

RESEARCH ARTICLE

A mathematical model and inference method for bacterial colonization in hospital units applied to active surveillance data for carbapenem-resistant enterobacteriaceae

Karen M. Ong^{1,2*}, Michael S. Phillips¹, Charles S. Peskin²¹ New York University School of Medicine, New York, New York, United States of America, ² Courant Institute of Mathematical Sciences, New York, New York, United States of America* karen.ong@med.nyu.edu

Abstract

Widespread use of antibiotics has resulted in an increase in antimicrobial-resistant microorganisms. Although not all bacterial contact results in infection, patients can become asymptotically colonized, increasing the risk of infection and pathogen transmission. Consequently, many institutions have begun active surveillance, but in non-research settings, the resulting data are often incomplete and may include non-random testing, making conventional epidemiological analysis problematic. We describe a mathematical model and inference method for in-hospital bacterial colonization and transmission of carbapenem-resistant Enterobacteriaceae that is tailored for analysis of active surveillance data with incomplete observations. The model and inference method make use of the full detailed state of the hospital unit, which takes into account the colonization status of each individual in the unit and not only the number of colonized patients at any given time. The inference method computes the *exact* likelihood of all possible histories consistent with partial observations (despite the exponential increase in possible states that can make likelihood calculation intractable for large hospital units), includes techniques to improve computational efficiency, is tested by computer simulation, and is applied to active surveillance data from a 13-bed rehabilitation unit in New York City. The inference method for exact likelihood calculation is applicable to other Markov models incorporating incomplete observations. The parameters that we identify are the patient–patient transmission rate, pre-existing colonization probability, and prior-to-new-patient transmission probability. Besides identifying the parameters, we predict the effects on the total prevalence (0.07 of the total colonized patient-days) of changing the parameters and estimate the increase in total prevalence attributable to patient–patient transmission (0.02) above the baseline pre-existing colonization (0.05). Simulations with a colonized versus uncolonized long-stay patient had 44% higher total prevalence, suggesting that the long-stay patient may have been a reservoir of transmission. High-priority interventions may include isolation of incoming colonized patients and repeated screening of long-stay patients.

OPEN ACCESS

Citation: Ong KM, Phillips MS, Peskin CS (2020) A mathematical model and inference method for bacterial colonization in hospital units applied to active surveillance data for carbapenem-resistant enterobacteriaceae. PLoS ONE 15(11): e0231754. <https://doi.org/10.1371/journal.pone.0231754>

Editor: Yang Yang, University of Florida, UNITED STATES

Received: March 10, 2019

Accepted: March 31, 2020

Published: November 12, 2020

Copyright: © 2020 Ong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information files](#).

Funding: Support was provided by the New York University Medical Scientist Training Program (NIH training grant T32GM007308) and the Minkowitz Scholarship (a private scholarship). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Carbapenem-resistant Enterobacteriaceae (CRE) are a rising global health threat [1–7]. CRE are a family of antibiotic-resistant, gram-negative enteric bacteria [8] that harbor resistance to many “last line” treatment drugs [1] and include human pathogens such as *Klebsiella pneumoniae*, *Enterobacter cloacae*, and *Escherichia coli* [8]. The unusual combination of pathogenicity [9] and antimicrobial resistance [1] of CRE renders it a major cause of morbidity and mortality in hospitalized patients [9] and, increasingly, otherwise healthy hosts [10]. These pathogens are not only resistant to the carbapenems, but to many or most other classes of antibiotics that are effective and safe [11], such as broad-spectrum cephalosporins. Treatment of CRE with polymyxins and aminoglycosides is complicated by efficacy, pharmacokinetics, and toxicity [12]. Recently introduced treatment options such as meropenem-vaborbactam [13] or ceftazidime-avibactam are expensive [14], and clinical experience with their use is limited [13].

Although not all bacterial contact results in infection, patients can become asymptotically colonized [15], leading to a higher probability of transmission to other patients [16], extended length of stay, increased carbapenem exposure, or infection, which increases mortality risk [17, 18] as high as 70% for patients in intensive care units [9]. As a result, the Centers for Disease Control has recommended active surveillance of patients at high risk [19]. Active surveillance, or screening without suspicion of infection, can identify asymptomatic carriers [20, 21] for subsequent intervention with infection control measures such as the use of contact precautions, decolonization, patient cohorting, or minimization of the use of invasive devices [15, 19], and may decrease rates of nosocomial transmission [22] and reduce morbidity, mortality, and hospitalization costs [23, 24].

Studies using active surveillance data typically use standard epidemiological methods such as case-control studies [25] or prospective observational studies [26] to determine a global statistic of infection or colonization and an assessment of risk factors and their importance. Because the infectious process is only partially observable and the data are highly dependent, mechanistic mathematical models that allow parameter estimation and hypothesis testing [27, 28] can be an important addition to the toolbox of epidemiologists, clinicians, and public health policy-makers. These methods can allow analysis of the dynamics of an outbreak and can use information about locations and times of colonization that are often obscured when data is aggregated for use with traditional methods. However, variants on the classic Kermack-McKendrick epidemic model [29]—which describes the spread of infectious disease in a population of uniformly-mixed, homogeneous individuals who may be disease susceptible, infective, or resistant (an SIR model)—are often poorly suited for describing pathogen spread within small populations with high patient turnover [30].

Stochastic models are naturally suited (and increasingly used [31]) to describe transmission and its variance amongst the small numbers of patients within hospital units. For example, several groups have used discrete-time [32] or continuous-time Markov chains to model and analyze hospital active surveillance data [32–40] for organisms such as methicillin-resistant *Staphylococcus aureus* (MRSA) [41, 42], vancomycin-resistant Enterococci (VRE) [38, 41, 43], carbapenem-resistant Enterobacteriaceae (CRE) [32], cephalosporin-resistant gram-negative rods (RGNR) [41], *Pneumococcus* [33], and/or *Pseudomonas* [38]. Earlier modeling efforts moved beyond simple Poisson or unstructured hidden Markov models, which assumed infections to be independent (appropriate for RGNR) [41], to incorporate two mechanisms of colonization: exogenous (patient cross-transmission) versus endogenous (arising from antibiotic pressure) for VRE [38, 41], *Pseudomonas* [38], MRSA [41], or CRE [32]. Later models incorporated other mechanisms such as healthcare workers as vectors [42, 44, 45], within-family versus community-acquired transmission [33], sporadic [37] or background colonization [35, 39, 40],

[46], pre-existing colonization [32, 34, 37, 41, 47, 48], transmission from other rooms [46], or bacterial acquisition between hospital stays [34]. Notably, López-García & Kypraios created a generalized model of nosocomial spread that included multiple sources of transmission including patients, staff, and environment, and allowed calculation of the basic reproduction number [36].

Many models used point prevalence [34, 37, 44] from regular surveillance testing [38], but Cooper and Lipsitch used sparse microbiological and clinical culture results, although fitting parameters required multiple years of data [41]. Less frequently, groups directly tracked individual colonization statuses in the state of the model [32, 33]. Numerous groups subsequently applied Markov Chain Monte-Carlo (MCMC) methods to allow incorporation of prior knowledge about parameters [41, 49], estimation of sensitivity [34, 35, 39, 40, 47], and/or augmentation of data with individual patients' possible but unobserved colonization times [34, 35, 39, 47]. Calculation of exact likelihood while incorporating partial observation data and unknown times of patient colonization was deemed an "intractable" problem by some groups [39, 47], although Bootsma et al. [32] estimates maximum likelihood in a discrete-time Markov model that uses test results to reduce the space of probable states and allows for varying occupancy of the unit.

In this work, we present a hybrid continuous-time/discrete-time Markov susceptible- infective-susceptible (SIS) model that allows direct calculation of likelihood while incorporating exact exit/entry times and incomplete observation data. This model tracks the discrete state of the unit, a binary string representing the colonization status of each patient within the unit. The state evolves continuously between events (tests or patient turnover) with possible patient transmission, but it can change instantaneously and discretely at the time of patient turnover with the entry of a new, potentially colonized patient and possible prior-to-new patient colonization event via a contaminated bed or surroundings. From a mechanistic standpoint, we describe a model of bacterial transmission and colonization tailored to analyze incomplete active surveillance data for carbapenem-resistant Enterobacteriaceae (CRE). Our mathematical model examines three mechanisms of bacterial colonization: **patient-patient transmission** (via healthcare providers or other vectors) [15], **prior-to-new patient transmission** (also referred to as patient-bed-patient transmission because it may occur via contaminated bed linens [26] or via the immediate surroundings [50]), and **pre-existing colonization** (in which patients have been colonized prior to hospital unit entry by exposure during previous hospitalizations or stays in long-term care facilities) [51, 52]. We will discuss 1) formulation of a hybrid discrete-time/continuous-time Markov model analogous to a SIS model (in which beds can be in either a colonized/uncolonized state), 2) event-driven simulation using data for patient exit and entry, 3) an inference method designed for incomplete active surveillance data that calculates parameter likelihood given all possible sequences of states consistent with (partial) observations, 4) methods to speed calculation of the likelihood function, 5) comparison with a reduced-state model similar to previously published SIS models [32, 34, 36], 6) results from maximum-likelihood parameter estimation, and 7) the contributions of this work in the context of past modeling and parameter inference efforts.

From an epidemiological standpoint, the model and methods presented in this paper can be used to improve estimation of patient-patient transmission from sparse data by using the individual time, location, and test result information for each patient, allowing parameter estimation from individual rooms and small hospital units. Furthermore, the incorporation of multiple routes of transmission allows breakdown of the overall colonization prevalence (total colonized patient-days) into the components attributable to mechanisms in excess of pre-existing colonization. From a technical standpoint, the methods for inference and parameter estimation allow exact determination of maximum likelihood parameters without requiring

likelihood function estimation or sampling the distribution of possible realizations using MCMC techniques, even when the model has a large number of possible states. This inference method is generally applicable to discrete-state, discrete- or continuous-time Markov models applied to partial observation data. Although calculation of the likelihood function for any given set of parameters is expensive because it incorporates a matrix exponential [53], use of the full detailed state tracking individual patient colonization statuses and the exact entry/exit times enabled by the continuous-time model allows better parameter estimation compared to use of a reduced state model that tracks the number of colonized patients. We outline a number of mathematical and computational methods to make this calculation both feasible and practical. (Details of many of these methods are available in the appendices following the main body of the paper).

Materials and methods

Ethics statement

The Institutional Review Board of the NYU School of Medicine approved this study (IRB #08 818) on the “Epidemiology of KPC producing Enterobacteriaceae in New York City” with a waiver of consent: “Active surveillance for multi-drug resistant bacteria is a routine, well-accepted method for enhancing infection control in areas of high risk in the hospital setting. Only routine clinical specimens obtained for the active surveillance of KPC-E are used in this study.” The data were analyzed anonymously.

Active surveillance data. We used de-identified hospital surveillance and census data (available in [S1 Data](#)) for CRE colonization within a rehabilitation unit in a New York City academic center, a subset of the data used for a CRE case-control study by Swaminathan et al. [25]. The surveillance data consisted of perirectal swabs taken shortly after patient entry into the hospital unit and approximately weekly thereafter until exit. All events were drawn from an electronic medical record which placed timestamps on tests and entry/exit events to at least the date, hour, and minute. The swabs were cultured for CRE (primarily *Klebsiella pneumoniae*). Also available were clinical microbiological test results from cultures performed upon suspicion of infection. These data were combined with census data consisting of patient bed locations and turnover times for the study duration of 417 days. (The average patient length of stay was 14.1 days). We limited our analysis to 13 of 25 beds in the rehabilitation unit that were consistently used, as the remainder were either temporary or rarely used beds that stayed empty for the majority of the study duration. Additional information is available in [S2 Appendix](#).

Models

Hybrid discrete-time/continuous-time Markov model

Consider a hospital unit of n beds filled with patients, all of whom are initially uncolonized with pathogen. We model the process of colonization and transmission as a hybrid discrete-time/continuous-time, discrete-state Markov chain [54, 55] in which a discrete-time Markov process is used to describe events occurring at patient turnover, but a continuous-time Markov chain is used to describe events occurring during time intervals in which no patients enter or leave.

The **detailed state** of the unit at time t is represented as a binary vector (or bit string) $\mathbf{b}(t)$. Each component $b_k \in \{0, 1\}$ corresponds to the status of the patient in bed k at time t , for all $k \in \{1, 2, \dots, n\}$. Patients are considered uncolonized ($b_k = 0$) if they do not harbor pathogen, but colonized ($b_k = 1$) if they do, even if asymptomatic. These states are equivalent to the

susceptible and infective states of an SIS model, which is a two-state variant of the classic three state susceptible-infective-resistant (SIR) model [29]. We assume beds are never empty, so there are only two possible states for each bed and only 2^n possible states of the unit. The **reduced state** of the unit is an integer representing the total number of colonized patients in the unit at a given time. In the detailed model, we choose to track the status of individual patients and not just the *number* of colonized patients, even though the detailed state determines the reduced state—but not conversely—simply by counting the number of colonized patients.

We assume three mechanisms in the model by which individual patients become colonized: **pre-existing colonization**, **prior-to-new patient colonization**, and **patient-patient transmission**. Here, pre-existing colonization is defined as carriage of pathogen upon entry into the unit. Prior-to-new patient colonization is transmission of pathogen from the prior bed occupant to the new patient immediately following, which can occur via contamination of the bed [26] or immediate surroundings [50]. Finally, patient-patient transmission can occur between any two patients who are simultaneously present in the unit via mechanisms such as the use of shared equipment or contact with health-care providers [15, 56]. Of the three processes, note that two occur (or can be thought of as occurring) instantaneously *during* patient turnover (pre-existing and prior-to-new patient colonization), whereas the third (patient-patient transmission) is considered to occur at any time *between* patient turnover events.

During exit/entry events, colonization can occur via prior-to-new patient colonization. Let S be a dichotomous random variable denoting the unknown colonization state of an entering patient, and let ϕ be the probability of **pre-existing colonization** (Fig 1b). $S = 0$ codes for uncolonized; $S = 1$ codes for colonized; $\Pr\{S = 1\} = \phi$; and $\Pr\{S = 0\} = 1 - \phi$. When the colonization state is known (not random), it is denoted by s , where (similarly) $s = 0$ for uncolonized (Fig 1a), and $s = 1$ for colonized (Fig 1c). If the new patient is uncolonized ($s = 0$ if status known, otherwise $S = 0$ with probability $1 - \phi$) and enters a bed in which the previous patient was colonized (Fig 1e), there is a probability ψ of **prior-to-new patient colonization**, which may occur via contamination of the bed or immediate surroundings. Alternately, the patient may remain uncolonized with probability $(1 - \psi)$. We assume that exit, entry, and possible prior-to-new patient colonization occurs instantaneously during an **exit/entry event**.

Between exit/entry events, **patient-patient transmission** can occur from colonized to uncolonized patients. For any given uncolonized patient, the fixed probability per unit time of

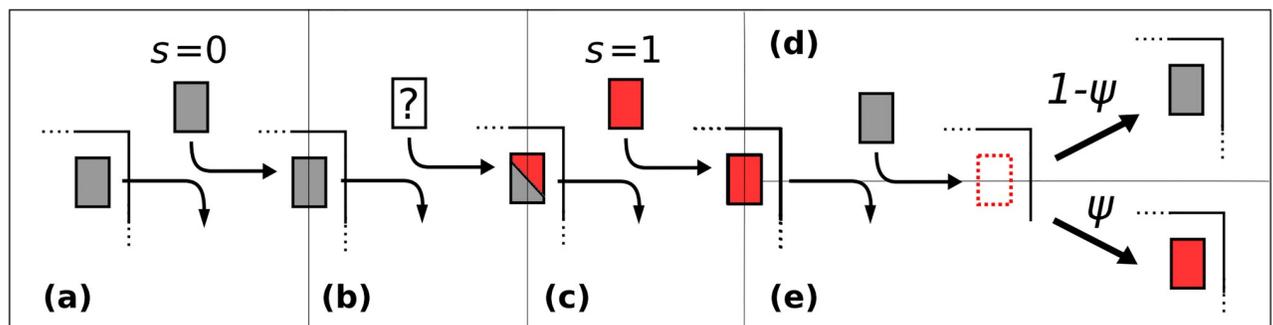


Fig 1. Possible exit/entry scenarios. Possible exit/entry scenarios, with s as an indicator variable for patient colonization (if known), ϕ as the probability of pre-existing colonization (if the new patient's status is unknown), and ψ as the probability of prior-to-new patient colonization given that the prior patient was colonized: (a) replacement of an uncolonized patient by an uncolonized patient ($s = 0$ if status known); (b) replacement of an uncolonized patient by a patient of unknown status (colonized with probability ϕ); (c) replacement of a patient of any status with a colonized patient ($s = 1$ if status known); and replacement of a colonized patient by an uncolonized patient ($s = 0$ if status known, otherwise uncolonized with probability $1 - \phi$), who (d) remains uncolonized with probability $1 - \psi$ or (e) becomes colonized via prior-to-new patient colonization (probability ψ).

<https://doi.org/10.1371/journal.pone.0231754.g001>

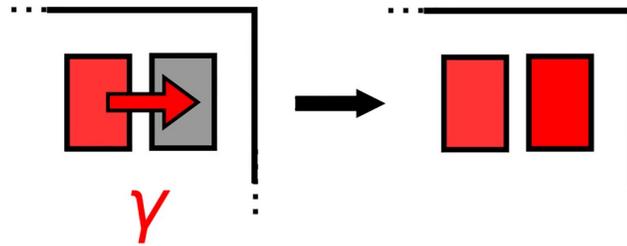


Fig 2. Transmission from colonized to uncolonized patient. Transmission from colonized to uncolonized patient at rate γ per day per colonized-uncolonized patient pair.

<https://doi.org/10.1371/journal.pone.0231754.g002>

a colonization event per colonized patient in the hospital unit is γ (Fig 2). Thus, we are assuming that the overall patient–patient transmission rate in the hospital unit is linearly proportional to the number of *pairs* of colonized and uncolonized patients present at a given time. Although continuous-time transmission is “restarted” at each exit/entry event, the model is a “memoryless” Markov model [55] that depends only on the present state to determine the next state without regard to history. At the discontinuity of an exit/entry event, the number of colonized patients may instantaneously increase or decrease by 1, which then would change the unit transmission rate ($i\gamma$, where i is the number of colonized patients present).

Of note, we use the term “rate” (the expected number of events per unit time) interchangeably with the term “probability per unit time,” but distinguish it from “rate constant,” which is a reaction rate coefficient in chemical kinetics that is *not* equivalent to the probability per unit time [57, 58]. We use “probability per unit time” as a shorthand for *instantaneous* probability per unit time, defined as the probability per unit time as the time interval decreases to an infinitesimal size (equivalent to a hazard function from survival analysis [59]):

$$\lim_{\Delta t \rightarrow 0} \frac{P(\text{event occurs in the time interval } (t, t + \Delta t))}{\Delta t}$$

We define γ as the instantaneous probability per unit time of a patient-patient transmission event occurring from the initial time at which patient-patient transmission can occur (here assumed to be $t_0 = 0$) to time t . The probability of no event occurring during the time interval $(0, t)$ is $e^{-\gamma t}$ (equivalent to a cumulative survival function [59]), so the probability of at least 1 event occurring during the time interval is $1 - e^{-\gamma t}$.

Changes in the state of hospital units. In the next section, we describe how probabilities evolve for the change in state of the entire hospital unit at and between events as a result of changes in the colonization status of individual patients. Let T be a vector of arbitrary length of the known, arbitrary times at which patient turnover occurs. We assume that patient replacement is instantaneous and the status of the z -th incoming patient may be known or unknown. A patient that leaves at time t_z^- is immediately replaced a new patient at time t_z^+ (Fig 3). Here, t_z^- is the limit approaching t_z from the left and t_z^+ is the limit approaching t_z from the right, given that time is pictured as flowing from left to right as in Fig 3.

Let $\mathbf{p}(t)$ be the probability distribution at time t of the different possible states of the unit. That is, $\mathbf{p}(t)$ is a vector with 2^n entries, each of which refers to a state of the n -bed unit as a whole and gives the probability that the hospital unit is in the corresponding state. Notice that in our model, pre-existing and prior-to-new patient colonization occur only during patient exit/entry events, but patient–patient transmission occurs solely *between* exit/entry events during **non-turnover intervals**. The probability distribution changes in a discrete manner at an exit/entry event, which begins at the time of the prior patient’s exit (t_z^-) and ends after the new

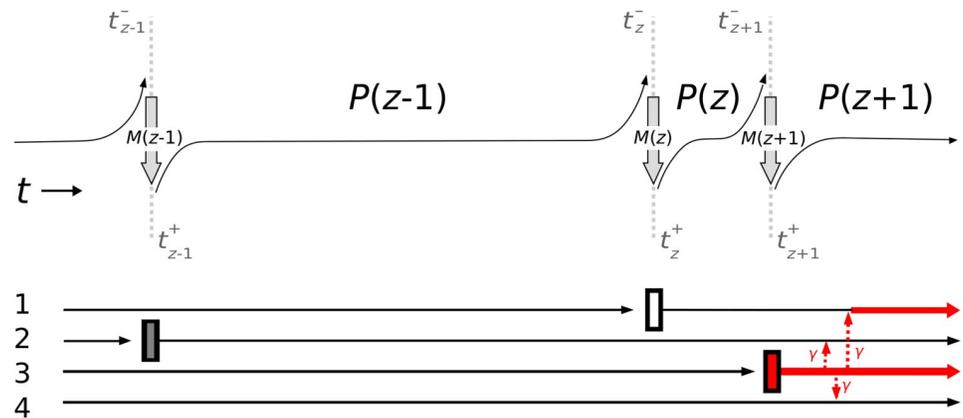


Fig 3. Overview of all processes of colonization and transmission. Overview of all processes of colonization and transmission. The top diagram shows the two types of transition matrices. $M(z)$ is the exit/entry probability matrix, which describes probabilities of various outcomes upon patient turnover and possible prior-to-new patient transmission at an exit/entry event. $P(z-1)$ is the non-turnover transition probability matrix, which is derived from the patient-patient transmission rate matrix and describes what happens from one turnover event to the next. The bottom diagram shows the statuses of individual beds within an example four-bed hospital unit. The vertical bars represent a patient turnover event. The vertical bar color represents the status of the incoming patient: white if uncolonized, red if colonized, and gray if unknown. The horizontal lines depict the colonization status of the patient in a particular bed: a thin black line means the patient is uncolonized, but the thick red line means the patient is colonized. The dashed red arrows represent the probability per unit time of patient-patient transmission between pairs of colonized-uncolonized patients.

<https://doi.org/10.1371/journal.pone.0231754.g003>

patient's entry and possible prior-to-new patient colonization (t_z^+). Between exit/entry events, no patient turnover occurs, and the probability distribution evolves continuously (Fig 3).

Transition probabilities for changes in the state of the hospital unit that occur *during* the z -th exit/entry event are represented as entries within a **discrete-time Markov transition probability matrix** $M(z)$ (described in S1F Appendix). In contrast, the transition rates for state changes that occur *between* patient exit/entry events are represented as entries within the **continuous-time transition rate matrix** R (described in S1G Appendix). The cumulative effect of these changes can be written [60, 61] as a **probability matrix** $P(z)$ that examines the change in the state of the hospital unit from the beginning to the end of a discrete time interval, as shown in S1H Appendix.

Simulation

Given input parameters γ , ϕ , and ψ ; a list of $Z-1$ exit/entry events at arbitrary times $T = \{t_1, \dots, t_{Z-1}\}$ that occur in beds $K = \{k_1, k_2, \dots, k_Z\}$; and, optionally, a list of arbitrary states $S(z)$ of the entering patients at time t_z , we can create a simulation of the hospital unit for the period t_0 to t_Z . (Here, we assume that at time t_0 , the hospital unit is completely uncolonized, and the end of the simulation period occurs after the time of the last exit/entry event at time t_{Z-1} . The simulation can also be performed with other initial states if desired).

Simulation of the hybrid continuous-time/discrete-time model requires three steps: 1) choosing a list of **arbitrary exit/entry times** from data or simulation; 2) **simulating the exit/entry events** of pre-existing colonization and possible prior-to-new patient colonization using the probabilities ϕ and ψ , and 3) using **event-driven simulation** (also known as the Gillespie Method [57, 62, 63]) and the parameter γ to simulate patient-patient transmission *between* exit/entry events.

First, we used the exact exit/entry times from the rehabilitation unit census data. However, it is also possible to create simulations using turnover times from other data sets or other methods, such as generating these times using the assumption that exit/entry is a Poisson process governed by the turnover time (the inverse of the mean length of stay [64]).

Second, we assign patients an initial colonization status upon entry into the unit and a final status after contact with the bed. If the patient's colonization status was previously known to be uncolonized or colonized, the initial colonization status was tentatively set to be 0 or 1, respectively. Incoming patients whose colonization status was unknown were given an initial status of 0 with probability $1 - \phi$ or status of 1 with probability ϕ . Then, uncolonized new patients (initial status 0) entering a bed previously occupied by a colonized patient were given a final colonization status of 0 or 1 with probability $1 - \psi$ or probability ψ , respectively. Patients who entered initially colonized remained colonized after contact with the bed, regardless of the previous patient's colonization status, and were given a final status of 1. Details about exit/entry event simulation are found in [S1B Appendix](#).

Last, we performed stochastic simulations of patient-patient transmission using **event-driven simulation** [57, 62, 63]. Recall that the probabilities per unit time of patient-patient transmission are represented as a continuous-time Markov chain for the interval between two exit/entry events at times t_{z-1} and t_z in which no patient turnover occurs. During this interval, the only possible event is patient-patient transmission, which occurs at rate γ . Event-driven simulation can be used to determine the time and location of patient-patient transmission events using a variant [63] on the Gillespie method [57, 62]. This method uses transition rates from a current state to multiple possible final states to create a realization of the final state of the system after transition. Because this simulation method uses random numbers as input, individual realizations will be different, but in aggregate, their statistics will approach that of the original transition rates. Details for simulating patient-patient transmission between exit/entry events are described in [S1C Appendix](#).

Inference method

The inference method in this paper is designed for use with the partial observations and arbitrary test and patient turnover times often found in active surveillance data. As shown in [Fig 4](#), entering and exiting patients may have an unknown status (gray vertical bar) or be known to be uncolonized (white bar) or colonized (red bar). Patients are tested at arbitrary times (open circles), but typically not every patient in the hospital unit is tested at the designated time(s). The possible *true* state of the hospital unit can only be known if all patients are tested at the same time, but in practice, the state of the unit is often partially rather than fully observed. All test results obtained are assumed to be accurate.

In the next section, we describe cleaning and pre-processing of the active surveillance dataset, including **data augmentation**, a method for extrapolating test results forward and backward in time based on some simple assumptions (similar to the assumptions of Bootsma et al [32]). We then describe a likelihood function, that is, the probability of a single sequence of observations given a set of arbitrary parameters. Finally, we describe an inference method that maximizes this likelihood function. Because direct calculation of the likelihood function is computationally expensive, we outline multiple techniques that can be used to streamline this calculation, including a method to “compress” a patient-patient transmission matrix $P(Z)$ from size 2^n to $n + 1$ square ([S1N Appendix](#)); use of Jordan form to find the analytical solution to the matrix exponential of the rate matrix $R(t)$ ([S1O Appendix](#)); and other computational techniques to reduce the required memory and time required for a single likelihood

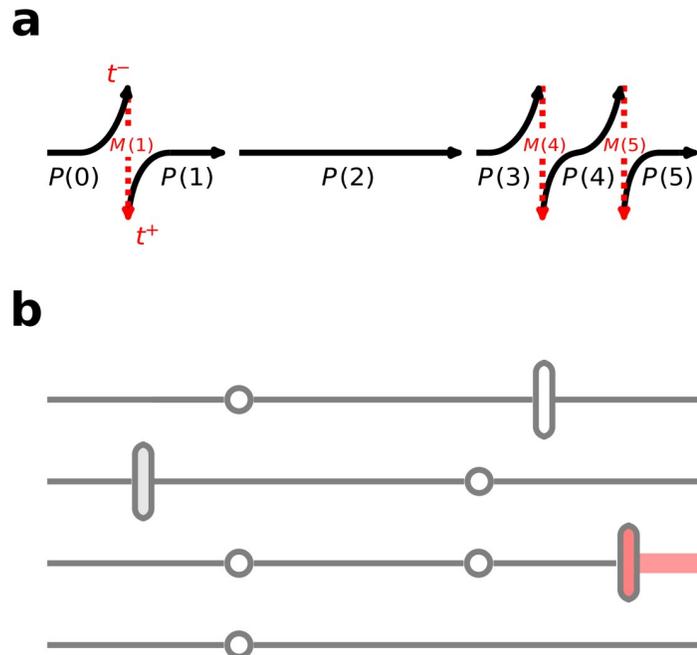


Fig 4. Example event timeline for a surveillance study performed on a 4-bed unit (with time flowing left to right). (a) Surveillance study timeline divided into non-turnover intervals (black arrows) by tests (spaces between black arrows) and exit/entry events (dotted red arrows). Each non-turnover matrix for the interval t_z to t_{z+1} has a corresponding matrix $P(z)$, and each exit/entry event at time t_z has a corresponding $M(z)$ matrix. (b) The events of a surveillance study in which each horizontal line represents the history of a bed, which may include exit/entry events (vertical bars) or tests (open circles). The color of the vertical bar indicates whether the incoming patient was uncolonized (white), colonized (red), or untested with unknown status (gray).

<https://doi.org/10.1371/journal.pone.0231754.g004>

calculation. We used these methods in conjunction with MATLAB's built-in optimization function `fmincon` to calculate the maximum-likelihood parameters for our data set.

Data pre-processing. Because the mathematical model assumes that no time elapses between exit and entry, we assumed that the “exit/entry” events of the model occurred at the recorded time of patient entry in the dataset. We assumed that once colonized, patients remained colonized throughout their stay, because even intentional attempts at decolonization have limited efficacy [65, 66]. Negative test results following positive test results were assumed to be false negatives, as intestinal excretion of CRE can be intermittent [67]. Patients who are uncolonized at one test time were assumed to be uncolonized at all times *prior* to the test, from and including the time of entry; patients found to be colonized at a particular test time were assumed to be colonized for all times at and *after* the test until exit. (The patient colonization status is still unknown after a negative test, before a positive test, and between a negative and a positive test because the exact time of conversion, if one occurred, is unknown). Thus, at each time at which testing was performed within the hospital unit, inferred test results were given to all other patients whose colonization statuses could be inferred. The final augmented data set of actual and inferred test results was used for all parameter estimation within this paper. Additional statistics and information on data pre-processing are available in [S2 Appendix](#).

Probability of a sequence of states. Consider a sequence of states and their corresponding matrices (either exit/entry $M(z)$ or patient–patient transmission probability $P(z)$ matrices) that occur at given times t_z . If all patients are tested during an observation, the state of the hospital unit will be completely known at that time. If not all patients are tested, only incomplete data regarding the state of the hospital unit will be available, as shown in the example testing

scheme of Fig 4. For convenience, we will re-index the states $\mathbf{b}(z)$ and the corresponding matrices in sequential order with integer indices j , allowing us to rid our notation of limits (such as t_z^- or t_z^+).

The re-indexed states and matrices will be notated as $\mathbf{a}(j)$ and $W(j)$, where $W(j)$ is a generic transition matrix that denotes either an exit/entry matrix $M(z)$ or the patient-patient transmission matrix $P(z)$ describing the probability of transitioning from state $j - 1$ to state j . Notice that $M(z)$ describes a state transition occurring instantaneously at exit/entry/bed contact, but $P(z)$ describes the probability of the state transition between two events at times t_{j-1} and t_j . The $P(z)$ probability matrix is created by exponentiating the continuous-time patient-patient transmission rate matrix over any interval of time between events such as exit/entry or tests. Thus, different $P(z)$ matrices may represent probabilities of state transitions over time intervals of varying lengths, unlike a typical discrete-time Markov model which calculates the probability over fixed time intervals such as a day.

Here, p_0 is the initial probability distribution, so $p_0(\mathbf{a})$ is the probability that $\mathbf{a}(0) = \mathbf{a}$, e.g., that if no time elapses, the state of the unit is the initial state. $W_{\mathbf{a}(i)\mathbf{a}(j)}(i)$ is the probability of transitioning from state $\mathbf{a}(i)$ at the i th event to state $\mathbf{a}(j)$ at the j th event. Notice that any amount of time can elapse between these events, so multiple instances of patient-patient transmission may occur between exit/entry and/or test events.

In the case of partial testing, there will be multiple possible states consistent with observations. For example, consider the first set of tests at time t_2 . If we assume all test results (for the white open circles shown) are negative, then the partially observed state of the unit is 00?0. The two possible underlying states consistent with test results are {0000, 0010}.

Let $B(j)$ be the set of all possible states consistent with test results at event j . Each state vector $\mathbf{b}(j)$ in the set $B(j)$ will have some components b_k (for k in the set of tested beds) whose status is known from test results, but the remaining components can take on any combination of values (0 or 1). Notice that if all tests are performed, the state of the system will be known and $B(j)$ will have only one element. If no tests are performed, all states are possible, so $B(j)$ will have 2^n elements. Thus, every observation with m tests performed (for $m \in \{0, 1, \dots, n\}$) yields a set of 2^{n-m} possible states consistent with its test results. The true state of the system $X(j)$ at event j is contained within the set of possible states $B(j)$. Therefore, the whole set of observations gives us the following information about the sequence of true system states at all of the test times:

$$[X(0) \in B(0)] \text{ and } [X(1) \in B(1)] \text{ and } \dots \text{ and } [X(\zeta) \in B(\zeta)]$$

The likelihood (a priori probability) \mathcal{P} of this set of observations is

$$\mathcal{P} = \Pr([X(0) \in B(0)] \text{ and } [X(1) \in B(1)] \text{ and } \dots \text{ and } [X(\zeta) \in B(\zeta)])$$

For brevity, we denote the sequence of states $\mathbf{i} = \mathbf{a}(0)$, $\mathbf{j} = \mathbf{a}(1)$, $\mathbf{k} = \mathbf{a}(2)$, $\mathbf{y} = \mathbf{a}(\zeta - 1)$, and $\mathbf{z} = \mathbf{a}(\zeta)$. Then the likelihood equation can be written as

$$\mathcal{P} = \sum_{\mathbf{i} \in B(0)} p_0(\mathbf{i}) \left(\sum_{\mathbf{j} \in B(1)} W(0)_{ij} \left(\sum_{\mathbf{k} \in B(2)} W(1)_{jk} \dots \left(\sum_{\mathbf{z} \in B(\zeta)} W(\zeta - 1)_{yz} \right) \right) \right) \tag{1}$$

(Note that this method is similar to Baum’s recursion method as used in the work of McBryde et al. [37]. The following method of obtaining submatrices is similar to the method by Bootsma et al. of reducing the number of possible states by incorporating observations using forward and backward vectors [32]).

Eq 1 can be written in matrix notation (and evaluated that way in MATLAB) if we introduce the following subvector and submatrices. Let $p_0(B(0))$ be the row vector with entries of

initial probabilities for each of the possible initial states in set $B(0)$. Let $W(j, B(j), B(j + 1))$ be the rectangular submatrix of $W(j)$ containing the rows corresponding with the states specified in $B(j)$ and the columns corresponding with the states specified by $B(j + 1)$. Finally, let $U(B(\zeta))$ be a column vector with all elements equal to 1 and the same number of elements as contained in $B(\zeta)$. Then the overall probability is

$$\mathcal{P} = p_0(\mathbf{a}(0)) W(0, B(0), B(1)) W(1, B(1), B(2)) \dots W(\zeta - 1, B(\zeta - 1), B(\zeta)) U(B(\zeta)) \quad (2)$$

To find the parameters for the prior-to-new patient probability ψ and the transmission rate γ , we must maximize the likelihood \mathcal{P} with respect to a particular sequence of observations: patient exit/entry events, the corresponding new-patient colonization statuses, and the test results of particular beds at given times.

Finding maximum likelihood parameters and error ranges. We used the function $\mathcal{P}(B; \gamma, \phi, \psi)$ (Eq 1) as the objective function to determine the likelihood of a sequence of observations B given input parameters γ, ϕ, ψ . The observations consisted of test times, locations, and results from active surveillance data for an n -bed hospital unit as well as census information (exit/entry times, the corresponding beds in which turnover occurred, and the colonization statuses of incoming patients if available). We determined the best-fit parameters by finding the maximum of the log-likelihood (the logarithm of Eq 1) using the MATLAB function `fmincon` with the following constraints: ϕ and ψ were probabilities with values between 0 and 1. (The parameters themselves remained on a linear scale). The patient-patient transmission rate γ is a probability per unit time that can have any non-negative value depending on the units chosen. Because estimates from epidemiological studies [68] or mathematical models [35, 39] suggested that the overall rate of transmission was less than 25 cases per 1000 patient-days of exposure (0.025), we chose a generous parameter search range between 0 and 2. An initial search of the space suggested a single maximum lay between 0 and 0.5, so the final parameter search was restricted to values between 0 and 0.5. Example plots from the likelihood landscape for $\gamma \in [0, 1]$ are shown in S11 Appendix. Error ranges were computed as described in Fig 5.

Techniques to improve inference. The continuous-time patient-patient transmission rate matrix allows use of exit/entry and test events at arbitrary times. However, this flexibility comes at a price: matrix exponentiation must be performed for each of the time intervals between events (approximately 500 intervals of different lengths for the rehabilitation unit data).

Calculating a matrix exponential is an inherently difficult problem because confluent or nearly confluent eigenvalues leads to a loss of accuracy, but algorithms that avoid eigenvalues tend to require more computer time and may be adversely affected by roundoff error [69, 70]. Finding the matrix exponential of the full detailed patient-patient transmission rate matrix becomes computationally expensive as the number of beds (n) increases because the number of matrix entries increases exponentially (as 2^{2n}). For units with a small number of beds (for example, $n = 4$ beds), the time needed is trivial because the matrices are 16 by 16, but for an $n = 13$ bed room, the full detailed state rate matrix has a size of 8192 by 8192. Consequently, if it is possible to decrease the dimensions of the transmission rate matrix or to factor the rate matrix so the matrix exponential can be calculated using only scalar exponentials, this step can be sped up significantly.

As described in S1N Appendix, we “compress” the patient-patient transmission rate matrix from size 2^n by 2^n to size $n + 1$ by $n + 1$ by converting entries in the full detailed matrix to their reduced matrix equivalents, similar to previous work analyzing SIS epidemic dynamics on a network of fully “connected” individuals [71]. Matrix exponentiation can then be performed

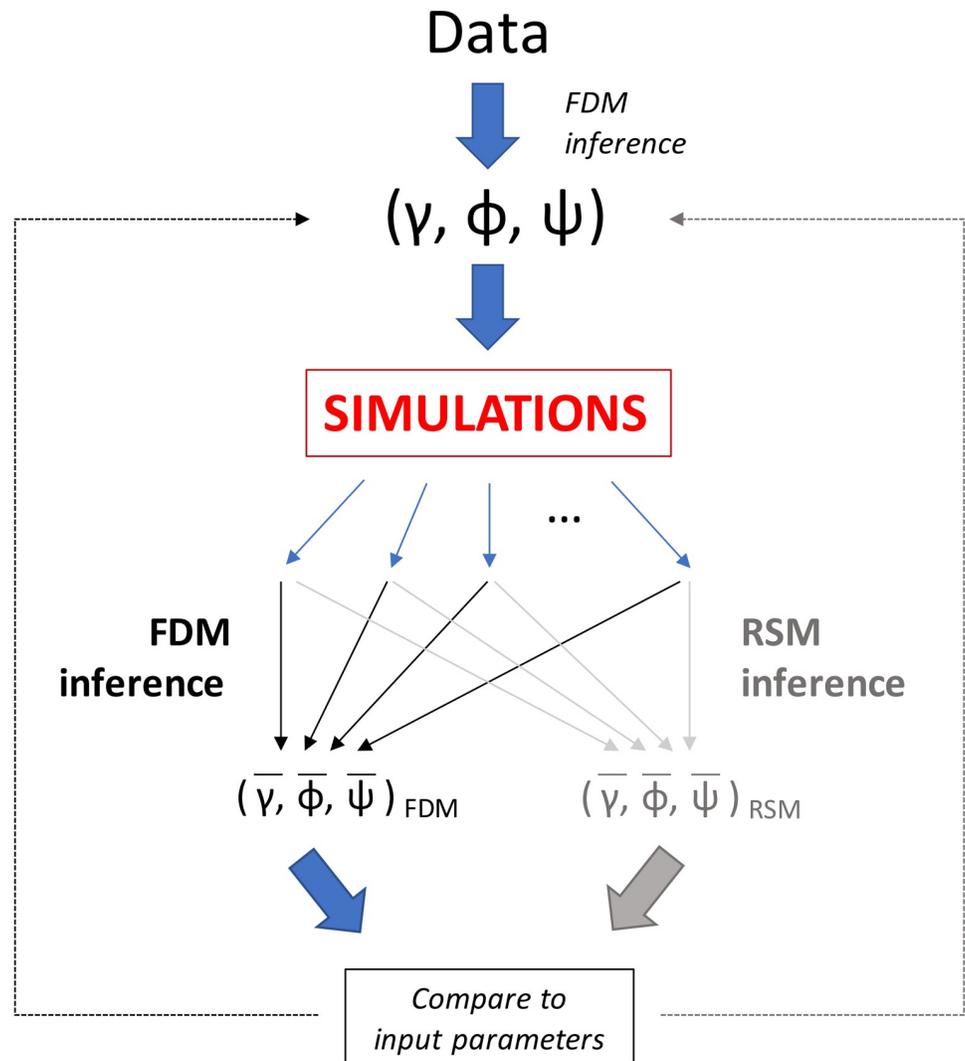


Fig 5. Comparison of inference methods of the full detailed model (FDM) versus the reduced state model (RSM) using simulations generated from the same set of input parameters (γ, ϕ, ψ) . First, the dataset from the rehabilitation unit was used to find the best-fit parameters using the FDM. These parameters were used to generate simulations. The FDM and RSM inference methods were used to generate two sets of best-fit output parameters. The mean best-fit output parameters $(\bar{\gamma}, \bar{\phi}, \bar{\psi})_{FDM}$ and $(\bar{\gamma}, \bar{\phi}, \bar{\psi})_{RSM}$ from the FDM and RSM inference methods, respectively, were found and then compared to the input parameters (γ, ϕ, ψ) .

<https://doi.org/10.1371/journal.pone.0231754.g005>

on the smaller matrix and the entries converted and “exploded” back to their correct locations in the full detailed matrix. This is a tremendous advantage because the number of reduced states increases only linearly with the number of beds (as $n + 1$) [71]. Although we cannot diagonalize the matrix because it is defective, we were able to find the analytic solution for the Jordan matrix exponential, as shown in [S10 Appendix](#). This solution allows us to calculate scalar exponentials of the Jordan matrix to find the matrix exponential.

Implementing the inference method. The surveillance data, microbiological test results, and hospital census data were pre-processed using Python and MATLAB (R2015b). The simulation and inference methods were implemented in MATLAB. Parameter estimation and simulation were performed in parallel using a combination of MATLAB and Bash scripts on Phoenix (now BigPurple), the high-performance computing cluster at New York University

Medical Center. Parallel processing was used for initial searches of parameter space and parallel simulations of the forward model.

Reduced state model

Reduced state model. The reduced state model is a continuous-time SIS Markov model in which the patient turnover is approximated as a Poisson process [72] in which patient exit/entry events occur continuously and independently at a constant average rate β . In essence, the reduced state model is a continuous-time classic birth and death process [73] applied to a susceptible-infective-susceptible epidemic model. The time between these exit/entry events is described by the exponential distribution, a memoryless probability distribution in which $1/\beta$ describes the mean time between exit/entry events (e.g., the average length of stay) [64]. Additional information and data supporting this assumption are described in **Fig B in S2 Appendix**.

The rate of increase (f_i) in the number of colonized patients from i to $i + 1$ is $f_i = \gamma i(n - i) + \beta(n - i)\phi$. Notice that f_i is governed by parameters for the turnover rate β , the patient-patient transmission rate γ , and the probability of pre-existing colonization ϕ . However, the rate of decrease is governed only by the parameters ϕ and ψ (the prior-to-new patient colonization probability): $g_i = \beta i(1 - \phi)(1 - \psi)$.

Let P_i be the probability of being in state i . As time passes, the probability of being in a particular state will change. For the state with i patients colonized, the differential equation for the change in P_i is $dP_i/dt = f_{i-1} P_{i-1} + g_{i+1} P_{i+1} - f_i P_i - g_i P_i$. At steady state, there is no change in the probability of being in any particular state, so the time derivatives of P_i can be set equal to zero. If we solve $dP_0/dt = g_1 P_1 - f_0 P_0 = 0$ for P_1 , we find that $P_1 = f_0 P_0/g_1$. By induction, we find the equations for $i \in \{2, \dots, n\}$ are $P_i = f_{i-1} P_{i-1}/g_i$. This equation shows the probability for the next higher state P_{i+1} in terms of the present state P_i . Substituting in recursively and solving for the probability P_i in terms of P_0 , for $i \in \{1, 2, \dots, n\}$, we find

$$P_i = \prod_{j=0}^{i-1} \frac{f_j}{g_{j+1}} P_0 \tag{3}$$

Because the sum of probabilities over all possible states is 1, we can solve for P_0 :

$$P_0 = \frac{1}{1 + \sum_{i=1}^n \left(\prod_{j=0}^{i-1} \frac{f_j}{g_{j+1}} \right)} \tag{4}$$

The equations for P_i and P_0 are general and hold regardless of the actual rates f_j or g_j . A detailed description of the reduced model is found in **S1K Appendix**.

The probability of all possible sequences of states consistent with (possibly incomplete) observations is analogous to the likelihood equation of the full detailed model (Eq 2, but the matrices have dimensions of $n + 1$ by $n + 1$, not 2^n by 2^n). In the reduced model inference method (described in **S1K Appendix**), we fit the probability distribution of states for the theoretical values against the distribution of states from actual data to estimate the minimum-error parameters.

Total prevalence. Total prevalence (TP), also known as the colonization pressure, can be defined as the fraction of patient-days during which patients are colonized. The theoretical total prevalence can be calculated from the reduced state model for any given set of parameters.

The total prevalence is the sum of the probabilities of each state weighted by the number of colonized patients in each state: $TP = 0P_0 + 1P_1 + 2P_2 + \dots + nP_n$. In terms of the rates of increase (f_i) or decrease (g_i), the total prevalence can be written as

$$TP = \sum_{i=0}^n iP_i = \sum_{i=0}^n \left(\prod_{j=0}^{i-1} \frac{f_j}{g_{j+1}} \right) iP_0 \quad (5)$$

Total prevalence, or the probability that at least 1 patient is colonized, is also equivalent to $1 - P_0$ (with P_0 defined in Eq 4). We can estimate the theoretical contribution of each colonization mechanism to total prevalence by varying the input parameters γ , ϕ , and ψ that govern the rates of increase f_i and decrease g_i in the number of colonized patients.

Comparison of reduced and full detailed model results

Fig 5 shows the process used to estimate error and check performance of our model. First, we used the best-fit parameters from the actual dataset (Table 1) to generate synthetic data. For this, we used the identical scheme of exit/entry events and testing times/locations as the true data from the rehabilitation unit. Then, we applied the inference method for the full detailed model to the simulated data sets to find the estimated parameters, their mean recovered values, and standard deviations for γ , ϕ , and ψ . Finally, for each parameter, we found the mean value over the 1265 trials and used the standard deviation to compute an error range (Table 1).

Results

We applied the full detailed state model (FDM) inference method to the 13-bed rehabilitation unit to determine maximum-likelihood parameters. We estimated the maximum-likelihood parameters for the 13-bed rehabilitation unit to be $\gamma = 0.00203/\text{day}$ per colonized/uncolonized patient pair, $\phi = 0.0497$, and $\psi = 0.000946$. For a scenario in a 13-bed unit in which 1 patient is colonized, there are 12 colonized-uncolonized patient pairs, so the probability per day of a transmission from the colonized patient to any uncolonized patient in the entire unit is about 0.02. A pre-existing colonization probability of about 5% suggests that about 5 out of every 100 patients is colonized. Finally, having a very small best-fit parameter for bed-patient colonization suggests that the probability of transmitting bacteria from a colonized patient to the bed or environment and subsequently to the next patient entering the bed (if uncolonized) essentially does not occur or is not detectable using the FDM method with available data.

In order to evaluate the reliability and reproducibility of these estimates, we used the best-fit parameters, exit/entry times, and testing schema from the actual hospital data as starting points for simulations. "Observations" (using actual hospital test times and locations) were performed on the simulation data, and the inference methods using the full detailed model or reduced state model were applied to the simulated data to find best-fit parameters. The inferred parameters were then compared with the input parameters for the simulations to

Table 1. Comparison of input parameters for simulations versus the mean estimated value (MEV) and standard deviations (SD) of parameters recovered using the inference method from the full detailed state model (FDM) or the reduced state model (RSM).

	Input	FDM MEV \pm SD	FDM error	RSM MEV \pm SD	RSM error
γ (per day)	0.002029	0.001869 \pm 0.000834	0.0788	0.002310 \pm 0.000417	0.1386
ϕ	0.049700	0.048655 \pm 0.012559	0.0210	0.066839 \pm 0.015914	0.3448
ψ	0.000946	0.022321 \pm 0.036288	22.6023	0.353808 \pm 0.004715	373.119
SSE		0.000458		0.124806	

Here, SD = standard deviation, SSE = sum of squares error. The model error per parameter is simply the fraction error $|x - x'|/x$, for true value x versus estimated value x' .

<https://doi.org/10.1371/journal.pone.0231754.t001>

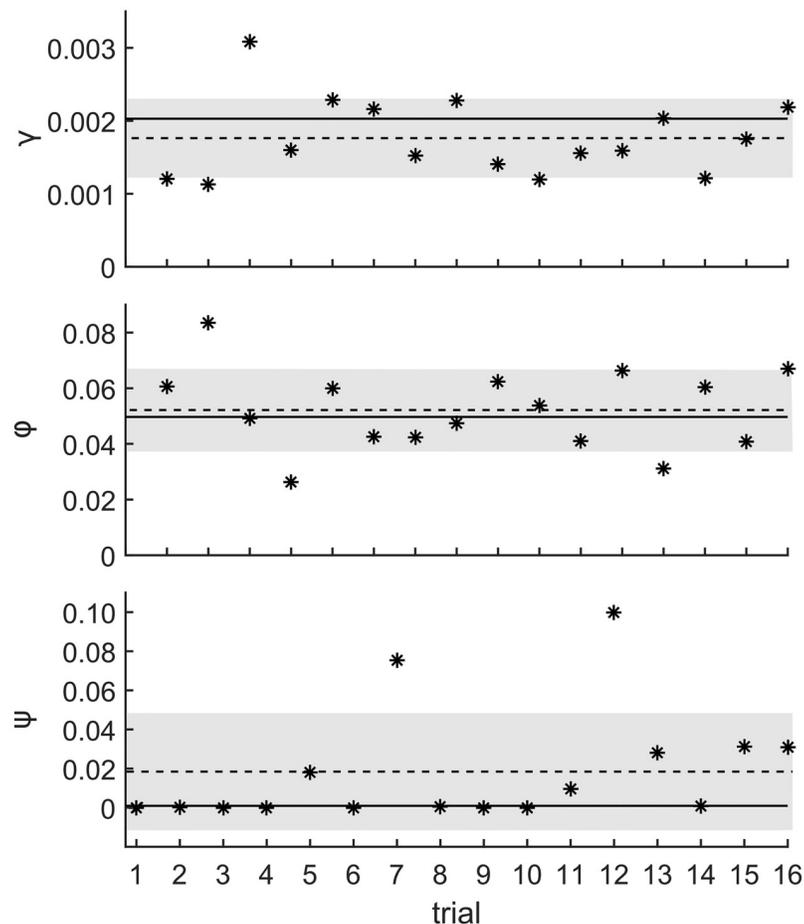


Fig 6. Range of parameter estimates inferred from 1265 simulations of the full detailed model. Range of parameter estimates inferred from 1265 simulations of the full detailed model created using the input parameters $\gamma = 0.00203$ per colonized/uncolonized patient pair per day, $\phi = 0.0497$, and $\psi = 0.000946$. Each star represents the estimated maximum-likelihood parameter for an individual trial, a dashed line represents the mean best-fit parameter over the 1265 trials, and the gray region shows the standard deviation from the mean best-fit parameter. The bold line shows the input parameter value.

<https://doi.org/10.1371/journal.pone.0231754.g006>

evaluate how well each inference method worked. A summary table of parameters, error ranges, and the sum of squares error for the full detailed model and the reduced model are shown in Table 1. Visualizations of best-fit parameter results for individual simulations are available in Fig 6 for the FDM and Fig G-Fig I in S1L Appendix for the RSM. Fig 6 shows the results of parameter estimation using the full detailed state model (FDM) inference method applied to 1265 simulations that were created using the above maximum-likelihood parameters as input. The mean recovered parameters were $\gamma = 0.0018 \pm 0.0008$ per day per colonized/uncolonized patient pair, $\phi = 0.05 \pm 0.01$, and $\psi = 0.02 \pm 0.04$, where the error range is the standard deviation of estimated parameters. Applying the reduced state model (RSM) inference method on 1265 simulations (Fig G-Fig I in S1L Appendix), we found the average estimated best-fit parameters to be $\gamma = 0.0023 \pm 0.0004$ per day for each colonized/uncolonized patient pair, $\phi = 0.07 \pm 0.02$, and $\psi = 0.35 \pm 0.005$. Note that for the FDM, the “true” (input) parameters are all contained within the error range. The RSM also has relatively good estimates for the input γ and ϕ , both of which are contained within the error range. However, the RSM estimate for ψ is highly—although consistently—inaccurate (with a narrow error range); this suggests

that the RSM model is not capable of accurately inferring ψ . As described at length in the discussion regarding the **non-equivalence of the full and detailed states for inference**, patient-patient transmission for γ involves only the *number* of simultaneously colonized patients in the unit, which is captured in the state data for both the FDM and RSM. However, determination of whether colonization of a new patient from a prior patient has occurred requires knowing *which* patients in the unit were colonized and when, not just how many—information that is captured by the FDM but not RSM.

For the full detailed model, the error in the estimate for ψ is high and remains so despite the large number of simulations performed, each of which was created with the exact exit/entry events, observations, and length of time as the actual data from the study. These simulations are statistically identical and have a finite amount of data. Because the error involved in parameter fitting has components of both bias and statistical error, the large number of trials performed will only eliminate the statistical error, not the bias. In the case of ψ , the parameters are constrained to be non-negative, introducing a bias, especially as the value for ψ is effectively zero. These results reveal that the mechanism of prior-to-new patient colonization is relatively unimportant compared to pre-existing colonization and patient-patient transmission. This suggests that for a realistic hospital surveillance situation with moderate testing compliance and low frequency of prior-to-new patient colonization, estimating the *value* of ψ will be difficult using the FDM method. At larger values of ψ , however, the FDM method may be able to produce better quantitative estimates.

Effect of a long-stay patient

In the actual rehabilitation unit data, a single patient was observed to have an unusually long length of stay (259 days) and to have a positive test for colonization approximately mid-way through the hospital stay. To investigate the possible influence of a long-stay patient on other patients in the unit, we performed 16 sample simulations using actual entry/exit and observation times but simulation colonization and transmission events. The disproportionate effect of this long-stay patient is apparent in 16 sample simulations shown in [S1E Appendix](#).

In 7 of 16 simulations, the long-stay patient becomes colonized with a mean time to colonization of 128 days, essentially the midpoint of the patient's stay, and an average time spent uncolonized of 130 days. The mean total prevalence of simulations in which the long-stay patient became colonized was 526 colonized patient-days. In 3 of the 7 simulations in which the long-stay patient becomes colonized, colonization occurs before the mid-point of the patient's stay after 38, 20, and 42 days, respectively.

After excluding the days colonized (if any) of the long-stay patient from each simulation's total prevalence, there was a mean of 396 colonized patient-days over the 7 simulations with a colonized long-stay patient and a mean of 275 colonized patient days over the 9 simulations in which the long-stay patient remained uncolonized, resulting in a difference of 120 colonized patient days between simulations with colonized versus non-colonized long-stay patients. This resulted in an increase of 44%. The times and duration of the simulated long-stay patient's colonization (if any) are shown in [S1D Appendix](#).

Total prevalence

Colonization pressure, or the fraction of colonized patients in a hospital unit, is one risk factor for a patient becoming colonized or infected [25, 74]. We can estimate the mean colonization pressure, or the fraction of colonized patient-days, with the theoretical total prevalence calculated from the reduced state model ([Fig 5](#)) and input parameter values. Thus, we can determine the expected fraction of colonized patient-days over the course of the study at particular

parameter values. The theoretical total prevalence calculated from the best-fit parameters was 0.07 of the total colonized patient-days of the entire active surveillance study (the **bold star** in all panels). Thus, for the case where active surveillance was performed in a hospital unit of 13 beds and in which complete occupancy is assumed, the model predicts that 7% or about 380 of 5421 total patient-days would be colonized.

Furthermore, we can estimate the fraction of total colonized patient-days over the course of the study that are attributable to various mechanisms of colonization. Fig 7 shows how the total prevalence would vary if one parameter was fixed at the best-fit parameter value and the other two parameters were varied. (Expanded versions of each plot are available in [SIM Appendix](#)). For example, in Fig 7f, ψ is fixed at $\psi = 0.0009$, and the plot shows how total prevalence varies with increasing ϕ and γ . The bold star shows the total prevalence (0.07) with all parameters at the best-fit values ($\gamma = 0.002$, $\phi = 0.05$, and $\psi = 0.0009$). Because prior-to-new transmission was found to be effectively zero, in absence of patient-patient transmission, the total prevalence increases linearly with and is equal to the pre-existing prevalence, shown as the linear diagonal line for $\gamma = 0$ in (Fig 7f). Thus, given a theoretical total prevalence of 7%, with $\phi = 0.05$, 5% of mean colonization-days can be attributed to pre-existing colonization; the remaining 2% can be attributed to patient-patient transmission, with prior-to-new colonization contributing a negligible amount. Note that while ϕ describes the fraction of entering patients that are colonized, those patients will remain colonized for the entirety of their stay, so the colonized patient days will compose 5% of total patient days.

Fig 7a shows the effect of prior-to-new patient colonization probability (ψ) on total prevalence at various levels of patient-patient colonization (γ) but fixed pre-existing colonization ($\phi = 0.05$). These results suggests that the total prevalence is not very sensitive to the effect of prior-to-new patient colonization at the best-fit parameter values (bold star). However, as the patient-patient transmission rate increases, the effect of prior-to-new bed-patient transition becomes more pronounced, with the greatest slope (increase in total prevalence with increasing ψ) occurring with $\gamma = 0.008$ to 0.01.

Fig 7b shows the relationship between the prior-to-new patient colonization probability (ψ) and the total prevalence with varying pre-existing colonization probabilities (ϕ) but a patient-patient transmission rate fixed at the best-fit value ($\gamma = 0.002$ /day per colonized/uncolonized patient pair). Like Fig 7a, the total prevalence is quite insensitive to changes in ψ at the best-fit values of $\phi = 0.05$ and $\gamma = 0.002$ /day per colonized/uncolonized patient pair, but it has the greatest increase with increases in ψ (has the greatest slope) at $\phi = 0.5$. However, at $\phi = 0$ or $\phi = 1$, in which no patients or all patients are colonized, respectively, the total prevalence is constant at 0 or 1, regardless of the prior-to-new patient transmission probability.

Fig 7c shows the total prevalence at different values of the patient-patient transmission rate (γ) and pre-existing colonization (ϕ), with ψ fixed at the best-fit value of $\psi = 0.0009$. With zero ($\phi = 0$) or all ($\phi = 1$) patients entering colonized, the patient-patient transmission rate is irrelevant because the total prevalence is, respectively, 0 or 1. At the best-fit parameter values (bold star), the total prevalence is about 0.07, but doubling the patient-patient transmission rate to $\gamma = 0.004$ /day (per colonized/uncolonized patient pair) increases the total prevalence to 0.11, an increase of 0.04. However, if the pre-existing colonization ϕ were to double from $\phi = 0.05$ to $\phi = 0.10$, the total prevalence would increase to 0.14. At $\phi = 0.10$, doubling γ to $\gamma = 0.004$ /day increases the total prevalence to 0.20. At $\phi = 0.25$, doubling γ increases the total prevalence from 0.32 to 0.48, a much greater absolute increase of 0.16. Thus, modest increases in pre-existing colonization will magnify the effects of small changes in patient-patient transmission.

Fig 7d shows the effect of patient-patient transmission and prior-to-new colonization on total prevalence with pre-existing colonization fixed at the best-fit value of $\phi = 0.05$. At the best-fit parameters (bold star), changing the patient-patient transmission does not greatly

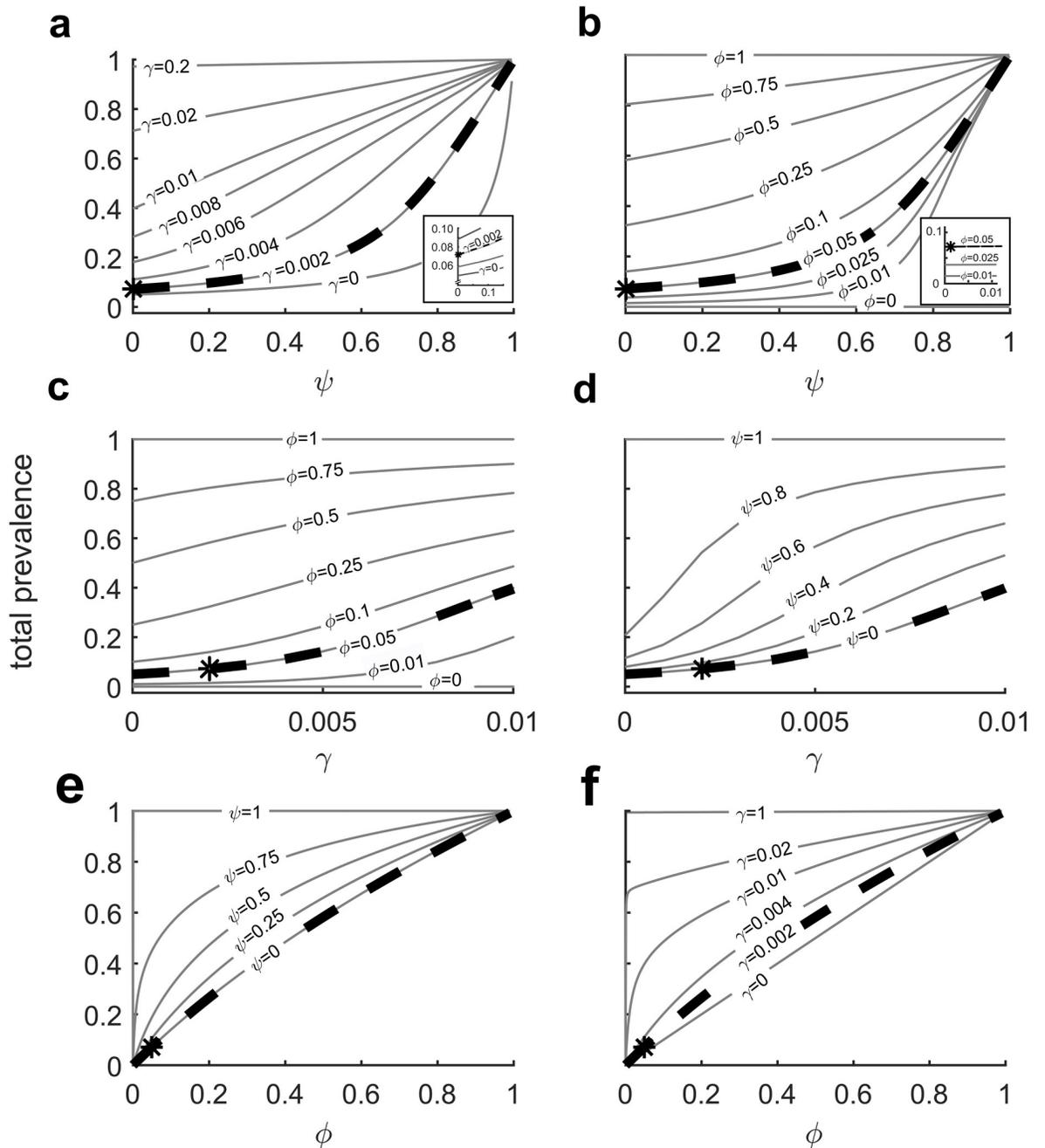


Fig 7. Predicted total prevalence of colonization in a 13-bed hospital unit as a function of the model parameters γ , ϕ , and ψ . Each panel shows the predicted total prevalence as a function of one model parameter for selected values of a second model parameter, with the third parameter held constant at the best-fit value for the rehabilitation data (inferred using the full detailed model inference method). The bold star (*) shows the point at which all parameters have the inferred best-fit values ($\gamma = 0.002029/\text{day}$, $\phi = 0.0497$, and $\psi = 0.000946$). The bold dashed curve shows the case in which two of the three parameters have their best-fit inferred values while the third parameter is varied. The panels show total prevalence as a function of (a) ψ and γ , (b) ψ and ϕ , (c) γ and ϕ , (d) γ and ψ , (e) ϕ and ψ , and (f) ϕ and γ . The insets in panels (a) and (b) show enlargements of the area near the origin. In the inset of panel (a), the values of γ for the contours from bottom to top are $\gamma = 0, 0.001, 0.002, 0.003$. Enlarged versions of all panels are available in S1M Appendix.

<https://doi.org/10.1371/journal.pone.0231754.g007>

increase the total prevalence, but as the prior-to-new patient colonization probability increases, small increases in patient–patient transmission are magnified and subsequently have a greater effect on total prevalence.

Fig 7e shows the effect of prior-to-new patient colonization on total prevalence with fixed patient–patient transmission at the best fit value of $\phi = 0.05$. The bold dotted line showing the total prevalence at the best-fit prior-to-new patient colonization ($\psi = 0.0009$) is essentially the diagonal line, suggesting that the amount of total prevalence attributable to prior-to-new bed-patient colonization is effectively zero.

Fig 7f, as discussed above, shows the relationship between pre-existing prevalence (ϕ) and total prevalence at different patient–patient transmission rates, with prior-to-new colonization fixed at the best fit value of essentially zero ($\psi = 0.0009$). The diagonal line at $\gamma = 0/\text{day}$ shows that total prevalence equals pre-existing prevalence if there is no patient–patient transmission. The increasing values of γ above the diagonal show the effect of patient–patient transmission on total prevalence. At the best-fit parameter values (bold star), the increase in total prevalence contributed by patient–patient transmission is small. However, as the patient–patient transmission rate increases, the fraction of total prevalence contributed by patient–patient transmission becomes greater. Additionally, at very high values of γ , the initial slopes become nearly vertical, suggesting that colonization spreads rapidly (an outbreak caused by entry of an initially colonized patient) and the majority of total prevalence is attributable to patient–patient transmission.

Discussion

In this paper, we developed a stochastic mathematical susceptible-infective (SI) model of bacterial transmission within hospital units that uses the full detailed state model (FDM). The FDM tracks individual patient colonization statuses, and it includes three mechanisms of colonization: patient–patient transmission, prior-to-new patient colonization, and pre-existing colonization. The FDM employs a stochastic hybrid continuous-time/discrete-time Markov model so that it can incorporate arbitrary times for patient exit/entry and testing. We used the FDM to develop an inference method that incorporates active surveillance data and hospital census data to distinguish parameters for three routes of transmission, including from prior to new patients. This inference method allows incorporation of incomplete test results and calculates the total probability of *all* possible sequences of states consistent with observation. We also compare the FDM with a reduced state model (RSM) similar to previously published models for in-hospital transmission [9, 32, 34, 37, 38, 41–43, 46, 47, 75] that track the *number* of colonized or infected patients rather than the status of each individual patient.

In the context of previous stochastic models [31] of patient colonization within hospital units (compared in S1 Table), this work incorporated five major modeling choices: 1) using a detailed rather than reduced state model (i.e., tracking individual patient statuses versus counting the number of colonized patients); 2) modeling evolution of states with a hybrid of discrete-time and continuous-time approaches; 3) incorporating three mechanisms of patient colonization, including prior-to-new (bed-patient) colonization; 4) allowing for incomplete observations and patient turnover at arbitrary times; and 5) calculating likelihood directly while incorporating partial observations.

The majority of models of in-hospital transmission use the reduced state, counting the number of individuals in each compartment [32, 34, 37, 38, 41–43, 46–48, 75]; the full detailed state is used far less frequently in forward model formulation [33] although individual patient statuses may be tracked during MCMC algorithms [34, 35, 39]. Of note, the general model of López-García & Kypraios [36], if formulated with each patient as a compartment, would be

equivalent to using a detailed state model with individual transmission rates (which would not be able to be represented as a reduced model). Although the reduced state may be sufficient for inferring patient-patient transmission rates—and in fact, the “compression” method described in [S1N Appendix](#) effectively transforms the full detailed state to the equivalent reduced state *between* events—the full detailed model is required at exit/entry and test events to track individual patient colonization statuses, which improves the model’s ability to uniquely identify parameters for bed-patient and pre-existing colonization. The reduction of the full detailed model is equivalent to the approach of graph-automorphism driven “lumping” by Simon and Kiss, who show that the exact solution of a continuous-time Markov Chain individual-based epidemic model converges to that of an ordinary differential equation-based mean field (e.g. reduced) model [76]. However, although the full detailed model and reduced model can be shown to be equivalent in formulation of the *forward* model (both statistically and in simulation), for purposes of *inference* they are not equivalent. The full detailed model performs better because it retains information about individual patients that can distinguish pre-existing or prior-to-new patient colonization from patient-patient colonization.

Most stochastic hospital transmission models use either continuous time [37, 38, 41–43, 48] or discrete time [32, 33, 40, 47] Markov models. To our knowledge, this model is the first in-hospital transmission model to incorporate a hybrid continuous-time/discrete-time approach, and also the only model that explicitly incorporates bed-patient colonization, although previous models included an environmental transmission component [36]. However, like many other groups [77, 78], we incorporated transmission from infected or colonized patients to susceptible patients [29] and included pre-existing colonization [32, 38, 42, 46–48].

Although multiple models incorporate incomplete observations [32, 33, 34, 42, 75], our approach is most similar to previous efforts that incorporated arbitrarily timed tests [34]. Similar to previous methods, we assume perfect specificity [35, 39], but unlike other groups, we do not directly estimate parameter sensitivity [34, 35, 39, 40, 47]. We instead use a “once colonized, always colonized” approach to testing, assuming that even if colonization-positive patients test negative, they nevertheless remain colonized because decolonization occurs on a timescale of months to years [67] and patients can retest positive after one [19] or even multiple [67] negative swabs.

We build on previous work directly calculating the likelihood of continuous-time Markov models [38, 41] and combine it with other approaches such as a discrete-time method that uses observations to reduce the number of possible states [32] and incorporates exact times of testing and turnover [36, 35, 39, 46–48]. A hybrid continuous-time/discrete-time formulation allows the flexibility to incorporate events that occur at varying time intervals while also accounting for events that are modeled as occurring instantaneously, such as patient exit/entry and exposure to a potentially contaminated bed. This eliminates a parameter and avoids assumptions that turnover times have a constant rate [30, 38] or exponential distribution [36]. The technique of compressing the continuous-time portions of the model incorporates techniques similar to previous work reducing the dimensionality of an exact stochastic SIS epidemic network with 2^n possible states [71] and “lumping” over a complete graph using automorphism [76]. The overall approach to calculating likelihood is generalizable to any Markov model that can ultimately be represented as a series of transition probability matrices. Although direct calculation of the likelihood of all possible permutations of events consistent with observed data is difficult or even “intractable” [39, 47], the problem is simplified when represented in matrix form (similar to previous work [32]): test results reduce the number of possible states and thus the dimensions of the sub-matrices representing the probability of transitions from event to event. Unlike approaches incorporating MCMC augmentation of data with “guesses” about the times of colonization events [34, 35, 39], the likelihood of all

observed test results and unobserved colonizations is already incorporated into the likelihood and does not require additional computation during parameter estimation. However, use of MCMC methods could allow incorporation of prior information about parameters.

Mechanisms of colonization

To our knowledge, this model is the first in-hospital transmission model to explicitly include prior-to-new transmission [79], although many previous stochastic models of in-hospital transmission have included pre-existing colonization and patient-patient transmission [32, 34, 38, 41, 48]. Although antibiotic exposure was found to be a risk factor for CRE colonization and infection in long-term acute care facilities [2, 80] and hospitals [18, 81–83], patients were generally not admitted into the rehabilitation unit while on intravenous or parenteral antibiotics, so we choose to assume that all patients have equal susceptibility to bacterial transmission. We also assumed that patient-patient transmission rates are linearly proportional to the number of colonized-uncolonized patient pairs, consistent with findings that the odds ratio of transmission in long-term acute care hospitals increased approximately linearly with increased colonization pressure [74] and that a linear patient-patient transmission model fit better than a non-linear model for medical hospital units [35]. Our model did not include background colonization, as two previous modeling efforts found no significant difference in model output or goodness-of-fit between mathematical models that did or did not include background colonization [35, 39]; one even found a slight preference for a no-background model in hospital medical units [39]. The true background transmission rate is unclear as most datasets included only 1-2 colonization mechanisms without prior-to-new patient colonization, but the fraction of acquired colonization attributable to mechanisms other than patient-patient transmission may range from none [39] to approximately half [84] or three-quarters [39, 68]. In our model, inclusion of a background colonization rate would be equivalent to adding a random noise term for sources including spontaneous emergence of resistance [85], unmasking of existing colonization by antibiotics [74], and transmission from outside sources such as visitors or equipment [47]. De novo creation of resistant strains seemed unlikely as the dataset used showed primarily clonal transmission of the ST258 strain of *K. pneumoniae* [25]. Unmasking of colonization was also less likely in a rehabilitation unit than in an intensive care unit, as the patients were required as a condition of admission to be able to participate in their care, and thus were rarely on parenteral or intravenous antibiotics. Plasmid movement was also not included as it occurs infrequently [86]) and therefore likely does not significantly affect results given the short average patient length of stay. However, if the model was applied to a dataset tracking mobile genetic elements for resistance [1] rather than a specific strain of CRE, transmission could be interpreted as the spread of those genetic elements [87, 88] from patient to patient.

Non-equivalence of the full and detailed states for inference

The classic SIR compartment model represents the state of a population as the numbers of susceptible, infective, and resistant patients. Most models of bacterial spread within hospital units [32, 38, 41] count only susceptible and infective patients and use longitudinal prevalence as the state of the unit, with the exception of one model of *S. pneumoniae* carriage within families, which tracked the infection statuses of individual members within households [33]. As shown in Fig 5, we compared the FDM against a simpler continuous-time, reduced-state model (RSM) in which only the prevalence, i.e., the number of colonized or uncolonized patients, was considered (similar to a susceptible-infected-susceptible model). In the RSM, the same three routes of transmission were included, but turnover was assumed to occur at a constant

rate β which was estimated from patient census data. Surprisingly, although the two models are equivalent for the forward problem of creating a simulation using given input parameters, the FDM and RSM are not equivalent for the inverse problem of parameter inference. Using the same parameters and initial state of the hospital unit, both the FDM and RSM give equivalent statistics and predictions about future states, such as the distribution of colonized patients at steady state. In the context of the inverse problem, however, the FDM has a greater ability to distinguish and uniquely identify parameters for patient-patient transmission versus pre-existing colonization from data.

For example, consider a scenario in which hospital unit testing occurs after each of three patient turnover events. At each test time, there is a single colonized patient within the unit (the reduced state), but inspection of the full detailed states could show different scenarios, such as a series of three patients in the *same* bed being colonized (001, 001, 001) or three patients in *different* beds being colonized (001, 100, 010). The former scenario suggests prior-to-new colonization, but the latter suggests pre-existing colonization. However, with the reduced model, both scenarios yield an identical series of states (1, 1, 1), discarding location information that could help distinguish the colonization mechanisms. This may account for some of the difficulties of past modeling efforts to uniquely identify parameters for patient-patient transmission versus spontaneous colonization [38] or pre-existing colonization, as aggregating data into counts diminishes the available information [41].

Cooper et al. used a hidden Markov model that required at least 40 months of contiguous data with at least 1 case per month. For a series with fewer than 20–30 observations, maximum likelihood estimates may fail to converge, and collinearity between parameter estimates can lead to identifiability problems. This may result from a loss of information that occurs with data aggregation [41]. We observed a similar effect in comparing the ability of FDM and RSM models to recover input parameters from “observations” sampled from simulated data created with those parameters and the identical test scheme used in hospital surveillance. The difference between input and recovered parameters for each inference method reflects the effects of incomplete information—including encoding state as the point prevalence (the number of colonized patients at a given test time)—and randomness.

Usage of the full detailed state reduces the problem of parameter collinearity or non-identifiability. The FDM model could easily distinguish parameters for patient-patient transmission and pre-existing colonization in simulated data, and its estimate of the prior-to-new-patient colonization probability was improved compared to that of the RSM, which consistently overestimated ψ , as shown in Table 1. The FDM model had a smaller estimated error range than the RSM model, although the estimate of ψ still had large uncertainty, likely because of the small or near-zero parameter value (see additional comments in “Epidemiological Relevance” below). In general, the overall parameter estimation error of the FDM (0.0005) was much smaller than that of the RSM (0.12). These results show that the inference method can in fact recover estimated parameter values reasonably close to the parameter values used to create the simulated data. The sum of squares error (SSE) for the FDM is smaller than that of the RSM, suggesting that parameter estimation for the FDM is better than for the RSM. Thus, tracking individual patient information of the detailed state can improve parameter identification for different mechanisms of transmission.

Computational challenges of partial observation

The approach of looking at all particular states consistent with observations has been applied to reduced state models [32, 37]. Although conceptually similar, calculation of the likelihood function is far more difficult in the FDM case because the number of states consistent with an

observation increases *exponentially* with the number of unobserved patients. If all beds are tested at a given observation time, there is only one possible state, but in reality, most surveillance studies do not have 100% compliance and test only *some* beds in a unit at a given observation time. Testing reduces the number of possible hospital unit states conditioned on the test results: if m of n patients in the unit are tested, then there are 2^{n-m} (FDM) or $2(n - m) + 1$ (RSM) possible states consistent with the test results at that observation time. The number of possible states was reduced even further by using the assumption of “once colonized, always colonized” (similar to previous approaches [32, 37]). A number of groups have used MCMC methods to sample over the space of states with times of colonization consistent with test results [34, 35, 39, 47].

Despite the exponentially larger number of possible states in the detailed model, the inference method presented in this paper still permits exact likelihood calculation that takes into account all possible sequences of hospital unit states consistent with partial observations, although this may be computationally expensive for larger units. The likelihood calculation is generalizable to any Markov model that can be reduced to a series of probability transition matrices. (In this paper, we transform the rate matrices from the continuous-time portions of the hybrid model into probability matrices for state changes over unequal time intervals). Likelihood calculations were made computationally tractable through the use of matrix “compression” (S1N Appendix) and a method for exact matrix exponentiation based on the Jordan form (S1O Appendix). These techniques drastically reduce the time and computer memory needed for calculating the multiple large matrix exponentials required for the continuous-time portions of the model between exit/entry events. After implementing these changes, calculation of the likelihood of a single set of parameters over the study period data (weekly sampling of 13 beds over 1 year) decreased from hours to approximately 7 minutes using lightly optimized MATLAB code on the Center for Health Informatics and Bioinformatics cluster. Use of a compiled rather than interpreted language or a system with additional random access memory (RAM) would likely decrease computation time significantly.

Epidemiological relevance

Colonization pressure is the proportion of patients already colonized or infected with pathogen [16]. In this study, we define it as the fraction of colonized patient-days of total patient-days in a hospital unit for a given period of time such as a month or year (as opposed to point prevalence, estimated as the fraction of positive tests at a given testing time [37]). This is equivalent to the theoretical total prevalence from Eq 5, the probability that at least one patient is colonized. (Note that the total prevalence is not equivalent to the probability of pre-existing colonization, which is the fraction of *incoming* patients that are colonized). Estimating the colonization pressure can be problematic if not all patients were tested and/or if colonized patients are tested at a different frequency than untested patients; in this work, testing is assumed to be arbitrary rather than random, so using clinical test results for patients suspected of having colonization or infection will not bias the results. Using best-fit parameters from the FDM, we were able to estimate the total prevalence at different parameter values and also estimate the relative contributions of the different mechanisms of colonization to the total prevalence. The theoretical total prevalence was estimated to be approximately 7%, similar to an estimate of 5.4% for New York City hospitals [25]. We were able to estimate the total prevalence attributable to patient–patient transmission (2%) over the baseline pre-existing colonization (5%). The best-fit patient-patient transmission rate of 0.002 per colonized-uncolonized patient pair per day, or 2 per 1000 colonized-uncolonized patient pairs per day, is consistent with the work of Hilty et al., which found a rate of between 0.0006 and 0.002 per colonized uncolonized patient-pair per day for extended-spectrum beta-lactamase *K. pneumoniae* [68], assuming that the colonized study patients had

been exposed to only one index case at a time. The estimate of pre-existing colonization (e.g., colonization of patients prior to entry into the hospital unit) is consistent with an estimate of community CRE prevalence of 5.6-10.8% in the United States [10] and observations that many patients were already colonized upon entry [1, 89]. Patients with pre-existing colonization may have become colonized during previous hospitalizations [67] (as was also previously found for VRE [90]), or within the community [68]. This suggests that infection control procedures within the rehabilitation unit should focus on identifying and isolating colonized patients on admission, and that improvements in handwashing [91] and other measures targeted at decreasing patient-patient transmission can only lower total prevalence to the baseline of pre-existing colonization. However, if pre-existing colonization prevalence increases, measures to reduce patient-patient transmission become increasingly important.

Our model found that the contribution of prior-to-new-patient colonization to total prevalence was effectively zero, suggesting that environmental contamination of CRE in the rehabilitation unit may be a lower-yield target for intervention. This is consistent with results from a study which showed that environmental contamination was a minor contributor to overall transmission, despite an increased risk of acquisition for patients admitted to rooms previously occupied by patients colonized by MRSA or VRE [79]. Our result differs from a study of ICU colonization by *K. pneumoniae*, which showed that occupying a bed previously occupied by a colonized patient was a major risk factor for colonization (odds ratio 4.8) [26]. However, this may reflect a difference in illness severity of ICU patients compared to rehabilitation patients or local factors such as cleaning methods [92] and room layout. In the rehabilitation unit of this study, patients were frequently transported in their beds from other locations rather than being placed into the same bed as the previous patient. In this case, the prior-to-new-patient colonization parameter would only describe colonization from surfaces surrounding the bed (such as sink drains [93], ventilators [94], tables or curtains) that remained in the same location and may have had fewer opportunities for patient contact. This is consistent with a study of CRE environmental contamination that found that 75% of the contamination occurred in beds and 25% in the surrounding environment, although there was a large variation in the degree of environmental contamination between different patients and even the same patient at different times [95]. Thus, the poor estimation of ψ may have been a result of the rarity of event occurrence or a mismatch between model assumptions and hospital conditions; likely this estimate would improve in situations with fewer confounding factors or in situations in which prior-to-new-patient colonization occurred more frequently. The model does not account for prior-to-new patient transmission occurring at distant time intervals or via methods other than patient-bed-patient transmission, as one institution found delayed transmission between patients inhabiting the same room even months apart [96].

The simulation results suggest that patients with longer-than-average hospital stays may become reservoirs of transmission. These long-stay patients may contribute disproportionately to outbreaks by acting as a source for multiple patient-patient transmission events. Thus, targeting long-stay patients for additional screening and possible isolation or decolonization (if possible) may be a high-yield infection-control measure, as colonized patients with shorter stays have fewer opportunities for transmission. Furthermore, additional infection-control measures may be considered to prevent colonization of long-stay patients to protect *other* hospital patients.

Other limitations

Our model identifies as “transmission” only cases in which patients are simultaneously present in the hospital unit for some length of time. For pathogens in which colonization or its detection might be delayed significantly, the model may not be appropriate.

Although use of the hybrid continuous-time/discrete-time model allows use of arbitrary exit/entry times and testing times, it is computationally expensive to calculate the likelihood function. The real “trade-off” in the choice to represent state as a scalar (the number of colonized/infected patients) versus vector (a list of the statuses of the patients in a unit) is that the size of the matrix increases linearly versus exponentially, respectively. However, the assumption that patients are equally likely to colonize or be colonized by any other patients within the unit may be unwarranted in hospital units with large numbers of patients. Thus, there will be an upper hospital unit size limit in applicability of this method because of the assumptions, regardless of the computational cost.

Additionally, some of the mathematical and computational techniques presented in this paper and its appendices exploit the specific structure and assumptions of our colonization and transmission model—in particular, the methods of matrix “compression”, factorization of γ , and use of Jordan form. Future modifications to the mathematical model that would eliminate some of the symmetries in patient–patient transmission (such as incorporation of additional parameters for spatial transmission or random colonization) would prevent use of these matrix compression methods on the continuous-time portions of the model. However, removal of these symmetries, which make the matrix defective, will allow use of standard techniques for matrix exponentiation such as diagonalization.

We did not calculate test sensitivity for reasons described previously, but incorporation of sensitivity into the model either as part of the likelihood framework or via a Bayesian method would be a valuable future extension of the model.

Conclusion

Although new tools such as KlebSeq [97] promise to make sequencing more widely available, hospitals likely will continue to have older methods of CRE detection in place. Our mathematical model and inference techniques provide a method for hospitals to estimate the fraction of colonization attributable to nosocomial transmission without need for sequencing, even with incomplete testing of incoming patients. (Should sequencing become available, data from individual strains or for particular resistance plasmids can be used as input, although there is some minimum amount of data required for parameter estimation). Because hospitals in the United States are penalized by the government for patient infections [98, 99], this method to distinguish pre-existing from hospital-acquired colonization may prove valuable, especially as it does not require additional resources and personnel for genetic sequencing [88] but can use existing testing methods. It is applicable to any nosocomial pathogen for which appropriate surveillance data exists and for which the mechanisms and assumptions of the models described in this paper are applicable. In particular, the model may be well-suited for describing MRSA, which colonizes the epithelium [100] and can be transmitted by direct skin and fomite contact [101, 102]. Other potential pathogens for which the model may be well-suited include *C. difficile*, *Pseudomonas* [102], and *E. coli*, which persist on contaminated surfaces in concentrations high enough for transmission [102, 103]. It may also be applicable to SARS-CoV-2. The model assumptions are less well-suited for pathogens that persist for shorter times on surfaces, for which person-person transmission or long-term carriage does not occur, and/or in which an extended non-infectious incubation time occurs. The inference method will work better for pathogens whose colonization prevalence falls in the intermediate range between non-existent and complete colonization.

In summary, the full detailed mathematical model and inference method enables estimation of parameters for three possible routes of colonization from active surveillance data. Using these parameters, we can estimate the importance of patient–patient transmission and

effectiveness of interventions on colonization pressure, and we can also simulate realistic scenarios with actual patient exit/entry times. Crucially, the inference method allows direct maximum likelihood estimation of parameters from *all* possible states consistent with incomplete and nonrandom observations without need for MCMC techniques. We also demonstrate the utility of tracking the colonization statuses of individual patients for unique parameter identification. Incorporation of additional routes of colonization and transmission, including exogenous colonization from sources other than the patients (such as visitors), is an area for future research.

Supporting information

S1 Appendix. Mathematical methods.

(ZIP)

S2 Appendix. Data description.

(ZIP)

S1 Table. Comparison of mathematical models.

(ZIP)

S1 Data. Data used for analysis.

(XLS)

Acknowledgments

Thanks to David Sontag for the important suggestion that the detailed state model may be preferable to the reduced model for parameter identification, even though the two models are equivalent when used for simulation with known parameters; the former NYULMC Center for Health Informatics and Bioinformatics for use of the Phoenix (now BigPurple) computing cluster; New York University's high performance computing facilities; Eric Peskin for advice on high-performance computing and searching for optimal parameters; and Loren Koenig for advice on high-performance computing, diagrams and text, and the transmission rate.

Author Contributions

Conceptualization: Karen M. Ong, Michael S. Phillips, Charles S. Peskin.

Data curation: Karen M. Ong.

Formal analysis: Karen M. Ong, Charles S. Peskin.

Investigation: Karen M. Ong, Michael S. Phillips, Charles S. Peskin.

Methodology: Karen M. Ong, Michael S. Phillips, Charles S. Peskin.

Project administration: Michael S. Phillips, Charles S. Peskin.

Resources: Karen M. Ong, Michael S. Phillips, Charles S. Peskin.

Software: Karen M. Ong, Charles S. Peskin.

Supervision: Michael S. Phillips, Charles S. Peskin.

Validation: Karen M. Ong, Charles S. Peskin.

Visualization: Karen M. Ong, Charles S. Peskin.

Writing – original draft: Karen M. Ong, Charles S. Peskin.

Writing – review & editing: Karen M. Ong, Michael S. Phillips, Charles S. Peskin.

References

1. Cerqueira G. C., Earl A. M., Ernst C. M., Grad Y. H., Dekker J. P., Feldgarden M., et al. (2017). Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. *Proceedings of the National Academy of Sciences*, (pp. 201616248). <https://doi.org/10.1073/pnas.1616248114> PMID: 28096418
2. Logan L. K. & Weinstein R. A. (2017). The Epidemiology of Carbapenem-Resistant Enterobacteriaceae: The Impact and Evolution of a Global Menace. *The Journal of Infectious Diseases*, 215 (suppl_1), S28–S36. <https://doi.org/10.1093/infdis/jiw282> PMID: 28375512
3. Maltezou H. C., Giakkoupi P., Maragos A., Bolikas M., Raftopoulos V., Papahatzaki H., Vrouhos G., Liakou V., et al. (2009). Outbreak of infections due to KPC-2-producing *Klebsiella pneumoniae* in a hospital in Crete (Greece). *Journal of Infection*, 58(3), 213–219. <https://doi.org/10.1016/j.jinf.2009.01.010> PMID: 19246099
4. Pournaras S., Protonotariou E., Voulgari E., Kristo I., Dimitroulia E., Vitti D., et al. (2009). Clonal spread of KPC-2 carbapenemase-producing *Klebsiella pneumoniae* strains in Greece. *Journal of Antimicrobial Chemotherapy*, 64(2), 348–352. <https://doi.org/10.1093/jac/dkp207>
5. Schwaber MJ & Carmeli Y (2008). Carbapenem-resistant enterobacteriaceae: A potential threat. *JAMA*, 300(24), 2911–2913. <https://doi.org/10.1001/jama.2008.896> PMID: 19109119
6. Schwaber M. J., Lev B., Israeli A., Solter E., Smollan G., Rubinovitch B., et al. (2011). Containment of a Country-wide Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* in Israeli Hospitals via a Nationally Implemented Intervention. *Clinical Infectious Diseases*, 52 (7), 848–855. <https://doi.org/10.1093/cid/cir025> PMID: 21317398
7. Woodford N., Tierno P. M., Young K., Tysall L., Paleou M.-F. I., Ward E., et al. (2004). Outbreak of *Klebsiella pneumoniae* Producing a New Carbapenem-Hydrolyzing Class A β -Lactamase, KPC-3, in a New York Medical Center. *Antimicrobial Agents and Chemotherapy*, 48(12), 4793–4799. <https://doi.org/10.1128/AAC.48.12.4793-4799.2004> PMID: 15561858
8. Gillespie, S. H. (2014). *Medical Microbiology Illustrated*. Butterworth-Heinemann.
9. Tischendorf J., de Avila R. A., & Safdar N. (2016). Risk of infection following colonization with carbapenem-resistant Enterobacteriaceae: A systematic review. *American Journal of Infection Control*, 44(5), 539–543. <https://doi.org/10.1016/j.ajic.2015.12.005> PMID: 26899297
10. Kelly A. M., Mathema B., & Larson E. L. (2017). Carbapenem-resistant Enterobacteriaceae in the community: a scoping review. *International Journal of Antimicrobial Agents*, 50(2), 127–134. <https://doi.org/10.1016/j.ijantimicag.2017.03.012> PMID: 28647532
11. van Duin D., Kaye K. S., Neuner E. A., & Bonomo R. A. (2013). Carbapenem-resistant Enterobacteriaceae: a review of treatment and outcomes. *Diagnostic Microbiology and Infectious Disease*, 75(2), 115–120. <https://doi.org/10.1016/j.diagmicrobio.2012.11.009> PMID: 23290507
12. Thaden J. T., Pogue J. M., & Kaye K. S. (2017). Role of newer and re-emerging older agents in the treatment of infections caused by carbapenem-resistant Enterobacteriaceae. *Virulence*, 8(4), 403–416. <https://doi.org/10.1080/21505594.2016.1207834> PMID: 27384881
13. Wunderink R. G., Giamarellos-Bourboulis E. J., Rahav G., Mathers A. J., Bassetti M., Vazquez J., et al. (2018). Effect and Safety of Meropenem-Vaborbactam versus Best-Available Therapy in Patients with Carbapenem-Resistant Enterobacteriaceae Infections: The TANGO II Randomized Clinical Trial. *Infectious Diseases and Therapy*, 7(4), 439–455. <https://doi.org/10.1007/s40121-018-0214-1> PMID: 30270406
14. Bartsch S. M., McKinnell J. A., Mueller L. E., Miller L. G., Gohil S. K., Huang S. S., et al. (2017). Potential economic burden of carbapenem-resistant Enterobacteriaceae (CRE) in the United States. *Clinical Microbiology and Infection*, 23 (1), 48.e9–48.e16. <https://doi.org/10.1016/j.cmi.2016.09.003>
15. Tacconelli E., Cataldo M. A., Dancer S. J., De Angelis G., Falcone M., Frank U., et al. (2014). ESCMID guidelines for the management of the infection control measures to reduce transmission of multidrug-resistant Gram-negative bacteria in hospitalized patients. *Clinical Microbiology and Infection*, 20, 1–55. <https://doi.org/10.1111/1469-0691.12427>
16. Temkin E., Adler A., Lerner A., & Carmeli Y. (2014). Carbapenem-resistant Enterobacteriaceae: biology, epidemiology, and management. *Annals of the New York Academy of Sciences*, 1323(1), 22–42. <https://doi.org/10.1111/nyas.12537> PMID: 25195939
17. Barbier F., Pommier C., Essaied W., Garrouste-Orgeas M., Schwebel C., Ruckly S., Dumenil A.-S., et al. (2016). Colonization and infection with extended-spectrum β -lactamase-producing Enterobacteriaceae in ICU patients: what impact on outcomes and carbapenem exposure? *Journal of Antimicrobial Chemotherapy*, (pp. dkv423). PMID: 26755492

18. Wang Q., Zhang Y., Yao X., Xian H., Liu Y., Li H., et al. (2016). Risk factors and clinical outcomes for carbapenem-resistant Enterobacteriaceae nosocomial infections. *European Journal of Clinical Microbiology & Infectious Diseases*, 35(10), 1679–1689. <https://doi.org/10.1007/s10096-016-2710-0> PMID: 27401905
19. Centers for Disease Control (2015). *Facility Guidance for Control of Carbapenem-resistant Enterobacteriaceae (CRE) “November 2015 Update CRE Toolkit*. Technical report, Centers for Disease Control.
20. Abbott I. J., Jenney A. W. J., Spelman D. W., Pilcher D. V., Sidjabat H. E., Richardson L. J., et al. (2015). Active surveillance for multidrug-resistant Gram-negative bacteria in the intensive care unit. *Pathology*, 47(6), 575–579. <https://doi.org/10.1097/PAT.0000000000000302> PMID: 26308128
21. Debby B. D., Ganor O., Yasmin M., David L., Nathan K., Ilana T., Dalit S., et al. (2012). Epidemiology of carbapenem resistant *Klebsiella pneumoniae* colonization in an intensive care unit. *European Journal of Clinical Microbiology & Infectious Diseases*, 31(8), 1811–1817. <https://doi.org/10.1007/s10096-011-1506-5>
22. Ben-David D., Maor Y., Keller N., Regev-Yochay G., Tal I., Shachar D., et al. (2010). Potential Role of Active Surveillance in the Control of a Hospital-Wide Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* Infection. *Infection Control and Hospital Epidemiology*, 31(6), 620–626. <https://doi.org/10.1086/652528> PMID: 20370465
23. Karchmer T. B., Durbin L. J., Simonton B. M., & Farr B. M. (2002). Cost-effectiveness of active surveillance cultures and contact/droplet precautions for control of methicillin-resistant *Staphylococcus aureus*. *Journal of Hospital Infection*, 51(2), 126–132. <https://doi.org/10.1053/jhin.2002.1200> PMID: 12090800
24. Muto C. A., Giannetta E. T., Durbin L. J., Simonton B. M., & Farr B. M. (2002). Cost-Effectiveness of Perirectal Surveillance Cultures for Controlling Vancomycin-Resistant *Enterococcus*. *Infection Control & Hospital Epidemiology*, 23(08), 429–435. <https://doi.org/10.1086/502080> PMID: 12186207
25. Swaminathan M., Sharma S., Blash S. P., Patel G., Banach D. B., Phillips M., et al. (2013). Prevalence and Risk Factors for Acquisition of Carbapenem-Resistant Enterobacteriaceae in the Setting of Endemicity. *Infection Control & Hospital Epidemiology*, 34(08), 809–817. <https://doi.org/10.1086/671270> PMID: 23838221
26. Parker, V. A., Logan, C. K., & Currie, B. (2014). *Carbapenem-Resistant Enterobacteriaceae (CRE) Control and Prevention Toolkit*. Technical Report AHRQ Publication No. 14-0028-E F, Agency for Healthcare Research and Quality, Rockville, MD. (Prepared by Boston University School of Public Health and Montefiore Medical Center under Contract No. 290-2006-0012-I)..
27. Becker N. G. & Britton T. (1999). Statistical studies of infectious disease incidence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61 (2), 287–307. <https://doi.org/10.1111/1467-9868.00177>
28. Cooper B. S. (2007). Confronting models with data. *Journal of Hospital Infection*, 65, Supplement 2, 88–92. [https://doi.org/10.1016/S0195-6701\(07\)60022-X](https://doi.org/10.1016/S0195-6701(07)60022-X) PMID: 17540249
29. Kermack W. O. & McKendrick A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772), 700–721.
30. Cooper B. S., Medley G. F., & Scott G. M. (1999). Preliminary analysis of the transmission dynamics of nosocomial infections: stochastic and management effects. *Journal of Hospital Infection*, 43(2), 131–147. <https://doi.org/10.1053/jhin.1998.0647> PMID: 10549313
31. van Kleef E., Robotham J. V., Jit M., Deeny S. R., & Edmunds W. J. (2013). Modelling the transmission of healthcare associated infections: a systematic review. *BMC Infectious Diseases*, 13, 294. <https://doi.org/10.1186/1471-2334-13-294> PMID: 23809195
32. Bootsma M. C. J., Bonten M. J. M., Nijssen S., Fluit A. C., & Diekmann O. (2007). An Algorithm to Estimate the Importance of Bacterial Acquisition Routes in Hospital Settings. *American Journal of Epidemiology*, 166(7), 841–851. <https://doi.org/10.1093/aje/kwm149> PMID: 17644823
33. Auranen K., Arjas E., Leino T., & Takala A. K. (2000). Transmission of Pneumococcal Carriage in Families: A Latent Markov Process Model for Binary Longitudinal Data. *Journal of the American Statistical Association*, 95(452), 1044–1053. <https://doi.org/10.1080/01621459.2000.10474301>
34. Cooper B. S., Medley G. F., Bradley S. J., & Scott G. M. (2008). An Augmented Data Method for the Analysis of Nosocomial Infection Data. *American Journal of Epidemiology*, 168(5), 548–557. <https://doi.org/10.1093/aje/kwn176> PMID: 18635575
35. Kypraios T., O'Neill P. D., Huang S. S., Rifas-Shiman S. L., & Cooper B. S. (2010). Assessing the role of undetected colonization and isolation precautions in reducing Methicillin-Resistant *Staphylococcus aureus* transmission in intensive care units. *BMC Infectious Diseases*, 10(1), 29. <https://doi.org/10.1186/1471-2334-10-29> PMID: 20158891

36. López-García M. & Kypraios T. (2018). A unified stochastic modelling framework for the spread of nosocomial infections. *Journal of The Royal Society Interface*, 15(143), 20180060. <https://doi.org/10.1098/rsif.2018.0060> PMID: 29899157
37. McBryde E., Pettitt A., Cooper B., & McElwain D. (2007). Characterizing an outbreak of vancomycin-resistant enterococci using hidden Markov models. *Journal of the Royal Society Interface*, 4(15), 745–754. <https://doi.org/10.1098/rsif.2007.0224> PMID: 17360254
38. Pelupessy I., Bonten M. J. M., & Diekmann O. (2002). How to assess the relative importance of different colonization routes of pathogens within hospital settings. *Proceedings of the National Academy of Sciences*, 99(8), 5601–5605. <https://doi.org/10.1073/pnas.082412899> PMID: 11943870
39. Wei Y., Kypraios T., O'Neill P. D., Huang S. S., Rifas-Shiman S. L., & Cooper B. S. (2016). Evaluating hospital infection control measures for antimicrobial-resistant pathogens using stochastic transmission models: Application to vancomycin-resistant enterococci in intensive care units. *Statistical Methods in Medical Research*.
40. Worby C. J., Jeyaratnam D., Robotham J. V., Kypraios T., O'Neill P. D., De Angelis D., et al. (2013). Estimating the Effectiveness of Isolation and Decolonization Measures in Reducing Transmission of Methicillin-resistant *Staphylococcus aureus* in Hospital General Wards. *American Journal of Epidemiology*, 177(11), 1306–1313. <https://doi.org/10.1093/aje/kws380> PMID: 23592544
41. Cooper B. & Lipsitch M. (2004). The analysis of hospital infection data using hidden Markov models. *Biostatistics*, 5(2), 223–237. <https://doi.org/10.1093/biostatistics/5.2.223> PMID: 15054027
42. Drovandi C. C. & Pettitt A. N. (2008). Multivariate Markov Process Models for the Transmission of Methicillin-Resistant *Staphylococcus Aureus* in a Hospital Ward. *Biometrics*, 64(3), 851–859. <https://doi.org/10.1111/j.1541-0420.2007.00933.x> PMID: 18047536
43. Austin D. J., Bonten M. J. M., Weinstein R. A., Slaughter S., & Anderson R. M. (1999). Vancomycin-resistant enterococci in intensive-care hospital settings: Transmission dynamics, persistence, and the impact of infection control programs. *Proceedings of the National Academy of Sciences*, 96(12), 6908–6913. <https://doi.org/10.1073/pnas.96.12.6908> PMID: 10359812
44. Austin D. J., Anderson R. M. (1999). Studies of antibiotic resistance within the patient, hospitals and the community using simple mathematical models. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 354(1384), 721–738. <https://doi.org/10.1098/rstb.1999.0425> PMID: 10365398
45. Vanhems P., Barrat A., Cattuto C., Pinton J.-F., Khanafer N., Régis C., et al. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS one*, 8(9), e73970. <https://doi.org/10.1371/journal.pone.0073970> PMID: 24040129
46. Starr J. M. & Campbell A., Renshaw E., Poxton I. R., & Gibson G. J. (2012). Spatio-temporal stochastic modelling of *Clostridium difficile*. *Journal of Hospital Infection*, 71(1), 49–56.
47. Forrester M. L., Pettitt A. N., & Gibson G. J. (2007). Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data. *Biostatistics*, 8(2), 383–401. <https://doi.org/10.1093/biostatistics/kxl017> PMID: 16926230
48. Wolkewitz M., Dettenkofer M., Bertz H., Schumacher M., & Huebner J. (2008). Statistical epidemic modeling with hospital outbreak data. *Statistics in Medicine*, 27(30), 6522–6531. <https://doi.org/10.1002/sim.3419> PMID: 18759371
49. O'Neill P. D. & Roberts G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1), 121–129. <https://doi.org/10.1111/1467-985X.00125>
50. Ulrich, R., Zimring, C., Quan, X., Joseph, A., & Choudhary, R. (2004). *Role of the Physical Environment in the Hospital of the 21st Century*. Technical report, Center for Health Design.
51. O'Fallon E., Kandell R., Schreiber R., & D'Agata E. (2010). Acquisition of Multidrug-Resistant Gram-Negative Bacteria: Incidence and Risk Factors within a Long-Term Care Population. *Infection Control and Hospital Epidemiology*, 31(11), 1148–1153. <https://doi.org/10.1086/656590> PMID: 20923286
52. Schechner V., Kotlovsky T., Tarabeia J., Kazma M., Schwartz D., Navon-Venezia S., et al. (2011). Predictors of Rectal Carriage of Carbapenem-Resistant Enterobacteriaceae (CRE) among Patients with Known CRE Carriage at Their Next Hospital Encounter. *Infection Control & Hospital Epidemiology*, 32 (05), 497–503. <https://doi.org/10.1086/659762> PMID: 21515981
53. Bertolazzi, E. (2009). *Matrix exponential: Integration lectures for the Course: Numerical Methods for Dynamical System and Control*. Technical report, Tech. report, UNITN.
54. Dobrow, R. P. (2016a). Continuous-Time Markov Chains. In *Introduction to Stochastic Processes With R* (pp. 265–319). John Wiley & Sons, Inc.
55. Dobrow, R. P. (2016b). Markov Chains: First Steps. In *Introduction to Stochastic Processes With R* (pp. 40–75). John Wiley & Sons, Inc.

56. Pittet D, Dharan S, Touvneau S, Sauvan V, & Perneger TV (1999). Bacterial contamination of the hands of hospital staff during routine patient care. *Archives of Internal Medicine*, 159(8), 821–826. <https://doi.org/10.1001/archinte.159.8.821> PMID: 10219927
57. Gillespie D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4), 403–434. [https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3)
58. Gillespie D. T. (1977). Concerning the validity of the stochastic approach to chemical kinetics. *Journal of Statistical Physics*, 16(3), 311–318. <https://doi.org/10.1007/BF01020385>
59. Sasieni P. D. & Brentnall A. R. (2014). Survival Analysis. In W. Ahrens & I. Pigeot (Eds.), *Handbook of Epidemiology* (pp. 1195–1239). New York, NY: Springer.
60. Kolmogoroff A. (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104(1), 415–458. <https://doi.org/10.1007/BF01457949>
61. Shiryaev A. N. (2012). Kolmogorov equation. In *Encyclopedia of Mathematics*. Kluwer Academic Publishers.
62. Gillespie D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25), 2340–2361. <https://doi.org/10.1021/j100540a008>
63. Gillespie D. T. (1978). Monte Carlo simulation of random walks with residence time dependent transition probability rates. *Journal of Computational Physics*, 28(3), 395–407. [https://doi.org/10.1016/0021-9991\(78\)90060-8](https://doi.org/10.1016/0021-9991(78)90060-8)
64. Sevast'yanov BA (2002). Exponential distribution. In *Encyclopedia of Mathematics*. Kluwer Academic Publishers.
65. Bar-Yoseph H., Hussein K., Braun E., & Paul M. (2016). Natural history and decolonization strategies for ESBL/carbapenem-resistant Enterobacteriaceae carriage: systematic review and meta-analysis. *The Journal of Antimicrobial Chemotherapy*, 71(10), 2729–2739. <https://doi.org/10.1093/jac/dkw221> PMID: 27317444
66. Rieg S., Küpper M. F., de With K., Serr A., Bohnert J. A., & Kern W. V. (2015). Intestinal decolonization of Enterobacteriaceae producing extended-spectrum beta-lactamases (ESBL): a retrospective observational study in patients at risk for infection and a brief review of the literature. *BMC Infectious Diseases*, 15. <https://doi.org/10.1186/s12879-015-1225-0> PMID: 26511929
67. Lübbert C., Becker-Rux D., Rodloff A., Laudi S., Busch T., Bartels M., et al. (2014). Colonization of liver transplant recipients with KPC-producing *Klebsiella pneumoniae* is associated with high infection rates and excess mortality: A case-control analysis. *Infection*, 42(2), 309–316. <https://doi.org/10.1007/s15010-013-0547-3> PMID: 24217959
68. Hilty M., Betsch B. Y., Bogli-Stubler K., Heiniger N., Stadler M., Kuffer M., et al. (2012). Transmission Dynamics of Extended-Spectrum β -lactamase-Producing Enterobacteriaceae in the Tertiary Care Hospital and the Household Setting. *Clinical Infectious Diseases*, 55(7), 967–975. <https://doi.org/10.1093/cid/cis581> PMID: 22718774
69. Moler C. & Van Loan C. (1978). Nineteen Dubious Ways to Compute the Exponential of a Matrix. *SIAM Review*, 20(4), 801–836. <https://doi.org/10.1137/1020098>
70. Moler C. & Van Loan C. (2003). Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*, 45(1), 3–49. <https://doi.org/10.1137/S00361445024180>
71. Economou A., Gómez-Corral A., & López-García M. (2015). A stochastic SIS epidemic model with heterogeneous contacts. *Physica A: Statistical Mechanics and its Applications*, 421, 78–97. <https://doi.org/10.1016/j.physa.2014.10.054>
72. Dekking F. M. (2005). The Poisson process. In Kraaikamp C., Lopuhaä H.P., & Meester L. E. (Eds.), *A Modern Introduction to Probability and Statistics: Understanding Why and How* (pp. 167–179). London: Springer London.
73. Allen L. J. S. (2011). *An Introduction to Stochastic Processes with Applications to Biology*, Second Edition. Hoboken: Taylor and Francis, 2nd ed. edition.
74. Okamoto K., Lin M. Y., Haverkate M., Lolans K., Moore N. M., Weiner S., et al. (2017). Modifiable Risk Factors for the Spread of *Klebsiella pneumoniae* Carbapenemase-Producing Enterobacteriaceae Among Long-Term Acute-Care Hospital Patients. *Infection Control & Hospital Epidemiology*, 38(6), 670–677. <https://doi.org/10.1017/ice.2017.62> PMID: 28397615
75. Gibson G. J. & Renshaw E. (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *Mathematical Medicine and Biology*, 15(1), 19–40. <https://doi.org/10.1093/imamb/15.1.19>
76. Simon P. L., Taylor M., & Kiss I. Z. (2011). Exact epidemic models on graphs using graph-automorphism driven lumping. *Journal of Mathematical Biology*, 62(4), 479–508. <https://doi.org/10.1007/s00285-010-0344-x> PMID: 20425114

77. Brauer F. (2005). The Kermack–McKendrick epidemic model revisited. *Mathematical Biosciences*, 198(2), 119–131. <https://doi.org/10.1016/j.mbs.2005.07.006> PMID: 16135371
78. Brauer F., van den Driessche P., Wu J., Morel J. M., Takens F., & Teissier B., Eds. (2008). *Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg.
79. Huang S. S., Datta R., & Platt R. (2006). Risk of acquiring antibiotic-resistant bacteria from prior room occupants. *Archives of internal medicine*, 166(18), 1945–51. <https://doi.org/10.1001/archinte.166.18.1945> PMID: 17030826
80. Mills J. P., Talati N. J., Alby K., & Han J. H. (2016). The Epidemiology of Carbapenem-Resistant *Klebsiella pneumoniae* Colonization and Infection among Long-Term Acute Care Hospital Residents. *Infection Control & Hospital Epidemiology*, 37(1), 55–60. <https://doi.org/10.1017/ice.2015.254> PMID: 26455382
81. Borer A., Saidel-Odes L., Eskira S., Nativ R., Riesenber K., Livshiz-Riven I., et al. (2012). Risk factors for developing clinical infection with carbapenem-resistant *Klebsiella pneumoniae* in hospital patients initially only colonized with carbapenem-resistant *K pneumoniae*. *American Journal of Infection Control*, 40(5), 421–425. <https://doi.org/10.1016/j.ajic.2011.05.022> PMID: 21906844
82. van Loon K., Voor in't Holt A. F. V., & Vos M. C. (2018). A Systematic Review and Meta-analyses of the Clinical Epidemiology of Carbapenem-Resistant Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy*, 62(1), e01730–17. <https://doi.org/10.1128/AAC.01730-17> PMID: 29038269
83. Wiener-Well Y., Rudensky B., Yinnon A. M., Kopuit P., Schlesinger Y., et al. (2010). Carriage rate of carbapenem-resistant *Klebsiella pneumoniae* in hospitalised patients during a national outbreak. *Journal of Hospital Infection*, 74(4), 344–349. <https://doi.org/10.1016/j.jhin.2009.07.022> PMID: 19783067
84. Harris A. D., Perencevich E. N., Johnson J. K., Paterson D. L., Morris J. G., Strauss S. M., et al. (2007). Patient-to-Patient Transmission Is Important in Extended-Spectrum β -Lactamase–Producing *Klebsiella pneumoniae* Acquisition. *Clinical Infectious Diseases*, 45(10), 1347–1350. <https://doi.org/10.1086/522657> PMID: 17968833
85. Mathers A. J., Peirano G., & Pitout J. D. D. (2015). The Role of Epidemic Resistance Plasmids and International High-Risk Clones in the Spread of Multidrug-Resistant Enterobacteriaceae. *Clinical Microbiology Reviews*, 28(3), 565–591. <https://doi.org/10.1128/CMR.00116-14> PMID: 25926236
86. Mathers A. J., Stoesser N., Chai W., Carroll J., Barry K., Cherunvanky A., et al. (2017). Chromosomal Integration of the *Klebsiella pneumoniae* Carbapenemase Gene, blaKPC, in *Klebsiella* Species Is Elusive but Not Rare. *Antimicrobial Agents and Chemotherapy*, 61(3), e01823–16. <https://doi.org/10.1128/AAC.01823-16> PMID: 28031204
87. Pecora N. D., Li N., Allard M., Li C., Albano E., Delaney M., Dubois A., et al. (2015). Genomically Informed Surveillance for Carbapenem-Resistant Enterobacteriaceae in a Health Care System. *mBio*, 6(4), e01030–15. <https://doi.org/10.1128/mBio.01030-15> PMID: 26220969
88. Salabi A. E., Walsh T. R., & Chouchani C. (2013). Extended spectrum beta-lactamases, carbapenemases and mobile genetic elements responsible for antibiotics resistance in Gram-negative bacteria. *Critical Reviews in Microbiology*, 39(2), 113–122. <https://doi.org/10.3109/1040841X.2012.691870> PMID: 22667455
89. Gomez-Simmonds A., Hu Y., Sullivan S. B., Wang Z., Whittier S., & Uhlemann A.-C. (2016). Evidence from a New York City hospital of rising incidence of genetically diverse carbapenem-resistant Enterobacter cloacae and dominance of ST171, 2007–14. *Journal of Antimicrobial Chemotherapy*, (pp. dkw132). <https://doi.org/10.1093/jac/dkw132> PMID: 27118776
90. Brodrick H. J., Raven K. E., Harrison E. M., Blane B., Reuter S., Török M. E., et al. (2016). Whole-genome sequencing reveals transmission of vancomycin-resistant *Enterococcus faecium* in a health-care network. *Genome Medicine*, 8(1), 4. <https://doi.org/10.1186/s13073-015-0259-7> PMID: 26759031
91. World Health Organization (2009). WHO Guidelines on Hand Hygiene in Health Care.
92. Datta R., Platt R., Yokoe D. S., & Huang S. S. (2011). Environmental cleaning intervention and risk of acquiring multidrug-resistant organisms from prior room occupants. *Archives of internal medicine*, 171(6), 491–4. <https://doi.org/10.1001/archinternmed.2011.64> PMID: 21444840
93. Palmore T. N. & Henderson D. K. (2013). Managing transmission of carbapenem-resistant enterobacteriaceae in healthcare settings: a view from the trenches. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 57(11), 1593–9. <https://doi.org/10.1093/cid/cit531> PMID: 23934166
94. Snitkin E. S., Zelazny A. M., Thomas P. J., Stock F., Program N. C. S., Henderson D. K., et al. (2012). Tracking a Hospital Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* with Whole-Genome Sequencing. *Science Translational Medicine*, 4(148), 148ra116–148ra116. <https://doi.org/10.1126/scitranslmed.3004129> PMID: 22914622

95. Lerner A., Adler A., Abu-Hanna J., Cohen Percia S., Kazma Matalon M., & Carmeli Y. (2015). Spread of KPC-producing carbapenem-resistant Enterobacteriaceae: the importance of super-spreaders and rectal KPC concentration. *Clinical Microbiology and Infection*, 21(5), 470.e1–470.e7. <https://doi.org/10.1016/j.cmi.2014.12.015> PMID: 25684452
96. Zhou K., Lokate M., Deurenberg R. H., Tepper M., Arends J. P., Raangs E. G. C., et al. (2016). Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β -lactamase producing ST15 *Klebsiella pneumoniae*. *Scientific Reports*, 6. <https://doi.org/10.1038/srep20840> PMID: 26864946
97. Bowers J. R., Lemmer D., Sahl J. W., Pearson T., Driebe E. M., Engelthaler D. M., et al. (2016). Kleb-Seq: A Diagnostic Tool for Healthcare Surveillance and Antimicrobial Resistance Monitoring of *Klebsiella pneumoniae*. *bioRxiv*, (pp. 043471).
98. Centers for Medicare & Medicaid Services (2008). Medicare program: changes to the hospital inpatient prospective payment systems and fiscal year 2009 rates; payments for graduate medical education in certain emergency situations; changes to disclosure of physician ownership in hospitals and physical self-referral rules; updates to the long-term care prospective payment system; updates to certain IPPS-excluded hospitals; and collection of information regarding financial relationships between hospitals; final rule. *Federal Register*, 73(161), 48434–49083.
99. Graves N. & McGowan J. E. (2008). Nosocomial Infection, the Deficit Reduction Act, and Incentives for Hospitals. *JAMA*, 300(13), 1577–1579. <https://doi.org/10.1001/jama.300.13.1577>
100. Otto M. (2012). MRSA virulence and spread. *Cellular Microbiology*, 14(10), 1513–1521. <https://doi.org/10.1111/j.1462-5822.2012.01832.x> PMID: 22747834
101. Cohen P. R. (2005). Cutaneous community-acquired methicillin-resistant *Staphylococcus aureus* infection in participants of athletic activities. *Southern Medical Journal*, 98(6), 596–. <https://doi.org/10.1097/01.SMJ.0000163302.72469.28> PMID: 16004165
102. Otter J. A., Yezli S., & French G. L. (2011). The Role Played by Contaminated Surfaces in the Transmission of Nosocomial Pathogens. *Infection Control & Hospital Epidemiology*, 32(7), 687–699. <https://doi.org/10.1086/660363> PMID: 21666400
103. Kramer A., Schwebke I., & Kampf G. (2006). How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infectious Diseases*, 6(1), 130. <https://doi.org/10.1186/1471-2334-6-130> PMID: 16914034