* 15.8 (Aug. 2018), p. 1699. DOI: `10 . 3390 / ijerph15081699`.

[15] Boone, SA and Gerba, CP. The occurrence of influenza A virus on household and day care center fomites. In: *Journal of Infection* 51.2 (2005), pp. 103–109. DOI: `10.1016/j.jinf.2004.09.011`.

[16] Kwok, YLA, Gralton, J, and McLaws, ML. Face touching: A frequent habit that has implications for hand hygiene. English. In: *American Journal of Infection Control* 43.2 (Feb. 2015), pp. 112–114. DOI: `10.1016/j.ajic.2014.10.015`.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. 2020. `https://www.r-project.org/`.

[18] Willmott, M, Nicholson, A, Busse, H, et al. Effectiveness of hand hygiene interventions in reducing illness absence among children in educational settings: a systematic review and meta-analysis.

In: *Archives of disease in childhood* 101.1 (2016), pp. 42–50.

[19] Sandora, TJ, Taveras, EM, Shih, MC, et al. A randomized, controlled trial of a multifaceted intervention including alcohol-based hand sanitizer and hand-hygiene education to reduce illness transmission in the home. In: *Pediatrics* 116.3 (2005), pp. 587–594.

[20] White, CG, Shinder, FS, Shinder, AL, et al. Reduction of illness absenteeism in elementary schools using an alcohol-free instant hand sanitizer. In: *The Journal of School Nursing* 17.5 (2001), pp. 248–265.

[21] Zomer, TP, Erasmus, V, Looman, CW, et al. A hand hygiene intervention to reduce infections in child daycare: a randomized controlled trial. In: *Epidemiology \& Infection* 143.12 (2015), pp. 2494–2502.

[22] Kutter, JS, Spronken, MI, Fraaij, PL, et al. Transmission routes of respiratory viruses among humans. In: *Current opinion in virology* 28 (2018), pp. 142–151.

Supplementary material

Table of Contents

S1 Model

We assumed that hands of susceptible individuals get contaminated at random. These contamination events are assumed to occur independently of each other, and follow a Poisson distribution with a mean of λ_c events per hour. The probability of the virus to persist on hands at time t after contamination, $P(t)$, is assumed to decay exponentially with a half-life of $T_{1/2}$. This is consistent with experimental data for influenza A [9]. We further assumed that, in a given time interval $[0, T]$ individuals touch their face at random times $t_1, \ldots, t_F$ leading to potential infection events that are assumed to occur independently of each other, and follow a Poisson distribution with a mean of λ_f events per hour. The probability that a single face-touching contact with contaminated hands actually leads to transmission is denoted by ϵ. Thus, the probability that a single face-touching contact leads to transmission accounting for the probability of virus persistence is $\epsilon P(t_i)$ The cumulative probability of infection over the time period T is given by:

$$1 - e^{-\sum_{i=1}^{F} \epsilon P(t_i)}$$

We assume that when hand washing is performed after the last hand contamination event and before a face-touching event at time t_i, the respective probability of virus persistence $P(t_i)$ is reduced to zero.

Probability of viral persistence on contaminated hands

The decay of the probability of viral persistence on contaminated hands is modeled as an exponential decay with probability distribution:

$$f_{\text{decay}}(t) = \lambda_d e^{-\lambda_d t} \tag{2.1}$$

where λ_d is the decay constant. The probability that virus will die off within time t is given by the integral of the decay distribution function from 0 to t:

$$\int_0^t f_{\text{decay}}(t) = \int_0^t \lambda_d e^{-\lambda_d t}$$

The probability that the virus will persist at time t is one minus the probability that it will die off within the same period:

$$P(t) = 1 - \int_0^t f_{\text{decay}}(t)$$
$$= 1 - \int_0^t \lambda_d e^{-\lambda_d t}$$
$$= e^{-\lambda_d t}$$

The average survival time (or mean lifetime) is given by:

$$\tau = \frac{1}{\lambda_d} = \frac{T_{1/2}}{\ln 2} \tag{2.2}$$

S2 Estimation of virus half-life

We estimated the half-life of viral survival on contaminated hands using experiments conducted by Thomas et al,[9] where $2\,\mu$L and $30\,\mu$L of influenza A (H3N2) viral suspension mixed with respiratory secretions were deposited on finger tips. The half-lives were calculated using data from both the $2\,\mu$L ([9] Figure 2) and $30\,\mu$L ([9] Figure 3) H3N2 viral inoculum experiments with an exponential decay model:

$$n(t) = n_0 \cdot e^{-\lambda_d t}$$

where λ_d is the decay rate. The decaying quantity, $n(t)$, represents the number of fingers with recoverable infectious viral particles and is assumed to have an initial value of n_0 at time zero.

In the experiment with $2\,\mu$L inoculum, 18 contaminated fingers from six individuals were tested for the presence of infectious virus at 1, 3, 5, 15 and 30 min after initial contamination. Figure S1 depicts the data and the fitted curve for the $2\,\mu$L inoculum. The decay rate was estimated to be $\lambda_d^{(1)} \approx 0.1279$. The half-life is therefore given by $T_{1/2}^{(1)} = \frac{\ln(2)}{\lambda_d^{(1)}} = 5.4\,\text{min}$.

For $30\,\mu$L of viral inoculum, a total of 12 fingers were contaminated and the presence of H3N2 was tested after 15 min. We estimated the half-life by using these

two data points (see Figure 3 in [9]). Thus, $\lambda_d^{(2)} = -\frac{\ln(9/12)}{15} \approx 0.0192$. Therefore, $T_{1/2}^{(2)} = \frac{\ln(2)}{\lambda_d^{(2)}} = 36.1\,\text{min}$.
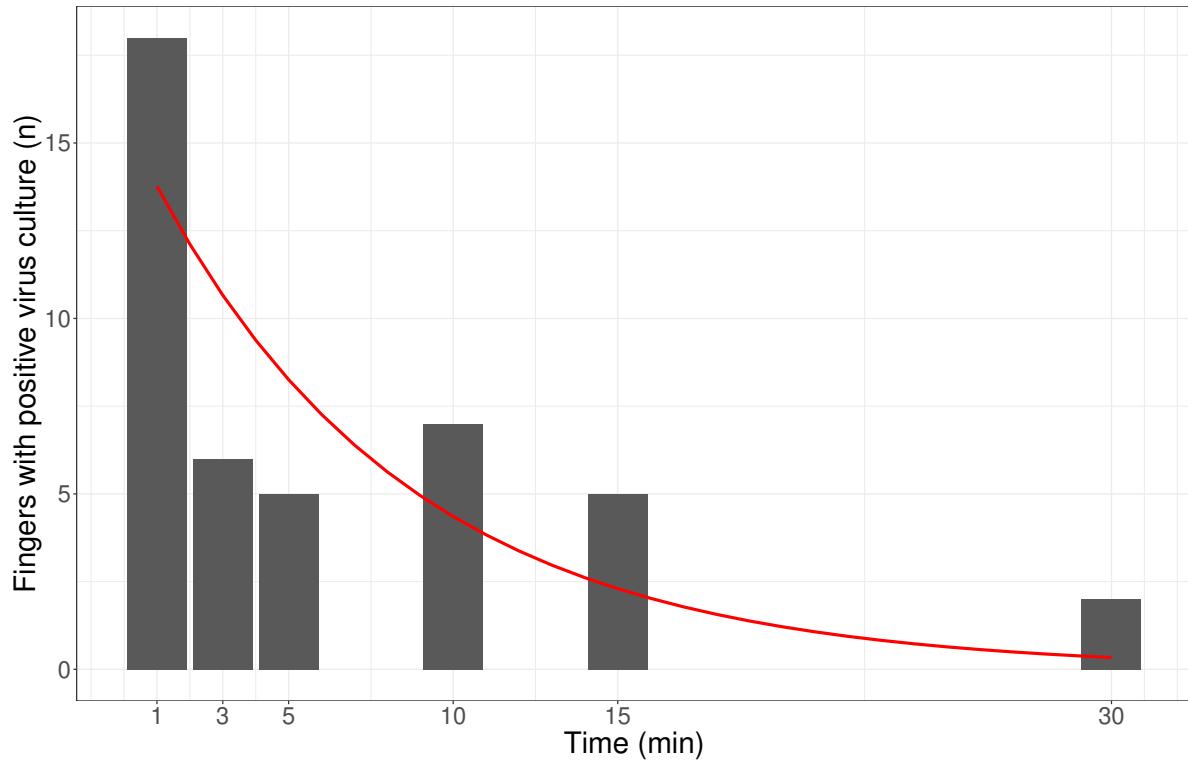
Figure S1: **Influenza A(H3N2) virus survival on fingers over time.** Data was retrieved from [9]. A 2 μL drop of influenza A (H3N2) viral suspension mixed with respiratory secretions was deposited on fingertips. Bars represent the absolute number of fingers from which infectious virus was recovered. The red line represents the exponential decay curve $n(t) = 15.65e^{-0.1279t}$ fitted to this data .

S3 Hand washing and half-life of virus persistence

The shorter the half-life of virus persistence, the higher the frequency of hand washing necessary in order to prevent 50% of infections (see Figure S2). In addition, the time intervals between hand contamination and hand washes have to be shorter in order to prevent 50% of the infections (see Figure S3.
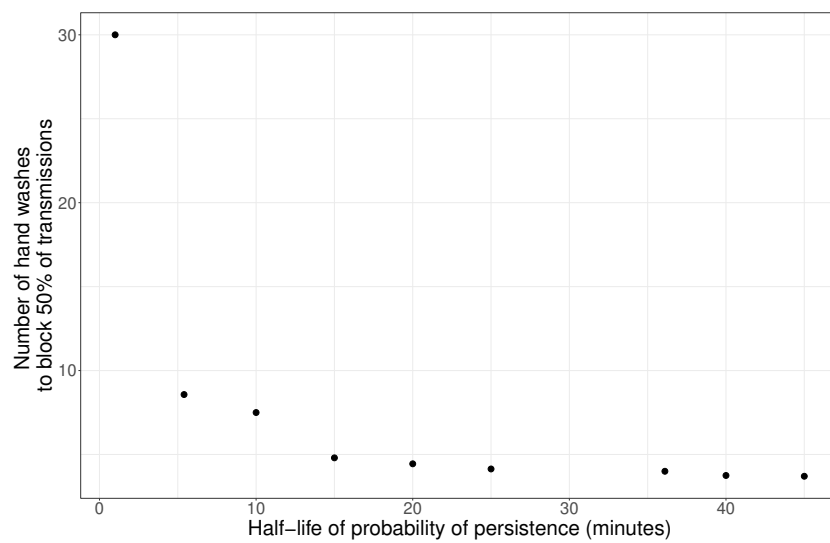
Figure S2: **Number of fixed-time hand washes necessary to prevent 50% of transmissions.** For each half-life value of virus persistence, the number of hand washes that is necessary to prevent 50% of transmission was computed for a baseline probability of infection of 10%. Hand contamination events are assumed to occur on average 4 times per hour.

Figure S3: **Cumulative time between hand contamination events and fixed-time hand washes to prevent 50% of transmissions.** For each half-life value of virus persistence, the cumulative time between hand contamination events and hand washes for preventing 50% of transmission was computed for a baseline probability of infection of 10%. Hand contamination events are assumed to occur on average 4 times per hour.
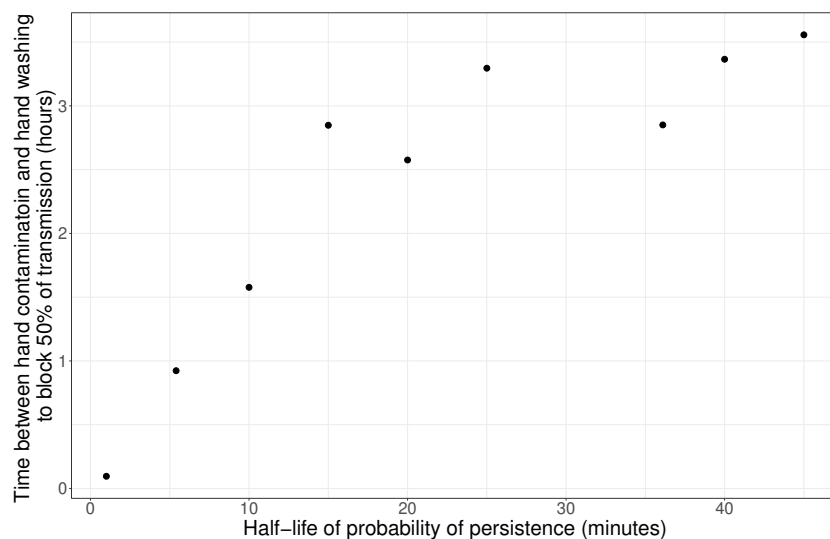
S4 Transmission probability per contact and half-life of virus persistence

Figure S4 shows that the shorter the virus persists on hands, the higher the probability of transmission per face-touching contact has to be if the cumulative probability of infection is assumed to be fixed.
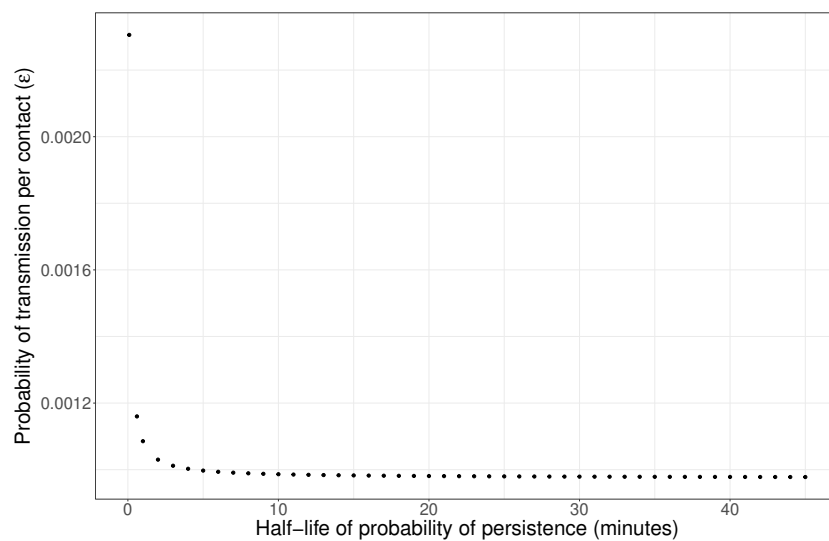
Figure S4: **Probability of transmission per face-touching contact for different half-lives of virus persistence.** For a baseline cumulative probability of infection of 10% and each half-life value of virus persistence, the probability of transmission per single face-touching contact was cmputed. Hand contamination events are assumed to occur on average 4 times per hour.

S5 Sensitivity analyses

Cumulative probability of infection

We performed sensitivity analyses for different probabilities of infection and present here the results for probability of infection = 30% and probability of infection = 50% (see Figure S5–S6).

Figure S7 shows the impact of hand contamination rate on the number of hand washes that are necessary to prevent 50% of transmissions. A baseline probability of infection of 30% was used.

Figure S5: **Impact of half-life of probability of virus persistence on probability of infection for different hand washing schemes and frequencies.** (A) Fixed-time hand washing (B) Event-prompted hand washing. The dashed lines represent the half-life of probability of persistence for H3N2 for viral inoculum volumes of 2 μL and 30 μL (calculated from [9]). For each half-life value of the virus, the probability of transmission per face-touching event ϵ was determined for a probability of infection = 30% in the case of no hand washing. The probability of infection for the different hand washing frequencies/delays was then computed using this ϵ value. Hand contamination events are assumed to occur on average 4 times per hour.
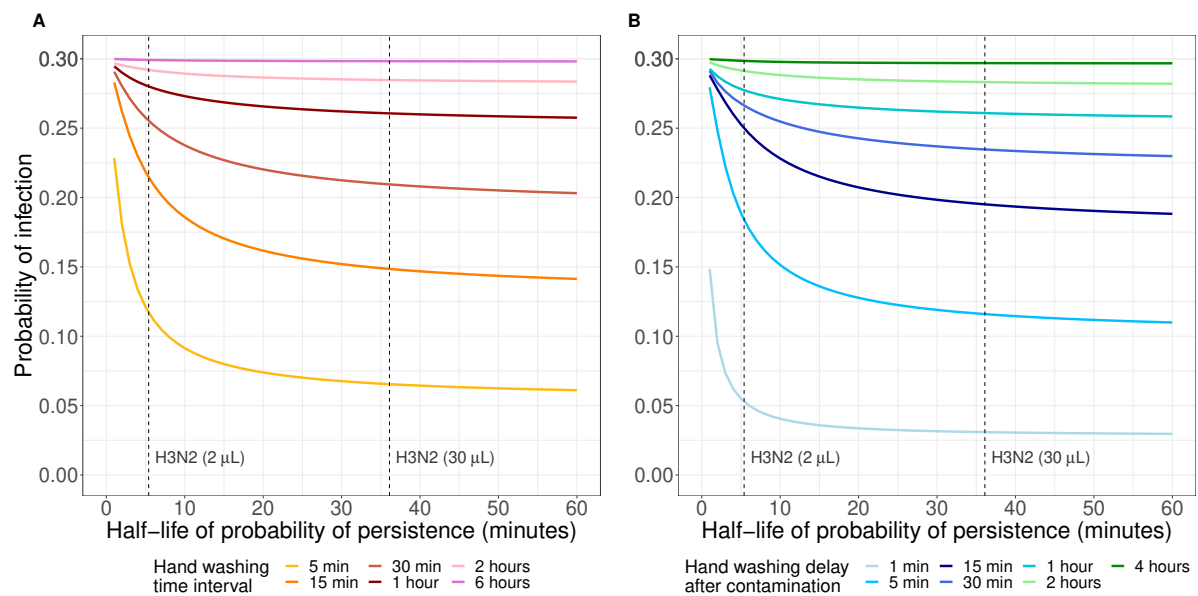
Figure S6: **Impact of half-life of probability of virus persistence on probability of infection for different hand washing schemes and frequencies.** (A) Fixed-time hand washing (B) Event-prompted hand washing. The dashed lines represent the half-life of probability of persistence for H3N2 for viral inoculum volumes of 2 μL and 30 μL (calculated from [9]). For each half-life value of the virus, the probability of transmission per face-touching event ϵ was determined for a probability of infection = 50% in the case of no hand washing. The probability of infection for the different hand washing frequencies/delays was then computed using this ϵ value. Hand contamination events are assumed to occur on average 4 times per hour.

Baseline probability of infection: 30%

Figure S7: **Number of hand washes necessary to decrease the cumulative probability of infection by 50%.**

Hand contamination event rate

We performed sensitivity analyses for different rates of hand contamination events and present the results for hand contamination rates of 1, 10 and 60 times per hour. The less frequently hands get contaminated, the larger the impact of increasing hand washing frequencies or reducing the delay of hand washing after hand contamination events and the larger the impact of the half-life of the probability of persistence of the virus on the actual probability of infection reduction. Figure S8 shows the results for a hand contamination rate of $\lambda_c = 1$ hour^{-1}. The conclusions drawn from the Results section are applicable in this scenario as well. Figure S9-S10 depict the results for a hand contamination rate of 10 and 60 times per hour, respectively. When hand contamination occurs very frequently, fixed-time and event-prompted hand washing have almost identical effects. For both hand washing schemes, increasing the hand washing uptake has only a small impact on the probability of infection unless hand washing is performed every 5 minutes or the time delay of hand washing after hand contamination events is decreased to one or five minutes. However, due to the the high rate of hand contamination events of every 4 minutes or every minute, respectively, such an
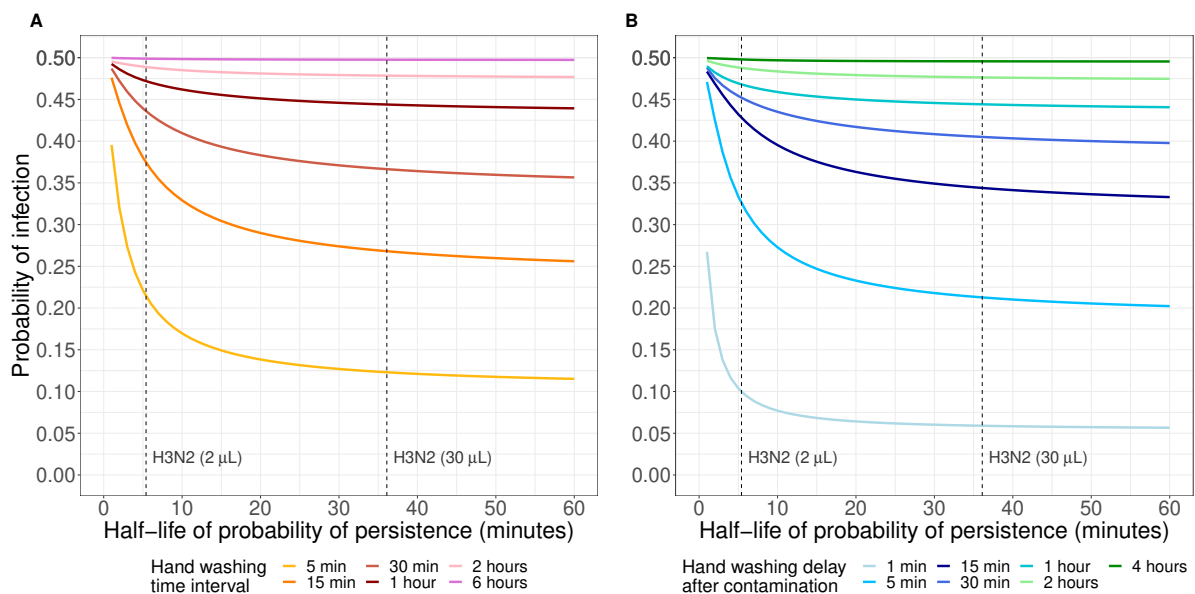
Figure S8: **Impact of half-life of probability of virus persistence on probability of infection for different hand washing schemes and frequencies.** (A) Fixed-time hand washing (B) Event-prompted hand washing. The dashed lines represent the half-life of probability of persistence for H3N2 for viral inoculum volumes of 2 μL and 30 μL (calculated from [9]). For each half-life value of the virus, the probability of transmission per face-touching event ϵ was determined for a probability of infection = 30% in the case of no hand washing. The probability of infection for the different hand washing frequencies/delays was then computed using this ϵ value. Hand contamination events are assumed to occur on average once per hour.
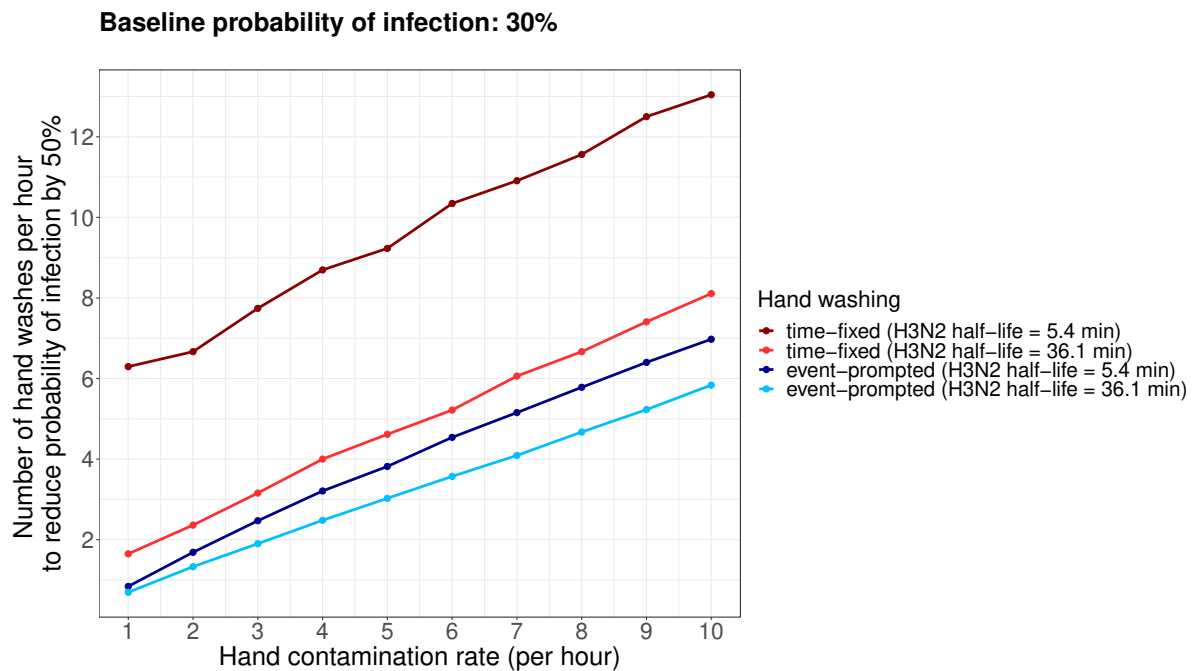
uptake seems infeasible. Hence, when susceptible individuals are exposed to continuous contamination, the best strategy would be to wash their hands as frequently as possible, especially after touching potentially contaminated surfaces, and to reduce the rate of contamination by, e.g., cleaning surfaces in their environment or isolating the infectious person.
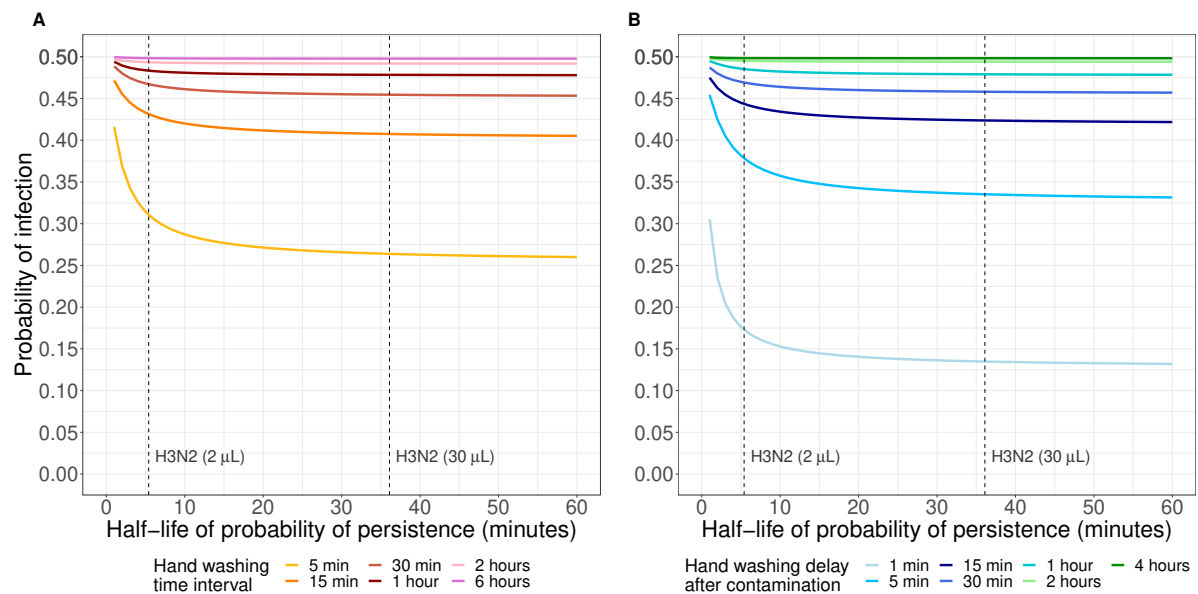
Figure S9: **Impact of half-life of probability of virus persistence on probability of infection for different hand washing schemes and frequencies.** (A) Fixed-time hand washing (B) Event-prompted hand washing. The dashed lines represent the half-life of probability of persistence for H3N2 for viral inoculum volumes of 2 μL and 30 μL (calculated from [9]). For each half-life value of the virus, the probability of transmission per face-touching event ϵ was determined for a probability of infection = 50% in the case of no hand washing. The probability of infection for the different hand washing frequencies/delays was then computed using this ϵ value. Hand contamination events are assumed to occur on average 10 times per hour.
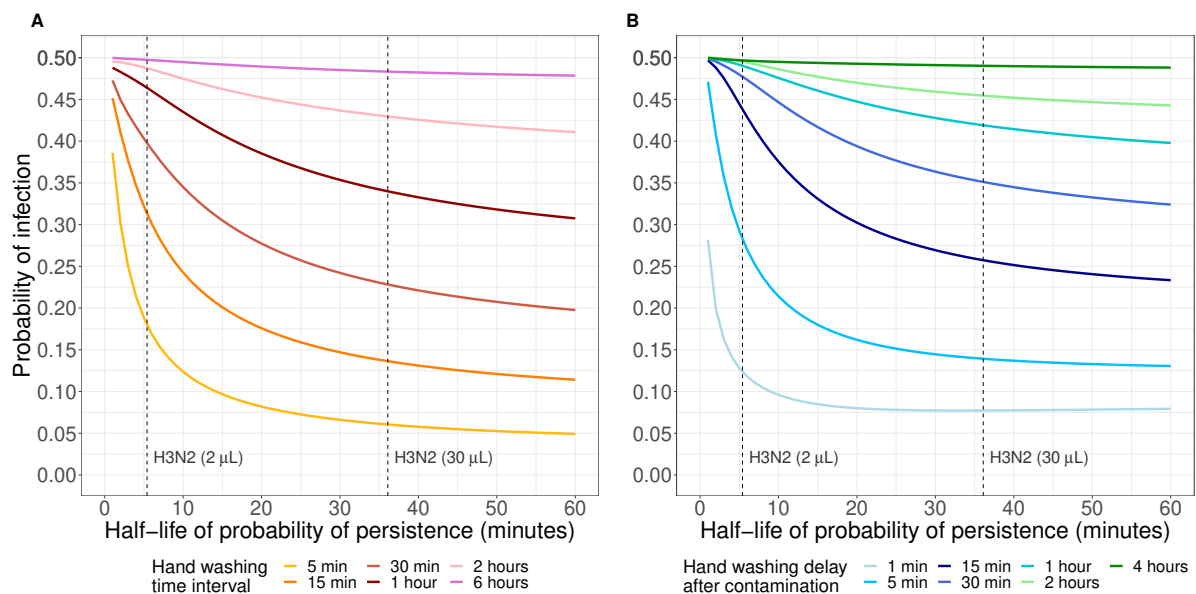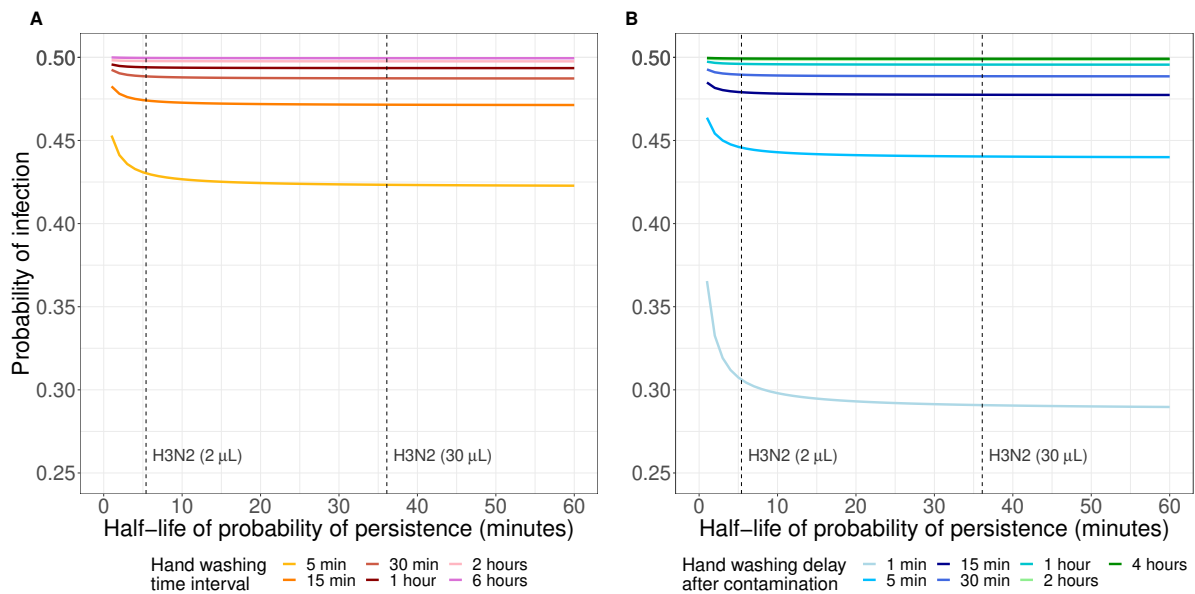
Figure S10: **Impact of half-life of probability of virus persistence on probability of infection for different hand washing schemes and frequencies.** (A) Fixed-time hand washing (B) Event-prompted hand washing. The dashed lines represent the half-life of probability of persistence for H3N2 for viral inoculum volumes of 2 μL and 30 μL (calculated from [9]). For each half-life value of the virus, the probability of transmission per face-touching event ϵ was determined for a probability of infection = 50% in the case of no hand washing. The probability of infection for the different hand washing frequencies/delays was then computed using this ϵ value. Hand contamination events are assumed to occur on average 60 times per hour.

Part II

Mathematical models to control the spread of infectious diseases in hospital settings

Chapter 3

Tracking Pseudomonas aeruginosa transmissions due to environmental contamination after discharge in ICUs using mathematical models.

Thi Mui Pham, Mirjam Kretzschmar, Xavier Bertrand, Martin Bootsma, on behalf of
COMBACTE-MAGNET Consortium

Abstract

Background: *Pseudomonas aeruginosa* (*P. aeruginosa*) Pseudomonas aeruginosa (*P. aeruginosa*) is an important cause of healthcare-associated infections, particularly in immunocompromised patients. Understanding how this multi-drug resistant pathogen is transmitted within intensive care units (ICUs) is crucial for devising and evaluating successful control strategies. While it is known that moist environments serve as natural reservoirs for *P. aeruginosa*, there is little quantitative evidence regarding the contribution of environmental contamination to its transmission within ICUs. Previous studies on other nosocomial pathogens rely on deploying specific values for environmental parameters derived from costly and laborious genotyping. Using solely longitudinal surveillance data, we estimated the relative importance of *P. aeruginosa* transmission routes by exploiting the fact that different routes cause different pattern of fluctuations in the prevalence.

Methods: We developed a mathematical model including background transmission, cross-transmission and environmental contamination. Patients contribute to a pool of pathogens by shedding bacteria to the environment. Natural decay and cleaning of the environment lead to a reduction of that pool. By assigning the bacterial load shed during an ICU stay to cross-transmission, we were able to disentangle environmental contamination during and after a patient's stay. Based on a data-augmented Markov Chain Monte Carlo method the relative importance of the considered acquisition routes is determined for two ICUs of the University hospital in Besançon (France). We used information about the admission and discharge days, screening days and screening results of the ICU patients.

Results: Both background and cross-transmission play a significant role in the transmission process in both ICUs. In contrast, only about 1% of the total transmissions were due to environmental contamination after discharge.

Conclusions: Based on longitudinal surveillance data, we conclude that cleaning improvement of the environment after discharge might have only a limited impact regarding the prevention of *P.A.* infections in the two considered ICUs of the University hospital in Besançon. Our model was developed for *P. aeruginosa* but can be easily applied to other pathogens as well.

Introduction

Hospital-acquired infections are a major cause of morbidity and mortality worldwide [1]. In industrialized countries, about $5 - 10\%$ of admitted acute-care patients are affected whereas the risk is even higher in developing countries [2].

Due to its intrinsic resistance to multiple antibiotics, *Pseudomonas aeruginosa* (short *P. aeruginosa* or *P. A.*) is an important contributor to nosocomial infections [3–5]. The most serious *P. aeruginosa* infections lead to bacteremia, pneumonia, urosepsis, wound infection as well as secondary infection of burns [6]. In 2018, the World Health Organization has recognized *P. aeruginosa* as a serious health-care threat by including it in the list of antibiotic-resistant highest priority pathogens [7].

Given the severe consequences of *P. aeruginosa* infections, in particular for critically-ill patients, it is clear that strategies preventing infections are seen as a key priority. However, infections are recognized as only the tip of the iceberg, while colonizations represent the true load of pathogens carried by patients in the intensive-care unit (ICU). Understanding the dynamics of *P. aeruginosa* colonizations is therefore crucial for developing and evaluating infection control policies.

There are several modes of transmission for colonizations. An overview of the reservoirs and modes of *P. aeruginosa* transmission can be found, e. g. in [8]. Potential sources of colonization can be categorized into those with endogenous and exogenous origin. Colonization from endogenous sources is due to e. g. antibiotic selection pressure and was regarded as the most important route of *P. aeruginosa* [9–13]. However, more and more evidence has emerged on the importance of exogenous sources: Cross-transmission usually caused by temporarily contaminated hands of health-care workers (HCWs) has been identified as an additional source of transmission [14–19]. It is furthermore known that moist environments (e. g. soil and water) may serve as natural reservoirs of *P. aeruginosa* and that it can persist for months on dry inanimate surfaces [20]. Several studies have been performed to asses the sources of environmental contamination leading to cross-colonization. A rapid systematic review is given by [21].

Quantifying the relative importance of routes of transmission may serve as an essential tool in designing effective and tailored control strategies. There is little quantitative evidence in the scientific literature regarding the relative contribution of environmen-

tal contamination within the transmission dynamics of *P. aeruginosa* especially for non-epidemic situations. Prior investigations for *P. aeruginosa* are molecular epidemiological rather than modeling studies. Others have been modeling the importance of contaminated surfaces on the transmission of other nosocomial pathogens, e. g., for Methicillin-resistant Staphylococcus aureus (MRSA) and Vancomycin-Resistant Enterococci (VRE) [22–26]. However, they rely on deploying specific values for model parameters corresponding to the environment. Such information was obtained from previous studies that conducted extensive epidemiological surveillance in combination with costly, laborious as well as time-consuming methods of genotyping. Thus, these methods cannot be easily applied to other nosocomial pathogens without this cumbersome preliminary work. Therefore, an important question emerged: Can we quantify the impact of environmental contamination of *P. aeruginosa* on the transmissions within ICUs after the discharge of patients, using only longitudinal data?

In this paper, we present a mathematical transmission model that differentiates between three modes of transmission based only on longitudinal routine surveillance data. In particular, we are interested in estimating the relative contribution of environmental contamination after discharge. We used data from two ICUs of the University hospital in Besançon to estimate the parameters that characterize the transmission routes. The estimation procedure is based on a data-augmented Markov chain Monte Carlo simulation [27]. To our knowledge, this is the first quantitative analysis of the impact of environmental contamination after discharge on *P. aeruginosa* transmissions in ICUs using solely routine surveillance data.

Materials and methods

In this section, we present our framework for modeling the transmission routes of *P. aeruginosa* including environmental contamination, as well as the method for computing the relative contributions of the routes. We further elaborate on the procedure that we used to estimate the relevant transmission parameters. A brief introduction to the data used for the analysis is given. We describe the model selection as well as model assessment procedures that are used to compare the developed models and to assess the model fit to the data.

Transmission models

The underlying model for our algorithm is a SI-model (e.g. [28]). All patients are admitted to an ICU and either belong to the susceptible (*P. aeruginosa* negative) or colonized (*P. aeruginosa* positive) compartment at any given time. The latter includes patients with asymptotic carriage and those with *P. aeruginosa* infection.

A susceptible patient may become colonized at a certain transmission rate, which depends on the colonization pressure in the ward at the time. The corresponding transmission process is modeled by three different modes of transmission through which colonization can be acquired. They are distinguished based on the different patterns in the prevalence time series induced by each of them. *Background transmission* is independent of other patients and is represented as a constant rate. Sources may be antibiotic selection pressure as well as the introduction by visitors or permanently contaminated environments, such as sinks or air-conditioning. Consequently, this route comprises endogenous and exogenous sources that lead to a prevalence which fluctuates around the mean value. The corresponding probability of acquisition for an uncolonized patient is therefore assumed to be constant during the time period. *Cross-transmission*, usually occurring via temporarily contaminated hands of health-care workers, is proportional to the fraction of colonized patients in the wards. The probability of colonization due to cross-transmission is high if the number of colonized patients is high and vice versa. *Environmental contamination* is modeled on a ward-level represented as a general pool of bacteria linked to objects contaminated by colonized patients. We focus on the bacterial load that may persist in the environment even after the discharge of patients. This leads to higher probabilities of acquiring colonization after outbreaks, even when the number of colonized patients is low.

The force of infection $\lambda(t)$, i.e. the probability per unit of time t for a susceptible patient to become colonized, is modeled as

$$\lambda(t) = \alpha + \beta \frac{I(t)}{N(t)} + \epsilon E(t) \tag{3.1}$$

where $I(t)$ is the number of colonized patients, $N(t)$ the total number of patients and $E(t)$ is a compartment tracking the overall bacterial load present in the ward at time

t. The parameters α, β and ϵ are transmission parameters linked to the background transmission term, fraction of colonized patients and the environmental bacterial load, respectively. Under the assumption of a force of infection $\lambda(x)$ at time x, the cumulative probability of any given susceptible person of becoming colonized in $[0, t]$ is $1 - e^{\int_0^t \lambda(x)dx}$ (see e.g. [29]). A schematic of the transmission model is presented in Fig 1.

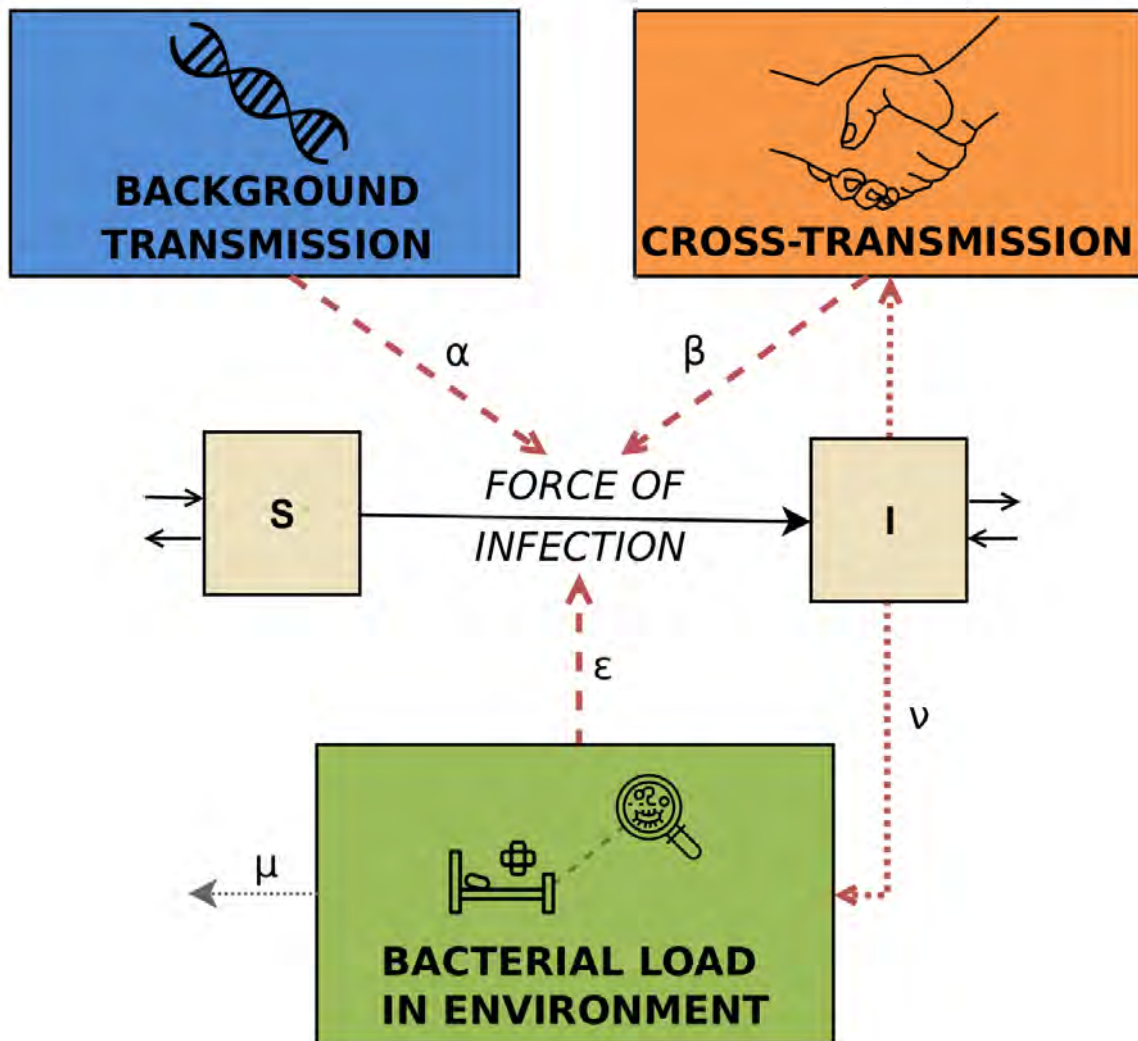**Figure 1**: **Schematic of the full transmission model.** It represents the three different routes, i.e. background transmission, cross-transmission and environmental contamination.

The described model is subject to the following further assumptions:

- Once colonized, patients remain colonized during the rest of the stay. This assumption is appropriate when the average length of stay of patients does not exceed the

duration of colonization, as is the case for *P. aeruginosa*.

- Colonization is assumed to be undetectable until a certain detectable bacterial level is reached. We do not distinguish between several levels of colonization. Furthermore, the detection of carriage in specimen is assumed to be the same for each screening separately.

- Assuming that HCWs are contaminated for a short period of time (typically until the next disinfection) in comparison with the length of carriage for patients, we use a quasi-steady state approximation [28]. This means that contact patterns between patients and HCWs are not explicitly modeled and we assume direct patient-to-patient transmission.

- All strains of *P. aeruginosa* are assumed to have the same transmission characteristics. We therefore assume that all colonized patients may be a source of transmission and contribute equally to the colonization pressure.

- All susceptible patients are assumed to be equally susceptible.

In order to analyze the impact of environmental contamination after the discharge of colonized patients, we model the underlying mechanism leading to the presence of pathogens in the environment after discharge. Patients contribute to the overall bacterial load by shedding *P. aeruginosa* at a rate ν during their stay. Furthermore, natural clearance and cleaning lead to a reduction of *P. aeruginosa* bacteria in the environment at a rate μ. The change of environmental contamination can be described by

$$\frac{dE}{dt} = \nu \frac{I(t)}{N(t)} - \mu E(t). \tag{3.2}$$

The differential equation (3.2) is solved by assuming $I(t) = I_t$ and $N(t) = N_t$ are known piece-wise constant functions with steps at times $t_0, t_1, \ldots, t_N$. Solving (3.2) using separation of variables leads to the overall bacterial load in the ward at time t:

$$E(t_i) = E_{t_{i-1}} e^{-\mu(t_i - t_{i-1})} + \frac{\nu}{\mu} \frac{I_{t_{i-1}}}{N_{t_{i-1}}} \left(1 - e^{-\mu(t_i - t_{i-1})}\right) \tag{3.3}$$

for $t_i \in \{t_0, \ldots, t_N\}$ and

$$E(t) = E_{\lfloor t \rfloor} e^{-\mu(t - \lfloor t \rfloor)} + \frac{\nu}{\mu} \frac{I_{\lfloor t \rfloor}}{N_{\lfloor t \rfloor}} \left(1 - e^{-\mu(t - \lfloor t \rfloor)}\right) \tag{3.4}$$

for $\lfloor t \rfloor := \max\{x \in \{t_0, \ldots, t_N\} \mid x \leq t\}$ and $t \in \mathbb{R} \setminus \{t_0, t_1, \ldots, t_N\}$. The initial amount of bacterial load is denoted by $E_0 := E(t_0)$. The full details of deriving Eq. (3.3) and (3.4) from (3.2) are given in S1 Text.

Given the number of colonized patients at a certain time t, the bacterial load $E(t)$ is deterministic. The acquisitions are stochastic based on the force of infection in (3.1). Our developed transmission model is therefore a hybrid of a stochastic and deterministic model.

All parameters, namely $\alpha, \beta, \epsilon, \mu, \nu$ and E_0 are assumed to be non-negative. By setting certain transmission parameters (α, β or ϵ) to zero, model variants may be defined. In this paper, we additionally consider a submodel with $\epsilon = 0$, where environmental contamination is not explicitly modeled and therefore only two transmission routes are considered. The force of infection for this transmission model with two acquisition routes is then given by $\lambda(t) = \alpha + \beta \frac{I(t)}{N(t)}$.

Relative contributions of transmission routes

For the prevention of colonization or infection with *P. aeruginosa*, specific intervention control strategies can be designed dependent on the relative importance of the transmission routes. However, for each observed acquisition of colonization, the responsible transmission route is unknown. And yet, for every acquisition, the probability that the colonization was due to a certain route can be estimated given that parameter values, the level of environmental contamination and the number of colonized patients are known. Thus, by estimating the transmission parameters $\alpha, \beta, \epsilon, \mu$ and ν, we were able to approximate the relative contributions of each transmission route to the total number of acquisitions.

The probability of acquisition can be approximated by the force of infection. It consists of different terms that can be assigned to the transmission routes under consideration, i.e.

$$\lambda(t) = \lambda_{\text{background}}(t) + \lambda_{\text{cross-transmission}}(t) + \lambda_{\text{environment}}(t)$$

The primary aim of this paper is to estimate the relative contribution of environmental contamination after discharge in order to estimate the role of terminal environmental cleaning among ICU patients. According to our full model, bacterial load is produced by a colonized patient currently present. The cumulative bacterial load increases over time until the respective patient is discharged. After discharge, shedding of that particular patient stops and decreases over time. The bacterial load shed during a patient's stay (which may then be transmitted via HCWs to other patients) is assigned to cross-transmission as in practice, it may not be distinguished from the classical definition of cross-transmission. The bacterial load persisting after discharge is the variable of interest and represents the impact of already discharged patients on the current transmissions in the ICU. A schematic of the bacterial load of a single patient over time and its attribution to the different transmission routes is visualized in Fig 2.
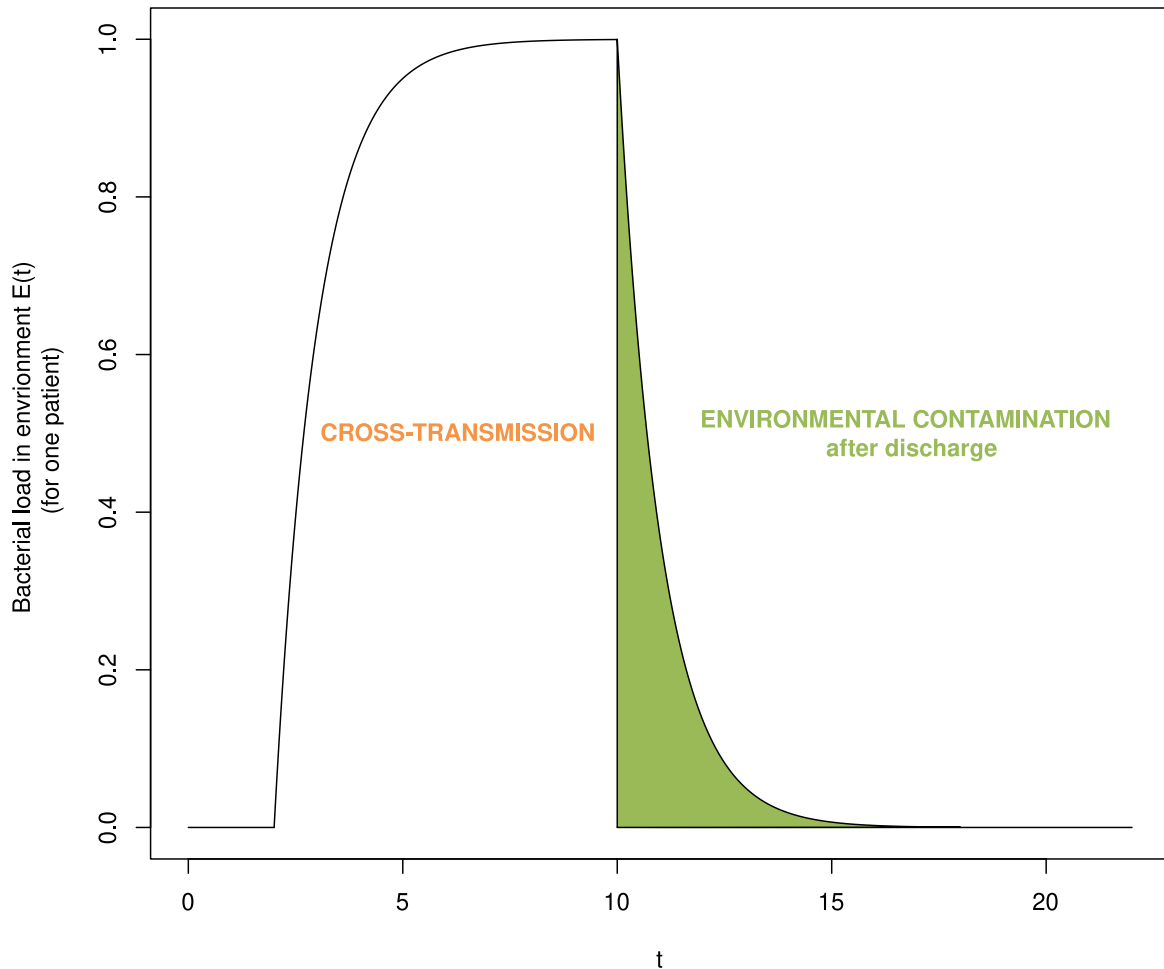
Figure 2: **Schematic of the bacterial load shed by a patient developing over time.**
The bacterial load that is shed during a patient's stay is assigned to cross-transmission.
Environmental contamination after discharge accounts only for the bacterial load persisting after the discharge of that patient.

The previous explanation leads to the following attribution of the terms to the different acquisition routes

$$\lambda(t) = \alpha + \beta \cdot \frac{I(t)}{N(t)} + \epsilon \cdot E(t)$$

$$= \underbrace{\alpha}_{\lambda_{\text{background}}} + \underbrace{\left[\beta \cdot \frac{I(t)}{N(t)} + \epsilon \sum_{i_p} \frac{E_{i_p}(t)}{N(t)}\right]}_{\lambda_{\text{cross-transmission}}(t)} + \epsilon \underbrace{\left[\sum_{i_d} \frac{E_{i_d}(t)}{N(t)} + E_0 e^{-\mu t}\right]}_{\lambda_{\text{environment}}(t)} \qquad (3.5)$$

where i_p indicates a colonized patient that is present at time t and i_d a colonized patient that has been colonized prior to t but was already discharged. The bacterial load produced by patient i at time t is given by

$$E_i(t) = \begin{cases} 0 & \text{for } t < t_i^c \\ \frac{\nu}{\mu}\left(1 - e^{-\mu(t-t_i^c)}\right) & \text{for } t_i^c \leq t < t_i^d \\ E_i(t_i^d)e^{-\mu(t-t_i^c)} & \text{for } t \geq t_i^d \end{cases}$$

where t_i^c is the time of colonization and t_i^d the time of discharge of patient i.

In continuous time, the relative contribution of a specific route to the overall number of acquired colonizations is determined by the ratio of the probability of colonization due to that route and the probability of colonization:

$$\text{Contribution of route } j = R_j = \frac{\sum_{i=1}^{I} \frac{P(\text{colonization at time } t_i^c \text{ due to route } j)}{P(\text{colonization at time } t_i^c)}}{\text{Number of acquisitions}}$$

$$= \frac{\sum_{i=1}^{I} \frac{\lambda_j(t_i^c)}{\lambda(t_i^c)}}{\text{Number of acquisitions}} \qquad (3.6)$$

where I is the number of colonized patients, $t_1^c, \ldots, t_I^c$ represent the times of colonization and j can be either of the three considered routes. The relative contributions are then given by:

- Contribution of background transmission $= R_{\text{background}} = \dfrac{\sum_{i=1}^{I} \frac{\alpha}{\lambda(t_i^c)}}{\text{Number of acquisitions}}$

- Contribution of cross-transmission $= R_{\text{crossT}} = \dfrac{\sum_{i=1}^{I} \frac{\beta \cdot \frac{I(t_i^c)}{N(t_i^c)} + \epsilon \sum_{i_p} \frac{E_{i_p}(t_i^c)}{N(t_i^c)}}{\lambda(t_i^c)}}{\text{Number of acquisitions}}$

- Contribution of environmental contamination $= R_{\text{env}} = \dfrac{\sum_{i=1}^{I} \dfrac{\epsilon \left[\sum_{i_d} \frac{E_{i_d}(t_i^c)}{N(t_i^c)} + E_0 e^{-\mu t_i^c} \right]}{\lambda(t_i^c)}}{\text{Number of acquisitions}}$

For the submodel including only background and cross-transmission, the computation of the relative contribution is derived from above by setting $\epsilon = 0$.

More details on the calculations can be found in S3 Text. In practice, colonization events are observed only in discrete times. The formulas for the transmission model and the relative contribution are adapted for this discrete time assumption and are elaborated in S2-S3 Texts. Since the calculations for the relative contributions of the transmission routes in the discrete-time scenario require the use of the gamma function and therefore become computationally intensive, we use the continuous-time formulas as approximations. Since values of the force of infection $\lambda(t)$ are typically small (< 0.25), the force of infection itself is a good approximation of the probability of infection as $1 - e^{-\int_0^t \lambda(x)dx} \approx \lambda(t)$ for small values of $\lambda(t)$. Hence, the discrete-time formulas for the relative contributions can be approximated by the continuous-time formulas evaluated at discrete time steps.

Estimation procedure

We assume that a patient is admitted to the ICU at time t_i^a and discharged at time t_i^d. The probability that a patient is admitted already colonized is described by the importation probability f. The rate at which a susceptible patient transitions to being colonized is given by Eq. (3.1). The colonization state of an individual patient is determined from screening information. We suppose that for each patient i a set of screening results $X_i = \left(X_i^{(1)}, \ldots, X_i^{(m)} \right)$, taken on days $t_i^{(1)}, \ldots, t_i^{(m)}$ is available. The set of all screening results is denoted by $X = \{X_1, \ldots, X_n\}$ where n is the total number of patients. Since screening tests are typically intermittent and imperfect, we define the test sensitivity ϕ, i.e. probability that a colonized patient has a positive result.

The aim is to estimate the model parameters $\alpha, \beta, \epsilon, \mu, \nu$ and E_0 as well as the sensitivity of the screening test ϕ and the importation rate f based on longitudinal data. The relative contributions of the transmission routes can then be estimated following the description in (3.6). The key idea of the estimation procedure is to fit a stochastic transmission model to the observed data. It is based on certain patterns of fluctuations in the

prevalence linked to the different transmission routes (as previously described in section).

In the analysis, we use the following input data for each patient:

- day of admission

- day of discharge

- screening days and results.

Thus, we use a day as the smallest time unit in our model and assume that events occur in daily intervals. In principle, other time units may be chosen for an analysis if the required information on admission, discharge and culturing is available. However, smaller units may increase the computational time.

If transmission dynamics were perfectly observed, it would be straightforward to calculate the likelihood of the data given parameters $\theta = \{\alpha, \beta, \epsilon, \mu, \nu, \phi, f\}$. However, the true colonization time of a patient is typically unobserved which leads to uncertainty about the true prevalence at any given time. Hence, the likelihood is analytically intractable. The method developed by [27] overcomes this problem by augmenting the parameter space with the unobserved colonization times and sampling over this space using an Markov-chain Monte Carlo (MCMC) algorithm. We adapted this method for our purposes to estimate the posterior distributions of the model parameters. The joint likelihood is determined using three models: an observation model, a transmission and importation model, and a prior model. The observation model describes the imperfect observation of the transmission dynamics for given the (augmented) colonization times. The transmission and importation model describe the probabilities of the realizations given the model parameters. The prior model determines the distribution of the parameters a priori.

The augmented data consists of a set of colonization statuses and times as well as importation markers. At each iteration, imperfectly observed colonization times are imputed and model parameters θ sampled that are consistent with the observed culture data. This approach accounts for imperfect and infrequent screening, missing admission and discharge swabs and leads to an estimation of the true (rather than the observed) prevalence on admission. Precise details of the analysis can be found in S5 Text. The algorithm was implemented in C++ and was tested using simulated data. Convergence

of the MCMC chains were verified using visual inspection.

We used uninformative exponential priors Exp(0.001) for the transmission parameters α, β, ϵ and μ. Parameters for the proposal distribution were tuned in order to ensure rapid convergence. Similar to [30], we estimated the sensitivity ϕ and importation parameter f using uninformative beta prior distributions Beta(1, 1). The initial bacterial load E_0 was approximated by $\frac{\nu}{\mu}\bar{I}$ with $\bar{I}$ being the mean prevalence in the ward. A discussion concerning the choice of parameters for the prior distribution is left to the supplementary section S8.

The MCMC algorithm was run for 500, 000 iterations following a burn-in of 30, 000 iterations. The MCMC iterations were then thinned by a factor of 10, leaving 50, 000 iterations for inference. In each iteration, 20 data-augmentation steps were performed with each augmentation chosen at random.

During the estimation process, several assumptions are made.

- Incorporating both sensitivity and specificity parameters in a model may cause iden-tifiability issues. Thus, test specificity was assumed to be 100%, meaning that pos-itive results were assumed to be true positive. Experimental results indicate the specificity of screening tests to be close to 100% [31].

- The initial bacterial load E_0 is assumed to be the environmental contamination at the beginning of the study period. The effect of E_0 diminishes proportionally to $\exp(-\mu)$ per day. It is therefore sufficient to use an approximation rather than including it as a parameter in the estimation process. We use the equilibrium state of (3.2) as an approximation, i.e.

$$E_0 \approx \frac{\nu}{\mu}\bar{I}$$

where $\bar{I}$ represents the mean prevalence in the ward.

- The environmental contribution to the force of infection at time t is $\epsilon \cdot E(t)$. As the total amount of environmental contamination $E(t)$ is unobserved, it is only

possible to estimate the product $\epsilon \cdot E(t)$. For $t = 0, 1, 2, \ldots$ it holds

$$\epsilon \cdot E(t) = \epsilon E_0 e^{-\mu t} + \frac{\epsilon \cdot \nu}{\mu} \sum_{i=0}^{t-1} \frac{I_i}{N_i} (1 - e^{-\mu})(e^{-\mu})^{t-1-i}.$$

The parameter ν is always integrated in the product $\epsilon \cdot \nu$. Hence, instead of estimating ϵ and ν separately, it is sufficient to estimate the product $\epsilon \cdot \nu$.

- Colonization was defined as the presence of bacteria at the screening sites as reported in the available data. Admission and screening are assumed to occur at 12:00 pm and discharge at 11:59 am.

- Re-admissions are not accounted for. Instead every new admission is treated as a new patient. The probability to be positive on admission is therefore identical for all patients, irrespective whether it is a readmission or not. Since we are interested in the overall prevalence and overall relative contribution of the acquisition routes rather than individual predictions, we do not expect this to have a major influence on our results.

- Since the smallest time unit is one day, colonization events occurring on a particular day are assumed to be independent.

- A negative result on the day of colonization is considered to be a false negative result.

- It is assumed that colonized patients contributed to the total colonized population from the day after colonization onwards, or for importations, from the day of admission. This assumption leads to an underestimation of the number of acquisitions for colonization times at the beginning of the day (but just after screening). On the other hand, since pathogenic bacteria such as P. aeruginosa undergo a *lag phase* during their growth cycle, in which the bacteria adapt to the new environment and are not yet able to divide, onward transmission events are likely to be rare during the early stages of colonization. Therefore, the number of onward transmissions are likely to be overestimated for colonizations occurring at the end of the day.

Data

The data used in the current analysis were collected from two ICUs, denoted by A and B, between 1999 and 2016 at University Hospital of Besançon, eastern France, in the framework of a systematic screening for *P.aeruginosa*. The data sets include admission and discharge dates as well as dates, sites and results of culturing of adult patients. ICU A is a surgical ICU that comprised 15 beds in the time period 1999-2008 and 20 from 2010 till 2016. The ICU was renovated between 2008 and 2009 and the number of beds was increased after completion of the renovation work. ICU B, a non-surgical ICU, had 15 beds from 2000 till 2011 and increased to 20 beds afterwards. Rectal and nose swabs were obtained upon admission (during the first 48 hours) and once a week thereafter. A positive result on one of the swabs was counted as a positive culture. A negative culture resulted from a negative culture on both swabs taken at the specific day. More than 84% of admitted patients were screened. As HCWs, including physicians, were (with minor exceptions) working only in one of the ICUs during the whole study period, the two ICUs can be treated independently in the analysis.

Since 2000, the hand hygiene procedures recommended in both ICUs is rubbing with alcohol-based gels, or solutions (ABS). Cleaning of the rooms is done daily by using the detergent-disinfectant Aniosurf®. The sinks were cleaned daily before pouring the detergent-disinfectant Aniosurf® into the U-bends. Plumbing fittings were descaled weekly.

In our main analysis, data for each ICU and each time period (before and after renovation) was treated as distinctive data sets, resulting in four different analyses. No pooling of the results were performed. In a second analysis, the data for the different time periods and different ICUs were combined. The results are compared with the main analysis and are presented in S1-S2 Tables. Each data set was analyzed using

- the full model including background transmission, cross-transmission and environmental contamination after discharge,

- the submodel with only background and cross-transmission.

Patient data were anonymized and de-identified prior to analysis.

Model selection

To assess the relative performance of a given model, we used a version of the deviance information criterion (DIC) based on [32]. For an estimated parameter set θ and observed data set x it is computed as the expected deviance plus the effective number of parameters: $\text{DIC} = \overline{D_x(\theta)} + p_D$. A lower value indicates a better fit. The effective number of parameters p_D represents a complexity measure and is calculated by the difference of the posterior mean deviance and the deviance at the posterior mean: $\overline{D_x(\theta)} - D_x(\tilde{\theta})$. In this paper, we use the approximation $p_D = \frac{1}{2}\text{var}(D_x(\theta))$ introduced by [32].

The DIC is a simple measure that can be used to compare hierarchical models. Furthermore, it allows determining whether two data sets may be concatenated or should be treated separate. The idea is to distinguish two models: one that includes one parameter set for both ICUs (and therefore treats them as concatenated) and one that includes different parameter sets for each ICU (and thus treats them as separate). The first scenario leads to one analysis and one DIC value whereas the second model results in two independent analyses and hence two DIC values. The sum of the DICs of the latter may be compared to the DIC value of the first scenario. A smaller DIC value is preferred. More details can be found in S6 Text.

Model assessment

We chose to check the adequacy of the models using the following approach. The ability of the model to predict the probability of acquisition based on the predicted force of infection was assessed. The computed numerical values for the force of infection are assigned to a bin representing the segment covering the numerical value. For a given value λ of the force of infection, the theoretical probability of acquisition p_{acq} per susceptible patient is computed by $1 - \exp(-\lambda)$. The predicted fraction of acquisitions f_{acq} is computed by dividing the number of acquisitions N_{acq} by the number of susceptible patients N_{susc}. We compute 95% confidence intervals assuming that the number of acquisitions follows a binomial distribution of $\text{Bin}(N_{susc}, f_{acq})$. The described method is performed for 100 MCMC updates. Coverage probabilities are computed to determine the actual proportion of updates for which the interval contains the theoretical probability of acquisition. We set the nominal confidence level to 0.95. A good fit is given when the actual coverage probability is (more or less) equal to the nominal confidence level. In order to avoid

the coverage probability tending to zero when p_{acq} tends to 0 or 1, Jeffreys confidence intervals are used (as recommended in [33]). When $N_{acq} = 0$ the lower limit is set to 0, and when $N_{acq} = N_{susc}$ the upper limit is set to 1.

Results

Descriptive analysis of data

The descriptive statistics of the data sets corresponding to ICU A and B with respect to the number of admissions, lengths of stay and colonization characteristics are shown in Table 1. The time period referred to as *before renovation* (short before) is defined as $20/04/1999 - 23/01/2008$ (approx. 8.8 years) for ICU A and $11/01/2000 - 12/01/2011$ (approx. 11 years) for ICU B. The time period referred to as *after renovation* (short after) is defined as $31/05/2008 - 30/12/2016$ (approx. 6.4 years) for ICU A and $13/01/2011 - 13/09/2016$ (approx. 5.7 years) for ICU B. In ICU A, the number of beds decreased during the renovation. Hence, we decided to remove the renovation period from the analysis for ICU A.

In total, 13,065 patients (6,061 admitted to ICU A and 7,004 to ICU B) and 37,738 screening results (14,631 in ICU A and 23,107 in ICU B) were included in the analysis. The number of readmissions is higher for ICU B than for ICU A. In our analysis, every admission was treated separately (as a new patient) resulting in 14,403 admissions (6,659 admitted to ICU A and 7,744 to ICU B).

The corresponding median length of stay was 8.0 days for both ICUs before and after renovation, respectively. Hence, there is hardly any difference between the ICUs, nor between the two time periods regarding the median length of stay.

The fraction of patients who were positive on admission was slightly higher after renovation in ICU A. The reverse is true for ICU B. The observed fraction of patients who acquired colonization slightly decreased after renovation in both ICUs. There were 1,519 patients (620 in ICU A and 899 in ICU B) observed to be colonized during their stay and 388 patients (137 in ICU A and 251 in ICU B) observed to be colonized on admission. The percentage of patients admitted positively on admission and with acquired colonization is higher in ICU B than in ICU A.

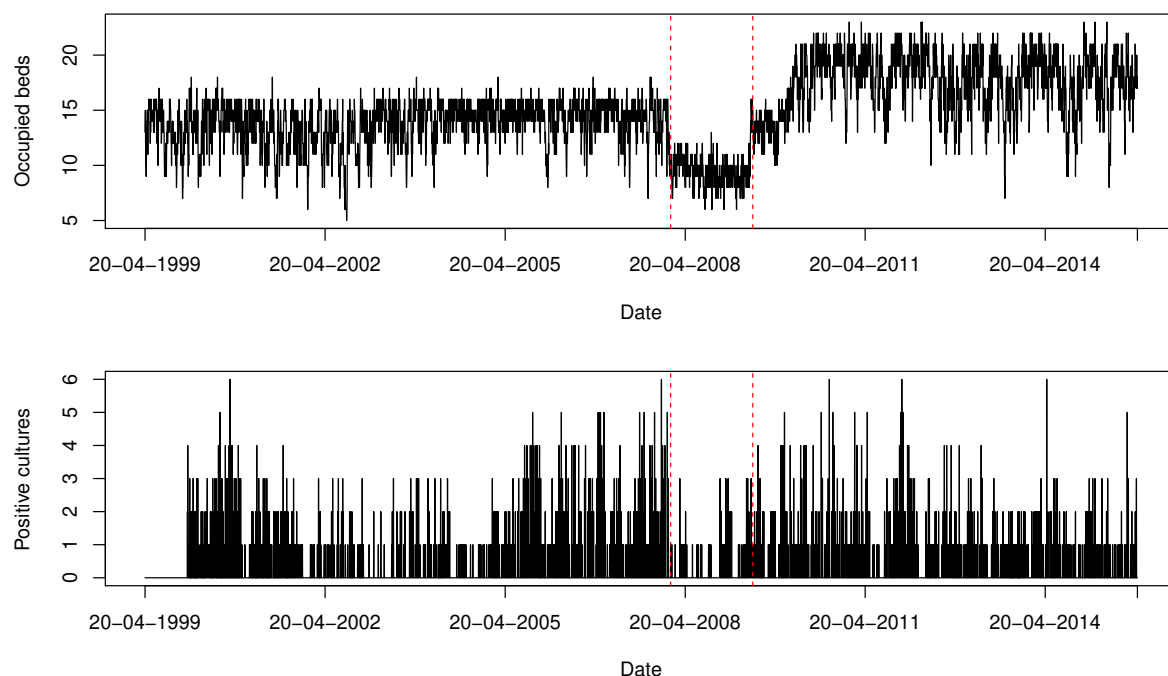The total number of patients per ICU and the number of positive cultures are visualized in Figs 3 and 4.

Figure 3: **Number of occupied beds and positive isolates cultured from patients per swab day for ICU A.** The red dotted lines indicate the time points that splits the study period into *before renovation* (20/04/1999 − 23/01/2008) and *after renovation* (31/05/2009 − 30/12/2016). Since the number of beds decreased during the renovation, the period is removed from the analysis.

Estimated model parameters

Two model variants were fitted to the Besançon ICU data aiming to estimate the set of parameters $\theta_1 = \{\alpha, \beta, \phi, f\}$ and $\theta_2 = \{\alpha, \beta, \epsilon, \mu, \phi, f\}$ corresponding to the submodel with only two and the full model with all three transmission routes, respectively.

Submodel: Two transmission routes

Posterior estimates of the model parameters for each ICU and each time period are reported in Table 2. Acceptance probabilities for proposed updates to the augmented data ranged from 3.2% (ICU B after renovation) to 11.1% (ICU A before renovation). Pairwise scatter plots indicated little correlation between parameter values, with the exception of a negative correlation between α and β (see S1). Histogram and trace
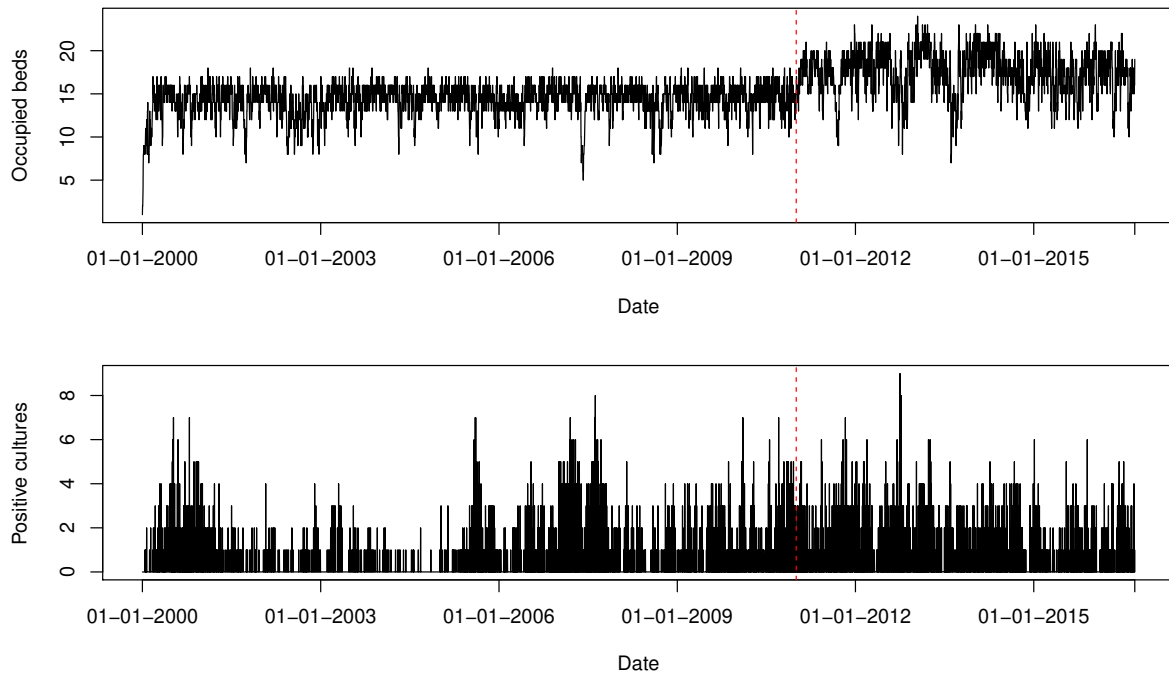
Figure 4: **Number of occupied beds and positive isolates cultured from patients per swab day for ICU B.** The red dotted line indicates the time point that splits the study period into *before renovation* $(11/01/2000 - 12/01/2011)$ and *after renovation* $(13/01/2011 - 13/09/2016)$.

plots of the posterior estimates are given in S2-S5 Figs and show that the MCMC chains rapidly mix and quickly converge to their stationary distribution. We found our estimates to be robust to the choice of priors for transmission parameters.

The probability of being colonized with *P. aeruginosa* on admission and the screening test sensitivity varied between the two ICUs and the time periods. For both ICUs, the median estimates of the importation probability f is higher in the data set after renovation than before, i.e. 4.5% and 6.2% for ICU A and 6.0% and 9.9% for ICU B. The difference between the time periods is only significant for ICU B. We estimated the median of the prevalence of *P.A.* to be 24.4% and 19.9% for ICU A and 22.3% and 24.4% for ICU B before and after renovation, respectively. Median estimates for the screening test sensitivity were 50.9% and 50.2% for ICU A and 61.8% and 58.6% for ICU B. Since the credibility intervals of the sensitivity estimates do not overlap with respect to the two ICUs, we can conclude that there is a 95% probability that the test sensitivity is higher in ICU B than in ICU A. Our possible explanation is based on the

Table 1: **Descriptive statistics for the P. aeruginosa carriage data collected from two ICUs at University Hospital Besançon, France, 1999-2016**

	No.			Median (IQR)[†]		%	
	Before[‡]	After[§]	Total	Before	After	Before	After
ICU A							
Study length, days	3200	2340	5540				
Readmissions	320	278	598			9.8	8.2
Admissions*	3261	3398	6659				
Length of stay, days				8.0 (3.0-19.0)	8.0 (3.0-16.0)		
Importations[¶]	50	87	137			1.5	2.6
Observed *P.A.* acquisitions[‖]	350	270	620			10.7	7.9
ICU B							
Study length, days	4020	2079	6099				
Readmissions	406	334	740			9.3	9.9
Admissions	4360	3384	7744				
Length of stay, days				8.0 (3.0-18.0)	8.0 (3.0-15.0)		
Importations	127	124	251			3.7	2.9
Observed *P.A.* acquisitions	504	395	899			11.6	11.7

*Each ICU stay was counted separately, even in case of a multiple ICU stay within a given hospitalization.
[†] Interquartile range
[‡] $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.
[§] $31/05/2009 - 30/12/2016$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.
[¶] Patients positive on admission; false negative results are not taken into account.
[‖] An acquisition is when a patient test negative on admission and had a postive result before discharge; false negative results are not taken into account.

Table 2: **Summary statistics of the marginal posterior distributions for parameters of the submodel based on the analysis of the Besançon data**

Parameter	Symbol	Median (95% credibility interval)*			
		ICU A		ICU B	
		Before[†]	After[‡]	Before	After
Background coefficient	α	0.011 (0.006, 0.016)	0.013 (0.009, 0.016)	0.007 (0.005, 0.01)	0.014 (0.009, 0.018)
Cross-transmission coefficient	β	0.043 (0.021, 0.064)	0.008 (0, 0.026)	0.046 (0.032, 0.06)	0.011 (0, 0.029)
Sensitivity	ϕ (%)	50.9 (47.8, 54.2)	50.2 (46.0, 54.4)	61.8 (59.6, 64.0)	58.6 (56.1, 61.1)
Importation probability	f (%)	4.5 (3.1, 6.1)	6.2 (4.8, 7.8)	6.0 (4.8, 7.1)	9.9 (8.2, 11.6)
Fraction colonized	p_{col} (%)	24.4 (23.1, 25.8)	19.9 (18.7, 21.2)	22.3 (21.6, 23.0)	24.4 (23.7, 25.1)
Contributions					
Background	$R_{background}$ (%)	53.6 (32.8, 75.9)	89.3 (67.9, 100)	43.4 (29.1, 58.7)	84.5 (60.9, 100)
Cross-transmission	R_{crossT} (%)	46.4 (24.1, 67.2)	10.7 (0, 32.1)	56.6 (41.3, 70.9)	15.5 (0, 39.1)

*Highest posterior density interval
[‡] $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.
[§] $31/05/2009 - 30/12/2016$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.

fact that the ICUs differ in their patient population. As a medical ward, ICU B contains patients with longer lengths of stay and more readmissions. Patients who are exposed to an ICU environment for a longer period of time may have a higher probability to get

colonized at a detectable level. However, our explanation is only hypothetical and the true reason for the difference is not known.

The relative importance of the two considered transmission routes per ICU and time period is depicted in Fig 5 (a) and (b). For ICU A, the median relative contribution of background transmission is 53.6% (95% CrI : 32.8 − 75.9%) and 89.3% (95% CrI : 67.9−100%) leaving 46.4% (95% CrI : 24.1−67.2%) and 10.7 (95% CrI : 0−32.1%) of the acquisitions assigned to cross-transmission before and after renovation, respectively. For ICU B, 43.4% (95% CrI : 29.1 − 58.7%) and 84.5% (95% CrI : 60.9 − 100%) of the acquisitions were due to the background and cross-transmission accounted for 56.6% (95% CrI : 41.3 − 70.9%) and 15.5% (95% CrI : 0 − 39.1%) of the acquisitions before and after renovation, respectively.
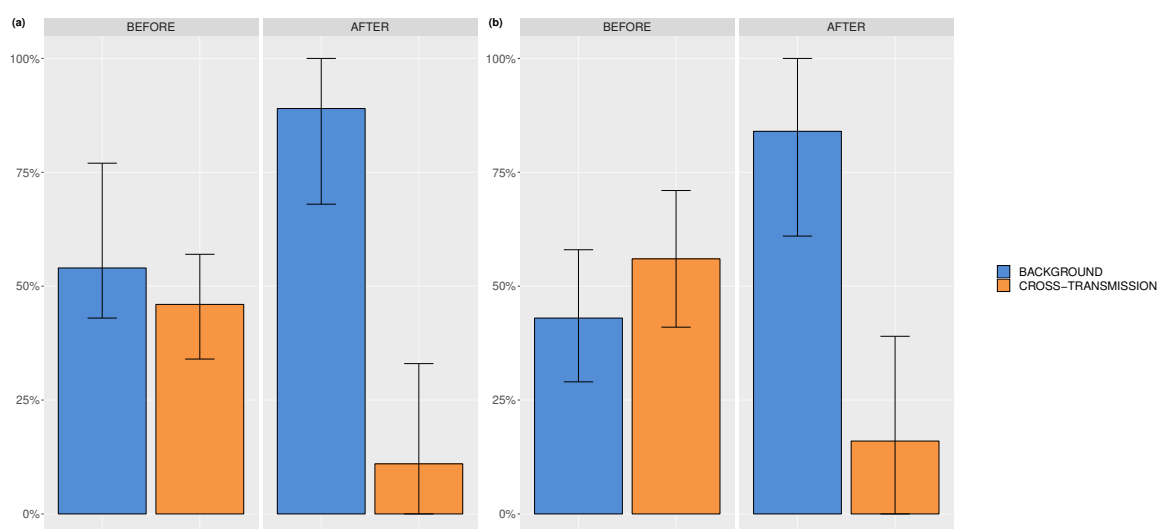

Figure 5: **Relative contributions of background and cross-transmission.** (a) ICU A before and after renovation, (b) ICU B before and after renovation.

The results suggest that both routes have an important impact on the acquisitions in both ICUs. The median estimates of the relative contribution of background transmission are higher after than before renovation in both ICUs. Thus, there is a tendency for lower contribution of cross-transmission route after renovation in both ICUs. Possibly, hygiene was improved after renovating the ICUs. However, since the credibility interval for background transmission overlap before and after renovation, there is no evidence that the relative contributions differ between the time periods. Before renovation, the credibility intervals of the relative contributions for background and cross-transmission

overlap. Thus, we conclude that no route considerably predominates the transmissions before renovation. On the other hand, the respective credibility intervals do not overlap after renovation. Hence, background transmission predominates the transmissions after renovation. Comparing the results across ICUs, we can see that the credibility intervals of the relative contributions overlap leading to the conclusion that the two ICUs do not seem to be different regarding the relative importance of the transmission routes.

Full model: Three transmission routes

Posterior estimates of the model parameters for each ICU are reported in Table 3. The estimates and interpretations for the importation rate f, the screening test sensitivity ϕ and the mean prevalence stay roughly the same when adding environmental contamination as an additional route. The same holds for the median relative contributions of background and cross-transmission. The median relative contribution of environmental contamination after discharge is less than 1% ranging from 0.3% to 0.5% for both ICUs and both time periods. The relative importance of the three considered transmission routes per ICU and time period is depicted in Fig 6 (a) and (b). Acceptance probabilities
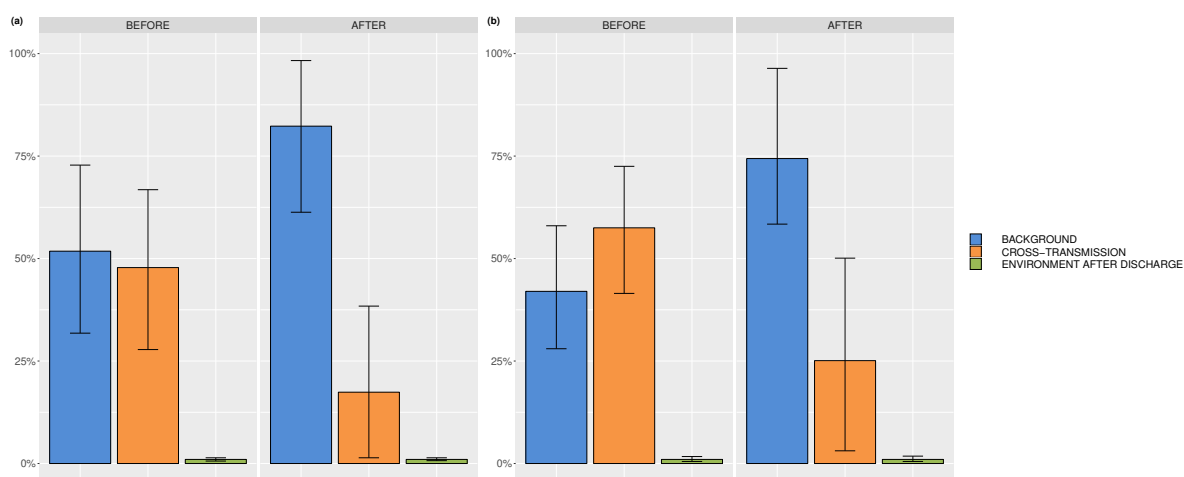
Figure 6: **Relative contributions of background transmission, cross-transmission and environmental contamination after discharge.** (a) ICU A before and after renovation, (b) ICU B before and after renovation.

for proposed updates to the augmented data ranged from 7.2% (ICU B after renovation) to 90% (ICU A before renovation). Pairwise scatter plots indicated strong correlations between α and β, β and ϵ and between ϵ and μ (see S14 Fig). The correlation coefficient of the latter pair ranged from 0.531 to 0.561. Furthermore, it can be seen in Table 3

that the credibility intervals for the parameters ϵ and μ are large. Nevertheless, histogram and trace plots of the posterior estimates show that the MCMC chains rapidly mixed and quickly converged to their stationary distribution as can be seen in S6-S13 Figs. The rapid convergence could be achieved by tuning the parameters of the proposal distribution for μ. In contrast, a flat prior for the decay rate μ in combination with a small initial standard deviation for its proposal distribution resulted in large acceptance ratios close to 1. The MCMC chain mixed too slowly and therefore hindered the identifiability of the likelihood. This can be explained by the fact that our developed model is overparametrized when colonizations of patients are not or hardly influenced by environmental contamination. Small values of the transmission parameter ϵ as well as high values of the decay rate μ would reflect the aforementioned situation. As a result, the respective likelihood might not be or only weakly identifiable. Our sensitivity analyses and artificial data simulations demonstrated similar pairwise scatter plots and wide credibility intervals for the parameters ϵ and μ in case of a small contribution of environmental contamination to the transmissions (more details can be found in S9 Text). Hence, we can conclude that the role of environmental contamination after discharge within the transmission process of *P. aeruginosa* in the two ICUs A and B is small before as well as after renovation.

Model selection

In total, 14 analyses were performed. For each ICU, three data sets were created - one for each time period and one combining the data sets before and after renovation. Additionally, the ICUs and time periods were combined in one data set. Each of the seven data sets were analyzed using the submodel and the full model. The DIC values for each model analysis can be found in Table 4. The analysis combining both ICUs and time periods shows smaller DIC values, i.e. 136507.8 and 130693, than the sum of the DICs for separate analyses (152428.8 and 150914.2) for both the submodel and full model, respectively. The full model results in a smaller DIC value for the analysis of the combined data set. Hence, based on the DIC, it would be sufficient to analyze the combined data set using the full model including endogenous route, cross-transmission and environmental contamination. Nevertheless, it can be seen in S10 Text that the posterior estimates of the different analyses are similar, especially for the relative contribution of environmental contamination after discharge.

Table 3: **Summary statistics of the marginal posterior distributions for parameters of the full model based on the analysis of the Besançon data**

Parameter	Symbol	Median (95% credibility interval)*			
		ICU A		ICU B	
		Before†	After‡	Before	After
Background coefficient	α	0.011 (0.006, 0.015)	0.012 (0.009, 0.016)	0.007 (0.004, 0.01)	0.012 (0.007, 0.017)
Cross-transmission coefficient	β	0.023 (0, 0.048)	0.006 (0, 0.022)	0.027 (0.0, 0.05)	0.008 (0, 0.027)
Environmental coefficient	ϵ	187.3 (0.041, 753.4)	76.0 (0002, 436.6)	48.8 (0.009, 227.9)	90.6 (0.017, 502.4)
Decay rate	μ	1202.7(24.4, 3184.9)	1567.3(34.9, 4453.7)	319.2 (13.8, 873.3)	1514.2(44.4, 4335.1)
Sensitivity	$\phi\,(\%)$	50.9 (47.8, 54.2)	49.5 (45.5, 53.5)	61.8 (59.5, 64.0)	58.7 (56.0, 61.3)
Importation probability	$f\,(\%)$	4.5 (3.1, 6.1)	6.1 (4.9, 7.6)	6.0 (4.9, 7.2)	10.1 (8.4, 11.9)
Fraction colonized	$p_{col}\,(\%)$	24.5 (23.1, 25.9)	20.1 (19.0, 21.4)	22.3 (21.7, 23.0)	24.4 (23.7, 25.1)
Contributions					
Background	$R_{backgr}\,(\%)$	51.8 (32.7, 73.0)	82.3 (61.0, 98.7)	42.0 (27.5, 58.0)	74.4 (48.1, 96.8)
Cross-transmission	$R_{crossT}\,(\%)$	47.8 (26.9, 66.9)	17.4 (1.3, 38.6)	57.5 (41.8, 72.1)	25.1 (2.9, 50.7)
Env. cont. after discharge	$R_{env}\,(\%)$	0.3 (0.0, 0.8)	0.2 (0, 0.7)	0.5 (0.0, 1.2)	0.4 (0.0, 1.3)

*Highest posterior density interval
‡ $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.
§ $31/05/2009 - 30/12/2016$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.

Model assessment

For each bin of the force of infection the coverage probabilities are plotted and can be found in S15-S16 Figs. It can be seen that the coverage probabilities are approximately (sometimes higher, sometimes smaller) equal to the nominal confidence level of 0.95. Thus, both the full model as well as the submodel gave adequate fits to the four data sets. In Fig 7, the predicted fraction of acquisitions are plotted against the binned force of infection for one exemplary MCMC update. The red lines indicate the relationship between the probability of acquisition and force of infection assumed by our models. For this example, it is always contained by the confidence intervals (blue lines).

Discussion

To our knowledge, our study is the first attempt to estimate the relative contribution of environmental contamination after discharge for *P. aeruginosa* based on mathematical modeling and using only admission, discharge and screening data. The three different routes, background transmission, cross-transmission and environmental contamination

Table 4: **Deviance information criterion for the different models**

	Submodel		Full model	
ICU A				
Before[†]	42961.72	$\left.\right\}\sum^{\S}$ 86899.53	35356.31	$\left.\right\}\sum$ 63057.37
After[‡]	43937.81		27701.06	
Combined*	88650.69		63779.12	
ICU B				
Before	35356.31	$\left.\right\}\sum$ 63057.37	34670.46	$\left.\right\}\sum$ 62568.88
After	27701.06		27898.42	
Combined	63778.12		59290.73	
Sum Combined[‖]	152428.8			
ICUs combined[¶]	136507.8			

[‡] $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.
[§] $31/05/2009 - 30/12/2016$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.
* Combined time periods
[§] Σ indicates that the sum of the respective columns in the previous row is calculated.
[¶] ICUs as well as time period (before, after renovation) are combined in one data set.
[‖] The sum of the DICs for ICU A (before and after renovation combined) and ICU B is computed.

after discharge, are distinguished by the resulting patterns of the prevalence that they induce. We estimated that environmental contamination after discharge accounts for at most 1% of the total *P. aeruginosa* transmissions in the two ICUs of the University hospital in Besançon before and after renovation. In contrast, background as well as cross-transmission are both essential for the transmission dynamics of *P. aeruginosa*. This suggests, that improvement of cleaning of the environment *after discharge* would have only a limited benefit regarding the prevention of *P. aeruginosa* colonization in the two considered ICUs of the University hospital in Besançon.

Previously, studies have been conducted to investigate the role of environmental contamination for colonizations of *P. aeruginosa*. For instance, Panagea et al. performed environmental studies to determine the extent of environmental contamination with an epidemic strain of *P. aeruginosa* [34]. They concluded that the transmissibility of the epidemic strain cannot be explained solely on the basis of improved environmental survival. Our results likewise demonstrate that the decay of *P. aeruginosa* is already
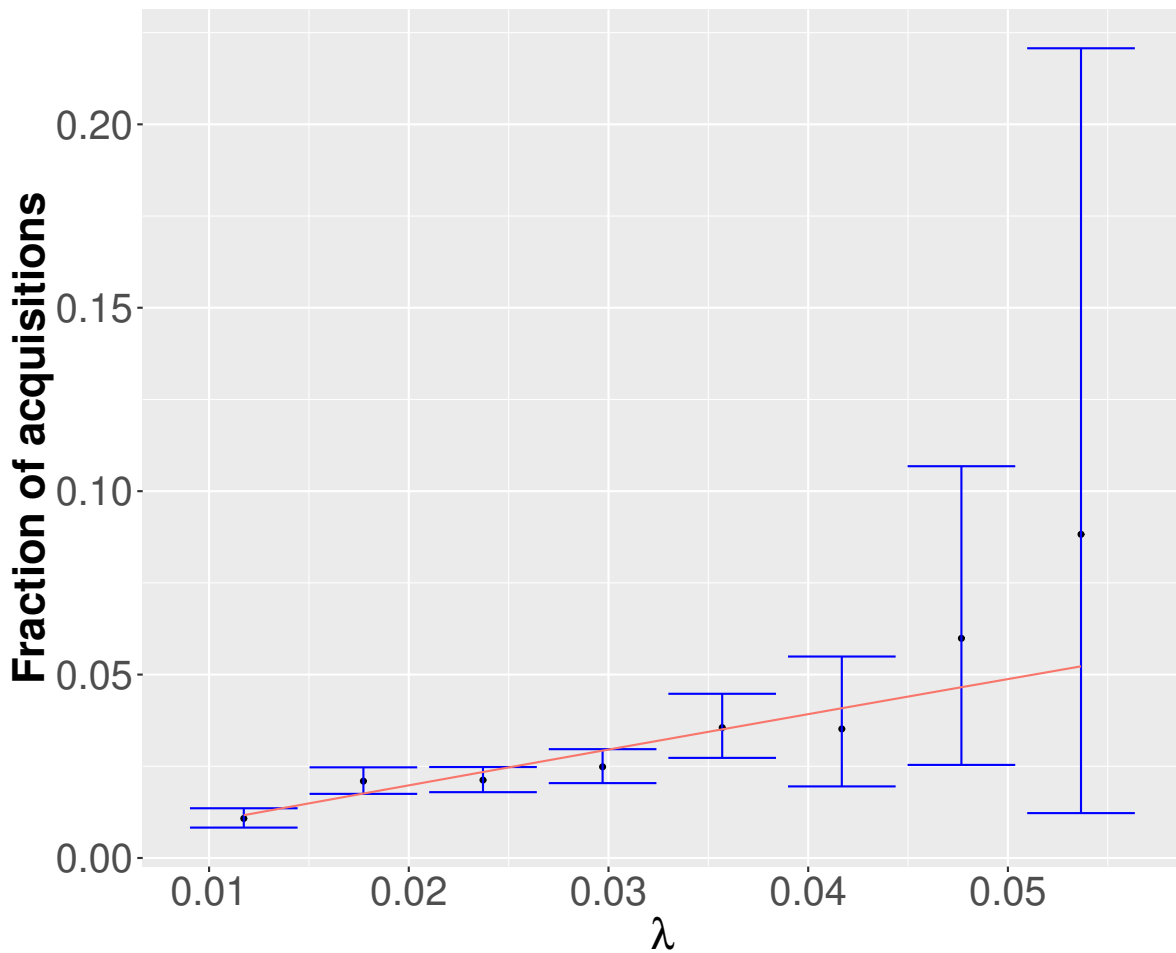
Figure 7: **Exemplary model assessment plot for one MCMC update using the sub-model applied to ICU A before renovation.** The predicted fraction of acquisition is plotted against the theoretical force of infection. The red line indicates the theoretical relation between the force of infection and the probability of acquisition. The blue lines indicate 95% credibility intervals.

rapid enough to limit its survival in the environment.

While our approach is efficient in determining the relative contribution of environmental contamination after discharge requiring merely longitudinal surveillance data, it has several limitations that may restrict its practical applicability.

Our conclusions on the impact of cleaning applies only to the environment after the discharge of patients. Permanently contaminated reservoirs in ICUs, such as sinks, may still serve as sources for colonization. In our model they are assigned to background transmission. Thus, while the effect of cleaning improvement after discharge might be limited

for the two considered ICUs, general cleaning improvement of the environment might be important to reduce permanent reservoirs for environmental contamination. Several studies based on molecular typing techniques suggest that contaminated taps and sinks in the environment may serve as a non-negligible source in the acquisition of *P. aeruginosa* colonization (see e. g. [8, 21, 35]). Since genotyping information is not available for the data set that we have analyzed, further studies for validation would increase the understanding of our conclusions.

In addition, we assume in our model that a patient's contribution to environmental contamination affects *all* patients present in the ward. This assumption might not be realistic as the patient admitted to the same room after the discharge of a colonized patient might be at a higher risk than patients in other rooms. In S11, we have investigated the possible influence of prior colonized bed occupants for the Besançon data sets. The results show that for these data sets, the impact of prior colonized bed occupants is limited ($< 6\%$). While prior bed occupants may pose serious risks for colonization in general, this hypothesis cannot be confirmed for the data sets we have analyzed. Further models that explore bedwise environmental contamination in more detail would constitute interesting extensions of our methodology.

The results of our analysis build on a data-augmented MCMC algorithm [27, 30]. Markov chain Monte Carlo sampling is a powerful tool to estimate posterior parameter distribution whenever the likelihood is analytically intractable. And yet, the inherent disadvantage of this sampling scheme is that it may take prohibitively many iterations to converge to the posterior distribution. The convergence properties of MCMC sampling in high-dimensional posterior distributions can be particularly problematic and sensitive to the choice of prior and proposal distributions. Thus, tuning of the MCMC parameters becomes crucial for its application. Our developed full model is overparametrized when colonizations of patients are not or hardly influenced by environmental contamination. As a result, the respective likelihood might not be identifiable or only weakly identifiable. Here, a flat prior for the decay rate μ in combination with a small initial standard deviation for its proposal distribution resulted in large acceptance ratios close to 1. The MCMC chain mixed too slowly and therefore hindered the identifiability of the likelihood. We were able to tune the parameters of the proposal distribution for μ such that rapid convergence to the posterior distribution could be assessed using visual inspection of histograms and trace plots. However, as presented in the section, pairwise scatter plots showed strong cor-

relations in particular between ϵ and μ. Simulation studies confirmed that this can be explained by an absence of environmental contamination in the investigated data sets. This supports our finding that an impact of environmental contamination after discharge on the transmission of *P. aeruginosa* may be neglected.

Moreover, colonization is assumed to remain until discharge. While this assumption is true for *P. aeurginosa* it does not hold true for all antibiotic-resistant nosocomial pathogens. However, intermittent carriage may be readily included allowing the method to be generalized to other pathogens.

We assumed no difference in transmissibility between different strains of *P. aeruginosa* and that all colonized patients are equally likely to transmit the pathogen. While information on antibiotic resistance or microbial genotyping in combination with epidemiological data may aid in distinguishing different strains and identifying specific transmission events, only the uncertainty of the estimates would be affected. In particular, the widths of the credibility intervals are likely to be reduced, but we do not expect a large effect on the parameter estimates.

Assessing the fit of the model to the data is crucial to model building. The true relative importance of the different transmission routes in ICUs is generally unknown. Genotyping data that might be used to demonstrate the source of the acquired colonization is scarce and was not available for the data used in our analysis. While the posterior predictive *p*-value is a popular method for assessing model fit, it has been increasingly criticized for its self-fulfilling nature [36]. Furthermore, the choice of the test statistic is crucial in order to adequately summarize discrepancies between datasets. Rather than relying on a suitable summary statistic, we presented a model assessment method that evaluates whether the estimated force of infection adequately represents the transmission dynamics in the ward. However, while the corresponding coverage probabilities may depict discrepancies per bin of the force of infection, the sample size is not controlled by choosing the number MCMC updates. It might well occur that specific patients (and their acquisition events) appear in more than one MCMC update simultaneously. Thus, the true sample size is estimated to be smaller. In addition, both the estimated force of infection and the number of acquisitions N_{acq} are obtained based on the data augmentation step. Thus, the theoretical probability of acquisition and the predicted fraction of acquisition are not independent. And yet, a large deviation of the model from the data would be reflected in the coverage probabilities since the augmented data is dependent on the observed data.

Further improvement of the method presented here or development of other methods would be a vital topic for assessing epidemic models.

Model selection was performed using the DIC which is known to display poor performance (i.e. identifying the correct model) for complex likelihood functions such as those corresponding to epidemic models. Comparing the plausibility of different models is crucial for selecting the model that describes the dynamics of the observed system best. Nevertheless, model choice for stochastic epidemic models is far from trivial. All known approaches for model selection exhibit advantages as well as disadvantages [36] which makes selecting the most suitable model comparison technique not straightforward. We selected the well-known DIC-method that was easy to use and implement. Our main results regarding environmental contamination after discharge do not depend on the model choice. And yet, the development of a suitable and robust model selection procedure in a data-augmented Bayesian framework would be an interesting and important topic for future research.

Finally, like all models, ours is a simplification of the truth as it is unlikely that all relevant variables are already included. Adding covariates such as antibiotic use, sex or age may improve the model fit.

Our work may be used or further extended for assessing the relative importance of different transmission routes within intensive-care units not only for *P. aeruginosa* but for hospital pathogens in general. Based on these results, consequential decisions for tailored interventions or policies may be deduced, aiding in improving infection prevention and control and therefore reducing morbidity, mortality and related costs in hospitals.

Declaration of interests

We declare that we have no conflicts of interest.

Acknowledgments

References

[1] Cassini, A, Plachouras, D, Eckmanns, T, et al. Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study. In: *PLoS Medicine* 13.10 (Oct. 2016), e1002150. doi: `10.1371/journal.pmed.1002150`.

[2] Pittet, D, Allegranzi, B, Storr, J, et al. *Infection control as a major World Health Organization priority for developing countries*. Apr. 2008. doi: `10.1016/j.jhin.2007.12.013`.

[3] Jarvis, WR and Martone, WJ. Predominant pathogens in hospital infections. In: *Journal of Antimicrobial Chemotherapy* 29.suppl_A (Jan. 1992), pp. 19–24. doi: `10.1093/jac/29.suppl{\_}A.19`.

[4] Spencer, RC. Predominant pathogens found in the European prevalence of infection in intensive care study. In: *European Journal of Clinical Microbiology and Infectious Diseases* 15.4 (1996), pp. 281–285. doi: `10.1007/BF01695658`.

[5] Tantracheewathorn, T, Vititpatarapak, N, and Phumisantiphong, U. Epidemiologic study of nosocomial bacterial infection of pediatric patients at BMA Medical College and Vajira Hospital. In: *Journal of the Medical Association of Thailand* 90.2 (2007), pp. 258–265.

[6] Kerr, KG and Snelling, AM. Pseudomonas aeruginosa: a formidable and ever-present adversary. In: *Journal of Hospital Infection* 73.4 (Dec. 2009), pp. 338–344. doi: `10.1016/J.JHIN.2009.04.020`.

[7] Tacconelli, E, Carrara, E, Savoldi, A, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. In: *The Lancet Infectious Diseases* 18.3 (Mar. 2018), pp. 318–327. doi: `10.1016/S1473-3099(17)30753-3`.

[8] Blanc, DS, Francioli, P, and Zanetti, G. Molecular Epidemiology of Pseudomonas aeruginosa in the Intensive Care Units - A Review. In: *The Open Microbiology Journal* 1.1 (Dec. 2007), pp. 8–11. doi: `10.2174/1874285800701010008`.

[9] Gruner, E, Kropec, A, Huebner, J, et al. Ribotyping of Pseudomonas aeruginosa Strains Isolated from Surgical Intensive Care Patients. In: *Journal of Infectious Diseases* 167.5 (May 1993), pp. 1216–1220. doi: `10.1093/infdis/167.5.1216`.

[10] Kropec, A, Huebner, J, Riffel, M, et al. Exogenous or endogenous reservoirs of nosocomial Pseudomonas aeruginosa and Staphylococcus aureus infections in a surgical intensive care unit. In: *Intensive Care Medicine* 19.3 (Mar. 1993), pp. 161–165. doi: `10.1007/BF01720533`.

[11] Bonten, MJ, Bergmans, DC, Speijer, H, et al. Characteristics of polyclonal endemicity of Pseudomonas aeruginosa colonization in intensive care units: Implications for infection control. In: *American Journal of Respiratory and Critical Care Medicine* 160.4 (Dec. 1999), pp. 1212–1219. doi: `10.1164/ajrccm.160.4.9809031`.

[12] Speijer, H, Savelkoul, PH, Bonten, MJ, et al. Application of different genotyping methods for Pseudomonas aeruginosa in a setting of endemicity in an intensive care unit. In: *Journal of Clinical Microbi-*

ology 37.11 (1999), pp. 3654–3661. doi: 10.1128/jcm.37.11.3654-3661.1999.

[13] Berthelot, P, Grattard, F, Mahul, P, et al. Prospective study of nosocomial colonization and infection due to Pseudomonas aeruginosa in mechanically ventilated patients. In: *Intensive Care Medicine* 27.3 (Mar. 2001), pp. 503–512. doi: 10.1007/s001340100870.

[14] Bergmans, DC, Bonten, MJ, Van Tiel, FH, et al. Cross-colonisation with Pseudomonas aeruginosa of patients in an intensive care unit. In: *Thorax* 53.12 (Dec. 1998), pp. 1053–1058. doi: 10.1136/thx.53.12.1053.

[15] Bertrand, X, Thouverez, M, Talon, D, et al. Endemicity, molecular diversity and colonisation routes of Pseudomonas aeruginosa in intensive care units. In: *Intensive Care Medicine* 27.8 (July 2001), pp. 1263–1268. doi: 10.1007/s001340100979.

[16] Thuong, M, Arvaniti, K, Ruimy, R, et al. Epidemiology of Pseudomonas aeruginosa and risk factors for carriage acquisition in an intensive care unit. In: *Journal of Hospital Infection* 53.4 (Apr. 2003), pp. 274–282. doi: 10.1053/jhin.2002.1370.

[17] Vallés, J, Mariscal, D, Cortés, P, et al. Patterns of colonization by Pseudomonas aeruginosa in intubated patients: A 3-year prospective study of 1,607 isolates using pulsed-field gel electrophoresis with implications for prevention of ventilator-associated pneumonia. In: *Intensive Care Medicine* 30.9 (Sept. 2004), pp. 1768–1775. doi: 10.1007/s00134-004-2382-6.

[18] Lashéras, A, Guisset, O, Boulestreau, H, et al. Réservoirs et transmission de Pseudomonas aeruginosa en réanimation médicale. In: *Medecine et Maladies Infectieuses* 36.2 (Feb. 2006), pp. 99–104. doi: 10.1016/j.medmal.2005.11.014.

[19] Agodi, A, Barchitta, M, Cipresso, R, et al. Pseudomonas aeruginosa carriage, colonization, and infection in ICU patients. In: *Intensive Care Medicine* 33.7 (July 2007), pp. 1155–1161. doi: 10.1007/s00134-007-0671-6.

[20] Kramer, A, Schwebke, I, and Kampf, G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. In: *BMC Infectious Diseases* 6.1 (Aug. 2006), pp. 1–8. doi: 10.1186/1471-2334-6-130.

[21] Loveday, HP, Wilson, JA, Kerr, K, et al. Association between healthcare water systems and Pseudomonas aeruginosa infections: A rapid systematic review. In: *Journal of Hospital Infection* 86.1 (Jan. 2014), pp. 7–15. doi: 10.1016/j.jhin.2013.09.010.

[22] McBryde, ES and McElwain, DL. A mathematical model investigating the impact of an environmental reservoir on the prevalence and control of vancomycin-resistant enterococci. In: *Journal of Infectious Diseases* 193.10 (May 2006), pp. 1473–1474. doi: 10.1086/503439.

[23] Wolkewitz, M, Dettenkofer, M, Bertz, H, et al. Environmental Contamination as an Important Route for the Transmission of the Hospital Pathogen VRE: Modeling and Prediction of Classical Interventions. In: *Infectious Diseases: Research and Treatment* 1 (July 2008), IDRT.S809. doi: 10.4137/idrt.s809.

[24] Wang, X, Xiao, Y, Wang, J, et al. A mathematical model of effects of environmental contamination and presence of volunteers on hospital infections in China. In: *Journal of Theoretical Biology* 293 (Jan. 2012), pp. 161–173. doi: 10.1016/j.jtbi.2011.10.009.

[25] Plipat, N, Spicknall, IH, Koopman, JS, et al. The dynamics of methicillin-resistant Staphylococcus aureus exposure in a hospital model and the potential for environmental intervention. In: *BMC Infec-*

tious Diseases 13.1 (Dec. 2013), pp. 1–11. doi: `10.1186/1471-2334-13-595`.

[26] Wang, L and Ruan, S. Modeling Nosocomial Infections of Methicillin-Resistant Staphylococcus aureus with Environment Contamination. In: *Scientific Reports* 7.1 (Apr. 2017), pp. 1–12. doi: `10.1038/s41598-017-00261-1`.

[27] Cooper, BS, Medley, GF, Bradley, SJ, et al. An augmented data method for the analysis of nosocomial infection data. In: *American Journal of Epidemiology* 168.5 (Sept. 2008), pp. 548–557. doi: `10.1093/aje/kwn176`.

[28] Diekmann, O, Heesterbeek, H, and Britton, T. Mathematical tools for understanding infectious diseases dynamics. In: (2013).

[29] Britton, T and Andersson, H. Stochastic Epidemic Models and Their Statistical Analysis. Lecture Notes in Statistics. New York, NY: Springer New York, 2000, pp. IX, 156. doi: `10.1198/jasa.2002.s483`.

[30] Worby, CJ, Jeyaratnam, D, Robotham, JV, et al. Estimating the effectiveness of isolation and decolonization measures in reducing transmission of methicillin-resistant staphylococcus aureus in hospital general wards. In: *American Journal of Epidemiology* 177.11 (June 2013), pp. 1306–1313. doi: `10.1093/aje/kws380`.

[31] Lavenir, R, Jocktane, D, Laurent, F, et al. Improved reliability of Pseudomonas aeruginosa PCR detection by the use of the species-specific ecfX gene target. In: *Journal of Microbiological Methods* 70.1 (July 2007), pp. 20–29. doi: `10.1016/j.mimet.2007.03.008`.

[32] Gelman, A, Carlin, JB, Stern, HS, et al. Bayesian Data Analysis. Chapman and Hall/CRC, Nov. 2013. doi: `10.1201/b16018`.

[33] Brown, LD, Cai, TT, and Das Gupta, A. Interval estimation for a binomial proportion. In: *Statistical Science* 16.2 (May 2001), pp. 101–117. doi: `10.1214/ss/1009213286`.

[34] Panagea, S, Winstanley, C, Walshaw, MJ, et al. Environmental contamination with an epidemic strain of Pseudomonas aeruginosa in a Liverpool cystic fibrosis centre, and study of its survival on dry surfaces. In: *Journal of Hospital Infection* 59.2 (Feb. 2005), pp. 102–107. doi: `10.1016/j.jhin.2004.09.018`.

[35] Trautmann, M, Lepper, PM, and Haller, M. *Ecology of Pseudomonas aeruginosa in the intensive care unit and the evolving role of water outlets as a reservoir of the organism.* June 2005. doi: `10.1016/j.ajic.2005.03.006`.

[36] Gibson, GJ, Streftaris, G, and Thong, D. Comparison and assessment of epidemic models. In: *Statistical Science* 33.1 (Feb. 2018), pp. 19–33. doi: `10.1214/17-STS615`.

[37] R Core Team. *R: A Language and Environment for Statistical Computing.* 2020. `https://www.r-project.org/`.

Supplementary material

Table of Contents

S1 Fig: Pairwise plots of samples from the posterior distribution for the transmission parameters of the submodel.

S2 Fig: Histograms for ICU A before renovation using the submodel with background and cross-transmission.

S3 Fig: Traceplots for ICU A before renovation using the submodel with background and cross-transmission.

S4 Fig: Histograms for ICU A after renovation using the submodel with background and cross-transmission.

S5 Fig: Traceplots for ICU A after renovation using the submodel with background and cross-transmission.

S6 Fig: Histograms for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S7 Fig: Histograms for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in$ {background, cross-transmission, environment}.

S8 Fig: Traceplots for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S9 Fig: Traceplots for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in$ {background, cross-transmission, environment}.

S10 Fig: Histograms for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S11 Fig: Histograms for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation

probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in$ {background, cross-transmission, environment}.

S12 Fig: Traceplots for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S13 Fig: Traceplots for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in$ {background, cross-transmission, environment}.

S14 Fig: Pairwise plots of samples from the posterior distribution for the transmission parameters of the full model. The plots were generated from the data of ICU A before renovation using Exp(0.001) prior and the full model with background, cross-transmission and environmental contamination.

S15 Fig: Coverage probabilities for the submodel using Jeffreys prior. (a) - (b) ICU A before and after renovation, respectively. (c) - (d) ICU B before and after renovation, respectively.

S16 Fig: Coverage probabilities for the full model using Jeffreys prior. (a) - (b) ICU A before and after renovation, respectively. (c) - (d) ICU B before and after renovation, respectively.

S17 Fig: Pairwise plots of samples from the posterior distribution for the transmission parameters of the full model. The plots were generated from the data of ICU A using U(0,2) prior and the full model with background, cross-transmission and environmental contamination after discharge.

S18 Fig: Coverage probabilities for simulated data set using Jeffreys prior. The data was simulated using $\alpha = 0.015$, $\beta = 0.055$, $\mu = 1/7$, $\epsilon = 0.15$, $f = 0.05$, $\phi = 1$. The analysis assumed only one route, i.e. background transmission. The plot shows large discrepancies between the expected and the computed coverage probabilities, pointing to a misspecified model.

S1 Text. Environmental contamination.

The full model includes environmental contamination on a ward-level. The bacterial load at any given time t is based on the differential equation

$$\frac{dE}{dt} = \nu \frac{I(t)}{N(t)} - \mu E(t) \tag{3.7}$$

Solving the above differential equation requires discretizing over t, resulting in a finite number of time steps $t_0, t_1, \ldots, t_N$. We then assume $I(t) = I_t$ and $N(t) = N_t$ to be constant within a time step and use it as initial conditions. Separating variables leads to

$$\frac{dE}{\nu \cdot \frac{I_t}{N_t} - \mu E(t)} = dt$$

and thus

$$\int \frac{dE}{\nu \frac{I_t}{N_t} - \mu E(t)} = \int dt$$

$$\Rightarrow -\frac{1}{\mu} \log \left| \nu \frac{I_t}{N_t} - \mu E(t) \right| = t + C$$

$$\Rightarrow \log \left| \nu \frac{I_t}{N_t} - \mu E(t) \right| = -\mu(t + C)$$

$$\Rightarrow \left| \nu \frac{I_t}{N_t} - \mu E(t) \right| = \exp[-\mu(t + C)] = \exp(-\mu t) \underbrace{\exp(-\mu C)}_{A_t}$$

Now, two cases have to be distinguished.

1. Case: $\nu \frac{I_t}{N_t} - \mu E(t) \geq 0$

$$\Rightarrow -\mu E(t) = A_t \cdot \exp[-\mu t] - \nu \frac{I_t}{N_t}$$

$$\Rightarrow E(t) = \underbrace{-\frac{A_t}{\mu}}_{B_t} \exp[-\mu t] + \frac{\nu}{\mu} \frac{I_t}{N_t}$$

$$\Rightarrow E(t) = B_t \cdot \exp[-\mu t] + \frac{\nu}{\mu} \frac{I_t}{N_t}.$$

2. Case: $\nu \frac{I_t}{N_t} - \mu E(t) < 0$

$$\Rightarrow \mu E(t) = A_t \cdot \exp[-\mu t] + \nu \frac{I_t}{N_t}$$

$$\Rightarrow E(t) = \underbrace{\frac{A_t}{\mu}}_{B_t} \exp[-\mu t] + \frac{\nu}{\mu} \frac{I_t}{N_t}$$

$$\Rightarrow E(t) = B_t \cdot \exp[-\mu t] + \frac{\nu}{\mu} \frac{I_t}{N_t}.$$

Determine B_{t_0} for initial condition $E(t_0) = E_{t_0}$:

$$E_{t_0} = E(t_0) = B_0 + \frac{\nu}{\mu} \frac{I_{t_0}}{N_{t_0}}$$

and therefore

$$B_{t_0} = E_{t_0} - \frac{\nu}{\mu} \frac{I_{t_0}}{N_{t_0}}. \tag{3.8}$$

For $t_0 \leq t \leq t_1$ the environmental load can be then computed by

$$E(t) = \left(E_{t_0} - \frac{\nu}{\mu} \frac{I_{t_0}}{N_{t_0}} \right) e^{-\mu t} + \frac{\nu}{\mu} \frac{I_{t_0}}{N_{t_0}}$$

$$= E_{t_0} e^{-\mu t} + \frac{\nu}{\mu} \frac{I_{t_0}}{N_{t_0}} (1 - e^{-\mu t}).$$

For $t_0 \leq t_i \leq t_N$ it holds

$$B_{t_i} = \frac{E_{t_i} - \frac{\nu}{\mu} \frac{I_{t_i}}{N_{t_i}}}{e^{-\mu t_i}} \tag{3.9}$$

and therefore, it holds for $\lfloor t \rfloor := \max\{t_0 \leq x \leq t_N \,|\, x \leq t\}$ and $t \in \mathbb{R} \setminus \{t_0, t_1, \ldots, t_N\}$

$$E(t) = \frac{E_{\lfloor t \rfloor} - \frac{\nu}{\mu} \frac{I_{\lfloor t \rfloor}}{N_{\lfloor t \rfloor}}}{e^{-\mu \lfloor t \rfloor}} \cdot e^{-\mu t} + \frac{\nu}{\mu} \frac{I_{\lfloor t \rfloor}}{N_{\lfloor t \rfloor}}$$

$$= E_{\lfloor t \rfloor} e^{-\mu(t - \lfloor t \rfloor)} + \frac{\nu}{\mu} \frac{I_{\lfloor t \rfloor}}{N_{\lfloor t \rfloor}} \left(1 - e^{-\mu(t - \lfloor t \rfloor)} \right).$$

and

$$E(t_i) = E_{t_{i-1}} e^{-\mu(t_i - t_{i-1})} + \frac{\nu}{\mu} \frac{I_{t_{i-1}}}{N_{t_{i-1}}} \left(1 - e^{-\mu(t_i - t_{i-1})} \right) \quad \text{for } 0 \leq i \leq N.$$

S2 Text. Discrete-time transmission model.

For the discrete-time transmission model, we assume that the number of colonized patients $I(t)$, the total number of patients $N(t)$ and the bacterial load $E(t)$ is constant during the day. It is assumed that admission and screening occur at 12:00 pm on each day T determining I_T and N_T. Given all the information (at 12:00 pm), the environmental contamination on day T is determined. The force of infection on day T is then given by

$$\lambda(T) = \alpha + \beta \frac{I_T}{N_T} + \epsilon E(T)$$
$$= \alpha + \beta \frac{I_T}{N_T} + \epsilon \cdot \left[E_{T-1} e^{-\mu} + \frac{\nu}{\mu} \frac{I_{T-1}}{N_{T-1}} (1 - e^{-\mu}) \right].$$

S3 Text. Relative contribution.

The computations in section *Relative contributions of transmission routes* were developed for continuous-time models. In our discrete-time model, we assume that events such as, admission, colonization and discharge of patients and screening occur on a daily basis. However, we do assume that the level of environmental contamination changes continuously. Computing the relative contributions of the different transmission routes becomes more laborious in this scenario. Let t_i^c be the acquisition time of patient $i \in \{1, \ldots, n\}$. The contribution of a route j is the ratio of the probability that the acquisition was due to route j and the total probability of acquisition:

$$\text{Contribution of route } j = R_j = \frac{\sum_{i=1}^{n} \frac{P(\text{infection during day } t_i^c \text{ due to route } j)}{P(\text{infection during day } t_i^c)}}{N_{\text{acq}}}$$

where N_{acq} is the total number of occured colonizations and R_j with $j \in \{\text{background}, \text{crossT}, \text{env}\}$ indicate the endogenous, cross-transmission or environmental route, respectively. The route-specific probabilities can be determined by

$$P(\text{infection during day } T \text{ by } R_{\text{background}}) = \int_{T}^{T+1} P(\text{patient still susceptible at time } t) \cdot \alpha \, dt$$

$$P(\text{infection during day } T \text{ by } R_{\text{crossT}}) = \int_{T}^{T+1} P(\text{patient still susceptible at time } t) \cdot \beta \frac{I_T}{N_T} \, dt$$

$$P(\text{infection during day } T \text{ by } R_{\text{env}}) = \int_{T}^{T+1} P(\text{patient still susceptible at time } t) \cdot \epsilon E(t) \, dt$$

where environmental contamination during a patient's stay is assigned to the environmental route. In the main part of our manuscript we consider only the bacterial load remaining after discharge as environmental contamination. All the formulas then change according to Eq. (3) of the main text.

The results are dependent on $\nu \frac{I_T}{N_T} - \mu E(T)$ and are given by

1. Case: $\nu \frac{I_T}{N_T} - \mu E(T) \geq 0$

$$P(\text{infection during day } T \text{ by } R_{\text{background}}) = \alpha \cdot \tilde{E} \cdot \left[\gamma \left(\frac{A}{\mu}, \tilde{B} \right) - \gamma \left(\frac{A}{\mu}, \tilde{B} e^{-\mu} \right) \right]$$

$$P(\text{infection during day } T \text{ by } R_{\text{crossT}}) = \beta \frac{I_T}{N_T} \cdot \tilde{E} \cdot \left[\gamma \left(\frac{A}{\mu}, \tilde{B} \right) - \gamma \left(\frac{A}{\mu}, \tilde{B} e^{-\mu} \right) \right]$$

$$P(\text{infection during day } T) = 1 - e^{-A - \tilde{B} \left(e^{-\mu} - 1 \right)}$$

2. Case: $\nu \frac{I_T}{N_T} - \mu E(T) < 0$

$$P(\text{infection during day } T \text{ by } R_{\text{background}}) = \alpha \cdot \hat{E} \cdot \left[\sum_{i=0}^{\infty} \frac{(-\tilde{B})^i}{i!} \cdot \frac{1}{A/\mu + i} \cdot \left(1 - e^{-A - \mu i} \right) \right]$$

$$P(\text{infection during day } T \text{ by } R_{\text{crossT}}) = \beta \frac{I_T}{N_T} \cdot \hat{E} \cdot \left[\sum_{i=0}^{\infty} \frac{(-\tilde{B})^i}{i!} \cdot \frac{1}{A/\mu + i} \cdot \left(1 - e^{-A - \mu i} \right) \right]$$

$$P(\text{infection during day } T) = 1 - e^{-A - \tilde{B} \left(e^{-\mu} - 1 \right)}$$

and

$$P(\text{infection during day } T \text{ by } R_{\text{background}})$$
$$= P(\text{infection during day } T) - \sum_{k \in \{\alpha, \beta\}} P(\text{infection during day } T \text{ by } k)$$

where

$$A = \alpha + \left(\beta + \frac{\epsilon \nu}{\mu} \right) \frac{I_T}{N_T}, \qquad \tilde{B} = \frac{\epsilon}{\mu} \left(\frac{\nu}{\mu} \frac{I_T}{N_T} - E_T \right)$$

$$\tilde{E} = \frac{1}{\mu} \cdot e^{\tilde{B}} \cdot \tilde{B}^{-\frac{A}{\mu}}, \qquad \hat{E} = \frac{e^{\tilde{B}}}{\mu},$$

$$G = \frac{\epsilon \nu}{\mu} \frac{I_T}{N_T}$$

and $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function. Note that the derivations are omitted here but can be requested from the first author.

S4 Text. Approximation of relative contribution in discrete-time.

Large values of the force of infection $\lambda(t)$ are very unlikely. Under the assumption of small $\lambda(t)$, the following simplifications and approximations can be made:

$$e^{-\int_0^t \lambda(x)dx} \approx 1 - \lambda(t)$$

$$1 - e^{-\int_0^t \lambda(x)dx} \approx \lambda(t)$$

$$\frac{\lambda(t)}{1 - \lambda(t)} \approx \lambda(t).$$

Therefore, the force of infection itself may be a good approximation of the probability of infection and the probability of acquiring colonization due to route j may be approximated by the respective sub-term of the force of infection assigned to route j:

$$P(\text{infection during day } T \text{ due to route } j) \approx \lambda_j$$

with $j \in \{\text{background}, \text{crossT}, \text{env}\}$. As an approximation of the relative contribution we compute the ratio of the transmission rate and the force of infection for each acquired colonization:

- Contribution of endogenous route $= R_{\text{background}} = \dfrac{\sum_{i=1}^{n} \frac{\alpha}{\lambda(t_i^c)}}{N_{\text{acq}}}$

- Contribution of cross-transmission $= R_{\text{crossT}} = \dfrac{\sum_{i=1}^{n} \frac{\beta \cdot \frac{I(t_i^c)}{N(t_i^c)} + \epsilon \sum_{i_p} \frac{E_{i_p}(t_i^c)}{N(t_i^c)}}{\lambda(t_i^c)}}{N_{\text{acq}}}$

- Contribution of environmental contamination $= R_{\text{env}} = \dfrac{\sum_{i=1}^{n} \frac{\epsilon \left[\sum_{i_d} \frac{E_{i_d}(t_i^c)}{N(t_i^c)} + E_0 e^{-\mu t_i^c} \right]}{\lambda(t_i^c)}}{N_{\text{acq}}}$

where t_i^c is the day of colonization of patient where $i \in \{1, \ldots, n\}$ and N_{acq} the total number of occured colonizations. Furthermore, i_p indicates a colonized patient that is present at time t_i^c and i_d a colonized patient that has been colonized prior to t_i^c but was already discharged. It holds $R_{\text{background}} + R_{\text{crossT}} + R_{\text{env}} = 1$.

S5 Text. Adapted data-augmented MCMC algorithm.

We model the transmission process using a two-state Markov model, where each patient can be either *susceptible* (*P. A.* negative) or *colonized* (*P. A.* positive). A patient i is admitted to the ICU on day t_i^a and discharged on day t_i^d. The probability that a patient is admitted already colonized is described by parameter f. The rate at which a susceptible patient transitions to being colonized is described in section *Materials and Methods*. The colonization state of an individual patient is determined from screening information. We suppose that for each patient i a set of screening results $X_i = X_i^{(1)}, \ldots, X_i = X_i^{(m)}$, taken on days $t_i^{(1)}, \ldots, t_i^{(m)}$ is available. The set of all screening results is denoted by $X = \{X_1, \ldots, X_N\}$ where N is the total number of patients. Since screening tests are typically intermittent and imperfect, we define the test sensitivity ϕ (i.e. probability that a colonized patient has a positive result). We assume that the specificity (i.e. probability that an uncolonized patient has a negative result) is 100%.

We implemented an adapted version of the data-augmented MCMC algorithm to analyze the data. The transmission and importation model, as well as the data-augmentation method is closely based on the approach of [27, 30] but adapted for the transmission routes presented in this paper.

The algorithm was implemented in C++ and the analysis of the output was performed in R (Version 3.5.1) [37].

The aim of our analysis was to estimate the set of parameters $\theta = \{\alpha, \beta, \epsilon, \mu, f, \phi\}$. The prior distribution were chosen as follows:

$$f, \phi \propto \mathrm{Beta}(a, b)$$

$$\alpha, \beta, \epsilon, \mu \propto \mathrm{Exp}(\lambda)$$

where $\mathrm{Exp}(\lambda)$ represents the exponential distribution with rate λ, and $\mathrm{Beta}(a, b)$ the beta distribution with shape parameters a and b. Having fixed $a = b = 1$ and $\lambda = 0.001$, we use uninformative priors in our analysis.

The data-augmentation procedure accounts for unobserved colonization times by augmenting the parameter space with $A = \{t^c, s^a\}$, a set comprising of the unobserved

colonization times t^c and admission states s_a of all n patients. An admission state of a patient is 1 if the patient is colonized upon admission and 0 otherwise. If the patient j becomes colonized during his/her stay, the colonization time may take an integer value between the time of admission t_j^a and time of discharge t_j^d (inclusive). If a patient does not acquire colonization, the respective value t_j^c takes a dummy value of -1. The augmented posterior density relation can be determined using Bayes' Theorem:

$$P(A, \theta \mid D) \propto P(D, A, \theta) = P(D \mid A, \theta) P(A \mid \theta) P(\theta) \tag{3.10}$$

$$= P(D \mid t^c, s^a, \theta) P(s^a \mid \theta) P(t^c \mid s^a, \theta) P(\theta) \tag{3.11}$$

where $P(D \mid A, \theta)$ is the likelihood of the observed data D, $P(A \mid \theta)$ is the likelihood of the augmented data and $P(\theta)$ is the joint prior distribution of the parameter set θ. All terms in (3.10) can be explicitly calculated. It holds

$$P(D \mid t^c, s^a, \theta) = \phi^{TP(X)} (1 - \phi)^{FN(X, t^c)}$$

where $TP(X)$ and $FN(X, A)$ are the total number of true positive and false negative swab results, given the colonization times t^c, respectively. It represents the imperfect observation of the transmission dynamics. Assuming that lost colonization can be excluded, we consider any negative result after the time of colonization as a false negative. Since false positive results are impossible, the $TP(X)$ is not dependent on the augmented data and can be determined directly from the observed data. The probability of the set of importations, given the importation probability f is given by

$$P(s^a \mid \theta) = f^{\sum_i s_i^a} (1 - f)^{n - \sum_i s_i^a}.$$

The transmission model itself is reflected in the probability of the colonization times given the admission states and the parameters

$$P(t^c \mid s^a, \theta) = \prod_{i=1}^{n} \exp\left(-\sum_{t=t_i^a}^{\min(t_i^d, t_i^c - 1)} \lambda(t)\right) \left(\prod_{j: t_j^c \neq -1} (1 - \exp(-q(t_j^c)))\right) \cdot f^{\sum_i s_i^a} (1 - f)^{n - \sum_i s_i^a}.$$

$$\tag{3.12}$$

To update the importation rate f and the sensitivity ϕ, we use Gibbs sampling as we can sample directly from the full conditional distributions. The transmission parameters

$\alpha, \beta, \epsilon, \mu$ are updated using an adapted version of the Metropolis-Hastings algorithm. Regular MCMC methods based on the Metropolis-Hastings algorithm tend to be very slow in high dimensions as a result of slow mixing and therefore inefficient convergence towards the target distribution. In high-dimensional spaces the volume outside is much larger than the volume of our target distribution. Thus, traditional MCMC methods such as the Metropolis-Hastings algorithm, spend considerable amount of time of traversing space away from the mode of the target distribution. Our adapted MCMC algorithm aims in exploring the target distribution more efficiently.

The Metropolis-Hastings algorithm generates a Markov chain $\theta^{(1)}, \ldots, \theta^{(N)}$ which converges to a target distribution $\pi(\cdot)$ if N is large enough. In each update of the Markov chain, a candidate point, θ^* is sampled from a proposal density $q(\theta^* \,|\, \theta^{(i)})$, which gives the probability density of proposing θ^*, given the current, i^{th} value. With a certain probability or so-called acceptance ratio

$$a(\theta^*, \theta^{(i)}) = \min\left(1, \frac{q(\theta^* \,|\, \theta^{(i)})\pi(\theta^*)}{q(\theta^{(i)} \,|\, \theta^*)\pi(\theta^{(i)})}\right),$$

the proposed value is accepted.

The Metropolis algorithm is a special case of the Metropolis-Hastings algorithm where the proposal function is symmetrical. Since a symmetrical proposal distribution simplifies the calculation of the acceptance ratio to $a(\theta^*, \theta^{(i)}) = \min\left(1, \pi(\theta^*)/\pi(\theta^{(i)})\right)$, it is often used for updating parameters. The proposal function has a great influence on the speed of convergence and hence efficiency of the algorithm. We suggest a proposal distribution that speeds up the convergence towards the target distribution while limiting the additional computational effort. The idea behind our method is as follows: For each estimated parameter set θ there is a corresponding force of infection $\lambda(t)$ for each time t. It can be assumed that the mean force of infection $\overline{\lambda}$ is approximately constant over the number of iterations. The rationale behind it is that there is *true* mean force of infection that should be approximated by the MCMC algorithm. Proposing new parameter candidates depending on the mean force of infection reduces the volume that has to be traversed in order to converge to the target distribution. The resulting proposal density is not symmetric anymore and thus the procedure requires an adjustment of the acceptance ratio. The adapted Metropolis-Hastings algorithm to

update the transmission parameters runs as follows:

Two transmission routes

1. Set initial values $\theta^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}\right)$, and the number of iterations N.

2. Sample new parameter values α^*, β^* as follows:

   (a) Propose candidate α^* by sampling from $\alpha^{(i)} + \mathcal{N}(0, \sigma_\alpha^2)$

   (b) Propose candidate β^* assuming $\lambda^* \overset{!}{\approx} \lambda^{(i)} = \alpha^{(i)} + \beta^{(i)} \cdot \overline{Prev}$:

   Sample β^* from $\frac{\lambda^{(i)} - \alpha^*}{\overline{Prev}} + \mathcal{N}(0, \sigma_\beta^2)$, i.e. $\mathcal{N}\left(\frac{\lambda}{\overline{Prev}} - \frac{\alpha^*}{\overline{Prev}}, \frac{\sigma_\alpha^2}{\overline{Prev}^2} + \sigma_\beta^2\right)$

   (c) With probability

   $$\alpha(\theta^*, \theta^{(i)}) = \min\left(1, \frac{q(\theta^* \mid \theta^{(i)})\pi(\theta^*)}{q(\theta^{(i)} \mid \theta^*)\pi(\theta^{(i)})}\right)$$

   where $\frac{q(\theta^* \mid \theta^{(i)})}{q(\theta^{(i)} \mid \theta^*)} = e^{\frac{\beta^{*2} - \beta^{(i)^2} + \mu_\beta(\beta^{(i)} - \beta^*)}{2\sigma_\beta^2}}$, accept the proposed value and set $\theta^{(i+1)} = \theta^*$, else set $\theta^{(i+1)} = \theta^{(i)}$.

3. If $i < N$, then go to step 2.

Three transmission routes

1. Set initial values $\theta^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}, \epsilon^{(0)}, \mu^{(0)}\right)$, and the number of iterations N.

2. Sample new parameters $\theta^* = (\alpha^*, \beta^*, \epsilon^*, \mu^*)$ from a proposal density $q\left(\theta^* \mid \theta^{(i)}\right)$ as follows:

   (a) Propose candidate α^* by sampling from $\alpha^{(i)} + \mathcal{N}(0, \sigma_\alpha^2)$

   (b) Propose candidate β^* by sampling from $\beta^{(i)} + \mathcal{N}(0, \sigma_\beta^2)$

   (c) Propose candidate μ_1 by sampling from $\mu^{(i)} + \mathcal{N}(0, \sigma_\mu^2)$

   (d) Update $E_0^{(i+1)}$ to $\frac{\nu}{\mu^{(i)}} \cdot \overline{Prev}$

   (e) Propose candidate ϵ^* assuming $\lambda^* \overset{!}{\approx} \lambda^{(i)} = \alpha^{(i)} + \beta^{(i)} \cdot \overline{Prev} + \epsilon^{(i)} \cdot E_0^{(i)}$:
   Sample ϵ^* from $\frac{\lambda^{(i)} - \alpha^* - \beta^* \cdot \overline{Prev}}{E^{(i+1)}} + \mathcal{N}(0, \sigma_\epsilon^2)$,
   i.e. $\mathcal{N}\left(\frac{\lambda^{(i)} - \alpha^* - \beta^* \cdot \overline{Prev}}{E^{(i+1}}, \frac{\sigma_\alpha^2}{E^{(i+1)^2}} + \frac{\sigma_\beta^2}{\left(\frac{E^{(i+1)}}{\overline{Prev}}\right)^2} + \sigma_\epsilon^2\right)$.

(f) With probability

$$\alpha(\theta^*, \theta^{(i)}) = \min\left(1, \frac{q(\theta^* \mid \theta^{(i)})\pi(\theta^*)}{q(\theta^{(i)} \mid \theta^*)\pi(\theta^{(i)})}\right)$$

where $\frac{q(\theta^* \mid \theta^{(i)})}{q(\theta^{(i)} \mid \theta^*)} = e^{\frac{\epsilon^{*2} - \epsilon^{(i)2} + 2\mu_\epsilon(\epsilon^{(i)} - \epsilon^*)}{2\sigma_\epsilon^2}}$, accept the proposed value and set $\theta^{(i+1)} = \theta^*$, else set $\theta^{(i+1)} = \theta^{(i)}$.

3. If $i < N$, then go to step 2.

S6 Text. Model selection.

We would like to assess whether we can concatenate the Besançon data e.g. before and after the renovation of the ICUs in one large data set to increase the power of our method. The idea is to compare the DICs for two different scenarios:

- Consider only one model including one parameter set $\theta = \{\alpha, \beta, f, \phi\}$ where α is the endogenous, β the cross-transmission parameter, f the importation rate and ϕ the test sensitvity. The analysis is then performed on all the data X of the two ICUs and the two time periods (before and after renovation). The DIC is then given by

$$\mathrm{DIC}_{X,\theta} = \overline{D_X(\theta)} + \frac{1}{2}\mathrm{Var}(D_X(\theta)).$$

- Consider a model including a parameter set consisting of separate parameters for each time period:

  - $\theta_1 = \{\alpha_1, \beta_1, f_1, \phi_1\}$

  - $\theta_2 = \{\alpha_2, \beta_2, f_2, \phi_2\}$

The parameter set of the model is then:

$$\theta = \theta_1 \cup \theta_2 = \{\alpha_1, \beta_1, f_1, \phi_1, \alpha_2, \beta_2, f_2, \phi_2\}.$$

The parameters in θ_1 are updated for the data set before renovation whereas the parameters in θ_2 are updated for the data set after renovation. The deviance for this model is determined by

$$
\begin{aligned}
D_X(\theta) &= -2\log\pi(X \mid \theta) \\
&= -2\log\pi(\{X_1, X_2\} \mid \theta_1 \cup \theta_2) \\
&= -2\log\left[\pi(X_1 \mid \theta_1 \cup \theta_2) \cdot \pi(X_2 \mid \theta_1 \cup \theta_2)\right] \\
&= -2\log\left[\pi(X_1 \mid \theta_1) \cdot \pi(X_2 \mid \theta_2)\right] \\
&= -2\left[\log\pi(X_1 \mid \theta_1) + \log\pi(X_2 \mid \theta_2)\right] \\
&= D_{X_1}(\theta_1) + D_{X_2}(\theta_2),
\end{aligned}
$$

where X_1 is the data set for the time period before and X_2 for the time period after renovation. Thus, the DIC for the model including separate parameters for each time period can be calculated as

$$\text{DIC}_{X,\theta} = \text{DIC}_{X_1,\theta_1} + \text{DIC}_{X_2,\theta_2}.$$

S7 Text. Model assessment.

To assess our method's ability to detect large discrepancies between the data and model assumptions, we simulated a data set with a substantial contribution of background, cross-transmission and environmental contamination after discharge (27%, 24% and 49%). We analyzed the data set with our MCMC procedure including only background transmission as a transmission route. Thus, the model in the MCMC process assumed a constant force of infection. In Fig S18, we can see that the expected coverage probabilities are not met. Hence, it can be asserted that there is a large discrepancy between the model and the data. The data can be found on `https://github.com/tm-pham/transmissionPA`.

S8 Text. Prior distributions.

Sensitivity analyses were performed using different priors. We performed the main analyses of the Besançon (as presented in the *Results* section) with uninformative exponential priors and small intitial values for the standard deviation of the proposal distribution. Further analyses were performed using a weakly informative exponential prior Exp(1.0) and a uniform prior U(0, 2) for the decay rate μ. The histograms and traceplots using Exp(1.0) show that these results are not different from results using an uninformative exponential prior. However, for the uniform prior $U(0, 2)$, the MCMC chain shows signs of non-convergence. The values for μ have a strong tendency towards the upper boundary and a strong correlation with ϵ (see S17). This behaviour was also observed when the full model was applied to simulated data sets with no environmental contamination after discharge (see S9) and can be explained as follows: A scenario with no environmental contamination after discharge is indistinguishable from a scenario with environmental contamination but very short bacterial persistence in the environment (i.e. high values of μ). In such a case, several combinations of ϵ and μ and β reflect the same situation. In particular, any high value of μ may reflect the absence of environmental contamination. The results of our sensitivity analyses confirm that for the two data sets of the Besançon hospital, environmental contamination after discharge is only of minor influence. Further elaborations on the influence of different values of the decay rate μ can be found in S9.

S9 Text. Simulation studies.

Several data sets were simulated to test the algorithm's ability to identify the correct relative contribution. We performed two types of simulations:

1. The duration of persistence of bacterial load is varied.

2. The relative contribution of envionmental contamination after discharge is varied.

For the first simulation study, we fixed values for α, β and ν and varied values for μ. The parameter ϵ is set to $0.6 \cdot \mu$. Three main scenarios regarding the duration of persistence of bacterial load in the environment are analyzed:

1. Long: $\mu \in \{1/7, 1/14\}$

2. Medium: $\mu \in \{1, 1/3\}$

3. Short: $\mu \in \{2, 5\}$

For each of the three scenarios the importation rate is varied within $\{0.01, 0.05, 0.1\}$. We observed that chain convergence cannot be attained in a reasonable amount of time using uninformative exponential priors. For medium duration of bacterial persistence convergence could be achieved using either a weakly informative prior $\text{Exp}(1.0)$ and a uniform prior $U(0, 2)$. The reasons for non-convergence for long or short bacterial persistence in the environment as well as the justification of the weakly informative and uniform prior are based on the same reasoning.

If bacterial persistence is set to be long (longer than the average length of stay of patients), then environmental contamination after discharge stays approximately at one level and the resulting probability of colonization due to this route is approximately constant. Thus, the induced fluctuations in the prevalence can be hardly distinguished from fluctuations due to background transmission. On the other hand, a short duration of bacterial persistence (much shorter than the average length of stay) leads to difficulties in distinguishing the resulting model from one with a higher contribution of cross-transmission and smaller contribution of environmental contamination. Hence, based on the fluctuations of the prevalence, only a medium length of bacterial persistence is meaningful and the restriction of the parameter space or the use of more informative priors is justified. For the second simulation, we varied the relative contribution of environmental contamination after discharge. In particular, when environmental contamination after discharge

was not present in the simulated data, the results resembled our analysis of the Besançon data shown in the *Results* section.

The histograms and plots corresponding to the described simulation studies can be found on `https://github.com/tm-pham/transmissionPA`. Further information on our simulation studies may be requested from the first author.

S10 Text. Secondary analyses.

In addition to the analyses presented in *Results*, six further analyses were performed. For each ICU, the time periods before and after renovation were combined. Finally, all available data was concatenated into one big data set and analyzed at once. The results of these analyses using the submodel as well as the full model are presented in S1 - S2 Tables. The posterior estimates of the model parameters and the corresponding relative contributions are similar to the ones presented in the *Results* section.

S11 Text. Impact of prior colonized bed occupants.

In a first step, we performed Fisher's exact test to study the association between colonization status of current and prior bed occupants for the data sets of the University Hospital of Besançon. The 2×2 tables can be found in S3 and S4.

Interestingly, there is a significant association for ICU B (OR: 1.54, 95% cI: (1.24, 1.91), *p*-value: 0.0001) but none for ICU A (OR: 0.95, 95% cI: (0.7, 1.27), *p*-value: 0.7707). However, such a simple test may be confounded. Consecutive colonized patients may be a result due to cross-transmission rather than an increased risk of prior bed occupants. In order to disentangle the effects of cross-transmission and prior bed occupants, we analyzed the data sets of the University Hospital of Besançon using the following simple model: Each bed occupant i faces a force of infection λ_i depending on the colonization status of the prior bed occupant:

$$\lambda_i(t) = \alpha + \beta \frac{I(t)}{N(t)} + p \cdot \mathbb{K}(c_i^{\text{prior}}) \tag{3.13}$$

where c_i^{prior} is the colonization status of the prior bed occupant and $\mathbb{K}(c_i^{\text{prior}}) = 1$ if $c_i^{\text{prior}} = 1$ and 0 otherwise. It represents the increased risk that patients experience when occupying a room/bed of a prior colonized bed occupant. The parameter estimates and the relative contributions can be found in Table S5.

The results show that the influence of prior bed occupants is only limited ($< 6\%$) for both ICUs of the University Hospital of Besançon. Simulation studies confirm that a significant impact of this route would be detected by this model. The code of the MCMC procedure for this analysis, the histogram and traceplots for the simulations studies and the data sets of Besançon can be found on `https://github.com/tm-pham/transmissionPA`.

S12 Text. COMBACTE-MAGNET membership list.

Please find below the list of COMBACTE-MAGNET consortium partners:

Albert Vermaas, Alex Waehry, Angela Supplitt, Anne Adams, Antonio Portolés Pérez, Aurore Drecq, Brian Allen, Christophe Misse, Cornelia Mockwitz, Fanny Senez, Felicity Jane Gabbay, Freek De Jong, Gabriella Monaco, Gill Wells, Heather Rogers, Henrik Landstrom, Hermann Hayn, Holger Schmoll, Jaime Caro Aguirre, Jantine Spithoven, Josep M. Campistol, Juergen Dreyer, Karine Clement, Karl-Heinz Müller, Lauren Fleming, Lea Pais, Lynsey Keig, Malcolm Skingle, Maria Carol Sanjurjo, Marion Do Maria, Markus Jäger, Michael Browne, Nicola Williams, Pascal Savary, Patricia Gizecki, Renaud Mazy, Sarah Everett-Cox, Sofia Karakostas, Tommaso Rupolo, Ursula Theuretzbacher, Virginia Nieto Guerrero, Wilfried Reincke, Alain Verschoren, Alexander Affeldt, Alfredo García Díaz, Andreas Rothfuss, Carlo Giaquinto, Christine Clerici, Daniel Wyler, Denis Hochstrasser, Dieter Kaufmann, Dirkjan Masman, Frank Miedema, Helen Steel, Holger Zimmermann, Jaap Verweij, Jan-Olof Jacke, Jean-François Lefebre, John Graham, José Francisco Soto Bonel, Jose Manuel Aranda Lara, Josep M. Campistol Plana, Laurence Lomme, Marcel Levi, María Dolores Acón, Markus Müller, Maya Saïd, Nicola Sartor, Noureddine Farah, Pastora Martinez Samper, Pierre-François Leyvraz, Renaud Mazy, Ron Scott, Yves Geysels, Andreas Kümin, Anthony Latte, Clemens Lässing, Elena Ferragut Roig, Eleonora Zuolo, Esther Bettiol, Eva Lindgren, Eveline Bielser, Gülseren Yalvac, Jenny Lawson, José Ángel Freire Astray, Jose Soto Bonel, Julia Lloyd-Parks, Jürgen Dreyer, Malgorzata Kielbasa, Marco Perdon, Markus Zeitlinger, Michaela Schuhmacher, Michiel Gerlagh, Olivier Brun, Pam Neagle, Patricia Schott, Rebecca Smith, Sally Miles, Sophie Monteau, Susanna Montalto, Thierry Borloz, Wouter Roobol, Xavier Fretille, Abdel Oualim, Alasdair Macgowan, Andreas Voss, Andrew Lovering, Anne Witschi, Antoni Torres, Antonio Oliver, Bruno Francois, Craig Maclean, Cuong Vuong, David Evans, Evelina Tacconelli, Hasan Jafri, Ingrid Klingmann, Jan Beyersmann, Jean Chastre, Jean-François Timsit, Jesús Rodriguez Baño, Johan Mouton, Kim Gilchrist, Leonard Leibovici, Leonhard Held, Marc Bonten, Martin Wolkewitz, Mervyn Singer, Miguel Sanchez, Mike Sharland, Miquel Pujol Rojo, Philippe Eggimann, Philippe Montravers, Pierre-François Laterre, Richard Bax, Richard Fitzgerald, Stephan Harbarth, Surbhi Malhotra-Kumar, Tom van der Poll, William Hope.

S1 Table. Summary statistics of the marginal posterior distributions for parameters of the submodel based on the analysis of the Besançon data.

Parameter	Symbol	Median (95% credibility interval)*		
		ICU A	ICU B	ICUs combined
Background coefficient	α	0.011(0.008, 0.015)	0.008(0.006, 0.011)	0.009(0.007, 0.011)
Cross-transmission coefficient	β	0.03 (0.012, 0.046)	0.038(0.006, 0.011)	0.034(0.024, 0.044)
Sensitivity	$\phi\,(\%)$	50.7 (48.0, 53.2)	60.5 (58.8, 62.1)	57.5 (56.2, 58.9)
Importation probability	$f\,(\%)$	5.5 (4.5, 6.5)	7.6 (6.7, 8.6)	6.5 (5.8, 7.2)
Fraction colonized	$p_{\mathrm{col}}\,(\%)$	22.2 (21.2, 23.1)	23.2 (22.7, 23.7)	22.4 (21.9, 22.8)
Contributions				
Background contribution	$R_{\mathrm{backgr}}\,(\%)$	65.1 (46.4, 84.6)	51.1 (35.5, 66.9)	57.6 (46.0, 68.9)
Cross-transmission contribution	$R_{\mathrm{crossT}}\,(\%)$	34.9 (15.4, 53.6)	48.9 (33.1, 64.5)	42.4 (31.1, 54.0)

*Highest posterior density interval

S2 Table. Summary statistics of the marginal posterior distributions for parameters of the full model based on the analysis the Besançon data.

Parameter	Symbol	Median (95% credibility interval)*		
		ICU A	ICU B	ICUs combined
Background coefficient	α	0.011 (0.008, 0.014)	0.008 (0.006, 0.011)	0.009 (0.007, 0.011)
Cross-transmission coefficient	β	0.013 (0, 0.033)	0.022 (0, 0.011)	0.018 (0, 0.034)
Environmental coefficient	ϵ	201.1 (0.018, 832.829)	176.7 (0.009, 811.6)	209.0 (0.02, 784.5)
Decay rate	μ	1415.8 (65.6, 4273.8)	1396.9 (26.4, 3992.2)	1419.2 (43.7, 4524.0)
Sensitivity	ϕ (%)	50.6 (48.0, 53.1)	60.5 (58.9, 62.1)	57.6 (56.2, 58.9)
Importation probability	f (%)	5.5 (4.5, 6.6)	7.5 (6.6, 8.4)	6.4 (5.7, 7.2)
Fraction colonized	p_{col} (%)	22.2 (21.3, 23.1)	23.2 (22.7, 23.6)	22.4 (21.9, 22.8)
Contributions				
Background	R_{backgr} (%)	62.2 (44.8, 79.7)	50.4 (35.6, 65.2)	57.6 (45.5, 67.9)
Cross-transmission	R_{crossT} (%)	37.2 (20.4, 54.9)	49.0 (34.0, 63.1)	42.7 (31.9, 54.0)
Env. cont. after discharge	R_{env} (%)	0.006 (0, 0.013)	0.005 (0, 0.013)	0.006 (0, 0.012)

*Highest posterior density interval
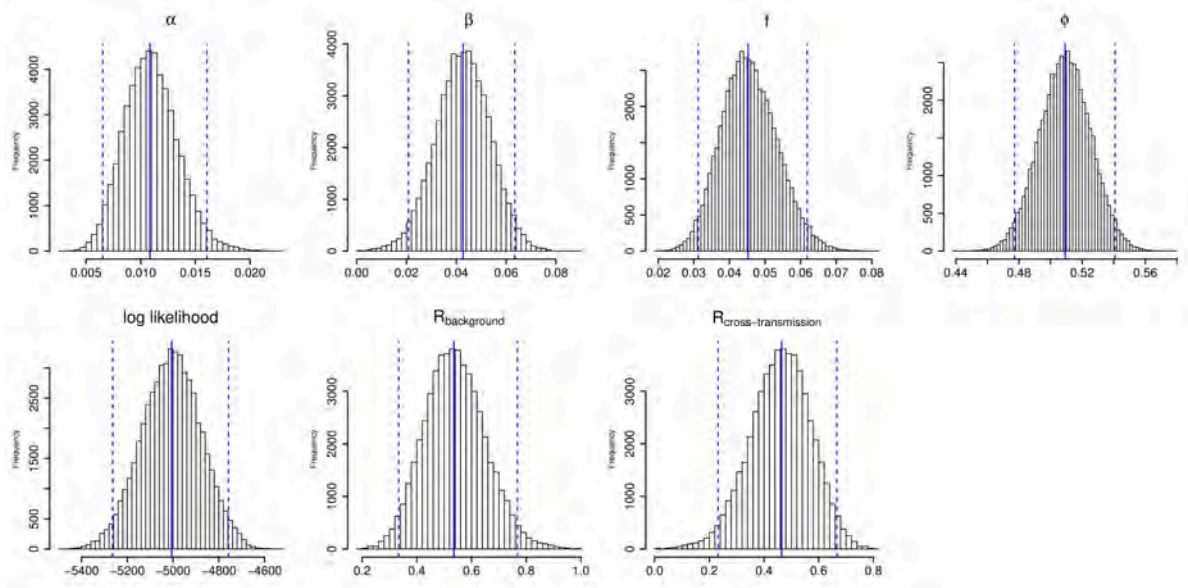
S3 Table. Association of colonization statuses of consecutive bed occupants in ICU A of the University Hospital of Besançon.

		Current bed occupant	
		Colonized	Non-colonized
Prior bed occupant	Colonized	57	478
	Non-colonized	481	3826

S4 Table. Association of colonization statuses of consecutive bed occupants in ICU B of the University Hospital of Besançon.

| | | Current bed occupant | |
		Colonized	Non-colonized
Prior bed occupant	Colonized	123	586
	Non-colonized	610	4479

S5 Table. Summary statistics of the marginal posterior distributions for parameters of model (3.13) **based on the analysis of the Besançon data.**

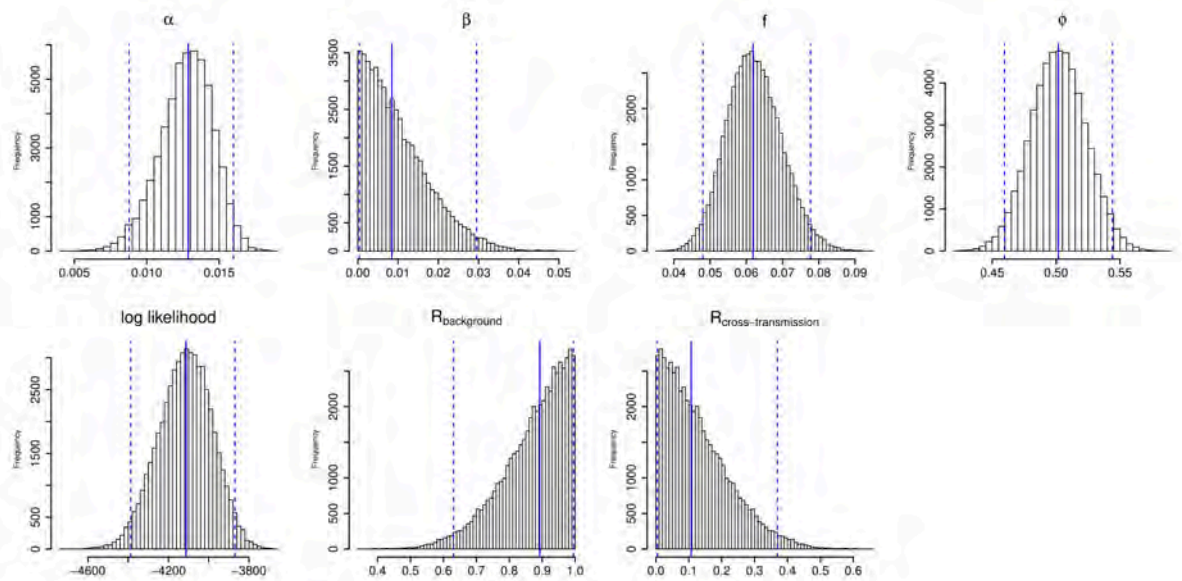| Parameter	Symbol	Median (95% credibility interval)*	
		ICU A	ICU B
Background coefficient	α	0.011 (0.008, 0.014)	0.008 (0.006, 0.011)
Cross-transmission coefficient	β	0.027 (0.013, 0.041)	0.036 (0.023, 0.048)
Prior bed occupant coefficient	p	0.004 (0, 0.007)	0.003 (0, 0.006)
Sensitivity	$\phi\,(\%)$	50.6 (48.0, 53.1)	60.5 (58.9, 62.1)
Importation probability	$f\,(\%)$	5.5 (4.5, 6.6)	7.5 (6.6, 8.4)
Fraction colonized	$p_{col}\,(\%)$	22.2 (21.3, 23.1)	23.2 (22.7, 23.6)
Contributions			
Background	$R_{backgr}\,(\%)$	65.1 (48.8, 81.7)	50.4 (35.7, 66.6)
Cross-transmission	$R_{crossT}\,(\%)$	31.8 (14.6, 47.2)	46.8 (31.3, 62.1)
Prior bed occupants discharge	$R_{prior}\,(\%)$	3.1 (0, 6)	2.6 (0, 5.6)

*Highest posterior density interval
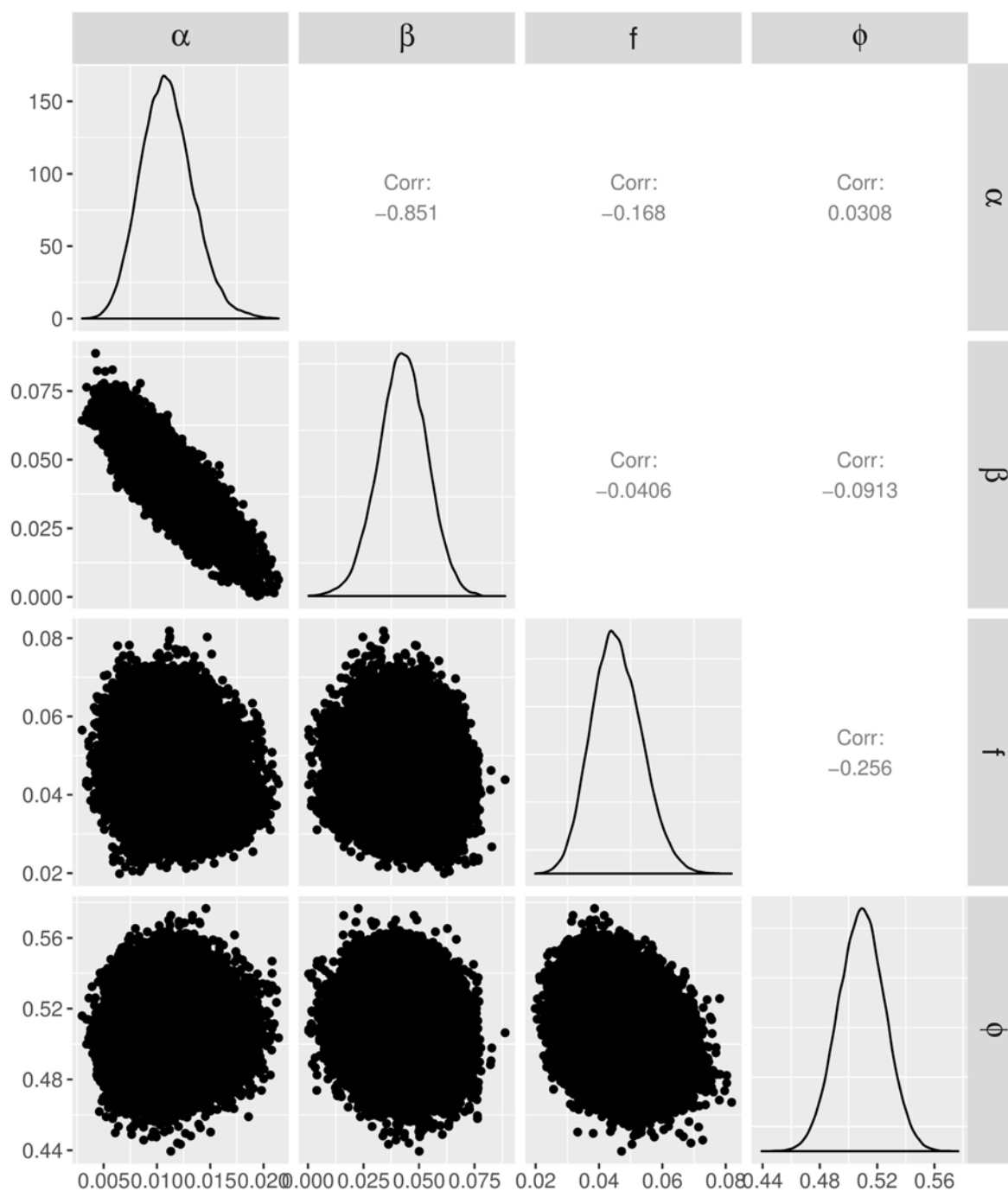
S1 Fig. Pairwise plots of samples from the posterior distribution for the transmission parameters of the submodel. The plots were generated from the data of ICU A before renovation using the submodel with background and cross-transmission.
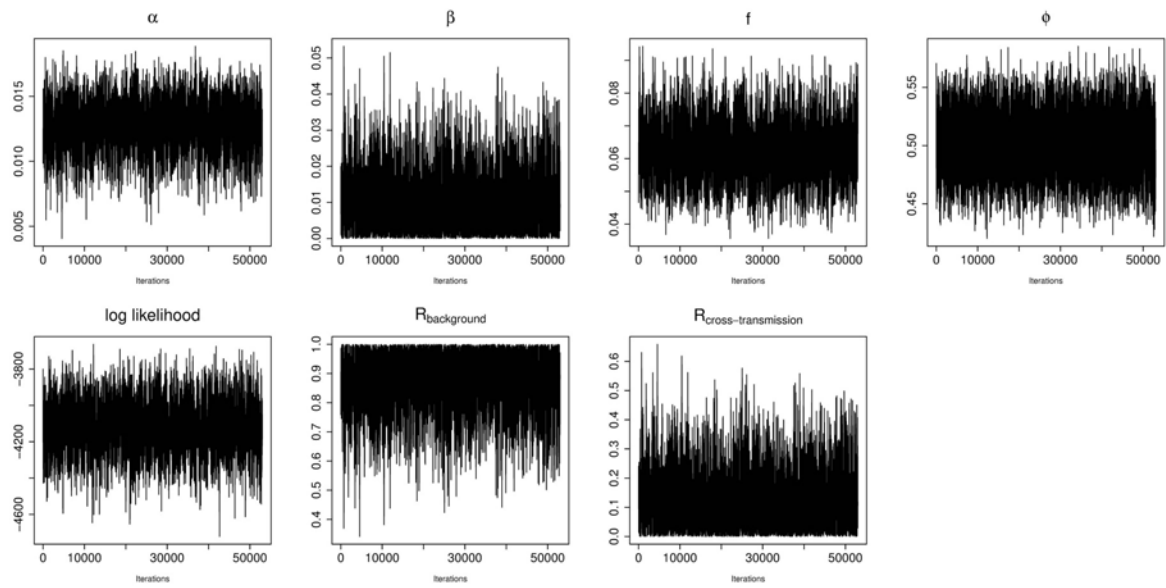
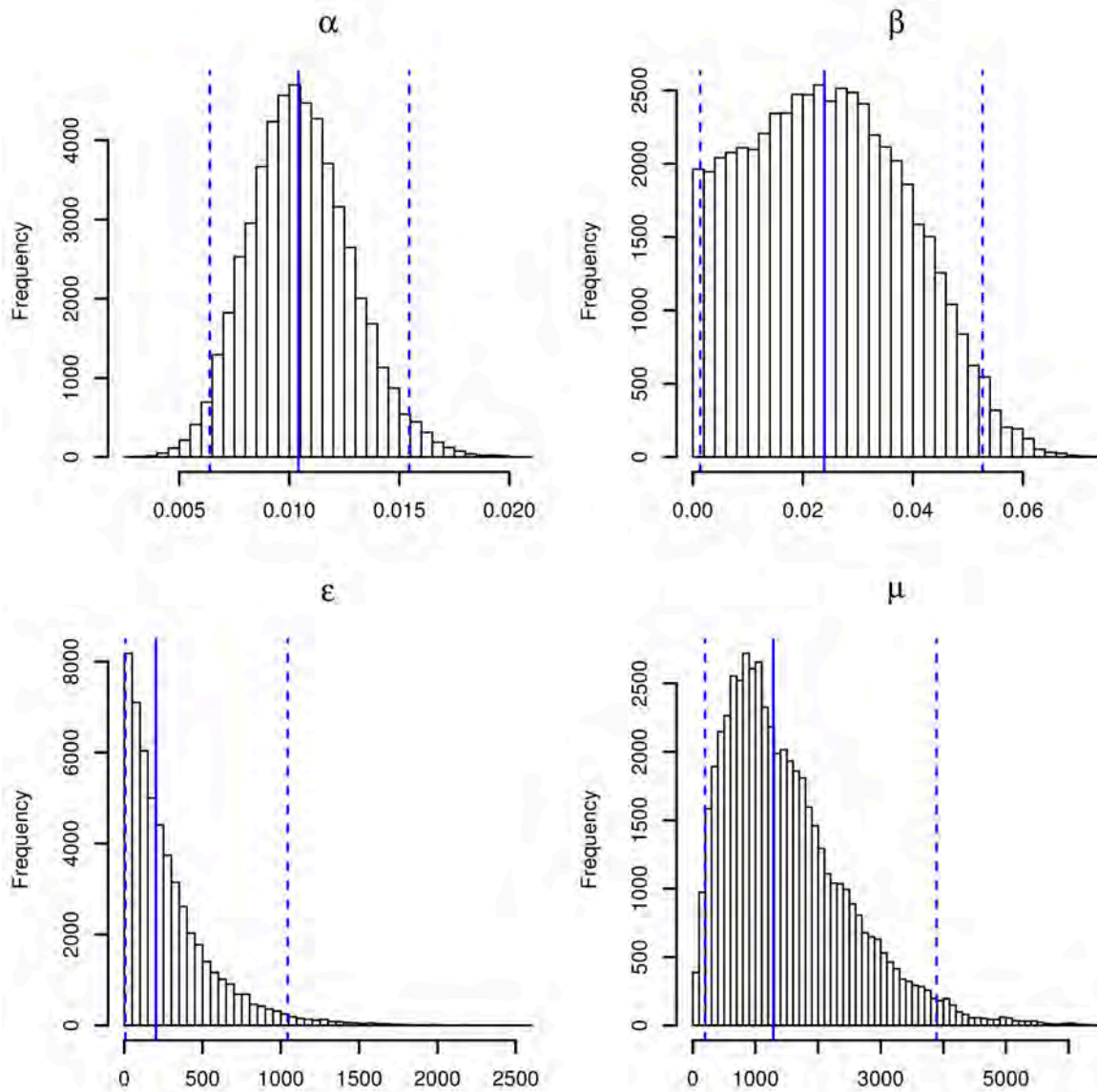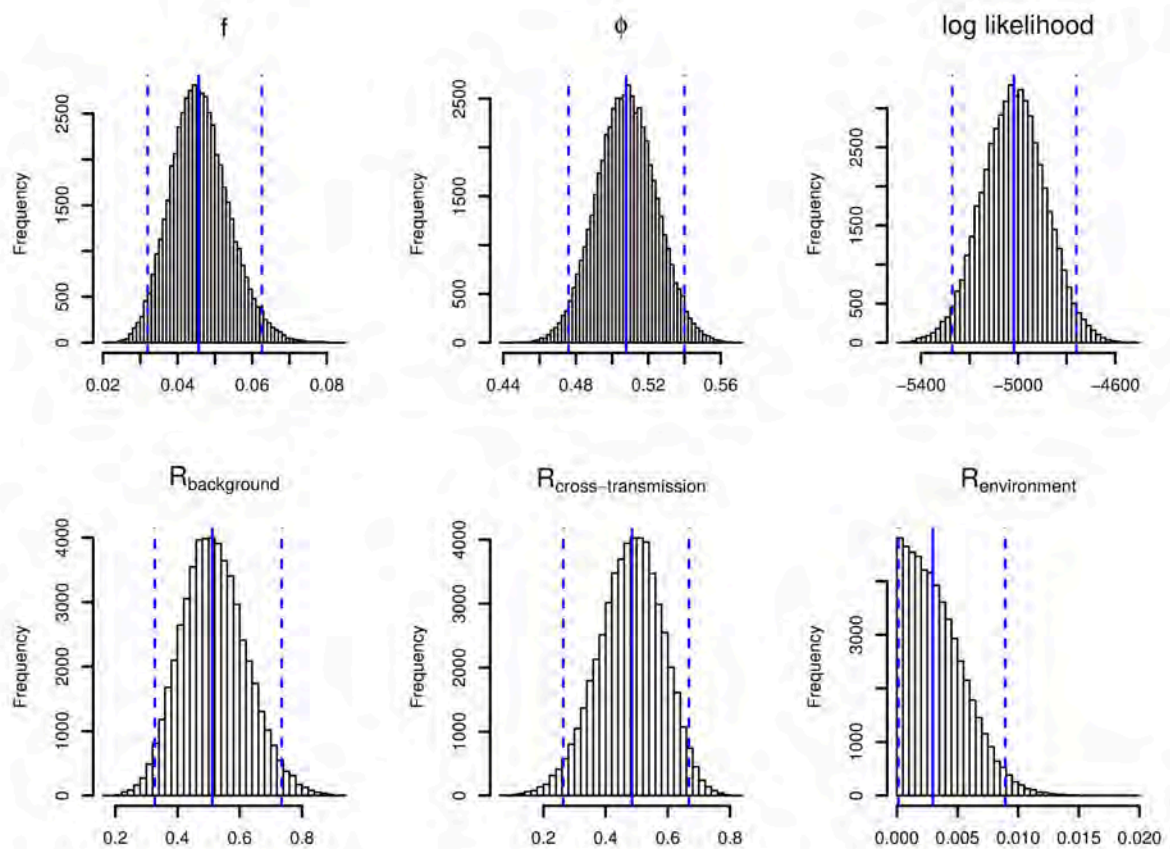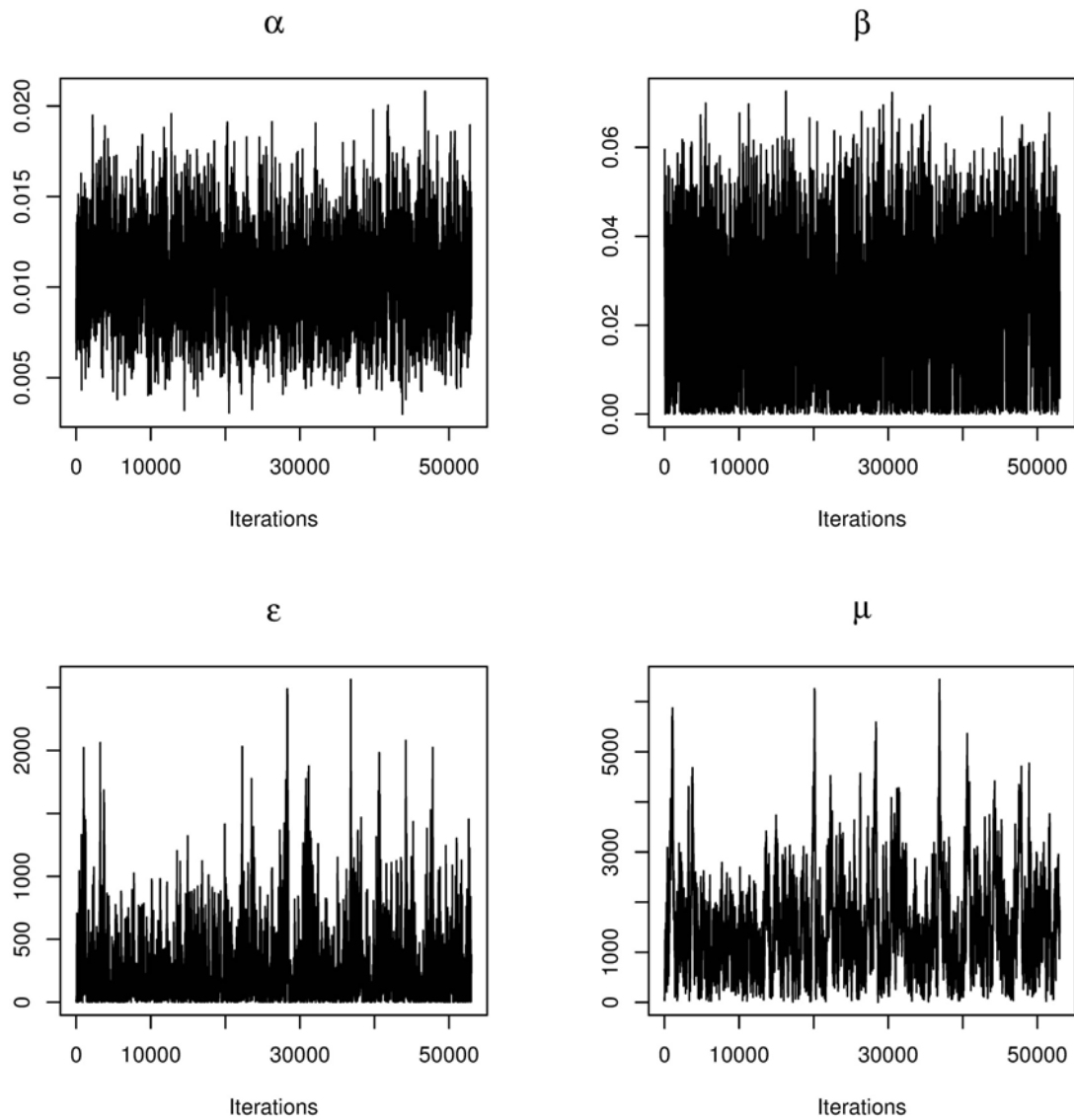**S2 Fig. Histograms for ICU A after renovation using the submodel with background and cross-transmission.**

S3 Fig. Traceplots for ICU A before renovation using the submodel with background and cross-transmission.

S4 Fig. Histograms for ICU A after renovation using the submodel with background and cross-transmission.

S5 Fig. Traceplots for ICU A after renovation using the submodel with background and cross-transmission.
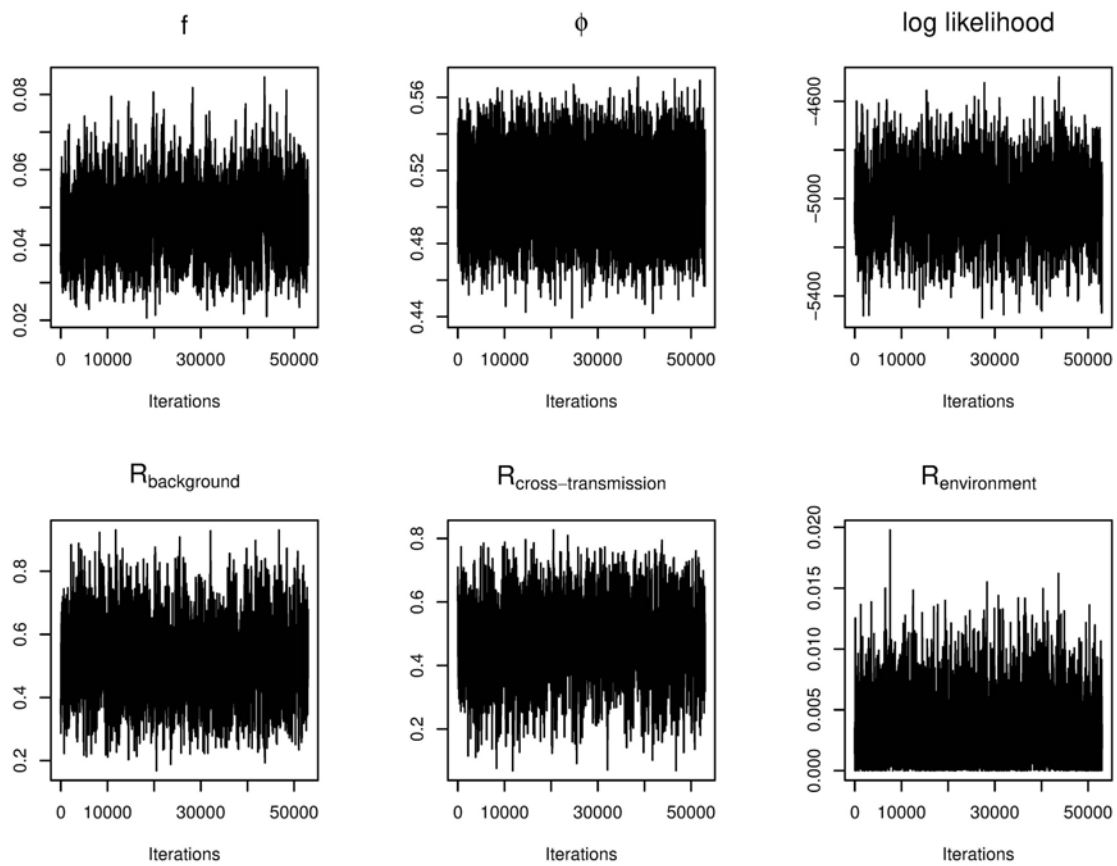
S6 Fig. Histograms for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.
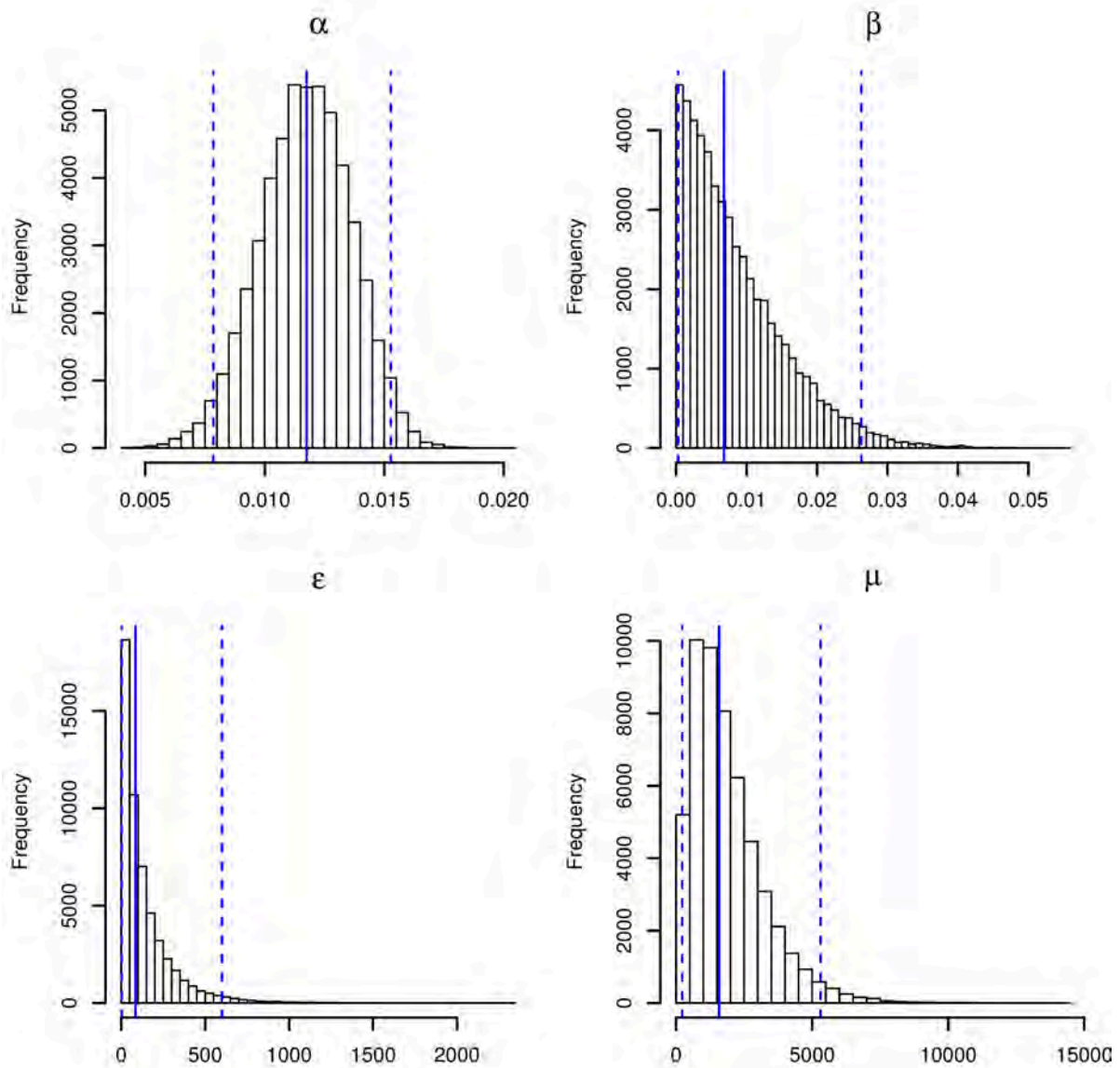
S7 Fig. Histograms for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in \{$background, cross-transmission, environment$\}$.

S8 Fig. Traceplots for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.
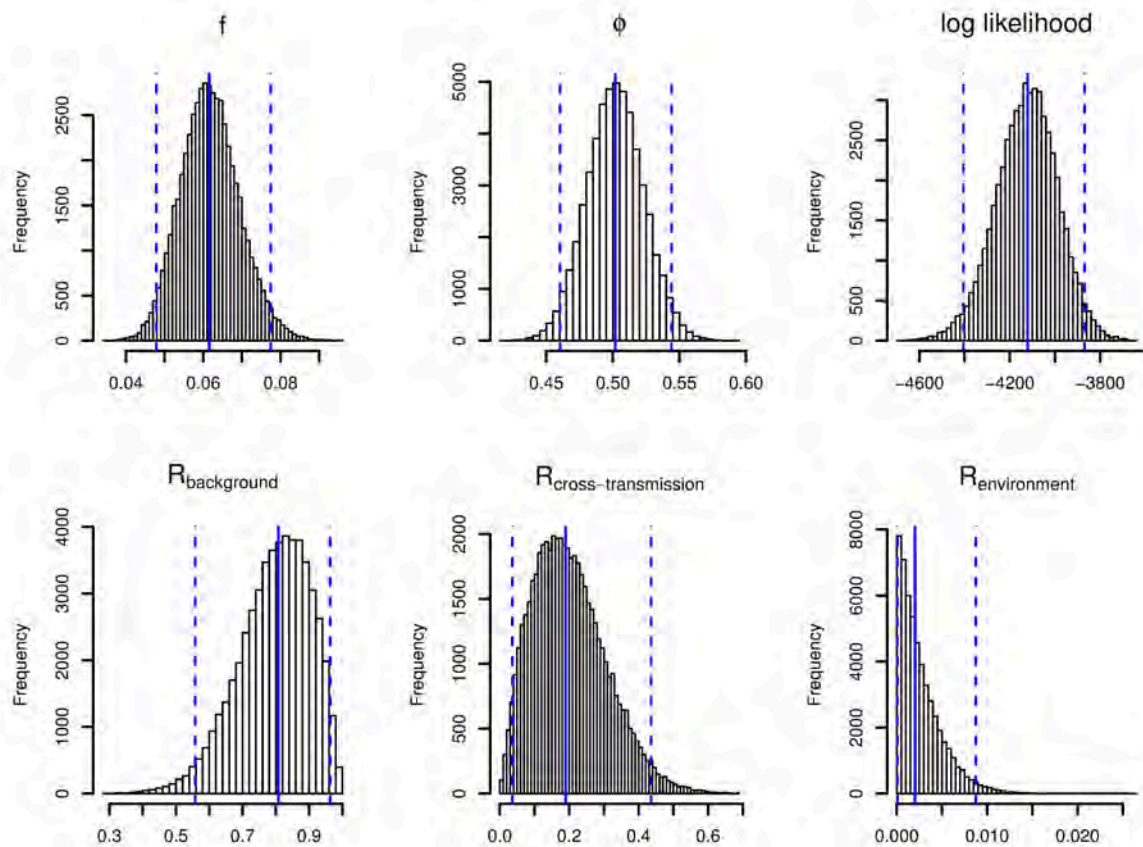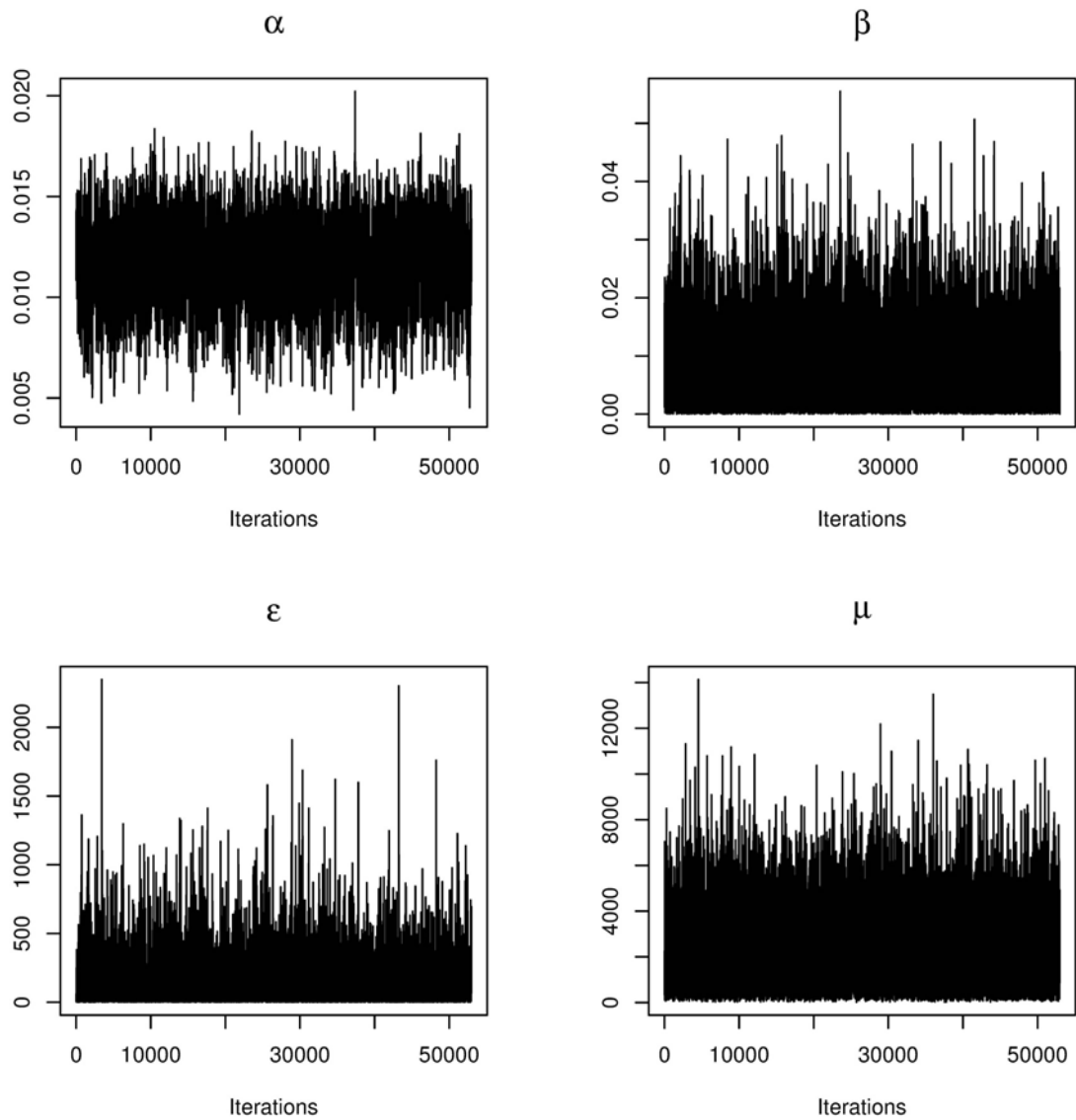
S9 Fig. Traceplots for ICU A before renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in \{\text{background, cross-transmission, environment}\}$.

S10 Fig. Histograms for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.
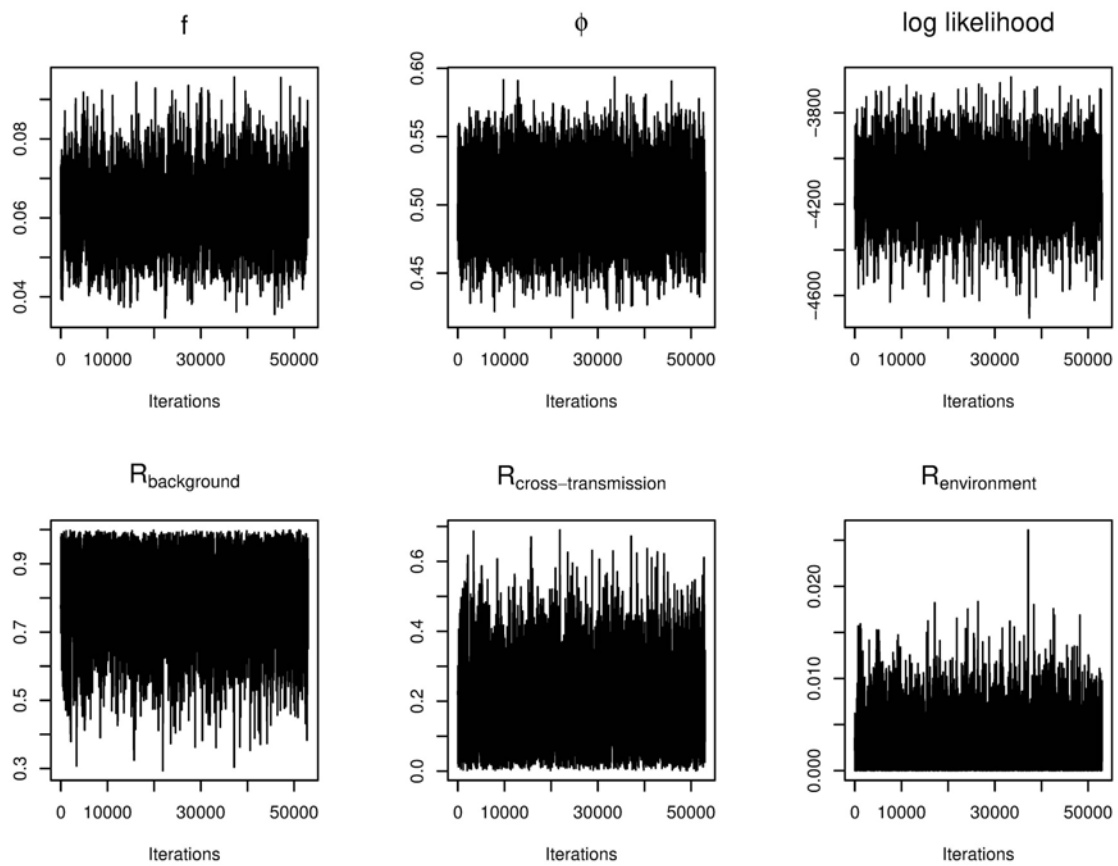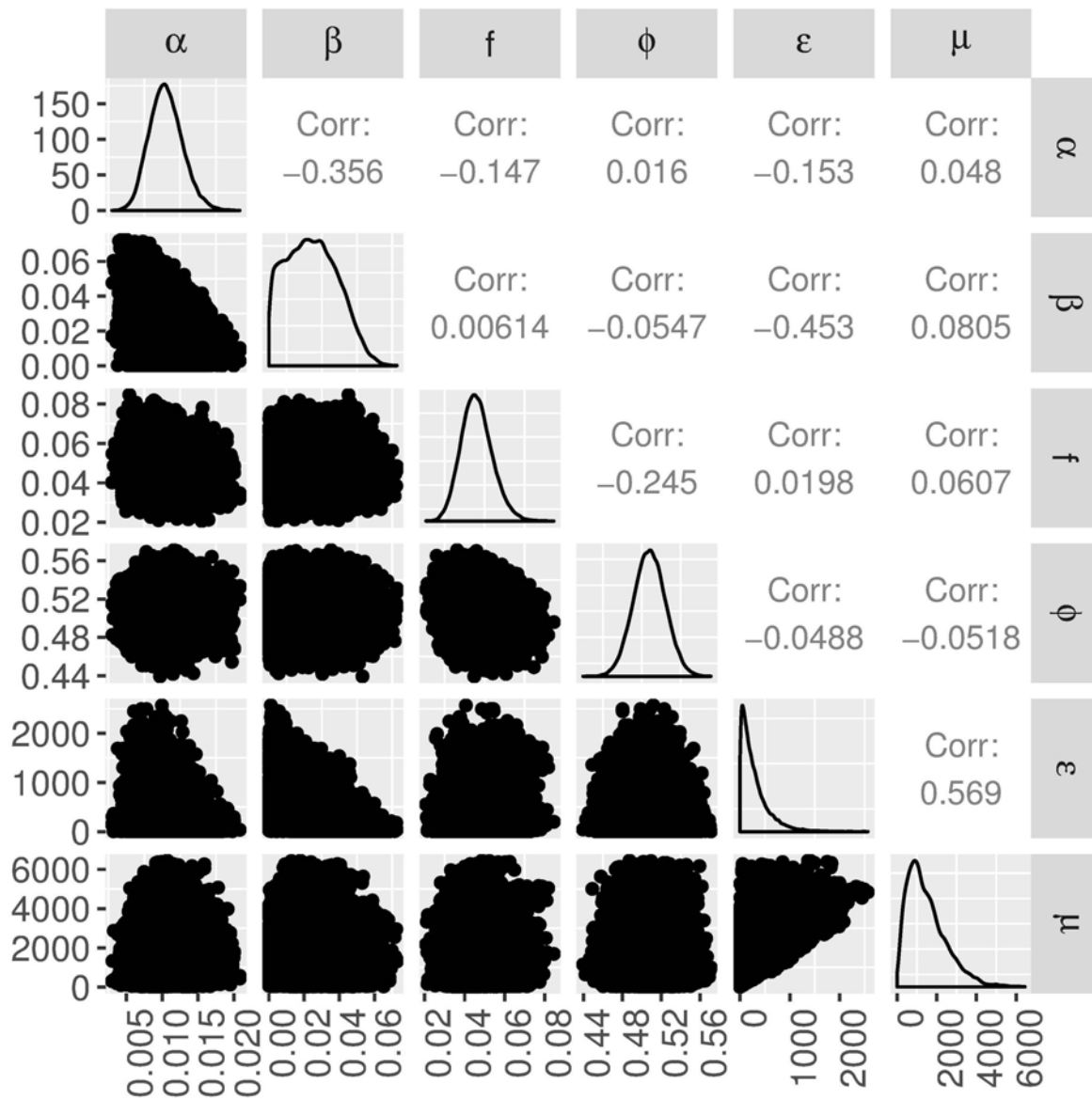
S11 Fig. Histograms for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in \{\text{background, cross-transmission, environment}\}$.

S12 Fig. Traceplots for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.
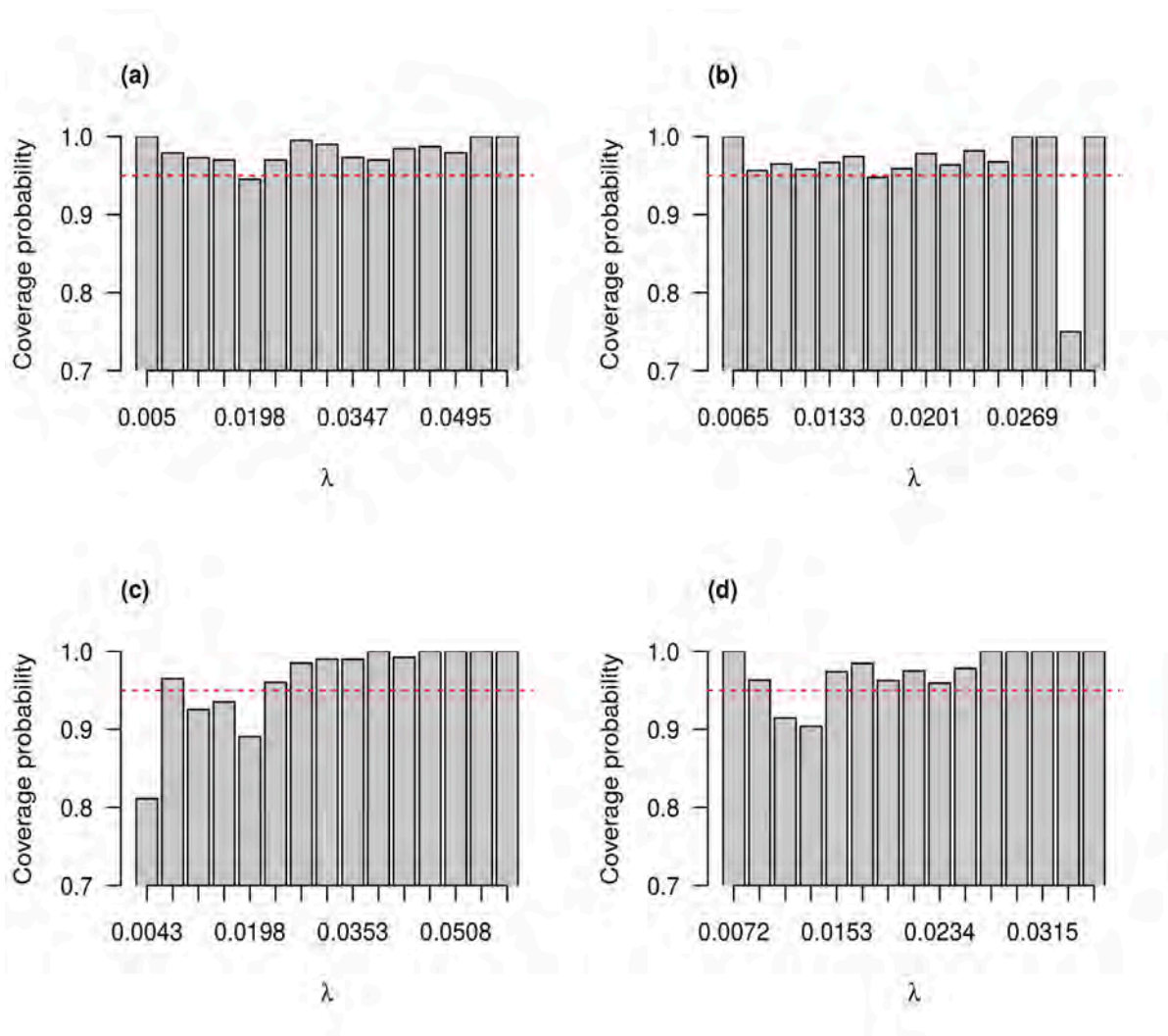
S13 Fig. Traceplots for ICU A after renovation using the full model with background, cross-transmission and environmental contamination. The results are displayed for the importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i, where $i \in \{\text{background, cross-transmission, environment}\}$.

S14 Fig. Pairwise plots of samples from the posterior distribution for the transmission parameters of the full model. The plots were generated from the data of ICU A before renovation using Exp(0.001) prior and the full model with background, cross-transmission and environmental contamination.

S15 Fig. Coverage probabilities for the submodel using Jeffreys prior. (a) - (b) ICU A before and after renovation, respectively. (c) - (d) ICU B before and after renovation, respectively.

S16 Fig. Coverage probabilities for the full model using Jeffreys prior. (a) - (b) ICU A before and after renovation, respectively. (c) - (d) ICU B before and after renovation, respectively.
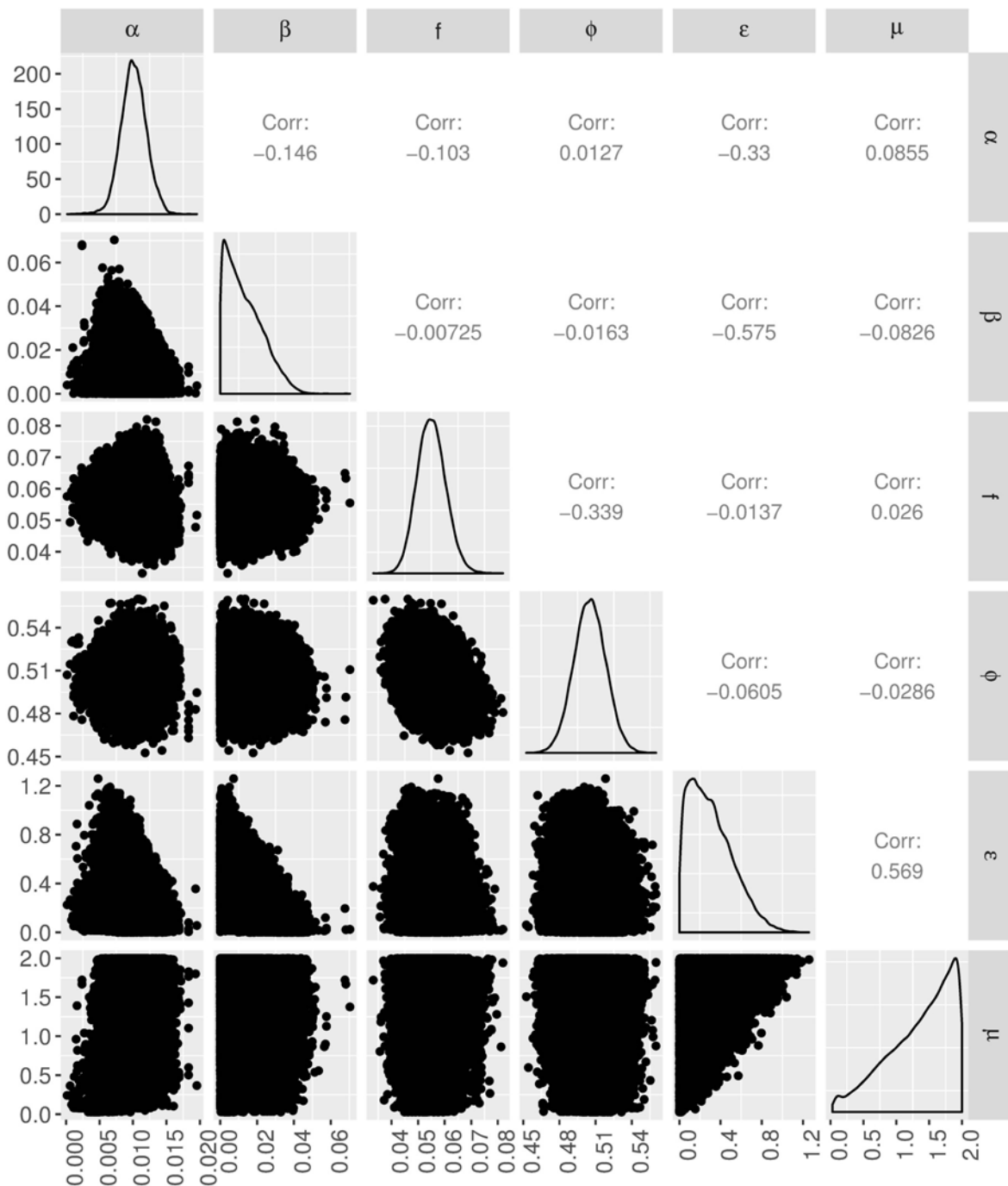
S17 Fig. Pairwise plots of samples from the posterior distribution for the transmission parameters of the full model. The plots were generated from the data of ICU A using U(0,2) prior and the full model with background, cross-transmission and environmental contamination after discharge.

S18 Fig. Coverage probabilities for simulated data set using Jeffreys prior. The data was simulated using $\alpha = 0.015$, $\beta = 0.055$, $\mu = 1/7$, $\epsilon = 0.15$, $f = 0.05$, $\phi = 1$. The analysis assumed only one route, i.e. background transmission. The plot shows large discrepancies between the expected and the computed coverage probabilities, pointing to a misspecified model.

Chapter 4

Modes of transmission of VIM-positive Pseudomonas aeruginosa in adult intensive care units - analysis of 9 years of surveillance at a university hospital using a mathematical model

Thi Mui Pham[*], Andrea C. Büchler[*], Anne F. Voor in 't holt, Juliëtte A. Severin, Martin C. J. Bootsma, Mirjam E. Kretzschmar[†], Margreet C. Vos[†]

[*] These authors contributed equally to this work.
[†] These authors contributed equally to this work and are cosenior authors.

Abstract

Background: Hospital outbreaks of multidrug resistant *Pseudomonas aeruginosa* (*P. aeruginosa*) are often caused by *Pseudomonas aeruginosa* (*P. aeruginosa*) clones which produce metallo-β-lactamases, such as Verona Integron-encoded Metallo-β-lactamase (VIM). Although different sources have been identified, the exact transmission routes often remain unknown. However, quantifying the role of different transmission routes of VIM-PA is important for tailoring infection prevention and control measures. The aim of this study is to quantify the relative importance of different transmission routes by applying a mathematical transmission model using admission and discharge dates as well as surveillance culture data of patients.

Methods: We analyzed VIM-PA surveillance data collected between 2010 and 2018 of two intensive-care unit (ICU) wards for adult patients of the Erasmus University Medical Center Rotterdam using a mathematical transmission model. We distinguished two transmission routes: Direct cross-transmission and a persistent environmental route. Based on admission, discharge dates, and surveillance cultures, we estimated the proportion of transmissions assigned to each of the routes.

Results: Our study shows that only 13.7% (95% credibility interval: 1.4%, 29%) of the transmissions that occurred in these two ICU wards were likely caused by cross-transmission, leaving the vast majority of transmissions (86.3%, 95% credibility interval: 71%, 98.6%) due to persistent environmental contamination.

Conclusions: Our results emphasize that persistent contamination of the environment may be an important driver of nosocomial transmissions of VIM-PA in ICUs. To minimize the transmission risk from the environment, potential reservoirs should be regularly and thoroughly cleaned and disinfected, or redesigned.

Keywords: Drug Resistance, Multiple; *Pseudomonas aeruginosa* (*P. aeruginosa*); Critical Care; Epidemiological monitoring; Models, Statistical.

Introduction

Multidrug resistant (MDR) microorganisms are an emerging problem worldwide. The most emerging threat is the spread of carbapenem-resistant Enterobacterales and carbapenem-resistant non-fermenting microorganisms, such as *Acinetobacter baumannii* (*A. baumannii*) and *Pseudomonas aeruginosa* (*P. aeruginosa*) [1]. *P. aeruginosa* is one of the most common nosocomial pathogens [2, 3]. It can cause serious infections in patients with underlying conditions, such as immunosuppression, cystic fibrosis, and patients admitted to the intensive care unit (ICU). The morbidity and mortality of *P. aeruginosa* bloodstream infections is high, especially in immunocompromised patients [4–6]. Due to its intrinsic and acquired resistance to multiple antibiotics, *P. aeruginosa* is not only a common cause of nosocomial infections but also diffcult to treat. Multidrug resistance mechanisms in *P. aeruginosa* are loss or alteration of outer membrane porins, increased efflux pump activity and carbapenemase production [2] with the latter being the most common underlying mechanism of MDR *P. aeruginosa* involved in in-hospital outbreaks [7]. Among the carbapenemases, the Verona Integron-encoded Metallo-beta-lactamase (VIM) is most dominant, and most widely disseminated [8].

Identifying the pathways of transmission of *P. aeruginosa* in hospital outbreaks is key for targeted and timely infection prevention and control (IPC) measures. Although the exact transmission route often remains unknown, different modes of transmission are described in the literature. For *P. aeruginosa*, water-related devices such as sinks are the most common environmental source [9, 10]. Quantifying the relative importance of transmission routes may serve as an essential tool in outbreak investigation as well as in designing effective and tailored IPC strategies.

Models for inference of transmission parameters for different transmission routes have been developed for various MDR bacteria [11–14]. Pham et al [14] developed a mathematical transmission model including three different routes of transmission for *P. aeruginosa* using ICU data from two ICUs of a French hospital in Besançon. The authors estimated the relative contribution of background transmission, cross-transmission and environmental contamination after discharge using an extensive surveillance data set. It

was shown that environmental contamination due to colonized patients that persisted after their discharge likely had a small contribution ($< 1\%$) to the overall number of transmissions. Persistent environmental contamination was included in "background transmission" for which the relative contribution was significantly higher. While this route could have played an important role in the transmission process, it could not be distinguished from other routes that could have caused a similar constant risk of colonization.

In this paper, we present the application of a similar mathematical transmission model to surveillance data of VIM-producing *P. aeruginosa* (VIM-PA) at the Erasmus University Medical Center Rotterdam (Erasmus MC), the Netherlands. In this hospital, since 2003, VIM-PA colonized and infected over 150 patients, with most patients being identified at the ICU [15]. Multiple sources and transmission routes have been identified since; with sinks as main source [15, 16]. However, the contribution of each transmission route remains unknown. Therefore, the aim of this study is to quantify the relative importance of each route at the ICU by applying a mathematical transmission model using admission and discharge dates as well as surveillance culture data of patients.

Methods

Setting

This retrospective study was conducted at the adult ICU wards of the Erasmus MC in Rotterdam, the Netherlands, using data from January 1st, 2010 until May 18th, 2018. The end date of this period was due to the move to a new hospital. In this 1200-bed university hospital, all medical specialties are available. The adult ICU comprised two high-level ICU wards located on the third and the tenth floor of the adults' hospital building, and consisted of a total of 34 single-occupancy rooms, of which 7 with anteroom (i.e., isolation rooms). At the ICU, patients expected to be on a mechanical ventilator for > 48 hours or anticipated to be admitted to the ICU for > 72 hours received selective digestive tract decontamination (SDD) [17]. During the study period, the SDD regimen did not change, nor did the empirical antibiotic therapy regimen.

General IPC measures were installed after each VIM-PA case was identified (e.g., contact isolation; using gloves and gowns when entering the patient room). However, in 2011 these measures were intensified, and twice-weekly screening for VIM-PA was implemented. An overview of all IPC measures implemented or executed during the study period is available in Supplement 1.

Written approval to conduct this study was received from the medical ethical research committee from the Erasmus MC University Medical Center (Erasmus MC), Rotterdam, the Netherlands (MEC-2015-306). All data were anonymized before analysis.

Data

We included all admission data and surveillance cultures from two distinct ICU wards in the time period 01/01/2010 till 18/05/2018. Since HCWs were not shared between wards and no movement of patients were recoded between them, these ICU wards were treated as separate entities with no transmission between them. If the admission date of a patient preceded the study period, it was set to the beginning of the study period. If the discharge date of a patient lied outside the study period, it was set to the end of the study period. We included all results from throat and rectum cultures that were part of regular VIM-PA surveillance. Non-surveillance, clinical cultures were excluded to avoid the introduction of selection bias. All data were de-identified and anonymized prior to the analysis.

Mathematical model

The underlying model is a Susceptible-Infected (SI) model (e.g., [18]). We assumed that all patients admitted to an ICU ward either belong to the susceptible (VIM-PA negative) or colonized (VIM-PA positive) compartment at any given time. The latter includes patients with asymptotic carriage as well as those with a VIM-PA infection. As such, we did not distinguish colonization and infection. In addition, we assumed that every admission is a new patient and once colonized, patients remained colonized with the same level of infectiousness throughout their stay. Events were assumed to occur in daily intervals.

A susceptible patient may enter the ICU already colonized (with probability ϕ), or may become colonized at a certain transmission rate . We assumed two different modes of transmission through which colonization can be acquired. The schematic illustration of the model and the transmission routes is given in Figure 1. Each route induces different patterns in the prevalence time series on the basis of which they may be distinguished statistically. Cross-transmission, i.e., colonization caused by (direct) transmissions from other colonized patients present at the same time on the same ward, is dependent on the fraction of colonized patients in the ward. The probability of colonization due to cross-transmission is therefore proportional to the number of colonized patients present in the ward. Since the mobility of patients in ICUs is usually restricted (due to their health status), cross-transmission typically occurs via temporarily contaminated hands of health-care workers (HCWs). We did not model the population of HCWs explicitly but rather assumed direct patient-to-patient transmission with HCWs representing vectors of transmission. Next to cross-transmission, patients may become colonized at a constant per capita rate α. In general, this transmission route may be due to, persistent environmental contamination, or introductions from other parts of the hospital, or rarely long term HCW carriers. For VIM-PA the main sources of this transmission route are persistently contaminated environments, such as sinks. We will therefore refer to this route as environmental route. The force of infection, i.e., the per capita rate of colonization, is modeled as

$$\lambda(t) - \alpha + \beta \frac{I(t)}{N(t)}$$

where $I(t)$ is the number of colonized patients, $N(t)$ is the total number of patients currently present in the ward at time t, and α and β are the transmission rates for the environmental route and cross-transmission, respectively. Based on these parameters, the proportion of acquired colonizations assigned to each route, i.e., the relative contribution of the transmission routes to the overall number of acquired colonizations can be estimated (e.g., [14]).

Estimation procedure

In the analysis, we used the day of admission and discharge, and the day and result of surveillance cultures as input data for the model. Patients may be admitted to the ward
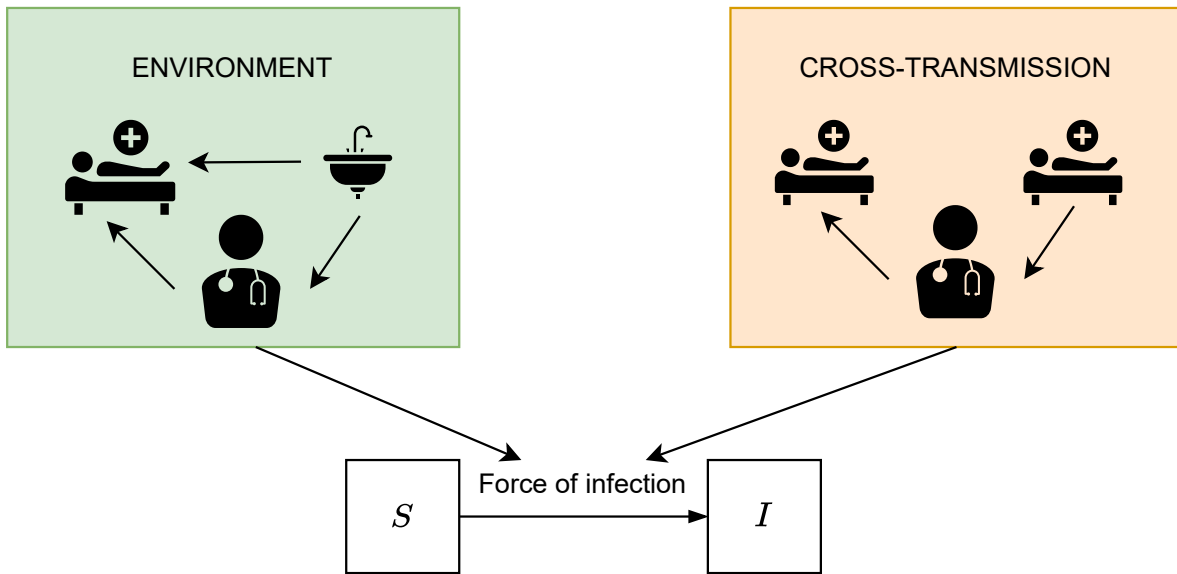
Figure 1: Illustration of the considered transmission routes and the basic transmission model. Patients are assumed to be either susceptible (S) or colonized (I) with VIM-PA. The rate at which susceptible patients may become colonized is represented by the force of infection and dependent on the routes of transmission. Two transmission routes are distinguished: Environmental route (green), mainly caused by transmissions from persistent environmental sources and cross-transmission (orange), i.e., transmissions from other colonized patients. In both routes, HCWs represent vectors of transmission.

either uncolonized or already colonized. The probability of the latter is defined as the importation probability f. The rate at which a susceptible patient may transition to being colonized is given by Eq. (1). The colonization state of a patient is determined by the surveillance cultures provided to the model. Since these culture results are typically intermittent and imperfect, we allow false negative results and colonization results to be imputed in our model. We define the test sensitivity ϕ, i.e., the probability that a colonized patient has a positive result.

We estimated the transmission parameters, the relative contribution of the corresponding transmission routes as well as the importation probability and test sensitivity based on a Bayesian framework using a data-augmented Markov chain Monte Carlo (MCMC) simulation method [11]. The parameters are estimated by fitting the stochastic transmission model to observed data. The main idea is to fit the prevalence pattern resulting from the model to the *observed* timeseries patterns of the prevalence.

Implementation

The MCMC algorithm was run for 1,000,000 iterations. A thinning factor of 10 and a burn-in of 30,000 iterations were used. In each iteration, 20 data-augmentation steps were performed with each augmentation chosen at random. The MCMC algorithm was implemented in C++ and the analysis of the output was performed in *R* (Version 4.0.1) [19]. The data and code are publicly available from:

`https://github.com/tm-pham/transmission_routes_erasmusMC`.

Results

Descriptive data analysis

An overview of the data used in the analysis can be found in Table 1. Since the two considered ICU wards do not differ from each other in terms of admitted patients (i.e., patients were allocated randomly to one of the two ICU wards), we used a combined data set comprising data of both ICU wards for the estimation. The ICU wards were treated as distinct wards with no transmission between them. Data was collected over a study period of 3058 days. In total, 8814 patients were included in the analysis. There were 62 patients with at least one positive culture and 7487 patients with only negative cultures. In total, 829 patients who were admitted to one of the two ICU wards did not have a culture result. The overall median length of stay was 3.0 days. Patients with an observed colonization had a median length of stay of 13.0 days whereas patients that only had negative culture results had a median length of stay of 3.0 days. The number of patients with positive cultures over time are shown in Figure 2.

Inference results

The estimated parameters are reported in Table 2. We estimated that the majority of the VIM-PA colonizations occurred as acquisitions on the wards and that the majority of these transmissions were due to persistent environmental contamination (Figure 3). In particular, of the estimated 58 (95% credibility interval: 45, 72) acquisitions, approximately 50 acquisitions (86.3%, 95% credibility interval: 71%, 98.6%) occurred via this route leaving 8 (13.7%, 95% credibility interval: 1.4%, 29%) acquisitions due to cross-transmission.

Table 1: Descriptive statistics for VIM-PA colonization data collected at Erasmus MC, 2010-2018

	No./Median (IQR)	Percentage (%)
Study period, days	3058	
Admissions	10408	
Number of included patients	8814	
Number of patients with readmissions	1128	
Observed number of patients with positive culture(s) for VIM-PA	62	0.7%
Number of patients with only negative cultures for VIM-PA	7487	84.9%
Number of patients with no cultures	1265	14.4%
Length of stay, days	3.0 (2.0-7.0)	
Observed colonized patients	13.0 (5.0-31.0)	
Observed uncolonized patients	3.0 (2.0-7.0)	
Number of cultures per included patient	6.0 (4.0-15.0)	
Number of cultures per admission	2.0 (1.0-3.0)	

Abbreviations: VIM-PA; Verona Integron-encoded Metallo-beta-lactamase (VIM)-producing *Pseudomonas aeruginosa* (*P. aeruginosa*), IQR; interquartile range.
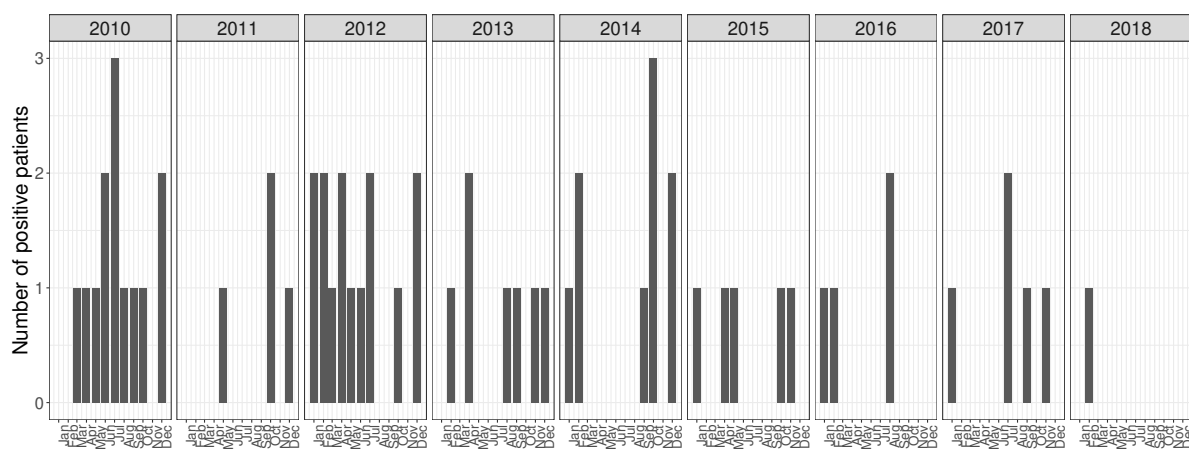
Figure 2: Number of VIM-PA positive patients in the ICU for adults at Erasmus MC, 2010-2018. The date of first positive culture was used. Data for the two ICUs were combined.

Discussion

Our results show that the minority of the transmissions that occurred in the two considered ICU wards was due to cross-transmission. By exclusion, most of the transmissions are estimated to have occurred through persistent environmental contamination. To our

Table 2: Summary statistics of the estimated parameters.

Parameter	Symbol	Mean (95% credibility interval)
Environmental contamination coefficient	α	$6.4 \cdot 10^{-4}(4.1 \cdot 10^{-4}, 9.2 \cdot 10^{-4})$
Cross-transmission coefficient	β	$7.1 \cdot 10^{-3}(5.6 \cdot 10^{-4}, 1.7 \cdot 10^{-2})$
Probability to be colonized on admission (%)	f	0.3 (0.2, 0.5)
Fraction colonized (%)		1.7 (1.6, 1.9)
Test sensitivity (%)	ϕ	98.8 (95.6, 100)
Number of acquisitions		58 (45, 72)
Number of importations[*]		32 (24, 41)
Contributions		
Environmental route (%)	R_α	86.3 (71, 98.6)
Cross-transmission (%)	R_β	13.7 (1.4, 29)

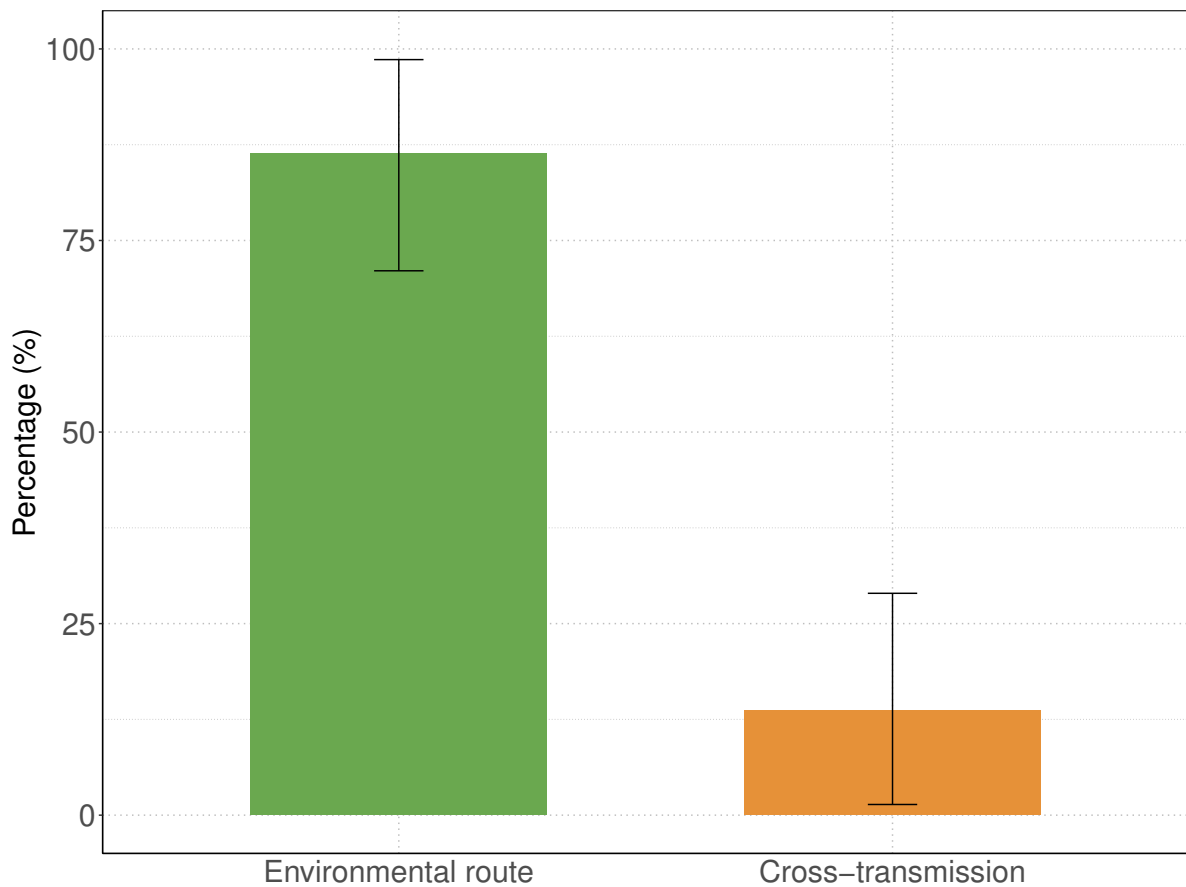[*]Colonizations prior to admission

Figure 3: Estimated relative contributions of transmission routes. The height of the bar shows the mean value, the error bars represent the corresponding 95% credibility intervals for the relative contributions of the transmission routes.

knowledge, this is the first study that quantifies the relative contributions of different transmission routes for VIM-PA and confirms the assumption expressed in Voor in 't holt et al [15], that persistent sources in the hospital environment were the main cause of VIM-PA colonizations.

VIM-PA colonizations have been linked to environmental reservoirs such as sinks in other ICUs (e.g., [20–23]). Kizny Gordon et al [24] summarized studies reporting outbreaks with carbapenem-resistant organisms with a link to the hospital water environment in a systematic review. The authors found that such outbreaks usually involved intensive care settings, the majority of these were caused by *P. aeruginosa*, and that drains, sinks, and faucets were most frequently colonized. Focusing specifically on carbapenem-resistant *P. aeruginosa* outbreaks and all reported sources, Voor in 't holt et al. [9] also showed an over-representation of sinks as reservoirs. While our method is not able to pinpoint to the exact source of colonizations, we were able to show that cross-transmission, and therefore direct transmission from other patients, was an unlikely cause for the majority of transmissions. In fact, we showed that most transmissions were due to sources that caused a constant risk of colonization independent from other colonized patients. HCW themselves may be such risk as was shown by Foca et al [25] who described three HCWs with persistent carriage of *P. aeruginosa* on their hands. However, this was associated with nail extenders, candida onychomycosis and an active otitis externa [25]. In the Erasmus MC, we cultured the hands of ICU HCW on two moments (Supplement 1; February 2010 and May 2011). VIM-PA was not detected in any of these. Furthermore, artificial nails and nail extenders are forbidden in our hospital and were also not observed during the culturing of hands. Therefore, long-term HCW carriers are an unlikely cause and transmissions due to (temporarily) contaminated hands of HCWs have to be linked to other colonized patients present in the respective ward. Visitors may introduce and transmit microorganisms. However, for VIM-PA, we consider this an unlikely cause. VIM-PA among hospitalized patients is already $< 1\%$, among non-hospitalized persons this would be even rarer [26]. Thus, by exclusion, the majority of transmissions are assumed to have occurred by persistent environmental sources, confirming the likely role of environmental contamination in the transmission process of VIM-PA in ICUs. These results may be used in the investigation for outbreaks. In fact, environmental sampling of sinks during this study period revealed that many sinks were found to be persistently

contaminated [16].

Our study encompasses several simplifying assumptions. Firstly, we assumed that every new admission is a new patient. Secondly, while we distinguish two different transmission routes, it is possible that other transmission routes exist that are not included in the model. As explained above, other routes than the environment, such as persistent colonization of HCWs, are highly unlikely. The environment as an exclusion-per-definition-category includes a broad range of sources including equipment and inventory. Microbial genotyping data of surveillance samples would allow the identification of specific transmission routes and more detailed quantification of the relative contribution of the transmission routes. Thirdly, we assumed that the environmental route affects all patients in the ICU ward equally. In reality, patients located close to an environmental reservoir may have an increased risk of colonization that will also depend on the microbial load present in the reservoir. Fourthly, non-surveillance or clinical cultures were excluded from our analysis to avoid selection bias. While this excludes potential information, this would likely only affect the uncertainty of our estimates as the data-augmented MCMC method we used imputes missing colonization times. Finally, we did not include risk factors of different patients and assumed that all patients are equally susceptible to colonization. While the model could be extended to account for the simplifications, this would likely only affect the uncertainty of our estimates but not the main results and conclusions regarding the relative contribution of the transmission routes. We, thus, opted for the simpler model to answer our research question.

In conclusion, using a large longitudinal data set on admission and discharge times as well as surveillance cultures of patients in two ICUs of the Erasmus MC, we were able to quantify the relative importance of cross-transmission and persistent environmental contamination. Our study contributes to the evidence that persistently contaminated environments in hospital wards may be a major cause of VIM-PA outbreaks. To minimize the transmission risk in wards, reservoirs in the environment should be regularly cultured, thoroughly cleaned, and disinfected. In addition, well-designed sinks and taps may minimize the risk of contamination and consequently spill-over from the environment to patients.

Ethics approval and consent to participate

Written approval to conduct this study was received from the Medical Ethical Research Committee of the Erasmus MC University Medical Center, Rotterdam, the Netherlands (MEC-2015-306), and was not subject to the Medical Research Involving Human Subjects Act.

Availability of data and materials

The anonymized data and the full code are publicly available from:
`https://github.com/tm-pham/VIM-PA_transmission_routes_erasmusMC.git`.
More details can be provided by the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' contributions

Conceptualization: TMP, ACB, MK, MV, MCJB, AV, JS
Collecting data: DG, ACB, AV, MV, JS
Analyzed the data: TMP
Interpretation of the data: TMP, ACB, AV, MV, MK, MCJB, JS
Drafted the manuscript: TMP, AV, ACB
All authors read, reviewed, and approved the final manuscript, and all authors have read and agreed to the published version of the manuscript.

Acknowledgements

References

[1] Tacconelli, E, Carrara, E, Savoldi, A, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. In: *The Lancet Infectious Diseases* 18.3 (Mar. 2018), pp. 318–327. doi: `10.1016/S1473-3099(17)30753-3`.

[2] Behzadi, P, Baráth, Z, and Gajdács, M. *It's not easy being green: A narrative review on the microbiology, virulence and therapeutic prospects of multidrug-resistant pseudomonas aeruginosa*. Jan. 2021. doi: `10.3390/antibiotics10010042`.

[3] Saleem, Z, Godman, B, Hassali, MA, et al. Point prevalence surveys of healthcare-associated infections: a systematic review. In: *Pathogens and Global Health* 113.4 (May 2019), pp. 191–205. doi: `10.1080/20477724.2019.1632070`.

[4] Persoon, MC, Voor In'T Holt, AF, Wielders, CC, et al. Mortality associated with carbapenem-susceptible and Verona Integron-encoded Metallo-β-lactamase-positive Pseudomonas aeruginosa bacteremia. In: *Antimicrobial Resistance and Infection Control* 9.1 (Feb. 2020), pp. 1–8. doi: `10.1186/s13756-020-0682-4`.

[5] Persoon, MC, Voor In 'T Holt, AF, Van Meer, MP, et al. Mortality related to Verona Integron-encoded Metallo-β-lactamase-positive Pseudomonas aeruginosa: Assessment by a novel clinical tool. In: *Antimicrobial Resistance and Infection Control* 8.1 (June 2019), pp. 1–8. doi: `10.1186/s13756-019-0556-9`.

[6] Montero, MM, Montesinos, IL, Knobel, H, et al. Risk Factors for Mortality among Patients with Pseudomonas aeruginosa Bloodstream Infections: What Is the Influence of XDR Phenotype on Outcomes? In: *Journal of Clinical Medicine* 9.2 (Feb. 2020), p. 514. doi: `10.3390/JCM9020514`.

[7] Nordmann, P and Poirel, L. Epidemiology and Diagnostics of Carbapenem Resistance in Gram-negative Bacteria. In: *Clinical Infectious Diseases* 69.Supplement_7 (Nov. 2019), S521–S528. doi: `10.1093/cid/ciz824`.

[8] Yoon, EJ and Jeong, SH. Mobile Carbapenemase Genes in Pseudomonas aeruginosa. In: *Frontiers in Microbiology* 0 (Feb. 2021), p. 30. doi: `10.3389/FMICB.2021.614058`.

[9] Voor in 't holt, AF, Severin, JA, Lesaffre, EM, et al. A Systematic Review and Meta-Analyses Show that Carbapenem Use and Medical Devices Are the Leading Risk Factors for Carbapenem-Resistant Pseudomonas aeruginosa. In: *Antimicrobial Agents and Chemotherapy* 58.5 (2014), pp. 2626–2637. doi: `10.1128/AAC.01758-13`.

[10] Breathnach, AS, Cubbon, MD, Karunaharan, RN, et al. Multidrug-resistant Pseudomonas aeruginosa outbreaks in two hospitals: Association with contaminated hospital waste-water systems. In: *Journal of Hospital Infection* 82.1 (Sept. 2012), pp. 19–24. doi: `10.1016/j.jhin.2012.06.007`.

[11] Cooper, BS, Medley, GF, Bradley, SJ, et al. An augmented data method for the analysis of nosocomial infection data. In: *American Journal of Epidemiology* 168.5 (Sept. 2008), pp. 548–557. doi: `10.1093/aje/kwn176`.

[12] Forrester, ML, Pettitt, AN, and Gibson, GJ. Bayesian inference of hospital-acquired infectious diseases and con-

trol measures given imperfect surveillance data. In: *Biostatistics* 8.2 (Apr. 2007), pp. 383–401. doi: `10 . 1093 / biostatistics/kxl017`.

[13] Worby, CJ, Jeyaratnam, D, Robotham, JV, et al. Estimating the effectiveness of isolation and decolonization measures in reducing transmission of methicillin-resistant staphylococcus aureus in hospital general wards. In: *American Journal of Epidemiology* 177.11 (June 2013), pp. 1306–1313. doi: `10 . 1093 / aje / kws380`.

[14] Pham, TM, Kretzschmar, M, Bertrand, X, et al. Tracking Pseudomonas aeruginosa transmissions due to environmental contamination after discharge in ICUs using mathematical models. In: *PLoS Computational Biology* 15.8 (2019), e1006697. doi: `10 . 1371 / journal . pcbi.1006697`.

[15] Voor in 't holt, AF, Severin, JA, Hagenaars, MB, et al. VIM-positive Pseudomonas aeruginosa in a large tertiary care hospital: Matched case-control studies and a network analysis. In: *Antimicrobial Resistance and Infection Control* 7.1 (Feb. 2018), pp. 1–10. doi: `10 . 1186 / s13756-018-0325-1`.

[16] Pirzadian, J, Harteveld, SP, Ramdutt, SN, et al. Novel use of culturomics to identify the microbiota in hospital sink drains with and without persistent VIM-positive Pseudomonas aeruginosa. In: *Scientific Reports 2020 10:1* 10.1 (Oct. 2020), pp. 1–12. doi: `10 . 1038 / s41598-020-73650-8`.

[17] De Smet, AMG, Kluytmans, JA, Blok, HE, et al. Selective digestive tract decontamination and selective oropharyngeal decontamination and antibiotic resistance in patients in intensive-care units: An open-label, clustered group-randomised, crossover study. In: *The Lancet Infectious Diseases* 11.5 (May

2011), pp. 372–380. doi: `10 . 1016 / S1473-3099(11)70035-4`.

[18] Diekmann, O, Heesterbeek, H, and Britton, T. Mathematical tools for understanding infectious diseases dynamics. In: (2013).

[19] R Core Team. *R: A Language and Environment for Statistical Computing*. 2020. `https://www.r-project.org/`.

[20] Ambrogi, V, Cavalié, L, Mantion, B, et al. Transmission of metallo-β-lactamase-producing Pseudomonas aeruginosa in a nephrology-transplant intensive care unit with potential link to the environment. en. In: *Journal of Hospital Infection*. Multi-drug Resistant Gram Negative Bacteria 92.1 (Jan. 2016), pp. 27–29. doi: `10.1016/j.jhin.2015.09.007`.

[21] Catho, G, Martischang, R, Boroli, F, et al. Outbreak of Pseudomonas aeruginosa producing VIM carbapenemase in an intensive care unit and its termination by implementation of waterless patient care. In: *Critical Care 2021 25:1* 25.1 (Aug. 2021), pp. 1–10. doi: `10.1186/S13054-021-03726-Y`.

[22] Geyter, DD, Vanstokstraeten, R, Crombé, F, et al. Sink drains as reservoirs of VIM-2 metallo-β-lactamase-producing Pseudomonas aeruginosa in a Belgian intensive care unit: relation to patients investigated by whole-genome sequencing. In: *Journal of Hospital Infection* 115 (Sept. 2021), pp. 75–82. doi: `10.1016/ J.JHIN.2021.05.010`.

[23] Knoester, M, Boer, Md, Maarleveld, J, et al. An integrated approach to control a prolonged outbreak of multidrug-resistant Pseudomonas aeruginosa in an intensive care unit. In: *Clinical Microbiology and Infection* 20.4 (Apr. 2014), O207–O215. doi: `10.1111/1469-0691. 12372`.

[24] Kizny Gordon, AE, Mathers, AJ, Cheong, EYL, et al. The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections—A Systematic Review of the Literature. In: *Clinical Infectious Diseases* 64.10 (May 2017), pp. 1435–1444. doi: 10 . 1093 / CID / CIX132.

[25] Foca, M, Jakob, K, Whittier, S, et al. Endemic Pseudomonas aeruginosa Infection in a Neonatal Intensive Care Unit. In: *New England Journal of Medicine* 343.10 (Aug. 2000), pp. 695–700. doi: 10.1056/nejm200009073431004.

[26] Greef, Sd and Mouton, J. Consumption of antimicrobial agents and antimicrobial resistance among medically important bacteria in the Netherlands; Monitoring of Antimicrobial Resistance and Antibiotic Usage in Animals in the Netherlands in 2012. In: *RIVM official reports* (2018), p. 88. doi: 10.21945/RIVM-2018-0046.

Supplementary material

Table of Contents

Supplement 1: Overview of infection prevention and control (IPC) measures installed to prevent transmission of Verona Integron-encoded metallo-beta-lactamase (VIM)-producing Pseudomonas aeruginosa (VIM-PA) at the 2 included intensive care units (ICU); January 2010 until May 18, 2018.

Table S1: ICP measures implemented for a period of time at the ICU.

IPC measure	Start date (month-year)	End date (month-year)
General use of re-usable gowns by HCW	Jan-10	Sep-11
General use of gloves and gowns by HCW and visitors (preemptive contact isolation) of all patients admitted at the ICUs	Oct-11	Mar-12
General use of gloves and gowns by HCW when having physical patient contact (i.e., not when changing an intravenous fluid bag)	Apr-12	May-18
The patient room of every patient identified with VIM-PA: daily cleaning, and disinfection with 250 ppm chlorine after discharge	Aug-10	May-18
Emphasis (communication by the IPC team) on separation of clean materials from dirty sinks	Feb-11	May-18
VIM-PA screening (throat and rectum) of patients on admission and at discharge	Aug-11	May-18
VIM-PA screening (throat and rectum) of patients during hospitalization (twice weekly)	Aug-11	Sep-14
VIM-PA screening (throat and rectum) of patients during hospitalization (weekly)	Oct-14	Dec-14
VIM-PA screening (throat and rectum) of patients during hospitalization (twice weekly)	Jan-15	May-18
Electronic flagging in the electronic patient fil of all VIM-PA positive patients	Sep-11	May-18
Only allowed to use single-use wash gloves at the ICU. Dec-11 May -18 Discontinuation usage of tap water at the ICU, only usage of bottled water allowed.	Dec-11	May-18
Installation of sink drain plugs as physical barriers against splashing to prevent transmission of VIM-PA from drain reservoirs to the surrounding sink environment		
Pre-intervention phase	Jan-13	Aug-13
Intervention phase	Jul-13	Sep-13
Post-intervention phase	Aug-13	Jun-14

Abbreviations: HCW; healthcare workers, ICU; intensive care unit, VIM-PA; Verona Integron-encoded metallo-beta-lactamase (VIM)-producing Pseudomonas aeruginosa, OMT; outbreak management team, CVVH; Continuous Veno-Venous Hemofiltration, ppm; parts per million, IPC; infection prevention and control, H2O2; hydrogen peroxide.
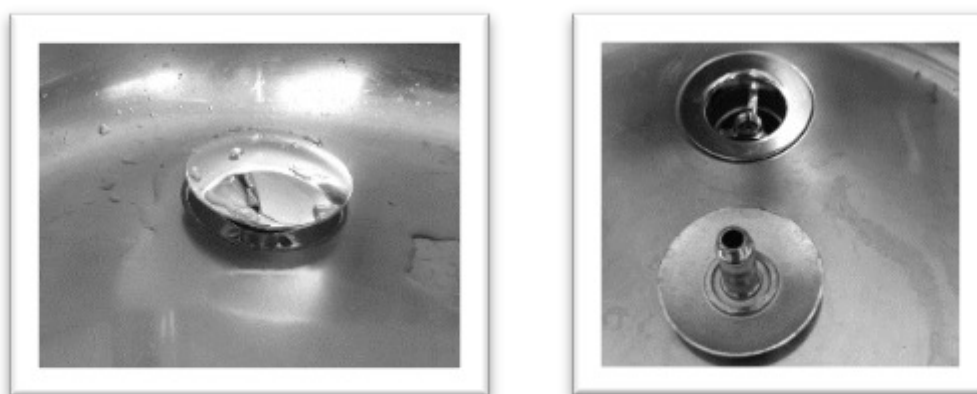
Figure S1: Sink drain plugs as physical barriers against splashing to prevent transmission of VIM-PA from drain reservoirs to the surrounding sink environment.

Table S2: ICP measures implemented for a period of time at the ICU.

IPC measure	Month-year
VIM-PA screening of the ICU environment	Jan-10, Apr-10, Aug-10, Sep-10, Oct-10, May-11, Aug-11, Sep-11, Oct-11, Nov-11, Dec-11, Jan-12, Jan-13, Jun-13, Dec-13, Feb-14, Jun-15, Nov-15, Jan-16, Feb-16, Mar-16, Apr-16, Jul-17, Aug-17
OMT meetings	Sep-10, Jan-11, Feb-11, Mar-11, Apr-11, May-11, Jul-11, Aug-11, Sep-11, Oct-11, Nov-11, Dec-11, Feb-12, Mar-12, Jun-12, Aug-12, Sep-12, Jan-13, Mar-14, Aug-14, Sep-14, Oct-14, Apr-15
VIM-PA screening of hands of healthcare workers (all HCW employed at one ICU)	Feb-10
Sampling of CVVH machines	Feb-10, Aug 12
Replacement of all siphons and drains, including cleaning and disinfection of patient rooms afterwards.	Oct-10
VIM-PA screening (throat and rectum) of all admitted ICU patients	Apr-11
VIM-PA screening of enteral feeding	Aug-11
H2O2 disinfection of patient rooms of one ICU ward	Aug-11
VIM-PA screening of throat and rectum of all employed HCW at both ICUs	Oct-11, Nov-11
International expert meeting	Dec-11
Regional feedback meeting	Dec-14

Abbreviations: HCW; healthcare workers, ICU; intensive care unit, VIM-PA; Verona Integron-encoded metallo-beta-lactamase (VIM)-producing Pseudomonas aeruginosa, OMT; outbreak management team, CVVH; Continuous Veno-Venous Hemofiltration, ppm; parts per million, IPC; infection prevention and control, H2O2; hydrogen peroxide.

Chapter 5

The contribution of hospital-acquired infections to the COVID-19 epidemic in England in the first half of 2020

Gwenan M. Knight, Thi Mui Pham, James Stimson, Sebastian Funk, Yalda Jafari, Diane Pople, Stephanie Evans, Mo Yin, Colin S. Brown, Alex Bhattacharya, Russell Hope, Malcolm G. Semple, ISARIC4C Investigators*, CMMID COVID-19 working group+, Jonathan M. Read, Ben S. Cooper, Julie V. Robotham

Abstract

Background: SARS-CoV-2 spreads in hospitals, but the contribution of infections acquired in hospitals to the total burden at a national level is unknown.

Methods: We used comprehensive national English datasets to determine the number of COVID-19 patients with identified hospital-acquired infections (with symptom onset >7 days after admission and before discharge) in acute English hospital facilities up to August 2020. As patients may leave the hospital prior to detection of infection or have rapid symptom onset, we combined measures of the length of stay and estimates of the incubation period distribution, to estimate how many hospital-acquired infections could be missed. Combining these two measures, we used simulations to estimate the total number (identified and unidentified) of symptomatic hospital-acquired infections, as well as infections due to onward transmissions from missed hospital-acquired infections, to 31st July 2020.

Results: In our dataset of hospitalized COVID-19 patients in acute English Trusts with a recorded symptom onset date (n = 65,028), 7% were classified as hospital-acquired. We estimated that only 30% (range across weeks and 200 simulations: 20-41%) of symptomatic hospital-acquired infections would be identified, with up to 15% (mean, 95% range over 200 simulations: 14.1%-15.8%) of cases currently classified as community-acquired COVID-19 potentially linked to hospital transmission.
We estimated that 26,600 (25,900 to 27,700) individuals acquired a symptomatic SARS-CoV-2 infection in an acute Trust in England before 31st July 2020, resulting in 15,900 (15,200-16,400) or 20.1% (19.2%-20.7%) of all identified hospitalized COVID-19 cases.

Conclusions: Transmission of SARS-CoV-2 to hospitalized patients likely caused approximately a fifth of identified cases of hospitalized COVID-19 in the "first wave" in England, but less than 1% of all infections in England. Using symptom onset as a detection method likely misses a substantial proportion (>60%) of hospital-acquired infections.

Keywords: COVID-19, SARS-CoV-2, nosocomial transmission, mathematical modelling

Introduction

The SARS-CoV-2 pandemic is a global public health priority [1]. Based on experience with other highly pathogenic coronaviruses within-hospital transmission can occur and hospitals may play an important role in amplifying transmission [2]. Moreover, many patients acquiring SARS-CoV-2 in hospitals are at high risk for severe outcomes and subsequent mortality [3]. Quantifying hospital-acquired transmission of SARS-CoV-2 is thus important both for prioritising control efforts and for understanding the contribution of hospitals to sustaining the community epidemic.

SARS-CoV-2 transmission in healthcare settings has been reported in many countries [3–6]. As the precise time of infection is rarely known, establishing whether an infection is hospital-acquired remains a challenge. For SARS-CoV-2, hospital-acquired infections are usually defined by comparing the time of admission and subsequent symptom onset [7] or first positive test [8]. If the delay is much longer than the incubation time, then it is likely that an infection is hospital-acquired. Thus, the proportion of patients with a hospital-acquired SARS-CoV-2 infection will depend on the definition used, with uncertainty driven by the unobservable nature of infection and the incubation period distribution. Records for all hospitals in England, using testing data and standard definitions (of first positive test more than 14 days from admission), indicate that 15% of detected SARS-CoV-2 infections in hospitalized patients could be attributed to hospital-acquired transmission [8] with analysis of data from single hospitals suggesting a similar level [3, 9].

In the absence of frequent universal testing of all inpatients, many hospital-acquired SARS-CoV-2 infections will not be identified by hospitals prior to discharge. Even with regular PCR testing of all inpatients regardless of symptoms we would expect to miss many infections because of short patient stays and potentially low PCR sensitivity 1-2 days after infection [10].

In the spring of 2020 in England, the majority of inpatient testing only occurred in those with symptoms, either on admission or during hospital stay [11]. Many patients who develop a symptomatic infection will do so after discharge (Figure 1) as hospital stays are typically shorter than the interval from infection to symptom onset (median length of stay = 2.4 days, standard deviation = 0.4 days, for non-COVID patients in England vs. incubation period average of 5.1 days [12]). Thus, there may be a considerable propor-

tion of hospital-acquired infections that remained unidentified. Its magnitude and further transmission to the community has been difficult to quantify. Additionally, a substantial proportion of infected individuals never progress to be symptomatic [13].

In this analysis, we used national, patient-level data sets of patients hospitalized with COVID-19 to estimate the contribution of hospital settings to the first wave of COVID-19 in England. We estimated the proportion of symptomatic hospital-acquired infections that have not been identified as hospital-acquired and modelled onward transmission from these unidentified infections to the community. We hence quantified the likely contribution of hospital-acquired infections to the first wave of SARS-CoV-2 infections in England.
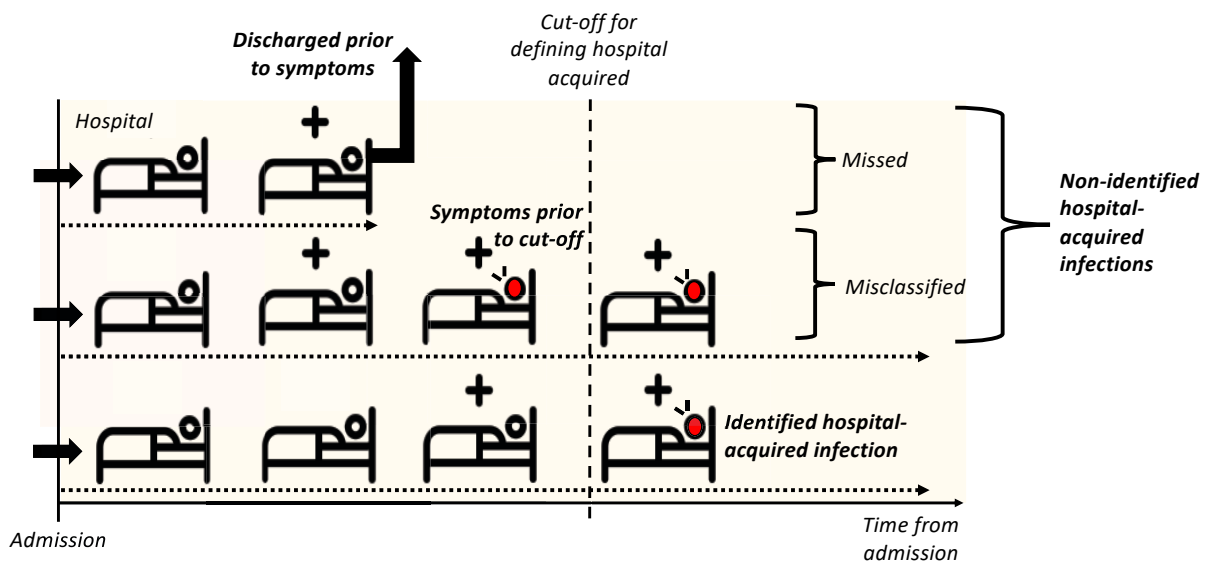
Figure 1: How might we underestimate hospital-acquired (HA) infections? With no asymptomatic screening in hospitals, detection of a hospital-acquired case relies on symptom onset prior to patient discharge. In the schematic a "+" above the bed denotes a hospital-acquired infection, and a red patient denotes one with symptoms. A patient with COVID-19 identified as being due to a hospital-acquired infection is one with symptom onset after a defined cut-off (e.g. > 7 days from admission to symptom onset but prior to discharge, bottom row patient). Patients with unidentified hospital-acquired infections are those with a symptom onset after discharge (top row patient, "missed") or those with symptom onset prior to the defined cut-off (middle row patient, "misclassified"). We focus on symptomatic infections: there will also be unidentified asymptomatic hospital-acquired infection which we do not include. We estimated that fewer than 1% of individuals with symptom onset > 7 days from admission will have been infected in the community.

Methods

Our primary aim was to estimate the total number of symptomatic hospital-acquired SARS-CoV-2 infections in England from 1st January to 31st July 2020. For each identified hospital-acquired infection, we estimated how many were unidentified. Our secondary aim was to estimate the contribution of these unidentified hospital-acquired infections to the community epidemic.

All analyses were conducted in R version 4.0.3 (14) with code available on Github [14]. The steps in the analysis (a - e) are outlined below and illustrated in Figure 2.

Data sources

The healthcare system in the UK is represented by the National Health System (NHS). NHS services are mainly provided by NHS Trusts, i.e., collections of hospitals (departments, buildings and facilities) that function as a single administrative unit. Acute medical care Trusts are defined as an NHS Trust with only acute hospitals (as opposed to Community or Mental Health facilities). In this study, we used two data sources on COVID-19 patients admitted to NHS Trusts (Supplementary 2). The first is the ISARIC4C UK COVID-19 Clinical Information Network (CO-CIN) study [15], a national cohort of COVID-19 patients collected in 208 acute Trusts in England, Scotland, and Wales up to 3rd December 2020, representing approximately two thirds of COVID-19 UK admissions during the first wave of SARS-CoV-2 infection. While not all NHS Trusts are represented in the data (as some have specialist roles that do not involve inpatient acute medical care), our CO-CIN extract comprised 208 of 223 acute medical care Trusts [16, 17]. We included 126 English Trusts and filtered the dataset for patients with a symptom onset before 1st August 2020. CO-CIN recorded admission date, discharge date, and earliest date of symptom onset for patients. We excluded CO-CIN participants without a recorded admission and symptom onset date (Supplementary 2).

The second is the SUS dataset [18] which contains data on all patient admissions and discharges for all Trusts in England. The SUS data were linked with testing data (Second

Generation Surveillance System (SGSS)) [18] to derive length of stay distributions for non-COVID-19 patients and total COVID-19 hospital admissions by week and NHS Trust.

These two data sources have their respective strengths and limitations. The CO-CIN data include information on the date of symptom onset [19] but is only a subset, albeit the majority, of all hospitalized COVID-19 patients, while the linked SUS/SGSS data include all known hospitalized COVID-19 patients but lack information on symptom onset date. Symptom onset dates do not rely on knowledge of testing regimens which vary over time and between Trusts. To address these different shortcomings, we decided to use SUS data to adjust CO-CIN information to account for enrolment variation between settings, resulting in a database combining the best features of both.

Setting

Our baseline population is all acute English Trusts in CO-CIN. These are aggregated as a single "England" population for our main analysis. A sensitivity analysis modelled the individual acute Trust level prior to aggregation (Supplementary 12).

Length of stay distribution

We used empirical length of stay (LoS) estimates for non-COVID-19 patient stays from SUS for each English acute Trust in CO-CIN for patients admitted each week (Supplementary 2). For a LoS distribution for England, LoS estimates across all including Trusts were pooled by week. The average length of stay was between 1.5 and 2.5 days.

Analysis steps

a. Identifying COVID-19 cases as infected in a hospital

The number of hospital-acquired COVID-19 cases per day in each Trust was estimated by comparing the dates of symptom onset and hospital admission for each patient provided by CO-CIN. Our analysis used a 7 day cut-off: we defined an *identified* hospital-acquired infection as an inpatient with symptoms onset more than 7 days after admission (Table 1) aligned with English definitions and the ECDC definition for a Probable (8-14) and Definite (> 14 days) healthcare-associated COVID-19 case [7, 20]. In sensitivity analyses we explored cut-offs of 4 and 14 days.

b. Accounting for enrolment into CO-CIN

We accounted for the fact that only a subset of all hospitalized COVID-19 patients was enrolled in CO-CIN as follows: We calculated the proportion of COVID-19 patients recorded in SUS in a given week that were included in the corresponding CO-CIN data. We then weighted the weekly estimates of the number of hospital-acquired infections from the CO-CIN data using the inverse of these weekly proportions to obtain estimates of identified hospital-acquired COVID-19 cases corrected for under-reporting in CO-CIN (Supplementary 4). Our method assumes that there is no bias in enrolment of hospital versus community-onset cases.

c. Proportion of hospital-acquired infections that are identified

Not all cases of COVID-19 are identified (e.g., some individuals are infected with SARS-CoV-2 in a hospital and subsequently have symptoms that are not confirmed to be COVID-19). All identified cases of COVID-19 with symptom onset in a hospital setting are classified as either hospital- or community-acquired. However, some are *misclassified* (e.g., those that are infected in a hospital but have a symptom onset prior to the cut-off threshold for defining hospital-acquired cases). Our aim was to estimate both overlooked symptomatic SARS-CoV-2 infections that were not identified and that were misclassified (Figure 1, Table 1). We did not consider those who acquire an infection but remain asymptomatic.

To calculate the proportion of symptomatic hospital-acquired infections that were identified as such, we calculated the probability that a patient with a hospital-acquired infection has a symptom onset that falls within the definition period, i.e., before discharge and after the cut-off threshold (Figure 1). The calculations were based on the incubation period of SARS-CoV-2 (Table 2), length of stay distribution of non-COVID-19 patients and assumed that all infections led to a symptom onset: hence it is the proportion of hospital-acquired infected individuals that will ever have symptoms and are identified (Supplementary 5). Uncertainty was included by sampling from parameter distributions (Table 2, Supplementary 10).

We estimated that fewer than 1% of inpatients with symptom onset 5 or more days

after admission were latently infected when admitted i.e., hospital-onset, community-acquired (Table 1, Supplementary 3). In short, this low number is due to relatively low prevalence in non-COVID-19 admissions during the first wave of the COVID-19 epidemic in the UK, the short length of stay and the fact that, in this work, we are only concerned with symptomatic infections. Hence, our definition of "misclassified" only considers those infections that are hospital-acquired and misclassified as community-acquired.

d. Reclassifying community-acquired COVID-19 cases as hospital-acquired

The number of patients with unidentified hospital-acquired infections was calculated by multiplying the number of identified hospital-acquired cases by the inverse of the proportion that were estimated to be identified (Figure 2). To determine the contribution of these unidentified hospital-acquired infections to the hospital burden of COVID-19 cases, we simulated their return as a COVID-19 hospital admission: We estimated the entire disease progression trajectory for each unidentified "missed" hospital-acquired infection by sampling from known natural history distributions (Figure 2). For each patient estimated to have had an unidentified "missed" hospital-acquired infection, we sampled a time from infection to discharge using the length of stay distribution of non-COVID patients (Supplementary 8) and assumed a date of discharge of five days before the detection date of the associated identified COVID-19 case (Figure 2c). This corresponds to the difference in the average length of stay of identified SARS-CoV-2 positive cases (approx. 7 days) and those thought to be SARS-CoV-2 negative (approx. 2 days) in SUS. In a sensitivity analysis, we explored the impact of this parameter by setting it to one day. From this date of discharge, we estimated what proportion of these unidentified "missed" infections would have been expected to return as a hospitalized COVID-19 case and when this would be. The proportion expected to return varied for each simulation (Figure 2, Supplementary 6). Recalling exact dates of symptom onset is difficult, hence we used a scenario analysis to explore three different distributions for the symptom onset to hospitalisation parameter (Table 2, Supplementary 7).

e. Hospital-linked cases

We defined a "hospital-linked infection" as an infection that occurred in the community but caused by a patient that was estimated to have had an unidentified "missed"

hospital-acquired infection. This time series of community infections was calculated by estimating four generations of onward infections under varying assumptions about the reproduction number (Supplementary 6). This is approximately the number of infections caused within one month after discharge (approx 6.7-day serial interval, Supplementary 6).

We explored three reproduction number values: 1) a constant value of 0.8, 2) a constant value of 1.2 both with a range generated as +/- 5% of the constant value, and 3) a time-varying estimate "R_t" for which we used upper/lower bounds for the 50% credible interval from a publicly available repository [21] (Supplementary 9).
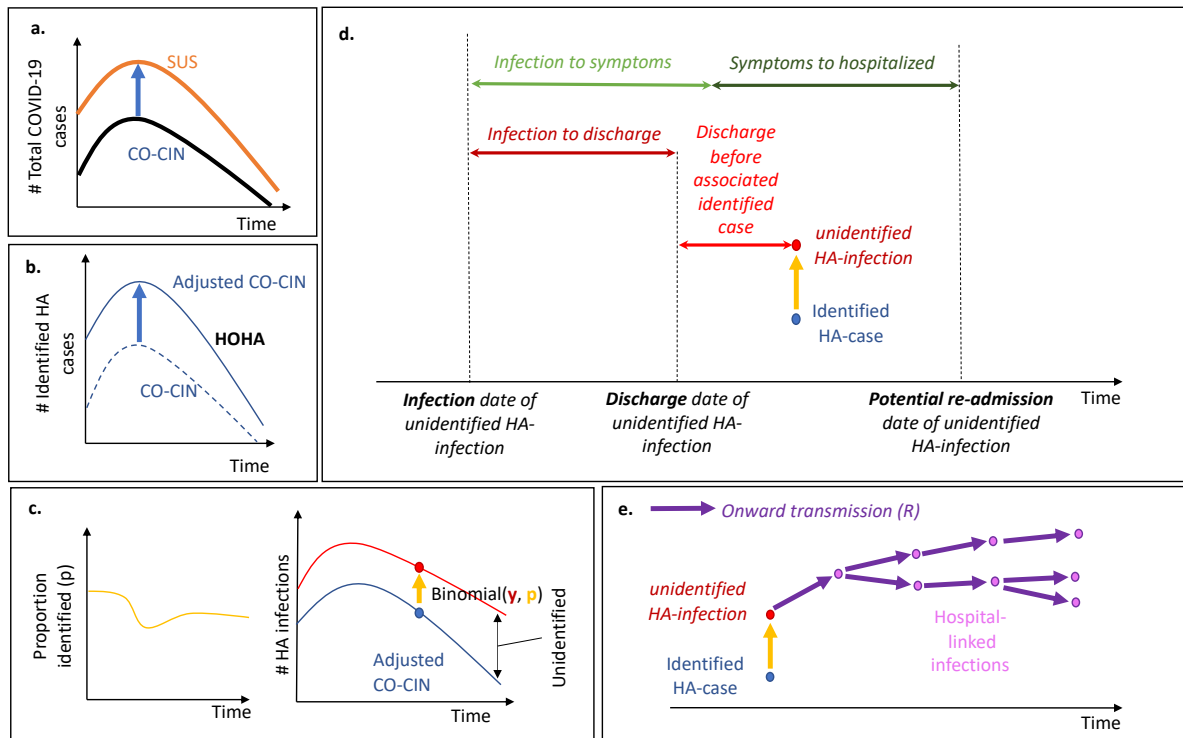
Figure 2: The analysis steps: (a) CO-CIN is inflated to match total COVID-19 hospitalized cases in SUS. (b) The same weekly adjustment is used to estimate the number of identified hospital-onset, hospital-acquired (HOHA) cases. (c) The length of stay for non-COVID-19 hospital patients and incubation period distribution is used to generate estimates of the proportion of hospital-acquired infections that would be identified (Figure 1). This proportion (p) is used to estimate how many unidentified hospital-acquired infections there would be for each identified hospital-onset hospital-acquired infection by assuming a Binomial distribution and calculating the number of "trials" or "unidentified" hospital-acquired infections there were. (d) The unidentified hospital-acquired infections with symptom onset after discharge ("missed") may return to hospital as a COVID-19 case: the trajectory of their disease is calculated to determine their contribution to hospitalized cases. (e) These "missed" unidentified hospital-acquired infections are assumed to contribute to onward transmission in the community: here we capture four generations of transmission to estimate the number of hospital-linked infections and subsequent hospitalized cases under different R estimates.

Table 1: **Case definitions.** Terms are distinguished between surveillance definitions and quantities estimated in the analysis. Additional definitions are given in Supplementary 1.

	Term	Acronym	Classification	Explanation
Surveillance	hospital-onset, hospital-acquired case	HOHA	An individual hospitalized with COVID-19 with symptom onset after a defined cut-off of days from admission and prior to discharge	An individual identified with COVID-19 in a hospital that was presumed to be infected with SARS-CoV-2 in the hospital.
Surveillance	community-onset, community-acquired case	COCA	A hospitalized COVID-19 case with a symptom onset before a defined cut-off of days from admission and prior to discharge	An individual with identified COVID-19 in the hospital or community that was presumed to be infected with SARS-CoV-2 in the community.
Surveillance	Cut-off days for definition of hospital-acquired infection (in identified cases)		7 days (4 or 14 used in sensitivity analysis)	If symptoms onset occurs after this number of days from admission but before discharge then the case is identified as hospital-acquired.
To be estimated	unidentified hospital-acquired infection	"missed"	A person infected with SARS-CoV-2 during a hospital stay but not identified as symptom onset was after the patient was discharged	Our model estimates how many patients with hospital-acquired infections would be unidentified by using a definition of hospital-acquired that relies on symptom onset prior to discharge. We do not consider asymptomatic infections. We did not consider community-acquired infections "misclassified" as hospital-acquired as the percentage is very small after only a few days from admission (Supplementary 3, see HOCA below).
		"misclassified"	A person infected with SARS-CoV-2 during a hospital stay but not identified as symptom onset was before the defined cut-off.	
To be estimated	Total number of patients with hospital-acquired infections		A person infected with SARS-CoV-2 during a hospital stay	The combined total of identified (those with symptom onset after a defined cutoff) and unidentified infections ("missed" and "missclassified").
To be estimated	community-onset, hospital-acquired case	COHA	A hospitalized community-onset COVID-19 case that has a community-acquired classification but was actually a unidentified hospital-acquired infection.	Our model prediction of how many unidentified hospital-acquired infections would return as a hospitalized COVID-19 case. These need to be re-classified as hospital- not community-acquired.

| To be estimated | community-onset, hospital-linked case | COHL | A hospitalized community-onset COVID-19 case that was infected in a chain of four generations of transmission that started with an unidentified hospital-acquired infection. | Our model prediction of the contribution of unidentified hospital infections to onward community transmission approximately one month after discharge. |
| Minimal | Hospital-onset, community-acquired case | HOCA | Symptoms after the cutoff for defining hospital-acquired, but infection was in the community. | We estimate that less than 1% of those with symptom onset more than 5 days from admission would have a community-acquired infection (Supplementary 3) |

Table 2: Parameters values used in the model. See Supplementary 6 for more details.

Definition	Values/Distributions	References
Proportion of individuals with unidentified hospital-acquired infections that will be subsequently admitted to hospital with COVID-19	Unif(range = 0.1-0.15)	[22–24]
Proportion of community infections that will be hospitalized cases of COVID-19	Normal(0.035, 0.0005)	[22]
Proportion of community infections that will be hospitalised cases of COVID-19	norm(0.035, 0.0005)	[22]
Time to symptom onset from infection (incubation distribution)	mean distribution: lognormal(mean = 1.62, sd = 0.4) standard deviation in estimates of mean and standard deviation: 0.064 (mean), 0.0691 (sd)	[12]
Time to hospitalisation from symptom onset	*Scenario 1:* Lognormal (mean = 1.66, sd = 0.89) *Scenario 2:* gamma(shape = 7, scale = 1) *Scenario 3:* Lognormal (mean = 1.44, sd = 0.72)	Supplementary 7, [19, 24, 25]
Time from infection to hospitalisation	Sum of means of infection to symptom onset and symptom onset to hospitalisation = 5.1 + 7 = 12.1 days	
Average number of secondary infections from one infected individual in the community (R)	0.8, 1.2 and "rt"	Supplementary 9, [21]
Time period over which an infected individual is infectious	gamma(shape = 4, scale = 0.875)	[24]
Number of days before associated identified hospital-acquired case detection that a patient with a unidentified "missed" hospital-acquired infection is discharged from hospital	5 or 1	Assumptions

Results

Identified and classified hospital-acquired cases

In CO-CIN, using a symptom onset-based definition, we found 7% of COVID-19 symptomatic cases (i.e., 4552 of n = 65,028) in acute English Trusts were identified and classified as hospital-acquired (having a symptom onset 8 or more days after admission and before discharge) before 31st July 2020. By adjusting for enrolment in CO-CIN (Figure 2b), we estimated that with this same cut-off, there were 6,640 "hospital-onset, hospital-acquired" identified cases across acute English Trusts up to the 31st July 2020.

Proportion of infections identified

We estimated 30% (20-41%, range across weeks and sampling, Supplementary 10) of symptomatic hospital-acquired infections (using a 7 day cut-off) were identified using a symptom onset based definition for England. Across all acute English Trusts the range was 0-82% (Figure 3). The proportion identified decreased with increasing cut-off day from admission (Figure 3c). The estimates are highly sensitive to LoS distributions (Supplementary 2). These results imply that for every single identified hospital-acquired COVID-19 case (using a 7 day cut-off) there were, on average, two unidentified symptomatic hospital-acquired infections (Figure 1 and 2).

Contribution of missed infections

We estimated that across England, 20,000 (mean; 95% range over 200 simulations: 19,200, 21,100) hospital-acquired infections were unidentified from acute Trusts if a 7 day symptom-based cut-off was used to identify hospital-acquired cases (C + D in Figure 5). The majority of patients with unidentified hospital-acquired infections were not identified due to the discharge of the infected patient prior to symptom onset ("missed") (Figure 1 and 3c): 12,300 (11,400, 13,400) in total.

A proportion of the patients with unidentified hospital-acquired infections that have symptom onset after discharge will return as hospitalized cases: we found 1,500 (1,200, 1,900) or 2.1% (1.7%, 2.6%) of cases originally classified as "community-onset,
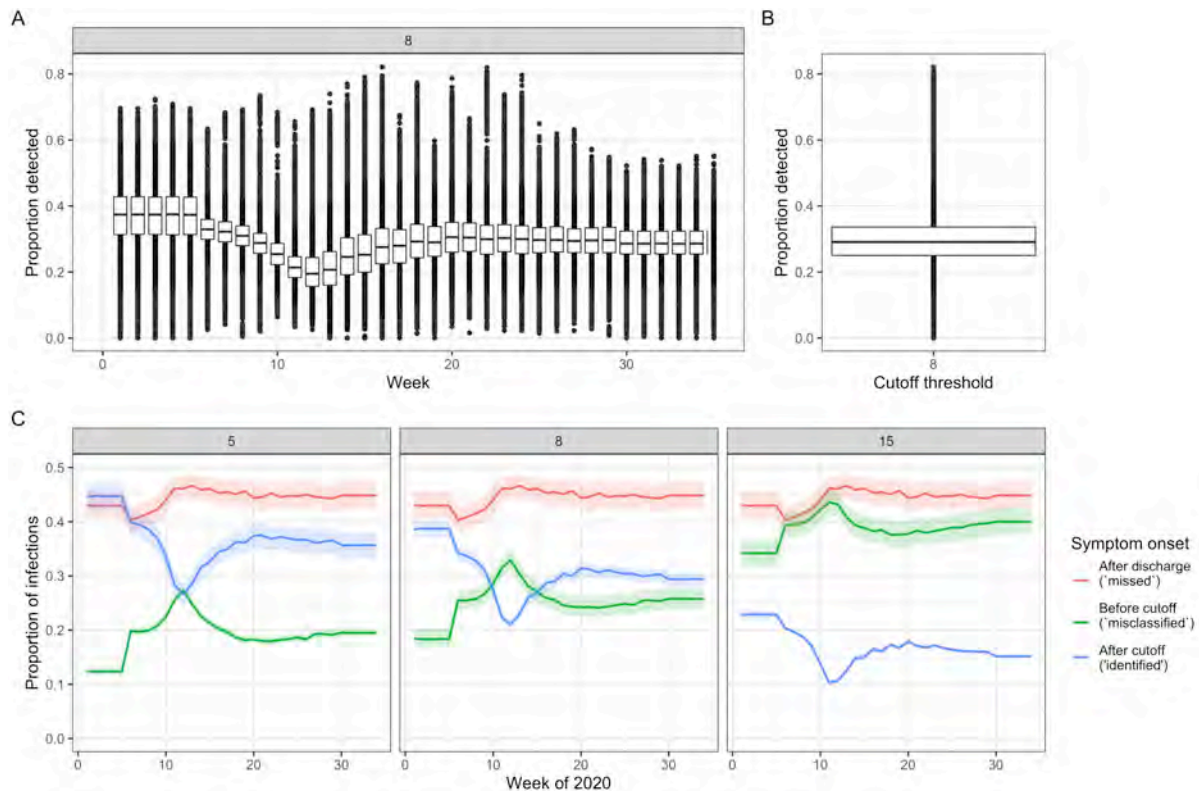
Figure 3: Proportion of symptomatic hospital-acquired infections identified, (A) given by week and (B) over all weeks using a 7 day cut-off, for all acute English Trusts. Each datapoint is the value from a single Trust for each of 200 samples. The boxplot highlights the median and 25th-75th quantile. (C) For England (the aggregate setting) the proportion of patients with hospital acquired infections split by those that are identified (blue) due to a symptom onset starting at a set number of days from admission (grey box) and before discharge, and those unidentified with symptom onset after discharge ("missed", red) or before the cut-off ("misclassified", green). The coloured lines represent the mean, and the shaded areas the 95% percentiles over the 200 samples.

community-acquired" should be classified as "community-onset, hospital-acquired" when a 7 day cut-off is used.

We found that there could have been 47,400 (mean; 95% range over 600 simulations: 45,000, 50,000 for the time varying R value) infections of individuals in the community secondary to patients with unidentified infections acquired in the hospital which had symptom onset after discharge ("missed") over the first wave. We estimated that these would result in 1,600 (1,600, 1,700) "community-onset, hospital-linked" cases with a 7 day cut-off. The values are reduced by one-third with a constant reproduction number of $R = 0.8$ (Supplementary 11). These contribute 2.3% (2.1%, 2.4%) of "community-onset, community-acquired" cases over the first wave with a 7 day cut-off and under both scenario 1 or 2 (Supplementary 11).

This contribution of community-linked cases to hospital admissions with COVID-19 varied depending on the timing of hospital admission post symptom onset (captured here by Scenarios 1-3, Table 2, Figure 4). The proportion of COVID-19 hospital admissions due to hospital-transmission was greatest when total case numbers first declined (peak in COHL in Figure 4D at approx. 4% in late April).

The number of unidentified hospital-acquired infections and hence reclassification levels increased or decreased under a 14- or 4-day cut-off, respectively (Supplementary 11).

Contribution of hospital settings to cases, infections and onward transmission

To summarise, we estimated that there have been a total of 26,600 (mean, 95% range over 200 simulations: 25,900, 27,700) hospital-acquired SARS-CoV-2 infections in acute English Trusts (E, Figure 5) with a 7 day cut-off prior to August 2020. Of these, a total of 15,900 (15,200, 16,400) infections correspond to patients with COVID-19 that were identified as symptomatic cases in hospitals (B+C, Figure 5): as such 60% of hospital-acquired infections were identified (but a proportion of the identified were misclassified). Over the whole first wave, 15% (14.1%, 15.8%) of cases originally classified as community-acquired were estimated to be hospital-acquired or
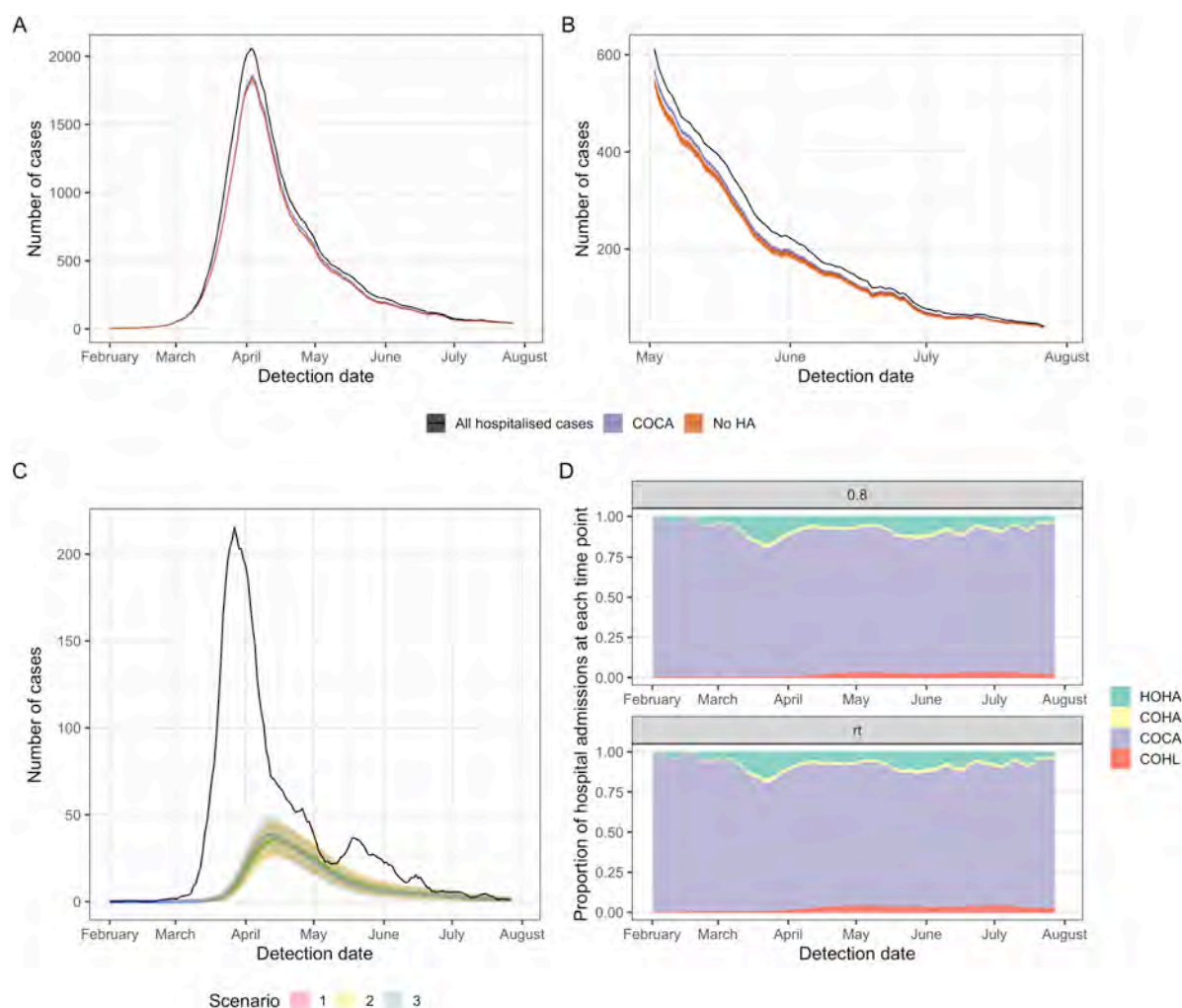
Figure 4: (A) Total COVID-19 admissions with model-adjusted definitions from "community-onset, community-acquired" (COCA) for Scenario 1 for the whole study period (January - 31st July 2020) and (B) for the end of the study period (May - 31st July 2020). The counterfactual of no hospital transmission ("No HA", orange) is compared to the adjusted model estimate of COCA (purple) and total admissions (black) for a time-varying R estimate. (C) The number of hospital-onset, hospital-acquired (HOHA) cases (black) is similar in magnitude to the number of community-onset hospital-linked (coloured lines, COHL) under the three scenarios for hospital admission after symptom onset. (D) The proportion of all hospital admissions in England that were estimated to be HOHA (green), community-onset, hospital-acquired (COHA, yellow), COCA (purple) and COHL (red) under two example R values (constant: 0.8 and time-varying R_t) and Scenario 1. All outputs take a threshold cut-off value for defining hospital-acquired as a symptom onset more than 7 days from admission. All outputs are the rolling 7-day mean for the mean over 200 simulations with 5-95% ranges in shaded areas in (C).

hospital-linked ((C + F) / (A - B), Figure 5).

The estimated percentage of identified COVID-19 cases in hospitals that were hospital-acquired is then 20.1% (19.2%, 20.7%) ((B + C)/ A, Figure 5). Accounting for onward transmission from unidentified "missed" hospital-acquired infections, we estimated that 22.1% (21.2%, 22.9%) of hospitalized COVID-19 cases were hospital-acquired or hospital-linked ((B + C + F)/A, Figure 5) using the median time-varying R value.

If 20.1% of COVID-19 cases identified in hospitals were hospital-acquired then, assuming that 3% of symptomatic cases were hospitalized, we estimated that hospital-acquired infections likely contributed to fewer than 1% of infections of the overall English epidemic of COVID-19 in wave 1.

Assuming similar levels of hospital transmission in non-acute English trusts suggests approximately 31,100 (30,300, 32,400) symptomatic infections could have been caused in total by hospital-acquired transmission in England.

Trust level and Sensitivity analysis

When aggregated, the results from the individual Trust level predicted a slightly higher proportion of cases to be hospital-acquired (25% vs 20%) (Supplementary 12). Varying the day of discharge of the unidentified "missed" infections had little impact on total case numbers, but did affect hospital-linked cases (Supplementary 11).

Discussion

We estimated that before 31st July 2020 20.1% (19·2%, 20·7%) of identified COVID-19 cases in hospitals were likely to have been hospital-acquired infections and that within-hospital transmission likely contributed directly to 26,600 (mean, 95% range over 200 simulations: 25,900, 27,700) symptomatic infections, and a further 47,400 (45,000, 50,000) hospital-linked infections. These results are based on a 7 day cut-off for symptom onset from admission and prior to discharge for defining an identified hospital-acquired case.
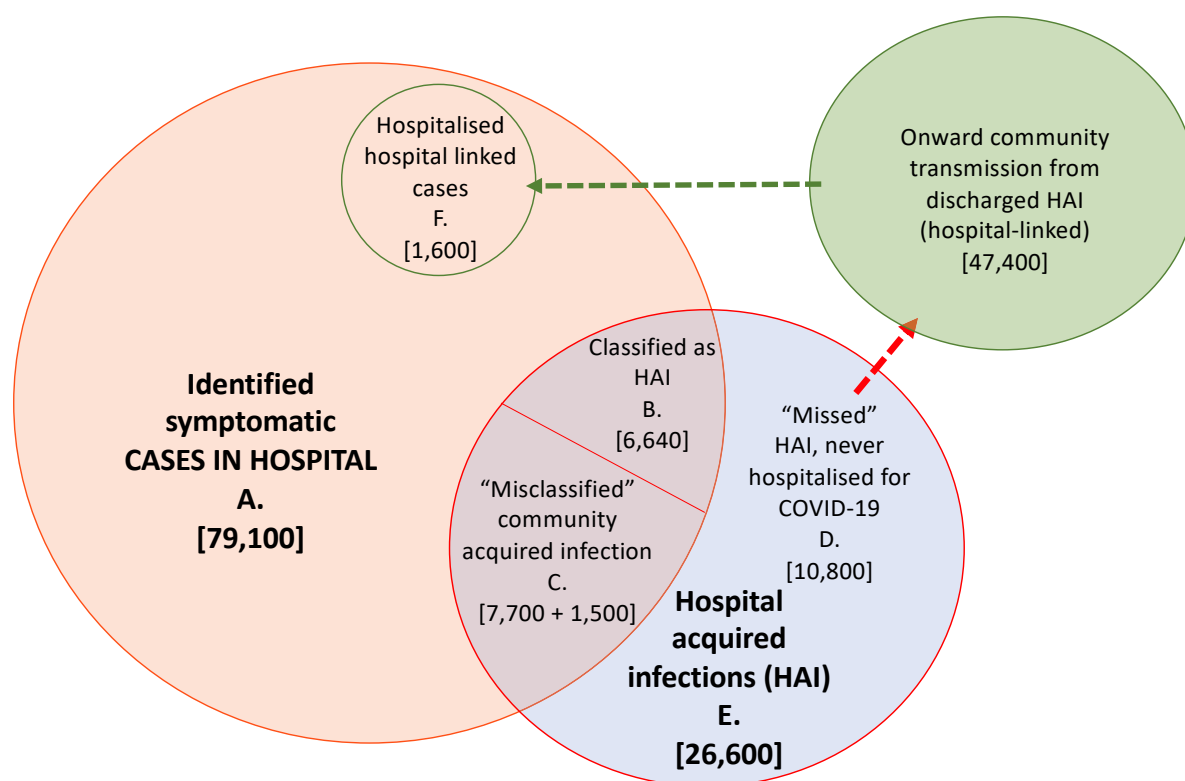
Figure 5: Summary figure of estimated values for patients with hospital-acquired symptomatic infections and onward community transmission with a 7 day cut-off for symptom onset after admission and prior to discharge for defining a patient with hospital-acquired infection. Note here that the "misclassified" (C) includes those "missed" unidentified infections that return to hospital later as a hospitalized COVID-19 case (1,500 "community-onset, hospital-acquired" cases).

Despite these levels of infection, we estimated hospital transmission to patients caused fewer than 1% of all infections in England in the first wave (prior to 31st July 2020). To some extent this reflects effective infection prevention within hospital settings with over 4 million non-COVID-19 patients being cared for in hospital settings during this period. However, the high proportion of hospital cases that were due to hospital-acquired infections is worrying as these are the most vulnerable members of our society and hence may have the most severe consequences. In addition, we did not account for the substantial proportion of asymptomatic infections in our analysis and thus, the impact of hospital transmission on the community epidemic is likely an underestimate [13].

This is the first study to estimate the total number of symptomatic hospital-acquired infections (not just the percentage of known cases that are hospital-acquired) and their

wider contribution to community transmission prior to 31st July 2020. In particular, we found that the contribution of symptomatic hospital-acquired infections to the epidemic likely varied over time, increasing in importance as community infections initially dropped, emphasizing the need to determine where most infections are occurring at any time during an epidemic.

Our results show that relying on symptom onset as a detection method for hospital-acquired SARS-CoV-2 may miss a substantial proportion (> 60%) of hospital-acquired infections even when asymptomatic infections are not accounted for. This depends on the length of stay for non-COVID admissions but suggests that in many settings estimates of the number of infections due to transmission in hospital settings will be substantial underestimates. For example, Read et al (2021) [26] acknowledged that the estimated proportion of nosocomial infections during the first epidemic wave of COVID-19 in the UK that was based on symptom onset data, is likely to be higher if accounted for unidentified cases. This is particularly relevant for low-resource settings with short lengths of stay for non-COVID patients and that rely on symptom onset screening for SARS-CoV-2 infection.

An alternative detection method is routine testing of patients, which will confirm symptomatic as well as detect pre-symptomatic and asymptomatic SARS-CoV-2 infections. However, even with screening on admission, independent of symptoms, and retesting three days after admission, a proportion of infections will likely not be detected due to short lengths of stay. Our estimates of the proportion of hospital cases that are due to hospital-acquired infection are higher than those from an England wide study [8] and those from single hospital settings in the UK [3, 9, 27–29], as we estimated all symptomatic hospital-acquired infections whether identified or not during their hospital stay. Our estimates of all infections are similar to previous modelling work using an SEIR model which estimates that nosocomial transmission was responsible for 20% (IQR 14.4, 27.1%) of infections in inpatients [30].

Our work implies that it may be effective to screen patients upon hospital discharge to detect infection, or to quarantine hospital patients on discharge to prevent further transmissions into the community: we estimated this would detect approximately 40%

of hospital-acquired infections that would become symptomatic (that would otherwise be "missed" in Figure 3c). Hence, depending on the test sensitivity by time from infection, up to 70% of symptomatic hospital-acquired infections could be detected. Onward community transmission from these infections may be especially important as community prevalence of SARS-CoV-2 infection decreases.

Currently, much more routine screening and testing is implemented in English hospitals contributing to the detection of infections prior to symptom onset or discharge [31]. However, screening will need to be conducted with high frequency to avoid missing those infected prior to discharge, or to screen on discharge. Our work is directly linked to the situation prior to August 2020 where little routine testing was in place and our estimates would be affected substantially by the new pandemic situation with new variants and vaccination. However, our conclusion that symptomatic screening has limited efficacy in detecting nosocomial transmission is still highly relevant to support the need for ongoing regular screening of non-symptomatic hospital patients and to emphasize potential missing infections.

Further work is needed to determine the precise risk of returning as a hospital case for those infected in hospitals. If our values (10-15%) are found to be conservative, then this percentage could increase substantially. If it were found to be higher, reflecting the poorer health of hospitalized patients and hence potentially increased susceptibility, then the proportion of hospital cases that are hospital-acquired could increase to 30-40%.

The interpretation of our results is limited by several simplifications. Firstly, we did not explicitly capture disease and hospital attendance variation by age. Future work could stratify our estimates to account for an older and more vulnerable hospital population. Secondly, we likely underestimated the total number of hospital-acquired infections as we modelled only those that progress to symptoms since these are the ones contributing directly to hospital burden. This decision was made as our definition of what was a hospital-acquired case was dependent on symptom onset and asymptomatic proportion estimates are highly variable [13]. Thirdly, we assumed a fixed number of four generations for onward transmission in the community, and did not account for infections in healthcare workers, nor in the setting to which hospitalized patients were

discharged to, such as long-term care facilities. The impact of onward transmission from hospital-acquired infections may be underestimated in this work since these settings may have high levels and large heterogeneity in onward transmission or overestimated if four generations is longer than the average chain from recently hospitalized individuals. Fourthly, we assumed that equal levels of infection control policies were in place in all NHS Trusts during this time period as we had no data to inform variation. Moreover, some of the "missed" cases may have been detected by community screening although there was little in place in England during this time (prior to August 2020).

Finally, identification of hospital infection using CO-CIN relied on symptom onset date, which may be unreliably recorded potentially leading to bias in the patient population. While we cannot assess the biases, it is reasonable to expect that symptoms were recorded well in a clinical setting, and frequently (approx. 65,000 patients included). An alternative definition of hospital-acquired infection reliant on the date of first positive swab would have its own limitations: patients could enter with symptoms and not test positive until more than a week into their stay for example [27].

Conclusions

Due to the delay from infection to symptom onset, hospital-acquired transmission of SARS-CoV-2 may be missed under common definitions of a hospital-acquired infection. We estimated that nearly 20% of symptomatic COVID-19 patients in hospitals in England in the first wave acquired their infection in hospital settings. Whilst this is likely to have contributed little to the overall number of infections in England, the vulnerability of the hospital community means that this is an important area for further focus. Increased awareness and testing, especially of patients at discharge, as is now commonly in place in the UK, is needed to prevent hospitals becoming vehicles for SARS-CoV-2 transmission.

Ethics approval and consent to participate

Ethical approval for ISARIC was given by the South Central - Oxford C Research Ethics Committee in England (Ref 13/SC/0149), the Scotland A Research Ethics Committee

(Ref 20/SS/0028) and the WHO Ethics Review Committee (RPC571 and RPC572, 25 April 2013).

Availability of data and materials

All code and links on how to access the data is available at the supporting Github repository at: `https://github.com/gwenknight/hai_first_wave.git`.

Competing interests

The authors declare that they have no competing interests.

Funding

Author contributions

GMK, JMR, BSC and JVR conceived of the project design. JS, SF, YJ, AB, DP, SE, MY, the ISARIC4C Investigators, the CMMID COVID-19 working group and MGS aided in the acquisition, analysis and interpretation of data. TMP and GMK were responsible for development of new software used in this work. CSB, AB and RH aided in the interpretation of results. GMK wrote the first draft and was responsible for the preliminary interpretation of data. TMP and JVR provided substantive revisions of the manuscript and interpretations of the data. All authors read and approved the final manuscript.

References

[1] World Health Organization. *Timeline: WHO's COVID-19 response*. en. 2011. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline.

[2] Chowell, G, Abdirizak, F, Lee, S, et al. Transmission characteristics of MERS and SARS in the healthcare setting: A comparative study. In: *BMC Medicine* 13.1 (Sept. 2015), pp. 1–12. doi: 10.1186/s12916-015-0450-0.

[3] Rickman, HM, Rampling, T, Shaw, K, et al. Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital. In: *Clinical Infectious Diseases* 72.4 (2021), pp. 690–693. doi: 10.1093/cid/ciaa816.

[4] Wang, X, Zhou, Q, He, Y, et al. Nosocomial outbreak of COVID-19 pneumonia in Wuhan, China. In: *European Respiratory Journal* 55.6 (June 2020). doi: 10.1183/13993003.00544-2020.

[5] Zhou, Q, Gao, Y, Wang, X, et al. Nosocomial infections among patients with COVID-19, SARS and MERS: a rapid review and meta-analysis. In: *Annals of Translational Medicine* 8.10 (May 2020), pp. 629–629. doi: 10.21037/atm-20-3324.

[6] Meredith, LW, Hamilton, WL, Warne, B, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. eng. In: *The Lancet Infectious Diseases* 20.11 (2020), pp. 1263–1272. doi: 10.1016/S1473-3099(20)30562-4.

[7] European Centre for Disease Prevention and Control. *Surveillance definitions for COVID-19*. 2021. https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions.

[8] Bhattacharya, A, Collin, SM, Stimson, J, et al. Healthcare-associated COVID-19 in England: a national data linkage study. In: *Journal of Infection* 0.0 (Aug. 2021). doi: 10.1016/j.jinf.2021.08.039.

[9] Taylor, J, Rangaiah, J, Narasimhan, S, et al. Nosocomial COVID-19: experience from a large acute NHS Trust in South-West London. English. In: *Journal of Hospital Infection* 106.3 (Nov. 2020), pp. 621–625. doi: 10.1016/j.jhin.2020.08.018.

[10] Hellewell, J, Russell, TW, Beale, R, et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. In: *BMC Medicine* (2020). doi: 10.1101/2020.11.24.20229948.

[11] Public Health England. *COVID-19: investigation and initial clinical management of possible cases*. 2020. https://www.gov.uk/government/publications/wuhan-novel-coronavirus-initial-investigation-of-possible-cases.

[12] Lauer, SA, Grantz, KH, Bi, Q, et al. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. In: *Annals of Internal Medicine* 172.9 (Mar. 2020), pp. 577–582. doi: 10.7326/M20-0504.

[13] Buitrago-Garcia, D, Egli-Gany, D, Counotte, MJ, et al. Asymptomatic SARS-CoV-2 infections: A living systematic review and meta-

analysis. en. In: *medRxiv* (July 2020), p. 2020.04.25.20079103. doi: `10.1101/2020.04.25.20079103`.

[14] Knight, GM. *Supporting Github repository*. Github, 2021. `https://github.com/gwenknight/hai_first_wave.git`.

[15] Docherty, AB, Harrison, EM, Green, CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. In: *The BMJ* 369 (May 2020). doi: `10.1136/bmj.m1985`.

[16] NHS Providers. *Introduction to NHS Foundation Trusts and Trusts*. 2015. `https://nhsproviders.org/media/1036/introduction_to_nhs_fts_and_trusts_-_nhs_providers_-_may_2015.pdf`.

[17] NHS Digital. *Organisation Data Service*. `https://digital.nhs.uk/services/organisation-data-service`.

[18] NHS Digital. *Secondary Uses Service (SUS)*. 2020. `https://digital.nhs.uk/services/secondary-uses-service-sus`.

[19] ISARIC 4C. *ISARIC 4C consortium*. en. `https://isaric4c.net`.

[20] Public Health England. *COVID-19: epidemiological definitions of outbreaks and clusters in particular settings*. 2020. `https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters`.

[21] Abbott, S, Hellewell, J, Thompson, RN, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. en. In: *Wellcome Open Research* 5 (Dec. 2020), p. 112. doi: `10.12688/wellcomeopenres.16006.2`.

[22] Verity, R, Okell, LC, Dorigatti, I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. English. In: *The Lancet Infectious Diseases* 20.6 (June 2020), pp. 669–677. doi: `10.1016/S1473-3099(20)30243-7`.

[23] NHS Digital. *Hospital Admitted Patient Care Activity 2020-21*. Apr. 2021. `https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2015-16`.

[24] Davies, NG, Kucharski, AJ, Eggo, RM, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. In: *The Lancet Public Health* 5.7 (July 2020), e375–e385. doi: `10.1016/S2468-2667(20)30133-X`.

[25] Boddington, NL, Charlett, A, Elgohari, S, et al. COVID-19 in Great Britain: Epidemiological and clinical characteristics of the first few hundred (FF100) cases: A descriptive case series and case control analysis. en. In: *medRxiv preprint* (May 2020), p. 2020.05.18.20086157. doi: `10.1101/2020.05.18.20086157`.

[26] Read, JM, Green, CA, Harrison, EM, et al. Hospital-acquired SARS-CoV-2 infection in the UK's first COVID-19 pandemic wave. In: *The Lancet* 398.10305 (Sept. 2021), pp. 1037–1038. doi: `10.1016/S0140-6736(21)01786-4`.

[27] Wake, RM, Morgan, M, Choi, J, et al. Reducing nosocomial transmission of COVID-19: Implementation of a COVID-19 triage system. In: *Clinical Medicine, Journal of the Royal College of Physicians of London* 20.5 (2020), E141–E145. doi: `10.7861/CLINMED.2020-0411`.

[28] Jewkes, SV, Zhang, Y, and Nicholl, DJ. Nosocomial spread of COVID-19: Lessons learned from an audit on a stroke/neurology ward in a UK district general hospital. In: *Clinical Medicine, Journal of the Royal College of Physicians*

of London 20.5 (2020), E173–E177. doi: `10.7861/CLINMED.2020-0422`.

[29] Marago, I and Minen, I. Hospital-Acquired COVID-19 Infection – The Magnitude of the Problem. en. In: *SSRN Electronic Journal* (June 2020). doi: `10.2139/ssrn.3622387`.

[30] Evans, S, Agnew, E, Vynnycky, E, et al. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1829 (July 2021), p. 20200268. doi: `10.1098/rstb.2020.0268`.

[31] NHS England. *Coronavirus » Secondary care*. 2020. `https://www.england.nhs.uk/coronavirus/secondary-care/`.

[32] Office for National Statistics. *Coronavirus (COVID-19) - Latest data and analysis on coronavirus (COVID-19) in the UK and its effect on the economy and society*. 2020. `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare`.

[33] ReAcT study. In: (2020). `https://www.reactstudy.org/`.

[34] Luo, C, Ma, Y, Jiang, P, et al. The construction and visualization of the transmission networks for COVID-19: A potential solution for contact tracing and assessments of epidemics. en. In: *Scientific Reports* 11.1 (Apr. 2021), p. 8605. doi: `10.1038/s41598-021-87802-x`.

[35] Plessis, L du, McCrone, JT, Zarebski, AE, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. en. In: *Science* 371.6530 (Feb. 2021), pp. 708–712. doi: `10.1126/science.abf2946`.

[36] Geoghegan, JL, Ren, X, Storey, M, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. en. In: *Nature Communications* 11.1 (Dec. 2020), p. 6351. doi: `10.1038/s41467-020-20235-8`.

[37] Knock, E, Whittles, L, Lees, J, et al. *Report 41 - The 2020 SARS-CoV-2 epidemic in England: key epidemiological drivers and impact of interventions*. en-GB. Dec. 2020. `https://doi.org/10.25561/85146`.

[38] McAloon, C, Collins, Á, Hunt, K, et al. Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. eng. In: *BMJ Open* 10.8 (Aug. 2020), e039652. doi: `10.1136/bmjopen-2020-039652`.

[39] Griffin, J, Casey, M, Collins, Á, et al. Rapid review of available evidence on the serial interval and generation time of COVID-19. en. In: *BMJ Open* 10.11 (Nov. 2020), e040263. doi: `10.1136/bmjopen-2020-040263`.

[40] Zinn, S and Würbach, A. A statistical approach to address the problem of heaping in self-reported income data. In: *Journal of Applied Statistics* 43.4 (Mar. 2016), pp. 682–703. doi: `10.1080/02664763.2015.1077372`.

[41] Delignette-Muller, ML and Dutang, C. fitdistrplus: An R package for fitting distributions. en. In: *Journal of Statistical Software* 64.4 (Mar. 2015), pp. 1–34. doi: `10.18637/jss.v064.i04`. `https://www.jstatsoft.org/index.php/jss/article/view/v064i04`.

[42] Rizzi, S, Gampe, J, and Eilers, PH. Efficient Estimation of Smooth Distributions from Coarsely Grouped Data. eng. In: *American Journal of Epidemiology* 182.2 (July 2015), pp. 138–147. doi: `10.1093/aje/kwv020`.

[43] Rizzi, S, Gampe, J, and Eilers, PH. Efficient Estimation of Smooth Distributions from Coarsely Grouped Data. eng. In: *American Journal of Epidemiology* 182.2 (July 2015), pp. 138–147. doi: `10.1093/aje/kwv020`.

[44] Akaike, H. A New Look at the Statistical Model Identification. In: *IEEE Transactions on Automatic Control* 19.6 (Dec. 1974), pp. 716–723. doi: `10.1109/TAC.1974.1100705`.

[45] Linton, NM, Kobayashi, T, Yang, Y, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. In: *Journal of Clinical Medicine* 9.2 (Feb. 2020). doi: `10.3390/jcm9020538`.

[46] Cao, B, Wang, Y, Wen, D, et al. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. In: *New England Journal of Medicine* 382.19 (May 2020), pp. 1787–1799. doi: `10.1056/nejmoa2001282`.

Supplementary material

Table of Contents

Supplementary 1: Definitions

Table S1: Common definitions

Term	Definition	Specifics for this analysis
Case	An individual that has COVID-19 (the disease due to SARS-CoV-2 infection)	
Identified hospital acquired infection	An individual with a SARS-CoV-2 infection that has been identified as hospital-acquired	In this work SARS-CoV-2 infection is detected by a case with symptom onset prior to 5, 8 or 15 days from admission in line with the ECDC definition [15]
unidentified hospital acquired infection	An individual with a SARS-CoV-2 infection that has not been identified as hospital-acquired	Some of these will be misclassified as community-acquired, some will be "missed" as the patient is discharged before symptom onset.
Symptom onset	The self-reported start date of COVID-19 symptoms	Here we mostly use the CO-CIN data which has a symptom onset defined by the ISARIC protocol.
Community-acquired	A patient with an infection with SARS-CoV-2 that is classified as being acquired outside of the hospital in the community setting	Individuals with a symptom onset before the cutoff date, including before admission, are classified as community-acquired in CO-CIN.

Hospital-linked	A patient with an infection that was acquired by transmission in the community from a four-generation chain of transmission originating with an unidentified "missed" hospital-acquired infection	We assume that every hospital-acquired infection that is "missed" is discharged into the community and can cause onward transmission. We calculated the number over approximately one month after discharge (4 x 6.7 days).
Classified	The assignation of "community-acquired" or "hospital-acquired" to the infection within a hospitalized patient with COVID-19	We use this to specify the current classification of a symptomatic infection. Hence a case could be classified as "community-acquired" but actually be "hospital-acquired". We chose to use classified as well as "identified" as some hospital acquired infections would not have been classified whilst some would.
Identified	The detection of hospital-acquired infection	
Detection date	the most recent of (1) date of symptom onset or (2) date of admission if this occurred after symptom onset for a patient with COVID-19, censored at date of discharge	For any "community-onset" case this was their admission date. For "hospital-onset, hospital-acquired" cases this was their date of symptom onset (Table 1).

Supplementary 2: Datasets

Trust and case number differences

For COCIN, we included 123 Trusts and 3 super-Trusts in the final data analysis (see Supplementary 4 for definition of super-Trusts, basically pooled Trusts to account for frequent transfers). SUS covers 589 Trusts in England. 319 of these reported a total of 91,319 COVID-19 cases up to 31st July 2020. 13,415 of these cases were not included in COCIN: suggesting that COCIN has a coverage of 85% of the total.

CO-CIN data inclusion

Using the 3rd December CO-CIN data extraction, there were 104,672 unique subject IDs. Of these 78% had a symptom onset and admission date. 62%, or 65,028/104,672 unique subject IDs were included in the final dataset. The included cases were those with (i) a symptom onset date, (ii) an admission date, (iii) a symptom onset date after the 12st January 2020 and (iv) a symptom onset date before the 31st July 2020. Most patients had a symptom onset before admission (Figure S1).
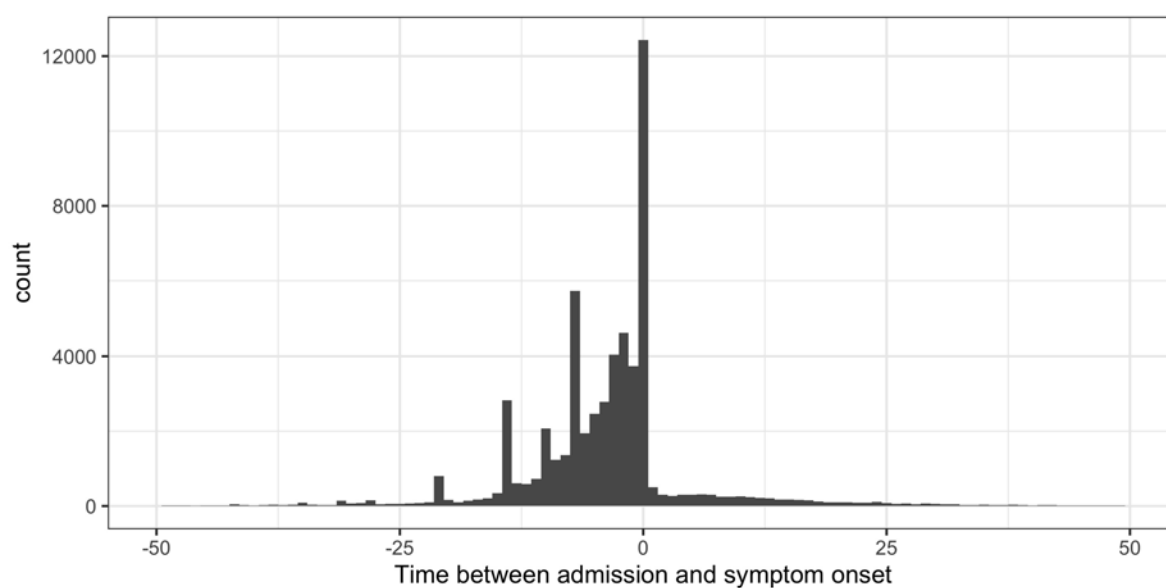
Figure S1: Data from CO-CIN on time between admission to hospital and symptom onset.

We defined a date of "detection" as the most recent of (1) date of symptom onset or (2) date of admission if this occurred after symptom onset for a patient with COVID-19, censored at date of discharge. For any "community-onset" case this was their admission date. For "hospital-onset, hospital-acquired" cases this was their date of symptom onset (Table 1).

LoS distributions

The length of stay (LoS) for non-COVID-19 positive patients is shown by week (in Figure S2) and over time (in Figure S3).
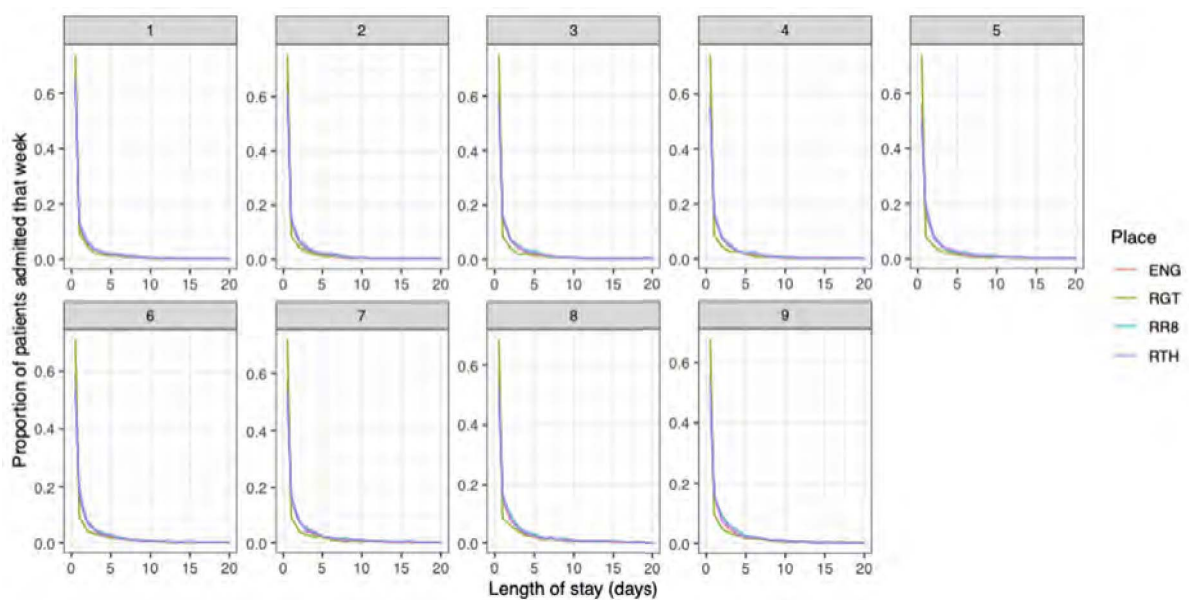
Figure S2: Length of stay variation by week (facet) and example Trusts (colour).
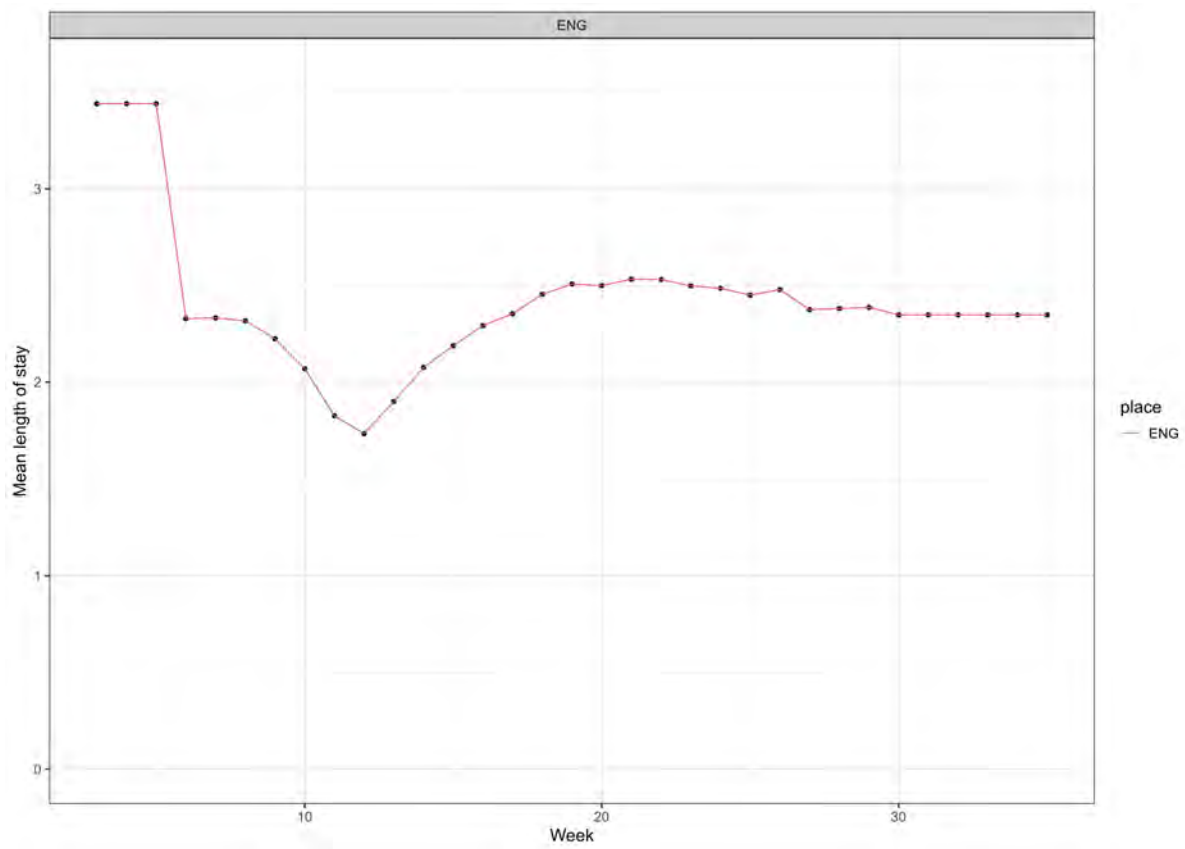
Figure S3: Average length of stay over time for England in top panel to compare to proportion identified for England (equivalent to Figure 3A from main text, bottom panel)).

Supplementary 3: Admission with infection levels

What proportion of hospitalized patients with symptom onset after the cut-off day *T* had been infected in the community and admitted to hospital for a non-covid reason while latently infected?

Data

The maximum prevalence of infection from seroprevalence surveys in the UK prior to September 2020 has been approximately:

0.5% from ONS (modelled, smoothed) [32]
0.3% from REACT1 [33]

Between the 27th April 10th May, ONS estimated prevalence of infection to be: 0.27 (0.17-0.41)%.

Model

The percentage of people at day *T* with COVID that acquired it in the community = Prevalence of infection at entry $\times$ probability still in hospital at day T $\times$ probability symptoms developed after day T =
(prev $\times$ (1-pexp(T,1/los)) $\times$ (1-plnorm(T,1.621, 0.418))$\times$100.

Baseline measures
For example, using the ONS data for early May:
0.0027 $\times$ (1-pexp(T,1/los)) $\times$ (1-plnorm(T,1.621, 0.418))$\times$100

For $T > 10$ this is zero due to very few patients remaining in hospital past this point (even assuming los for non-COVID of 7 days, which is an overestimate).

For $T = 5$, the value is 0.03 (0.02,0.04)%, 0.05 (0.03,0.08)% 0.07 (0.04,0.1)% for mean length of stays of 3, 5 or 7 days respectively. In conclusion < 0.1% of cases past

day 5 are likely to be acquired in the community currently.

At the maximum prevalence: 0.0054 × (1-pexp(T,1/los)) × (1-plnorm(T,1.621, 0.418))*100

For $T > 10$ this is zero due to very few patients remaining in hospital past this point (even assuming los for non-COVID of 7 days, which is an overestimate).

For $T = 5$, the value is 0.05 (0.04,0.07)%, 0.1 (0.08,0.13)% 0.14 (0.1,0.17)% for mean length of stays of 3, 5 or 7 days respectively. In conclusion $< 0.2\%$ of cases past day 5 are likely to be acquired in the community currently.

Conclusion: The prevalence was likely to be higher at the peak of the epidemic, but even at 10x higher this would be less than 1% of cases past day 5 being attributable to non-recent hospital transmission.

Supplementary 4: Comparing COCIN and SUS by week

There are several discrepancies between the Trusts enrolled in COCIN and SUS. The steps to calculate how to go from non-complete enrolment in CO-CIN to SUS (national COVID-19 case total data) are given below.

For each Trust in CO-CIN and each week (aggregated using lubridate::week (Grolemund and Wickham 2011)), the proportion of CO-CIN cases in SUS was calculated.

When the proportion of SUS in CO-CIN was less than 1 (expected as CO-CIN enrolment based)
The algorithm for a single Trust or England, for a set cutoff was

1. Calculate the weekly proportion of CO-CIN cases in SUS

2. Inverse this weekly proportion to give a multiplier

3. In the cleaned (removed those with no subject onset or admission date), one row per subject CO-CIN, enter the multiplier for the week of the admission date for each subject

4. Multiply each single hospital-acquired defined case by the multiplier for their week of admission to inflate the hospital-acquired case numbers. These were rounded to the nearest number.

5. Aggregate over individual case data to get total number of

   (a) hospital-acquired cases (by summing over the inflated case numbers at the individual level)

   (b) Total cases (by summing over the multipliers: each single entry needs inflating)

Code in: trust_number_noso_all.R in `https://github.com/gwenknight/hai_first_wave.git` [14].
When the proportion of SUS in CO-CIN was greater than 1 (unexpected as SUS should have all cases)
If this proportion was greater than 1 (i.e. unexpected more cases in CO-CIN than SUS),

then we explored the actual numerical difference in case numbers that was seen. If this difference in numbers was greater than 20% of the original total numbers in CO-CIN then we explored the difference further: 20 Trusts. The idea here is that especially in May / June there is a small number of cases admitted per week (< 5). It may be that a proportion >1 is then 2 in CO-CIN but only 1 in SUS. If their relative difference is not so big (< 20%) of the original CO-CIN data then we ignore this issue and set the proportion to 1.

For those to be explored further, we looked at the impact of capping the proportion at 1 and multiplying through the CO-CIN data to match the SUS data. If the total number of cases was greater than 150% of SUS then explored these further: this was the case for 5 Trusts.

In closer investigation we found that several of these Trusts had frequent transfers with other Trusts, for example three Trusts in one county, meaning that cases may be differently labelled as being in one Trust or the other in COCIN and SUS. This may be as SUS is based on test date and COCIN on symptom onset which may occur for a patient in different Trusts. To tackle this we aggregated Trusts with frequent transfers into super-Trusts. This results in three super-Trusts (R13, RR0, ESX) which included 2 (RT3, R1K), 2 (RRF, 02H), or 3 (RDD, RQ8, RAJ) Trusts and covered four of these problem Trusts. The fifth Trust (RBA) we removed from analysis as the discrepancy was substantial: more than 20 cases in COCIN than SUS at the peak and a secondary SUS peak that was not present in CO-CIN.

The resulting proportion of CO-CIN cases in SUS over time is shown in Figure S4.
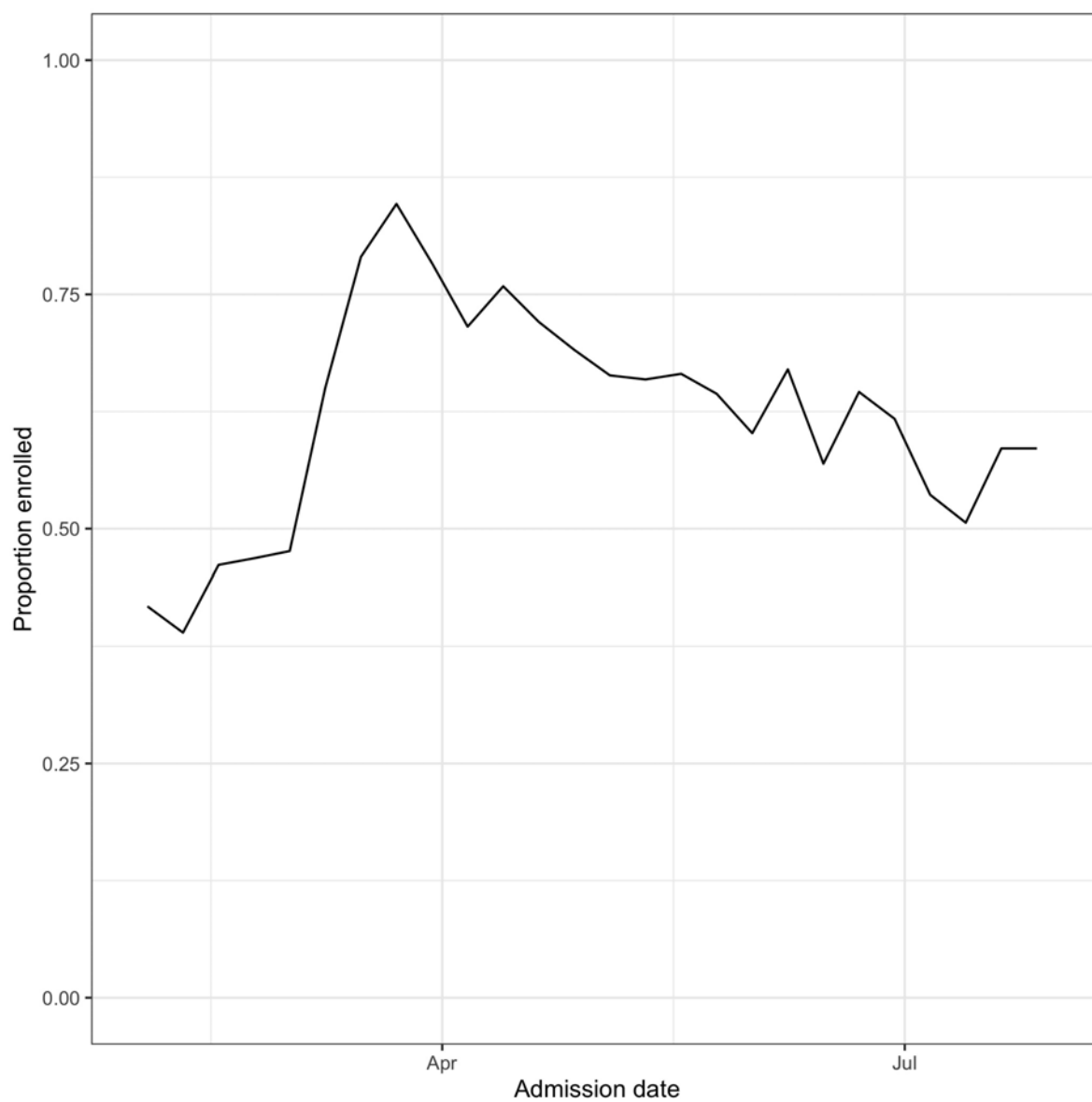
Figure S4: Proportion of CO-CIN cases in SUS over time for acute English Trusts

Supplementary 5: Calculations for the proportion of undetected hospital-acquired SARS-CoV-2 infections

Hospital-acquired infections are here defined as patients who have symptom onset after a certain cut-off value X after hospital admission. In particular, if T_{inf} is the time of infection, T_{inc} is the time from infection till symptom onset and X is number of days after hospital admission of a hospitalized patient, then the patient is classified/detected as a nosocomial case if $T_{\text{inf}} + T_{\text{inc}} \geq X$. Only a subset of all hospital-acquired infections will be detected by this method. We estimated the proportion of hospital-acquired cases that get detected in the hospital based on information available from the CO-CIN and SUS data set. From that we could deduce the proportion of hospital-acquired infections that would be *missed* by this method. We assumed that the cut-off value X is chosen large enough such that community-acquired cases can be excluded.

We implemented R functions for the calculations of the proportions of missed hospital-acquired infections based on the theoretical calculations below. The full code is available from: `https://github.com/tm-pham/covid-19_nosocomialdetection`.

CO-CIN Analysis

CO-CIN includes information on date of symptom onset of hospitalized patients. Let LoS be the random variable representing the length of stay of hospitalized (non-COVID-19) patients and estimated from empirical data from SUS. Three types of hospital-acquired cases can be distinguished:

1. Patients with symptom onset before the cut-off X days after admission, i.e. $\{T_{\text{inf}} + T_{\text{inc}} < X\}$

2. Patients with a symptom onset after discharge, i.e. $\{T_{\text{inf}} + T_{\text{inc}} > \text{LoS}\}$

3. Patients with a symptom onset after X days after admission but before discharge, and with a length of stay of at least X days, i.e. $\{T_{\text{inf}} + T_{\text{inc}} \geq X\} \cap \{\text{LoS} \geq T_{\text{inf}} + T_{\text{inc}}\}$

Only the last category of hospitsal-acquired cases will be detected by the method described above. On a given day, the probability that a hospital-acquired case is detected

(using a cut-off of X days) is given by

P(randomly selected patient is detected on a given day | patient is a nosocomial case)

$$= P(\text{randomly selected patient fulfills 3.}) \tag{5.1}$$

$$= P(X \leq T_{\text{inf}} + T_{\text{inc}} \leq LoS) \tag{5.2}$$

$$= \sum_{l=X}^{\infty} p_l P(X \leq T_{\text{inf}} + T_{\text{inc}} \leq l) \tag{5.3}$$

$$= \sum_{l=X}^{\infty} \sum_{t=1}^{l-1} p_l P(X \leq t + T_{\text{inc}} \leq l) \cdot P(T_{\text{inf}} = t) \tag{5.4}$$

$$= \sum_{l=X}^{\infty} \sum_{t=1}^{l-1} p_l P(X \leq t + T_{\text{inc}} \leq l) \cdot \frac{1}{l} \tag{5.5}$$

$$= \sum_{l=X}^{\infty} \sum_{t=1}^{l-1} \frac{p_l}{l} P(X - t \leq T_{\text{inc}} \leq l - t) \tag{5.6}$$

We adjusted for the fact that over a given period of time, patients with longer length of stays are more likely to be encountered and to be infected in the hospital than patients with short length of stays. Hence, the probability that on a given day, a randomly selected hospital-acquired case has LoS $= l$ is given by

$$p_l = \frac{P(\text{LoS} = l) \cdot l}{\sum_{l=1}^{\infty} l \cdot P(\text{LoS} = l)}$$

We, further, assumed a constant force of infection on each day, i.e., a non-COVID-19 patient is equally likely to get infected on each day and therefore $P(T_{\text{inf}} = t) = \frac{1}{l}$. This assumption was verified using data from Oxford (see below).

Probability of infection per day

In the above calculation, we assumed that a non-COVID-19 patient is equally likely to get infected on each day and therefore $P(T_{\text{inf}} = t) = \frac{1}{l}$. This assumption was based on hospital data from Oxfordshire. We fitted a generalised additive model with the probability of being tested positive for SARS-CoV-2 dependent on the day of hospitalisation accounting for age, gender, ward type, and ethnicity, using a logit link.
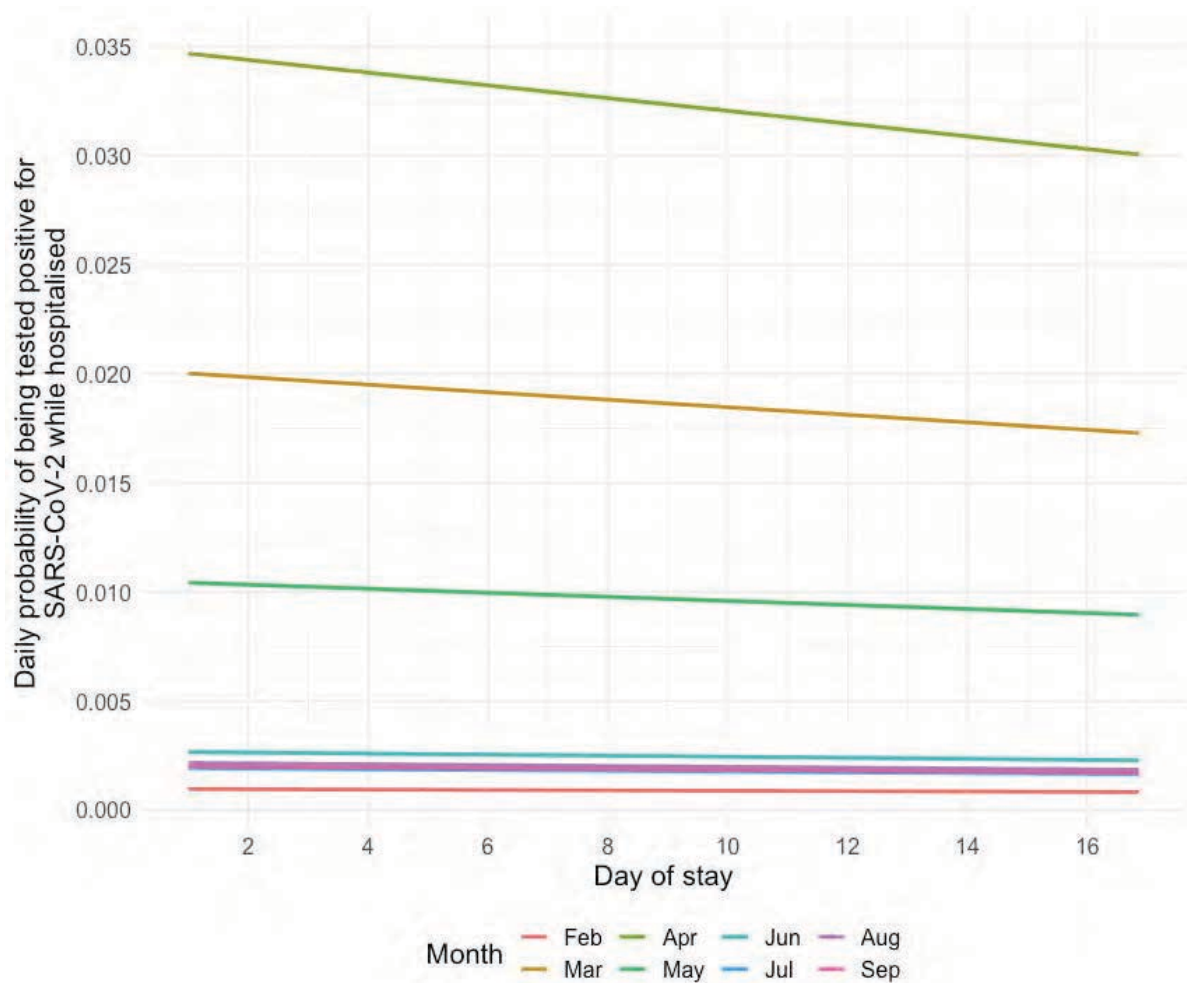
Figure S5: Daily probability of being tested positive for SARS-CoV-2 while hospitalized based on generalized additive model fitted to hospital data from Oxfordshire.

Supplementary 6: Parameterization and additional methods

Serial interval:

Latency period: mean of 5.1 days

Infectious period: mean of 3.4 days

Subsequent infection"   mean of 5.1 + uniform(0,1) * 3.4 = mean of 6.8 days

For each infection, a latency period, infectious period and uniform random number were sampled. An "R" number of subsequent infections were then generated at a time latency period plus the uniform random number times the infectious period.

We chose to look at approximately the first month of transmission after discharge to limit the number of onward cases. It is likely that chains of transmission are short: 4 generations in China [34], and suggested to be short from genomic data in the UK and New Zealand [35, 36].

Additional methods

d. Reclassifying community-acquired as hospital-acquired

To determine the contribution of unidentified hospital-acquired infections to hospitalized patient burden, we estimated when an unidentified "missed" hospital-acquired infection would return as a hospital admission by generating the entire disease progression trajectory for each unidentified "missed" hospital-acquired infection (Figure 2).

For the disease progression trajectory, the proportion returning to hospital was sampled using a Bernouilli trial and varied for each simulation (Table 2). For each individual that was expected to become a hospitalized case we sampled a time (i) from infection until discharge (ii) from infection to symptoms and (iii) from symptoms to potential hospitalisation (Figure 2, Table 2). The time since infection was subtracted from the time to hospitalisation (the sum of time to symptoms from infection and time from symptoms to hospitalisation) to calculate the time at which the unidentified "missed"

hospital-acquired infected individuals would be identified but currently misclassified as a "community" case at hospital admission (new "community onset, hospital-acquired" cases, Figure 2, Table 1).

e. Hospital-linked cases

To account for onward transmission in the community from patients with unidentified "missed" hospital-acquired infections (due to symptom onset after discharge) we estimated "hospital-linked infections": calculated as first-, second-, third- and fourth-generation infections. This is approximately the number of infections caused within one month after discharge (6.7 day serial interval, Supplementary 6) and assumes that most onward transmission chains are relatively short [34–36].

The time series for these was calculated by sampling a certain time to infection (a sum of a sample from the latency distribution and a sample from a uniform distribution on 0-1 multiplied by a sample from the distribution for the duration of clinical infectiousness ($\sim$ 3 days)), a number of secondary infections (using estimates for the reproduction number, R), a sampled proportion which progress to disease, a sampled proportion of infections that become hospitalized and a sampled time to hospitalisation (with different distributions for each symptom onset to hospitalisation scenario) (Figure 2, Table 2).

For the onward transmission, we explored three reproduction number values: a constant value of 0.8 or 1.2 with a range generated as +/- 5% of the constant value. For a time-varying estimate "Rt" we took upper/lower bounds for the 50% credible interval from a publicly available repository (39) (Supplementary 9). Mean and 95% ranges for onward transmission infections and case numbers are presented as over the 600 simulations generated from 200 simulations on each R value (estimate, upper and lower bound).

f. Reclassifying community-acquired to hospital-acquired

The number of unadjusted identified hospital-acquired COVID-19 cases is from the inflated CO-CIN dataset ("hospital-onset, hospital-acquired" cases, Figure 2, Table 1). The unadjusted community-acquired classifications were then defined as the difference between the total number of COVID-19 hospital admissions and the unadjusted

identified hospital-acquired COVID-19 cases.

We adjusted the number of hospital-acquired cases by adding our model estimates of (1) "community-onset, hospital-acquired" and (2) any hospital-linked cases, to the identified hospital-acquired case numbers ("adjusted" hospital-acquired assignations). The "adjusted" community-acquired classifications are then altered accordingly. We then calculated the proportion of community cases that were reassigned as (unadjusted community - adjusted community) / (unadjusted community).

To calculate the counterfactual of no transmission in hospital settings, we compared the original total number of hospitalized cases to the adjusted community number (i.e. those that we did not model as being acquired-in or linked-to hospital settings).

Total English burden

Acute Trusts in CO-CIN covered approximately 85% of the COVID-19 cases recorded in SUS. In order to give estimates for all English trusts, we multiplied our results by 1.17 and assumed similar levels of nosocomial transmission in non-acute English trusts.

Table S2: Parameters values used in the model. Extended version of Table 2.

Parameter in R code	Definition	Literature	Notes	Base case
prop_miss_hosp	Proportion of recently hospitalized patients with missed hospital-acquired infections that will be subsequently admitted to hospital with COVID-19	Infection hospitalisation ratio that ranged from < 5% in those aged 40 to > 40% in those aged 80+ [37]; 3,4,8,12,17,18% infections are hospitalized for 10yr age groups from 30 to 80+ respectively (Table 3, [22]); Non-COVID hospital population composed of 33% older than 70, 60% older than 50 (5 yr age group data used) [23]	Multiplying the proportion of the non-COVID hospital population in each age group by the risks in Knock et al [37] leads to an upper estimate of 15%. These patients have previously been hospitalized so have a higher risk of re-infection than others in their same age group. We assumed a uniform distribution between 10% and 15%. For each patient a Bernoulli trial then used this sample to assess whether the patient would return.	unif(0.1,0.15)
prop_comm_hosp	Proportion of community infections that will be hosp. cases of COVID-19	3.5% (95% CrI 3.3%-3.7%) of people infected needed hospitalization [37]	Assume normal distribution with mean from literature, and estimated standard deviation to match range	norm(0.035, 0.0005)
time_inf_to_symp_mean	Time to symptom onset from infection	Incubation period: mean of 5.1 days [12].	Use the Lauer distribution for analysis	1.62
time_inf_to_symp_sd		Log normal distribution, with mean of 5.8 (95% CI 5.0 to 6.7) days [38]		0.4
time_inf_to_symp_mean_sd	Standard deviation in estimates of mean and standard deviation time to symptom onset from infection	Incubation period: mean of 5.1 days [12]	Taken from range from Lauer et al [12]	0.064
time_inf_to_symp_sd_sd				0.0691
time_symp_to_hosp_meanlog	Time to hospitalisation from symptom onset	Gamma distribution, shape parameter equal to the mean of 7 days (standard deviation 2.65) [24]	Scenario (1) : Log normal distribution fitted to CO-CIN data (mean of 7 days, median 6 days) Scenario (2): Gamma distribution as in Davies et al (gamma(7,1)) Scenario (3): Log-normal distribution from FF100 data (log-normal(1.44, 0.72)) (See below)	1.66
time_symp_to_hosp_sdlog		Analysis of "first" wave CO-CIN gives a mean of 7.7 days, median = 6 days and a range of 1-129. Analysis of first few 100 cases in the UK [25]		0.89
	Time from symptom onset to hospitalization	Approximately 2 weeks	Sum of means = 7 + 5.1 = 12.1 days	
R	Average number of secondary infections from one infected individual in the community	Use the time varying estimates of R from epifore-casts.io as well as constants	Constant or time varying estimate	0.8, 1.2 and "rt"
infectious_shape	Time period over which an infected individual is infectious (time from exposure to infection)	Duration of clinical infectiousness: gamma (shape=3.5, scale = 4) [24]	Taken to be as in Davies et al as a balance between underestimating due to lack of pre-clinical period vs. an overestimate if look at serial interval estimates (3 - 6 days [39])	4
infectious_scale		Generation time estimates: 3.95 to 5.20 days [39]		0.875 (3.5/4)
cut-off_date	Days from admission cut off for defining hospital-acquired case	Assumed 5, 8, 10, 14	Will affect time series of hospital-acquired cases	5

Supplementary 7: Symptom onset to hospitalization

As this was a key parameter for our estimates we chose to perform a scenario analysis around this distribution.

Baseline scenario 1: "Best" fit to CO-CIN raw and smoothed data

With data on 38,168 patients from CO-CIN reporting a symptom onset prior to hospitalisation in Wave 1, we could estimate the best fit to the data. However, the data suffered from "heaping" issues where patients preferably reported symptom onset data 1 week, 10 days, a fortnight or 3 weeks before hospital admission (Figure S6). This has been seen for many types of participant reported data (e.g. income [40]). To account for this we fitted to (1) the raw data (Figure S6) below using the fitdistr R package [41] and (2) used a penalized composite link model [42, 43] to adjust for this heaping. We then compared the model fits using the Akaike Information Criterion (AIC) [44].

For both fitting to the raw and smoothed data the distribution with the smallest AIC value was the log-normal distribution (orange line in both Figure S6 and S7): AIC for the gamma distribution (next smallest AIC) was 228080 and 229646 for the smoother or raw data respectively, whilst for the log-normal distribution it was 225675 and 226842.

The values for the log-normal distribution fitted to the raw were:
Meanlog   1.662 (0.005)   SDlog   0.889 (0.003)
And smoothed data:

Meanlog   1.665 (0.005)   SDlog   0.894 (0.003)

We used a lognormal(1.66, 0.89) distribution in the base case Scenario 1.

Scenario 2: previous estimates

We also took a scenario which used a previous estimate of the time from symptom onset to hospitalisation as a gamma distribution with shape 7 and rate 1 [24] (grey line in Figure S6). This was calculated using international data from the first wave [45, 46].

Scenario 3: First Few 100 (FF100) cases in Great Britain

We used data from the first few 100 cases data from Public Health England [25]. This contains information on symptoms from the first 492 cases, 167 of which were hospitalized. At this time there was not a strict list of symptoms as there was later in 2020 (loss of taste / smell, continuous cough, fever). Fitting to this data suggested a best fit of logNormal distribution with mean log = 1.44, SD log = 0.72.
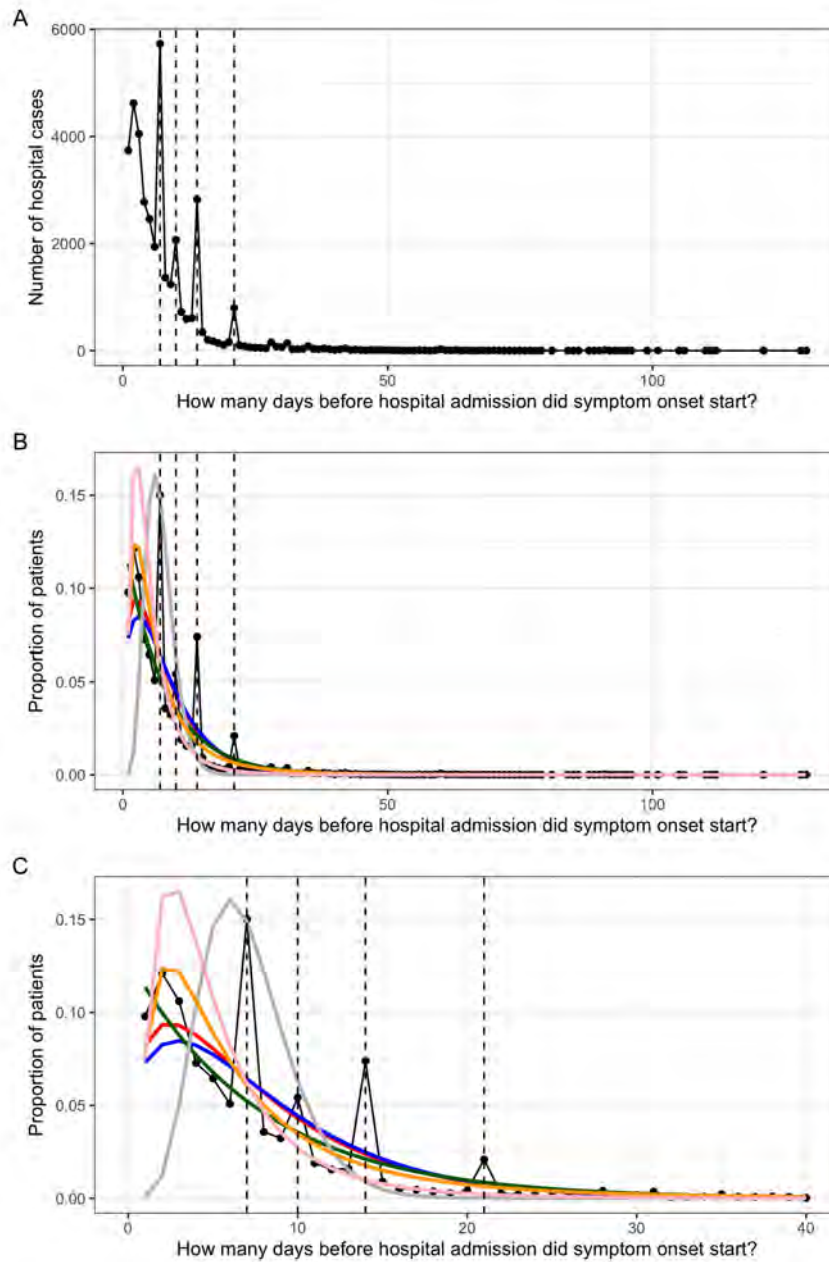
Figure S6: What is the distribution of symptom onset before hospitalisation? (A) CO-CIN data for 38,168 patients from Wave 1 in England with a symptom onset and hospital admission date. Dashed lines indicate heaps in the data at 7, 14, 10 and 21 days prior to admission. (B) Results of probability distribution fitting to the data: red = gamma, blue = negative binomial, dark green = exponential, orange = log-normal (Scenario 1). The grey line is the distribution from [24] (Scenario 2: ≈gamma(7,1)) and the pink line is the distribution from the FF100 data (Scenario 3: lognormal (1.44, 0.72) (C) Zoom in on (B) to show smaller differences in days between symptom onset and admission.
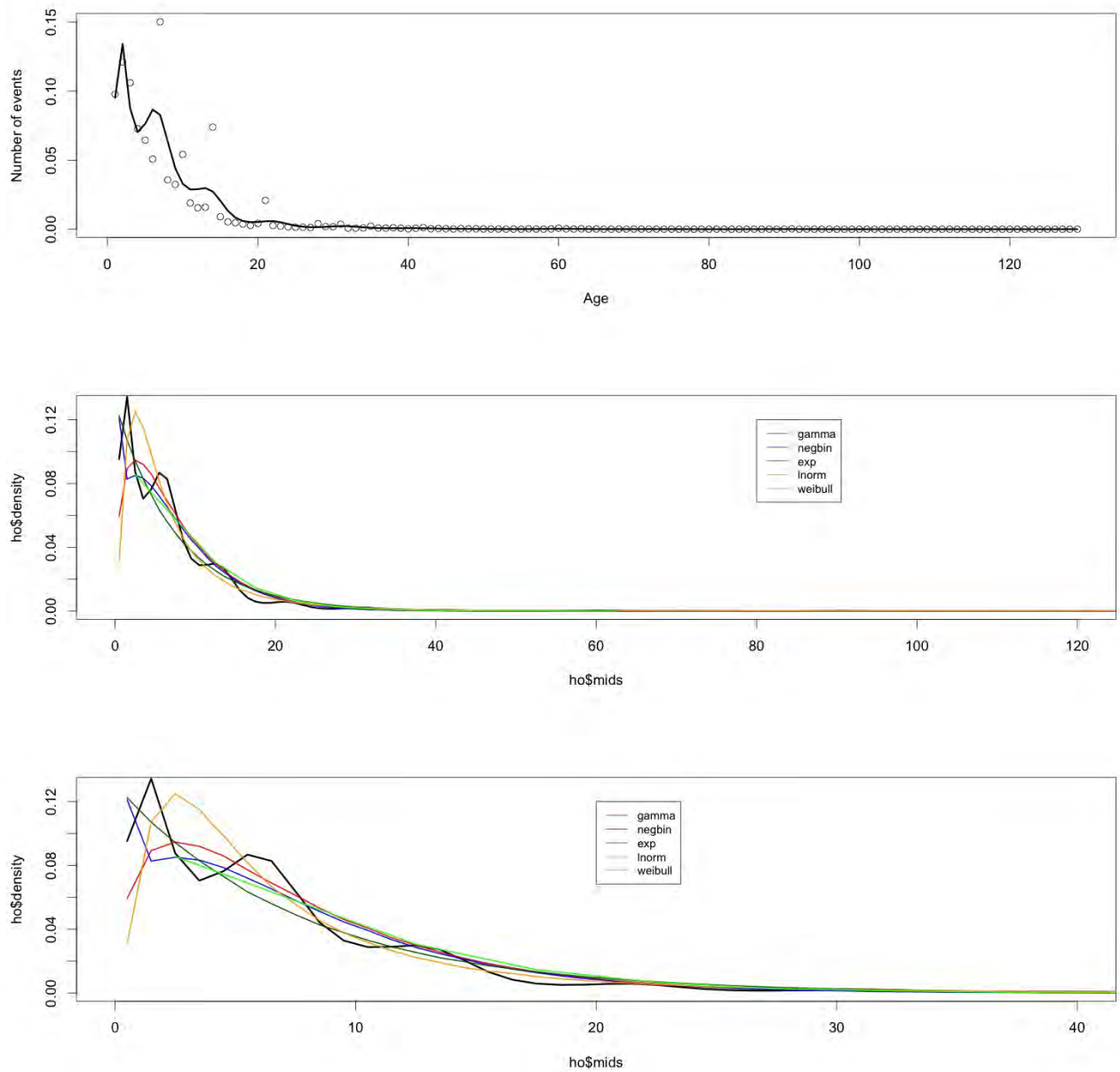
Figure S7: What is the distribution of symptom onset before hospitalisation? (A) CO-CIN data (dots) smoothed using a penalized composite link model to give the black line. (B) Results of probability distribution fitting to the smoothed data (black line) (C) Zoom in on smaller differences between symptom onset and admission.

Supplementary 8: Infection to discharge calculations

Distribution of time from infection until hospital discharge for pre-symptomatic and asymptomatic patients

Let $T_{\text{inf,dis}}$ be the time from infection until discharge for pre-symptomatic and asymptomatic patients. Aim is to determine when "missed infections" will be discharged into the community. Thus, we assume that the time of infection is before discharge of the patient, i.e., $T_{\text{inf}} \leq \text{LoS}$. Given a LoS$= l$, we assume that infection is equally likely to occur on any day of length-of-stay. The distribution of $T_{\text{inf,dis}}$ is given by

$$P(T_{\text{inf,dis}} = t) = P(\text{LoS} - T_{\text{inf}} = t) \tag{5.7}$$

$$= \sum_{l=1}^{\infty} P(\text{LoS} - T_{\text{inf}} = t \mid \text{LoS} = l) p_l \tag{5.8}$$

$$= \sum_{l=1}^{\infty} P(T_{\text{inf}} = l - t \mid \text{LoS} = l) \frac{P(\text{LoS} = l) \cdot l}{\sum_{l=1}^{\infty} l \cdot P(\text{LoS} = l)} \tag{5.9}$$

$$= \sum_{l=1}^{\infty} \frac{1}{l} 1(t \leq l) \frac{P(\text{LoS} = l) \cdot l}{\sum_{l=1}^{\infty} l \cdot P(\text{LoS} = l)} \tag{5.10}$$

$$= \sum_{l=1}^{\infty} 1(t \leq l) \frac{P(\text{LoS} = l)}{\sum_{l=1}^{\infty} l \cdot P(\text{LoS} = l)} \tag{5.11}$$

where p_l is the probability that on a given day, a randomly selected patient has LoS $= l$, i.e.

$$p_l = \frac{P(\text{LoS} = l) \cdot l}{\sum_{l=1}^{\infty} l \cdot P(\text{LoS} = l)}$$

Supplementary 9: Rt estimates
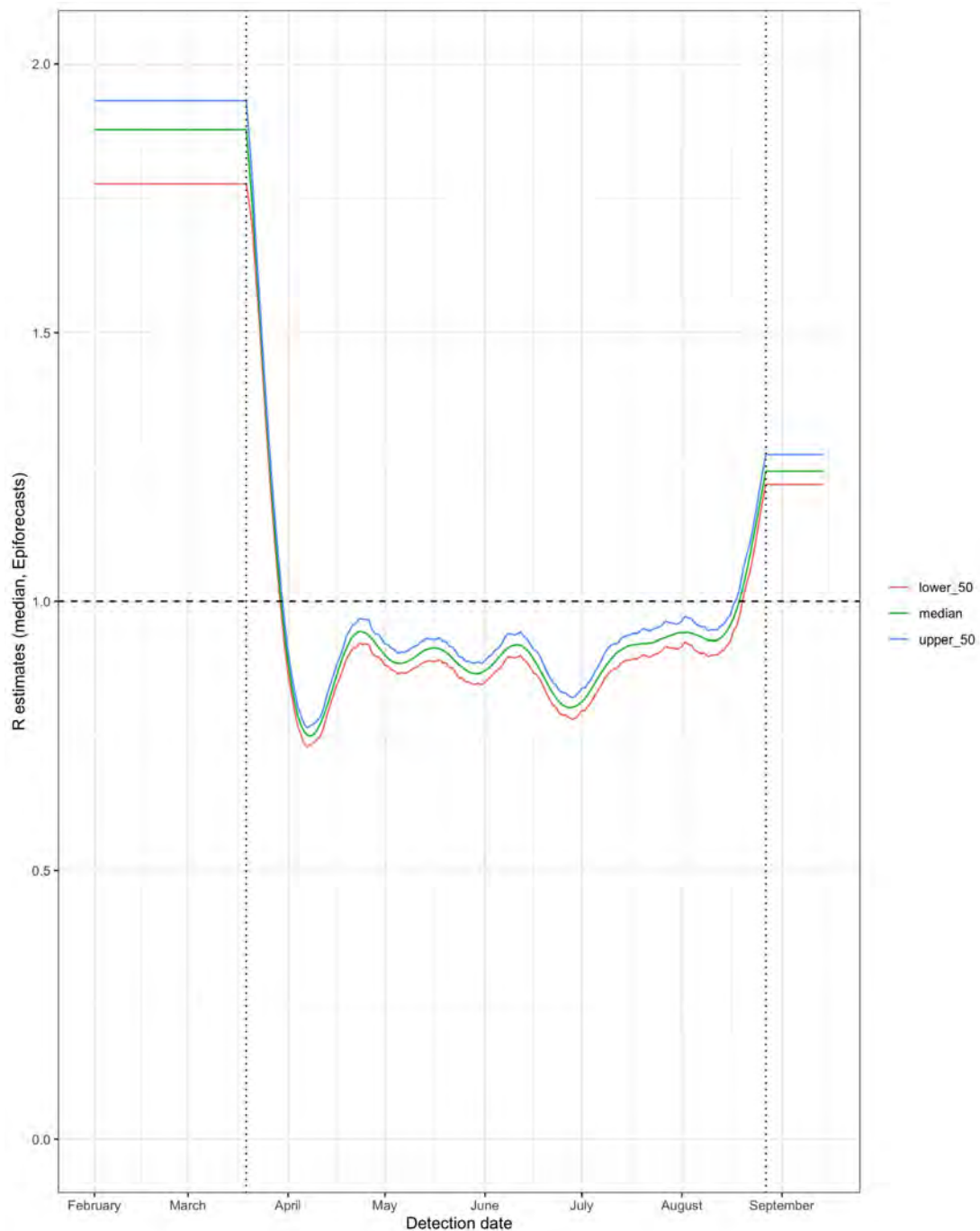
Figure S8: Time varying estimate of Rt taken from EpiForecast team: median estimated using hospitalized cases [34] with upper and lower bounds of the 50% credible intervals.

Supplementary 10: Uncertainty inclusion

200 simulations were generated. Each simulation included uncertainty from three stages:

Stage 1

As we generated estimates of the proportion identified by place and week, we included uncertainty from two elements each week:

(a) Length of stay distribution: bootstrap the distribution for that week from SUS. As there are so many patients (n = 237,981) in the data there is little variation produced by this variation (see Supplementary Figure S8 below, top left).

(b) Incubation period: sampled the parameters for the incubation period distribution (i.e. sample from the mean and standard deviation for the lognormal distribution from a normal distribution with the estimated mean and sd to give a different distribution for each sample for the time to symptom onset from infection, see Table 2).

This incubation period distribution and length of stay for non-COVID patients was used for the entire of the simulation. This is coded in "trustproportion_detect_by_week_all.R" [14]. It gives the variation in the proportion of hospital-acquired infections identified and is presented in Figure 3c, and shown again in Figure S9 for a cutoff of symptom onset more than 7 days from admission.

For example, towards the end of March: 250 hospital-acquired cases were identified in the inflated CO-CIN (Figure S9, bottom). At this stage it is likely that we were identifying between 20% and 22% of hospital-acquired cases (Figure S9, top). Hence this corresponds to between 840 and 1,000 missed cases.

Stage 2

To accounting for binomial sampling variation, the proportion identified for each sample and week (generated above) were used within a Bayesian framework as the binomial probability of identification to infer from the number of identified hospital-acquired

A

B

C
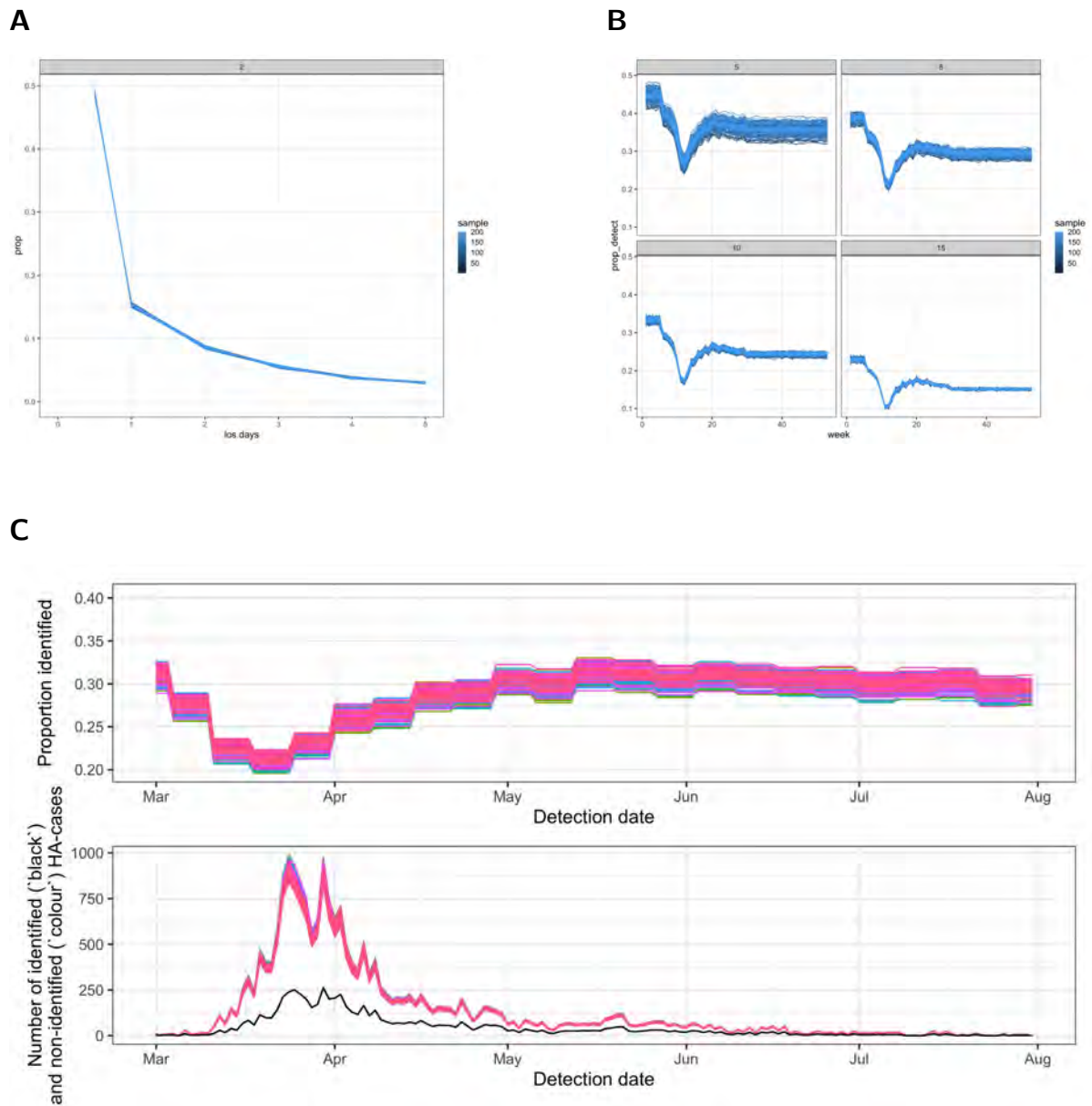
Figure S9: Uncertainty in the length of stay (A) and incubation period drive uncertainty in the proportion identified (B and middle). The inverse of this proportion multiplies the number of identified hospital-acquired cases per week (black, C) to calculate the number of unidentified infections (colour, C).

cases, the total number of hospital-acquired infections ("trials").

In more detail - using the distributions in step 1 within our function we could generate 200 samples of the proportion of true hospital acquired infections that were identified each week, i, and setting, j, from hospital data ($p_{i,j}$). Assuming the number of hospital acquired infections were binomially distributed, we estimated the weekly number of true hospital-acquired infections, $X_{i,j} \sim \text{Bin}(N_{i,j}, \ p_{i,j})$. Subtracting from this the identified weekly hospital-acquired infection numbers we can estimate the number of unidentified hospital-acquired infections.

This used the function in "binom_posterior.R" [14].

Stage 3

The uncertainty in the natural history trajectory for each of these unidentified hospital-acquired infections was then calculated (as shown in Figure 2d) by sampling from the relevant distributions for the probability (e.g. of returning as a hospitalized cases) and timings (e.g. symptom onset after infection).

This is coded in "perc_contribution_function_ trust_week.R" [14].

For each unidentified infection, the probability of returning as a COVID-19 case to hospital is a Bernoulli trial for each missed infection with weekly randomly sampled probability of returning taken from a uniform distribution over 10-15%. This probability of a "missed" unidentified infection returning of a community infection becoming hospitalized is fixed across each of the 200 simulations. Each of the following timings for each returning to hospital as a case unidentified hospital-acquired infection are then sampled from the relevant distributions (Table 2):

  (a) From infection to discharge

  (b) From infection to symptom onset (incubation period)

  (c) From symptom onset to hospitalisation (this is scenario dependent)

We decided to use 200 simulations as above approximately 150 simulations the output for key parameters (shown in Figure S12) stabilized.
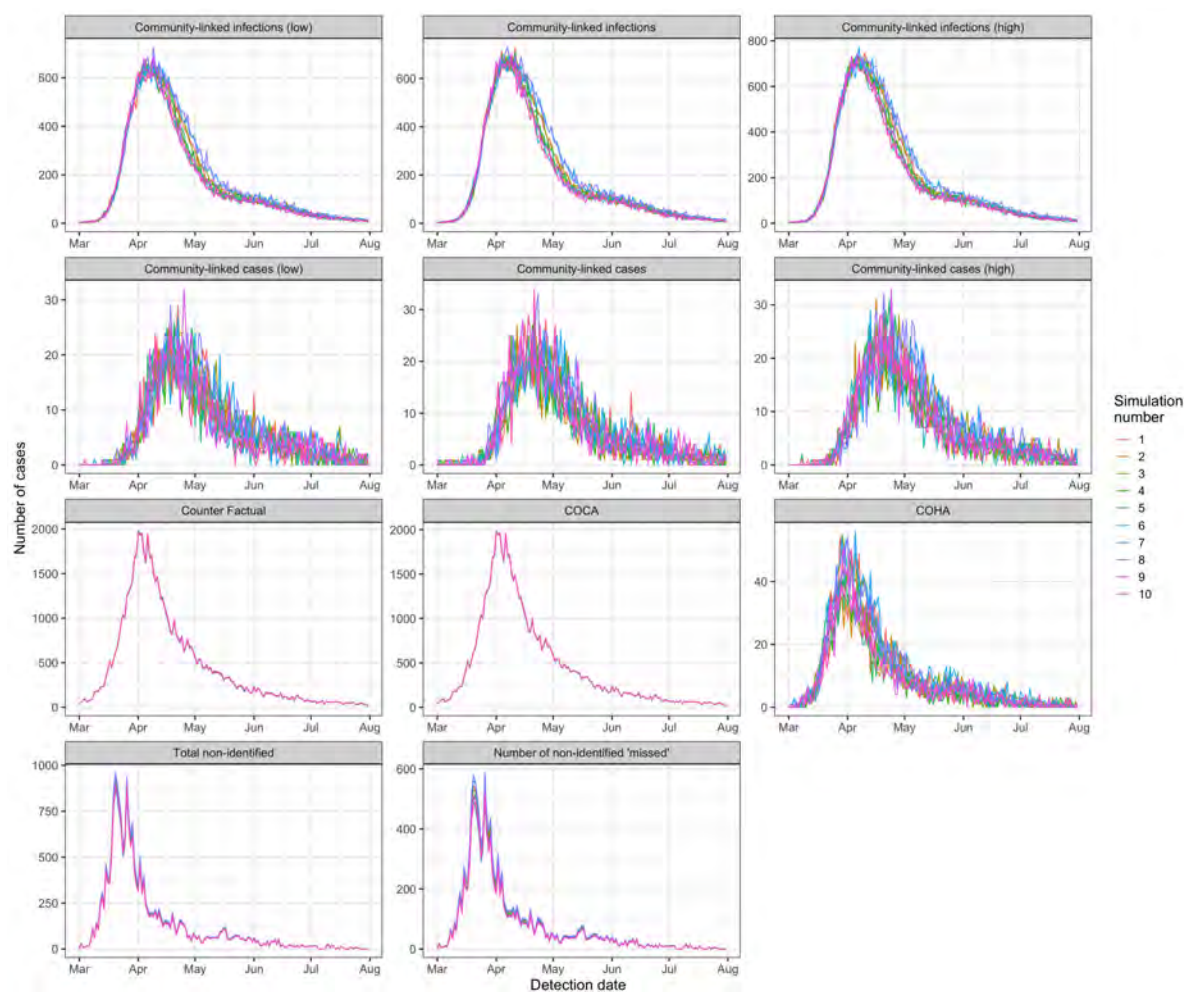
Figure S10: Example time series with a cutoff of at least 7 days from symptom onset to hospitalisation for defining a hospital-acquired case: the first 10 simulations for key model outputs are shown in the 10 colours in the above facets over time (detection date). The top two rows show the variation in community-onset, hospital-linked infections (first row) and subsequent cases (second row) at low, mean and high values of onward transmission (R = 0.76, 0.8, 0.84). The third row shows the counterfactual: the number of hospitalized cases there would be predicted to be without any hospital-acquisition of SARS-CoV-2, alongside the community-onset, community-acquired ("COCA") and community-onset, hospital-acquired ("COHA") case estimates. The final row shows the same variation shown in Figure S8: the total number of unidentified infections and the "missed" subset of these ("missed" due to discharge prior to symptom onset).

Conclusion

Uncertainty in our estimates was generated from sampling from a range of natural history distributions and the length of stay data. As we had data from SUS on the latter for a large number of non-COVID patients, we had little ambiguity in this key parameter for
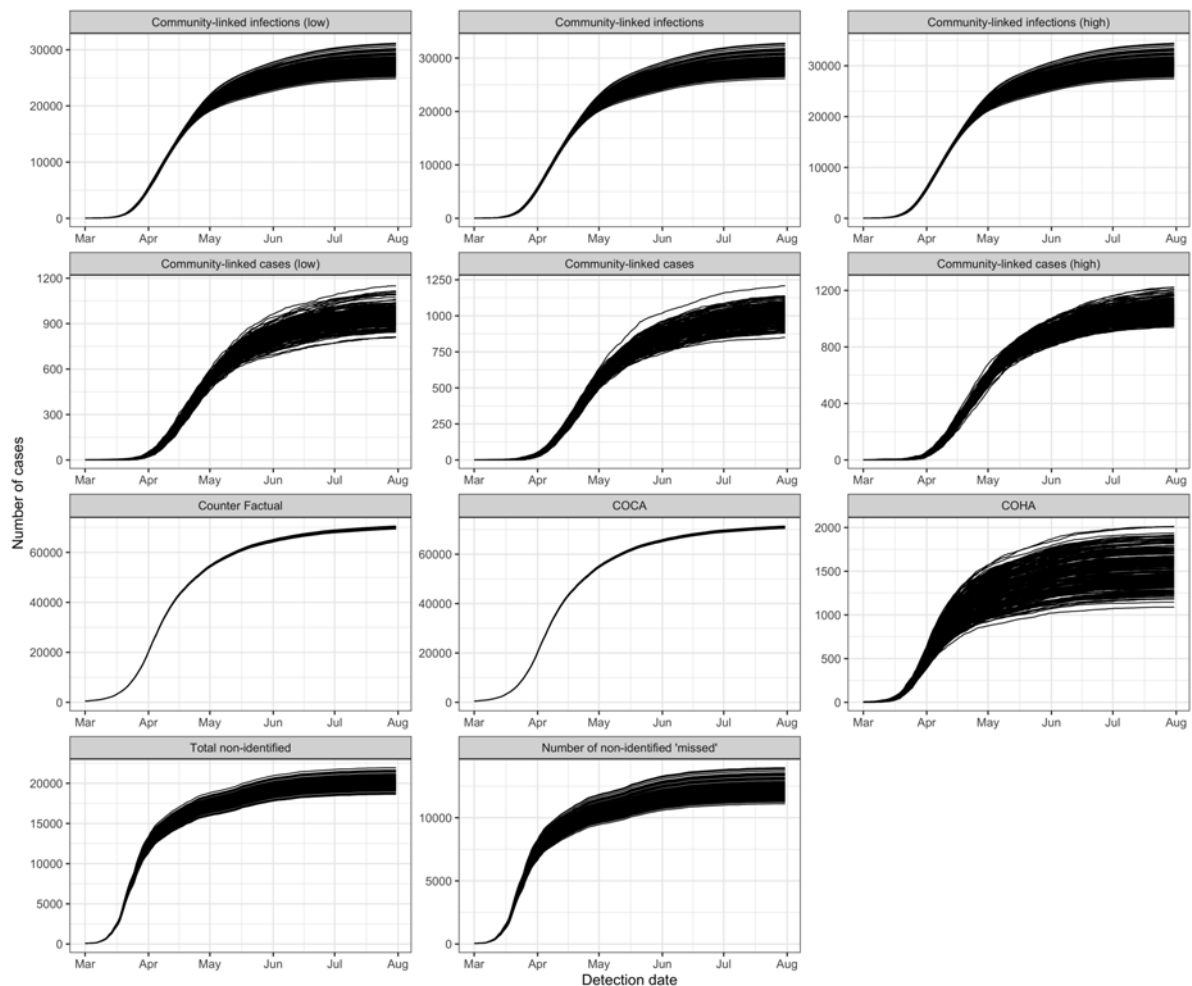
Figure S11: Example cumulative values as in Figure S10 for all 200 simulations (each black line) with a cutoff of at least 7 days from symptom onset to hospitalisation for defining a hospital-acquired case. The top two rows show the variation in community-onset, hospital-linked infections (first row) and subsequent cases (second row) at low, mean and high values of onward transmission (R = 0.76, 0.8, 0.84). The third row shows the counterfactual: the number of hospitalized cases there would be predicted to be without any hospital-acquisition of SARS-CoV-2, alongside the community-onset, community-acquired ("COCA") and community-onset, hospital-acquired ("COHA) case estimates. The final row shows the same variation shown in Figure S8: the total number of unidentified infections and the "missed" subset of these ("missed" due to discharge prior to symptom onset). Note the variation in the y axis values.
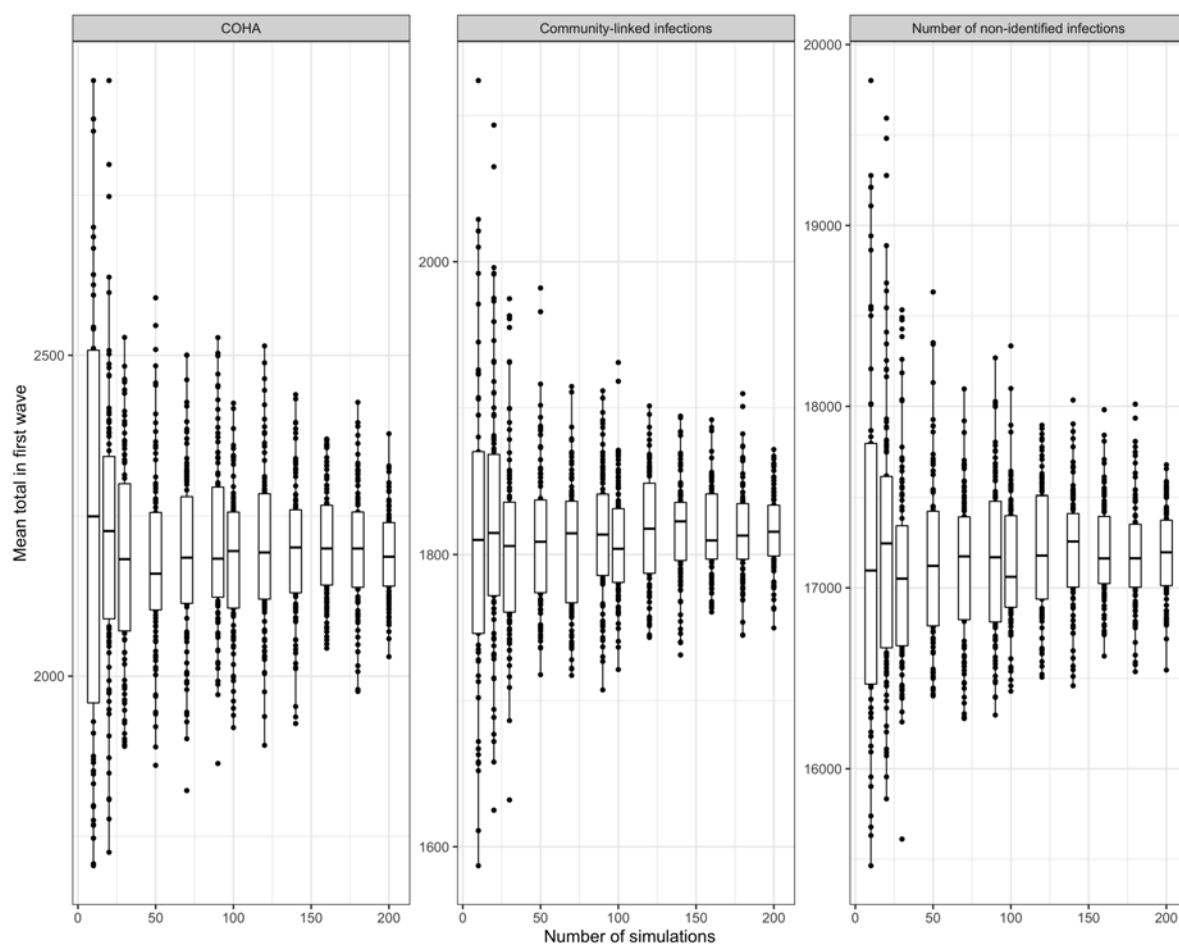
Figure S12: Boxplot of mean total value of key outcome variables over the "first wave" (to 31st July 2020) against the number of simulations. Left is "community onset, hospital-acquired" cases (COHA), middle are community-linked infections and right is the number of unidentified infections.

estimating the proportion of hospital-acquired infections identified. Moreover, much of the uncertainty was in the timing of events (symptom onset 2 or 5 days from infection for example), which, when aggregated over a 7-month period had little impact on the final aggregated results.

Supplementary 11: Additional results

- Figure 5 additional analysis

- Table S3: Additional reported results

- Table S4: Estimated percentage of "community onset, community acquired" infections that would be re-classified as "community onset, hospital acquired" infections

- Table S5: Estimated number of community onset, hospital-linked cases

- Figure S13: Impact of 1 vs 5 day discharge before associated identified hospital case

- Figure S14: Impact of R value variation over time (not just aggregated)

Figure 5 additional analysis

- Of all hospital patients who had a SARS-CoV-2 infection some time during their stay, 29.6% (28.9%, 30.5%) were hospital-acquired (E/(A+D), Figure 5).

- With the addition of hospital-linked infections, out of all hospital patients with a SARS-CoV-2 infection, 31.5% (30.6%, 32.4%) were estimated to have acquired their infection in hospitals or were hospital-linked ((E+F)/(A+D), Figure 5).

Table S3: Estimated additional main results for 14 and 4 day cut-offs in line with 7 day values in main text.

	Cutoff			
Estimate	7	14	4	Details
"hospital-onset, hospital-acquired" identified cases across acute English Trusts up to the 31st July 2020	6,640	4,440	7,830	From adjusted CO-CIN
unidentified hospital-acquired infections	20,000 (19,200, 21,100)	29,000 (28,400, 29,600)	17,500 (16,000, 19,300)	mean; 95% range over 200 simulations
Percentage of "community-onset, community-acquired" that should be classified as "community-onset, hospital-acquire"	2.1% (1.7%, 2.6%)	2.6% (2.1%, 3.1%)	2.1% (1.7%, 2.6%)	mean; 95% range over 200 simulations
"community-onset, hospital-linked" cases	1,600 (1,600, 1,700)	2,100 (2,000, 2,200)	1,600 (1,400, 1,700)	For the time varying R value mean; 95% range over 600 simulations

Table S4: Estimated mean and 95% quantile range over 200 simulations of the percentage of "community onset, community acquired" infections that would be re-classified as "community onset, hospital acquired" infections under different R values, hospital-acquired (HA) definition cutoffs (if symptom onset starts this many days from admission), discharge times from associated hospital-acquired case for unidentified hospital-acquired infection and scenarios for symptom onset to hospitalisation.

R value	HA definition cutoff (5,8,15)	Discharge time for unidentified hospital-acquired infection (1 or 5)	Symptom onset to hospitalisation scenario (1-3)	Mean	95% quantile range	
0.8	5	5	1	2.1	1.6	2.5
0.8	5	5	2	2.1	1.6	2.5
0.8	5	5	3	2.1	1.6	2.5
0.8	8	5	1	2.1	1.7	2.6
0.8	8	5	2	2.1	1.7	2.6
0.8	8	5	3	2.1	1.7	2.6
0.8	15	5	1	2.6	2.1	3.1
0.8	15	5	2	2.6	2.1	3.1
0.8	15	5	3	2.6	2.1	3.1
1.2	5	5	1	2.1	1.6	2.5
1.2	5	5	2	2.1	1.6	2.5
1.2	5	5	3	2.1	1.6	2.5
1.2	8	5	1	2.1	1.7	2.6
1.2	8	5	2	2.1	1.7	2.6
1.2	8	5	3	2.1	1.7	2.6
1.2	15	5	1	2.6	2.1	3.1
1.2	15	5	2	2.6	2.1	3.1
1.2	15	5	3	2.6	2.1	3.1
rt	5	5	1	2.1	1.6	2.6
rt	5	5	2	2.1	1.6	2.5
rt	5	5	3	2.1	1.6	2.6
rt	8	5	1	2.1	1.7	2.6
rt	8	5	2	2.1	1.7	2.6
rt	8	5	3	2.1	1.7	2.6
rt	15	5	1	2.6	2.1	3.1
rt	15	5	2	2.6	2.1	3.1
rt	15	5	3	2.6	2.1	3.1
0.8	5	1	1	2.1	1.6	2.5
0.8	5	1	2	2.1	1.6	2.5
0.8	5	1	3	2.1	1.6	2.5

0.8	8	1	1	2.1	1.7	2.6
0.8	8	1	2	2.1	1.7	2.6
0.8	8	1	3	2.1	1.7	2.6
0.8	15	1	1	2.6	2.1	3.1
0.8	15	1	2	2.6	2.1	3.1
0.8	15	1	3	2.6	2.1	3.1

Table S5: Estimated mean and 95% quantile range over 200 simulations number and percentage contribution of "community onset, hospital linked cases" (COHL) under different *R* values, hospital–acquired (HA) definition cutoffs (if symptom onset starts this many days from admission), discharge times from associated hospital-acquired infection and scenarios for symptom onset to hospitalization.

R value (0.8, 1.2, rt)	HA definition cutoff (5,8,15)	Discharge time for unidentified hospital-acquired infection (1 or 5)	Symptom onset to hospitalisation scenario (1-3)	Number of infections			Proportion of COHL		
				Mean	95% quantile range		Mean	95% quantile range	
0.8	5	5	1	1000	800	1100	1.3	1.2	1.5
0.8	5	5	2	1000	800	1100	1.3	1.2	1.6
0.8	5	5	3	1000	800	1100	1.3	1.2	1.5
0.8	8	5	1	1000	900	1100	1.4	1.2	1.5
0.8	8	5	2	1000	900	1100	1.4	1.3	1.5
0.8	8	5	3	1000	900	1100	1.4	1.2	1.5
0.8	15	5	1	1300	1200	1400	1.7	1.6	1.8
0.8	15	5	2	1300	1200	1400	1.7	1.6	1.8
0.8	15	5	3	1300	1200	1400	1.7	1.6	1.8
1.2	5	5	1	2600	2300	2900	3.6	3.2	4.1
1.2	5	5	2	2600	2300	2900	3.6	3.3	4.1
1.2	5	5	3	2600	2300	3000	3.7	3.3	4.2
1.2	8	5	1	2700	2500	3000	3.8	3.5	4.1
1.2	8	5	2	2700	2500	3000	3.8	3.5	4.1
1.2	8	5	3	2700	2500	3000	3.8	3.5	4.1
1.2	15	5	1	3400	3200	3700	4.6	4.3	4.9
1.2	15	5	2	3400	3300	3700	4.6	4.4	4.9
1.2	15	5	3	3500	3200	3700	4.6	4.3	5
rt	5	5	1	1600	1400	1700	2.2	2	2.4
rt	5	5	2	1600	1400	1700	2.2	2	2.4
rt	5	5	3	1600	1400	1700	2.2	2	2.4
rt	8	5	1	1600	1600	1700	2.3	2.1	2.4
rt	8	5	2	1600	1500	1700	2.3	2.1	2.4
rt	8	5	3	1600	1500	1800	2.3	2.1	2.4
rt	15	5	1	2100	2000	2200	2.8	2.7	2.9
rt	15	5	2	2100	2000	2200	2.8	2.7	2.9
rt	15	5	3	2100	2000	2200	2.8	2.7	2.9
0.8	5	1	1	900	800	1100	1.3	1.2	1.5

1.5	1.2	1.3	1100	800	1000	2	1	5	0.8
1.5	1.2	1.3	1100	800	1000	3	1	5	0.8
1.5	1.3	1.4	1100	900	1000	1	1	8	0.8
1.5	1.3	1.4	1100	900	1000	2	1	8	0.8
1.5	1.3	1.4	1100	900	1000	3	1	8	0.8
1.8	1.6	1.7	1400	1200	1300	1	1	15	0.8
1.8	1.6	1.7	1400	1200	1300	2	1	15	0.8
1.8	1.6	1.7	1400	1200	1300	3	1	15	0.8
4.1	3.2	3.6	2900	2300	2600	1	1	5	1.2
4.1	3.2	3.6	2900	2300	2600	2	1	5	1.2
4.1	3.3	3.6	2900	2300	2600	3	1	5	1.2
4.1	3.5	3.8	2900	2500	2700	1	1	8	1.2
4.1	3.5	3.8	3000	2500	2700	2	1	8	1.2
4.1	3.5	3.8	3000	2500	2700	3	1	8	1.2
4.9	4.3	4.6	3700	3200	3400	1	1	15	1.2
4.9	4.3	4.6	3700	3200	3400	2	1	15	1.2
4.9	4.3	4.6	3700	3200	3400	3	1	15	1.2
2.1	1.7	1.9	1500	1200	1300	1	1	5	rt
2.1	1.7	1.9	1500	1200	1300	2	1	5	rt
2.1	1.7	1.9	1500	1200	1300	3	1	5	rt
2.1	1.8	1.9	1500	1300	1400	1	1	8	rt
2.1	1.8	1.9	1500	1300	1400	2	1	8	rt
2.1	1.8	1.9	1500	1300	1400	3	1	8	rt
2.5	2.2	2.4	1900	1700	1800	1	1	15	rt
2.5	2.2	2.4	1900	1700	1800	2	1	15	rt
2.5	2.2	2.4	1900	1700	1800	3	1	15	rt

Impact of 1 - 5 day discharge

As shown in Figure S13, there is a minimal impact of varying the day of discharge of missed cases, except for the "community-onset, hospital-linked" (COHL) cases when using the time varying R estimates ("rt"). Cumulatively, up to the end of July 2020, this results in a less than 0.001% change in the number of "community-onset, community-acquired" cases but a $\sim 30\%$ higher number of "community-onset, hospital-linked" cases when using the time varying R estimates ("rt") and a 5 day discharge. This is due to a synergistic impact of the missed infections entering the community at peak R value (before early April).
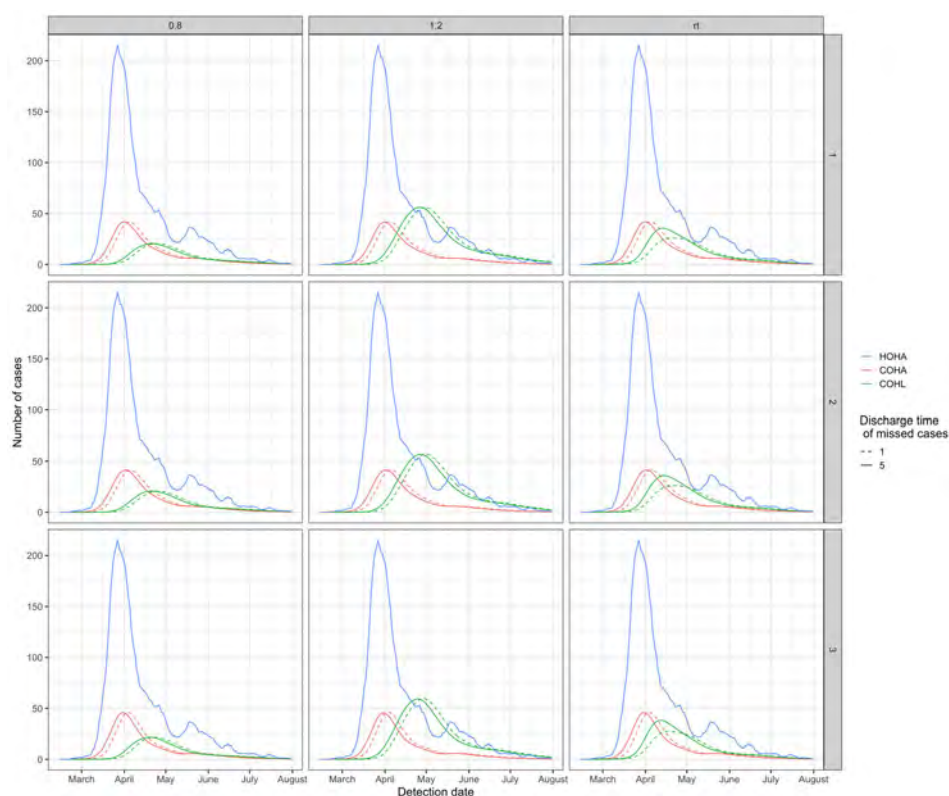
Figure S13: The impact of discharging missed cases 5 days (solid line, baseline) or 1 day (dashed line) before the associated identified hospital-acquired case at a cut-off threshold of 7 days from admission across different R values (columns) and Scenarios (rows) of symptom onset to hospitalisation. This is for "hospital-onset, hospital-acquired" (HOHA, blue), "community-onset, hospital-acquired" (COHA, red) and "community-onset, hospital-linked" (COHL, green) cases

Impact of R value variation
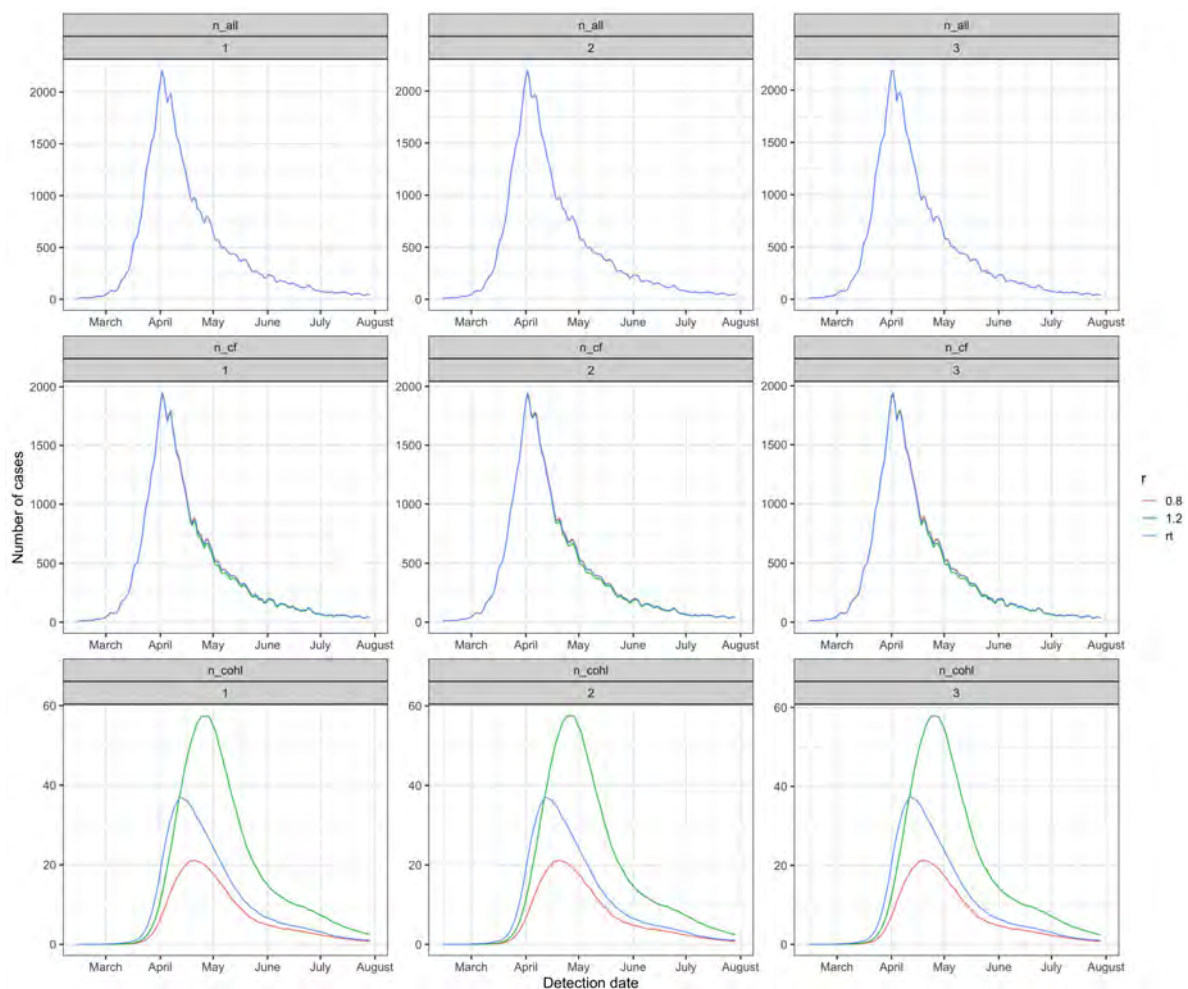
Figure S14: Time series of all hospitalized, counter-factual and community onset, hospital-linked cases under different onward transmission (R) values (median values shown here, colours). This is for a cutoff of 7 days from admission for the hospital-acquired definition and for the three scenarios (columns) for symptom onset to hospitalisation.

Supplementary analysis 12: Grouped Trust level analysis

Method: We applied the same analysis (shown in Figure 2) at the individual acute Trust level (n = 126) and then aggregated the results. We performed this analysis for all three cutoffs, three R values and for the first symptom onset to hospitalisation scenario with 50 simulations for each of the 126 Trusts to generate uncertainty ranges.

Results: 4 Trusts had no nosocomial cases recorded in the data over this time period. Two Trusts had no nosocomial cases recorded when using a 14 day cutoff for definition of a nosocomial case.

The proportion missed each week varied over trusts with a mean of 29% and a range between 0 and 88% over 50 simulations and all weeks and Trusts. 0.2% of the proportion detected estimates were zero.

Comparing the aggregated England setting (data pooled before doing analysis) to the grouped individual Trust (analysis performed at the Trust level and then aggregated) analysis shows similar results but the levels from the individual Trust analysis is higher (Table S5). Some variation would be expected due to rounding e.g. of the number of missed infections from the identified number of hospital-acquired cases. At the baseline cutoff of symptom onset 8 or more days from admission, the variation is relatively small, but it increases at a 15 or more days from admission cutoff, especially for "community-onset, hospital-linked" cases. The similarity in key indicators is shown in Figure S15-17 below. Using the grouped individual Trust analysis predicts that 25.5% (24.6%, 26.4%) of identified COVID-19 cases in hospitals were hospital-acquired, higher than the level predicted from the aggregated England setting: 20.1% (19.2%, 20.7%).

Comparison and interpretation: The proportion identified is predicted to be very small when there are few hospital-onset, hospital-acquired (HOHA) cases, as is often the case when doing the analysis at the individual trust level. Using the Bayesian framework to infer the total number of hospital-acquired infections ("trials") results in higher numbers ($\sim 50\%$) for the estimated number of unidentified hospital-acquired infections and hence onward case estimates (COHL / COHA). For example, 1 HOHA case, with a proportion detected of 0.005, is predicted to be linked to 524 hospital-acquired infections. However, 7 HOHA cases, with the same proportion detected, results in a predicted 1450 infections: an increase of 3x infections instead of 7x as might be expected from the increase in HOHA. We believe that the analysis at the Trust level suffers from issues of small numbers and

issues with using the empiric length of stay distributions. This leads to unrealistically small proportions detected and hence inflation to a greater number of missed infections.

Table S6: Comparison of England level vs. grouped trusts analysis for varying cutoff and R values for three key indicators under the first scenario for symptom onset to hospital admission. The cutoff is defined by the symptom onset on this day or later after admission. The values presented are the mean and 95% quantile over 200 simulations for the aggregated England setting and over 50 simulations for each Trust. Bold values are those with the baseline cutoff.

Setting	Cutoff for defining hospital-acquired	R value	No. of hospital-onset hospital-acquired identified cases (HOHA)	No. of unidentified hospital-acquired infections	No. of community-onset hospital-linked cases (COHL)
ENG	5	0.8	7,800 (7,800, 7,800)	17,400 (16,100, 19,100)	1,000 (900, 1,100)
grouped		0.8	7,400 (7,400, 7,400)	24,700 (22,500, 27,100)	1,500 (1,400, 1,700)
ENG		1.2	7,800 (7,800, 7,800)	17,400 (16,100, 19,100)	2,600 (2,400, 2,900)
grouped		1.2	7,400 (7,400, 7,400)	24,700 (22,500, 27,100)	3,400 (3,100, 3,900)
ENG		rt	7,800 (7,800, 7,800)	17,400 (16,100, 19,100)	1,600 (1,500, 1,700)
grouped		rt	7,400 (7,400, 7,400)	24,700 (22,500, 27,100)	2,400 (2,200, 2,600)
ENG	8	0.8	6,600 (6,600, 6,600)	20,000 (19,300, 21,000)	1,000 (900, 1,100)
grouped		0.8	6,200 (6,200, 6,200)	29,200 (28,100, 30,800)	1,700 (1,500, 1,800)
ENG		1.2	6,600 (6,600, 6,600)	20,000 (19,300, 21,000)	2,800 (2,500, 3,100)
grouped		1.2	6,200 (6,200, 6,200)	29,200 (28,100, 30,800)	3,800 (3,500, 4,200)
ENG		rt	6,600 (6,600, 6,600)	20,000 (19,300, 21,000)	1,700 (1,600, 1,800)
grouped		rt	6,200 (6,200, 6,200)	29,200 (28,100, 30,800)	2,500 (2,400, 2,700)
ENG	15	0.8	4,400 (4,400, 4,400)	29,100 (28,500, 29,700)	1,300 (1,200, 1,400)
grouped		0.8	4,000 (4,000, 4,000)	55,700 (51,500, 62,900)	2,600 (2,400, 3,000)
ENG		1.2	4,400 (4,400, 4,400)	29,100 (28,500, 29,700)	3,500 (3,300, 3,800)
grouped		1.2	4,000 (4,000, 4,000)	55,700 (51,500, 62,900)	6,600 (5,900, 7,400)
ENG		rt	4,400 (4,400, 4,400)	29,100 (28,500, 29,700)	2,100 (2,000, 2,200)
grouped		rt	4,000 (4,000, 4,000)	55,700 (51,500, 62,900)	4,100 (3,700, 4,500)
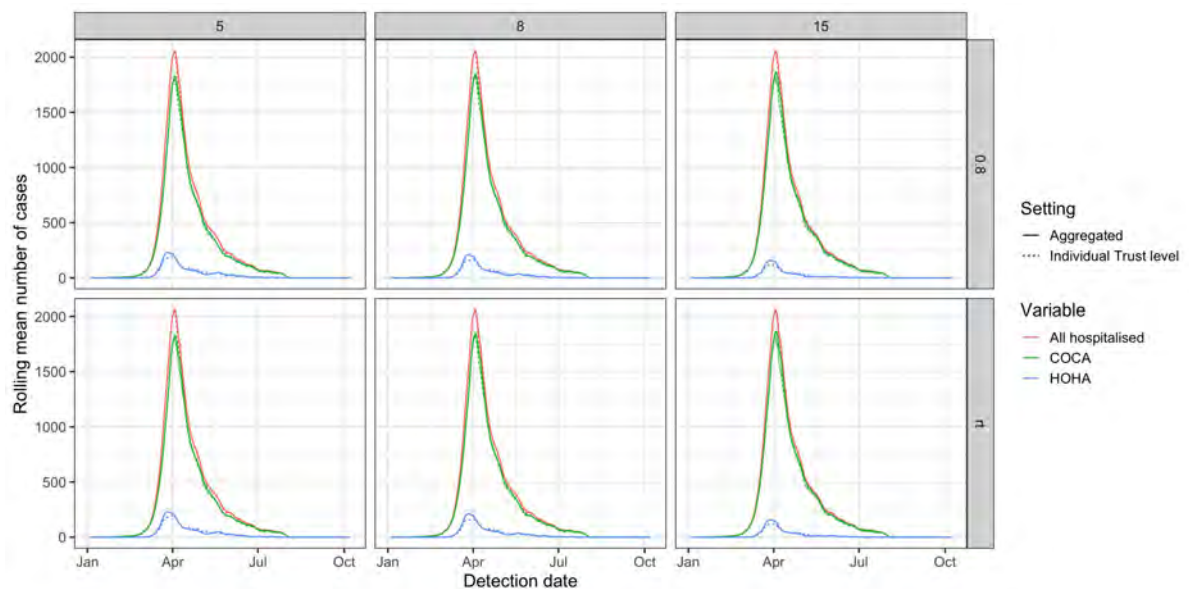
Figure S15: For the first symptom onset to hospitalisation scenario, there is little variation in the output of key case numbers (all hospitalized (red), community-onset, community-acquired (COCA, green) and hospital-onset, hospital-acquired (HOHA, blue)) if the analysis is performed on the aggregated England setting level (solid line, baseline) or at the individual Trust level and then aggregated (dashed line). The line here is the mean over 200 simulations for the aggregated England setting (50 simulations per Trust for the individual Trust analysis) and 95% range in shaded area.
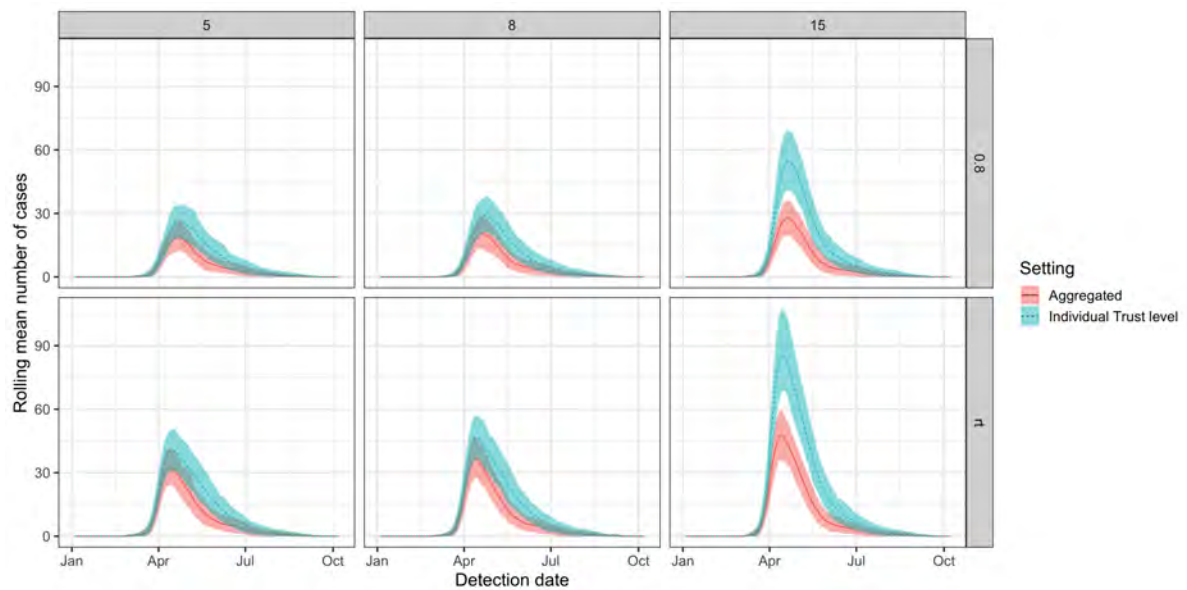
Figure S16: For the first symptom onset to hospitalisation scenario, there is some variation in the number of community-onset, hospital-linked cases if the analysis is performed on the aggregated England setting level (solid line, baseline, red) or at the individual Trust level and then aggregated (dashed line, blue). The line here is the mean over 200 simulations for the aggregated England setting (50 simulations per Trust for the individual Trust analysis) and 95% range given in the shaded area.
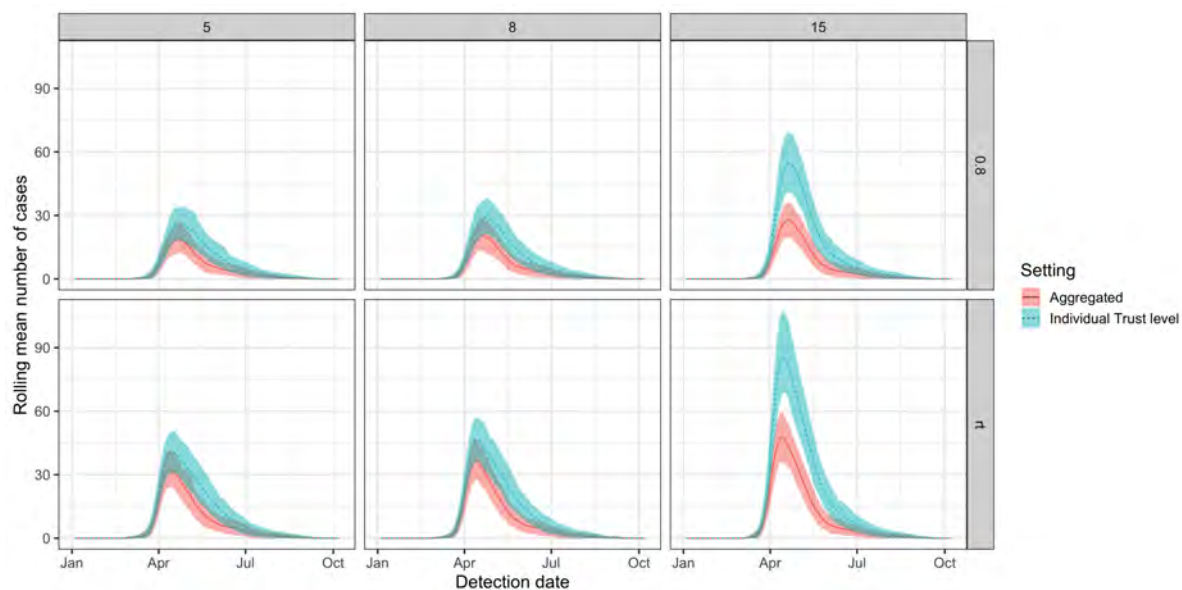
Figure S17: For the first symptom onset to hospitalisation scenario, there is some variation in the number of community-onset, hospital-acquired cases if the analysis is performed on the aggregated England setting level (solid line, baseline, red) or at the individual Trust level and then aggregated (dashed line, blue). The line here is the mean over 200 simulations for the aggregated England setting (50 simulations per Trust for the individual Trust analysis) and 95% range given in the shaded area.

Chapter 6

Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study

Thi Mui Pham[*], Hannan Tahir[*], Janneke H.H.M. van de Wijgert, Bastiaan Van der Roest, Pauline Ellerbroek, Marc J.M. Bonten, Martin C.J. Bootsma, Mirjam E. Kretzschmar

These authors contributed equally to this work.

Abstract

Background: Emergence of more transmissible SARS-CoV-2 variants requires more efficient control measures to limit nosocomial transmission and maintain healthcare capacities during pandemic waves. Yet, the relative importance of different strategies is unknown.

Methods: We developed an agent-based model and compared the impact of personal protective equipment (PPE), screening of healthcare workers (HCWs), contact tracing of symptomatic HCWs, and restricting HCWs from working in multiple units (HCW cohorting) on nosocomial SARS-CoV-2 transmission. The model was fit on hospital data from the first wave in the Netherlands (February until August 2020) and assumed that HCWs used 90% effective PPE in COVID-19 wards and self-isolated at home for seven days immediately upon symptom onset. Intervention effects on the effective reproduction number (R_E), HCW absenteeism and the proportion of infected individuals among tested individuals (positivity rate) were estimated for a more transmissible variant.

Results: Introduction of a variant with 56% higher transmissibility increased — all other variables kept constant — R_E from 0.4 to 0.65 (+63%) and nosocomial transmissions by 303%, mainly because of more transmissions caused by pre-symptomatic patients and HCWs. Compared to baseline, PPE use in all hospital wards (assuming 90% effectiveness) reduced R_E by 85% and absenteeism by 57%. Screening HCWs every three days with perfect test sensitivity reduced R_E by 67%, yielding a maximum test positivity rate of 5%. Screening HCWs every three or seven days assuming time-varying test sensitivities reduced R_E by 9% and 3%, respectively. Contact tracing reduced R_E by at least 32% and achieved higher test positivity rates than screening interventions. HCW cohorting reduced R_E by 5%. Sensitivity analyses for 50% and 70% effectiveness of PPE use did not change interpretation.

Conclusions: In response to the emergence of more transmissible SARS-CoV-2 variants, PPE use in all hospital wards might still be most effective in preventing nosocomial transmission. Regular screening and contact tracing of HCWs are also effective interventions, but critically depend on the sensitivity of the diagnostic test used.

Keywords: COVID-19; SARS-CoV-2; nosocomial transmission; agent-based modelling; infection control; contact tracing; healthcare worker screening; personal protective equipment; sensitivity; cohorting

Introduction

Effective interventions to limit nosocomial transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are pivotal to maintain healthcare capacities during pandemic waves [1, 2]. During the first epidemic wave many hospitals around the world restricted visits and canceled non-essential medical procedures in order to maintain adequate staffing levels for patients with COVID-19. In the Netherlands, specific infection control measures were implemented but nosocomial transmission may have been facilitated by temporary shortness of supplies of personal protective equipment (PPE), including gloves, goggles, face shields, gowns, and (N95) masks, at the onset of the pandemic.

Indeed, HCWs experienced a higher incidence of SARS-CoV-2 infections, compared to other professions, during the first pandemic wave [3–5]. Front-line HCWs in the UK and USA tested three times more frequently positive during the first epidemic wave than the general population after accounting for the frequency of testing [3]. Other studies from the UK and the Netherlands found higher SARS-CoV-2 incidences after the first epidemic wave among staff working in COVID-19 wards than staff working elsewhere in the hospital [5, 6]. In addition to direct contact with infectious patients, HCW-to-HCW transmission most likely also contributed to these elevated incidence rates.

Only a few studies incorporated modelling of SARS-CoV-2 transmission in healthcare set-tings [7–11]. In a stochastic within-hospital model, combined with a deterministic model reflecting SARS-CoV-2 transmission in the community, PPE use by HCWs and patients in the entire hospital substantially reduced nosocomial infections, while random weekly test-ing of asymptomatic HCWs and patients was less effective [9]. Moreover, strict cohorting of undiagnosed patients and HCWs in small units reduced the probability that SARS-CoV-2 introduction would lead to a large outbreak. In a deterministic within-hospital Susceptible-Exposed-Infectious-Recovered (SEIR) model isolating COVID-19 patients in single rooms or bays reduced infection acquisition in patients by up to 80% [8]. The model predicted that periodic testing of HWCs would have a smaller effect on the COVID-19 patient-burden than isolating patients but could reduce HCW infections by up to 64% and lead to a reduction of staff absenteeism. Both aforementioned models assumed a

time-invariant SARS-CoV-2 infectiousness and diagnostic PCR test with 100% sensitivity. An individual-based modelling study assessed the impact of different interventions for SARS-CoV-2 transmission in a non-COVID-19 hospital unit [11]. The model was calibrated to COVID-19 outbreak data in a neurosurgery hospital unit in Wuhan (January until February 2020). High-efficacy face-masks were shown to be most effective for reducing infection cases and workday loss. Reduction of contact rates had only a marginal effect on mitigating the outbreak in the long run. Another model (stochastic, individual-based, aimed at patients and HCWs in long-term care facilities (LTCF) did incorporate a test sensitivity that varies with time since infection [7]. This model concluded that pooled testing (combining clinical specimens from multiple individuals into a single biological sample for a single RT-PCR test) was the most effective and efficient surveillance strategy for resource-limited LTCFs.

While these previous studies investigated interventions such as the PPE use, physical distancing among HCWs, various testing strategies, and cohorting of patients and HCWs, the impact of contact tracing within hospital settings has not been modeled yet. Observational evidence from 5,700 HCWs in two large hospitals and 40 outpatient units in Milan, Italy, suggested that random testing (positivity rate of 2.6%) was less efficient than contact tracing (10%) [12].

In Dutch hospitals patients and HCWs were cohorted in COVID-wards, where HCWs used PPE during patient care, in addition to the basic infection control measures applied. With these measures, nosocomial transmission was considered well-controlled during the first wave of the pandemic, although outbreaks have been reported sporadically [13]. Yet, with the emergence of more transmissible variants, current infection control measures may become less effective. While COVID-19 vaccine rollout is underway, it is still unclear how they affect transmission and how their efficacy is affected by the new SARS-CoV-2 variants. We, therefore, explored the relative effectiveness of different infection prevention strategies for HCWs in hospitals in the absence of vaccination using an agent-based model of nosocomial SARS-CoV-2 transmission. First, we fitted the model to real-life data from the University Medical Center Utrecht (UMCU) during the period February-August 2020. Next, we evaluated the impact of various interventions on transmission, HCW absenteeism and test positivity as a marker of intervention efficiency for a more transmissible variant (e.g., B.1.1.7) and draw general conclusions for infection control in hospitals with a similar structure.

Methods

Agent-based model

We developed an agent-based model that describes the dynamics of SARS-CoV-2 transmission in a hospital allowing for importations of infections from the community (Fig 1A). We modeled a hospital comprising four ward types: 1) general COVID wards, 2) general non-COVID wards, 3) COVID intensive-care units (ICUs), and 4) non-COVID ICUs. Within the hospital we distinguish patients, nurses, and doctors. Patients are assumed to occupy a hospital bed in a single room. HCWs (nurses and doctors) work in duty shifts. HCWs meet patients in a number of rounds per shift (Additional File 1: Table S1), and HCWs meet other HCWs in the common staff room of each ward. Individuals may be in one of the disease states: susceptible (S), asymptomatically infected (I_A), pre-symptomatically infected (I_P), infected with moderate symptoms (I_M), infected with severe symptoms (I_S), and recovered (I_R). We did not explicitly model other respiratory tract infections with similar symptoms. Hence, all symptomatic individuals are necessarily infected with SARS-CoV-2. We did not model death in our simulations. All infected individuals are assumed to be infectious following a time-varying infectiousness curve. We denote infectiousness over time since infection τ by $\beta(\tau))$, i.e., it is the mean rate at which an individual infects others at time τ after its time of infection. The reproduction number R (average number of secondary cases caused by an infected individual) is given by integrating $\beta(\tau)$ over time since infection $R = \int_0^\infty \beta(\tau)d\tau$. Assuming the mean generation time () to be equivalent with the observed mean serial interval, we calculated the infectiousness profile by $\beta(\tau) = \omega(\tau)R$. Based on this, the individual's infectiousness follows a Weibull distribution with a mean of 6 days (Fig 1C) [14] and the reproduction number is a scaling factor of the infectiousness profile. We assumed the infectiousness to differ between asymptomatic and symptomatic infected individuals, defined by $\beta_A(\tau)$ and $\beta_S(\tau)$, respectively. Then $\beta(\tau)$ can be decomposed into

$$\beta(\tau) = P_A\beta_A(\tau) + (1 - P_A)\beta_S(\tau)$$

where P_A represents the proportion of asymptomatic infections. Asymptomatic individuals are assumed to have an infectiousness proportional to that of symptomatic individuals, i.e., $\beta_A = x_A \cdot \beta_S, x_A \leq 1$. Integrating over each of the two terms leads to the respective

contribution to the overall reproduction number:

$$R = R_A + R_S = \int_0^\infty P_A \cdot x_A \cdot \beta_S(\tau)d\tau + \int_0^\infty (1 - P_A)\beta_S(\tau)d\tau$$

Transmission events can occur between patients and HCWs, and among HCWs. We assumed no patient-to-patient transmission as patients are assumed to occupy single-bed rooms. Only HCWs in their asymptomatic or pre-symptomatic phase contribute to transmission. We assumed that the incubation period has a Gamma distribution with mean 5.5 days [15]. Patients may be admitted to the hospital for non-COVID reasons or with moderate or severe COVID-19 symptoms. In the first case, they may be susceptible, pre-symptomatically, or asymptomatically infected. Symptomatically infected patients are admitted to COVID wards (moderate symptoms) or COVID ICUs (severe symptoms). Patients in non-COVID wards that develop symptoms during their stay are immediately transferred to COVID wards. We assumed that moderately and severely infected patients recover after 14 and 35 days, respectively [16].

Data and parametrization

We used data from the UMCU to parametrize the number of wards and beds per ward (Additional File 1 pp. 2). We used the number of patients admitted to the UMCU for non-COVID reasons and their length of stay for the time period 2014-2017 and assumed a 50% decrease in admissions during the study period (Additional File 1: Table S1). The daily number of COVID-19 hospitalizations and their length of stay distribution was based on UMCU data from 27 February until 24 August 2020 (Additional File 1: Figure S1-S2). The simulations started on 30 December 2019 with a hospital at 100% occupancy without any SARS-CoV-2-infected individuals. The first COVID-19 admissions occurred on 27 February 2020. To account for admissions of patients that are infected but not (yet) symptomatic and HCWs who were (unknowingly) infected in the community, we used daily national numbers of SARS-CoV2 infectious individuals estimated by the Dutch National Institute for Public Health and the Environment (RIVM) from 17 February until 24 August 2020 (Additional File 1 pp. 2) [17]. We additionally used publicly available age-specific hospitalization rates in the Netherlands in 2012 and age-specific SARS-CoV-2 infection incidence rates in Utrecht province to scale the daily probability of being infected in the community for non-COVID patients and HCWs arriving in the hospital
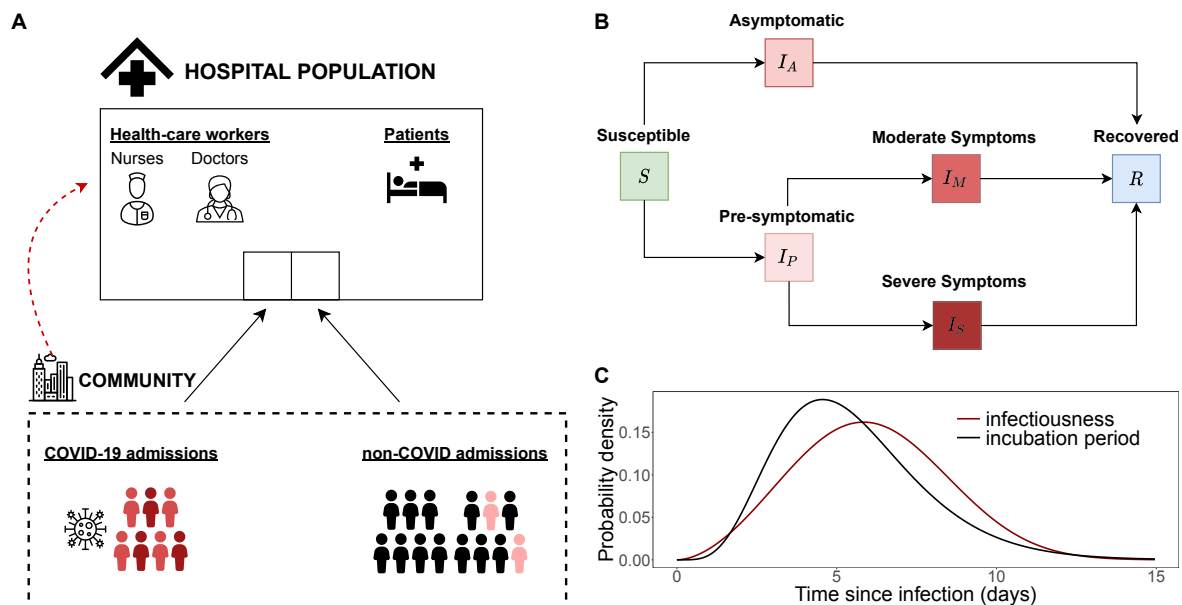
Figure 1: **Schematics for agent-based model.** (A) Diagram of the agent-based model including the agents in the main environment (hospital) and community importations. The hospital population is divided into healthcare workers (nurses and doctors) and patients. Patients may be admitted from the community either with moderate (red) or severe (dark red) COVID-19 symptoms or for non-COVID reasons. Patients may be in a pre-symptomatic stage (light red) when hospitalized to non-COVID wards. Healthcare workers may get infected in the community (red dashed line). (B) Disease progression diagram. Individuals are in either of the following categories: Susceptible (S), Asymptomatically Infected (I_A), Pre-symptomatically infected (I_P), Moderately infected (I_M), Severely infected (I_S), and Recovered (R). (C) Probability density of infectiousness of an infected individual and incubation period over time since infection.

[**CBS2019Ziekenhuisopnamen2019**, 18]. Based on a published meta-analysis, we assumed that a fixed percentage of 20% and 31% of SARS-CoV-2 infections in patients and HCWs, respectively, were asymptomatic (see also Table 1) [19]. First, we chose the basic reproduction numbers R_S and R_A such that the numbers of occupied beds by COVID-19 patients predicted by our model were in good agreement with real-life UMCU data on the number of COVID-19 patients at UMCU during the first epidemic wave by visual inspection (Table 1 and Fig 2A). During this calibration, a change in the basic reproduction numbers R_S and R_A resulted in a change of the individual's infectiousness per time unit and thus the probability of transmission per contact. The remaining parameters did not change. These reproduction numbers incorporated the effects of typical (but not COVID-specific) infection prevention measures in the hospital. We will refer to the model parameterized with these reproduction numbers as the wild-type scenario.

This scenario also assumed that HCWs use 90% effective PPE (i.e., 90% reduction in infectiousness) in COVID wards and isolate at home immediately upon symptom onset for seven days, after which they return recovered to work. Next, we introduced a more transmissible SARS-CoV-2 variant into the hospital, keeping all other parameters – including PPE use in COVID wards and self-isolation after symptom-onset – the same. Based on recent results for B.1.1.7, we assumed a 56% increase in transmissibility [20]. We will refer to the model parameterized with these higher reproduction numbers as our baseline scenario. Various intervention scenarios were compared to this baseline scenario.

Diagnostic performance of the PCR test

We assumed a PCR test specificity of 100% and distinguished two scenarios for the test sensitivity: 1) a time-invariant perfect sensitivity of 100%; and 2) a sensitivity increasing with time since infection with a maximum sensitivity of 93.1% close to symptom onset and declining afterward (time-varying sensitivity) [14]. We considered two sensitivity analyses to test the impact of PCR test sensitivity assumptions on our results (Additional File 1 pp.3 and Fig S1). Hospital staff typically self-quarantine from symptom onset, get tested and receive their test results within hours (based on UMCU data). We, therefore, assumed no delay between testing and receiving test results, and that HCWs do not contribute to virus transmission after symptom onset.

Infection control interventions

Baseline scenario

In the baseline scenario, HCWs were assumed to use PPE in COVID wards when attending to patients, but not during breaks or in other parts of the hospital. The baseline reduction factor (PPE effectiveness) was assumed to be 90%, which includes both perfect-use PPE efficacy and expected PPE use adherence level. We assumed that 95% of the HCWs work in the same ward as during their previous shift.

All interventions described below were in addition to the baseline scenario. An overview of all scenarios is given in Fig 2.

Intervention: PPE in all wards

In this scenario, all HCWs used 90% effective PPE in all (non-COVID and COVID) wards. However, no PPE was used when HCWs meet each other off-ward. We performed sensitivity analyses assuming PPE effectiveness of 50% and 70%.

Intervention: HCW cohorting (no ward change)

This scenario restricted HCWs to work only in specific wards and did not allow any ward changes. This scenario represents the most optimistic scenario where both nurses as well as physicians are assumed to be eligible for cohorting to the same degree.

Intervention: Regular HCW screening

All HCWs were tested for SARS-CoV-2 either with a) a test with perfect sensitivity every three days, or a test with time-varying sensitivity, b) every three days, or c) every seven days. If tested positive, HCWs were assumed to immediately self-isolate at home for seven days.

Intervention: HCW contact-tracing

If a HCW developed symptomatic SARS-CoV-2 infection, all contacts in the hospital during a time window of either two or seven days before symptom onset were traced and tested. We will refer to these scenarios as 2-day Contact tracing and 7-day contact tracing. For 2-day contact tracing, contacts were always tested assuming a time-varying test sensitivity. For 7-day contact tracing, we distinguished between perfect and time-varying sensitivity sub-scenarios. In the perfect sensitivity sub-scenario, contacts were instantaneously tested on the day of symptom onset of the index (the HCW). In the time-varying test sensitivity sub-scenario, the test was performed on the day of symptom onset if the contact with the index was more than five days ago. Otherwise, it was performed on day five after the contact. Exposed HCWs awaiting tests were assumed to wear PPE during contact with any patient and with other HCWs. In case of a positive test, patients were moved to a COVID ward while infected HCWs were sent home for self-isolation for seven days and replaced by susceptible HCW. We did not model any absences of HCWs with disease symptoms caused by other respiratory pathogens.

Scenarios	Virus type	PPE use (90% effectiveness)	HCW cohorting	HCW screening	HCW contact tracing	PCR test sensitivity
Wild-type scenario	Wild-type	COVID wards	X	X	X	—
Baseline scenario	more transmissible variant	COVID wards	X	X	X	—
No HCW ward change	more transmissible variant	COVID wards	✓	X	X	—
PPE in all wards	more transmissible variant	COVID wards non-COVID wards	X	X	X	—
Screening 3 days perfect sens	more transmissible variant	COVID wards	X	✓	X	time-invariant, perfect (100%)
Screening 3 days	more transmissible variant	COVID wards	X	✓	X	time-varying, imperfect
Screening 7 days	more transmissible variant	COVID wards	X	✓	X	time-varying, imperfect
7-day Contact tracing perfect sens	more transmissible variant	COVID wards	X	X	✓	time-invariant, perfect (100%)
7-day Contact tracing	more transmissible variant	COVID wards	X	X	✓	time-varying, imperfect
2-day Contact tracing	more transmissible variant	COVID wards	X	X	✓	time-varying, imperfect

Figure 2: **Overview of all simulated scenarios.** The main characteristics of the scenarios simulated in our agent-based model are presented.

Outcome measures

We computed the effective reproduction number R_E (average number of secondary cases caused by an infected individual) to evaluate an intervention's effectiveness. We calculated an overall R_E for an average individual (patients and HCWs combined) but also stratified R_E by patients, HCWs, and symptom status. The reproduction numbers of patients were calculated for those who eventually developed symptoms (R_S^{pat}) and those who remained without symptoms (R_A^{pat}). Since HCWs were assumed to immediately self-isolate upon symptom onset, we calculated R during pre-symptomatic (R_S^{hcw}) and asymptomatic states (R_A^{hcw}). To evaluate the maximum demand on hospital capacity, we considered the total number of nosocomial infections among patients and HCWs over time. In addition, we computed the percentage of absent HCWs due to self-isolation (because of symptom onset or detection via screening or contact-tracing) over time. We assessed the efficiency of screening and contact-tracing interventions by their positivity rates (percentage of detected infected individuals among tested individuals). We did not include individuals that developed symptoms prior to being tested in the positivity rate calculations since those were already detected and isolated in our model.

For every scenario and outcome measure, we calculated the mean and 95% percentiles over 100 simulation runs (95% uncertainty interval). We calculated positivity rates over time merging data from all simulation runs and computed 95% Bayesian beta-binomial credibility intervals.

A detailed description of the full model and the parameters can be found in the appendix. We performed sensitivity analyses to test the robustness of our results (Table 1) and the respective results are shown in the appendix. The data and full code are available from: `https://github.com/htahir2/covid_intra-hospital_model.git`.

Results

We observed good agreement between the number of patients in COVID wards predicted by our wild-type scenario and the real-life UMCU data during the first wave for $R_S = 1.25$ and $R_A = 0.5$. However, the model slightly overestimates hospitalizations for the second half of the first wave (Fig 2A). We subsequently assumed the introduction of a SARS-CoV-2 variant with a 56% increase in transmissibility (based on B.1.1.7 data), resulting in $R_S = 1.95$ and $R_A = 0.8$. Keeping all other parameters the same, including HCWs using PPE in COVID wards and self-isolating at symptom-onset, the total number of nosocomial transmissions increased by 303% (Fig 2B) and the overall effective reproduction number increased by 62.5% (Fig 2C). R_S^{hcw} and R_S^{pat} increased the most to 0.94 and 0.6, respectively (Fig 2D), indicating that pre-symptomatic individuals pose the highest risk for onward transmissions.

Intervention effects on reproduction numbers

In the context of this SARS-CoV-2 variant with higher transmissibility, the baseline scenario of 90% effective PPE use in COVID wards yielded an overall R_E of 0.65 (Fig 4A). Extending PPE use to non-COVID wards reduced R_E by an additional 85%, to 0.1. Restricting HCWs to work only in specific wards yielded a reduction in R_E of 5% (to 0.62). The effect of HCW screening on R_E highly depended on the test sensitivity. With time-varying test sensitivity, screening every three or seven days reduced R_E to 0.59 and 0.63 (reductions of 9% and 3%), respectively. When perfect sensitivity was assumed, screening every three days reduced R_E by 63%, to 0.24. The impact of contact-tracing

also depended on the test sensitivity assumptions, but to a lesser extent. For perfect test sensitivity, 7-day contact-tracing reduced R_E by 32%, to 0.44. For time-varying test sensitivity, the 2-day and 7-day contact-tracing scenarios reduced R_E to 0.41 and 0.39 (reductions of 37% and 40%), respectively. The additional reductions of R_E by the intervention scenarios over and above the baseline scenario were most prominent for pre-symptomatic HCWs (Fig 4B).

Intervention effects on numbers of nosocomial infections

PPE use in all wards or HCW screening every three days with perfect test sensitivity would prevent 93.7% and 82.7% of all transmissions, respectively (Fig 5), and both interventions would also prevent outbreaks among patients and HCWs (Fig 6). Reductions in nosocomial infections were much smaller for regular screening interventions with time-varying test sensitivity: screening every three days would lead to a 20.4% reduction and screening once a week to a 10.1% reduction. Testing with perfect test sensitivity followed by 7-day contact-tracing was more effective (55.8% reduction of transmissions) than regular screening every three or seven days. Testing with time-varying sensitivity followed by 2-day or 7-day contact tracing were similarly effective as testing with perfect sensitivity followed by 7-day contact tracing (reductions of 61.4% and 64.1%, respectively). HCW cohorting would decrease the total number of nosocomial infections by 13%. Note that our model predicted that approximately 30% of patients that either got admitted with SARS-CoV-2 or acquired the infection in the hospital were detected either due to testing at symptom onset or testing as part of an intervention (Additional File 2: Fig. S1). The remaining 70% of infected patients were discharged undiagnosed and without symptoms.

Intervention effects on HCW absenteeism

Our baseline scenario predicted a maximum HCW absenteeism of 5.4%, including absenteeism due to symptoms or home isolation (Fig 7). When comparing intervention scenarios to the baseline scenario, HCW absenteeism is lowest for PPE use in all wards (a maximum of 2.3%). The maximum absenteeism percentages were 5.2% for HCW cohorting, 5.1% for regular screening with perfect test sensitivity, 8.6% for regular screening with time-varying test sensitivity every seven days and 6.6% every three days, 4.0% for 7-day contact tracing with testing assuming perfect sensitivity, 3.6% for 2-day contact

tracing with testing assuming time-varying sensitivity, and 3.9% for 7-day contact tracing with testing assuming time-varying sensitivity.

Efficiency of screening and contact-tracing interventions

HCW screening every three days with a perfect test would lead to the lowest test positivity rate of all testing-based interventions (Fig 8A). Screening of HCWs every week compared to every three days yields higher positivity rates with its mean reaching a maximum value of 5.1%. The positivity rate of screening interventions linearly increases with increasing prevalence (Additional File 2: Figure S1).

Positivity rates for contact-tracing interventions are much higher than for screening interventions, reaching as high as 15.1% when a perfect test sensitivity is assumed (Fig 9A). The maximum positivity rates for 2-day and 7-day contact tracing with time-varying test sensitivities are only slightly lower at 11.3% and 10.4%, respectively (Fig 9B-C). Positivity rates of contact-tracing interventions are stable across prevalence values (Additional File 1: Figure S2).

Sensitivity analyses show that our findings do not change significantly when the assumed PPE effectiveness is reduced to 70%. When PPE effectiveness is assumed to be as low as 50%, screening every three days with perfect sensitivity becomes more effective than PPE use in all wards. However, PPE use in all wards is still more effective than all other interventions (Additional File 2 p. 2).

Discussion

During the first epidemic wave of the wild-type SARS-CoV-2 in the Netherlands, nosocomial transmission was considered to be of relative minor importance. Our results suggest that a more transmissible virus variant could significantly increase the total number of nosocomial transmissions if hospital prevention measures would not be expanded beyond those implemented during the first wave (HCWs using PPE with assumed 90% effectiveness in COVID-19 wards and self-isolating at home after symptom onset). Our findings suggest that universal PPE use in all hospital wards is the most effective intervention to reduce the reproduction number and absenteeism. These results are consistent with a previous modelling study and previous findings on significant

reductions of nosocomial-acquired SARS-CoV-2 infections after implementation of universal masking policies [1, 11, 13, 21–24]. In our model, HCW cohorting only had a small impact on nosocomial transmissions, which is due to the fact that we assumed 90% effective PPE use in the COVID wards in all scenarios. Several studies have reported elevated risks for HCWs working in COVID-19 patient care [5, 6]. Our results suggest that maintaining sufficient PPE supplies in hospital settings may reduce the need for implementing additional HCW cohorting strategies. Our model also suggested that regular screening of HCWs could have a strong impact, but only if the test sensitivity is high throughout the infectious period. Tests with imperfect time-varying sensitivity miss many infections during the pre-symptomatic phase. Indeed, our model identified pre-symptomatically infected HCWs as drivers of transmission both to patients and to other staff. This is consistent with a descriptive study on HCWs in France where contacts causing the transmissions took place in the pre-symptomatic phase of the index case in 30% of all cases and in almost 50% of HCW-HCW transmissions. Our results also agree with previous modelling studies suggesting that regular screening of HCWs was less effective than effective PPE use.

Contact tracing was highly effective in limiting nosocomial transmissions in our model, especially when traced contacts are tested at least five days after their exposure and precautionary measures are undertaken in the meantime. If traced HCWs are immediately tested, self-isolated, and replaced by susceptible HCWs, this can lead to increased transmission, a phenomenon that was also observed by Scarpino and colleagues [25]. The authors used a network model and evidence from data on influenza and dengue outbreaks to show that replacing infected individuals in essential societal roles with susceptibles may lead to accelerated transmission. Our results indicate that allowing traced HCWs to work with PPE in all hospital wards is more effective in limiting transmission. Finally, our model suggests that contact tracing yields higher positivity rates than screening interventions, not only at high prevalence but also during periods of low infection rates, making this also a potentially successful and cost-effective infection control strategy in hospital settings. Our findings reinforce the recommendation by Paltansing and colleagues to test all close contacts of a SARS-CoV-2 positive case immediately and subsequently on day 3 and 7 regardless of symptoms and to allow HCWs to work with surgical masks while awaiting their test results [13].

Our study has several limitations. First, we assumed that transmission occurs solely via HCWs in the absence of a direct patient-to-patient contact pathway, as has been used before in an individual-based model of nosocomial influenza transmission [26]. Assuming similar transmission modes for SARS-COV-2, we consider this assumption reasonable for hospital settings in Western countries where direct patient-to-patient contact is rare. When this assumption is violated, our estimated impact of HCW-based interventions is likely to be overestimated. Second, we considered SARS-CoV-2 as a cause of symptoms and neglected other respiratory tract infections. Thus, real-life positivity rates of contact tracing may be lower than presented in this study. Third, while we have included age-specific hospitalization rates for patients admitted with SARS-CoV-2 and different proportions of asymptomatic infections for HCWs and patients, we have neglected age-structure in our transmission model. A possible extension of our model would be the inclusion of age-dependent susceptibility and infectiousness parameters. However, since the considered interventions in our model are not differential with respect to age, we do not expect any impact on the relative effect of the interventions. Further, our HCW cohorting intervention scenarios assumes the same degree of cohorting both for nurses and physicians. In reality, cohorting strategies are only feasible for nurses. As such, the estimated effect of this intervention is likely to be overestimated. Since the estimated effect of HCW cohorting was estimated to be small, we expect it to be even smaller when implemented in the real world. Moreover, the duration of contacts, SARS-CoV-2 reinfections, visitors or other ancillary staff, delays between symptom onset and isolation, or delays between test application and test result were not included. Finally, while we identified one parameter set for which our model results fitted the available data well, it is possible that other parameter sets exist that would produce a comparable fit. We have not used formal fitting procedures to match our model results to the data given the large number of parameters. However, qualitatively, our conclusions were robust in sensitivity analyses to variation of the most important model parameters. While our model was developed using data of a large Dutch teaching hospital and of the first wave of the COVID-19 epidemic in the Netherlands, our results can be generalised to other hospitals with a similar structure and may be relevant for subsequent waves and future infectious disease outbreaks.

Conclusions

In conclusion, our model demonstrates that PPE use in all wards is the most effective measure to substantially reduce nosocomial spread of SARS-CoV-2 variants with higher transmissibility. However, contact-tracing and regular screening using high-sensitivity tests are also effective interventions, which might be preferred in some settings.

Availability of data and materials

The datasets used and/or analysed as well as the full code reproducing the results in the current study are available from `https://github.com/htahir2/covid_intra-hospital_model.git`.

Author contributions

TMP and HT have contributed equally to this work. TMP, HT, MK, MCJB, and JHHMvdW developed the conceptual framework of the study. TMP, HT, MK and MCJB developed the model. HT programmed the model and produced the output. HT and TMP produced the results of the model. TMP produced the visualization for the main text and the appendix. TMP, MK, BvdR and JHHMvdW conducted the literature research. PE and BvdR collected the data. TMP and HT have verified the underlying data. MK, MCJB, MB, JHHMvdW and PE contributed to interpretation of the results. TMP

wrote the original draft of the main text. TMP and HT wrote the appendix. All authors provided critical review of the manuscript, and approved its final version for submission.

Ethics approval and consent to participate

Not applicable (only unlinked aggregated data is used).

Competing interests

The authors declare that they have no competing interests.

Table 1: **Parameter values for the agent-based model.**

	Symbol	Description	Distribution/Value*	Source
Incubation period	$s(\tau)$	Time between infection and symptom onset	Gamma distribution shape = 5.807 scale = 0.948 mean = 5.510 SD = 2.284	Lauer et al [15]
Generation time	$\omega(\tau)$	Time between becoming infected and subsequent onward transmission events	Weibull distribution shape = 2.826 scale = 6.839 mean = 6	Grassly et al [14]
Proportion of asymptomatic infections among infected patients	P_A^p		20%	Buitraga-Garcia et al [19]
Proportion of asymptomatic infections among infected HCWs	P_A^h		31%	Buitraga-Garcia et al [19]
Proportion of severe symptomatic individuals	P_S	Proportion of exposed individuals that will develop severe symptoms	20%	Wu et al [27]
Reproduction number of asymptomatic infectees for wild-type variant	R_A^W	Mean number of infections caused by an individual asymptomatically infected with the wild-type SARS-CoV-2 variant	0.5	Calibrated to UMCU data

Reproduction number of symptomatic infectees for wild-type variant	R_S^W	Mean number of infections caused by an individual symptomatically infected with the wild-type SARS-CoV-2 variant	1.25	Calibrated to UMCU data
Reproduction number of asymptomatic infectees for new virus variant	R_A	Mean number of infections caused by an individual asymptomatically infected with the SARS-CoV-2 variant	0.8 (1.95)	Based on R_A^W with 56% higher transmissibility, varied in sensitivity analysis
Reproduction number of symptomatic infectees for new virus variant	R_S	Mean number of infections caused by an individual symptomatically infected with the SARS-CoV-2 variant	1.95	Based on R_A^W with 56% higher transmissibility
Maximum sensitivity of diagnostic PCR test			93.1% (79%)	Grassly et al [14], varied in sensitivity analysis
Proportion of HCWs that work in the same ward as during their previous shift			95% (baseline) 100% (intervention)	Assumed
PPE effectiveness		Reduction in infectiousness upon contact between an infected and susceptible individual (includes PPE efficacy and adherence)	90% (50%, 70%)	Suzuki et al [28], Qian et al [29], Bessesen et al [30], varied in sensitivity analysis

Isolation period for HCWs		Amount of time HCWs have to isolate after symptom onset or after being detected by screening or contact tracing	7 days	Assumed
Recovery time for asymptomatic infection		Mean duration of an asymptomatic infection	14 days Sensitivity analysis: Unif(9, 19)	Assumed
Recovery time for symptomatic (moderate, severe) infection		Mean duration of a symptomatic infection	14 days (moderate) 35 days (severe) Sensitivity analysis: Unif(9, 19) Unif(30, 40)	Liu et al [16]
LoS of non-COVID patients in ICU			Lognormal meanlog = 0.37 sdlog = 0.82 mean = 1.45 days sd = 2.27	Fitted distributions to UMCU data from 2014-2017
LoS of non-COVID patients in normal ward			Weibull shape = 0.92 scale = 4.18 mean = 4.35 days	Fitted distributions to UMCU data from 2014-2017
LoS of moderately infected patients			Gamma shape = 1.88 rate = 0.25 mean = 7.52 days sd = 30.08	Fitted distributions to UMCU data from 2020

			Gamma	
LoS of severely infected patients			shape = 1.59 rate = 0.05 mean = 31.8 days sd = 636	Fitted distributions to UMCU data from 2020

*Values given are fixed in the simulations. Values in brackets were used in sensitivity analyses.
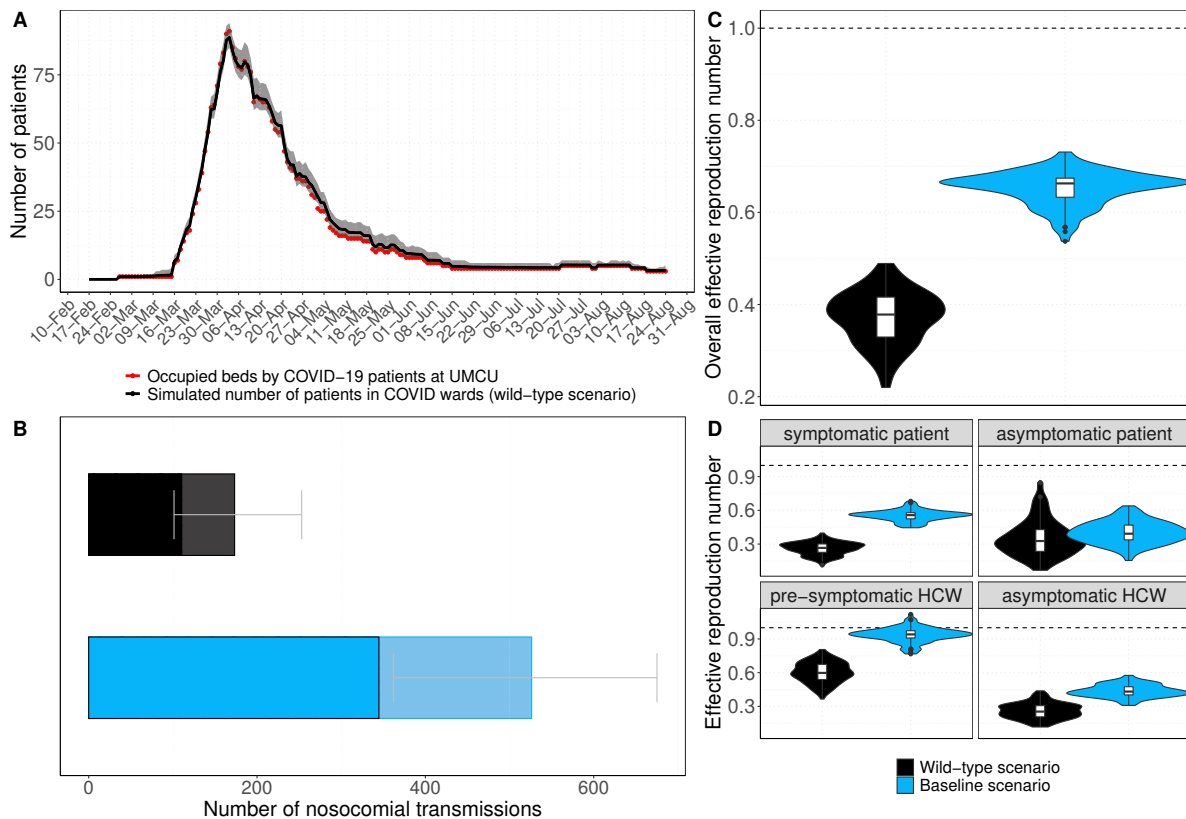
Figure 3: **Comparison of the scenarios with the wild-type and a more transmissible SARS-CoV-2 variant.** Both scenarios assume 90% effective PPE use in COVID wards. For the wild-type scenario (black), model simulations were performed with $R_S = 1.25$ (reproduction number of symptomatically infected individuals) and $R_A = 0.5$ (reproduction number of asymptomatically infected individuals). For the baseline scenario (blue), model simulations were performed with $R_S = 1.95$ and $R_A = 0.8$ (with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). Horizontal dashed lines represent a reproduction number of 1. Summary statistics were calculated for 100 simulations. (A) Simulated mean number of beds occupied by patients in COVID wards per day (black curve) and 95% uncertainty interval (grey shaded area). Red points represent real-life data on the daily number of beds occupied by COVID-19 patients at the UMCU between 27 February and 24 August 2020. (B) Number of nosocomial transmissions as predicted by the models. Full rectangular bar height represents the mean total number of noso-comial transmissions during the whole study period. Grey error bars represent 95% uncertainty intervals. Patients acquiring a SARS-CoV-2 nosocomial infection may be diagnosed in the hospital (due to symptom onset during hospital stay or detection by an intervention) or discharged to the community in a pre-symptomatic or asymptomatic state. Rectangular bars with black borders represent mean number of individuals (patients and HCWs) infected with SARS-CoV-2 and diagnosed in the hospital. Lighter rectangular bars represent the remaining mean number of patients discharged to community undiagnosed. (C) Violin and box plots of the overall effective reprduction numbers (R_E, for pre-/symptomatic and asymptomatic patients and HCWs combined) for the nosocomial spread in the wildtype and baseline scenario. (D) Violin and box plots of R_E for the nosocomial spread in the wildtype and baseline scenario (separate values for pre-/symptomatic and asymptomatic individuals). Since HCWs are assumed to immediately self-isolate upon symptom onset, the reproduction number is assigned to the pre-symptomatic state.
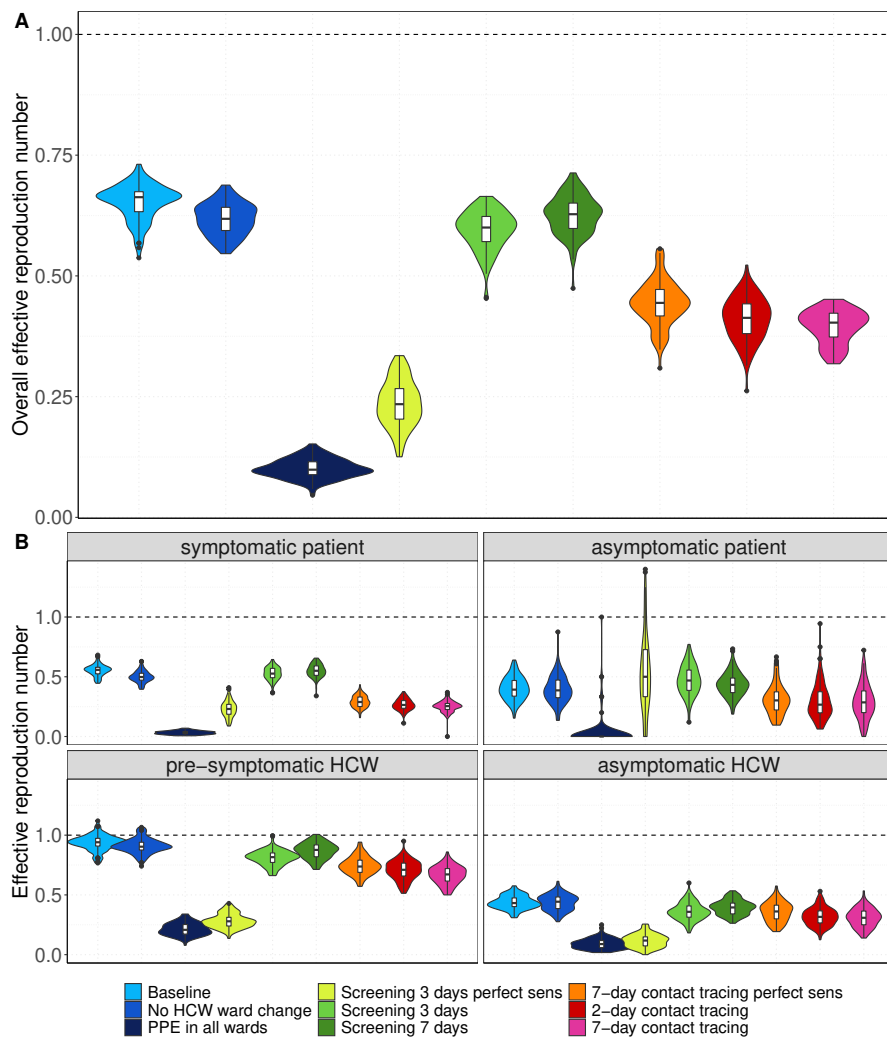
Figure 4: **Effective reproduction numbers for the nosocomial spread of the SARS-CoV-2 variant for each simulation scenario.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). Horizontal dashed lines represent a reproduction number of 1. Summary statistics were calculated for 100 simulations. (A) For each intervention scenario, violin and boxplots of the overall effective reproduction numbers (for pre-/symptomatic and asymptomatic patients and HCWs combined) are shown. (B) For each intervention scenario, violin and boxplots of the effective reproduction numbers for pre-/symptomatic and asymptomatic individuals are shown. Since HCWs are assumed to immediately self-isolate upon symptom onset, the reproduction number is assigned to the pre-symptomatic state. For screening every 3 days and 7-day contact tracing prior to symptom onset of SARS-CoV-2 infected HCWs, we considered two different test sensitivity scenarios: time-invariant perfect test sensitivity (perfect sens) and time-varying test sensitivity.

Figure 5: **Number of nosocomial transmissions of the SARS-CoV-2 variant for each simulation scenario.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). Summary statistics were calculated for 100 simulations. The full rectangular bar height represents the mean total number of nosocomial transmissions during the whole study period. The grey error bars represent the corresponding 95% uncertainty intervals. Patients that acquire a SARS-CoV-2 nosocomial infection may be diagnosed in the hospital (due to symptom onset during hospital stay or due to detection by an intervention) or discharged to the community in a pre-symptomatic or asymptomatic state. The rectangular bars with the black border represent the mean number of individuals (patients and HCWs) infected with SARS-CoV-2 and diagnosed in the hospital. The lighter rectangular bars represent the remaining mean number of patients discharged to community undiagnosed. For screening every 3 days and 7-day contact tracing, we considered two different test sensitivity scenarios: time-invariant perfect test sensitivity (perfect sens) and time-varying imperfect test sensitivity.
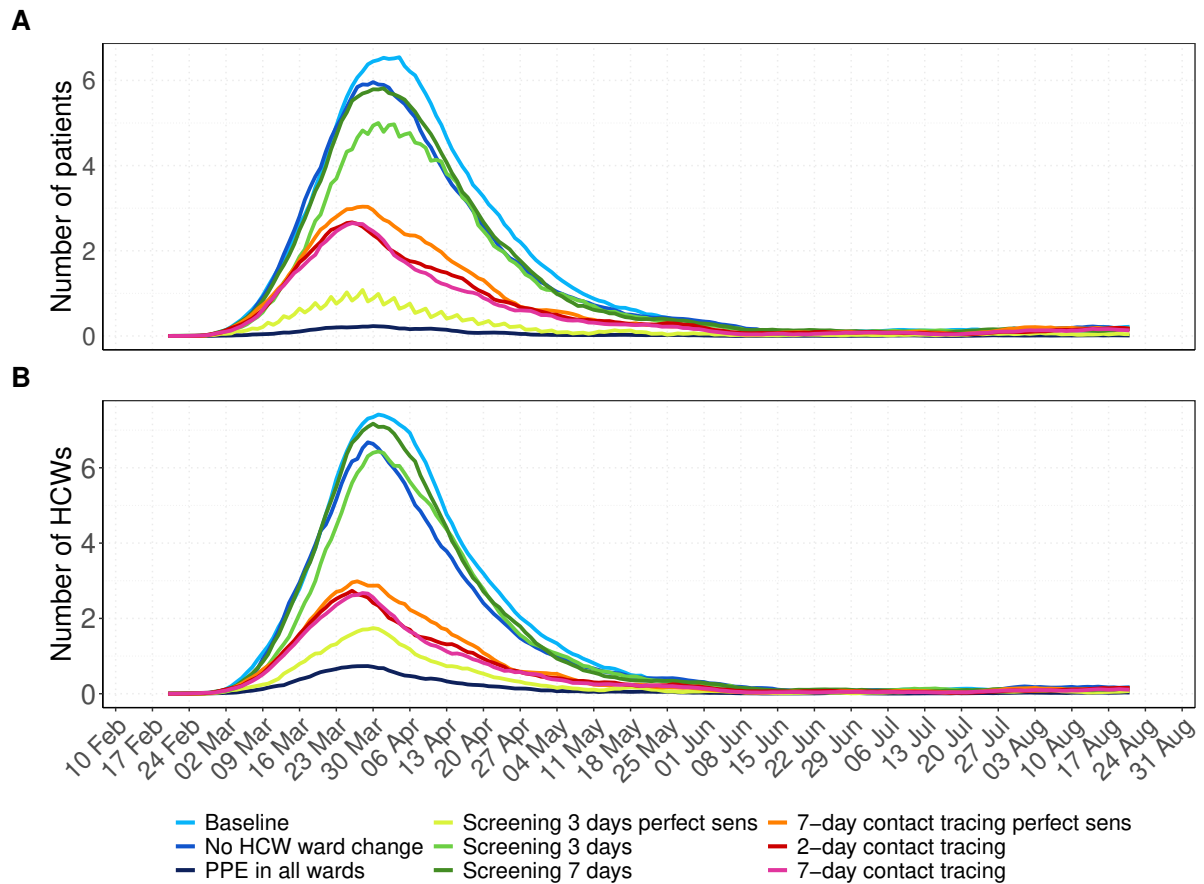
Figure 6: **Number of nosocomial infections among patients and HCWs over time for all simulation scenarios with the SARS-CoV-2 variant.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). For each scenario, the 7-day moving average of the mean prevalence (over 100 simulation runs) is shown. A) Number of hospital-acquired infections among patients. B) Number of hospital-acquired infections among HCWs. For screening every 3 days and contact tracing 7 days prior to symptom onset of SARS-CoV-2 infected HCWs, we considered two different test sensitivity scenarios: time-invariant perfect test sensitivity (perfect sens) and time-varying imperfect test sensitivity.
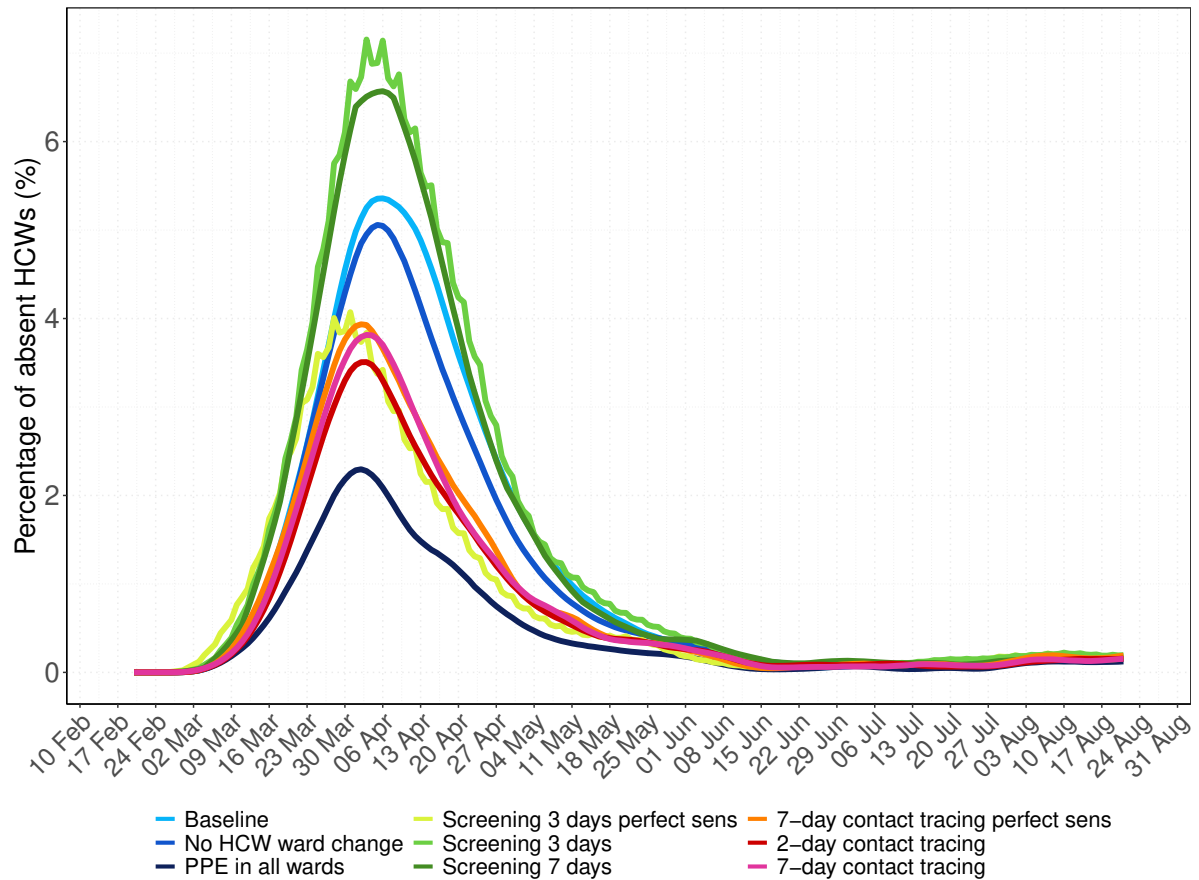
Figure 7: **Daily percentage of absent HCWs during the hospital epidemic for each simulation scenario.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). The 7-day moving average of the mean percentage (over 100 simulation runs) of HCWs absent from work due to symptom onset or a detected SARS-CoV-2 infection screening or contact tracing is shown. For screening every 3 days and contact tracing 7 days prior to symptom onset of SARS-CoV-2 infected HCWs, we considered two different test sensitivity scenarios: time-invariant perfect test sensitivity (perfect sens) and time-varying imperfect test sensitivity.
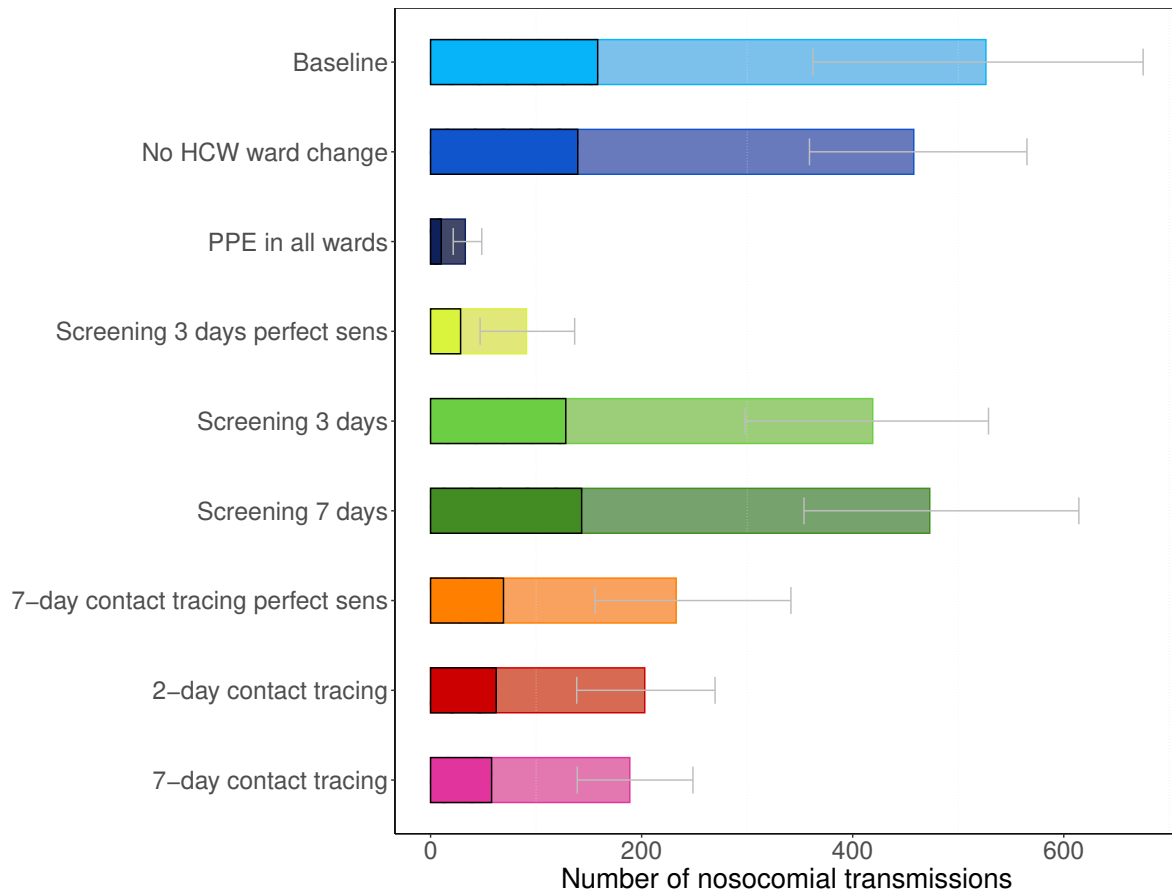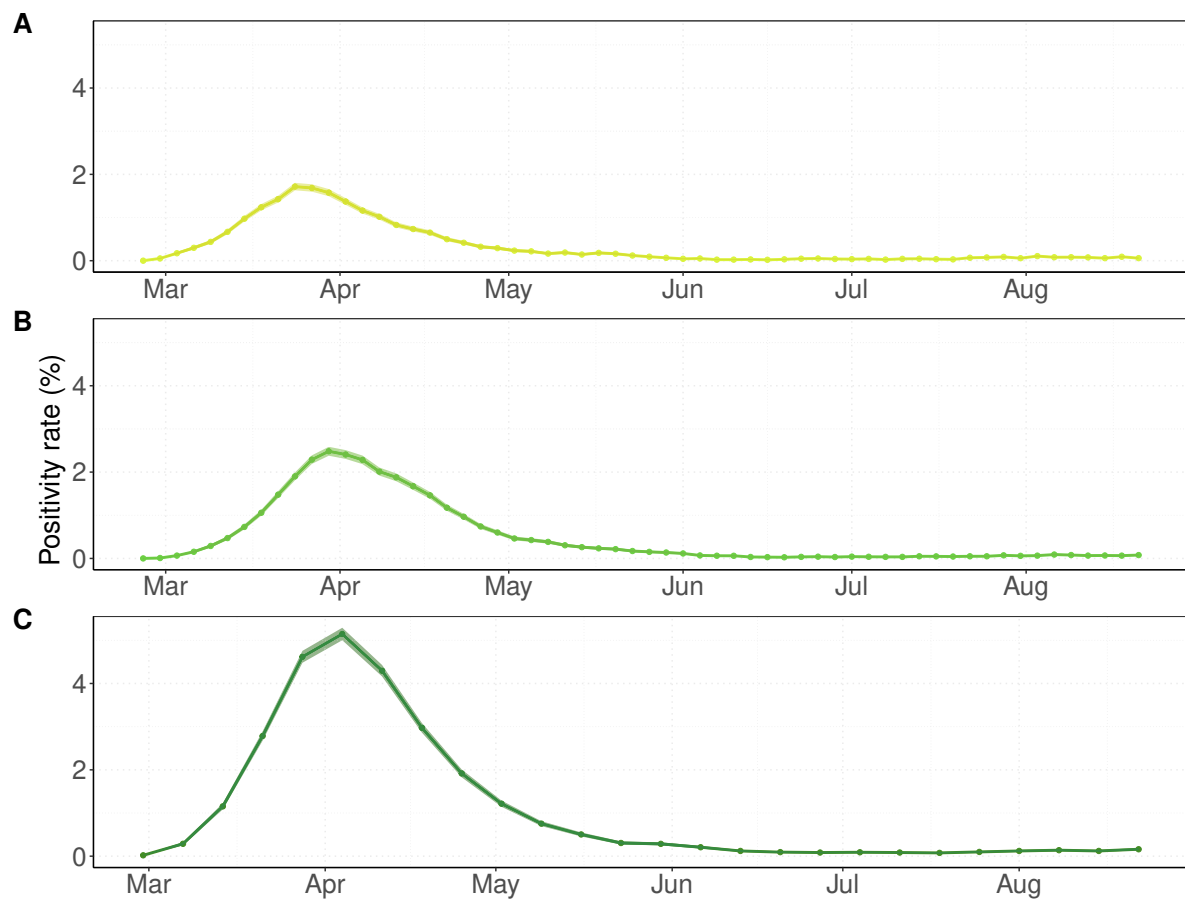
Figure 8: Positivity rates over time for screening interventions. Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). Positivity rates were calculated by the number of positive detected HCWs among the number of tested HCWs using data of all simulation runs combined (points). The shaded regions represent the 95% Bayesian beta-binomial credibility intervals. HCWs who developed symptoms prior to the day of testing were not included in the positivity rate as we assume that they were already correctly identified. (A) Screening every three days with time-invariant perfect test sensitivity. (B) Screening every three days with time-varying imperfect test sensitivity. (C) Screening every seven days with time-varying test sensitivity.
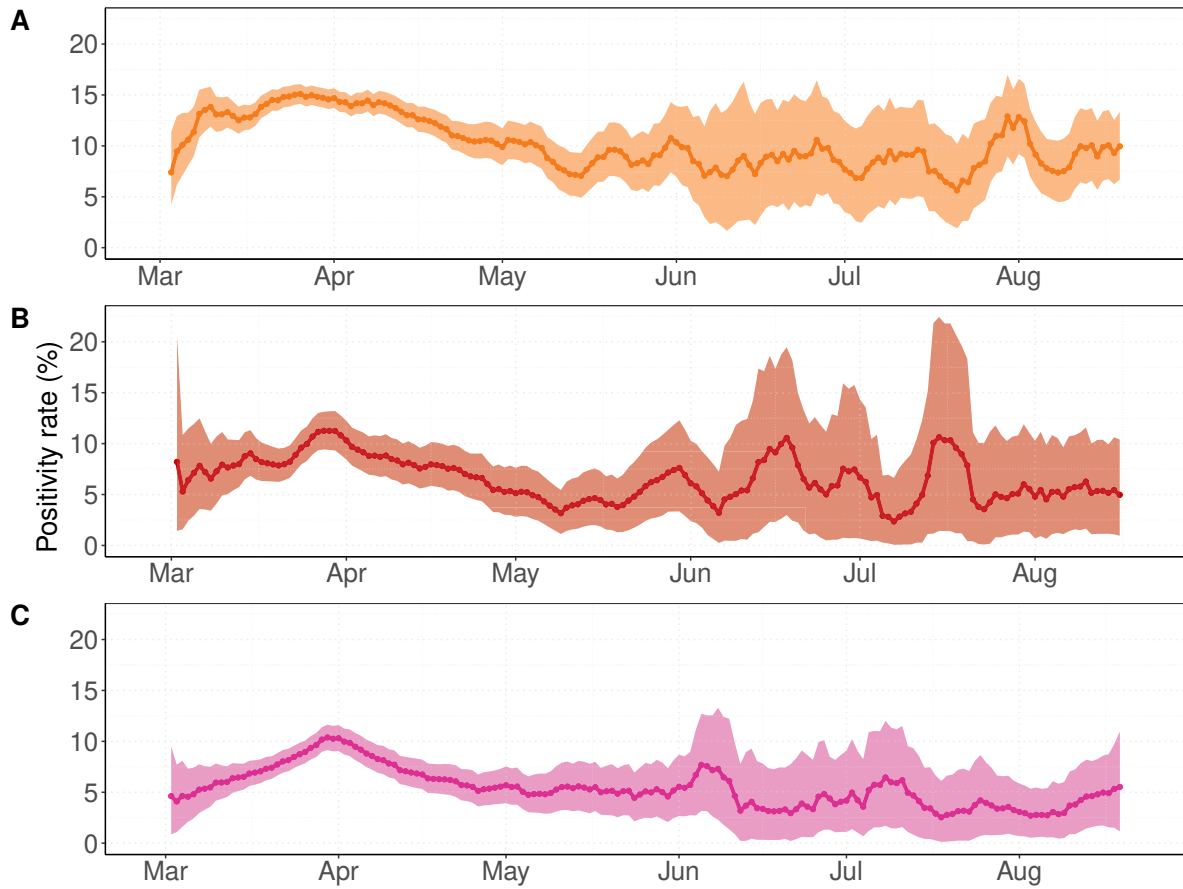
Figure 9: **Positivity rates over time for contact tracing interventions.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). The positivity rate is computed by the percentage of positive tested contacts among all traced contacts using data of all 100 simulation runs merged. Positivity rates are assigned to the day of symptom onset of the index case, i.e., HCW that developed symptoms due to a SARS-CoV-2 infection. Traced contacts who developed symptoms due to a SARS-CoV-2 infection are excluded from contact tracing as we assume that they are always correctly identified. The plot shows the 7-day moving average (colored line) and the 95% Bayesian beta-binomial confidence interval (shaded area). (A) Tracing contacts of symptomatically infected HCWs of the last two days before symptom onset using a diagnostic test with perfect test sensitivity. (B) Tracing contacts of symptomatically infected HCWs of the last two days before symptom onset with testing five days after contact with the index case assuming time-varying test sensitivity. (C) Tracing contacts of symptomatically infected HCWs of the last seven days before symptom onset with testing five days after contact with the index case assuming time-varying test sensitivity.

References

[1] Richterman, A, Meyerowitz, EA, and Cevik, M. *Hospital-Acquired SARS-CoV-2 Infection: Lessons for Public Health.* Dec. 2020. doi: `10.1001/jama.2020.21399`.

[2] Bielicki, JA, Duval, X, Gobat, N, et al. *Monitoring approaches for health-care workers during the COVID-19 pandemic.* English. Oct. 2020. doi: `10.1016/S1473-3099(20)30458-8`.

[3] Nguyen, LH, Drew, DA, Graham, MS, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. In: *The Lancet Public Health* 5.9 (2020), e475–e483. doi: `10.1016/S2468-2667(20)30164-X`.

[4] Gómez-Ochoa, SA, Franco, OH, Rojas, LZ, et al. COVID-19 in Health-Care Workers: A Living Systematic Review and Meta-Analysis of Prevalence, Risk Factors, Clinical Characteristics, and Outcomes. In: *American Journal of Epidemiology* 190.1 (Jan. 2020), pp. 161–175. doi: `10.1093/aje/kwaa191`.

[5] Sikkens, JJ, Buis, DTP, Peters, EJG, et al. Serologic Surveillance and Phylogenetic Analysis of SARS-CoV-2 Infection Among Hospital Health Care Workers. In: *JAMA Network Open* 4.7 (Jan. 2021), e2118554. doi: `10.1001/jamanetworkopen.2021.18554`.

[6] Eyre, DW, Lumley, SF, O'donnell, D, et al. Differential occupational risks to healthcare workers from SARS-CoV-2 observed during a prospective observational study. In: *eLife* 9 (Aug. 2020), pp. 1–37. doi: `10.7554/ELIFE.60675`.

[7] Smith, DR, Duval, A, Pouwels, KB, et al. Optimizing COVID-19 surveillance in long-term care facilities: a modelling study. In: *BMC Medicine* 18.1 (Dec. 2020), p. 386. doi: `10.1186/s12916-020-01866-6`.

[8] Evans, S, Agnew, E, Vynnycky, E, et al. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1829 (July 2021), p. 20200268. doi: `10.1098/rstb.2020.0268`.

[9] Qiu, X, Miller, JC, MacFadden, DR, et al. Evaluating the contributions of strategies to prevent SARS-CoV-2 transmission in the healthcare setting: a modelling study. In: *BMJ Open* 11.3 (Mar. 2021), e044644. doi: `10.1136/bmjopen-2020-044644`.

[10] Chin, ET, Huynh, BQ, Chapman, LAC, et al. Frequency of Routine Testing for Coronavirus Disease 2019 (COVID-19) in High-risk Healthcare Environments to Reduce Outbreaks. In: *Clinical Infectious Diseases* (Oct. 2020). doi: `10.1093/cid/ciaa1383`.

[11] Huang, Q, Mondal, A, Jiang, X, et al. SARS-CoV-2 transmission and control in a hospital setting: An individual-based modelling study. In: *Royal Society Open Science* 8.3 (Mar. 2021). doi: `10.1098/rsos.201895`.

[12] Mandić-Rajčević, S, Masci, F, Crespi, E, et al. Source and symptoms of COVID-19 among hospital workers in Milan. In: *Occupational Medicine* 70.9 (Dec. 2020), pp. 672–679. doi: `10.1093/occmed/kqaa201`.

[13] Paltansing, S, Sikkema, RS, Man, SJ de, et al. Transmission of SARS-CoV-2 among healthcare workers and patients in a teaching hospital in the Netherlands confirmed by whole-genome sequencing. English. In: *Journal of Hospital Infection* 110.0 (Feb. 2021), pp. 178–183. doi: `10.1016/j.jhin.2021.02.005`.

[14] Grassly, NC, Pons-Salort, M, Parker, EP, et al. Comparison of molecular testing strategies for COVID-19 control: a mathematical modelling study. English. In: *The Lancet Infectious Diseases* 0.0 (Aug. 2020). doi: `10.1016/S1473-3099(20)30630-7`.

[15] Lauer, SA, Grantz, KH, Bi, Q, et al. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. In: *Annals of Internal Medicine* 172.9 (Mar. 2020), pp. 577–582. doi: `10.7326/M20-0504`.

[16] Liu, Y, Yan, LM, Wan, L, et al. Viral dynamics in mild and severe cases of COVID-19. In: *The Lancet Infectious Diseases* 0.0 (2020). doi: `10.1016/S1473-3099(20)30232-2`.

[17] RIVM. *COVID-19 Aantallen gemeente per dag*. 2020. `https://geodata.rivm.nl/covid-19/COVID-19_aantallen_gemeente_per_dag.csv`.

[18] De Bruin, J. *Number of diagnoses with coronavirus disease (COVID-19) in The Netherlands (Version v2020.3.15) [data-municipal]*. `http://doi.org/10.5281/zenodo.3711575`.

[19] Buitrago-Garcia, D, Egli-Gany, D, Counotte, MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: A living systematic review and meta-analysis. en. In: *PLoS Medicine* 17.9 (Sept. 2020), e1003346. doi: `10.1371/journal.pmed.1003346`.

[20] Davies, NG, Abbott, S, Barnard, RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. In: *Science (New York, N.Y.)* (Mar. 2021). doi: `10.1126/science.abg3055`.

[21] Baker, MA, Fiumara, K, Rhee, C, et al. Low Risk of Coronavirus Disease 2019 (COVID-19) Among Patients Exposed to Infected Healthcare Workers. In: *Clinical Infectious Diseases* ciaa1269 (Aug. 2020). doi: `10.1093/cid/ciaa1269`.

[22] Seidelman, JL, Lewis, SS, Advani, SD, et al. *Universal masking is an effective strategy to flatten the severe acute respiratory coronavirus virus 2 (SARS-CoV-2) healthcare worker epidemiologic curve*. en. Dec. 2020. doi: `10.1017/ice.2020.313`.

[23] Barrett, ES, Horton, DB, Roy, J, et al. Prevalence of SARS-CoV-2 infection in previously undiagnosed health care workers in New Jersey, at the onset of the U.S. COVID-19 pandemic. In: *BMC Infectious Diseases* 20.1 (Nov. 2020), p. 853. doi: `10.1186/s12879-020-05587-2`.

[24] Rhee, C, Baker, M, Vaidya, V, et al. Incidence of Nosocomial COVID-19 in Patients Hospitalized at a Large US Academic Medical Center. en. In: *JAMA network open* 3.9 (Sept. 2020), e2020498. doi: `10.1001/jamanetworkopen.2020.20498`.

[25] Scarpino, SV, Allard, A, and Hébert-Dufresne, L. The effect of a prudent adaptive behaviour on disease transmission. en. In: *Nature Physics* 12.11 (Nov. 2016), pp. 1042–1046. doi: `10.1038/nphys3832`.

[26] Ong, BS, Chen, M, Lee, V, et al. "An individual-based model of influenza in nosocomial environments". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 5101 LNCS. PART 1. Springer, Berlin, Heidelberg, 2008, pp. 590–599. doi: `10.1007/978-3-540-69384-0{\_}64`.

[27] Wu, JT, Leung, K, and Leung, GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. In: *The Lancet* 395.10225 (2020), pp. 689–697.

doi: 10.1016/S0140-6736(20)30260-9.

[28] Suzuki, T, Hayakawa, K, Ainai, A, et al. Effectiveness of personal protective equipment in preventing severe acute respiratory syndrome coronavirus 2 infection among healthcare workers. In: *Journal of Infection and Chemotherapy* 27.1 (Jan. 2021), pp. 120–122. doi: 10.1016/j.jiac.2020.09.006.

[29] Qian, Y, Willeke, K, Grinshpun, SA, et al. Performance of N95 respirators: Filtration efficiency for airborne microbial and inert particles. eng. In: *American Industrial Hygiene Association Journal* 59.2 (Feb. 1998), pp. 128–132. doi: 10.1080/15428119891010389.

[30] Bessesen, MT, Savor-Price, C, Simberkoff, M, et al. *N95 respirators or surgical masks to protect healthcare workers against respiratory infections: Are we there yet?* en. May 2013. doi: 10.1164/rccm.201303-0581ED.

[31] RIVM. *Covid-19 besmettelijke personen per dag.* https://data.rivm.nl/covid-19/COVID-19_prevalentie.json.

[32] CBS. *Ziekenhuisopnamen; kerncijfers; geslacht, leeftijd, regio, 1981-2012.* Jan. 2019. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/71857ned/table?ts=1517582466533.

[33] De Bruin, J. *Number of diagnoses with coronavirus disease (COVID-19) in The Netherlands (Version v2020.3.15) [data-age].* http://doi.org/10.5281/zenodo.3711575.

[34] Woloshin, S, Patel, N, and Kesselheim, AS. False Negative Tests for SARS-CoV-2 Infection - Challenges and Implications. In: *The New England journal of medicine* 383.6 (Aug. 2020), e38. doi: 10.1056/NEJMp2015897.

[35] Bi, Q, Wu, Y, Mei, S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. In: *The Lancet Infectious Diseases* 20.8 (Apr. 2020), pp. 911–919. doi: 10.1016/S1473-3099(20)30287-5.

[36] Ferretti, L, Wymant, C, Kendall, M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. In: *Science* 368.6491 (Mar. 2020), eabb6936. doi: 10.1126/science.abb6936.

[37] He, X, Lau, EH, Wu, P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. en. In: *Nature Medicine* 26.5 (May 2020), pp. 672–675. doi: 10.1038/s41591-020-0869-5.

[38] Bernard, H, Fischer, R, Mikolajczyk, RT, et al. Nurses' contacts and potential for infectious disease transmission. In: *Emerging Infectious Diseases* 15.9 (Sept. 2009), pp. 1438–1444. doi: 10.3201/eid1509.081475.

[39] Masad, D and Kazil, J. "MESA: an agent-based modeling framework". In: Proceedings of the 14th Python in Science Conference, 2015. https://github.com/projectmesa/mesa.

[40] Thibon, P, Breton, P, Mouet, A, et al. Healthcare associated coronavirus disease 2019 among health care workers in Normandy, France: a multi-center study. en. In: *Infection Prevention in Practice* 3.1 (Mar. 2021), p. 100109. doi: 10.1016/j.infpip.2020.100109.

[41] Pham, TM, Tahir, H, Wijgert, JH van de, et al. Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study. In: *BMC Medicine* 19.1 (Dec. 2021), pp. 1–16. doi: 10.1186/s12916-021-02060-y.

Supplementary material

Table of Contents

Additional File 1: Supplementary methods

We simulated nosocomial COVID-19 epidemics using an agent-based model coded in Python. The code is available from:
`https://github.com/htahir2/covid_intra-hospital_model.git`.

First, we fitted the model to real-life data from the University Medical Center Utrecht (UMCU) during the period February-August 2020. Next, we evaluated the impact of various intervention strategies aimed at healthcare workers (HCWs) on the nosocomial spread of a more transmissible SARS-CoV-2 variant (e.g., B.1.1.7) in a hospital.

We first outline the data used to inform and parameterize our model. We continue by explaining the details of our agent-based model, the transmission model, and the underlying assumptions. We further describe the intervention strategies implemented in our model, the considered outcome measures, and the results of our sensitivity analyses. Lastly, we elaborate on the algorithm of our model.

I. Data

We used data from the University Medical Center Utrecht (UMCU), The Netherlands, and data provided by the National Institute for Public Health and the Environment (RIVM), The Netherlands, during the first wave of the SARS-CoV-2 epidemic to inform and parameterize our model.

Hospitalization data

We used unlinked anonymized hospitalization data of patients in COVID wards at the UMCU between 27 February and 24 August 2020. The data set comprises 167 admissions of which 82 patients were admitted to an intensive care unit, and 85 patients were admitted to regular wards. Based on information of the infection control department of the UMCU, we assumed that 95% of those admissions were COVID-19 admissions, leaving 5% of the patients admitted for non-COVID reasons but later diagnosed with a SARS-CoV-2 infection in the hospital. The number of admissions per day are shown in Figure S1. The data further comprises discharge dates of the respective patients. We used the resulting number of beds occupied by COVID-19 patients to fit

the reproduction numbers in our model. The data is available from:
`https://github.com/htahir2/covid_intra-hospital_model/blob/main/data/covid_`
`patient_admissions_los_UMCU.csv`.

Length of stay distributions

We calculated the respective length of stay (LoS) from admission and discharge times for each of the COVID-19 patients at UMCU. Data on patient admissions to the UMCU prior to the COVID-19 pandemic (2014-2017) were used to estimate the length of stay distributions for non-COVID admissions to the hospital. We fitted probability distributions to the length of stay data for admissions to regular wards and ICUs, separately for COVID or for non-COVID related admissions. We considered exponential, log-normal, gamma, and Weibull distribution and chose the best fit by visual inspection of the empirical vs theoretical densities, the Q-Q plot, and the P-P plot. The length of stay data and fitted distributions are shown in Figure S2. The respective parameters can be found in Table S1.

Importation from community

We assume that 40 patients are admitted to the hospital for reasons unrelated to COVID-19 per day in the time period 27 February to 24 August 2020 (Table S1). We based this number on UMCU admission data in the time period 2014-2017 and the assumption that admissions decreased by 50% during the first wave of the COVID-19 epidemic. Those admitted patients might be infected with SARS-CoV-2 due to transmissions in the community. We further assume that HCWs go home after each daily shift and therefore may acquire infection in the community as well. They may be in their pre-symptomatic phase or asymptomatically infected with SARS-CoV-2 when they arrive at work in the hospital. These patients and HCWs do not experience any symptoms (yet), and therefore do not know that they are infected. We approximate the probability of being infected in the community for non-COVID patient admissions and HCWs arriving at work as follows: We used data on the number of infectious people in the Netherlands estimated by the National Institute for Public Health and the Environment (RIVM) from 17 February till 24 August 2020 [31]. They used data from serological surveys in the Netherlands and related these to numbers of hospitalized cases (stratified by age group), leading to the number of "actual" infections per hospitalized case. They thereby included all infections that led to

an immunological response, not only those that were detected in real time by PCR testing. Note that this method might be less reliable when the number of hospitalizations is low. This estimated number of infectious individuals includes hospitalized COVID-19 patients as well as individuals that are isolated at home (e.g., due to detection in the community via testing or contact tracing). To roughly account for this, we subtracted the total number of reported cases in the province of Utrecht from the estimated number of infectious individuals (RIVM estimate described before). We hereby assume that all individuals in Utrecht are eligible for admission at the UMCU. We additionally used publicly available age-specific hospitalization rates of the Netherlands of 2012 and age-specific COVID-19 incidence rates in the Netherlands to scale the daily probability of being infected in the community for non-COVID patient admissions and HCWs arriving at work [18, 32, 33]. For HCWs, we only used age-specific COVID-19 incidence rates for age-groups between 20 and 65 years. Since age-specific prevalence values are not available to date, the previous calculation is based on the assumption that the distribution of age groups is roughly the same for incidence and prevalence. Furthermore, we assumed a catchment population size of 100,000 people for the hospital.

II. Model

Environment

We modelled a typical (Dutch) hospital comprising 28 wards which are divided as follows

- COVID ICU (4) with 17 beds each

- Normal (non-COVID) ICU (1) with 12 beds

- COVID ward (4) with 3×23 beds and one with 22 beds

- Normal (non-COVID) ward (19) with $2 \times 20, 4 \times 19, 13 \times 18$ beds

The total number of beds in the hospital is 521. The numbers are approximated in accordance to the number of beds and ward distribution of the UMCU (for patients who stayed at least one day at UMCU).

Agent-types

There are three different agents involved in the transmission process within the hospital of our model: Patients (non-COVID and COVID) and health-care workers (HCWs),

separated into nurses, and doctors. Visitors or ancillary workers are excluded from the model. Patients are assumed to occupy a hospital bed in a single room. This assumption is suitable for a setting where the transmission is mainly driven by HCWs as vectors and with no direct patient-to-patient transmission. HCWs have a number of duty shifts per day. We assume that they meet patients in a number of rounds per shift (see Table S1) and that HCWs meet other HCWs in the common staff room of each ward. The ratio between HCWs and patients and the time HCWs spend with a patient are ward specific and assumed. The number of HCW duty shifts per day, and the number of rounds per shift are independent from the ward and assumed. The respective parameters can be found in Table S1.

Disease progression

The disease progression of an infection with SARS-CoV-2 was modelled using a Susceptible-Exposed-Infectious-Recovered (SEIR) model and is shown in Figure 1c of the main text. Individuals who have not been infected with SARS-CoV-2 are susceptible (S), and may transition to being exposed (E) upon contact with an infected individual. A proportion $(1 - P_A)$ of infected individuals develop symptoms. We based the incubation period (time between infection and appearance of symptoms) on a Gamma distribution with mean 5.5 days as described by Lauer and colleagues [15]. Symptomatically infected individuals may develop moderate symptoms (I_M) or severe symptoms (I_S). All infected individuals will eventually recover and become immune (I_R). Asymptomatically infected (I_A) are assumed to recover after 14 days. We assume that moderately and severely infected patients recover after 14 and 35 days, respectively [16]. We assumed that symptomatically infected HCWs are perfectly isolated at home for seven days immediately upon developing symptoms, after which they return recovered to work. Based on a meta-analysis by Buitraga-Garcia and colleagues, we assumed the asymptomatic proportion of COVID-19 infections among patients to be 20% and the proportion of asymptomatic infections among HCWs to be 31% (see also Table S1) [19]. We used their overall estimate of the proportion of asymptomatic infected individuals for the patient population in our model, and their estimates obtained from studies with screened individuals for the HCW population in our model.

Hospital admissions

Patients can be hospitalized either for non-COVID reasons to normal wards and ICUs, or with moderate or severe COVID-19 symptoms to COVID wards or COVID-ICUs. The length of stay of a patient differs according to these four categories. Probability distributions were fitted to length of stay data of patients admitted to the UMCU. The data and fitted distributions are shown in Figure S1-S2. The respective parameters can be found in Table S1. Patients admitted to normal wards and ICUs who are later detected with a SARS-CoV-2 infection are immediately transferred to COVID wards and ICUs upon diagnosis. If they develop severe symptoms, their length of stay is prolonged according to the length of stay distribution of admitted severe COVID-19 patients.

Accuracy of the diagnostic test

In our model, patients and HCWs are assumed to be tested using reverse transcriptase polymerase chain reaction (RT-PCR) either when being screened or after being identified as a contact of a symptomatic infected individual in contact tracing. These diagnostic tests can be inaccurate either because of a false positive or a false negative result. The latter are considered to be more consequential with a potential high impact on onward transmission due to undetected cases. It has been documented that the sensitivity of the PCR test varies with time from exposure and symptom onset [34]. We assumed a time-varying imperfect sensitivity of the diagnostic test (Figure S3) based on the results reported in Grassly and colleagues [14]. These authors used published data from three meta-analyses of the test sensitivity over time since symptom onset. They assume the pre-symptomatic sensitivity to be proportional to the infectiousness curve such that the estimate on day 5 matched the empirical data from the day of symptom onset. We performed a sensitivity analysis assuming 1) a 15% lower sensitivity, and 2) a test sensitivity that stays at its peak value after reaching the maximum (see Figure S3). Finally, we assessed the impact of perfect sensitivity of 100% which stays constant over time on the results of our model. Throughout the simulations, we assume test sensitivity to be the same for symptomatic and asymptomatic infections, and we assume a specificity of 100%.

Infectiousness

We use a time-varying infectiousness profile following Grassly and colleagues and shown in Figure 1C of the main text [14]. Infectiousness is assumed to vary over time since infection and to follow a Weibull distribution, with a mean of 6 days [14]. The average duration of the infectious period is therefore assumed to be 6 days. This approximation is consistent with published estimates of the serial interval for SARS-CoV-2 [35–37]. We denote infectiousness over time since infection τ by $\beta(\tau)$. It is the mean rate at which an individual infects others at time after its time of infection. We use the infectiousness profile for calculating the probability of transmission from an infectious to a susceptible individual (see below). The reproduction number R (average number of secondary cases caused by an infected individual) is given by integrating $\beta(\tau)$ over time since infection $R = \int_0^\infty \beta(\tau) d\tau$. The generation time distribution $\omega(\tau)$ is given by unit normalisation such that $\omega(\tau) = \beta(\tau)/R$. Assuming the mean generation time to be equivalent with the observed mean serial interval, we calculate the infectiousness profile by $\beta(\tau) = \omega(\tau)R$. We assumed the infectiousness over time since infection to differ between asymptomatic and symptomatic infected individuals, defined by $\beta_A(\tau)$ and $\beta_S(\tau)$, respectively. Then $\beta(\tau)$ can be decomposed into

$$\beta(\tau) = P_A\beta_A(\tau) + (1 - P_A)\beta_S(\tau)$$

where P_A represents the proportion of asymptomatic infections. Asymptomatic individuals are assumed to have an infectiousness proportional to that of symptomatic individuals, i.e., $\beta_A = x_A \cdot \beta_S, x_A \leq 1$. Integrating over each of the two terms leads to the respective contribution to the overall reproduction number:

$$R = R_A + R_S$$

First, we chose the basic reproduction numbers R_S and R_A (values are given in Table S1) such that the numbers of occupied beds by COVID-19 patients predicted by our model were in good agreement with real-life UMCU data on the number of COVID-19 patients at UMCU during the first epidemic wave (Figure 1 and Figure 2A in the main text). These reproduction numbers incorporated the effects of typical (but not COVID-specific) infection prevention measures in the hospital. We will refer to the model

parameterized with these reproduction numbers as the wild-type scenario. This scenario also assumed that HCWs use 90% effective PPE in COVID wards and isolate at home immediately upon symptom onset for seven days, after which they return recovered to work. Next, we introduced a more transmissible SARS-CoV-2 variant into the hospital, keeping all other parameters – including PPE use in COVID wards and self-isolation after symptom-onset – the same. Based on recent results for B.1.1.7, we assumed a 56% increase in transmissibility [20]. We will refer to the model parameterized with these higher reproduction numbers as our baseline scenario. Various intervention scenarios were compared to this baseline scenario.

Transmission

Transmission events can occur between susceptible patients and HCWs, or between (asymptomatic or pre-symptomatic) HCWs. Thus, we assumed that there is no direct transmission between patients. Only HCWs in their pre-symptomatic stage, or HCWs who are asymptomatically infected, contribute to transmission, since we assume that HCWs are perfectly isolated at home for seven days immediately upon developing symptoms. Upon a contact between two individuals, a transmission may take place between an infected and a susceptible individual. The probability of transmission is dependent on the current infectiousness of the infected individual. If $\beta(\tau)$ is the infectiousness of the infected individual at time since infection, the average probability of transmission per contact with a susceptible person is given by

$$A(\tau) = \frac{\beta(\tau)}{c}$$

where c is the average contact rate of the individual which can be determined by computing the largest eigenvalue of the respective contact matrix

$$\begin{bmatrix} c_{n,n} & c_{n,p} & c_{n,d} \\ c_{p,n} & 0 & c_{p,d} \\ c_{d,n} & c_{d,p} & c_{d,d} \end{bmatrix}$$

where $c_{i,j}$ is the contact rate of an individual of type i with an individual of type j (see Table S1). Let N_n, N_{hc}, and N_p be the average number of nurses, doctors, and patients in the hospital population, respectively. Since the total number of contacts of individuals

of type i has to be the same as the total number of contacts of an individual of type j, $c_{n,p}N_n = c_{p,n}N_p$, $c_{d,p}N_p = c_{p,d}N_d$, and $c_{d,n}\ N_n = c_{n,d}N_d$ the contact matrix is given by

$$
\begin{bmatrix}
c_{n,n} & c_{n,p} & c_{n,d} \\
c_{n,p}\frac{N_p}{N_n} & 0 & c_{p,d} \\
c_{n,d}\frac{N_d}{N_n} & c_{p,d}\frac{N_d}{N_p} & c_{d,d}
\end{bmatrix}
$$

The values of the contact rates in the matrix are based on HCW to patient ratios and the number of rounds per shift of HCWs and were estimated once from our simulations assuming 100% patient occupancy in the hospital. These estimated parameters were later used for all the simulation scenarios. The respective values can be found in Additional File 1: Table S1. We compared these values to a prospective contact survey of nurses in five hospitals in the German federal state of Bavaria conducted by Bernard and colleagues [38]. The authors reported a median work-related contact rate of $c_n = 34$ of nurses during 24 hours. Most work-related contacts were with patients (51%) or other staff member/other persons (49% = 40% + 9%). Thus, nurses meet approximately 17.3 patients and 16.7 other staff members per 24 hours. The contact rate of nurses with patients per duty shift from our simulations ($c_{n,d} = 19.07$) is similar to the reported value by Bernard and colleagues. The contact rates between hospital staff are lower in our simulation than reported in the contact survey (see values in Additional File 1: Table S1) and based on our assumption that contacts between hospital staff decreased during the first wave of the COVID-19 pandemic in the Netherlands.

Time of infection

For individuals infected in the community, the time of infection is unknown. For asymptomatic individuals, we assume an infectious period of 14 days and draw the infection time uniformly from 0 to 14 days prior to admission. For individuals that will develop symptoms, we draw an incubation period t_{inc} from the respective distribution (see Table S1) and then draw the infection time uniformly from 0 to t_{inc} prior to admission. Note that this approach neglects the fact that in an early stage of an outbreak when the epidemic grows at an exponential rate, it is likely that there are many more recently infected individuals.

III. Calibration of parameters to data

In such a multi-dimensional parameter space, it is possible that there are multiple sets of parameters that would produce the fit. The parameter set we presented in this work is likely not the only one that could produce the presented fit (Figure 2A in the main text). For many parameters we have knowledge from empirical observations about the likely values for those parameters (e.g., incubation period, generation time, . . .). This already considerably limits the parameter region of the model. We then tried to identify the parameters that would most likely change the fit to the data. For example, when we extended the length of stay of hospital-acquired infected patients with mild symptoms, the simulation output did not fit to the data anymore ("width" of the curve did not fit). This confirmed information we received from UMC Utrecht that in the Netherlands, the LoS of patients who do not require intensive care would not be extended (but patients would be sent home for isolation). Similarly, we noticed that when we changed the reproduction number that this had a big impact on the "height" of the curve. Thus, we calibrated the basic reproduction numbers such that the height of the simulation curve matched the data (and the width matched already with the remaining parameters given in Table 1 in the main text and Additional File 1: Table S1). Fine-tuning could yield a better fit but this is out of scope of this work.

The following parameters were involved in the calibration of the parameters to the observed data on occupied beds of COVID-19 patients at UMC Utrecht:

- Reproduction numbers for asymptomatically and symptomatically infected individuals

- Length of stay adjustment for patients who acquired a SARS-CoV-2 infection in the hospital

- Isolation period of HCWs

IV. Infection control interventions

We used the model to evaluate the effect of several interventions aimed at HCWs on the hospital epidemic using data from the first wave of the epidemic in the Netherlands but assuming the introduction of a SARS-CoV-2 variant with higher transmissibility in

the hospital. As such, our model results show the impact of the interventions on the nosocomial spread of a new variant.

Throughout the simulations, we assume that HCWs use 90% effective PPE in COVID wards and isolate at home immediately upon symptom onset for seven days, after which they return recovered to work. Furthermore, we assume that there is no delay between testing and receiving the COVID-19 test result. This assumption is in particular reasonable for hospital staff tested in the hospital, as they receive their test result within hours (UMCU) and have to self-isolate until they receive the result. Thus, it can be assumed that they do not contribute to the transmission of the virus while waiting for the test result.

Baseline scenario

We assumed that HCWs used personal protective equipment (PPE) while working in COVID wards. PPE reduces the transfer of droplets or other body fluids onto HCWs' skin and clothing or directly onto the mucous membranes of the eye or nasopharynx. We define PPE efficacy as the percentage reduction of droplet transfer. Furthermore, we define the effectiveness of PPE as the reduction of infectiousness by a factor upon each contact between an infected and susceptible individual. This reduction factor includes PPE efficacy as well as adherence to adequate PPE use. In our baseline scenario, we assumed that all PPE use was 90% effective. We assumed HCWs do not use PPE when meeting each other in the common room and that per day 95% of the HCWs work in the same ward as during their previous shift.

Intervention: PPE in all hospital wards

In this intervention scenario, we assumed that all HCWs wore 90% effective PPE in all (non-COVID and COVID) wards. Note that no PPE is worn when HCWs meet each other. We performed sensitivity analyses assuming PPE effectiveness of 50% and 70%.

Intervention: HCW cohorting (no HCW ward change)

In this intervention scenario, we restricted HCWs to work only in specific wards and did not allow any ward change.

Intervention: Regular HCW screening

All HCWs were tested for SARS-CoV-2 either with a) a test with perfect sensitivity every three days, or a test with time-varying sensitivity, b) every three days, or c) every seven days. If tested positive, HCWs were assumed to immediately self-isolate at home for seven days.

Intervention: HCW contact tracing

If a HCW developed symptomatic SARS-CoV-2 infection, all contacts in the hospital during a time window of either two or seven days before symptom onset were traced and tested. We will refer to these scenarios as 2-day Contact tracing and 7-day contact tracing. For 2-day contact tracing, contacts were always tested assuming a time-varying test sensitivity. For 7-day contact tracing, we distinguished between perfect and time-varying sensitivity sub-scenarios. In the perfect sensitivity sub-scenario, contacts were instantaneously tested on the day of symptom onset of the index (the HCW). In the time-varying test sensitivity sub-scenario, the test was performed on the day of symptom onset if the contact with the index was more than five days ago. Otherwise, it was performed on day five after the contact. Exposed HCWs awaiting tests were assumed to wear PPE during contact with any patient and with other HCWs. In case of a positive test, patients were moved to a COVID ward while infected HCWs were sent home for self-isolation for seven days and replaced by susceptible HCW. We did not model any absences of HCWs with disease symptoms caused by other respiratory pathogens.

Outcome measures

We calculated the effective reproduction number R_E (average number of secondary cases caused by an infected individual) to evaluate an intervention's effectiveness in suppressing outbreak expansion in the hospital. We approximated R_E for an average individual (patients and HCWs combined) in the hospital (overall R_E) from our simulations by calculating the average number of secondary cases by an infected individual in our model. We further stratified this number by patients, HCWs, and symptom status. The reproduction numbers of patients were calculated for those who will eventually develop symptoms (R_S^{pat}) and those who will remain without symptoms (R_A^{pat}). Since HCWs were assumed to immediately self-isolate upon symptom onset, we calculated R during pre-symptomatic

(R_S^{hcw}) and asymptomatic states $(_A^{\text{hcw}})$ only. In order to evaluate the maximum demand on the hospital capacity, we considered the total number of hospital-acquired infections among patients and HCWs over time. In addition, we computed the proportion of absent HCWs due to self-isolation (because of symptom onset or detection via screening or contact tracing) over time. We assessed the efficiency of screening and contact tracing interventions with respect to their positivity rates (proportion of detected infected individuals among tested individuals). We did not include individuals that developed symptoms prior to being tested in the positivity rate calculations since those were already detected and isolated in our model. We determined the proportion of transmissions attributed to different transmission routes (HCW-to-HCW, HCW-to-patient, patient-to-HCW). For every scenario, we calculated the mean and 95% percentiles over 100 simulation runs (95% uncertainty interval). We calculated positivity rates over time merging data from all simulation runs and computed 95% Bayesian beta-binomial credibility intervals.

V. Implementation of the model

The model was built using Python (version 3.6) using the library Mesa which is an open source agent-based modelling framework [39]. The code is available from: `https://github.com/htahir2/covid_intra-hospital_model.git`.
An overview of all processes in the implemented model is shown in Figure S4. The individual processes are described below and in flowcharts in Figure S5-S12.

Initialization

The model is initialized with non-COVID patients admitted to normal ICU and normal wards. We assume 50% of the rooms (beds) in the normal ICU and normal wards are free at the moment of model initialization. There are three duty shifts in a day and in these duty shifts, HCWs (nurses and doctors) are assigned to all wards in the hospitals. We assume that the number of HCWs in the hospital remains constant throughout the simulation period. Every patient agent has its own unique characteristics such as ID, ward and room number, LoS, and disease state. Depending on the ward (normal ICU or normal ward), every patient is assigned a LoS from the given distributions at the time of admission (Table S1). HCWs also have unique characteristics such as ID, ward to which HCW is assigned, duty shift in which HCW is working, HCW wearing PPE or not, time being absent from work due to quarantine, and disease state. A patient or a HCW can

only be in one of the following disease states at a particular moment: susceptible, exposed, moderate symptomatic, severe symptomatic, asymptomatic, or recovered. However, at the model initialization, all patients and HCWs are in susceptible state.

Study period

The simulation is run for 239 days in total, with an initial period of 59 days to get a stable non-COVID patient population in the hospital. The first symptomatic COVID-19 patient is admitted to the hospital on day 60.

Scheduling

A time step in our agent-based model represents 10 minutes. The following processes occur during the run time:

New patient arrival

Given an average daily patient arrival rate for the UMCU (Table S1), patients arrive at the hospital following a Poisson process and are randomly admitted to normal ICU and normal wards. The majority of these new daily patients are in a susceptible state but as mentioned earlier in the section "Importation from community", we use a community-prevalence-dependent, age-specific importation rate of exposed and asymptomatic patients into the hospital. Therefore, some patients from the daily new patient arrivals come in an exposed or asymptomatic state. Since the disease status of such patients is not known at the time of admission, they are admitted to normal ICU or normal wards. Depending on the patient ward (normal ICU or normal ward), the LOS is drawn from the appropriate distributions (Table S1). The first symptomatic COVID-19 patient is admitted to the hospital on day 60 (based on UMCU COVID-19 admission data). Known symptomatic patients are admitted to either COVID wards or COVID ICUs depending on the severity of their symptoms. Moderate symptomatic patients are admitted to COVID wards, whereas severe symptomatic patients are admitted to COVID ICUs. For symptomatic COVID patients admitted to the hospital, LOS of the individual patient is sampled directly from the UMCU data.

Patient discharge

The remaining LOS of every patient is decremented at every time step. When the LOS of a patient reaches zero, the patient is discharged from the hospital. We do not model patient deaths.

HCWs visiting patients

At each time step, HCWs (nurses and doctors) from every ward visit individual patients. This is the moment where a contact between HCW and patient takes places. Single HCWs visit one single patient in one time step. When a HCW and a patient meet in a room and if one of them (patient or HCW) is infected with SARS-CoV-2, a transmission event can take place. As explained earlier in the section "Transmission", a Bernoulli trial using the average probability of transmission per contact with a susceptible person is carried out. If a trial is successful, a susceptible individual acquires infection. The patient may be in an exposed state and develop symptoms after an incubation period or be asymptomatically infected. All exposed individuals in the model follow symptomatic route whereas asymptomatic individuals follow asymptomatic route. To decide on this for a patient, a random number is drawn and if it is less than the specified proportion of asymptomatic patients (P_A^p), the patient state is changed to asymptomatic, otherwise exposed. Similarly, for a HCW if the random number is smaller than the specified proportion of asymptomatic HCWs (P_A^h), the HCW's state is changed to asymptomatic, otherwise the state of the susceptible HCW is changed to exposed. Infectiousness from symptomatic or asymptomatic individuals over time is different as explained earlier in the section "Infectiousness". For exposed patients and HCWs, an incubation period t_{inc} is drawn from the Gamma distribution ($s(\tau)$) with a mean of 5.5 days.

Exposed to infection

For every individual in an exposed state, the incubation period is then decremented by 1 at every time step. When the incubation time of an exposed individual (patient or HCW) reaches zero, the individual is either moved to moderate or severe symptomatic state depending on the proportion of individuals developing severe symptoms (P_s).

For patients who develop severe symptoms, a sample is drawn from the LoS of severely infected patients (based on UMCU data) and the LoS of the respected patient is extended

accordingly. If the drawn LoS is shorter than the original one, the original LoS remains unchanged. The LoS of moderately infected patients is not extended. This is based on the assumption that those patients do not require intensive care and they are sent home for isolation. Severely infected patients are then moved to one of the COVID-ICUs, whereas moderately infected patients are moved to any of the COVID wards.

A HCW who develops symptoms is sent home for an isolation period of seven days, and a susceptible HCW is added in the same ward and shift as a replacement.

Infection to recovery

Moderately or severely infected patients are assumed to recover either at the time of discharge or after a maximum period of 14 days or 35 days, respectively. Asymptomatically infected patients recover at the time of discharge or after maximum period of 14 days. Moderately or severely infected HCWs return to work as recovered after the end of the isolation period of seven days. When a recovered HCW returns to work to a specific ward, a HCW is removed from that ward as we assume a constant HCW population. To do that, we first look in the list of susceptible HCWs in the same ward and duty shift. If that list is not empty, a susceptible HCW is randomly chosen and removed from the hospital HCWs population. If there is no susceptible HCWs in that ward and duty shift, we look further into the list of exposed and asymptomatic HCWs in the same ward and duty shift. If that list is not empty, then a randomly chosen exposed or asymptomatic HCW is removed from the hospital population. If this is also not successful, we randomly choose a recovered HCW from the same ward and duty shift and remove him from the hospital population. These steps are required to maintain a constant HCW population.

HCWs meeting in the common areas

Every hour, two HCWs meet in the common areas of every ward. For this, we randomly pick two HCWs from the list of HCWs working in a ward in a shift. If one of the randomly chosen HCW is exposed or asymptomatic and the other HCW is susceptible, a transmission event can take place. As explained earlier in the section "HCWs visiting patients", a Bernoulli trial using the average probability of transmission per contact with a susceptible person is carried out. If the trial is successful, the susceptible HCW acquires infection and can either enter into an exposed state or an asymptomatic state. Next, a random number is drawn, and if smaller than the specified proportion of asymptomatic HCWs

(P_A^h), the HCW state is changed to asymptomatic, otherwise the state of the HCW is changed to exposed. For an exposed HCW, an incubation period t_{inc} is drawn from the Gamma distribution ($s(\tau)$) with a mean of 5.5 days.

HCWs ward swap

Before the start of each day, a certain proportion of HCWs (W_h) are randomly selected and their wards are changed (Table S1). In order to do that, we first loop over the list of active HCWs (the ones that are not isolated at home). Since we change wards of two HCW at the same time (ward swapping), we draw a random number for every HCW. If the random number is less than $W_h/2$, we select that specific HCW (HCW-A) to be moved to another ward. The next step is to find another HCW (HCW-B) in a ward different from the ward of HCW-A. For this, we again make a list of all the active nurses (if HCW-A was a nurse) or active doctors (if HCW-A was a doctor), and then randomly pick a HCW (HCW-B). Once we have selected both HCW-A and HCW-B, we can now swap the wards, duty shifts and PPE of the both HCWs. We repeat the above process for all active HCWs.

Table S1: **Model parameters**

Name	Symbol	Description	Distribution/Value*	Source
Incubation period	$s(\tau)$	Time between infection and symptom onset	Gamma distribution shape=5.807 scale = 0.948 mean = 5.510 SD = 2.284	Lauer and colleagues[5]
Generation time	$\omega(\tau)$	Time between becoming infected and subsequent onward transmission events	Weibull distribution shape = 2.826 scale = 6.839 mean = 6 days	Grassly and colleagues [14]
Proportion of asymptomatic infections among infected patients	P_A^p	Proportion of infected patients that will experience no symptoms	20%	Buitraga-Garcia and colleagues [19]
Proportion of asymptomatic infections among infected HCWs	P_A^h	Proportion of infected HCWs that will experience no symptoms	31%	Buitraga-Garcia and colleagues [19]
Proportion of severe symptomatic individuals	P_S	Proportion of exposed individuals that will develop severe symptoms	20%	Wu and colleagues [27]
Reproduction number of asymptomatic infectees for wild-type variant	R_A^W	Mean number of infections caused by an individual asymptomatically infected with the wild-type SARS-CoV-2 variant	0.5	Calibrated to UMCU data

Reproduction number of symptomatic infectees for wild-type variant	R_A	Mean number of infections caused by an individual symptomatically infected with the wild-type SARS-CoV-2 variant	1.25	Calibrated to UMCU data
Reproduction number of asymptomatic infectees for new virus variant		Mean number of infections caused by an individual asymptomatically infected with the SARS-CoV-2 variant	0.8 (1.95)	Based on with 56% higher transmissibility, varied in sensitivity analysis
Reproduction number of symptomatic infectees for new virus variant	R_S	Mean number of infections caused by an individual symptomatically infected with the SARS-CoV-2 variant	1.95	Based on with 56% higher transmissibility
Peak sensitivity of RT-PCR test for SARS-CoV-2		Maximum sensitivity of the RT-PCR diagnostic test for SARS-CoV-2	93.1% (79%)	Grassly and colleagues [14], varied in sensitivity analyses
Proportion of HCWs that work in the same ward as during their previous shift	W_h	Proportion of HCWs that change wards they were working in their previous shift	95% (baseline) 100% (intervention)	Assumed

PPE effectiveness		Reduction in infectiousness upon contact between an infected and susceptible individual (includes PPE efficacy and adherence)	90% (50%, 70%)	Assumed
Isolation period for HCWs		Amount of time HCWs have to isolate after symptom onset or after being detected by screening or contact tracing	7 days	Assumed
Recovery time for asymptomatic infection	γ_A	Mean duration of an asymptomatic infection	14 days Sensitivity analysis: Uniform(9,19)	Assumed
Recovery time for symptomatic (moderate, severe) infection	γ_S	Mean duration of a symptomatic infection	14 days (moderate) 35 days (severe) Sensitivity analysis: Uniform(9,19) Uniform(30,40)	Liu and colleagues [16]
LoS of non-COVID patients in ICU			Lognormal meanlog = 0.37 sdlog = 0.82 mean = 1.45 days sd = 2.27	Fitted distributions to UMCU data from 2014-2017

LoS of non-COVID patients in normal ward			Weibull shape = 0.92 scale = 4.18 mean = 4.35 days	Fitted distributions to UMCU data from 2014-2017
LoS of moderately infected patients			Gamma shape = 1.88 rate = 0.25 mean = 31.8 days sd = 30.08	Fitted distributions to UMCU data from 2020
LoS of severely infected patients			Gamma shape = 1.59 rate = 0.05 mean = 7.52 days sd = 636	Fitted distributions to UMCU data from 2020
Patient-nurse ratio		1:1 (COVID ICU), 2:1 (COVID ward), 1:1 (Normal ICU), 4:1 (Regular ward)		Assumed
Patient-doctor ratio		6:1 (COVID ICU, COVID ward, Normal ICU), 10:1 (Regular ward)		Assumed
Frequency of HCWs visiting patients (ward dependent)		Min 10 minutes, Max 30 minutes		
Duty shifts of HCWs per day		3 shifts		

Rounds per shift		Nurses: 6 Doctors: 2		
Contact rates (per shift)	$c_{n,n}$	Average number of contacts between nurses	4·6	From simulation
	$c_{n,p}$	Average number of contacts that a nurse has with patients	19·07	
	$c_{n,d}$	Average number of contacts that a nurse has with doctors	3	
	$c_{p,n}$	Average number of contacts that a patient has with nurses		
	$c_{p,d}$	Average number of contacts that a patient has with doctors	2	
	$c_{d,d}$	Average number of contacts between doctors	0·43	
	$c_{d,p}$	Average number of contacts a doctor has with patients	17·4	
	$c_{d,n}$	Average number of contacts a doctor has with nurses	3	
Daily arrival rate of non-COVID patients		Number of patients that arrive at the hospital per day for non-COVID related reasons	40 patients per day	Based on UMCU data from 2014-2017 assuming 50% decrease during study period

HCW population		Constant number of HCWs working in the hospital per day	870	
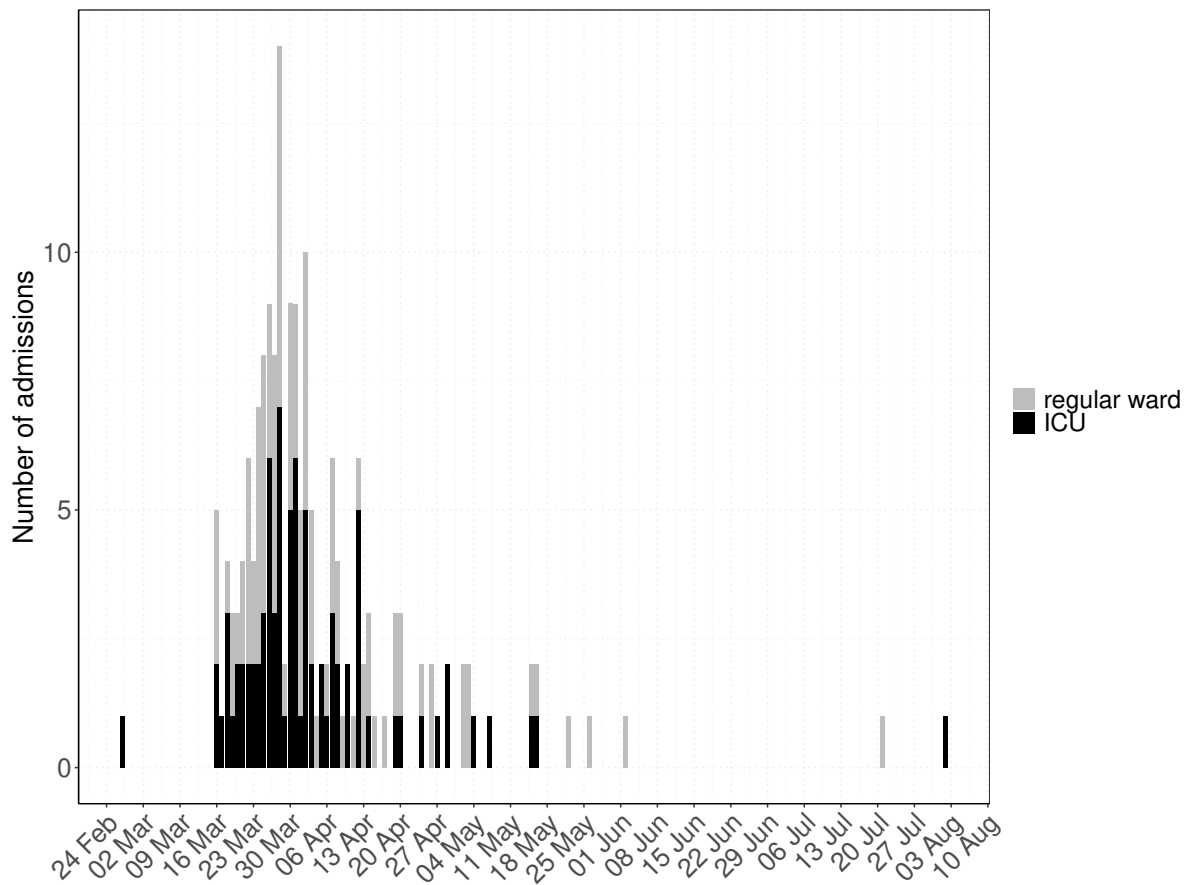
Figure S1: **Number of patients admitted to UMCU with a SARS-CoV-2 infection between 27 February and 2 August 2020.** Patients were either admitted to an ICU or to another ward in the hospital.
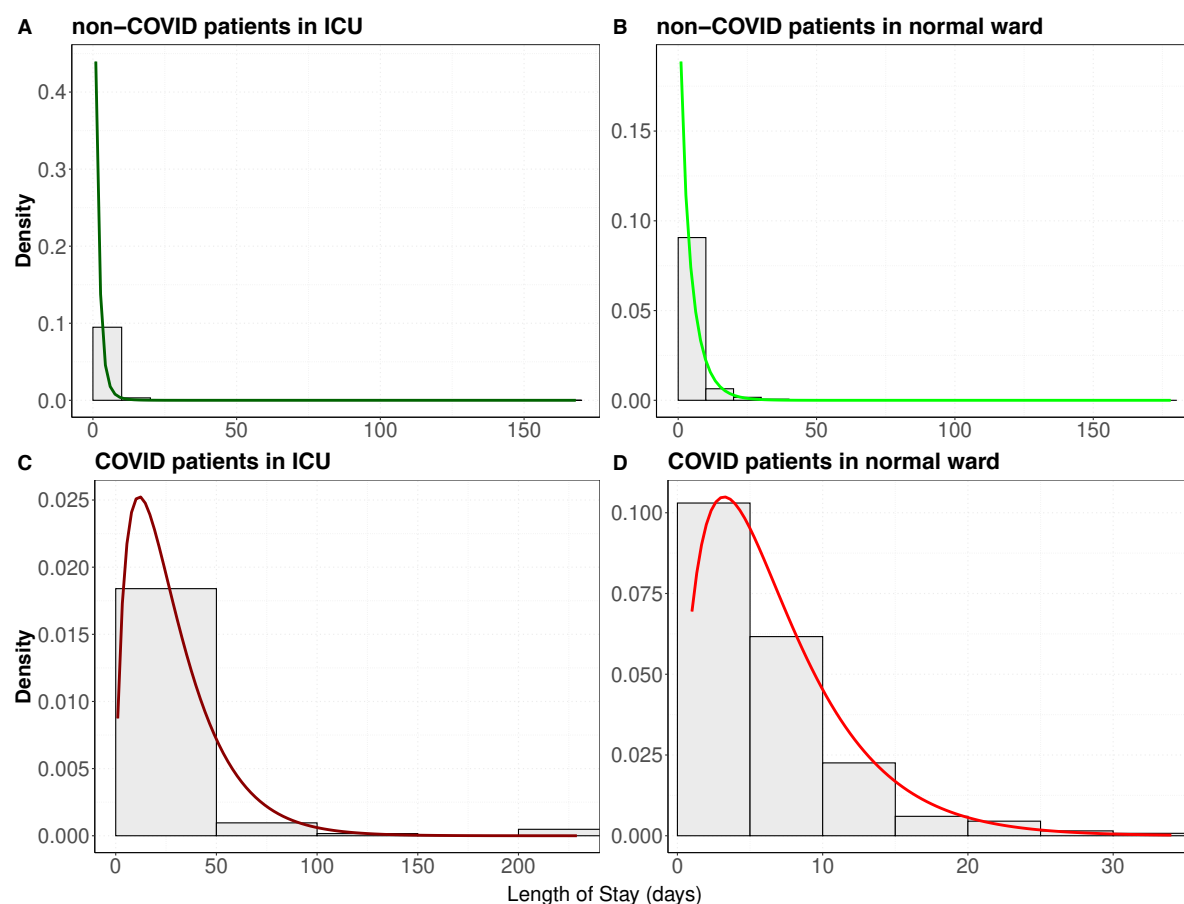
Figure S2: **Length of stay data of UMCU and fitted distributions for non-COVID and COVID patients in the hospital.** (A)-(B) Histograms show the length of stay distributions for patients admitted to the UMCU between 2014 and 2017. The bold lines represent the fitted probability distributions. The length of stay of patients admitted for non-COVID reasons to the ICUs and to normal wards follow a lognormal distribution and Weibull distribution, respectively. (C)-(D) Histograms show the length of stay distributions for patients admitted with a SARS-CoV-2 infection to the UMCU between 27 February and 24 August 2020. The bold lines represent the fitted probability distributions. The length of stay of patients admitted with a SARS-CoV-2 infection to ICUs and to normal wards follow gamma distributions. The parameters of the probability distributions can be found in Table S1.
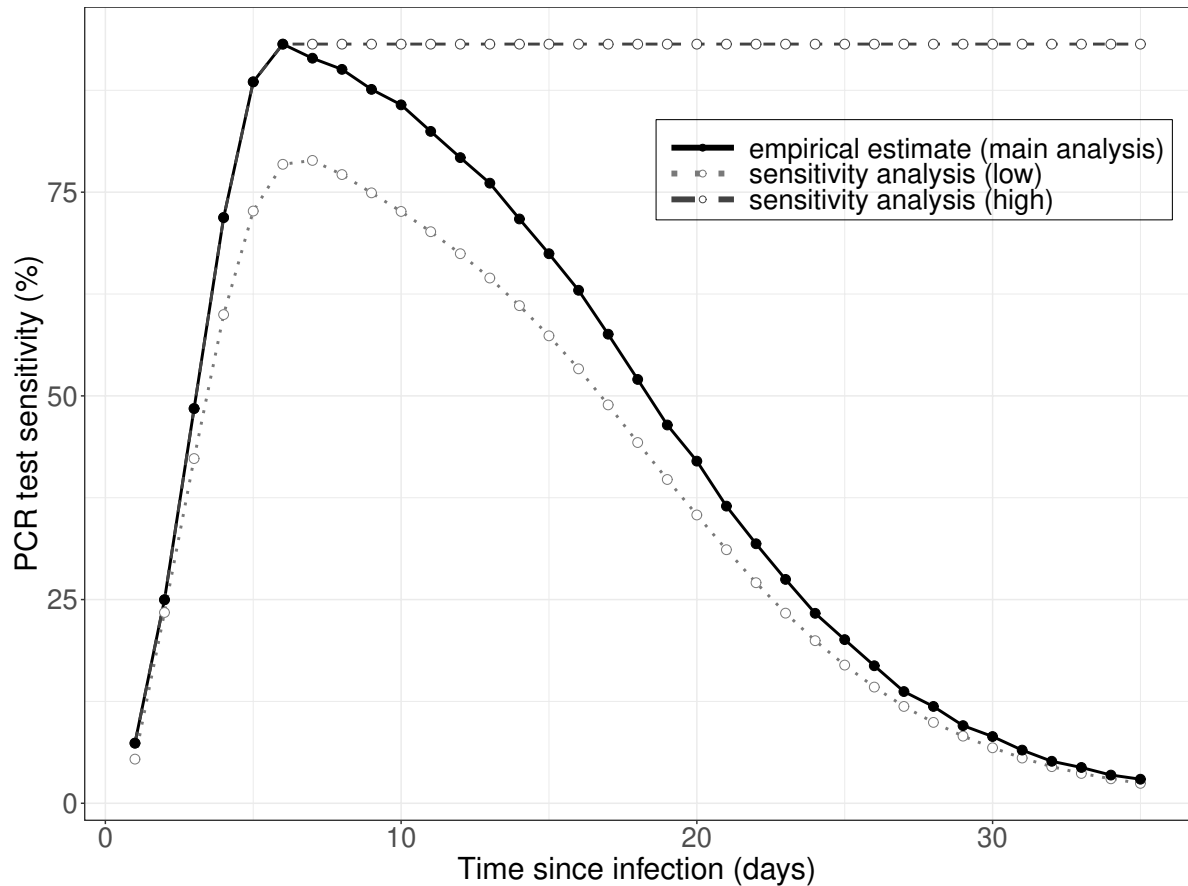
Figure S3: **PCR test sensitivity over time since infection.** The empirical estimate based on published data as reported by Grassly and colleagues is shown (black dots) Two sensitivity analyses were performed: 1) assuming the test sensitivity remains at the maximum after reaching its peak (dark grey dashed) and 2) test sensitivity curve of the main analysis scaled down to 85% (light grey dotted).
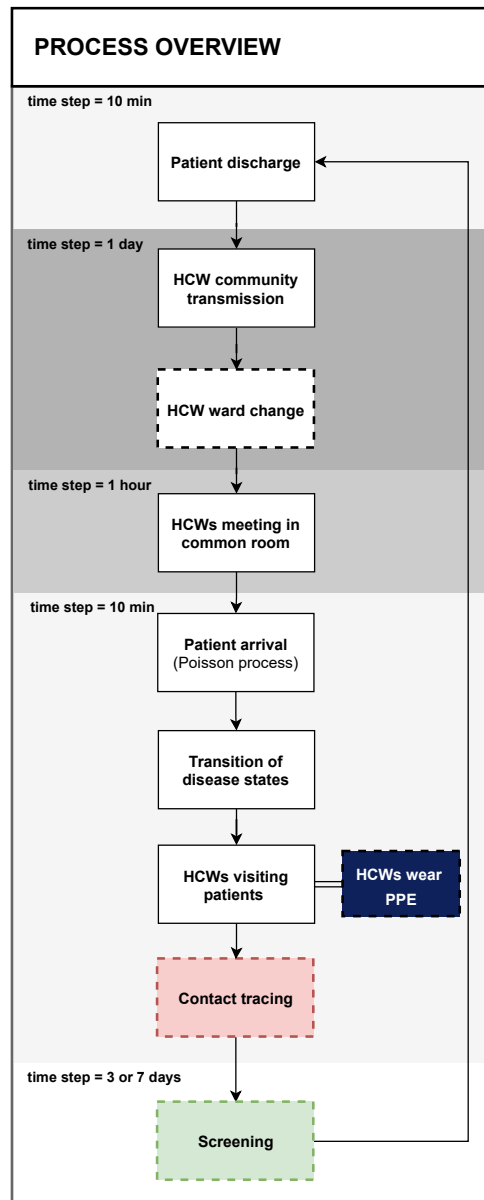
Figure S4: **Overview of processes in the agent-based model.** The dashed boxes indicate processes that are only performed for the respective intervention scenario. The time indicates when the process is called (e.g., every 10 min). The smallest time step in the model is 10 minutes.
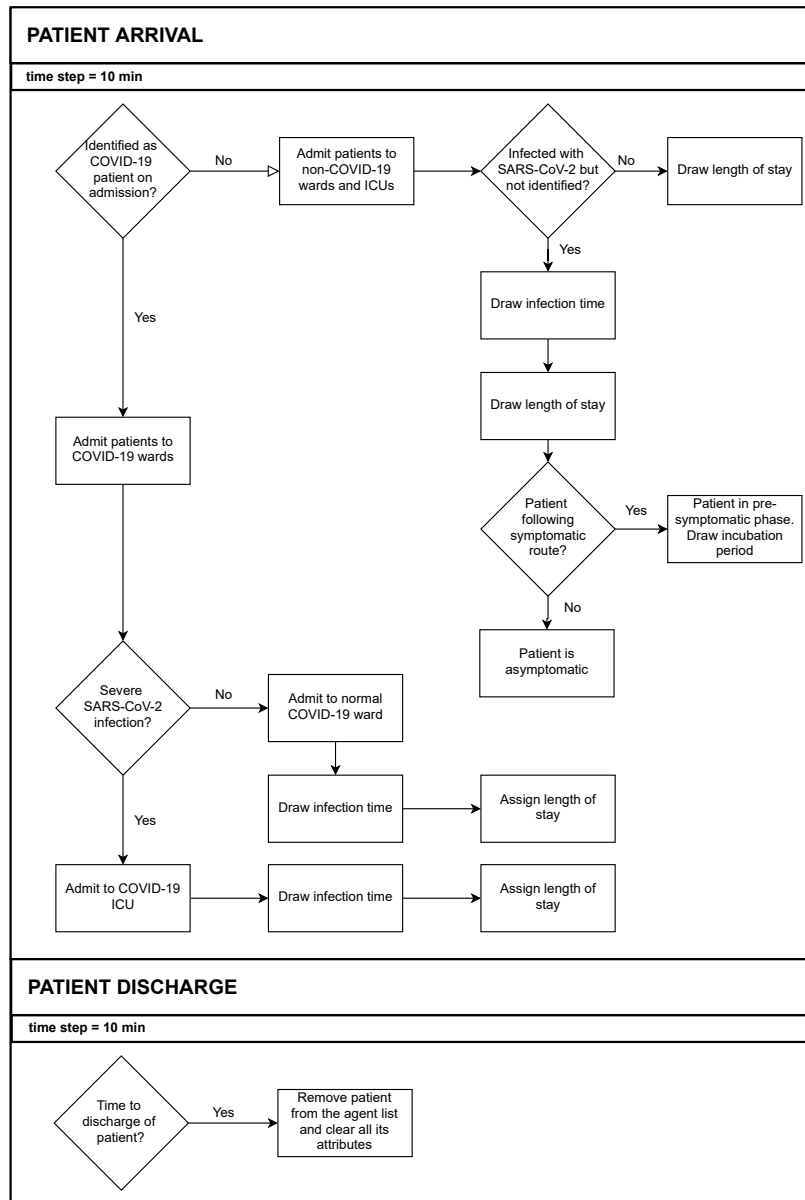
Figure S5: **Flowchart for patient arrival and patient discharge in the agent-based model.** The time indicates that the process is called every 10 minutes. Note that the smallest time step in the model is 10 minutes.
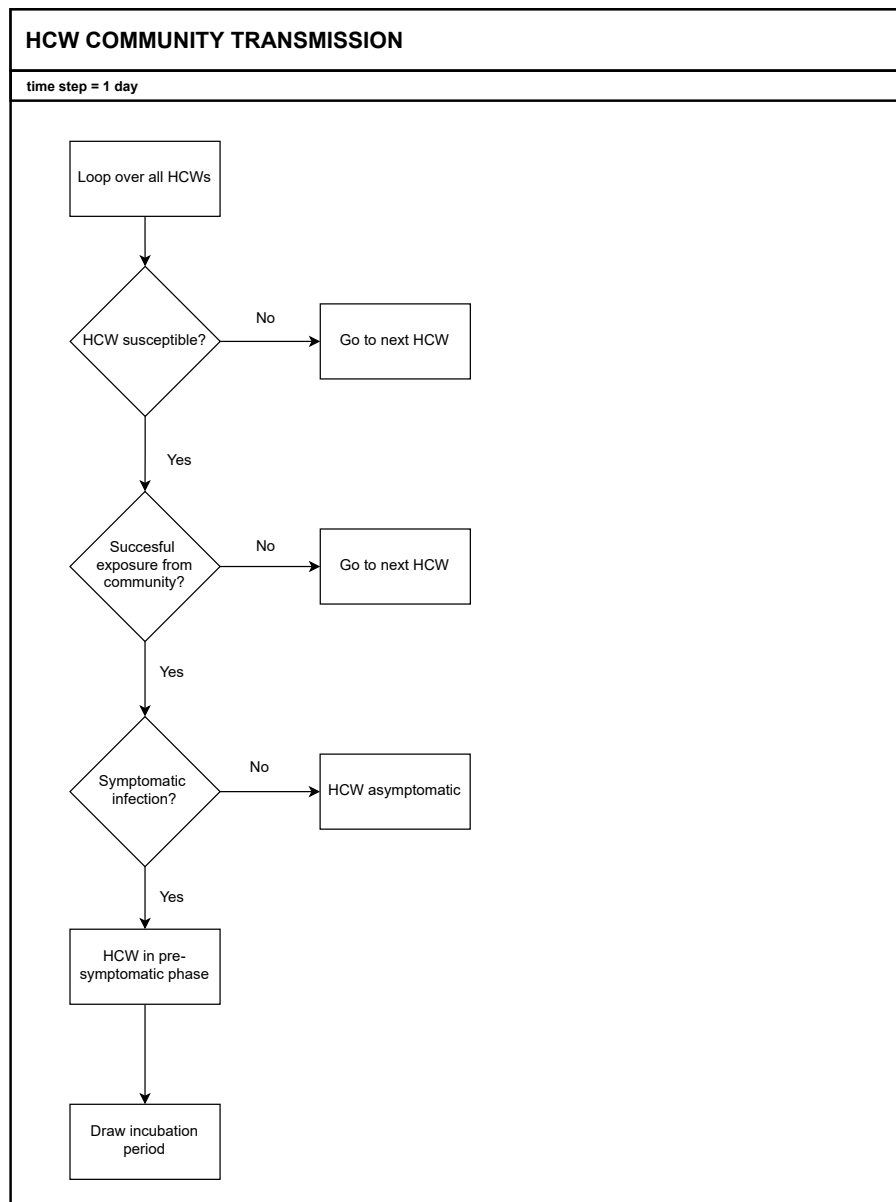
Figure S6: **Flowchart for HCW community transmission in the agent-based model.** The time indicates that this process is called once a day. Note that the smallest time step in the model is 10 minutes.
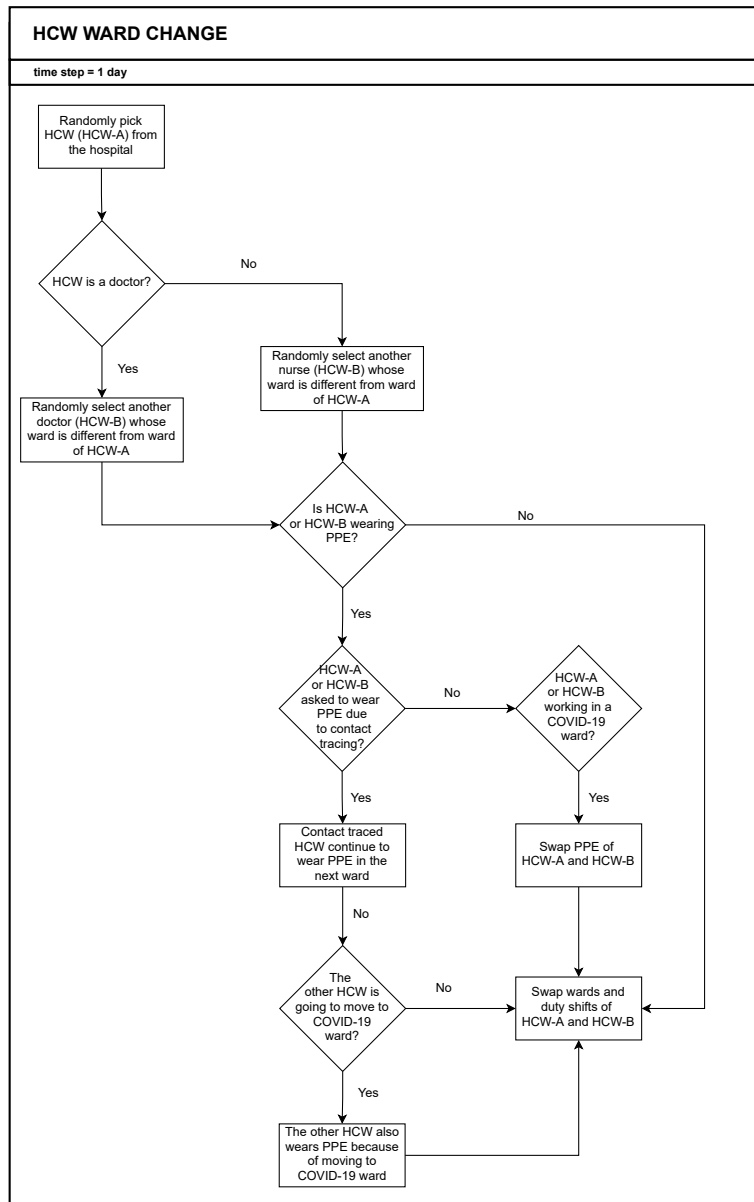
Figure S7: **Flowchart for HCW ward change in the agent-based model.** The time indicates that this process is called once a day. Note that the smallest time step in the model is 10 minutes.
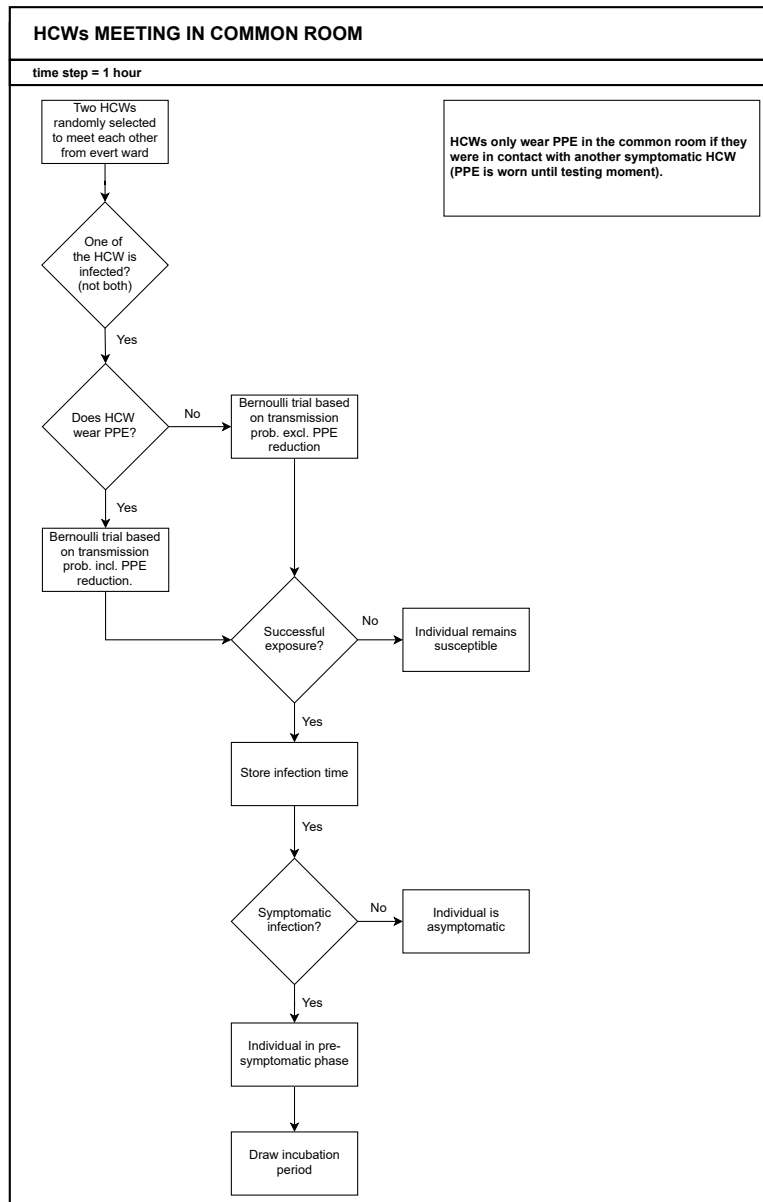
Figure S8: **Flowchart for HCWs meeting in common room in the agent-based model.**
The time indicates that this process is called every hour. Note that the smallest time
step in the model is 10 minutes.

Figure S9: **Flowchart for transition of disease states in agent-based model.** The time indicates that this process is called every 10 minutes. Note that the smallest time step in the model is 10 minutes.
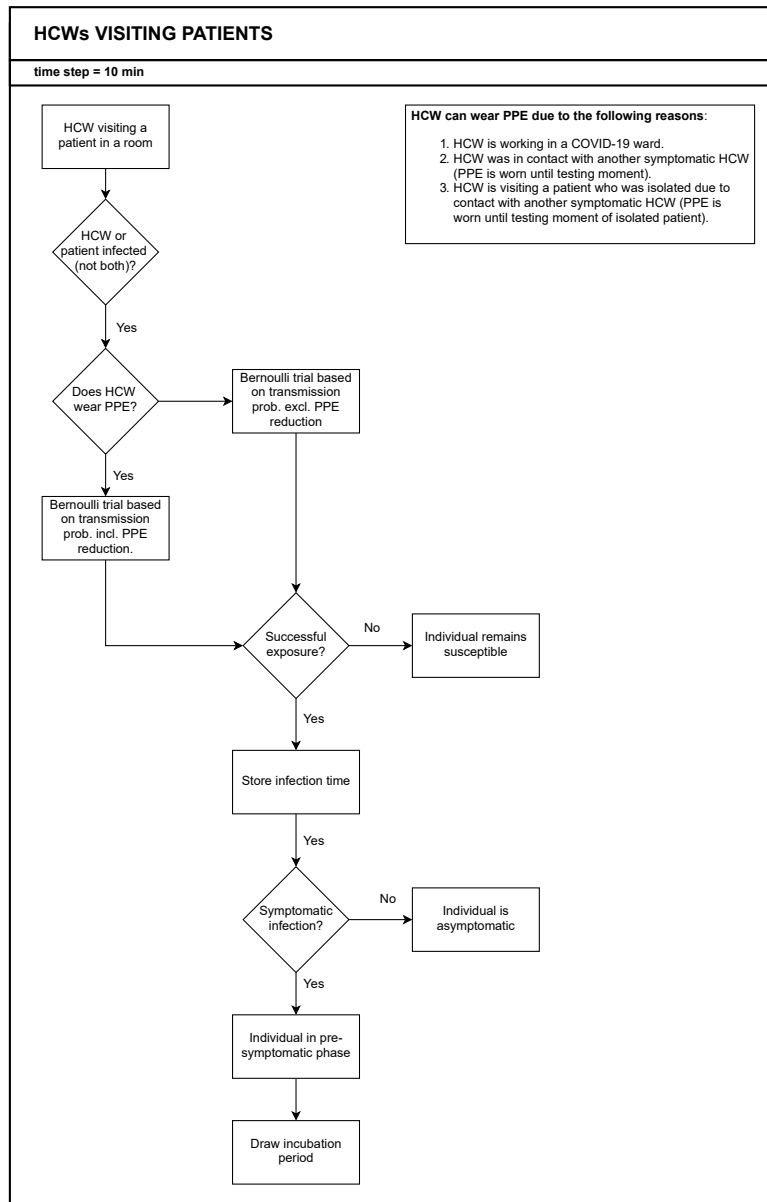
Figure S10: **Flowchart for HCWs visiting patients in the agent-based model.** The time indicates that this process is called every 10 minutes. Note that the smallest time step in the model is 10 minutes.

Screening

Test all HCWs at the screening moment

Given PCR test sensitivity, HCW detected positive?

Yes

Send HCW to quarantine

Add replacement HCW

The following screening moments are considered:
1. every 3 days
2. every 7 days

This intervention is applied with perfect test sensitivity as well as with time-varying test sensitivity.
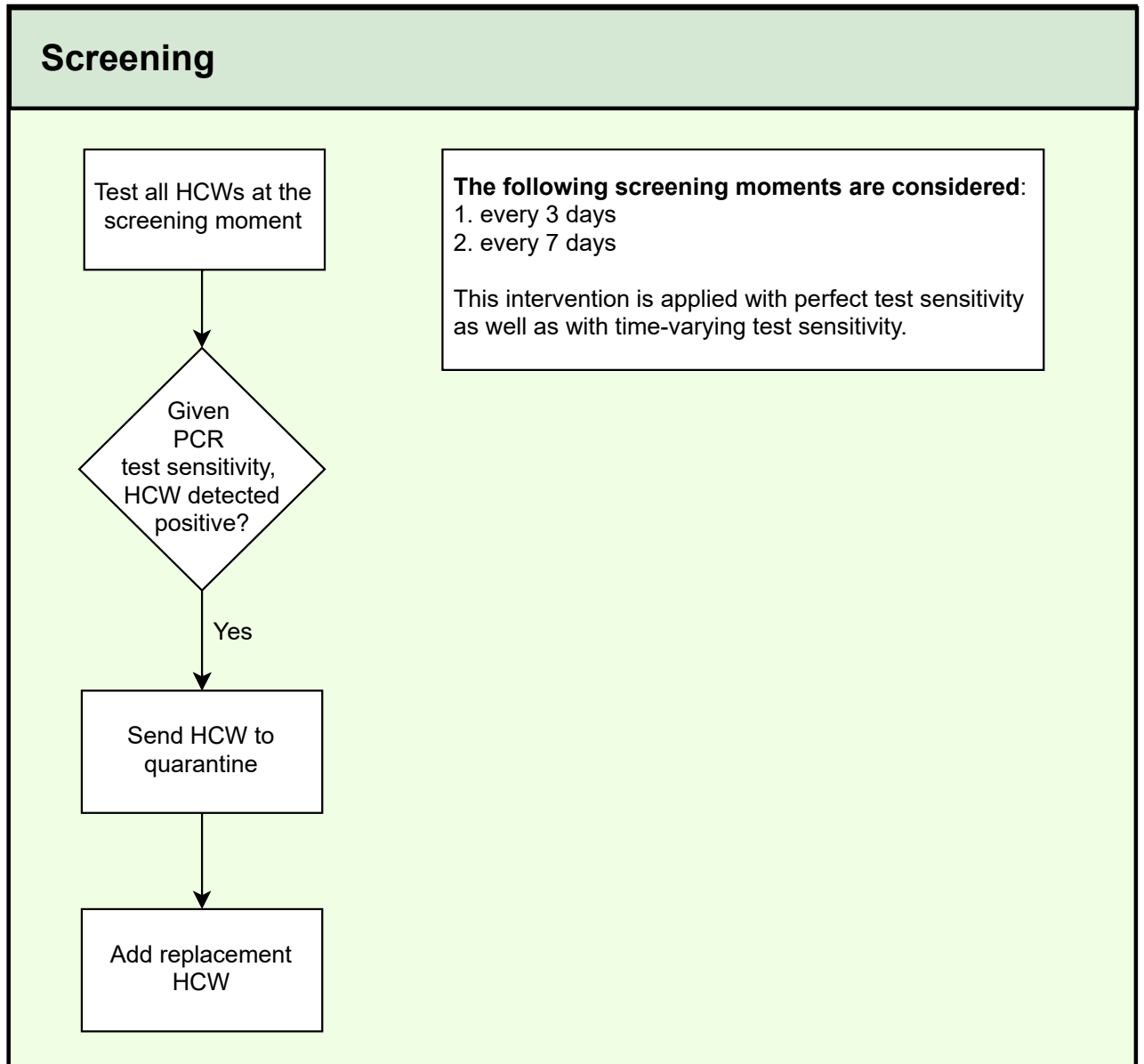
Figure S11: **Flowchart for contact tracing in the agent-based model.**
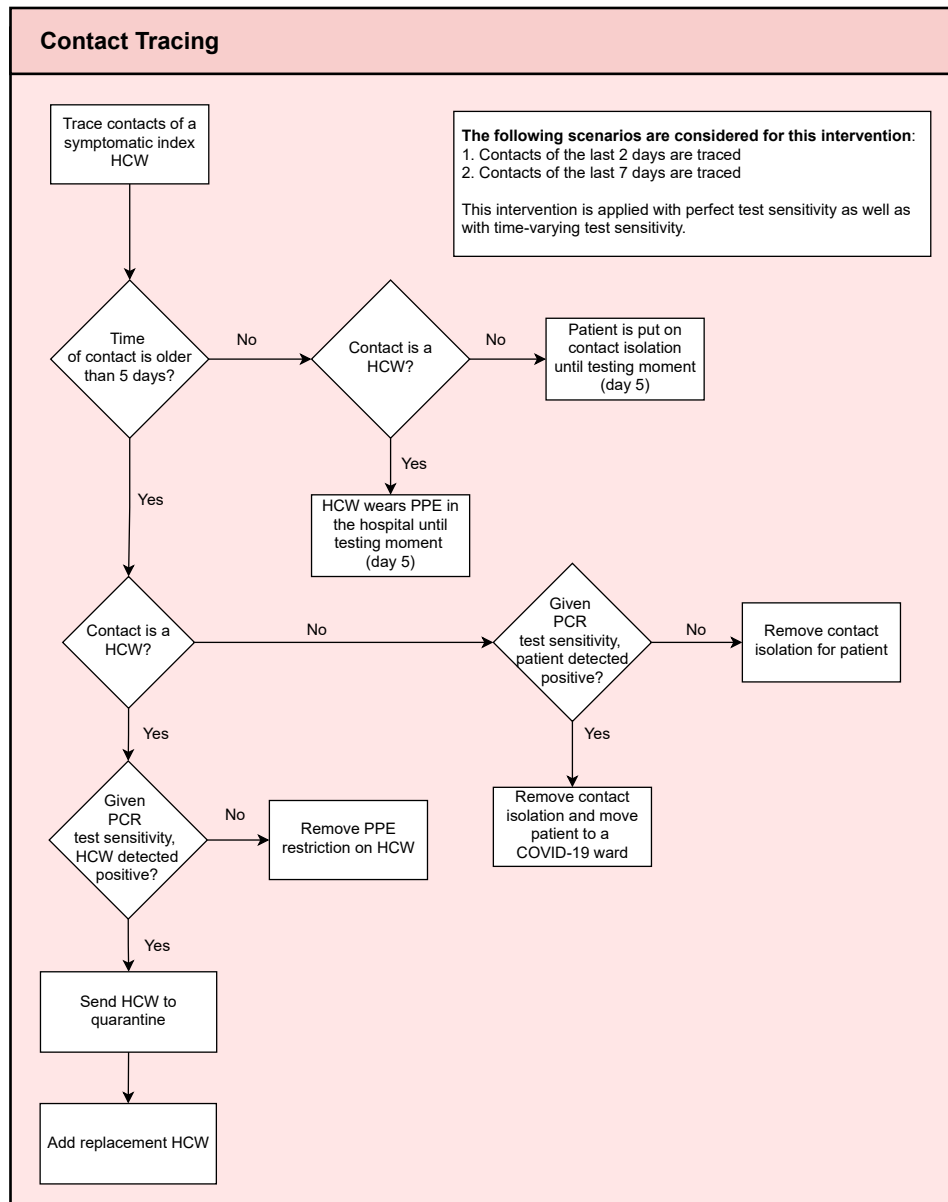
Figure S12: **Flowchart for HCW screening in the agent-based model.**

Additional File 2: Supplementary results

I. Results on transmission routes

Our results show that for the considered simulation scenarios most of the nosocomial transmissions of the SARS-CoV-2 variant is mainly driven by transmissions between patients and HCWs (Figure S4). This is expected as we assumed that there is no direct contact between patients and the majority of contacts of HCWs are with patients. Furthermore, for most of the intervention scenarios, over 90% of transmissions occur in non-COVID wards where no use of PPE is assumed in the baseline scenario (Figure S6). Since in our model infected patients are transferred to COVID wards and infected HCWs are assumed to self-isolate immediately upon symptom onset, most transmissions take place during the pre-symptomatic stage of an infected individual (dark-grey bars in Figure S7). This is in line with a French study where secondary cases were exposed mainly in the pre-symptomatic phase [40]. When PPE is used throughout the hospital or HCWs are screened assuming a perfect test sensitivity, most transmissions are prevented (Figure 5 of the main text). In particular, transmissions that occur during non-symptomatic states in non-COVID wards are significantly reduced, decreasing their contribution to the overall number of transmissions (Figure S6-Figure S7).

II. Results of sensitivity analyses

We evaluate the changes of our results with respect to changes in our model parameters. The figures for the sensitivity analyses can be found online in our publication [41] and in `https://github.com/htahir2/covid_intra-hospital_model.git`

PPE effectiveness

We performed two sensitivity analyses to test the impact of PPE effectiveness values on our results:

   a) 50% effective PPE

   b) 70% effective PPE

Our sensitivity analyses show that the effective reproduction numbers and the total number of nosocomial transmissions increase with lower PPE effectiveness and decrease with higher PPE effectiveness, in particular for the "PPE in all wards" intervention scenario (compare Figure S8-S7, Figures S11-S12, and Figures 4-5 of the main text). A similar effect can be observed for the daily percentage of HCW absenteeism (compare Figure S10, Figure S13, and Figure 6 of the main text). The relative impact of the different interventions on the reproduction number in comparison to the baseline scenario are similar to what we have observed in our main analysis. The only difference is that for a low value of PPE effectives of 50%, screening every three days with time-invariant perfect sensitivity is more effective in reducing the effective reproduction number, especially for pre-symptomatic HCWs. However, the use of 50% effective PPE in all wards still decreases the effective reproduction number more than the remaining interventions.

Reproduction number

We performed a sensitivity analysis to test the impact of equal reproduction numbers of symptomatically and asymptomatically infected individuals on our results (Figure S14-S16). Furthermore, we show the model results for the reproduction numbers resulting from calibrating our model to data on the number of occupied beds by COVID-19 patients at the UMCU (Figure S17-S19). Our sensitivity analyses show that the effective reproduction numbers, the total number of nosocomial transmissions as well as the daily percentage of HCW absenteeism increase with increasing basic reproduction number. In particular, when the reproduction number of asymptomatically infected individuals is as high as the one of symptomatically infected individuals, the respective effective reproduction numbers for asymptomatic patients and HCWs increase. The impact on the overall effective reproduction number is smaller, however, still notable. Qualitatively, our conclusions regarding the relative effect of the considered infection control interventions remain unchanged. For low reproduction numbers as it was the case for the nosocomial spread of the wild-type SARS-CoV-2 variant at UMCU, the numbers of nosocomial transmissions are very small and hence the relative impact of the intervention scenarios in comparison to each other and to the baseline scenario is smaller than for higher reproduction numbers. However, the qualitative conclusions remain unchanged.

Increased HCW-to-HCW contact rate

In our main analysis, we assume that HCWs meet other HCWs once every hour. In this sensitivity analysis, we relax this assumption by increasing the contact rates between HCWs to once every 30 minutes and evaluate the impact on our results (Figure S20-S22). The effective reproduction numbers, the total number of nosocomial transmissions, and the daily percentage of HCW absenteeism increase when the contact rate between HCWs is increased. In particular, the effective reproduction numbers for HCWs increase but not those for symptomatic patients (Figure S20). Qualitatively, our conclusions with respect to the impact of the interventions on the hospital epidemic do not change with respect to this parameter.

Test sensitivity

We performed two sensitivity analyses:

a) assuming the test sensitivity to remain at the maximum after reaching its peak (high test sensitivity scenario) and

b) reducing the test sensitivity curve of the main analysis by 15% (low test sensitivity scenario).

The respective test sensitivity curves varying from time since infection are shown in Additional File 1: Figure S1. There are only minor differences in our results for both sensitivity scenarios (Figure S23-S25 vs Figures 4-6 of the main text).

Recovery time

To test the impact of the recovery time of infected individuals (i.e., the time after which infected individuals are set to non-infectious and recovered in the model), we performed the simulations with a stochastic (instead of fixed) implementation of the recovery times. For this sensitivity analysis we assumed the following uniform distributions for the recovery times:

- Unif(9,19) for asymptomatic and moderately symptomatically infected individuals

- Unif(30,40) for severely symptomatically infected individuals.

The parameters in brackets represent the time since infection and serve as lower and upper bounds in the uniform distribution. Qualitatively, our results do not change with respect to this parameter (Figure S29-S31).

Table S2: Outcome measures for baseline and intervention scenarios.

Scenario	Effective reproduction number*					Total number of nosocomial transmissions*	Peak percentage of HCW absenteeism (%)†	Peak positivity rates (%)*
	Overall	Symptomatic patient	Asymptomatic patient	Pre-symptomatic HCW	Asymptomatic HCW			
Baseline	0.65 (0.57, 0.71)	0.55 (0.46, 0.66)	0.40 (0.23, 0.60)	0.94 (0.79, 1.07)	0.44 (0.33, 0.54)	526.2 (362.3, 675.2)	5.4	
HCW cohorting	0.62 (0.56, 0.67)	0.50 (0.40, 0.60)	0.39 (0.17, 0.60)	0.91 (0.79, 1.04)	0.44 (0.31, 0.55)	457.9 (359, 565.1)	5.2	
PPE in all wards	0.10 (0.07, 0.14)	0.03 (0.01, 0.06)	0.03 (0.00, 0.42)	0.21 (0.13, 0.32)	0.09 (0.02, 0.18)	32.9 (21.5, 48.5)	2.3	
Screening 3 days perfect sens	0.24 (0.14, 0.32)	0.23 (0.10, 0.36)	0.53 (0.00, 1.23)	0.28 (0.19, 0.40)	0.12 (0.03, 0.24)	90.9 (47, 136.5)	5.1	1.7 (1.6, 1.8)
Screening 3 days	0.59 (0.51, 0.65)	0.53 (0.42, 0.62)	0.47 (0.24, 0.71)	0.81 (0.68, 0.92)	0.36 (0.22, 0.51)	419 (298.1, 528.7)	8.6	2.5 (2.4, 2.6)
Screening 7 days	0.63 (0.54, 0.69)	0.55 (0.47, 0.64)	0.44 (0.25, 0.64)	0.87 (0.73, 0.98)	0.39 (0.29, 0.51)	473 (353.9, 614.3)	6.6	5.1 (5, 5.3)
7-day Contact tracing perfect sens	0.44 (0.36, 0.55)	0.29 (0.21, 0.38)	0.31 (0.11, 0.61)	0.74 (0.60, 0.90)	0.36 (0.22, 0.50)	232.7 (155.9, 341.4)	4	15.1 (14.1, 16.1)
2-day Contact tracing	0.41 (0.33, 0.48)	0.27 (0.18, 0.35)	0.30 (0.08, 0.65)	0.71 (0.55, 0.84)	0.32 (0.19, 0.47)	202.9 (138.5, 269.6)	3.6	11.3 (9.4, 13.1)
7-day Contact tracing	0.39 (0.33, 0.44)	0.25 (0.17, 0.34)	0.29 (0.05, 0.61)	0.67 (0.52, 0.80)	0.31 (0.17, 0.44)	188.7 (139, 248.6)	3.9	10.4 (9.1, 11.6)

*Mean values over 100 simulation runs are given. Values in brackets are the lower and upper bounds of 95% uncertainty intervals.
†7-day moving average of mean percentage (over 100 simulation runs)

Figure S1: **Positivity rate of screening interventions for different prevalence ranges.**
Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the
SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-
CoV-2 variant). (A) Screening every three days with constant perfect test sensitivity. (B)
Screening every three days with imperfect, time-varying test sensitivity. (C) Screening
every seven days with imperfect, time-varying test sensitivity. On each day when HCWs
were screened, the number of positive tested HCWs among the total number of screened
HCWs is computed. The prevalence values on the day when HCWs were screened is
divided into categories. For each prevalence category, the positivity rate was computed
by the total number of positive tested HCWs divided by the total number of screened
HCWs (merging values of all simulations) and is shown as a point. The error bars represent
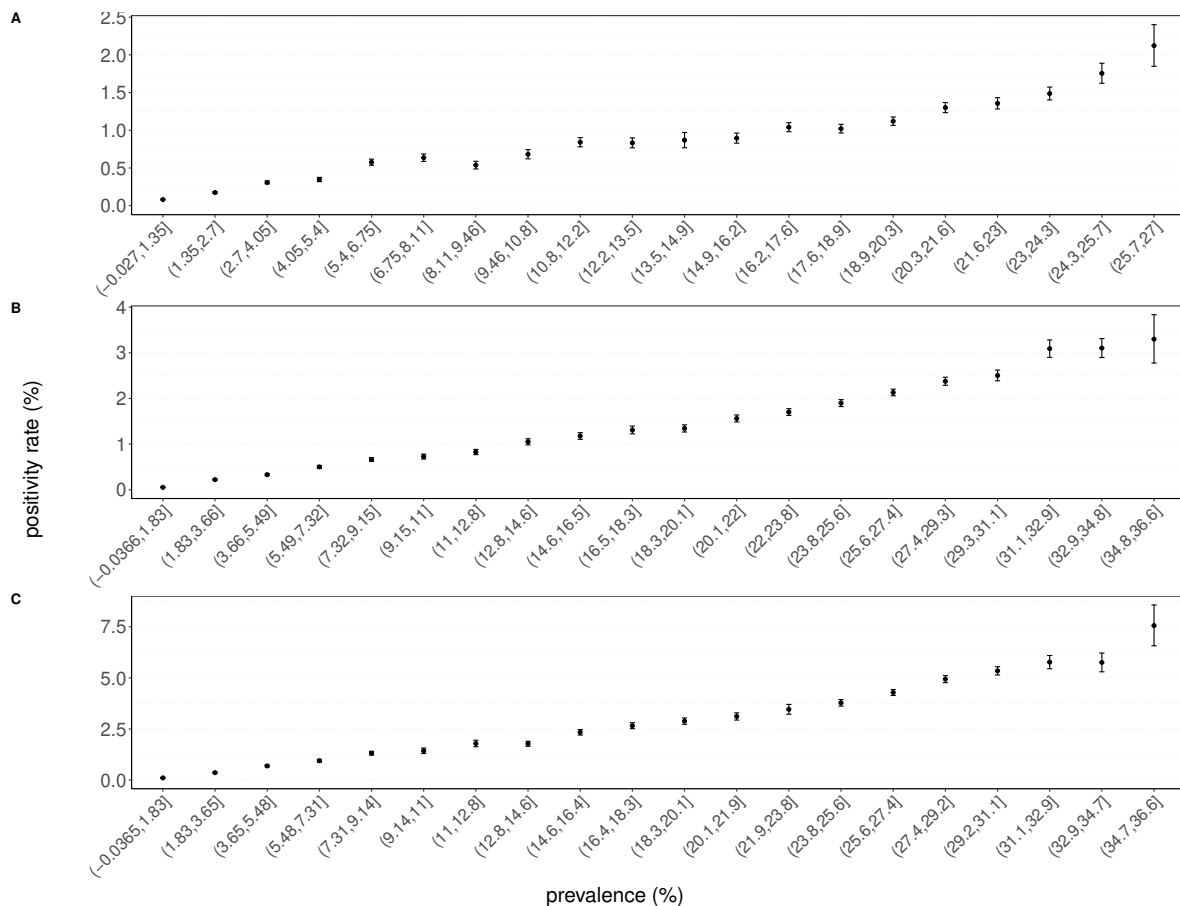the 95% binomial proportion confidence intervals.
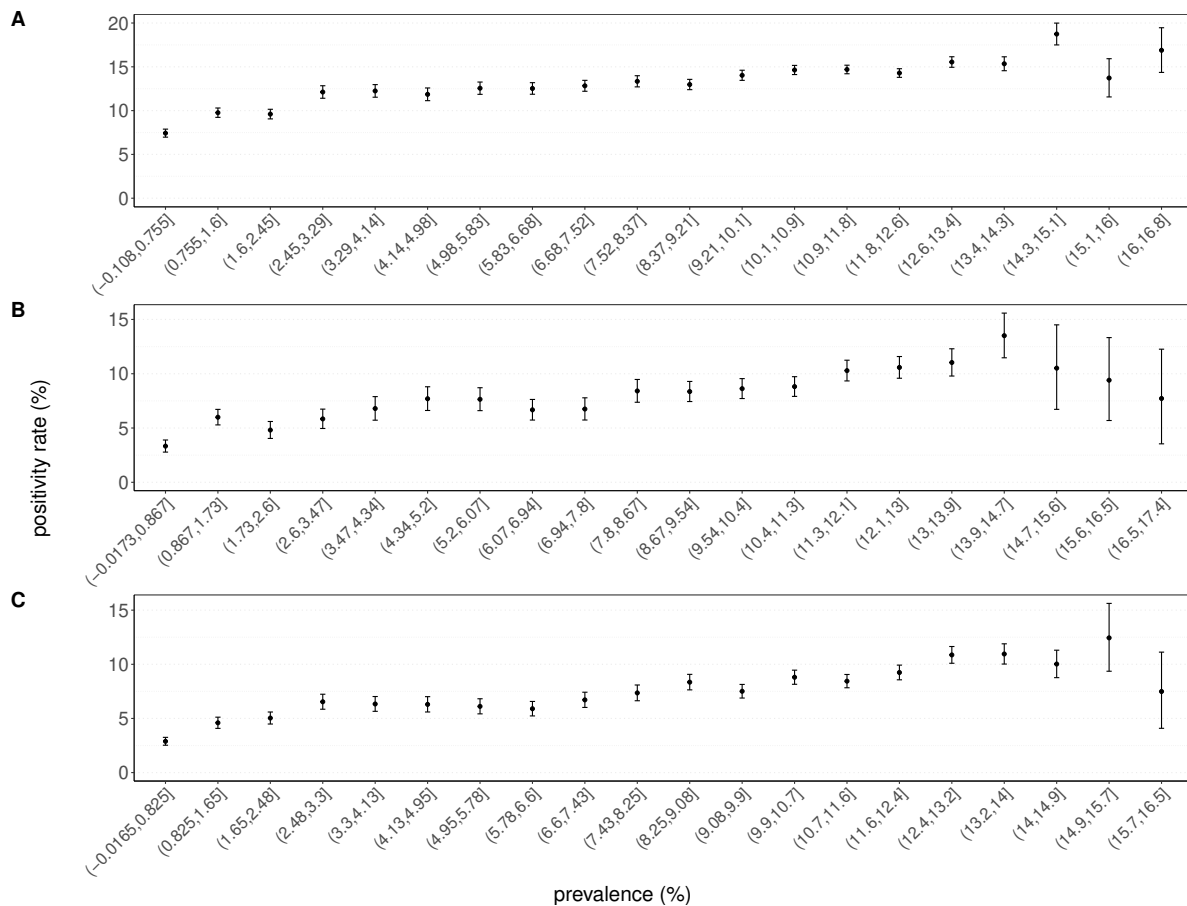
Figure S2: **Positivity rate of contact tracing interventions for different prevalence ranges.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). (A) Contact tracing of contacts two days prior to symptom onset with perfect test sensitivity. (B) Contact tracing of contacts two days prior to symptom onset with time-varying imperfect test sensitivity. (C) Contact tracing of contacts seven days prior to symptom onset with time-varying, imperfect test sensitivity. For each index case (symptomatically infected HCW), the number of positive contacts and total number of contacts that are traced is computed in each simulation. The prevalence values on the day when an index case was traced, is divided into categories. For each prevalence category, the positivity rate is computed by the total number of positive divided by the total number of traced contacts (all simulations merged) and is shown as a point. The error bars represent the binomial proportion confidence interval.
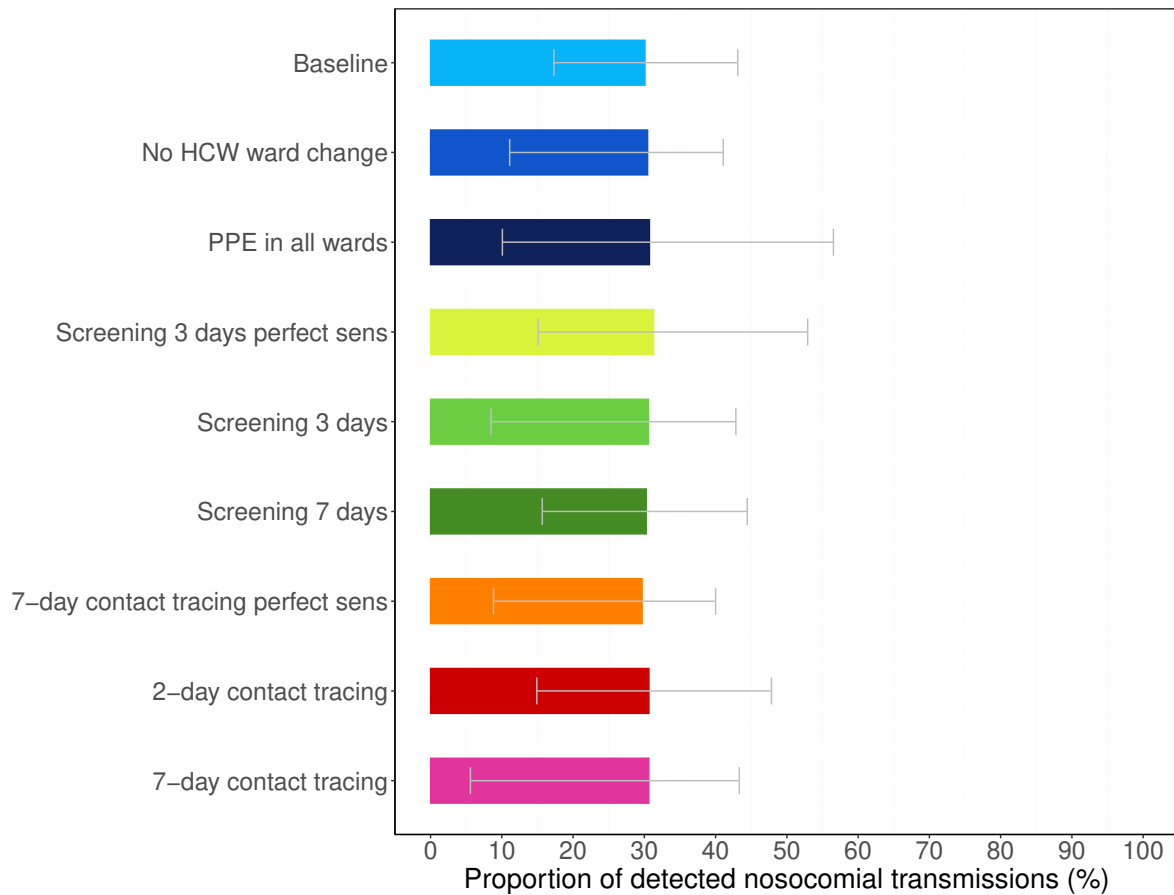
Figure S3: **Proportion of detected nosocomial transmissions of the SARS-CoV-2 variant for each simulation scenario.** Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant). The colored rectangular bars with black borders represent the mean proportion of patients detected with a SARS-CoV-2 infection in the hospital due to symptom onset or detection by an intervention (over 100 simulation runs). The denominator are patients either admitted with a SARS-CoV-2 infection (asymptomatic or pre-symptomatic) or acquired it in the hospital. The proportions of infected patients undetected comprise patients who are discharged to community in a pre-symptomatic or asymptomatic state. The grey error bars the respective 95% uncertainty intervals.
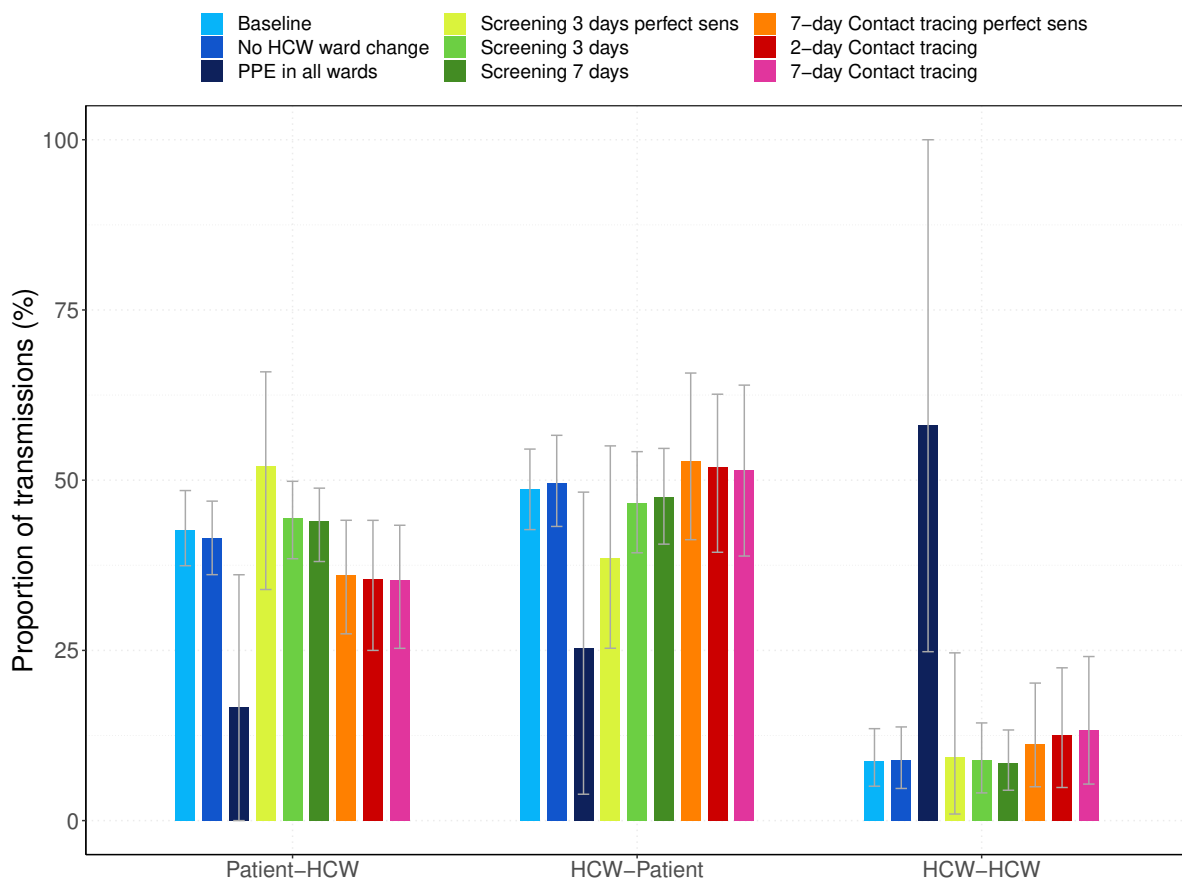
Figure S4: **Transmission route contributions for nosocomial transmissions of the SARS-CoV-2 variant for each simulation scenario.** Three different transmission routes are considered: From patient to HCW (Patient-HCW), from HCW to patient (HCW-patient), and from HCW to HCW (HCW-HCW). The colored rectangular bars represent the mean percentage of transmissions (averaged over 100 simulations) due to the respective transmission route for each simulation scenario. The grey error bars represent the respective 95% uncertainty intervals. For screening every 3 days and 7-day contact tracing, we considered two different test sensitivity scenarios: time-invariant perfect test sensitivity (perfect sens) and time-varying imperfect test sensitivity.
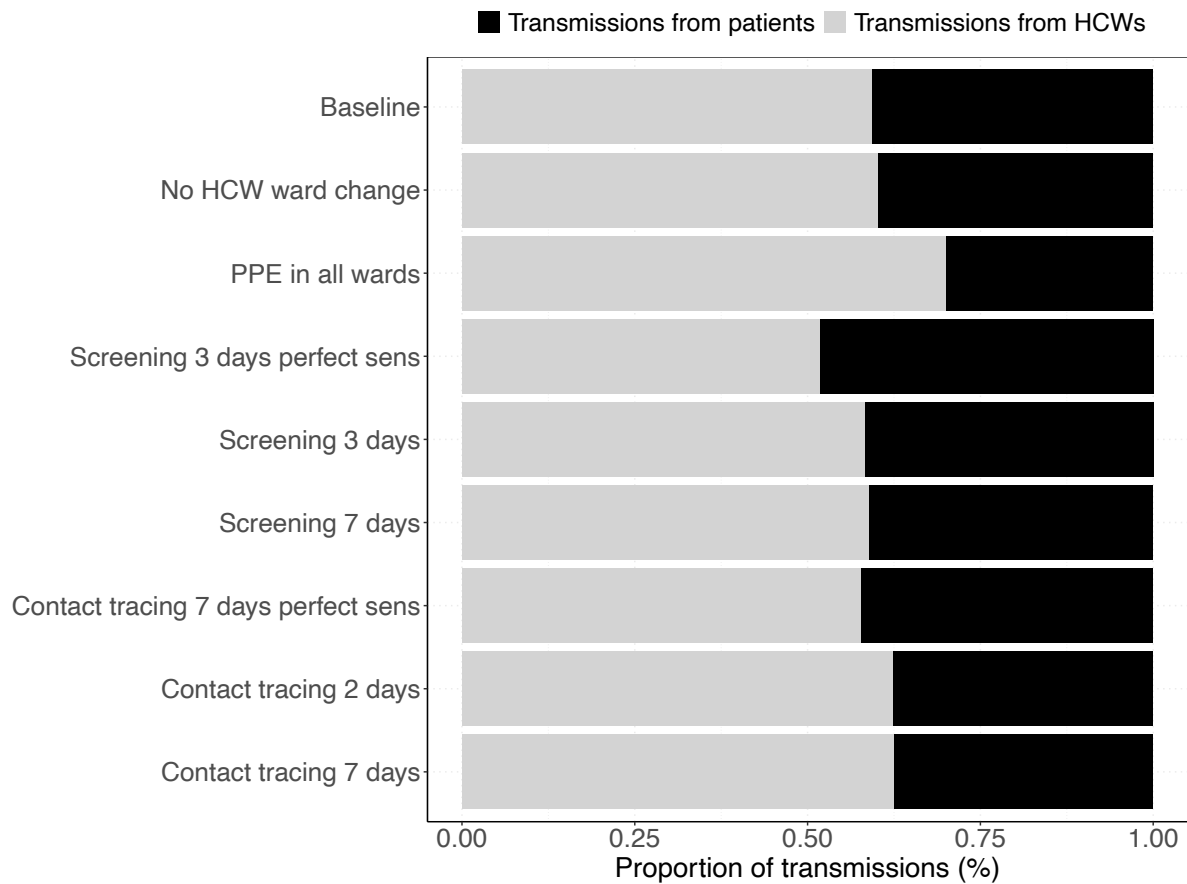
Figure S5: **Proportion of transmissions from HCWs and from patients for each simulation scenarios.** Mean percentage of total transmissions (averaged over 100 simulation runs) that occurred from HCWs vs from patients are shown in stacked bar plots. Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant).
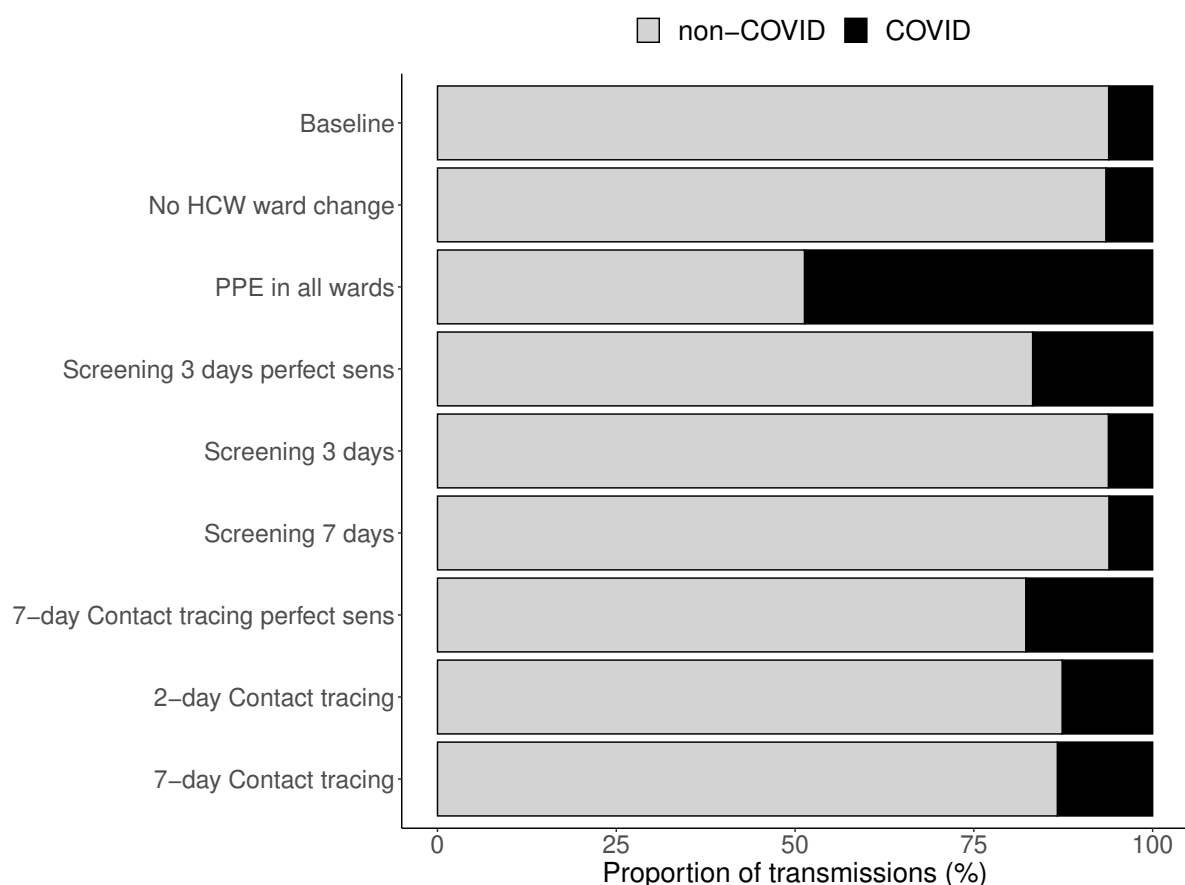
Figure S6: **Proportion of nosocomial transmissions in COVID- and non-COVID wards for each simulation scenario.** The mean percentages of nosocomial transmissions (averaged over 100 simulation runs) in COVID and non-COVID wards are shown in stacked bar plots. Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant).
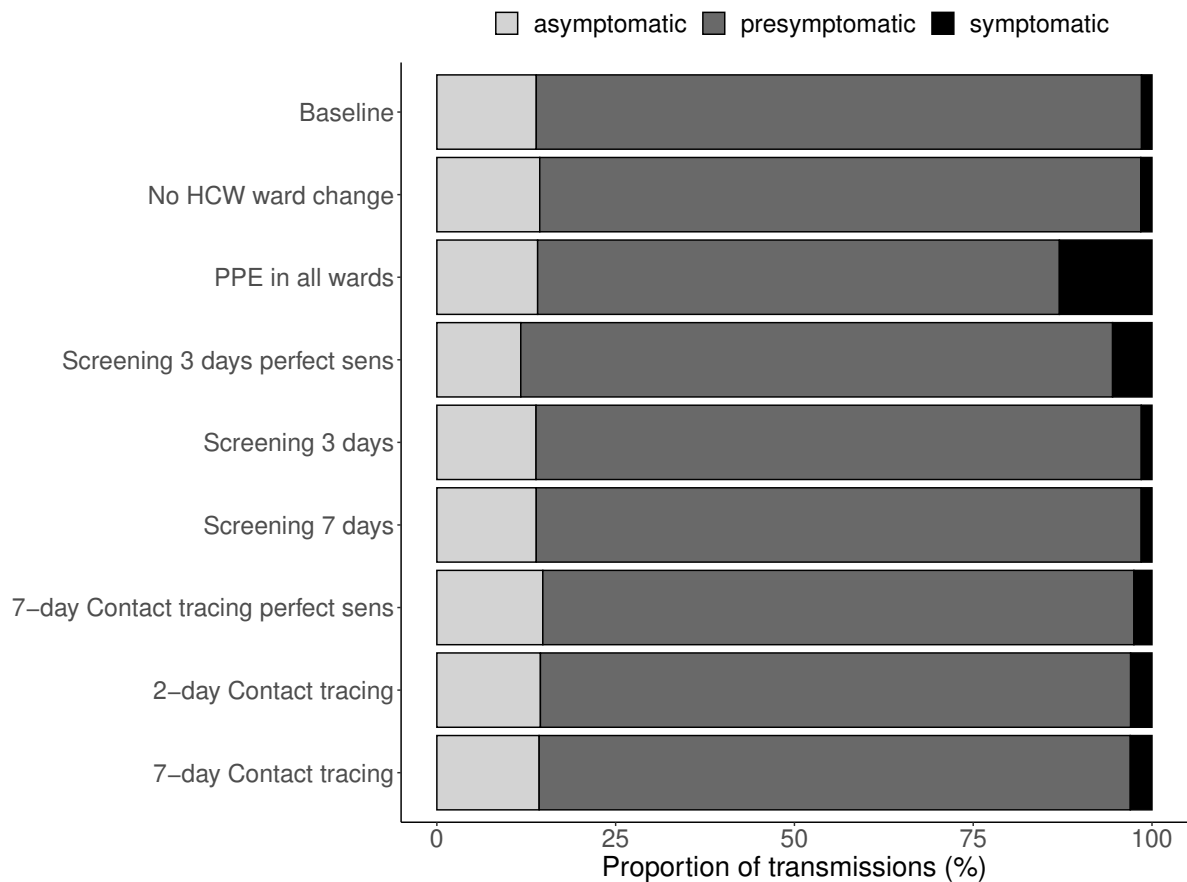
Figure S7: **Proportion of transmissions during different infection states for each simulation scenario.** The mean percentages of transmissions (averaged over 100 simulations) that occurred while the infected individual was in an asymptomatic, presymptomatic, or symptomatic state are shown in stacked bar plots. Results shown are based on $R_S = 1.95$ and $R_A = 0.8$ (reproduction numbers for the SARS-CoV-2 variant with 56% higher transmissibility with respect to the wild-type SARS-CoV-2 variant).

Discussion

Thi Mui Pham

In this thesis, we demonstrated how mathematical modeling can be a versatile and powerful tool to improve our understanding about the dynamics of infectious diseases as well as to explore a variety of methods to effectively control them. What follows is a summary of our most important findings, a brief discussion on what can be concluded, and suggestions for future research.

Coupling health-related preventive behavior and epidemic spread

The spread of an infectious disease depends not only on the biological characteristics of the pathogen, but also on the environment, and the behavior of the host population [1, 2]. In the absence of pharmaceutical interventions, the focus of public health response strategies aiming at containing the spread of an infectious disease lies on changing the behavioral patterns in a population. Understanding to which extent preventive behavior can influence infectious disease spread is, therefore, vital for informing public health policies. In Chapter 1 and 2 of this thesis, we presented two different approaches for modeling the impact of human preventive health behavior on the epidemic spread of infectious diseases. Both studies showed that 1) changes in health-related behavior during an epidemic can have important implications for infection control policies, 2) the effect of interventions may depend on the precise implementation and thus should be tailored for the situation at hand, and 3) reliable data on behavioral habits and changes of humans are generally scarce.

Changes in behavioral patterns in a population may be either imposed by the government or voluntarily undertaken as a reaction to an ongoing epidemic. In Chapter 1 [3], we compared the impact of these two approaches on the epidemic spread of SARS-CoV-2 in the general population. Additionally to different disease status, we explicitly included disease awareness and the subsequent voluntary uptake of preventive measures in the model, and coupled this behavioral response to disease prevalence. Our results highlight the vital role of disseminating evidence-based information about effective personal protective measures against SARS-CoV-2 transmission. Voluntary adoption of the latter as a reaction to high disease prevalence may be an effective strategy to mitigate and delay the COVID-19 epidemic, in particular in situations where government-imposed

measures are not accepted in the society or have to be avoided due to large societal and economic consequences. Furthermore, the rate at which these self-imposed measures are adopted determines whether a large epidemic could even be prevented rather than merely mitigated or delayed. Although previous work has emphasized that ignoring dynamic contact networks can lead to different predictions of the dynamics of disease spread and may have implications for assessing the success of infection control measures [4–8], the framework of behavioral change is incorporated in only few models for COVID-19, in particular in the beginning of the pandemic. Generally, (contact) behavior may be incorporated "exogenously" as a model parameter, or "endogenously" as a model variable [8]. In the first case, conducting longitudinal contact surveys and creating *observed* social contact matrices, as it has been done, e.g., in the CoMix study initially for the UK [9] and later extended to other European countries [10], are useful to inform mathematical models for infectious disease transmission without an explicit explanation of the behavior change. In the second case, the feedback loop between behavior itself and the disease dynamics are explicitly modelled in a mechanistic manner. Our study contributes to the endogenous representation of behavior change for COVID-19 transmission models. It serves as a theoretical motivation for increasing disease awareness as a measure of controlling the COVID-19 epidemic. Behavioral studies such as large-scale surveys on behavioral measures conducted by the RIVM, the Netherlands Municipal Public Health Services and Medical Assistance in Accidents and Disasters (GGD-GHOR) and the regional public health services (GGDs) [11], tracking the compliance to preventive measures such as quarantine and isolation, or staying 1.5 m apart as well as the change of behavioral patterns during an epidemic are insightful for assessing their impact. Due to the qualitative nature of such studies, direct translation of these results to inform model parameters for reactive behavioral responses during an ongoing epidemic requires careful consideration of the context and their generalizability remains challenging [12]. More research is needed to test crucial assumptions in behavior-disease models. For example, quantifying how disease awareness grows and wanes over time, the role of media on behavioral responses, and the actual interdependence on the epidemic itself would provide additional understanding on which model assumptions need refinement. Behavioral data obtained from digital media such as Twitter, Facebook, mobile phones, and search engines may be promising "new" sources that can be incorporated in disease-behavior models [8]. In particular, risk perception and corresponding change of

behavior may not be directly linked to disease prevalence but to media coverage of the epidemic [13]. However, approaches on how to overcome related challenges such as under-representation of certain age or socioeconomic groups [12], or search behavior triggered by widespread media coverage [14] remain an important field for future research.

The impact of self-imposed measures on disease spread crucially depends on their effectiveness. In Chapter 1, we quantified the threshold of efficacy necessary for self-imposed measures to be effective in preventing a large epidemic. Since our model was developed in the early stage of the pandemic when reliable data on the effectiveness of specific preventive measures were scarce, we explored the full range of efficacy values for all self-imposed measures. Meanwhile, evidence suggests a dominant role of aerosols in the spread of SARS-CoV-2 while contaminated surfaces rarely contributed to transmission [15–17]. In light of this new information, efficacy values for hand washing are likely to be at the lower end whereas the efficacy of face masks and their ability to interrupt SARS-CoV-2 transmission might be higher. However, high uncertainty around the true efficacy values of personal protective measures remain. In fact, the protective effect of self-imposed measures may depend not only on the efficacy of the method itself but also on how it is applied by the user. In Chapter 2 [18] of this thesis, we explored how hand hygiene behavior may be optimized to effectively reduce disease transmission. While the presented model was motivated by the COVID-19 pandemic, and key results were presented at SAGE to inform the policy response in the United Kingdom, the work has far wider relevance to the control of respiratory tract infections. We emphasized this broad relevance in the paper. Based on our results, providing hand sanitisers at places with a high risk of hand contamination (e.g., supermarkets, public transport, or work places) would greatly reduce infection risks, in particular for viruses that remain viable on hands for a long time. Thus, recommendations on hand hygiene measures should be tailored to the relevant context. This work may also serve as an impetus to test these hypotheses in empirical or experimental investigations such as clinical trials. Such empirical data may be useful for model fitting and estimation of parameters for which we had to make simplifying assumptions. For example, data on the rate of hand contamination events would decrease the uncertainty in our model assumptions and the need for extensive sensitivity analyses. The precise impact of personal protective measures, thus, depends on their implementation. It is determined by the rate at which the measures are adopted (as

shown in Chapter 1 of this thesis) and, for hand washing, by the frequency and moment at which it is applied, and the duration of persistence of the pathogen on hands.

Mathematical modeling and statistical inference of transmission routes

Understanding and quantifying the dynamics of nosocomial transmission is crucial for identifying and devising effective infection prevention and control measures. Mathematical modeling allows translating biological and clinical knowledge about the routes of transmission into a mathematical framework in which quantitative analyses can be performed using available data. These analyses can be useful to estimate transmission parameters and infer the relative contributions of the different transmission routes.

For *P. aeruginosa*, a significant role of environmental contamination in the transmission process has been suggested in several observational and experimental case studies [19–23]. However, the transmission dynamics of *P. aeruginosa* have not been modeled and the relative importance of environmental contamination to nosocomial transmission has not been quantified. In Chapter 3 and 4 of this thesis, we presented the first attempt to retrospectively estimate the role of environmental contamination for two different settings of *P. aeruginosa* transmission. While for both chapters we used the same Bayesian MCMC estimation method to infer transmission parameters and the relative contributions of the respective transmission routes, we investigated different aspects of environmental contamination in each study. In Chapter 3 [24], we explicitly model environmental contamination as a contribution of bacterial load to a general pool in the environment remaining after the discharge of patients. Using an extensive data set of *P. aeruginosa* surveillance from two ICUs of a French University hospital in Besançon over a time period of 17 years, we estimated that environmental contamination after the discharge of patients played only a minor role when compared to background and cross-transmission. Cross-transmission contributed to approximately half of the transmissions in these French ICUs (with background transmission making up the other half). In contrast, cross-transmission was only of minor importance in the two wards of the adult ICU of the Erasmus Medical Center in Rotterdam (as presented in Chapter 4). By exclusion, we concluded that persistent contamination from the environment was the

predominant route of transmission. In comparison with Chapter 3, we did not explicitly incorporate environmental contamination into the transmission model but inferred its contribution by exclusion. Thus, we demonstrated two different approaches on modeling environmental contamination as a route of nosocomial transmission of *P. aeruginosa* in ICUs and showed that the quantitative results for the relative importance of transmission routes depends on the specific setting. These settings need to be considered when conclusions for infection control are drawn.

While our transmission models were based on relatively simple compartmental stochastic models, more details such as hospital-specific interventions, contact patterns, heterogeneity in patient characteristics such as their demographics, and disease history could be included when using an agent-based modeling approach. In the absence of reliable data, the results of such detailed models strongly depend on the model assumptions and may not reflect realistic patterns. Detailed data on observed healthcare worker-patient contacts collected, e.g., via remote-based sensor networks, or on patient transfers within and between healthcare settings have been integrated in transmission models over recent years [25] and could help parameterize more complex models [25, 26]. However, the inclusion of such information would increase computational costs. Using Bayesian MCMC estimation methods, we inferred the relative contribution of transmission routes of *P. aeruginosa* based on statistical patterns in the prevalence over time. For such a method, we need longitudinal data over long time periods for the results to be informative but we can forego knowledge about individual transmission pathways. When exact transmission pathways (who-infected-whom) are of interest, routine data on genotyping, as increasingly collected in high income settings, will be highly valuable in linking transmission events, improving estimates, and reducing uncertainty bounds for transmission parameters [27, 28]. In lower income settings, these types of data and modeling studies of nosocomial transmission are scarce [26] and hence, modeling approaches as presented in Chapter 3 and 4 will be useful for tackling the problems of healthcare-associated infections.

Mathematical modeling for data interpretation and intervention strategies

The COVID-19 pandemic has highlighted challenges in infectious disease modeling that arise around data availability and reliability in the early stages of a pandemic [29]. Data collection during an ongoing epidemic is particularly challenging for healthcare institutions facing a high workload and the risk of being overwhelmed. Nevertheless, existing data bases may be linked and used to inform clinical management and hospital policy recommendations. In Chapters 5 and 6, we focus on modeling approaches addressing the problem of hospital-acquired SARS-CoV-2 infections using hospital data in the UK and in the Netherlands.

Based on experience with previous novel corona viruses [30], hospital-acquired infections have been recognized early on as a serious concern in the UK and discussed at SAGE 14 on 10th March 2020 [31]. In spring 2020, identification of infections acquired in English hospital facilities mainly relied on symptomatic testing (on admission or during hospital stay). A hospital-acquired infection was defined by a cutoff in days of symptom onset after admission. Symptomatic hospital-acquired infections may remain unidentified using this method due to short patient stays (in comparison with the incubation period) or due to a symptom onset prior to the cutoff. We accounted for these unidentified cases and provided the first national-level estimates of the total burden of symptomatic hospital-acquired infections and their contribution to the overall COVID-19 epidemic during the first wave in England in Chapter 5 [32]. We showed that defining a hospital-acquired COVID-19 case by a symptom onset cutoff of 7 days, would likely identify only 30% (range across 200 simulations: 20-41%) of symptomatic hospital-acquired infections. This finding suggests that under common definitions, as provided by the UK Health Security Agency [33] and the European Centre for Disease Prevention and Control (ECDC) [34], the number of hospital-acquired infections is significantly underestimated. Estimates of the proportion of nosocomial infections, e.g. presented in Read et al (2021) [35] for the UK that are based on symptom onset data, are therefore likely to be underestimates. We found, however, that the contribution of symptomatic hospital-acquired infections to the overall number of SARS-CoV-2 infections in England was likely small (less than 1%) but varied over time with an increasing importance in time periods with low community

prevalence. Our results imply that screening patients at hospital discharge, or quarantining discharged hospital patients may be effective in preventing further transmissions into the community. In fact, we estimated that approximately 40% of those symptomatic infections that would otherwise be missed could be detected by these measures. This is particularly relevant for low-resource settings that rely on symptom onset screening and where hospital stays are generally short. Our work contributed to understanding the magnitude of symptomatic nosocomial infections of SARS-CoV-2 and the importance of hospitals in sustaining the community epidemic, and was important for prioritizing infection control measures in the UK. The results were presented to SAGE on 22nd October 2020 [36] to inform infection control policy in hospitals in the UK for future epidemic waves.

Our analysis also illustrates how challenges arising from data sources that were not collected and hence not tailored for the data analysis itself could be addressed. To perform our analysis, we used two complementary data sources containing information on COVID-19 patients in NHS acute trusts (organisational units containing one or more acute care hospitals within the National Health Services of England and Wales). While CO-CIN [37] is an enrollment-based study including information on the date of symptom onset, it only represents a subset (albeit the majority) of COVID-19 patients hospitalized in NHS acute trusts. The SUS dataset [38] contains information on all patient admissions and discharges for all NHS acute trusts, but lacks symptom onset information. To overcome these issues, we merged the two data sets into one database integrating the best features of both. Since none of the two data sources included sufficient information about non-COVID-19 patients, we employed a simulation analysis rather than a full transmission model. If data on regular inpatient testing, whole genome sequences, or HCW-patient and patient-to-patient contact data was available, more complex models could be developed. In particular, genomic sequencing may assist hospital outbreak investigations and provide further evidence for specific transmission pathways for SARS-CoV-2, but the use of genomic data has its own limitations such as limited availability of samples, difficulty of producing sequences for low viral loads, or insufficient genetic variation [39]. Integrating epidemiological, contact network, and genomic datasets into probabilistic inference procedures (e.g., Bayesian data-augmented MCMC) could be used to infer the relative contribution of hospital-acquired infections, or the transmission parameters and pathways between HCWs and patients, or between patients, while accounting for incomplete data.

Since SARS-CoV-2 is likely to persist as an endemic seasonal virus in coming years, it is not only critical to understand its role in hospital transmission, but also how their burden can be minimized. We assessed the effectiveness of various hospital-based interventions in controlling nosocomial transmission of SARS-CoV-2 in Chapter 6 [40]. For the evaluation of interventions in hospital populations the effect of stochasticity is often non-negligible and compartmental models may be insufficient in capturing essential features. In these cases, contact networks and the individual's time since infection might influence the impact of certain measures and interventions. We showed how a detailed agent-based model can be applied to address these issues. Equipped with data on the hospital structure of the UMC Utrecht and the number of occupied beds of COVID-19 patients during the first wave in the Netherlands, we were able to compare the effect of several hospital-based interventions targeting HCWs on the nosocomial transmission of SARS-CoV-2. Our results showed that the use of highly effective PPE in all hospital wards was the most effective intervention due to a large proportion of asymptomatic and presymptomatic transmission. We also highlighted the importance of contact tracing beyond community settings, especially since it can achieve higher test positivity rates (proportion of infected among tested individuals) than regular screening interventions. Our conclusions focus on the relative effectiveness of hospital-based intervention strategies and are informative for future waves of the COVID-19 pandemic and outbreaks of newly emerging viruses similar to SARS-CoV-2. Our model could be improved if more detailed data were available, such as the combination of admission and discharge, symptom onset and testing data (and the reason for the test), genomic sequencing data, ward and ward transfer data for patients as well as for HCWs. This would allow the application of proper model fitting procedures and estimating transmission parameters and the within-hospital reproduction number similarly to what has been done for a long-term care facility in Paris, France [41]. In addition, the model would mimic the transmission dynamics in hospitals more realistically and allow a more precise estimation of the impact of the modeled interventions.

Both studies emphasize the need to invest in sharing and standardizing data in a sustainable and scalable way to improve future epidemic response, especially in the healthcare sector [29]. Proactively designing hospital databases that balance patient privacy and the usability for public health data analyses, and that are standardized to be used on a

national or international level will be crucial for effectively preparing and responding to future epidemics [29].

Conclusion

We presented a diverse set of mathematical models addressing various research questions, ranging from evaluating interventions and preventive measures for the newly emerged SARS-CoV-2 to understanding the transmission dynamics of common pathogens, such as *P. aeruginosa*, in hospital settings. Overall, this thesis provided new insights into how biological and clinical research problems may be translated into a mathematical or statistical framework, and how the level of detail and the model choice highly depends on the objective, but also on the availability of data. As mathematical modelers of infectious diseases, we face the trade-off between a simpler model that is easy to understand and to analyze, or following a more complex approach leading to a more realistic representation of the system, but that requires, in exchange, more intricate analysis tools and detailed input. If problems need to be addressed on a population level, simpler models with an aggregated structure (such as compartmental models) maybe be sufficient to capture the relevant transmission dynamics. Simple statistical or simulation models linked with appropriate data may provide first quantitative evidence without the need of a full mechanistic transmission model. To account for stochastic effects in smaller populations such as hospitals, more complex models such as agent-based models may be more appropriate but also need to be equipped with more detailed data. We identified gaps in data collection in particular on behavioral responses and hospital patient data that need to be addressed to improve infectious disease model building and to reduce the reliance on arbitrary assumptions. Model-informed data collection and the synthesis of multiple data sources will hugely improve epidemic response, but sustainable data sharing and thus the collaboration of many different research groups, disciplines, and countries will be essential [29]. Infectious diseases know neither boundaries nor borders. Preparing and responding to inevitable future disease epidemics will require us to similarly transcend disciplinary boundaries.

References

[1] Hollingsworth, TD. Controlling infectious disease outbreaks: Lessons from mathematical modelling. In: *Journal of Public Health Policy* 30.3 (Oct. 2009), pp. 328–341. doi: `10.1057/jphp.2009.13`.

[2] Cevik, M and Baral, SD. *Networks of SARS-CoV-2 transmission*. July 2021. doi: `10.1126/science.abg0842`.

[3] Teslya, A, Pham, TM, Godijk, NG, et al. Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study. In: *PLOS Medicine* 17.7 (July 2020), e1003166. doi: `10.1371/JOURNAL.PMED.1003166`.

[4] Ferguson, N. Capturing human behaviour. In: *Nature* 446.7137 (Apr. 2007), p. 733. doi: `10.1038/446733a`.

[5] Funk, S, Gilad, E, Watkins, C, et al. The spread of awareness and its impact on epidemic outbreaks. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.16 (Apr. 2009), pp. 6872–6877. doi: `10.1073/pnas.0810762106`.

[6] Fenichel, EP, Castillo-Chavez, C, Ceddia, MG, et al. Adaptive human behavior in epidemiological models. In: *Proceedings of the National Academy of Sciences* 108.15 (Apr. 2011), pp. 6306–6311. doi: `10.1073/PNAS.1011250108`.

[7] Perra, N, Balcan, D, Gonçalves, B, et al. Towards a characterization of behavior-disease models. In: *PLoS ONE* 6.8 (2011), e23084. doi: `10.1371/journal.pone.0023084`.

[8] Funk, S, Bansal, S, Bauch, CT, et al. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. In: *Epidemics* 10 (Mar. 2015), pp. 21–25. doi: `10.1016/j.epidem.2014.09.005`.

[9] Davies, NG, Kucharski, AJ, Eggo, RM, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. In: *The Lancet Public Health* 5.7 (July 2020), e375–e385. doi: `10.1016/S2468-2667(20)30133-X`.

[10] Verelst, F, Hermans, L, Vercruysse, S, et al. SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries. In: *BMC Medicine 2021 19:1* 19.1 (Sept. 2021), pp. 1–7. doi: `10.1186/S12916-021-02133-Y`.

[11] RIVM. *Applying behavioural science to COVID-19: Study on behavioural measures and well-being*. 2020. `https://www.rivm.nl/en/coronavirus-covid-19/research/behaviour`.

[12] Eggo, RM, Dawa, J, Kucharski, AJ, et al. The importance of local context in COVID-19 models. In: *Nature Computational Science* 1.1 (Jan. 2021), pp. 6–8. doi: `10.1038/s43588-020-00014-7`.

[13] Gozzi, N, Tizzani, M, Starnini, M, et al. Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. In: *Journal of Medical Internet Research* 22.10 (Oct. 2020), e21597. doi: `10.2196/21597`.

[14] Butler, D. When Google got flu wrong. In: *Nature* 494.7436 (Feb. 2013), pp. 155–156. doi: `10.1038/494155A`.

[15] Greenhalgh, T, Jimenez, JL, Prather, KA, et al. *Ten scientific reasons in support of airborne transmission of SARS-CoV-2.* May 2021. doi: `10.1016/S0140-6736(21)00869-2`.

[16] Heneghan, CJ, Spencer, EA, Brassey, J, et al. SARS-CoV-2 and the role of airborne transmission: a systematic review. In: *F1000Research* 10 (Mar. 2021), p. 232. doi: `10.12688/f1000research.52091.1`.

[17] Goldman, E. Exaggerated risk of transmission of COVID-19 by fomites. In: *The Lancet Infectious Diseases* 20.8 (Aug. 2020), pp. 892–893. doi: `10.1016/S1473-3099(20)30561-2`.

[18] Pham, TM, Mo, Y, and Cooper, BS. The Potential Impact of Intensified Community Hand Hygiene Interventions on Respiratory tract Infections: A Modelling Study. In: *medRxiv* (May 2020), p. 2020.05.26.20113464. doi: `10.1101/2020.05.26.20113464`.

[19] Catho, G, Martischang, R, Boroli, F, et al. Outbreak of Pseudomonas aeruginosa producing VIM carbapenemase in an intensive care unit and its termination by implementation of waterless patient care. In: *Critical Care 2021 25:1* 25.1 (Aug. 2021), pp. 1–10. doi: `10.1186/S13054-021-03726-Y`.

[20] Geyter, DD, Vanstokstraeten, R, Crombé, F, et al. Sink drains as reservoirs of VIM-2 metallo-β-lactamase-producing Pseudomonas aeruginosa in a Belgian intensive care unit: relation to patients investigated by whole-genome sequencing. In: *Journal of Hospital Infection* 115 (Sept. 2021), pp. 75–82. doi: `10.1016/J.JHIN.2021.05.010`.

[21] Knoester, M, Boer, Md, Maarleveld, J, et al. An integrated approach to control a prolonged outbreak of multidrug-resistant Pseudomonas aeruginosa in an intensive care unit. In: *Clinical Microbiology and Infection* 20.4 (Apr. 2014),

O207–O215. doi: `10.1111/1469-0691.12372`.

[22] Kizny Gordon, AE, Mathers, AJ, Cheong, EYL, et al. The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections—A Systematic Review of the Literature. In: *Clinical Infectious Diseases* 64.10 (May 2017), pp. 1435–1444. doi: `10.1093/CID/CIX132`.

[23] Voor in 't holt, AF, Severin, JA, Lesaffre, EM, et al. A Systematic Review and Meta-Analyses Show that Carbapenem Use and Medical Devices Are the Leading Risk Factors for Carbapenem-Resistant Pseudomonas aeruginosa. In: *Antimicrobial Agents and Chemotherapy* 58.5 (2014), pp. 2626–2637. doi: `10.1128/AAC.01758-13`.

[24] Pham, TM, Kretzschmar, M, Bertrand, X, et al. Tracking Pseudomonas aeruginosa transmissions due to environmental contamination after discharge in ICUs using mathematical models. In: *PLoS Computational Biology* 15.8 (2019), e1006697. doi: `10.1371/journal.pcbi.1006697`.

[25] Assab, R, Nekkab, N, Crépey, P, et al. Mathematical models of infection transmission in healthcare settings: Recent advances from the use of network structured data. In: *Current Opinion in Infectious Diseases* 30.4 (Aug. 2017), pp. 410–418. doi: `10.1097/QCO.0000000000000390`.

[26] Kleef, E van, Robotham, JV, Jit, M, et al. Modelling the transmission of healthcare associated infections: A systematic review. In: *BMC Infectious Diseases* 13.1 (June 2013), pp. 1–13. doi: `10.1186/1471-2334-13-294`.

[27] Worby, CJ, Lipsitch, M, and Hanage, WP. Shared Genomic Variants: Identification of Transmission Routes Using Pathogen Deep-Sequence Data.

In: *American Journal of Epidemiology* 186.10 (Nov. 2017), pp. 1209–1216. doi: `10.1093/aje/kwx182`.

[28] De Maio, N, Worby, CJ, Wilson, DJ, et al. Bayesian reconstruction of transmission within outbreaks using genomic variants. In: *PLoS Computational Biology* 14.4 (Apr. 2018), e1006117. doi: `10.1371/journal.pcbi.1006117`.

[29] Kucharski, AJ, Hodcroft, EB, and Kraemer, MUG. Sharing, synthesis and sustainability of data analysis for epidemic preparedness in Europe. In: *The Lancet Regional Health - Europe* 9 (Oct. 2021), p. 100215. doi: `10.1016/j.lanepe.2021.100215`.

[30] De Wit, E, Van Doremalen, N, Falzarano, D, et al. *SARS and MERS: Recent insights into emerging coronaviruses*. June 2016. doi: `10.1038/nrmicro.2016.81`.

[31] Public Health England. *SAGE 14 minutes: Coronavirus (COVID-19) response, 10 March 2020 - GOV.UK*. 2020. `https://www.gov.uk/government/publications/sage-minutes-coronavirus-covid-19-response-10-march-2020`.

[32] Knight, GM, Pham, TM, Stimson, J, et al. The contribution of hospital-acquired infections to the COVID-19 epidemic in England in the first half of 2020. In: *medRxiv* (Sept. 2021), p. 2021.09.02.21262480. doi: `10.1101/2021.09.02.21262480`.

[33] Public Health England. *COVID-19: epidemiological definitions of outbreaks and clusters in particular settings*. 2020. `https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters`.

[34] European Centre for Disease Prevention and Control. *Surveillance definitions for COVID-19*. 2021. `https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions`.

[35] Read, JM, Green, CA, Harrison, EM, et al. Hospital-acquired SARS-CoV-2 infection in the UK's first COVID-19 pandemic wave. In: *The Lancet* 398.10305 (Sept. 2021), pp. 1037–1038. doi: `10.1016/S0140-6736(21)01786-4`.

[36] SAGE. *Sixty-third SAGE meeting on Covid-19, 22nd October 2020*. 2020. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/935103/sage-63-meeting-covid-19-s0842.pdf`.

[37] Docherty, AB, Harrison, EM, Green, CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. In: *The BMJ* 369 (May 2020). doi: `10.1136/bmj.m1985`.

[38] NHS Digital. *Secondary Uses Service (SUS)*. 2020. `https://digital.nhs.uk/services/secondary-uses-service-sus`.

[39] Lumley, SF, Constantinides, B, Sanderson, N, et al. Epidemiological data and genome sequencing reveals that nosocomial transmission of SARS-CoV-2 is underestimated and mostly mediated by a small number of highly infectious individuals. In: *Journal of Infection* 83.4 (Oct. 2021), pp. 473–482. doi: `10.1016/j.jinf.2021.07.034`.

[40] Pham, TM, Tahir, H, Wijgert, JH van de, et al. Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study. In: *BMC Medicine* 19.1 (Dec. 2021), pp. 1–16. doi: `10.1186/s12916-021-02060-y`.

[41] Shirreff, G, Zahar, JR, Cauchemez, S, et al. How well does SARS-CoV-2 spread in hospitals? In: *medRxiv* (Sept. 2021), p. 2021.09.28.21264066. doi: `10.1101/2021.09.28.21264066`.

APPENDICES

Abbreviations

Brief summaries in English, Dutch, German and Vietnamese

Dutch summary/Nederlandse samenvatting

Contributing authors

List of publications

Acknowledgements

About the author

Acronyms

A. baumannii *Acinetobacter baumannii*.

P. aeruginosa *Pseudomonas aeruginosa*.

COVID-19 Coronavirus Disease 2019.

HCWs healthcare workers.

ICU Intensive care unit.

LTCF Long-term care facilities.

MCMC Markov chain Monte Carlo.

MRSA methicillin-resistant *Staphylococcus aureus*.

NPIs non-pharmaceutical interventions.

RIVM Rijksinstituut voor Volksgezondheid en Milieu (Dutch National Institute for Public Health and the Environment).

RT-PCR Reverse transcriptase polymerase chain reaction.

SAGE Scientific Advisory Group for Emergencies.

SARS-CoV-2 *severe acute respiratory syndrome coronavirus 2*.

SDD selective digestive tract decontamination.

VIM Verona Integron-encoded Metallo-beta-lactamase.

VRE vancomycin-resistant *Enterococci* (VRE).

WHO World Health Organization.

English summary

In this thesis, we elucidate the important role of mathematical models for infectious disease control with a focus on the transmission of SARS-CoV-2 in the community and in hospital settings, as well as the transmission of *Pseudomonas aeruginosa* in intensive care units.

Most important findings

- Self-imposed measures (hand washing, mask-wearing, physical distancing) as a reaction to an ongoing COVID-19 epidemic can be effective in mitigating and delaying it. They can even prevent a large epidemic if they are adopted fast and if the measure has a high efficacy.

- Hand hygiene may be optimized by the frequency and the moment at which it is applied to reduce viral respiratory tract infections. We found that immediate hand washing after contamination is consistently more effective than at fixed time intervals.

- Many hospital-acquired COVID-19 cases may remain unidentified if their definition relies on a cutoff of days after symptom onset: Patients may be discharged or have a symptom onset prior to this cutoff. Screening or quarantining hospital patients at discharge may be effective in preventing further transmissions into the community.

- In our simulation study, the use of highly effective personal protective equipment in all hospital wards was the most effective intervention to limit SARS-CoV-2 transmission in hospitals. We highlighted the role of contact tracing beyond community settings since it can detect more infections among tested individuals than regular screening interventions.

- We modeled environmental contamination as a transmission route for *Pseudomonas aeruginosa* and showed that its contribution depends on the specific setting.

Nederlandse samenvatting

In dit proefschrift lichten we de belangrijke rol toe van wiskundige modellen voor de beheersing van infectieziekten, met een focus op de overdracht van SARS-CoV-2 in de samenleving en in ziekenhuisomgevingen en van *Pseudomonas aeruginosa* in intensive care-afdelingen.

Belangrijkste bevindingen

- Zelfopgelegde maatregelen (handen wassen, gezichtsmaskers dragen, fysiekafstand houden) als reactie op een aanhoudende COVID-19-epidemie kunnen effectief zijn om deze te verkleinen en vertragen. Ze kunnen zelfs een grote epidemie voorkomen als de maatregelen voldoender effectiviteit zijn en snel geïmplementeerd worden.

- De frequentie en het moment van handhygiëne kan worden geoptimaliseerd om virale luchtweginfecties te verminderen. We ontdekten dat onmiddellijk handen wassen na besmetting consequent effectiever is dan op vaste tijdsintervallen.

- Veel in het ziekenhuis opgelopen COVID-19-gevallen kunnen ongedetecteerd blijven als de definitie van nosocomiale verspreiding gebaseerd is op een cut-off van dagen na het begin van de symptomen: patiënten kunnen voorafgaand aan de cut-off ontslagen of symptomatisch worden. Het screenen of in quarantaine plaatsen van ziekenhuispatiënten bij ontslag kan effectief zijn om verdere overdracht naar de samenleving te voorkomen.

- In onze simulatiestudie was het gebruik van zeer effectieve persoonlijke beschermingsmiddelen op alle ziekenhuisafdelingen de meest effectieve interventie om de transmissiet van SARS-CoV-2 in ziekenhuizen te beperken. We benadrukken de rol van bron- en contactonderzoek, juist ook in ziekenhuizen, omdat het meer infecties bij geteste personen kan detecteren dan een screeningstrategie op gezette tijden.

- We hebben besmettingen vanuit de omgeving gemodelleerd als een transmissieroute voor *Pseudomonas aeruginosa* en hebben aangetoond dat de bijdrage ervan afhangt van de specifieke setting.

Deutsche Zusammenfassung

In dieser Arbeit beleuchten wir die wichtige Rolle mathematischer Modelle für die Kontrolle von Infektionskrankheiten mit einem Fokus auf die Übertragung von SARS-CoV-2 im normalen gesellschaftlichen Umgang in der Bevölkerung und im Krankenhaus sowie die Übertragung von *Pseudomonas aeruginosa* auf Intensivstationen.

Wichtigste Erkenntnisse

- Selbst auferlegte Maßnahmen (Händewaschen, Tragen von Masken, soziale Distanzierung) als Reaktion auf eine anhaltende COVID-19 Epidemie können den Höhepunkt der Epidemie verzögern, die Epidemiekurve abflachen und die Zahl der Neuerkrankungen verringern. Ein große Epidemie kann sogar vollständig verhindert werden, wenn die Präventionsmaßnahmen schnell ergriffen werden und sie eine hohe Wirksamkeit aufweisen.

- Die Effektivität von Handehygiene kann durch die Häufigkeit und den Zeitpunkt der Anwendung optimiert werden, um virale Atemwegsinfektionen zu reduzieren. Wir haben festgestellt, dass sofortiges Händewaschen nach einer Kontamination durchweg effektiver ist als das Händewaschen in festen regelmäßigen Zeitabständen.

- Infektionen werden gewöhnlich mittels eines *Cut-offs* (Zeitdauer von der Aufnahme einesd Patienten in das Krankenhaus bis zum Auftreten der ersten Symptome der Infektion) als im Krankenhaus erworben klassifiziert ("nosokomial"). Bei SARS-CoV-2 Infektionen bleiben dadurch möglicherweise viele nosokomiale Infektionen unidentifiziert: Patienten könnten bereits vor Symptombeginn entlassen werden oder ihre Symptome könnten vor dem Cut-off auftreten und dadurch falsch klassifiziert werden. Das Screening oder die Quarantäne von Krankenhauspatienten bei der Entlassung kann effektiv sein, um weitere Übertragungen im normalen gesellschaftlichen Umgang zu verhindern.

- In unserer Simulationsstudie war der Einsatz von hochwirksamer persönlicher Schutzausrüstung auf allen Krankenhausstationen die effektivste Intervention, um die Übertragung von SARS-CoV-2 in Krankenhäusern einzudämmen. Die umfassende Rückverfolgung von Kontakten kann speziell in Krankenhäusern

ebenfalls eine wichtige Rolle spielen, vor allem da der Anteil der detektierten Infektionen bei getesteten Personen höher ist als beim regelmäßigen Screening und sie damit eine effizientere Maßnahme darstellt.

- Wir haben mathematische Modelle entwickelt um die Übertragung von *Pseudomonas aeruginosa* über Krankenhausumgebungen zu quantifizieren. Unsere Untersuchungen zeigten, dass *Pseudomonas aeruginosa* durchaus durch Erregerreservoire in der Krankenhausumgebung übertragen werden kann. Allerdings hängt der exakte Beitrag von der spezifischen Umgebung des Krankhauses ab.

Tóm tắt tiếng Việt ngắn gọn

Trong luận án này, chúng tôi phân tích vai trò quan trọng của các mô hình toán học trong việc kiểm soát bệnh truyền nhiễm; với trọng tâm là sự lây truyền của SARS-CoV-2 trong cộng đồng và trong môi trường bệnh viện, và sự lây truyền của *Pseudomonas aeruginosa* trong các đơn vị chăm sóc đặc biệt.

Những phát hiện quan trọng nhất

- Các biện pháp cá nhân (như rửa tay, đeo khẩu trang, giữ khoảng cách) để ứng phó với dịch COVID-19 có thể có hiệu quả trong việc giảm thiểu và trì hoãn lây truyền dịch bệnh. Các biện pháp này thậm chí có thể ngăn chặn đại dịch nếu được áp dụng nhanh chóng và khi biện pháp có hiệu quả cao.

- Việc rửa tay (bằng nước, cồn, xà phòng, dung dịch rửa tay khô) có thể được tối ưu hóa theo tần suất và thời điểm áp dụng để giảm thiểu nhiễm trùng đường hô hấp do vi rút. Chúng tôi nhận thấy rằng rửa tay ngay lập tức sau mỗi lần tay nhiễm bẩn tạo hiệu quả cao hơn so với rửa tay sau mỗi khoảng thời gian cố định.

- Nhiều trường hợp nhiễm COVID-19 từ bệnh viện khó được xác định nếu chỉ dựa vào số ngày giữa ngày phát triệu chứng và ngày nhập viện. Bệnh nhân có thể được xuất viện trước khi phát hiện triệu chứng hoặc có triệu chứng trước số ngày giới hạn. Việc sàng lọc hoặc cách ly bệnh nhân ngay khi xuất viện có thể hiệu quả trong việc ngăn ngừa lây truyền thêm vào cộng đồng.

- Trong nghiên cứu mô phỏng của chúng tôi, việc sử dụng thiết bị bảo hộ cá nhân hiệu quả cao ở tất cả các khu bệnh viện là biện pháp can thiệp hiệu quả nhất để hạn chế lây truyền SARS-CoV-2 trong bệnh viện. Chúng tôi nhấn mạnh vai trò của việc theo dõi tiếp xúc ở môi trường bệnh viện vì tỷ lệ nhiễm virut trong trường hợp này có thể cao hơn tỷ lệ nhiễm trong cộng đồng thông qua sàng lọc ngẫu nhiên.

- Khi tìm hiểu vai trò của nhiễm khuẩn dụng cụ trong bệnh viện trong việc lây truyền của *Pseudomonas aeruginosa*, kết quả mô hình cho thấy rằng sự đóng góp của nó phụ thuộc vào bối cảnh cụ thể.

Nederlandse wetenschappelijke samenvatting

Van de zwarte dood tot de nog voortgaande coronavirus pandemie, infectieziekten zijn altijd onder ons geweest. Ondanks een wereldwijde afname in doden door overdraagbare aandoeningen zoals in 2019 gerapporteerd door de Wereldgezondheidsorganisatie (WHO), behoren lagere luchtweginfecties en diarree-gerelateerde ziekten nog steeds tot de top tien doodsoorzaken ter wereld [1]. Met name in lage- en middeninkomenslanden blijven zij de belangrijkste oorzaak van ziekte en sterfte. Zes van de top tien doodsoorzaken in arme landen zijn overdraagbare ziekten [1]. Bovendien worden infecties veroorzaakt door antibioticaresistente bacteriën over het algemeen gezien als een groot risico voor de volksgezondheid [2–4]. In 2018 publiceerde de WHO een prioriteitenlijst voor onderzoek naar en ontwikkeling van nieuwe antibiotica voor gebruik tegen antibiotica-resistente bacteriën [5]. Naast dat er prioriteit gegeven moet worden aan onderzoek naar de multiresistente *Mycobacterium tuberculosis*, zijn multiresistente Gram-negatieve bacteriën, met name *Acinetobacter baumanni*, *Pseudomonas aeruginosa* (*P. aeruginosa*) and *Enterobacteriaceae* ook erg belangrijk voor toekomstig onderzoek.

Het belang van een grootschalige georganiseerde reactie op de infectieziektenproblematiek is recent weer duidelijk geworden, door het ontstaan van het *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2), het virus dat de coronavirus 2019 (COVID-19) pandemie veroorzaakt. Beleidsmakers op het gebied van de volksgezondheid moesten plotseling snel de meest effectieve interventie-strategieën vinden. Echter, om het gekozen beleid te verantwoorden zijn gevalideerde methoden nodig die de effectiviteit van eventuele interventiestrategieën bepalen. Wiskundig modelleren is een essentieel onderdeel geworden om de effectiviteit van interventiestrategieën te bepalen, met name tijdens de huidige COVID-19 pandemie.

In dit proefschrift lichten wij de belangrijke rol van wiskundig modelleren toe voor het bepalen van effectieve interventiestrategieën voor infectieziekten, met een focus op de overdracht van zowel SARS-CoV-2 en de nosocomiale transmissie van *P. aeruginosa*. Wij kijken zowel naar de transmissie in de samenleving als naar de transmissie in een ziekenhuis. Deze samenvatting geeft een kort overzicht van de wiskundige modellen van infectieziekten die in dit proefschrift beschreven staan alsook van de maatregelen voor

het beheersen ervan. Tot slot laten wij zien hoe de modellen en maatregelen verbonden zijn en geven wij een overzicht van onze belangrijkste bevindingen.

Wiskundig modelleren van infectieziekten in een notendop

Een wiskundig model kan gebruikt worden als middel om het dynamische gedrag van infectieziekten te begrijpen en kwantificeren. Een wiskundig model heeft drie doelen: 1. het *begrijpen* van de verspreiding van de ziekte; 2. het *voorspellen* van het vervolg van een uitbraak; en 3. het *evalueren* van maatregelen om de verspreiding van de ziekte te beheersen. Wij kunnen onze aannames over biologische processen vertalen naar een wiskundig model door het model te vergelijken met waarnemingen uit de werkelijkheid. Door deze vergelijking kunnen wij ons begrip van de epidemiologie van de ziekte toetsen. Wiskundige modellen complementeren traditionele experimentele aanpakken, in het bijzonder wanneer experimentele manipulatie van het te bestuderen systeem niet mogelijk is, zoals het geval is gedurende de uitbraak van infectieziekten [6].

Een klassiek voorbeeld van een model van een infectieuze ziekte is het **compartimenten SIR model**, waar voor elk punt in de tijd, t, de bevolking wordt verdeeld over de groepen vatbar $S(t)$, infectieus $I(t)$ en hersteld $R(t)$. Met hersteld bedoelen wij een individu die de ziekte heeft gehad en geen verspreider meer is. Deze modellen worden vaak gebruikt om de dynamiek van infectieziekten in een grote populatie te onderzoeken. Voor het begrijpen van kleinschalige uitbraken zoals die voorkomen in scholen of ziekenhuizen, denken wij dat een complexer model beter geschikt is, zeker omdat toevalsprocessen dan relevant zijn. Zo zijn **Agent-based modellen** de afgelopen jaren een populaire toevoeging geworden aan bestaande methodes [7, 8]. Een definiërende eigenschap is om het individu centraal in het model te zetten en hun bijbehorende eigenschappen en interacties te volgen. Dit stelt ons niet alleen in staat om gedrag op het niveau van het individu waar te nemen, maar ook om realistischere contact netwerken in het model te verwerken.

Een belangrijk deel van het ontwikkelen van onze modellen is het kalibreren van de parameters van het model, d.w.z., het bepalen de parameters waarbij het model de waarnemingen uit de praktijk zo goed mogelijk reproduceert. Model parameters kunnen worden gekozen op basis van 1) inschattingen door experts, 2) observaties of c) uit de literatuur.

Er bestaan statistische methoden om modelresultaten te fitten op observaties uit de praktijk. Als modelresultaten niet direct geverifieerd kunnen worden door een tekort aan data, dan kan een gevoeligheidsanalyses via een parameter-ruimte exploratie helpen om de belangrijke parameters te vinden en om model-voorspellingen kwalitatief te verifiëren.

Het beheersen van de overdracht van infectieziekten

Een van de belangrijkste parameters om de overdracht van infectieuze ziekten te meten is het basis reproductiegetal R_0, gedefinieerd als het gemiddelde aantal infecties dat een enkel besmet persoon kan veroorzaken gedurende zijn of haar besmettelijke periode, in een bevolking die volledig vatbaar is. Het effectieve reproductie getal, R_E, is van toepassing op een bevolking die deels immuun is voor de ziekte of waar overdrachtsbeperkende maatregelen zijn genomen. Zulke maatregelen hebben als doel om R_E onder 1 te krijgen, opdat op den duur de ziekte verdwijnt.

Niet-farmaceutische interventies

In dit proefschrift richting wij ons op niet-farmaceutische interventies: maatregelen voor de volksgezondheid die als doel hebben de overdracht van een ziekte te verminderen zonder medicatie te gebruiken.

Het **wassen van handen** met zeep en water, of met op alcohol gebaseerde handreinigingsmiddelen, vermindert ziekteoverdracht.

Gezichtsmaskers verminderen de transmissie door de lucht van een pathogeen.

Fysiek afstand houden vermindert het aantal contacten waardoor de ziekte zich kan verspreiden. Overheden kunnen "social distancing" opleggen door maatregelen als de bevolking verplichten thuis te blijven, het beperken van het toegestane aantal bezoekers op een locatie, of het sluiten van scholen. Individuen kunnen er ook voor kiezen om zelf afstand te houden tot anderen.

Quarantaine en isolatie voorkomt dat potentieel besmettelijke personen zich mengen met de rest van de bevolking en verminderen dus hun bijdrage aan de verspreiding van de ziekte. *Quarantaine* verwijst naar het beperken van de bewegingsvrijheid van potentieel

besmettelijke personen, terwijl *isolatie* verwijst naar het uit elkaar houden van besmette en niet besmette personen.

Regelmatig testen is een vorm van testen waarbij individuen op regelmatige tijdsintervallen getest worden, onafhankelijk van symptomen. Deze maatregel zelf heeft niet direct impact op transmissie, maar als gedetecteerde besmette personen gebruikmaken van transmissie beperkende maatregelen, dan kunnen transmissieketens doorbroken worden.

Bron- en contactonderzoek is een techniek die op basis van de geschiedenis van contacten tussen mensen, en de kennis dat bepaalde personen besmet zijn, probeert te achterhalen wie er nog meer besmet is. Besmette individuen worden geïnterviewd door volksgezondheid medewerkers om te achterhalen door wie ze besmet zijn, en wie ze nog meer besmet kunnen hebben. Isolatie of quarantaine wordt dan ofwel opgelegd of geadviseerd aan de besmette individuen. Als het bron- en contactonderzoek voldoende snel gedaan wordt, dan kan het secundaire transmissie door waarschijnlijk besmette individuen voorkomen.

Belangrijkste bevindingen van dit proefschrift

Het koppelen van gezondheidsgerelateerd preventief gedrag en epidemische verspreiding

In hoofdstuk 1 ontwikkelden wij een deterministisch gecompartimentaliseerd model dat de biologische ziekteprogressie koppelt aan het gezondheidsgerelateerd gedrag van mensen. Wij waren benieuwd naar de vrijwillige preventie maatregelen, zoals fysiek afstand houden, handen wassen, en het gebruik van gezichtsmaskers, als reactie op de voortgaande COVID-19 epidemie, de epidemie kan beïnvloeden. Deze maatregelen werden vergeleken met een overheidsmaatregel die eenmalig "social distancing" houden oplegde gedurende een korte periode. Onze resultaten laten zien dat vrijwillige maatregelen de piek van epidemie kunnen vertragen en verlagen, en dat deze de grootte van de uitbraak kunnen verlagen. Echter, de impact van deze maatregelen hangt sterk af van de mate waarin ze worden geadopteerd en hoe effectief ze zijn. Een grote epidemie kan worden voorkomen als maatregelen uitgeoefend worden door meer dan 90% van de bevolking met een

effectiviteit van 50%. Korte-termijn maatregelen die fysieke afstand creëren en door de overheid worden opgelegd, kunnen de de epidemische piek vertragen, maar verminderen de hoogte van de piek en de grootte van de uitbraak nauwelijks. Zo'n interventie is het meest wenselijk wanneer tijdrekken zinvol is, bijvoorbeeld omdat een vaccin ontwikkeld wordt of omdat zorginstellingen zich nog moeten voorbereiden. Wij raden overheden en volksgezondheidsinstanties aan om, in aanvulling op opgelegde maatregelen voor fysiek afstand houden, de bevolking te motiveren om zelf opgelegde maatregelen met bewezen effectiviteit te adopteren, met het doel COVID-19 te beheersen.

In hoofdstuk 2 van dit proefschrift hebben wij gekeken hoe handhygiëne het beste ingezet kan worden om de transmissie van lagere luchtweginfecties te verminderen. Wij keken naar het effect van verschillende tijdstippen en frequenties van het wassen van handen, zowel als naar hoe lang een virus op de handen levensvatbaar blijft. Als die periode kort is, zoals dat voor griep is waargenomen, dan moeten de handen zeer vaak gewassen worden of zelfs onmiddellijk na het besmetten van de handen om de kans van infectie significant te verminderen. Wanneer het virus voor langere tijd op de handen werkzaam blijft, bijvoorbeeld in de aanwezigheid van snot of in het geval van stevigere virussen, dan is een beperkte hoeveelheid handen wassen voldoende om de transmissiekans van het virus aanmerkelijk te verminderen. Het onmiddellijk wassen van de handen na een besmetting is veel effectiever dan dit met regelmatige frequentie doen. Het aanbieden van handreinigingsmiddelen op plekken met een grote kans van transmissie door handen, zoals supermarkten, openbaar vervoer en werkplekken, kan het risico op infectie sterk verminderen, in het bijzonder voor virussen die voor lange tijd op de handen blijven.

Deze twee hoofdstukken illustreren dat de precieze impact van persoonlijke beschermende maatregelen afhangt van de details. Het wordt bepaald door de snelheid waarmee de maatregelen geadopteerd worden (Hoofdstuk 1) en, voor het wassen van handen, door de frequentie en het moment waarop dit wordt gedaan, alswel de duur dat het pathogeen op de handen levensvatbaar blijft (Hoofdstuk 2).

Wiskundige modellering en statistische bevindingen van transmissieroutes

In de hoofdstukken 3 en 4 hebben wij een transmissie/overdrachts model ontwikkeld voor de verspreiding van *Pseudomonas aeruginosa*, een bacterie die bekend is voor zijn inhe-

rente resistentie tegen antibiotica, zijn alomtegenwoordigheid in vochtige omgevingen, en vermeld staat in de hoogste categorie van WHO´s prioriteitenlijst met betrekking tot zeer antibiotica-resistente bacteriën. Verspreiding van *P. aeruginosa* is met name problematisch in ziekenhuis omgevingen waar, zonder enige controle maatregelen, de kwetsbare populatie een verhoogd risico loopt op besmetting. Het onderzoeken van de transmissie-routes en het kwantificeren van hun relatieve bijdrage aan het totaal aantal transmissies zal helpen bij het evalueren van interventiestrategieën en bij het beslissen welk beleid voor infectiebeheersing in ziekenhuizen prioriteit moet krijgen.

In Hoofdstuk 3 modeleren wij besmettingen vanuit de omgeving door expliciet te modeleren dat patiënten hun omgeving besmetten met pathogenen en dat die pathogenen achterblijven ook nadat de patiënt ontslagen is. Met behulp van een uitgebreide dataset van *P. aeruginosa* observaties die plaats hebben gevonden in twee IC's van een Frans universitair ziekenhuis in Besançon over een periode van 17 jaar, schatten wij in dat omgevingsbesmettingen na het ontslag van patiënten slechts een ondergeschikte rol speelde in vergelijking met achtergrond en kruistransmissie. Kruistransmissie droeg bij aan ongeveer de helft van de transmissies in deze Franse IC's (waarbij achtergrond transmissie de andere helft uitmaakt). Daarentegen was kruistransmissie slechts van ondergeschikt belang op de twee afdelingen van de IC voor volwassenen van het Erasmus Medisch Centrum in Rotterdam (zoals gepresenteerd in hoofdstuk 4). Door uitsluiting concludeerden wij dat aanhoudende besmetting vanuit de omgeving de belanrijkste transmissieroute was. In tegenstelling tot hoofdstuk 3 hebben wij omgevingsbesmettingen niet expliciet in het transmissiemodel opgenomen, maar de bijdrage ervan afgeleid door uitsluiting. Wij hebben dus twee verschillende benaderingen gedemonstreerd voor het modelleren van omgevingsbesmettingen als een route van (nosocomial) overdracht van *P. aeruginosa* op IC's en toonden aan dat de kwantitatieve resultaten voor het relatieve belang van trans-missieroutes sterk afhangen van de situatie. Er moet dus rekening gehouden worden met de lokale situatie wil men effectieve infectiepreventiemaatregelen nemen.

Wiskundige modellering voor de interpretatie van gegevens en interventiestrategieën

In de hoofdstukken 5 en 6 richten wij ons op modellerings-benaderingen die het probleem van in ziekenhuis opgelopen SARS-CoV-2-infecties aanpakken met behulp van ziekenhuisgegevens in het Verenigde Koninkrijk en Nederland.

In hoofdstuk 5 geven wij de eerste schattingen van de totale last van symptomatische ziekenhuisinfecties en hun bijdrage aan de algehele COVID-19 epidemie op nationaal niveau tijdens de eerste golf in Engeland. In het voorjaar van 2020 werd in Engelse ziekenhuizen een COVID-19-infectie gedefinieerd als zijnde verkregen in het ziekenhuis als de symptomen minstens een aantal dagen na opname optraden ("cut-off"). Symptomatische ziekenhuisinfecties kunnen met deze methode ongeïdentificeerd blijven door een kort verblijf van de patiënt (in vergelijking met de incubatieperiode) of als de eerste symptomen optraden vóór de cut-off. Wij toonden aan dat als men aanneemt dat een patiënt die voor de achte opnamedag symptomen van COVID-19 ontwikkelt, reeds besmet was bij opname, dat men dan waarschijnlijk slechts 30% van de symptomatische ziekenhuisinfecties zou identificeren. Deze bevinding suggereert dat onder gebruikelijke definities, zoals verstrekt door de UK Health Security Agency [9] en het European Centre for Disease Prevention and Control (ECDC) [10], het aantal ziekenhuisinfecties aanzienlijk wordt onderschat. Wij ontdekten echter dat de bijdrage van symptomatische ziekenhuisinfecties aan het totale aantal SARS-CoV-2 infecties in Engeland waarschijnlijk klein was (minder dan 1%) maar, in verloop van tijd, sterk varieerde met een toenemend belang in tijdsperioden met een lage prevalentie in de samenleving. Onze resultaten impliceren dat het screenen van patiënten bij ontslag uit het ziekenhuis, of het in quarantaine plaatsen van ontslagen ziekenhuispatiënten, effectief kan zijn in het voorkomen van verdere besmettingen binnen de gemeenschap. Wij schatten zelfs dat ongeveer 40% van die symptomatische infecties, die onopgemerkt kunnen blijven, door deze maatregelen opgespoord zouden kunnen worden. Dit is met name relevant voor instellingen die weinig middelen tot hun beschikking hebben en afhankelijk zijn van symptoomscreening waarbij ziekenhuisopnames relatief kort zijn. Onze resultaten werden op 22 oktober 2020 [11] aan SAGE gepresenteerd om het infectiebeheersing- beleid in ziekenhuizen in het Verenigd Koninkrijk te informeren voor toekomstige epidemische golven.

Wij hebben de effectiviteit beoordeeld van verschillende interventies in ziekenhuizen bij het beheersen van nosocomiale overdracht van SARS-CoV-2 in hoofdstuk 6. Door gebruik te maken van de gegevens over de ziekenhuisstructuur van het UMC Utrecht en het aantal bezette bedden van COVID-19-patiënten tijdens de eerste golf in Nederland waren wij in staat om het effect op de nosocomiale transmissie van SARS-CoV-2 te vergelijken van verschillende interventies gericht op HCW's. Onze resultaten toonden aan dat het gebruik

van persoonlijke beschermingsmiddelen op alle ziekenhuisafdelingen de meest effectieve interventie was vanwege het grote aandeel van asymptomatische en presymptomatische overdracht. Wij benadrukten tevens het belang van bron- en contactonderzoek buiten de algehele samenleving, vooral omdat het hogere percentages van test-positiviteit (aandeel geïnfecteerden onder geteste personen) kan bereiken dan reguliere screening interventies. Onze conclusies richten zich op de relatieve effectiviteit van interventiestrategieën in ziekenhuizen en zijn informatief voor toekomstige golven van de COVID-19-pandemie en uitbraken van nieuw opkomende varianten van het SARS-CoV-2 virus.

Conclusies van dit proefschrift

Over het algemeen presenteert dit proefschrift wiskundige modellen die verschillende onderzoeksvragen op het gebied van infectieziekten aanpakken, variërend van het evalueren van interventies en preventieve maatregelen voor de nieuw opgekomen SARS-CoV-2 tot het begrijpen van de transmissie dynamiek van veel voorkomende pathogenen in ziekenhuisomgevingen, zoals *P. aeruginosa*. Wij toonden aan hoe de mate van detail en modelkeuze sterk afhankelijk is van de beoogde resultaten, maar tevens ook voor de beschikbaarheid van informatie en gegevens. Wanneer problemen moeten worden behandeld op een populatieniveau kunnen eenvoudigere modellen met een geaggregeerde structuur (zoals compartimenten-modellen) voldoende zijn om de relevante transmissie-dynamiek vast te leggen. Eenvoudige statistische of simulatiemodellen gekoppeld aan geschikte gegevens kunnen het eerste kwantitatieve bewijs leveren zonder dat een volledig mechanisch transmissiemodel nodig is. Om rekening te houden met stochastische effecten in kleinere populaties zoals ziekenhuizen, kunnen complexere modellen zoals agent-gebaseerde modellen geschikter zijn, maar deze zullen moeten worden gevoed met gedetailleerdere gegevens. Wij identificeerden hiaten in de gegevensverzameling, met name wanneer het gaat over het gedrag en de gegevens van ziekenhuispatiënten. Dit zal moeten worden aangepakt om de modelbouw voor infectieziekten te verbeteren en de afhankelijkheid van slecht-onderbouwde aannames te verminderen. Model-geïnformeerde dataverzameling en het verenigen van verschillende informatiebronnen zullen een grote invloed hebben op hoe efficiënt er op een epidemie gereageerd kan worden, maar hiervoor is het noodzakelijk/essentieel dat verschillende onderzoeksgroepen, disciplines en landen hun informatie met elkaar delen over een langere periode [12]. Infectieziektes houden tenslotte geen

rekening met grenzen. Daarom is het noodzakelijk dat wij evenzo disciplinaire grenzen overschrijden om ons beter voor te kunnen bereiden en adequaat te reageren op aankomende ziekte-epidemieën.

References

[1] World Health Organization. *Global health estimates: Leading causes of death.* 2020. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death.

[2] World Health Organization. Antimicrobial resistance: global report on surveillance. In: *World Health Organization* 61.3 (2014), pp. 383–94. http://www.ncbi.nlm.nih.gov/pubmed/22247201.

[3] Woolhouse, M and Farrar, J. Policy: An intergovernmental panel on antimicrobial resistance. In: *Nature* 509.7502 (May 2014), pp. 555–557. doi: 10.1038/509555a.

[4] Laxminarayan, R, Matsoso, P, Pant, S, et al. *Access to effective antimicrobials: A worldwide challenge.* Jan. 2016. doi: 10.1016/S0140-6736(15)00474-2.

[5] Tacconelli, E, Buhl, M, Humphreys, H, et al. Analysis of the challenges in implementing guidelines to prevent the spread of multidrug-resistant gram-negatives in Europe. In: *BMJ Open* 9.5 (May 2019), e027683. doi: 10.1136/bmjopen-2018-027683.

[6] Fitzpatrick, MC, Bauch, CT, Townsend, JP, et al. Modelling microbial infection to address global health challenges. In: *Nature Microbiology 2019 4:10* 4.10 (Sept. 2019), pp. 1612–1619. doi: 10.1038/s41564-019-0565-8.

[7] Bruch, E and Atwell, J. Agent-Based Models in Empirical Social Research. In: *Sociological Methods and Research* 44.2 (Oct. 2015), pp. 186–221. doi: 10.1177/0049124113506405.

[8] Elizabeth Hunter, BMN, Kelleher, JD, Namee, BM, et al. A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology. In: *2016:84:3* 20.3 (June 2017). doi: 10.18564/JASSS.3414.

[9] Public Health England. *COVID-19: epidemiological definitions of outbreaks and clusters in particular settings.* 2020. https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters.

[10] European Centre for Disease Prevention and Control. *Surveillance definitions for COVID-19.* 2021. https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions.

[11] SAGE. *Sixty-third SAGE meeting on Covid-19, 22nd October 2020.* 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/935103/sage-63-meeting-covid-19-s0842.pdf.

[12] Kucharski, AJ, Hodcroft, EB, and Kraemer, MUG. Sharing, synthesis and sustainability of data analysis for epidemic preparedness in Europe. In: *The Lancet Regional Health - Europe* 9 (Oct. 2021), p. 100215. doi: 10.1016/j.lanepe.2021.100215.

Contributing authors

Alex Bhattacharya

Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, UK

Andrea C. Büchler

Department of Medical Microbiology and Infectious Diseases, Erasmus MC University Medical Center Rotterdam, The Netherlands

Marc Bonten

1) Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands. 2) Department of Medical Microbiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Martin C.J. Bootsma

1) Department of Mathematics, Faculty of Science, Utrecht University, Utrecht, The Netherlands. 2) Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, The Netherlands

Colin S. Brown

Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, UK

Ben S. Cooper

Nuffield Department of Medicine, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK

Pauline Ellerbroek

Department of Internal Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Stephanie Evans

Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, UK

Sebastian Funk

Centre for mathematical modelling of infectious diseases, IDE, EPH, London School of Hygiene Tropical Medicine, London, UK

Noortje G. Godijk

Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, The Netherlands

Diederik Gommers

Department of Adult Intensive Care, Erasmus MC University Medical Center, Rotterdam, The Netherlands

Russell Hope

Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, UK

Yalda Jafari

Centre for mathematical modelling of infectious diseases, IDE, EPH, London School of Hygiene Tropical Medicine, London, UK

Gwenan M. Knight

Centre for mathematical modelling of infectious diseases, IDE, EPH, London School of Hygiene Tropical Medicine, London, UK

Mirjam E. Kretzschmar

1) Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands. 2) Center for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, Utrecht, The Netherlands

Yin Mo

1) Oxford Centre for Global Health Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK.
2) Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. 3) Division of Infectious Disease, Department of Medicine, National University Hospital, Singapore, Department of Medicine, National University of Singapore, Singapore

Diane Pople

Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, UK

Jonathan M. Read

Lancaster Medical School, Lancaster University, Lancaster, UK

Julie V. Robotham

1) Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, United Kingom. 2) NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with Public Health England, Oxford, UK

Ganna Rozhnova

1) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. 2) BioISI—Biosystems & Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

Malcolm G. Semple

1) NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK. 2) Respiratory Medicine, Alder Hey Children's NHS Foundation Trust, Liverpool, UK

Juliëtte A. Severin

Department of Medical Microbiology and Infectious Diseases, Erasmus MC University Medical Center Rotterdam, The Netherlands

James Stinson

1) Healthcare Associated Infections and Antimicrobial Resistance Division, National Infection Service, PHE, Colindale, London, UK

Hannan Tahir

Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands

Alexandra Teslya

Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Bastiaan Van der Roest

Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Janneke H. H. M. van de Wijgert

1) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. 2) Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Liverpool, UK

Anne F. Voor in 't holt

Department of Medical Microbiology and Infectious Diseases, Erasmus MC University Medical Center Rotterdam, The Netherlands

Margreet C. Vos

Department of Medical Microbiology and Infectious Diseases, Erasmus MC University Medical Center Rotterdam, The Netherlands

ISARIC4C Investigators listed at

https://isaric4c.net/about/authors/

Centre for Mathematical Modelling of Infectious Disease COVID-19 Working Group

Sam Abbott, Amy Gimma, Hamish P Gibbs, Kaja Abbas, Rosanna C Barnard, Frank G Sandmann, Nikos I Bosse, Paul Mee, Ciara V McCarthy, Matthew Quaife, Adam J Kucharski, Christopher I Jarvis, Joel Hellewell, Emilie Finch, Alicia Rosello, Mark Jit, Rachael Pung, Rosalind M Eggo, Akira Endo, Graham Medley, Damien C Tully, Kerry LM Wong, Yang Liu, Katharine Sherratt, James D Munday, Lloyd A C Chapman, Stéphane Hué, Kathleen O'Reilly, Nicholas G. Davies, Sophie R Meakin, Fiona Yueqian Sun, Oliver Brady, C Julian Villabona-Arenas, Katherine E. Atkins, Kiesha Prem, David Hodgson, Mihaly Koltai, Carl A B Pearson, William Waites, Simon R Procter, Rachel Lowe.

List of publications and preprints

Publications included in this thesis

Pham TM, Kretzschmar ME, Bertrand X, Bootsma MCJ, on behalf of COMBACTE-MAGNET Consortium (2019). Tracking Pseudomonas aeruginosa transmissions due to environmental contamination after discharge in ICUs using mathematical models. PLoS Comput Biol 15(8): e1006697.

Teslya A, **Pham TM**, Godijk NG, Kretzschmar ME, Bootsma MCJ, Rozhnova G (2020). Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study. PLoS Med 17(7): e1003166.

Pham TM, Tahir H, van de Wijgert JHHM, Van der Roest B, Ellerbroek P, Bonten MJM, Bootsma MCJ, Kretzschmar ME (2021) Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study. BMC Med 19, 211.

Preprints included in this thesis

Knight GM, **Pham, TM**, Stimson J, Funk S, Jafari Y, Pople D, Evans S, Mo Y, Brown CS, Bhattacharya A, Hope R, Semple MG, ISARIC4C Investigators, CMMID COVID-19 working group, Read JM, Cooper BS, Robotham JV (2020). The contribution of hospital-acquired infections to the COVID-19 epidemic in England in the first half of 2020. medRxiv 2021.09.02.21262480; doi: 10.1101/2021.09.02.21262480.

Pham TM, Mo Y, Cooper BS (2020). The Potential Impact of Intensified Community Hand Hygiene Interventions on Respiratory tract Infections: A Modelling Study. medRxiv 2020.05.26.20113464; doi: 10.1101/2020.05.26.20113464

Publications not included in this thesis

Jafari Y, Mo Yin, Lim C, Pople D, Evans S, Stimson J, **Pham TM**, Read JM, Robotham JV, Cooper BS, Knight GM, Effectiveness of infection prevention and control interventions, excluding personal protective equipment, to prevent nosocomial transmission of SARS-CoV-2: a systematic review and call for action, Infection Prevention in Practice, Volume 4, Issue 1, 2022, 100192, ISSN 2590-0889.

Preprints not included in this thesis

Cooper BS, Evans S, Jafari Y, **Pham TM** Yin M, Lim C, Pritchard MG, Pople D, Hall V, Stimson J, Eyre DW, Read JM, Donnelly CA, Horby P, Watson C, Funk S, Robotham JV, Knight GM (2021). The burden and dynamics of hospital-acquired SARS-CoV-2 in England. PREPRINT (Version 1) available at Research Square. doi: https://doi.org/10.21203/rs.3.rs-1098214/v110.21203/`rs.3.rs-1098214/v1`

Teslya A, Rozhnova G, **Pham TM**, van Wees, DA, Nunner H, Godijk NG, Bootsma MCJ, Kretzschmar ME (2020). The importance of sustained compliance with physical distancing during COVID-19 vaccination rollout. medRxiv 2021.09.22.21263944; doi: `10.1101/2021.09.22.21263944`

Acknowledgments

This thesis would not have been possible without the help, guidance, and support by many colleagues, friends, and family. During these past five years I have learned that science is far from being an individual sport but highly benefits from collaboration and team effort. I would like to thank everyone who has accompanied me on this journey and helped me to grow as a scientist as well as a person.

Dear **Mirjam**, thank you for giving me the opportunity to embark on this PhD journey and accepting me to your research group. During these five years, you have impressed me with your consistent ability to maintain the bird's-eye view, your capacity to combine biological and mathematical knowledge in a wide range of topics, and your ability to stay focused and share your optimism in challenging times. You made sure that our work was continuously improving through constructive feedback but never forgot to recognize the effort and progress already made. Thank you also for every recommendation letter you wrote for me, for your encouragement and support throughout the PhD. You always gave me the feeling that you have genuine interest in my scientific success and in training me to become an independent scientist.

Dear **Martin**, thank you for accompanying me for the last five years as my daily supervisor. From day one, I was impressed by your sharp eye for details. You never hesitated to spare your precious time for some in-depth mathematical or statistical discussions with me. In particular, I really enjoyed our (casual) discussions in your office in the Maths building - very traditional on a black board where you inspired me with your clear explanations. You would always check in whether I could follow your line of thoughts, encouraged me to ask questions but also to question ideas, even if they were your own. You always showed good research practice and emphasized an open, transparent and honest way of working. You helped when I needed help but gave me enough room and freedom to make up my own scientific mind.

Dear **Ganna**, it is not very common to have a colleague, a co-supervisor, a climbing partner, and most importantly, a friend combined in one person. Everything you do, you do with a lot of motivation, thoroughness, endurance and dedication. You always

encouraged me to push my limits and strive for the best (be it with respect to work or climbing), and supported me by selflessly sharing your past experience. You also guided me when I was overwhelmed with too many tasks and got me back on the right track. I am more than grateful to have worked with you, and I had a lot of fun during our dinners, movie nights, and climbing sessions together with **Victor**. Although I will not be your direct colleague for the next two years, I will still be your friend and I sincerely hope that we will continue to have opportunities to work, climb, and spend time together.

Dear **Assessment Committee**, **prof. dr. ir. Hans Heesterbeek**, **prof. dr. Niel Hens**, **prof. dr. Jan Kluijtmans**, **prof. dr. Rafael Mikolajczyk** and **prof. dr. Rob Willems**, thank you very much for taking the time to read and assess my thesis.

Dear **Ben**, when I met you at the SPHINx conference in Paris in 2019, little I would know that I would end up working so closely with you. My research visit in Oxford turned out completely different due to the COVID-19 pandemic. This did not stop you from teaching me how to think about a problem from a statistical point of view, how to make sure that our work is relevant for (clinical) practice and that its conclusion is communicated in a transparent and clear way. You impressed me with your in-depth statistical and technical coding skills as well as your extensive overview of the literature in the field of infectious diseases. It always seems that you know every paper in this field! I hope that we can tick off some items of the long list of potential projects in our future collaborations. I also really appreciated how you integrated me in your research group: I never felt like I was *just* a visitor but that you took me just as serious as your other PhD students, and just as serious as your other senior colleagues. Our occasional cycling trips around Oxford will always stay in my mind as fond memories. Lastly, I am very grateful for your constant endorsement despite the short time period we got to work together, and all the doors that you opened for me.

Dear **Mo Yin**, I got to know you as a colleague in Oxford as part of Ben's research group. But in the end, you were not only a colleague but also my friend and my companion. Our excellent restaurant-style dinners, long walks along the River Thames, bouldering and climbing sessions at Brookes, and hilarious game nights made my stay in Oxford a wonderful experience. You impressed me with your eagerness to learn, ability to

juggle multiple projects simultaneously, and your wide interest outside academia. Good luck and all the best wishes for your return to Singapore! I really hope to visit and climb with you there!

Dear **Sasha**, thank you so much for being not only such a competent colleague, but also a kind and open-minded friend. Having transitioned from a theoretical Maths degree to Applied Epidemiology, I was pleasantly reminded by you of the beauty of the mathematical theory of epidemics can offer. I especially enjoyed our collaboration at the beginning of the COVID-19 pandemic. It was an intense time and your constant motivation, your stamina, and team spirit hugely impressed me. I wish you all the best for your family and hope to get to know little Veronica in not too far future.

Dear **Infectious Disease Modelling group**, **Bastiaan, Daphne, Emil, Franco, Hannan, Hendrik, Ilse, Kim, Maartje, Michiel, Noor, Sascha, Maartje, Martin, Mirjam**. Thank you for all the interesting presentations during our meetings on Wednesday afternoon. You gave me insights to your work and broadened my scientific horizon. In turn, your critical questions and constructive criticism gave me the possibility to take a step back and improve my work. I also really enjoyed our social gatherings at Mirjam's house! **Hannan**, thank you for the great collaboration in Chapter 6. Despite having kids at home during the lockdown, you always tried to be available for our meetings, often late during the night. Good luck with your new job in Wageningen and all the best wishes to your family! **Noor**, thank you for being a fellow PhD student for the past 3 years. I particularly enjoyed the conferences and the summer school that we attended together!

Dear **SPI-M subgroup on nosocomial COVID**, **Ben, Gwen, James, Jon, Julie, Stephanie, Yalda, Diane and Yin**, thank you for giving me the opportunity to contribute to the work on nosocomial SARS-CoV-2 transmission for informing SPI-M in the UK. Despite being an external researcher I always had the feeling that you valued my opinion just as any other. **Gwen**, thank you for being such a great collaborator during the COVID-19 pandemic. You always spared time to discuss the nitty-gritty details of our work in Chapter 5 but also to have a chat about work-unrelated topics. Although I was not your personal PhD student, you gave me tips and provided me guidance in the final stages of my thesis. Crazy to believe that we have only met on Zoom. When

conferences resume, I would love to meet up in person!

Members of **WMM**, **Claudia, Darren, Denise, Emma, Fien, Fleur, Henri, Janneke Verberk, Janneke van de Wijgert, Lufang, Marc, Nienke, Patricia, Sharon, Tess, Tessa, Thijs, Tim en Valentijn**. Thank you all for the fruitful discussions and input during our Wednesday Morning Meetings. When I joined in 2017, I had very little background in the clinical and biological aspects that are necessary to built useful mathematical models. Time has past and I realized that there is even more that I don't know but also that there is plenty that I have learned. Getting a different perspective and your honest feedback on other and my own work was a big and important part of my PhD journey. **Marc**, you impressed me with your extensive clinical and microbiological knowledge, your witty questions even in a field that is not your expertise, your clear and interesting explanations, your integrity, and your ability to add the necessary icing on the cake. **Patricia**, thank you for the great collaboration on the COVID-19 school project. I was impressed how well you could grasp and communicate our findings, and how dedicated and motivated you stayed throughout our work! **Henri**, your genuine and keen interest in understanding also the mathematical and statistical gist of research questions in infectious diseases have made a big impression on me. Thank you for reviewing my first article of my PhD and for your thoughtful questions and feedback during the WMM meetings.

Dear all members of the **COMBACTE** project. Thank you for giving me the opportunity to joining the big COMBACTE team. Thank you all members of the **COMBACTE-MAGNET WP3B package**, **Jan Beyersmann, Jean-François Timsit, Leonard Held, Martin Wolkewitz, Maja van Cube, Marlieke de Kraker, Tobias Bluhmki and Stephan Harbarth**. I enjoyed our collaboration and in particular our fruitful discussions at conferences and general assemblies.

Dear **Xavier Bertrand**, thank you for the great collaboration in Chapter 3. Thank you, in particular, for providing the data for the underlying study and your clinical expertise.

Dear **Andrea, Anne, Diederik, Juliëtte and Margreet from Erasmus MC**, thank you for the great collaboration in Chapter 4. Disrupted by the COVID-19 pandemic, we still

managed to (almost) finish this work. Thank you for providing the data, for your clinical expertise, and the interesting discussions during our meetings.

Dear **GS-LS PhD council**, thank you for your important work, for representing each PhD of the Graduate School of Life Science, and for giving me the opportunity to contribute. I enjoyed co-organizing the PhD days but also our social gatherings.

Beste **Kamer 6.125**, **Anne-Floor, Carolien, Heleen, Michiel, Manon, Sabine en Yvonne**. Dank jullie wel voor alle hulp en ondersteuning vooral in het begin van mijn trajectie. Ik had altijd het gevoel dat we alles (leuk of verdrietig) kunnen met elkaar delen. Het was ook heel tof met jullie soms avonds samen te eten of te kletsen. **Anne-Floor**, we waren maar een paar weken kamergenoten in de Julius center. Maar toch hebben we contact kunnen houden nadat ik uit Oxford weer terug in Utrecht was. Het was heel leuk met je samen te koken, te eten, te kletsen, te fietsen, en te wandelen. Veel success nog met je PhD! **Manon**, ik heb het heel genoten met je samen te kletsen, de Utrecht Science Park Campus run samen te hardlopen, en vond het erg lief dat ik in 2019 een paar weken in jouw appartement kon logeren en met jouw kat Joppie kon spelen. Veel success nog met je PhD! **Sabine**, dankjewel voor al de interessante gespreken en voor je steun in moeilijke tijden. Ik was er altijd onder de indruk van je creativiteit, nieuwsgierigheid, en slimheid, en ben er echt zeker van dat je een heel briljante arts zult zijn.

Dear **Thao**, I got to know you briefly as a fellow Lindy Hopper, former colleague at the Julius, and above all very kind and helpful person. Thank you so much for translating my short summary into Vietnamese. It's great to see that Vietnamese people help each other out, no matter where they are!

Beste **Remy**, je bent mijn klimpartner, Lindy Hop danspartner, pizza sous-chef, en mijn paranimf. Je hebt me begeleid tijdens bijna mijn hele doctoraat in Utrecht. Ik heb er genoten van onze klimavonturen, danssessies, spel- en kookavonde. Ondanks ik soms heel chaotisch was, was je altijd geduldig met me and heb je me altijd veel hulp aangeboden, vooral in moeilijke tijden. Dankjewel ook voor de vertaling van mijn nederlandse wetenschappelijke samenvatting!

Dear **Francisco**, thank you so much for founding the philosophy round table in Café Lijn 4. I was immediately enthusiastic about the idea to discuss any philosophical and political topic where opposing views would be actually desirable. I could not know back then that this was the beginning of a great friendship. Thank you for all the intriguing conversations, for our fun cooking and (board)game nights, and above all, thank you for being a good friend and standing besides me as my paranimf!

Dear **Paul**, thank you for all the fascinating discussions, the occasional walks and trips, the dinners, and movie nights in the past few months. All these little things were healthy distractions that kept me sane during the final stages of my PhD. I wish you good luck and a lot of fun on your PhD journey!

Beste **Filar**, onze vriendschap begon met een internationale dinner bij de Voorkamer. Sinds dan hadden vele kookavonde, wandeltjes, en dinnerdates. Ik voelde me gelijk heel welkom in Utrecht en vind je inzet om nieuwe mensen en vluchtelingen in Utrecht te verwelkomen hel tof en lief. Hoewel ik binnenkort ga verhuizen hoop ik dat we nog in contact kunnen blijven en als ik weer terug ben in Europa weer van onze vega-kookavonde kunnen genieten, maar dan met jouw kleine kroost!

Beste **Marieke**, ook jou heb ik bij een internationale dinner bij de Voorkamer ontmoeten. Ik vond het altijd leuk met je samen te eten, te wandelen en te kletsen. Ik was er onder de indruk hoe warm je nieuwe mensen hebt verwelkomt. Veel success met jouw nieuwe kleine familie!

Dear **Alina**, it seems like yesterday that we met at the philosophy round table in Café Lijn 4. It's been such a pleasure to have spent time with you while you were living in Utrecht. I still have vivid and fond memories of our dinners, philosophical discussions, bouldering and tea sessions. It is nice to see that we do not need to live in one city to stay in contact and so, I am confident that our friendship will continue even with an ocean between us. Congratulations to your Marie-Curie fellowship which you have earned without any doubt. Good luck with your new job and new start in Dublin! I will be glad to pay you a visit.

Beste **Calvin**, oorspronkleijk hebben we elkaar tijdens het boulderen ontmoeten. Maar sinds dan ben je niet alleen een boulderbuddy maar ook een goede vriend. Ik herinner me graag aan je snurken tijdens onze vakantie in Portugal, jij onbevredigde verlangen naar Affogato in Dresden en je onbaatzuchtige hulp met mijn nieuwe kamer en mijn Murphy bed in Utrecht. Dankjewel voor de mooie tijd en je bent altijd welkom me te bezoeken!

Beste **bouldervrienden**, **Aaron, Cabinh, Chen, Cabinh, Dani, Dani Bodor, Pier, Sioulan, Terry**. Dank jullie wel voor de tof bouldersessies die ik soms echt nodig had om mijn hoofd vrij te kregen van all die werk.

A big thank to my friends and second family **DmCMS** (Anh **D**ung, **C**hristoph, **M**aria and **S**ebastian). It's hard to put in words what you mean to me. You are the evidence that the strong bonds of friendship can transcend any borders and that it can be as important as family. You have followed all my steps on my PhD journey, from accepting the position until handing in my thesis. You have been my pillar of support for the last 13 years and I am glad that we kept the tradition (with one year of exception) of meeting up during the Christmas holidays and New Year's Eve.

Dear **Samuel**, thank you for being such a great climbing buddy and friend. It's truly great to have a friend that enjoys multiday hikes, indoor and outdoor climbing, vegetarian cooking as well as discussing human rights topics. Our PhD retreat in the Swiss mountains was a lot of fun. Thank you for never complaining about my crazy working hours and for supporting me as best as you could! Good luck with finishing up your PhD and with your next adventure. Hopefully, we can still occasionally meet up for some climbing trips.

Dear **Linh Chi**, we have been friends now for about 12 years. We have grown so close to each other, despite being physically far apart. I always appreciated that although we do not meet very often, when we do, it's like we have never left. I gladly look back at our hiking and climbing trips that gave me the possibility to take a break from my research. I am very grateful to have you as a friend and happy to be the godmother of little Tim, and am looking forward to continue our outdoor holidays

with him. I am sure he will be able to teach me to ski in a few years! Dear **Dominik**, thanks for the fun and educating holidays to the Balkan region, and to Switzerland. I always appreciated your positive mind set and will always remember your words: Tip Top!

Beste **Huize Royaards**, **Anja, Rene, Ruurd, Eva** ik ben zo blij jullie als huisgenoten te hebben en ga jullie missen! Jullie hebben me altijd het gevoel gegeven dat we niet alleen huisgenoten maar ook een mucki-club een en kleine familie zijn! We konden altijd over alles praten en hadden altijd veel plezier met films, games en etentjes. Dankjewel **Anja**, dat ik samen met jouw familie kerst kon vieren in 2020. Je hebt me vanaf de eerste dag heel hartelijk in het huis ontvangen. **René**, dankjewel voor de vertaling van mijn Nederlandse samenvatting, voor de leuke boulder- en workout sessies en dat je altijd om mijn welzijn geeft. **Ruurd**, dankjewel dat je deur altijd open staat en dat we altijd leuk kunnen praten. Dankjewel ook vor het leuke samen boulderen. **Eva**, dankjewel voor het lekkere samen eten en leuke pratjes. **Desirée**, ondanks we niet meer samen wonen vond ik de tijd met jou in de huis heel leuk. Ik was er heel onder de indruk van je gaming skills!

My former flatmates in Oxford, **Alex, Charlotte, Chris, Elsa and Elénore**, thank you so much for letting me join your house on Reliance Way in Oxford. I was impressed how open-minded you were, accepting me as an international as your housemates although we have met only online. You quickly became more than just housemates, you became my friends, and even my little British family. I always have to smile when I remember our dinners, vivid discussions, (board) game nights, cycling trips, and dance sessions!

Dear **Peiqi**, thank you for being a great fellow Lindy Hopper and great friend. I really enjoyed all of our walks, dinners and game nights. Please come and visit me in Boston!

Dear **Swing in Utrecht**, thank you for organising these wonderful Lindy Hop lessons, the Wednesday evening social Hopspots and monthly band nights!

Dear **Jacob Seifert**, thank you for accompanying me during my first PhD years in Utrecht. I, particularly, enjoyed our deep conversations, our outdoor holidays, and (board) game nights. Good luck on your remaining PhD journey and all your future endeavours!

Mẹ yêu mến, cảm ơn mẹ đã luôn tin tưởng ở con, đã cho con mọi cơ hội tốt nhất, bất chấp mọi khó khăn mẹ gặp phải trong cuộc sống. Cảm ơn mẹ đã luôn ủng hộ và yêu thương con. Con yêu mẹ hơn bất cứ thứ gì trên trái đất này.

Bố yêu mến, cảm ơn bố đã tin tưởng và ủng hộ con. Con rất vui khi có thêm hai em Hải An, Hải Hiếu và gì Toàn trong đại gia đình của chúng mình. Con yêu bố rất nhiều.

Dear **Mai**, thank you for always being such a wonderful sister. Although we are physically separated, I always feel that our family bond connects us. I am more than grateful that we can just share everything and that we manage to be there for each other by calling and texting. I am sure we can make this work even after I have moved to the US. You have always given me your full support, wherever I went, whatever my ambition or endeavors were. Dear **Dennis**, I am glad to have you as a brother-in-law. Thank you making my sister happy and for being a great addition to our family!

About the author

Thi Mui Pham was born on 12th July 1991 in Hanoi, Vietnam. Her family moved to Dresden in Germany when she was six months old. She studied Mathematics at RWTH Aachen University in Germany. During her Master's degree, she was an Erasmus student at Queen Mary University of London in the UK. In 2016, she obtained her Master's degree in Mathematics with a minor in Computer Science.
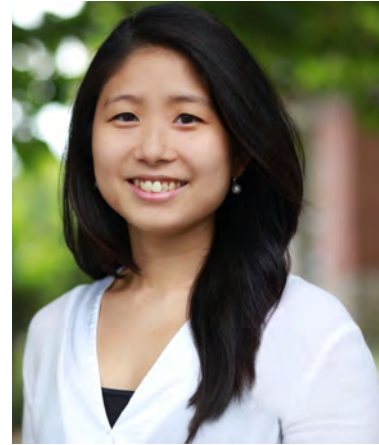
Mui started a PhD trajectory in 2017 at the Infectious Disease Modelling group of the Julius Center at the University Medical Center Utrecht led by Mirjam Kretzschmar. She was supervised by Mirjam Kretzschmar, Martin Bootsma and Ganna Rozhnova. In 2020, she was a visiting researcher at Ben Cooper's research group at the Nuffield Department of Medicine of the University of Oxford. Due to the COVID-19 pandemic, her work shifted from the mathematical modeling of transmission routes of antibiotic resistant bacteria in hospitals to the modelling of interventions to limit the transmission of SARS-CoV-2 during the second half of her PhD. From March till September 2021, she worked on modeling the transmission of SARS-CoV-2 in secondary schools on behalf of the Ministry of Education, Culture and Science of the Netherlands.

Concurrently to her PhD, Mui obtained a postgraduate master in Epidemiology, with a specialization in Epidemiology of Infectious diseases at Utrecht University. She was a member of the PhD Council of the Graduate School of Life Sciences where she was one of the representatives of the Epidemiology program.

Mui has accepted a postdoctoral research position with Marc Lipsitch and Yonatan Grad at Harvard T.H. Chan School of Public Health and will start her position in March 2022.