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

Thi Mui Pham [1*], Mirjam Kretzschmar [1,4], Xavier Bertrand [3], Martin Bootsma [1,2], on behalf of COMBACTE-MAGNET Consortium

1 Julius Center for Health Sciences and Primary Care of the UMC Utrecht, Utrecht, The Netherlands
2 Mathematical Institute, Utrecht University, Utrecht, The Netherlands
3 Centre Hospitalier Universitaire Besançon, Besançon, France
4 National Institute for Public Health and the Environment (RIVM), Utrecht, The Netherlands

* t.m.pham-2@umcutrecht.nl

Abstract

Pseudomonas aeruginosa (*P. aeruginosa*) is an important cause of healthcare-associated infections, particularly in immunocompromised patients. Understanding how this multi-drug resistant pathogen is transmitted within intensive care units (ICUs) is crucial for devising and evaluating successful control strategies. While it is known that moist environments serve as natural reservoirs for *P. aeruginosa*, there is little quantitative evidence regarding the contribution of environmental contamination to its transmission within ICUs. Previous studies on other nosocomial pathogens rely on deploying specific values for environmental parameters derived from costly and laborious genotyping. Using solely longitudinal surveillance data, we estimated the relative importance of *P. aeruginosa* transmission routes by exploiting the fact that different routes cause different pattern of fluctuations in the prevalence.
We developed a mathematical model including endogenous colonization, cross-transmission and environmental contamination. Patients contribute to a pool of pathogens by shedding bacteria to the environment. Natural decay and cleaning of the environment lead to a reduction of that pool. By assigning the bacterial load shed during an ICU stay to cross-transmission, we were able to disentangle environmental contamination during and after a patient's stay. Based on a data-augmented Markov Chain Monte Carlo method the relative importance of the considered acquisition routes is determined for two ICUs of the University hospital in Besançon (France). We used information about the admission and discharge days, screening days and screening results of the ICU patients.
Both cross-transmission and endogenous transmission play a significant role in the transmission process in both ICUs. In contrast, only about 1% of the total transmissions were due to environmental contamination after discharge.
Improved cleaning of the environment after discharge would have only a limited impact regarding the prevention of *P.A.* infections in the two considered ICUs. Our model was developed for *P. aeruginosa* but can be easily applied to other pathogens as well.

Author summary

Understanding the transmission dynamics of multi-drug resistant pathogens in intensive-care units is essential for designing successful infection control strategies. We developed a method that estimates the relative importance of several transmission routes of *Pseudomonas aeruginosa* (*P. aeruginosa*), a bacterium intrinsically resistant to multiple antibiotics and known to be a major contributor to hospital-acquired infections. Our model includes three different routes: endogenous colonization, cross-transmission and environmental contamination. Since moist environments may serve as natural reservoirs for *P. aeruginosa*, we focused our study on environmental contamination. Patients contribute to a pool of pathogens by shedding bacteria to the environment. Natural decay and cleaning of the environment lead to a reduction of that pool. By assigning the bacterial load shed during an ICU stay to cross-transmission, we disentangled environmental contamination during and after a patient's stay. Previous studies were able to assess the role of environmental contamination for specific hospitals using laborious and costly genotyping methods. In contrast, we only used surveillance screening data to estimate the relative contributions of the considered transmission routes. Our results can be used to tailor or assess the effect of interventions. Our model was developed for *P. aeruginosa* but can be easily applied to other pathogens as well.

Introduction

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

Due to its intrinsic resistance to multiple antibiotics, *Pseudomonas aeruginosa* (short *P. aeruginosa* or *P. A.*) is an important contributor to nosocomial infections [3–5]. The most serious *P. aeruginosa* infections lead to bacteremia, pneumonia, urosepsis, wound infection as well as secondary infection of burns [6]. In 2018, the World Health Organization has recognized *P. aeruginosa* as a serious health-care threat by including it in the list of antibiotic-resistant highest priority pathogens [7]. Given the severe consequences of *P. aeruginosa* infections, in particular for critically-ill patients, it is clear that strategies preventing infections are seen as a key priority. However, infections are recognized as only the tip of the iceberg, while colonizations represent the true load of pathogens carried by patients in the intensive-care unit (ICU). Understanding the dynamics of *P. aeruginosa* colonizations is therefore crucial for developing and evaluating infection control policies.

There are several modes of transmission for colonizations. The endogenous route, due to e. g. antibiotic selection pressure, was regarded as the most important route of *P. aeruginosa* [8, 9], but, more and more evidence has emerged on the importance of exogenous sources. Cross-transmission usually caused by temporarily contaminated hands of health-care workers (HCWs) has been identified as an additional source of transmission [10–12]. It is furthermore known that moist environments (e. g. soil and water) may serve as natural reservoirs of *P. aeruginosa* and that it can persist for months on dry inanimate surfaces [13]. Thus, an important question emerged: Does the contamination of the environment with *P. aeruginosa* strains have a substantial impact on the transmission within ICUs even after the discharge of patients?

Quantifying the relative importance of routes of transmission may serve as an essential tool in answering this question. There is little quantitative evidence in the scientific literature regarding the importance of environmental contamination within the transmission dynamics of *P. aeruginosa* especially for non-epidemic situations. Prior studies analyzing the importance of contaminated surfaces on the transmission of other

nosocomial pathogens have been conducted, e. g., for Methicillin-resistant Staphylococcus aureus (MRSA) and Vancomycin-Resistant Enterococci (VRE) [14–18]. However, they rely on deploying specific values for model parameters corresponding to the environment. Such information was obtained from previous studies that conducted extensive epidemiological surveillance in combination with costly, laborious as well as time-consuming methods of genotyping. Thus, these methods cannot be easily applied to other nosocomial pathogens without this cumbersome preliminary work.

In this paper, we present a mathematical transmission model that includes, besides endogenous and cross-transmission, environmental contamination as an additional route. Our primary aim is to estimate the relative contributions of the three transmission routes using only longitudinal prevalence data as input. In particular, we are interested in the relative importance of environmental contamination after discharge. We used data from two ICUs of the University hospital in Besançon to estimate the parameters that characterize the transmission routes. The estimation procedure is based on a data-augmented Markov chain Monte Carlo simulation [19]. To our knowledge, this is the first quantitative analysis of the impact of environmental contamination after discharge on *P. aeruginosa* transmissions in ICUs using solely routine surveillance data.

Materials and methods

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

Transmission models

The underlying model for our algorithm is a compartmental SI-model (e.g. [20]). All patients are admitted to an ICU and either belong to the susceptible (*P. aeruginosa* negative) or colonized (*P. aeruginosa* positive) compartment at any given time. The latter includes patients with asymptotic carriage and those with *P. aeruginosa* infection. A susceptible patient may become colonized at a certain transmission rate, which depends on the colonization pressure in the ward at the time.

A stochastic model was developed to reflect the transmission process. The model incorporates three different modes of transmission through which colonization can be acquired. They are distinguished based on the different patterns in the prevalence time series induced by each of them. The endogenous route is independent of other patients and is represented as a constant rate. Consequently, the corresponding prevalence is expected to fluctuate around the mean value and the probability of acquisition for an uncolonized patient is constant during the time period. Sources may be the introduction through visitors or permanently contaminated environments, such as sinks or air-conditioning. Cross-transmission, usually occurring via temporarily contaminated hands of health-care workers, is proportional to the fraction of colonized patients in the wards. The probability of colonization due to cross-transmission is high if the number of colonized patients is high and vice versa. Environmental contamination is modeled on a ward-level represented as a general pool of bacteria linked to objects contaminated by colonized patients. Bacterial load may persist in the environment even after the discharge of patients. This leads to higher probabilities of acquiring colonization after outbreaks, even when the number of colonized patients is low.

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

$$\lambda(t) = \alpha + \beta \frac{I(t)}{N(t)} + \epsilon E(t) \tag{1}$$

where $I(t)$ is the number of colonized patients, $N(t)$ the total number of patients and $E(t)$ is a compartment tracking the bacterial load present in the ward at time t. The parameters α, β and ϵ are transmission parameters linked to the endogenous term, fraction of colonized patients and the environmental bacterial load, respectively. A schematic of the transmission model is presented in Fig 1.

Fig 1. Schematic of the full transmission model. It represents the three different routes, i.e. endogenous route, cross-transmission and environmental contamination.

The described model is subject to the following further assumptions:

- Once colonized, patients remain colonized during the rest of the stay. This assumption is appropriate when the average length of stay of patients does not exceed the duration of colonization, as is the case for *P. aeruginosa*.

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

- Assuming that HCWs are colonized for a short period of time (typically until the next disinfection) in comparison with the length of carriage for patients, we use a quasi-steady state approximation [20]. This means that contact patterns between patients and HCWs are not explicitly modeled and we assume direct patient-to-patient transmission.

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

- All susceptible patients are assumed to be equally susceptible.

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

$$\frac{dE}{dt} = \nu \frac{I(t)}{N(t)} - \mu E(t). \tag{2}$$

The differential equation (2) is solved by assuming $I(t) = I_t$ and $N(t) = N_t$ are known piece-wise constant functions with steps at times $t_0, t_1, \ldots, t_N$. For $t_i \in \{t_0, \ldots, t_N\}$, it holds

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

while

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

for $\lfloor t \rfloor := \max\{x \in \{t_0, \ldots, t_N\} \mid x \leq t\}$ and $t \in \mathbb{R} \setminus \{t_0, t_1, \ldots, t_N\}$ (full details are given in S1 Text). The initial amount of bacterial load is denoted by $E_0 := E(t_0)$. Given the number of colonized patients at a certain time t, the bacterial load $E(t)$ is deterministic. The acquisitions are stochastic based on the force of infection in (1). Under the assumption of a force of infection $\lambda(x)$ at time x, the cumulative probability of any given susceptible person of becoming colonized in $[0, t]$ is $1 - e^{-\int_0^t \lambda(x)dx}$ (see e.g. [21]).

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

Relative contributions of transmission routes

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

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

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

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

Fig 2. Schematic of the bacterial load shed by a patient developing over time. The bacterial load that is shed during a patient's stay is assigned to cross-transmission. Environmental contamination after discharge accounts only for the bacterial load persisting after the discharge of that patient.

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

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

$$= \underbrace{\alpha}_{\lambda_{\text{endogenous}}} + \underbrace{\left[\beta \cdot \frac{I(t)}{N(t)} + \epsilon \sum_{i_p} \frac{E_{i_p}(t)}{N(t)} \right]}_{\lambda_{\text{cross-transmission}}(t)} + \underbrace{\epsilon \left[\sum_{i_d} \frac{E_{i_d}(t)}{N(t)} + E_0 e^{-\mu t} \right]}_{\lambda_{\text{environment}}(t)} \quad (3)$$

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

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

where t_i^c is the time of colonization and t_i^d the time of discharge of patient i. In continuous time, the relative contribution of a specific route to the overall number of acquired colonizations is determined by the ratio of the probability of colonization due to that route and the probability of colonization:

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

$$= \frac{\sum_{i=1}^{l} \frac{\lambda_j(t_i^c)}{\lambda(t_i^c)}}{\text{Number of acquisitions}} \quad (4)$$

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

- Contribution of endogenous route $= \dfrac{\sum_{i=1}^{l} \frac{\alpha}{\lambda(t_i^c)}}{\text{Number of acquisitions}}$

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

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

For the submodel including only endogenous and cross-transmission, the computation of the relative contribution is derived from above by setting $\epsilon = 0$. More details on the calculations can be found in S3 Text. In practice, colonization events are observed only in discrete times. The formulas for the transmission model and the relative contribution are adapted for this discrete time assumption and are elaborated in S2-S3 Texts. Since the calculations for the relative contributions of the transmission routes in the discrete-time scenario require the use of the gamma function and therefore become computationally intensive, we use the continuous-time formulas as approximations. Since values of the force of infection $\lambda(t)$ are typically small (< 0.25), the force of infection itself is a good approximation of the probability of infection as $1 - e^{-\int_0^t \lambda(x)dx} \approx \lambda(t)$ for small values of $\lambda(t)$. Hence, the discrete-time formulas for the relative contributions can be approximated by the continuous-time formulas evaluated at discrete time steps.

Estimation procedure

We assume that a patient is admitted to the ICU at time t_i^a and discharged at time t_i^d. The probability that a patient is admitted already colonized is described by the

importation probability f. The rate at which a susceptible patient transitions to being colonized is given by Eq. (1). The colonization state of an individual patient is determined from screening information. We suppose that for each patient i a set of screening results $X_i = \left(X_i^{(1)}, \ldots, X_i^{(m)} \right)$, taken on days $t_i^{(1)}, \ldots, t_i^{(m)}$ is available. The set of all screening results is denoted by $X = \{X_1, \ldots, X_n\}$ where n is the total number of patients. Since screening tests are typically intermittent and imperfect, we define the test sensitivity ϕ, i.e. probability that a colonized patient has a positive result. The aim is to estimate the model parameters $\alpha, \beta, \epsilon, \mu, \nu$ and E_0 as well as the sensitivity of the screening test ϕ and the importation rate f based on longitudinal data. The relative contributions of the transmission routes can then be estimated following the description in (4). The key idea of the estimation procedure is to fit a stochastic transmission model to the observed data. It is based on certain patterns of fluctuations in the prevalence linked to the different transmission routes (as previously described in section *Transmission models*).

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

- day of admission

- day of discharge

- screening days and results.

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

If transmission dynamics were perfectly observed, it would be straightforward to calculate the likelihood of the data given parameters $\theta = \{\alpha, \beta, \epsilon, \mu, \nu, \phi, f\}$. However, the true colonization time of a patient is typically unobserved which leads to uncertainty about the true prevalence at any given time. Hence, the likelihood is analytically intractable. We adapted the data-augmented Markov-chain Monte Carlo (MCMC) algorithm (based on [19]) to estimate the posterior distributions of the model parameters. At each iteration, imperfectly observed colonization times are imputed and model parameters θ sampled. This method accounts for imperfect and infrequent screening, missing admission and discharge swabs and leads to an estimation of the true (rather than the observed) prevalence on admission. Precise details of the analysis can be found in S5 Text. The algorithm was implemented in C++ and was tested using simulated data. Convergence of the MCMC chains were verified using visual inspection. We used uninformative exponential priors Exp(0.001) for the transmission parameters α, β, ϵ and μ. Parameters for the proposal distribution were tuned in order to ensure rapid convergence. Similar to [26], we estimated the sensitivity ϕ and importation parameter f using uninformative beta prior distributions Beta(1, 1). The initial bacterial load E_0 was approximated by $\frac{\nu}{\mu} \bar{I}$ with $\bar{I}$ being the mean prevalence in the ward.

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

During the estimation process, several assumptions are made.

- Incorporating both sensitivity and specificity parameters in a model may cause identifiability issues. Thus, test specificity was assumed to be 100%, meaning that positive results were assumed to be true positive. Experimental results indicate the specificity of screening tests to be close to 100% [22].

- The initial bacterial load E_0 is assumed to be the environmental contamination at the beginning of the study period. The effect of E_0 diminishes proportionally to $\exp(-\mu)$ per day. It is therefore sufficient to use an approximation rather than including it as a parameter in the estimation process. We use the equilibrium state of (2) as an approximation, i.e.

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

  where $\bar{I}$ represents the mean prevalence in the ward. The environmental contribution to the force of infection at time t is $\epsilon \cdot E(t)$. As the total amount of environmental contamination $E(t)$ is unobserved, it is only possible to estimate the product $\epsilon \cdot E(t)$ and therefore, we assume the shedding parameter ν to be fixed at 0.1. All remaining parameters are to be estimated from the data.

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

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

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

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

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

Data

The data used in the current analysis were collected from two ICUs, denoted by A and B, between 1999 and 2017 at University Hospital of Besançon, eastern France, in the framework of a systematic screening for *P. aeruginosa*. The data sets include admission and discharge dates as well as dates, sites and results of culturing of adult patients. ICU A is a surgical ICU that comprised 15 beds in the time period 1999-2008 and 20 from 2010 till 2017. The ICU was renovated between 2008 and 2009 and the number of beds was increased after completion of the renovation work. ICU B, a non-surgical ICU, had 15 beds from 2000 till 2011 and increased to 20 beds afterwards. Rectal and nose swabs were obtained upon admission (during the first 48 hours) and once a week thereafter. A positive result on one of the swabs was counted as a positive culture. A negative culture

resulted from a negative culture on both swabs taken at the specific day. More than 84% of admitted patients were screened. As HCWs, including physicians, were (with minor exceptions) working only in one of the ICUs during the whole study period, the two ICUs can be treated independently in the analysis.

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

In our main analysis, data for each ICU and each time period (before and after renovation) was treated as distinctive data sets, resulting in four different analyses. No pooling of the results were performed. In a second analysis, the data for the different time periods and different ICUs were combined. The results are compared with the main analysis and are presented in S9-S10 Tables. Each data set was analyzed using

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

- the submodel with only endogenous and exogenous transmission.

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

Model selection

To assess the relative performance of a given model, we used a version of the deviance information criterion (DIC) based on [23]. For an estimated parameter set θ and observed data set x it is computed as the expected deviance plus the effective number of parameters: $\mathrm{DIC} = \overline{D_x(\theta)} + p_D$. A lower value indicates a better fit. The effective number of parameters p_D represents a complexity measure and is calculated by the difference of the posterior mean deviance and the deviance at the posterior mean: $\overline{D_x(\theta)} - D_x(\tilde{\theta})$. In this paper, we use the approximation $p_D = \frac{1}{2}\mathrm{var}(D_x(\theta))$ introduced by [23].

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

Model assessment

We chose to check the adequacy of the models using the following approach. The ability of the model to predict the probability of acquisition based on the predicted force of infection was assessed. The computed numerical values for the force of infection are assigned to a bin representing the segment covering the numerical value. For a given value λ of the force of infection, the theoretical probability of acquisition p_{acq} per susceptible patient is computed by $1 - \exp(-\lambda)$. The predicted fraction of acquisitions f_{acq} is computed by dividing the number of acquisitions N_{acq} by the number of susceptible patients N_{susc}. We compute 95% confidence intervals assuming that the number of acquisitions follows a binomial distribution of $\mathrm{Bin}(N_{\mathrm{susc}}, f_{\mathrm{acq}})$. The described method is performed for 100 MCMC updates. Coverage probabilities are

computed to determine the actual proportion of updates for which the interval contains   295
the theoretical probability of acquisition. We set the nominal confidence level to 0.95. A   296
good fit is given when the actual coverage probability is (more or less) equal to the   297
nominal confidence level. In order to avoid the coverage probability tending to zero   298
when p_{acq} tends to 0 or 1, Jeffreys confidence intervals are used (as recommended   299
in [24]). When $N_{\text{acq}} = 0$ the lower limit is set to 0, and when $N_{\text{acq}} = N_{\text{susc}}$ the upper   300
limit is set to 1.   301

Results   302

Descriptive analysis of data   303

The descriptive statistics of the data sets corresponding to ICU A and B with respect to   304
the number of admissions, lengths of stay and colonization characteristics are shown in   305
Table 1. The time period referred to as *before renovation* (short BEFORE) is defined as   306
$20/04/1999 - 20/05/2008$ (approx. 9 years) for ICU A and $11/01/2000 - 12/01/2011$ 307
(approx. 11 years) for ICU B. The time period referred to as *after renovation* (short   308
AFTER) is defined as $21/05/2008 - 21/03/2017$ (approx. 7.2 years) for ICU A and   309
$13/01/2011 - 13/09/2016$ (approx. 5.8 years) for ICU B. In ICU A, the number of beds   310
decreased during the renovation. Hence, we decided to remove the renovation period   311
from the analysis for ICU A.   312
In total, 10,378 patients (5891 admitted to ICU A and 4487 to ICU B) and 37,638   313
screening results (14,631 in ICU A and 23,007 in ICU B) were included in the analysis.   314
The number of readmissions is higher for ICU B than for ICU A. In our analysis, every   315
admission was treated separately (as a new patient) resulting in 14,403 admissions   316
(6,659 admitted to ICU A and 7,744 to ICU B).   317
The corresponding median length of stay was 8.0 days for both ICUs before and after   318
renovation, respectively. Hence, there is hardly any difference between the ICUs, nor   319
between the two time periods regarding the median length of stay.   320
In both ICUs, the fraction of patients who were positive on admission was slightly   321
higher after renovation. In contrast, the observed fraction of patients who acquired   322
colonization slightly decreased after renovation. There were 1,644 patients (652 in ICU   323
A and 992 in ICU B) observed to be colonized during their stay and 516 patients (173   324
in ICU A and 343 in ICU B) observed to be colonized on admission. The percentage of   325
patients admitted positively on admission and with acquired colonization is higher in   326
ICU B than in ICU A. The total number of patients per ICU and the number of   327
positive cultures are visualized in Figs 3 and 4.   328

Fig 3. Number of occupied beds and positive isolates cultured from patients per swab day for ICU A. The red dotted lines indicate the time points that splits the study period into *before renovation* $(20/04/1999 - 23/01/2008)$ and *after renovation* $(06/11/2010 - 21/03/2017)$. Since the number of beds decreased during the renovation, the period is removed from the analysis.

Fig 4. Number of occupied beds and positive isolates cultured from patients per swab day for ICU B. The red dotted line indicates the time point that splits the study period into *before renovation* $(11/01/2000 - 12/01/2011)$ and *after renovation* $(13/01/2011 - 13/09/2016)$.

Table 1. Descriptive statistics for the *P. aeruginosa* carriage data collected from two ICUs at University Hospital Besançon, France, 1999-2017

	No.		Median (IQR)[†]		%	
	BEFORE[‡]	AFTER[§]	BEFORE	AFTER	BEFORE	AFTER
ICU A						
Study length, days	3200	2340				
Readmissions	367	463			9.3	11.8
Admissions*	3261	3398				
Length of stay, days			8.0 (3.0-19.0)	8.0 (3.0-16.0)		
Patients positive on admission[¶]	69	104			3.5	3.7
Observed P.A. acquisitions[‖][¶]	363	289			11.1	8.5
ICU B						
Study length, days	4020	2079				
Readmissions	891	1968			15.4	17.8
Admissions	4360	3384				
Length of stay, days			8.0 (3.0-18.0)	8.0 (3.0-15.0)		
Patients positive on admission	168	175			4.7	6.3
Observed P.A. acquisitions	569	423			13.1	12.5

*Each ICU stay was counted separately, even in case of a multiple ICU stay within a given hospitalization.

[†] Interquartile range

[‡] $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.

[§] $06/11/2010 - 21/03/2017$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.

[¶] False negative results are not taken into account.

[‖] An acquisition is when a patient test negative on admission and had a postive result before discharge.

Estimated model parameters

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

Submodel: Two transmission routes

Posterior estimates of the model parameters for each ICU and each time period are reported in Table 2. Acceptance probabilities for proposed updates to the augmented data ranged from 3.2% (ICU B after renovation) to 11.1% (ICU A before renovation). Pairwise scatter plots indicated little correlation between parameter values, with the exception of a negative correlation between α and β (see S11). Histogram and trace plots of the posterior estimates are given in S12-S15 Figs and show that the MCMC chains rapidly mix and quickly converge to their stationary distribution. We found our estimates to be robust to the choice of priors for transmission parameters. The probability of being colonized with *P. aeruginosa* on admission and the screening test sensitivity varied between the two ICUs and the time periods. For both ICUs, the median estimates of the importation probability f is higher in the data set after renovation than before, i.e. 4.5% and 6.2% for ICU A and 6.0% and 9.9% for ICU B. The difference between the time periods is only significant for ICU B. We estimated the median of the prevalence of *P.A.* to be 24.4% and 19.9% for ICU A and 22.3% and 24.4% for ICU B before and after renovation, respectively. Median estimates for the

Table 2. Summary statistics of the marginal posterior distributions for model parameters

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

*Highest posterior density interval

[‡] $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.

[§] $06/11/2010 - 21/03/2017$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.

screening test sensitivity were 50.9% and 50.2% for ICU A and 61.8% and 58.6% for ICU B. Since the credibility intervals of the sensitivity estimates do not overlap with respect to the two ICUs, we can conclude that there is a 95% probability that the test sensitivity is higher in ICU B than in ICU A. Our possible explanation is based on the fact that the ICUs differ in their patient population. As a medical ward, ICU B contains patients with longer lengths of stay and more readmissions. Patients who are exposed to an ICU environment for a longer period of time may have a higher probability to get colonized at a detectable level. However, our explanation is only hypothetical and the true reason for the difference is not known.
The relative importance of the two considered transmission routes per ICU and time period is depicted in Fig 5 (a) and (b). For ICU A, the median relative contribution of the endogenous route is 53.6% (95% CrI : $32.8 - 75.9\%$) and 89.3% (95% CrI : $67.9 - 100\%$) leaving 46.4% (95% CrI : $24.1 - 67.2\%$) and 10.7 (95% CrI : $0 - 32.1\%$) of the acquisitions assigned to cross-transmission before and after renovation, respectively. For ICU B, 43.4% (95% CrI : $29.1 - 58.7\%$) and 84.5% (95% CrI : $60.9 - 100\%$) of the acquisitions were due to the endogenous route and cross-transmission accounted for 56.6% (95% CrI : $41.3 - 70.9\%$) and 15.5% (95% CrI : $0 - 39.1\%$) of the acquisitions before and after renovation, respectively.

Fig 5. Relative contributions of endogenous route and cross-transmission.
(a) ICU A before and after renovation, (b) ICU B before and after renovation.

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

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

Full model: Three transmission routes

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

Fig 6. Relative contributions of endogenous route, cross-transmission and environmental contamination after discharge. (a) ICU A before and after renovation, (b) ICU B before and after renovation.

from 7.2% (ICU B after renovation) to 90% (ICU A before renovation). Pairwise scatter plots indicated strong correlations between α and β, β and ϵ and between ϵ and μ (see Fig 24). The correlation coefficient of the latter pair ranged from 0.531 to 0.561. Furthermore, it can be seen in Table 3 that the credibility intervals for the parameters ϵ and μ are large. Nevertheless, histogram and trace plots of the posterior estimates show that the MCMC chains rapidly mixed and quickly converged to their stationary distribution as can be seen in S16-S23 Figs. The rapid convergence could be achieved by tuning the parameters of the proposal distribution for μ. In contrast, a flat prior for the decay rate μ in combination with a small initial standard deviation for its proposal distribution resulted in large acceptance ratios close to 1. The MCMC chain mixed too slowly and therefore hindered the identifiability of the likelihood. This can be explained by the fact that our developed model is overparametrized when colonizations of patients are not or hardly influenced by environmental contamination. Small values of the transmission parameter ϵ as well as high values of the decay rate μ would reflect the aforementioned situation. As a result, the respective likelihood might not be or only weakly identifiable. Our sensitivity analyses and artificial data simulations demonstrated similar pairwise scatter plots and wide credibility intervals for the parameters ϵ and μ in case of a small contribution of environmental contamination to the transmissions (more details can be found in S7 Text). Hence, we can conclude that the role of environmental contamination after discharge within the transmission process of *P. aeruginosa* in the two ICUs A and B is small before as well as after renovation.

Model selection

In total, 14 analyses were performed. For each ICU, three data sets were created - one for each time period and one combining the data sets before and after renovation. Additionally, the ICUs and time periods were combined in one data set. Each of the seven data sets were analyzed using the submodel and the full model. The DIC values

Table 3. Summary statistics of the marginal posterior distributions for model parameters of the full model

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

*Highest posterior density interval
‡ $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.
§ $06/11/2010 - 21/03/2017$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.

for each model analysis can be found in Table 4. The analysis combining both ICUs and time periods shows smaller DIC values, i.e. 136507.8 and 130693, than the sum of the DICs for separate analyses (152428.8 and 150914.2) for both the submodel and full model, respectively. The full model results in a smaller DIC value for the analysis of the combined data set. Hence, based on the DIC, it would be sufficient to analyze the combined data set using the full model including endogenous route, cross-transmission and environmental contamination. Nevertheless, it can be seen in S8 Text that the posterior estimates of the different analyses are similar, especially for the relative contribution of environmental contamination after discharge.

Model assessment

For each bin of the force of infection the coverage probabilities are plotted and can be found in S25-S26 Figs. It can be seen that the coverage probabilities are approximately (sometimes higher, sometimes smaller) equal to the nominal confidence level of 0.95. Thus, both the full model as well as the submodel gave adequate fits to the four data sets. In Fig 7, the predicted fraction of acquisitions are plotted against the binned force of infection for one exemplary MCMC update. The red lines indicate the relationship between the probability of acquisition and force of infection assumed by our models. For this example, it is always contained by the confidence intervals (blue lines).

Table 4. Deviance information criterion for the different models

	Submodel			Full model		
ICU A						
BEFORE[†]	42961.72	$\Big\}\sum^{\S}$	86899.53	35356.31	$\Big\}\sum$	63057.37
AFTER[‡]	43937.81			27701.06		
COMBINED*	88650.69			63779.12		
ICU B						
BEFORE	35356.31	$\Big\}\sum$	63057.37	34670.46	$\Big\}\sum$	62568.88
AFTER	27701.06			27898.42		
COMBINED	63778.12			59290.73		
Sum COMBINED[‖]		152428.8				
ICUs combined[¶]		136507.8				

[‡] $20/04/1999 - 23/01/2008$ for ICU A and $11/01/2000 - 12/01/2011$ ICU B.

[§] $06/11/2010 - 21/03/2017$ for ICU A and $13/01/2011 - 13/09/2016$ for ICU B.

* Combined time periods

[§] Σ indicates that the sum of the respective columns in the previous row is calculated.

[¶] ICUs as well as time period (before, after renovation) are combined in one data set.

[‖] The sum of the DICs for ICU A (before and after renovation combined) and ICU B is computed.

Fig 7. Exemplary model assessment plot for one MCMC update using the submodel applied to ICU A before renovation. The predicted fraction of acquisition is plotted against the theoretical force of infection. The red line indicates the theoretical relation between the force of infection and the probability of acquisition. The blue lines indicate 95% credibility intervals.

Discussion

To our knowledge, our study is the first to use mathematical transmission models to estimate the relative contribution of environmental contamination after discharge for *P. aeruginosa* using only admission, discharge and screening data. The three different routes, endogenous route, cross-transmission and environmental contamination after discharge, are distinguished by the resulting patterns of the prevalence that they induce. We estimated that environmental contamination after discharge accounts for at most 1% of the total *P. aeruginosa* transmissions in the two ICUs of the University hospital in Besançon before and after renovation. In contrast, endogenous as well as cross-transmission are both essential for the transmission dynamics of *P. aeruginosa*. This suggests, that improvement of cleaning of the environment *after discharge* would have only a limited benefit regarding the prevention of *P. aeruginosa* colonization in the two considered ICUs of the University hospital in Besançon.

Previously, studies have been conducted to investigate the role of environmental contamination for colonizations of *P. aeruginosa*. For instance, Panagea et al. performed environmental studies to determine the extent of environmental contamination with an epidemic strain of *P. aeruginosa* [25]. They concluded that the transmissibility of the epidemic strain cannot be explained solely on the basis of improved environmental survival. Our results likewise demonstrate that the decay of *P. aeruginosa* is already rapid enough to limit its survival in the environment.

While our approach is efficient in determining the relative contribution of environmental contamination after discharge requiring merely longitudinal surveillance data, it has

several limitations that may restrict its practical applicability. 457

Our conclusions on the impact of cleaning only applies to the environment after the 458 discharge of patients. Permanently contaminated reservoirs in ICUs, such as sinks, may 459 still serve as sources for colonization. In our model they are assigned to the endogenous 460 route. Thus, while the effect of cleaning improvement after discharge might be limited 461 for the two considered ICUs, general cleaning improvement of the environment might be 462 important to reduce permanent reservoirs for environmental contamination. 463

The results of our analysis build on a data-augmented MCMC algorithm [19, 26]. 464 Markov chain Monte Carlo sampling is a powerful tool to estimate posterior parameter 465 distribution whenever the likelihood is analytically intractable. And yet, the inherent 466 disadvantage of this sampling scheme is that it may take prohibitively many iterations 467 to converge to the posterior distribution. The convergence properties of MCMC 468 sampling in high-dimensional posterior distributions can be particularly problematic 469 and sensitive to the choice of prior and proposal distributions. Thus, tuning of the 470 MCMC parameters becomes crucial for its application. Our developed full model is 471 overparametrized when colonizations of patients are not or hardly influenced by 472 environmental contamination. As a result, the respective likelihood might not be 473 identifiable or only weakly identifiable. Here, a flat prior for the decay rate μ in 474 combination with a small initial standard deviation for its proposal distribution resulted 475 in large acceptance ratios close to 1. The MCMC chain mixed too slowly and therefore 476 hindered the identifiability of the likelihood. We were able to tune the parameters of the 477 proposal distribution for μ such that rapid convergence to the posterior distribution 478 could be assessed using visual inspection of histograms and trace plots. However, as 479 presented in the *Results* section, pairwise scatter plots showed strong correlations in 480 particular between ϵ and μ. Simulation studies confirmed that this can be explained by 481 an absence of environmental contamination in the investigated data sets. This supports 482 our finding that an impact of environmental contamination after discharge on the 483 transmission of *P. aeruginosa* may be neglected. 484

Moreover, colonization is assumed to remain until discharge. While this assumption is 485 true for *P. aeurginosa* it does not hold true for all antibiotic-resistant nosocomial 486 pathogens. However, intermittent carriage may be readily included allowing the method 487 to be generalized to other pathogens. 488

We assumed no difference in transmissibility between different strains of *P. aeruginosa* 489 and that all colonized patients are equally likely to transmit the pathogen. While 490 information on antibiotic resistance or microbial genotyping in combination with 491 epidemiological data may aid in distinguishing different strains and identifying specific 492 transmission events, only the uncertainty of the estimates would be affected. In 493 particular, the widths of the credibility intervals are likely to be reduced, but we do not 494 expect a large effect on the parameter estimates. 495

Assessing the fit of the model to the data is crucial to model building. The true relative 496 importance of the different routes of colonization in ICUs is generally unknown. 497 Genotyping data that might be used to demonstrate the source of the acquired 498 colonization is generally scarce and was not available for the data used in our analysis. 499 While the posterior predictive p-value is a popular method for assessing model fit, it has 500 been increasingly criticized for its self-fulfilling nature [27]. Furthermore, the choice of 501 the test statistic is crucial in order to adequately summarize discrepancies between 502 datasets. Rather than relying on a suitable summary statistic, we presented a model 503 assessment method that evaluates whether the estimated force of infection adequately 504 represents the transmission dynamics in the ward. However, while the corresponding 505 coverage probabilities may depict discrepancies per bin of the force of infection, the 506 sample size is not controlled by choosing the number MCMC updates. It might well 507 occur that specific patients (and their acquisition events) appear in more than one 508

MCMC update simultaneously. Thus, the true sample size is estimated to be smaller. Further improvement of the method presented here or development of other methods would be a vital topic for assessing epidemic models.

Model selection was performed using the DIC which is known to display poor performance (i.e. identifying the correct model) for complex likelihood functions such as those corresponding to epidemic models. Comparing the plausibility of different models is crucial for selecting the model that describes the dynamics of the observed system best. Nevertheless, model choice for stochastic epidemic models is far from trivial. All known approaches for model selection exhibit advantages as well as disadvantages [27] which makes selecting the most suitable model comparison technique not straightforward. We selected the well-known DIC-method that was easy to use and implement. Our main results regarding environmental contamination after discharge do not depend on the model choice. And yet, the development of a suitable and robust model selection procedure in a data-augmented Bayesian framework would be an interesting and important topic for future research.

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

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

Acknowledgments

Supporting information

S1 Text. Environmental contamination. The full model includes environmental contamination on a ward-level. The bacterial load at any given time t is based on the differential equation

$$\frac{dE}{dt} = \nu \frac{I(t)}{N(t)} - \mu E(t) \tag{5}$$

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

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

and thus

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

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

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

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

Now, two cases have to be distinguished.

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

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

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

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

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

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

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

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

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

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

and therefore

$$B_{t_0} = E_{t_0} - \frac{\nu}{\mu} \frac{I_{t_0}}{N_{t_0}}. \tag{6}$$

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

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

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

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

$$B_{t_i} = \frac{E_{t_i} - \frac{\nu}{\mu} \frac{I_{t_i}}{N_{t_i}}}{e^{-\mu t_i}} \tag{7}$$

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

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

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

and

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

for $0 \le i \le N$. 540

S2 Text. Discrete-time transmission model. For the discrete-time transmission model, we assume that the number of colonized patients $I(t)$, the total number of patients $N(t)$ and the bacterial load $E(t)$ is constant during the day. It is assumed that admission and screening occur at 12:00 pm on each day T determining I_T and N_T. Given all the information (at 12:00 pm), the environmental contamination on day T is determined. The force of infection on day T is then given by

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

541

S3 Text. Relative contribution. The computations in section *Relative contributions of transmission routes* were developed for continuous-time models. In our discrete-time model, we assume that events such as, admission, colonization and discharge of patients and screening occur on a daily basis. However, we do assume that the level of environmental contamination changes continuously. Computing the relative contributions of the different transmission routes becomes more laborious in this scenario. Let t_i^c be the acquisition time of patient $i \in \{1, \ldots, n\}$. The contribution of a route j is the ratio of the probability that the acquisition was due to route j and the total probability of acquisition:

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

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

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

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

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

where environmental contamination during a patient's stay is assigned to the 542
environmental route. For the alternative assignment (i.e. only considering the bacterial 543
load remaining after discharge as environmental contamination), the formulas change 544
according to Eq. (3). The results are dependent on $\nu \frac{I_T}{N_T} - \mu E(T)$ and are given by 545

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

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

19/29

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

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

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

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

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

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

and

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

where

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

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

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

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

S4 Text. Approximation of relative contribution in discrete-time. Large values of the force of infection $\lambda(t)$ are very unlikely. Under the assumption of small $\lambda(t)$, the following simplifications and approximations can be made:

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

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

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

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

- Contribution of endogenous route $= R_{\mathrm{end}} = \dfrac{\sum_{i=1}^{n} \frac{\alpha}{\lambda(t_i^c)}}{N_{\mathrm{acq}}}$ 551

- Contribution of cross-transmission $= R_{\mathrm{crossT}} = \dfrac{\sum_{i=1}^{n} \frac{\beta \cdot \frac{I(t_i^c)}{N(t_i^c)} + \epsilon \sum_{i_p} \frac{E_{i_p}(t_i^c)}{N(t_i^c)}}{\lambda(t_i^c)}}{N_{\mathrm{acq}}}$ 552

- Contribution of environmental contamination $= R_{\mathrm{env}} = \dfrac{\sum_{i=1}^{n} \frac{\epsilon \left[\sum_{i_d} \frac{E_{i_d}(t_i^c)}{N(t_i^c)} + E_0 e^{-\mu t_i^c} \right]}{\lambda(t_i^c)}}{N_{\mathrm{acq}}}$ 553

where t_i^c is the day of colonization of patient $i \in \{1, \ldots, n\}$ and N_{acq} the total number   554
of occured colonizations. Furthermore, i_p indicates a colonized patient that is present at   555
time t_i^c and i_d a colonized patient that has been colonized prior to t_i^c but was already   556
discharged. It holds $R_{\mathrm{end}} + R_{\mathrm{crossT}} + R_{\mathrm{env}} = 1$.   557

S5 Text. Adapted data-augmented MCMC algorithm. We model the   558
transmission process using a two-state Markov model, where each patient can be either   559
susceptible (P.A. negative) or *colonized* (P.A. positive). A patient i is admitted to the   560
ICU on day t_i^a and discharged on day t_i^d. The probability that a patient is admitted   561
already colonized is described by parameter f. The rate at which a susceptible patient   562
transitions to being colonized is described in section *Materials and methods*. The   563
colonization state of an individual patient is determined from screening information. We   564
suppose that for each patient i a set of screening results $X_i = X_i^{(1)}, \ldots, X_i = X_i^{(m)}$,   565
taken on days $t_i^{(1)}, \ldots, t_i^{(m)}$ is available. The set of all screening results is denoted by   566
$X = \{X_1, \ldots, X_N\}$ where N is the total number of patients. Since screening tests are   567
typically intermittent and imperfect, we define the test sensitivity ϕ (i.e. probability   568
that a colonized patient has a positive result). We assume that the specificity (i.e.   569
probability that an uncolonized patient has a negative result) is 100%.   570
We implemented an adapted version of the data-augmented MCMC algorithm to
analyze the data. The transmission and importation model, as well as the
data-augmentation method is closely based on the approach of [19, 26] but adapted for
the transmission routes presented in this paper. The algorithm was implemented in
C++ and the analysis of the output was performed in R (Version 3.5.1) [28]. The aim of
our analysis was to estimate the set of parameters $\theta = \{\alpha, \beta, \epsilon, \mu, f, \phi\}$. The prior
distribution were chosen as follows:

$$f, \phi \propto \mathrm{Beta}(a, b)$$
$$\alpha, \beta, \epsilon, \mu \propto \mathrm{Exp}(\lambda)$$

where $\mathrm{Exp}(\lambda)$ represents the exponential distribution with rate λ, and $\mathrm{Beta}(a, b)$ the   571
beta distribution with shape parameters a and b. Having fixed $a = b = 1$ and $\lambda = 0.001$,   572
we use uninformative priors in our analysis.   573
The data-augmentation procedure accounts for unobserved colonization times by
augmenting the parameter space with $A = \{t^c, s^a\}$, a set comprising of the unobserved
colonization times t^c and admission states s_a of all n patients. An admission state of a
patient is 1 if the patient is colonized upon admission and 0 otherwise. If the patient j
becomes colonized during his/her stay, the colonization time may take an integer value
between the time of admission t_j^a and time of discharge t_j^d (inclusive). If a patient does
not acquire colonization, the respective value t_j^c takes a dummy value of -1. The
augmented posterior density relation can be determined using Bayes' Theorem:

$$P(A, \theta \mid D) \propto P(D, A, \theta) = P(D \mid A, \theta) P(A \mid \theta) P(\theta) \tag{8}$$
$$= P(D \mid t^c, s^a, \theta) P(s^a \mid \theta) P(t^c \mid s^a, \theta) P(\theta) \tag{9}$$

where $P(D \mid A, \theta)$ is the likelihood of the observed data D, $P(A \mid \theta)$ is the likelihood of the augmented data and $P(\theta)$ is the joint prior distribution of the parameter set θ. All terms in (8) can be explicitly calculated. It holds

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

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

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

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

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

$$(10)$$

To update the importation rate f and the sensitivity ϕ, we use Gibbs sampling as we can sample directly from the full conditional distributions. The transmission parameters $\alpha, \beta, \epsilon, \mu$ are updated using an adapted version of the Metropolis-Hastings algorithm. Regular MCMC methods based on the Metropolis-Hastings algorithm tend to be very slow in high dimensions as a result of slow mixing and therefore inefficient convergence towards the target distribution. In high-dimensional spaces the volume outside is much larger than the volume of our target distribution. Thus, traditional MCMC methods such as the Metropolis-Hastings algorithm, spend considerable amount of time of traversing space away from the mode of the target distribution. Our adapted MCMC algorithm aims in exploring the target distribution more efficiently.

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

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

the proposed value is accepted.

The Metropolis algorithm is a special case of the Metropolis-Hastings algorithm where the proposal function is symmetrical. Since a symmetrical proposal distribution simplifies the calculation of the acceptance ratio to $a(\theta^*, \theta^{(i)}) = \min\left(1, \pi(\theta^*)/\pi(\theta^{(i)})\right)$, it is often used for updating parameters. The proposal function has a great influence on the speed of convergence and hence efficiency of the algorithm. We suggest a proposal distribution that speeds up the convergence towards the target distribution while limiting the additional computational effort. The idea behind our method is as follows: For each estimated parameter set θ there is a corresponding force of infection $\lambda(t)$ for each time t. It can be assumed that the mean force of infection $\overline{\lambda}$ is approximately constant over the number of iterations. The rationale behind it is that there is *true*

mean force of infection that should be approximated by the MCMC algorithm. Proposing new parameter candidates depending on the mean force of infection reduces the volume that has to be traversed in order to converge to the target distribution. The resulting proposal density is not symmetric anymore and thus the procedure requires an adjustment of the acceptance ratio. The adapted Metropolis-Hastings algorithm to update the transmission parameters runs as follows:

Two transmission routes

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

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

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

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

   (c) With probability

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

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

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

Three transmission routes

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

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

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

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

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

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

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

       i.e. $\mathcal{N}\left(\frac{\lambda^{(i)} - \alpha^* - \beta^* \cdot \overline{Prev}}{E^{(i+1}}, \frac{\sigma_\alpha^2}{E^{(i+1)2}} + \frac{\sigma_\beta^2}{\left(\frac{E^{(i+1)}}{\overline{Prev}}\right)^2} + \sigma_\epsilon^2\right)$.

   (f) With probability

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

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

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

S6 Text. Model selection. We would like to assess whether we can concatenate the Besançon data e.g. before and after the renovation of the ICUs in one large data set to increase the power of our method. The idea is to compare the DICs for two different scenarios:

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

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

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

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

  The parameter set of the model is then:

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

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

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

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

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

S7 Text. Simulation studies. Several data sets were simulated to test the algorithm's ability to identify the correct relative contribution. We fixed values for α, β and ν and varied values for μ. The parameter ϵ is set to $0.6 \cdot \mu$. Three main scenarios regarding the duration of persistence of bacterial load in the environment are analyzed:

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

2. Medium: $\mu = 1/3$

3. Short: $\mu = 1$

For each of the three scenarios the importation rate is varied within $\{0.01, 0.05, 0.1\}$. We observed that for long and short durations of bacterial persistence, chain convergence cannot be attained in a reasonable amount of time using uninformative exponential priors. A long duration of bacterial persistence leads to a model that cannot be distinguished from one with a higher endogenous contribution and small contribution of environmental contamination. On the other hand, a short duration of bacterial persistence leads to difficulties in distinguishing the resulting model from one with a higher contribution of cross-transmission and smaller contribution of environmental contamination. However, informative priors on μ (e.g. uniform priors $U(0,2)$) led to more rapid convergence of the MCMC chain.

For a medium length of bacterial persistence, the model is able to estimate the simulated parameter values i.e. the true parameter values and respective relative contributions lie in the 95% credibility intervals, given the mean prevalence was large enough ($> 15\%$). We have performed further simulation studies where the relative contribution of environmental contamination after discharge varied. In particular, when environmental contamination after discharge was not present in the simulated data, the results resembled our analysis of the Besançon data shown in the *Results* section. Further information on our simulation studies may be requested from the first author.

S8 Text. Secondary analyses. In addition to the analyses presented in *Results*, six further analyses were performed. For each ICU, the time periods before and after renovation were combined. Finally, all available data was concatenated into one big data set and analyzed at once. The results of these analyses using the submodel as well as the full model are presented in S9 - S10 Tables. The posterior estimates of the model parameters and the corresponding relative contributions are similar to the ones presented in the *Results* section.

S9 Table. Summary statistics of the marginal posterior distributions for model parameters of the submodel.

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

*Highest posterior density interval

S10 Table. Summary Statistics of the marginal posterior distributions for model parameters of the full model.

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

*Highest posterior density interval

S11 Fig. Pairwise plots of samples from the posterior distribution for the transmission parameters of the submodel. The plots were generated from the data of ICU A before renovation using the submodel with endogenous and exogenous transmission.

S12 Fig. Histograms for ICU A before renovation using the submodel with endogenous route and cross-transmission.

S13 Fig. Traceplots for ICU A before renovation using the submodel with endogenous route and cross-transmission.

S14 Fig. Histograms for ICU A after renovation using the submodel with endogenous route and cross-transmission.

S15 Fig. Traceplots for ICU A after renovation using the submodel with endogenous route and cross-transmission.

S16 Fig. Histograms for ICU A before renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S17 Fig. Histograms for ICU A before renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i where $i \in \{$endogenous, cross-transmission, environment$\}$.

S18 Fig. Traceplots for ICU A before renovation using the full model with

endogenous route, cross-transmission and environmental contamination.
The results are displayed for transmission parameters α, β, ϵ and μ.

S19 Fig. Traceplots for ICU A before renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i where $i \in \{\text{endogenous, cross-transmission, environment}\}$.

S20 Fig. Histograms for ICU A after renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S21 Fig. Histograms for ICU A after renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i where $i \in \{\text{endogenous, cross-transmission, environment}\}$.

S22 Fig. Traceplots for ICU A after renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for transmission parameters α, β, ϵ and μ.

S23 Fig. Traceplots for ICU A after renovation using the full model with endogenous route, cross-transmission and environmental contamination. The results are displayed for importation probability f, sensitivity parameter ϕ, log-likelihood and relative contributions R_i where $i \in \{\text{endogenous, cross-transmission, environment}\}$.

S24 Fig. Pairwise plots of samples from the posterior distribution for the transmission parameters of the full model. The plots were generated from the data of ICU A before renovation using the full model with endogenous route, cross-transmission and environmental contamination.

S25 Fig. Coverage probabilities for the submodel using Jeffreys prior. (a) - (b) ICU A before and after renovation, respectively. (c) - (d) ICU B before and after renovation, respectively.

S26 Fig. Coverage probabilities for the full model using Jeffreys prior. (a) - (b) ICU A before and after renovation, respectively. (c) - (d) ICU B before and after renovation, respectively.

References

1. Cassini A, Plachouras D, Eckmanns T, Abu Sin M, Blank HP, Ducomble T, et al. Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study. PLoS Medicine. 2016;13(10). doi:10.1371/journal.pmed.1002150.

2. Pittet D, Allegranzi B, Storr J, Nejad SB, Dziekan G, Leotsakos A, et al. Infection control as a major World Health Organization priority for developing

countries. Journal of Hospital Infection. 2008;68(4):285–292. doi:10.1016/j.jhin.2007.12.013.

3. Jarvis WR, Martone WJ. Predominant pathogens in hospital infections. The Journal of Antimicrobial Chemotherapy. 1992;29 Suppl A:19–24.

4. Spencer RC. Predominant pathogens found in the European Prevalence of Infection in Intensive Care Study. European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology. 1996;15(4):281–285.

5. Tantracheewathorn T, Vititpatarapak N, Phumisantiphong U. Epidemiologic study of nosocomial bacterial infection of pediatric patients at BMA Medical College and Vajira Hospital. Journal of the Medical Association of Thailand = Chotmaihet Thangphaet. 2007;90(2):258–265.

6. Kerr KG, Snelling AM. Pseudomonas aeruginosa: a formidable and ever-present adversary. Journal of Hospital Infection. 2009;73(4):338–344. doi:10.1016/j.jhin.2009.04.020.

7. WHO | Guidelines for the prevention and control of carbapenem-resistant Enterobacteriaceae, *Acinetobacter baumannii* and *Pseudomonas aeruginosa* in health care facilities;. Available from: `http://www.who.int/infection-prevention/publications/guidelines-cre/en/`.

8. Gruner E, Kropec A, Huebner J, Altwegg M, Daschner F. Ribotyping of Pseudomonas aeruginosa strains isolated from surgical intensive care patients. The Journal of Infectious Diseases. 1993;167(5):1216–1220.

9. Bonten MJM, Bergmans DCJJ, Speijer H, Stobberingh EE. Characteristics of Polyclonal Endemicity ofPseudomonas aeruginosa Colonization in Intensive Care Units. American Journal of Respiratory and Critical Care Medicine. 1999;160(4):1212–1219. doi:10.1164/ajrccm.160.4.9809031.

10. Bertrand X, Thouverez M, Talon D, Boillot A, Capellier G, Floriot C, et al. Endemicity, molecular diversity and colonisation routes of *Pseudomonas aeruginosa* in intensive care units. Intensive Care Medicine. 2001;27(8):1263–1268. doi:10.1007/s001340100979.

11. Vallés J, Mariscal D, Cortés P, Coll P, Villagrá A, Díaz E, et al. Patterns of colonization by *Pseudomonas aeruginosa* in intubated patients: a 3-year prospective study of 1,607 isolates using pulsed-field gel electrophoresis with implications for prevention of ventilator-associated pneumonia. Intensive Care Medicine. 2004;30(9):1768–1775. doi:10.1007/s00134-004-2382-6.

12. Lashéras A, Guisset O, Boulestreau H, Rogues AM, Fiore M, Szajner S, et al. Réservoirs et transmission de Pseudomonas aeruginosa en réanimation médicale. Médecine et Maladies Infectieuses. 2006;36(2):99–104. doi:10.1016/j.medmal.2005.11.014.

13. Kramer A, Schwebke I, Kampf G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infectious Diseases. 2006;6:130. doi:10.1186/1471-2334-6-130.

14. McBryde ES, McElwain DLS. A Mathematical Model Investigating the Impact of an Environmental Reservoir on the Prevalence and Control of Vancomycin-Resistant Enterococci. The Journal of Infectious Diseases. 2006;193(10):1473–1474. doi:10.1086/503439.

15. Wolkewitz M, Dettenkofer M, Bertz H, Schumacher M, Huebner J. Environmental Contamination as an Important Route for the Transmission of the Hospital Pathogen VRE: Modeling and Prediction of Classical Interventions , Environmental Contamination as an Important Route for the Transmission of the Hospital Pathogen VRE: Modeling and Prediction of Classical Interventions. Infectious Diseases: Research and Treatment. 2008;1:IDRT.S809. doi:10.4137/IDRT.S809.

16. Wang L, Ruan S. Modeling Nosocomial Infections of Methicillin-Resistant Staphylococcus aureus with Environment Contamination*. Scientific Reports. 2017;7. doi:10.1038/s41598-017-00261-1.

17. Wang X, Xiao Y, Wang J, Lu X. A mathematical model of effects of environmental contamination and presence of volunteers on hospital infections in China. Journal of Theoretical Biology. 2012;293:161–173. doi:10.1016/j.jtbi.2011.10.009.

18. Plipat N, Spicknall IH, Koopman JS, Eisenberg JN. The dynamics of methicillin-resistant Staphylococcus aureusexposure in a hospital model and the potential for environmental intervention. BMC Infectious Diseases. 2013;13(1):595. doi:10.1186/1471-2334-13-595.

19. Cooper BS, Medley GF, Bradley SJ, Scott GM. An augmented data method for the analysis of nosocomial infection data. American Journal of Epidemiology. 2008;168(5):548–557. doi:10.1093/aje/kwn176.

20. Diekmann O, Heesterbeek H, Britton T. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press; 2013.

21. Andersson H, Britton T. Stochastic Epidemic Models and Their Statistical Analysis. Lecture Notes in Statistics. New York: Springer-Verlag; 2000. Available from: //www.springer.com/us/book/9780387950501.

22. Lavenir R, Jocktane D, Laurent F, Nazaret S, Cournoyer B. Improved reliability of Pseudomonas aeruginosa PCR detection by the use of the species-specific ecfX gene target. Journal of Microbiological Methods. 2007;70(1):20–29. doi:10.1016/j.mimet.2007.03.008.

23. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, Second Edition. CRC Press; 2003.

24. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Statistical Science. 2001;16(2):101–133. doi:10.1214/ss/1009213286.

25. Panagea S, Winstanley C, Walshaw MJ, Ledson MJ, Hart CA. Environmental contamination with an epidemic strain of Pseudomonas aeruginosa in a Liverpool cystic fibrosis centre, and study of its survival on dry surfaces. Journal of Hospital Infection. 2005;59(2):102–107. doi:10.1016/j.jhin.2004.09.018.

26. Worby CJ, Jeyaratnam D, Robotham JV, Kypraios T, O'Neill PD, De Angelis D, et al. Estimating the effectiveness of isolation and decolonization measures in reducing transmission of methicillin-resistant Staphylococcus aureus in hospital general wards. American Journal of Epidemiology. 2013;177(11):1306–1313. doi:10.1093/aje/kws380.

27. Gibson GJ, Streftaris G, Thong D. Comparison and Assessment of Epidemic Models. Statistical Science. 2018;33(1):19–33. doi:10.1214/17-STS615.

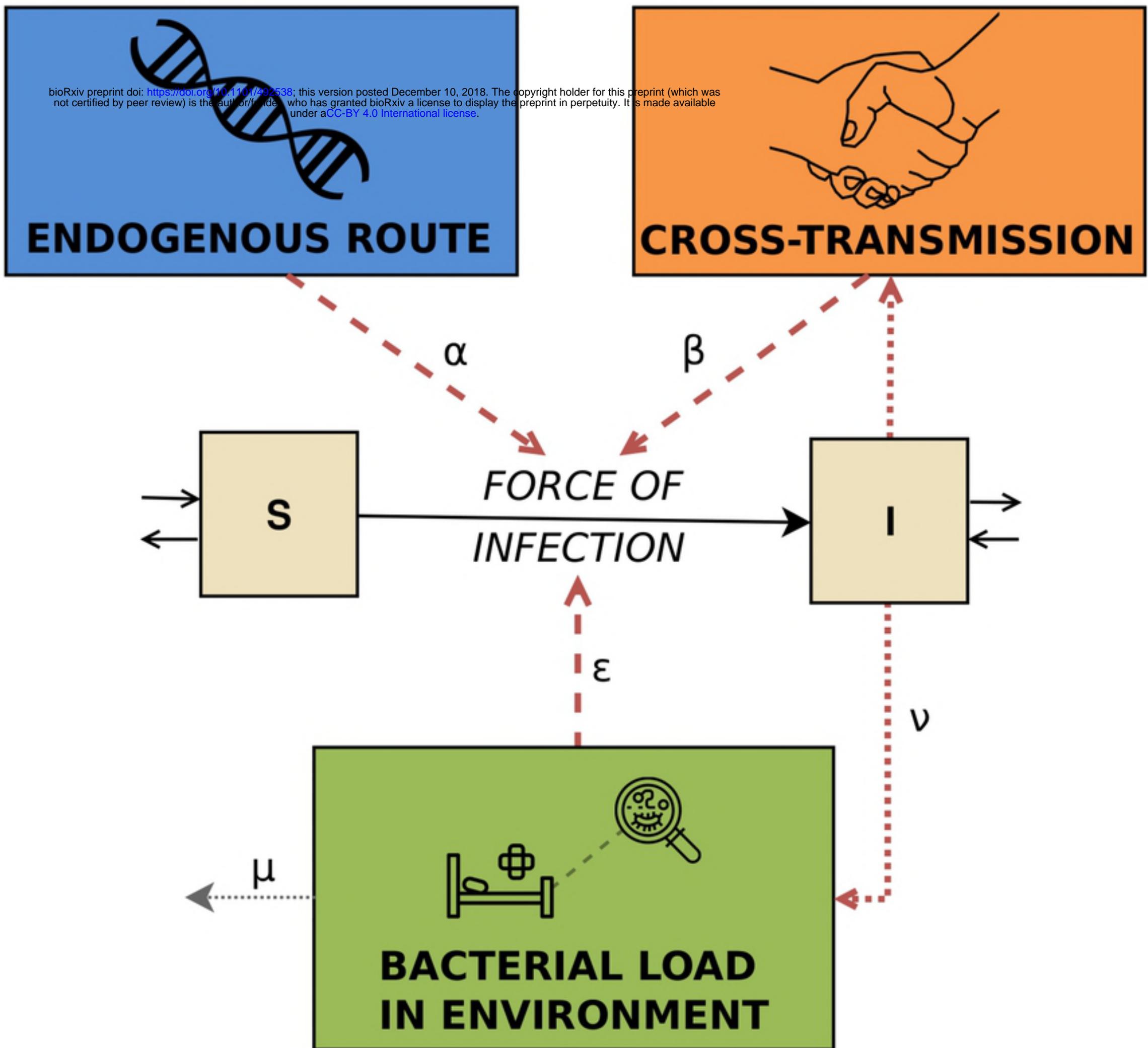28. R Core Team. R: A Language and Environment for Statistical Computing; 2018.
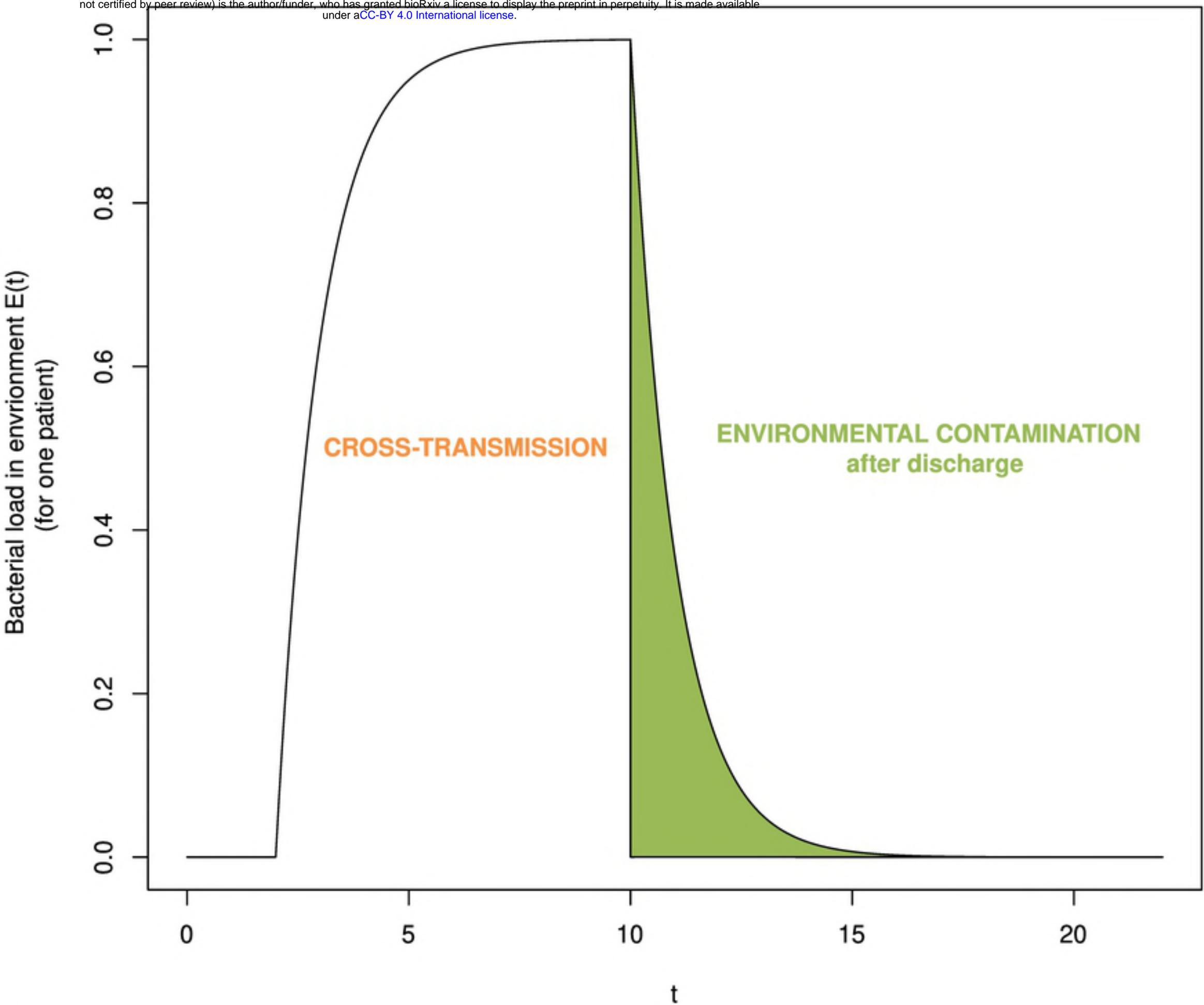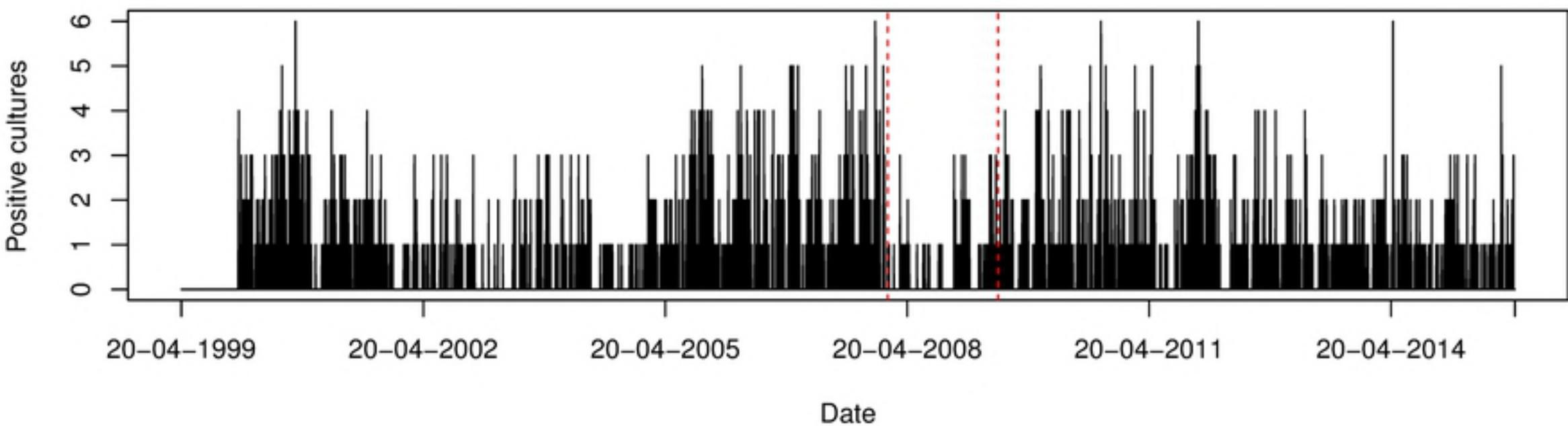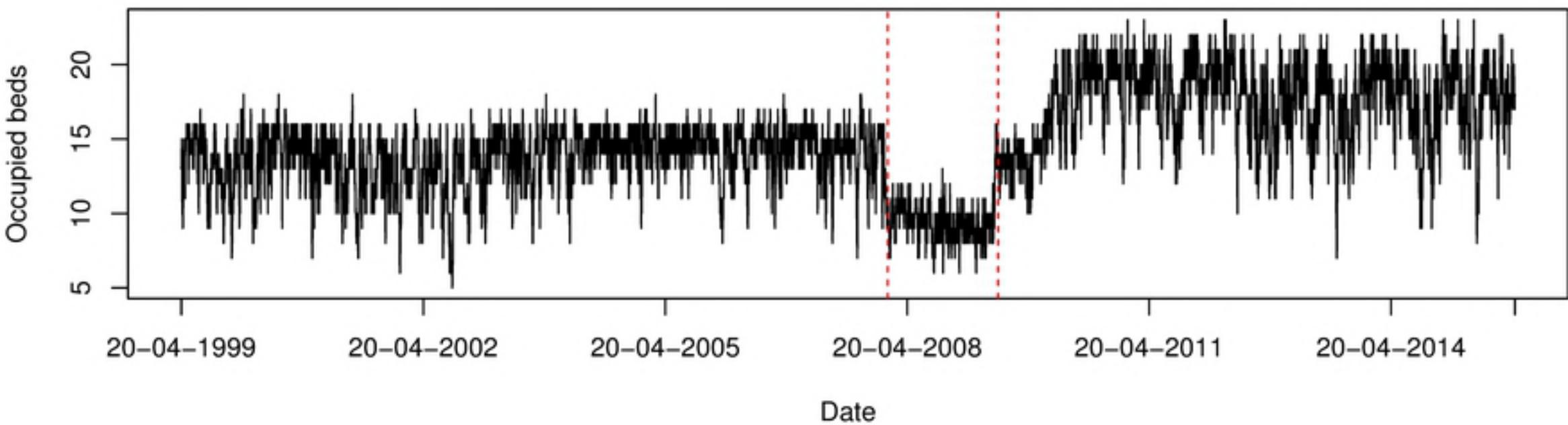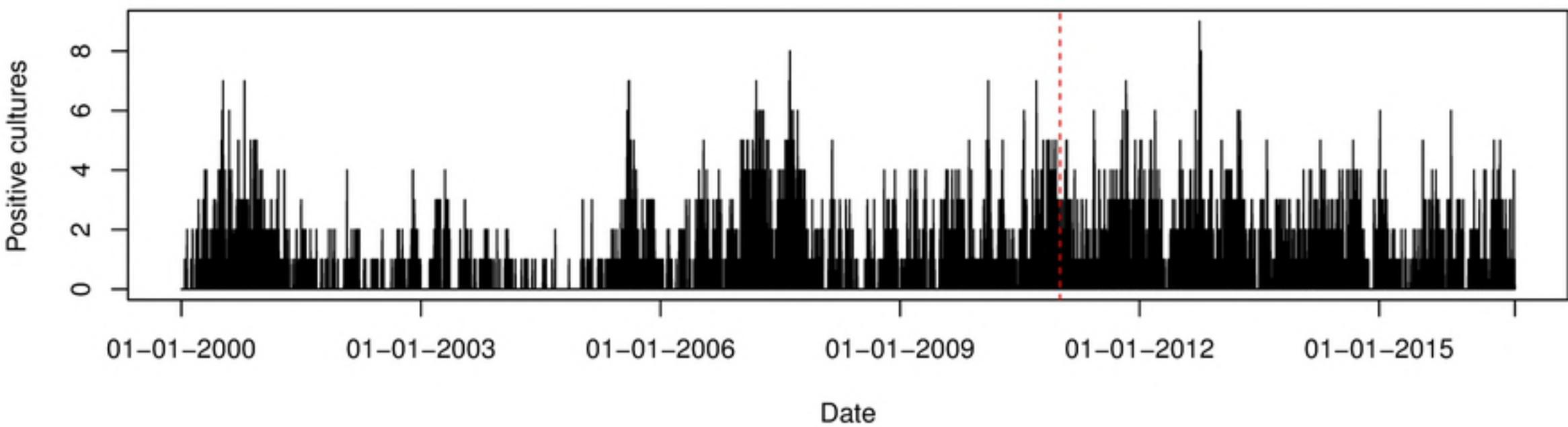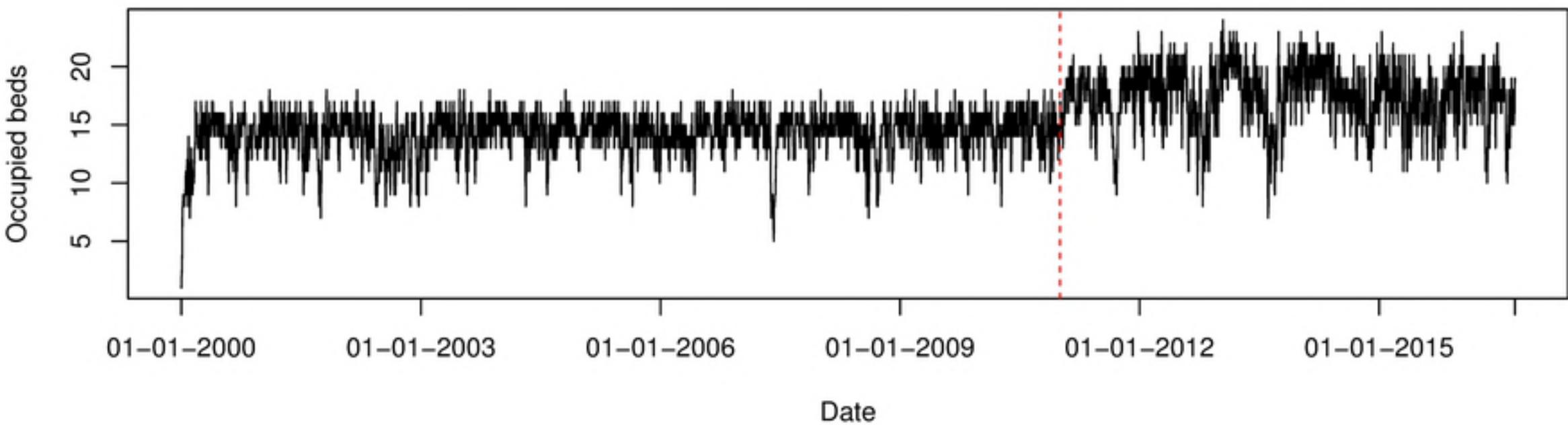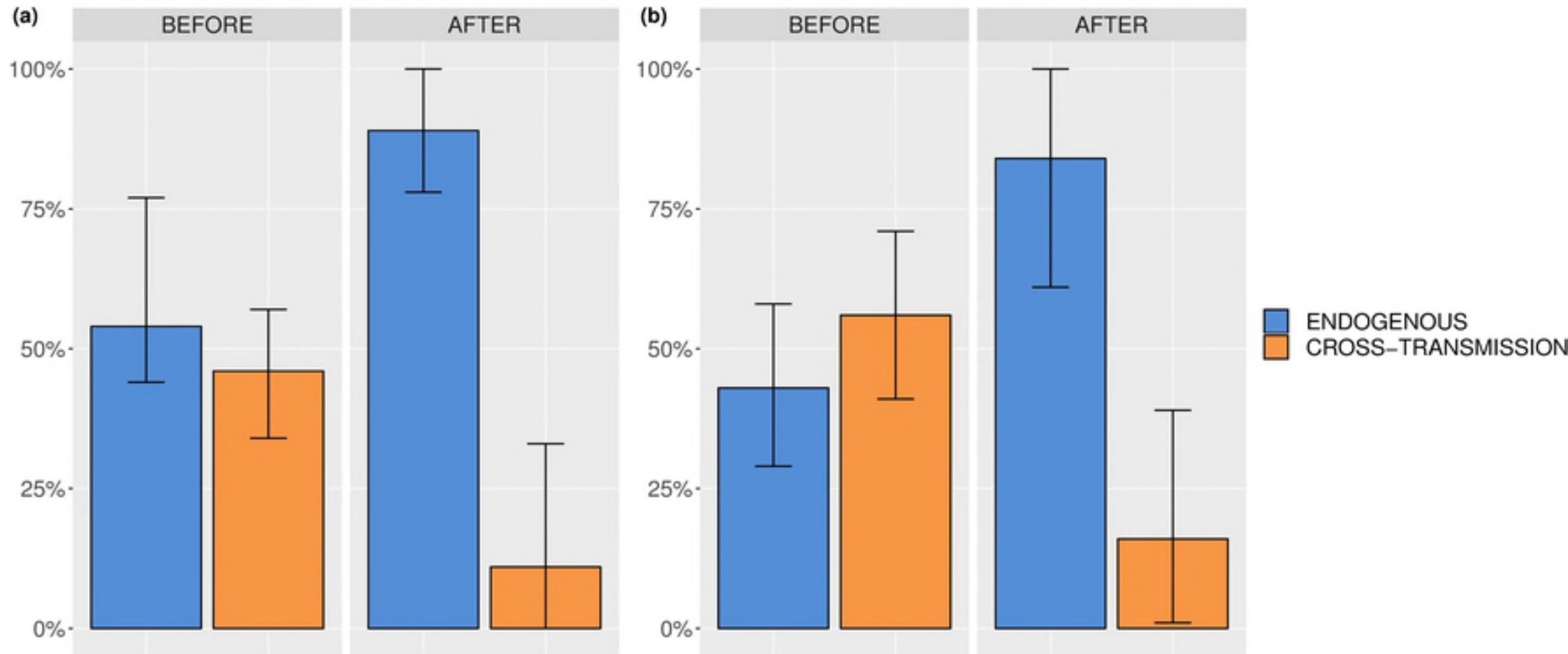   Available from: `https://www.R-project.org/`.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7