A unified stochastic modelling framework for the spread of nosocomial infections

Martín López-García[1*], Theodore Kypraios[2]

[1]*School of Mathematics, University of Leeds, LS2 9JT Leeds, UK*
[2]*School of Mathematical Sciences, University of Nottingham, NG7 2RD Nottingham, UK*

11^{th} May 2018

Abstract

Over the last years, a number of stochastic models have been proposed for analysing the spread of nosocomial infections in hospital settings. These models often account for a number of factors governing the spread dynamics: spontaneous patient colonization, patient-staff contamination/colonization, environmental contamination, patient cohorting, or health-care workers (HCWs) hand-washing compliance levels. For each model, tailor-designed methods are implemented in order to analyse the dynamics of the nosocomial outbreak, usually by means of studying quantities of interest such as the reproduction number of each agent in the hospital ward, which is usually computed by means of stochastic simulations or deterministic approximations. In this work, we propose a highly versatile stochastic modelling framework that can account for all these factors simultaneously, and which allows for the exact analysis of the reproduction number of each agent at the hospital ward during a nosocomial outbreak. By means of five representative case studies, we show how this unified modelling framework comprehends, as particular cases, many of the existing models in the literature. We implement various numerical studies via which we: i) highlight the importance of maintaining high hand-hygiene compliance levels by HCWs, ii) support infection control strategies including to improve environmental cleaning during an outbreak, and iii) show the potential of some HCWs to act as super-spreaders during nosocomial outbreaks.

Keywords: hospital-acquired or nosocomial infections; antibiotic resistant bacteria; infection control; stochastic model; Markov chain; reproduction number

1 Introduction

The risk of acquiring nosocomial infections is a recognised problem in health-care facilities worldwide [1]. It has been estimated that nosocomial infections affect more than 4 million patients in Europe each year, leading to €7 billion of direct medical costs [2]. Moreover, the emergence and spread of antibiotic resistance among these pathogens has posed a second major problem worldwide, stressing the need for understanding their transmission routes in health-care facilities, and to identify the most effective infection control strategies in these settings [3]. A paradigmatic example of an antibiotic resistant nosocomial pathogen is bacteria *Staphylococcus aureus* (SA), which is a normal inhabitant of the skin and mucosal surfaces, but can cause different infections when it flourishes in other areas (e.g., soft tissue, bloodstream or lung infections). SA resistance against Penicillin-like antibiotics arose a few years after the introduction of Penicillin. Moreover, Methicillin-resistant SA (MRSA) strains were reported in Europe after only two years of the introduction of Methicillin in 1959 [4]. Currently, new strains of MRSA have been reported which are also resistant to Vancomycin [4].

Health-care environments such as hospitals or nursing homes are ideal settings for the spread of multidrug-resistant bacteria (MDRB), due to, among other reasons, opportunities for bacteria to enter into the bloodstream or infect open wounds, the presence of immunocompromised and aged individuals, and the high exposure levels to antibiotics [5, 6]. The precise mode of transmission is uncertain for many nosocomial pathogens, but usually both exogenous (*e.g,* cross-colonization) and endogenous (*e.g.* selective pressure of antibiotics) routes are considered as feasible for these pathogens [3]. While for some nosocomial infections most of the transmission is considered to occur via HCW-patient contact routes [7], there is increasing recognition in the literature of the potential role played by environmental contamination and airborne spread [8, 9, 10].

Infection control strategies usually implemented in hospital settings include, among others, hand disinfection procedures, environmental cleaning, active screening for colonization among patients and isolation of colonized individuals, managing staffing levels, antibiotic prescription and decolonization procedures, or patient cohorting [11]. However, control procedures followed in health-care facilities world-

*Author for correspondence (m.lopezgarcia@leeds.ac.uk)

wide usually amount to combinations of the individual interventions listed above, so that the efficacy of each individual strategy is hard to measure. On the other hand, the application of classical epidemiology procedures for addressing this individual efficacy is often not feasible due to financial and ethical restrictions [4, 12]. Thus, mathematical modelling is one of the best tools available for understanding the role played by different factors on the emergence and spread of these pathogens and their antibiotic resistance, while measuring the impact of individual interventions [8, 13].

A wide range of deterministic and stochastic mathematical models for the spread of nosocomial pathogens have been developed during the last years [2]. Although deterministic models were originally proposed for capturing the main infection dynamics in single wards and hospitals, modelling efforts were soon redirected towards the stochastic perspective due to the small and highly heterogeneous populations usually present in these settings. From a stochastic perspective, most of the models proposed in the literature are based on Markov processes, where it is assumed that inter-event times are exponentially distributed. This simplifying assumption is usually crucial for analytically and computationally treating the processes under study; we refer the reader to Ref. [3] for a discussion on the advantages of stochastic (in particular, Markovian) approaches, and to Ref. [2, 14] for systematic reviews in this field. Stochastic models in this area can be classified as *compartment-based*, where the population of individuals is classified in groups according to their state against the disease, and wide homogeneities are assumed among the members within the same group, or *agent-based*, which keep track of the state of each individual within the population throughout time and allow one to model heterogeneities at the individual level [8]. Agent-based models can incorporate heterogeneity in, for example, transmission risk profiles of specific patients or HCWs [21], but are usually restricted to the implementation of stochastic simulations in small wards, and are computationally constrained [2].

When constructing and studying these stochastic models, efforts have been focused, and tailor-designed analytical and numerical methods have been implemented, in order to analyse the dynamics of the nosocomial outbreak when accounting for spontaneous colonization of patients, patient-to-staff and staff-to-patient contamination/colonization, environmental contamination, patient cohorting, room configuration of the hospital ward, staff hand-washing compliance levels, the presence of different types of HCWs or specific staff-patient contact network structures. This analysis is usually carried out by means of studying summary statistics directly related to the nosocomial outbreak, such as the reproduction number of each particular agent (*e.g.,* of a colonized patient or a contaminated health-care worker) in the hospital ward. This is usually computed in an approximative fashion, for example by means of stochastic simulations or in terms of deterministic approximations [15]. On the other hand, the limitations of analysing these processes by simulation approaches, and the convenience of following exact procedures instead when dealing with small populations (such as those usually involved in nosocomial outbreaks), have been highlighted in Ref. [16].

In this work, we propose a versatile stochastic modelling framework that can simultaneously account for all the factors listed above, and which allows in Section 2 for the exact and analytical study of the reproduction number of each agent at the hospital ward during the nosocomial outbreak. We make use of five representative case studies in Section 3, regarding both hypothetical and real nosocomial outbreaks at hospital wards, to show how this unified modelling framework comprehend, as particular cases, many of the existing models in the field. We conduct several numerical studies and our results in Section 3 highlight the importance of maintaining high hand-hygiene compliance levels by health-care workers, support control strategies including to improve environmental cleaning during nosocomial outbreaks, and show the potential of some health-care workers to act as *super-spreaders* during these outbreaks.

2 A unified stochastic modelling framework

In this Section, we propose the unified stochastic modelling framework for the spread of nosocomial infections, where agents represented in the model can be of different *type* (patients, HCWs, surfaces, patients located in different rooms,...). This general framework, which is constructed in terms of a continuous-time Markov chain, allows one to follow an exact and analytical approach for computing the reproduction number of each different *agent* playing a role in the infection spread, which measures the number of *infections* directly caused by this agent until the agent stops spreading the nosocomial pathogen. We also show how this reproduction number can be exactly analysed while deciphering among which individuals this agent is spreading the disease, so that this becomes a quantitative measure of the infectiousness of a given *agent* among individuals of different *type*. This then becomes a useful tool when analysing the role played by different routes of infection during a nosocomial outbreak in a given hospital ward, as shown in numerical results in Section 3.

2.1 The model

We consider model depicted in Figure 1, which amounts to a stochastic SIS epidemic model with multiple *compartmental levels*. In Case Studies 1-5 in Section 3, this modelling framework is used to represent the spread of nosocomial infections, such as MDRB, within a hospital ward, where the meaning of a compartmental level depends on the particular case study, showing the versatility and flexibility of this unified framework.

We consider the stochastic process $\mathcal{X} = \{\mathbf{X}(t) = (I_1(t), \ldots, I_M(t)) : \ t \geq 0\}$, where $I_j(t)$ amounts to the number of infectives in compartmental level j at time $t \geq 0$. We assume that the number of individuals at each compartmental level remains constant throughout time, which is directly related to standard assumptions when modelling nosocomial infections; see Section 3. This means that the

number $S_j(t)$ of susceptibles in compartmental level j at time t is given by $S_j(t) = N_j - I_j(t)$ for all $t \geq 0$. Process $\mathcal{X}$ evolves among states in $\mathcal{S} = \mathcal{C} \cup \{\Delta\}$, where

$$\mathcal{C} \;=\; \{(i_1, \ldots, i_M) \in \mathbb{N}_0^M :\; 0 \leq i_j \leq N_j,\; j \in \{1, \ldots, M\}\}.$$

State $(i_1, \ldots, i_M)$ represents the presence of i_j infected individuals at compartmental levels $1 \leq j \leq M$, while the final state Δ represents the detection and declaration of the outbreak in the hospital ward. In particular, process $\mathcal{X}$ transits among states in $\mathcal{S}$ according to the following transitions:

- *Removal* **at compartmental level** j**:** $(i_1, \ldots, i_M) \to (i_1, \ldots, i_j - 1, \ldots, i_M)$, occurring at rate $\mu_j(i_1, \ldots, i_M)$;
- *Infection* **at compartmental level** j**:** $(i_1, \ldots, i_M) \to (i_1, \ldots, i_j + 1, \ldots, i_M)$, occurring at rate $\lambda_j(i_1, \ldots, i_M)$;
- **Detection and declaration of the outbreak:** $(i_1, \ldots, i_M) \to \Delta$, occurring at rate $\delta(i_1, \ldots, i_M)$.

This unified model has been developed to account for patients, different types of HCWs and/or surfaces involved in a nosocomial outbreak in a hospital ward. The generality of functions $\lambda_j(i_1, \ldots, i_M)$, $\mu_j(i_1, \ldots, i_M)$ and $\delta(i_1, \ldots, i_M)$ allows for incorporating into the model a wide range of factors having an impact on the nosocomial spread dynamics. This means that the particular meaning of each compartmental level $1 \leq j \leq M$, as well as of each event (infections and removals represented by arrows in Figure 1) depends on the particular hospital ward and pathogen under analysis; see Section 3 where compartmental levels $1 \leq j \leq M$ can represent colonized/non-colonized patients, contaminated/non-contaminated HCWs, volunteers and surfaces, or can be related to the specific spatial configuration of the hospital ward under analysis, or the particular staff-patient contact network (*e.g.,* representing patient cohorting).

Outbreak detection and declaration rate $\delta(i_1, \ldots, i_M)$ allows one to analyse situations where a nosocomial pathogen is introduced for the first time in a given hospital ward (*e.g.,* by admission of a colonized patient), starting an outbreak, and the spread dynamics are analysed until the presence of this pathogen is detected by HCWs. By conveniently specifying the function rate $\delta(i_1, \ldots, i_M)$, different hospital surveillance policies (*e.g.,* detection by the first individual showing symptoms, by random screening of patients within the ward, or by systematic screening upon patient admission) can be considered. However, as illustrated in Section 3, scenarios where the interest is not in the spread dynamics until detection, but in the long-term infection dynamics of the pathogen (*e.g.,* endemic situations) and in assessing the infectiousness of each agent within this ward, can be analysed by setting $\delta(i_1, \ldots, i_M) = 0$. We note that setting $\delta(i_1, \ldots, i_M) = 0$ means deleting the final state Δ in Figure 1, so that the infection dynamics during the nosocomial outbreak would amount to the stochastic movement of individuals, throughout time, between the susceptible and infective compartments at the different compartmental levels in Figure 1; see case studies 2-5.

In subsection 2.1, and for a given initial state $(I_1(0), \ldots, I_M(0)) = (i_1, \ldots, i_M)$, we analyse the *exact reproduction number* for an infective individual in compartmental level j: the number of infections (understood in a broad sense, see Section 3) directly caused by this individual until he/she is removed or until the outbreak is detected, $R^{(j)}_{(i_1,\ldots,i_M)}$; see Refs. [17, 18, 19]. Since an infective individual at compartmental level j can infect individuals at compartmental levels $1 \leq k \leq M$, one can split $R^{(j)}_{(i_1,\ldots,i_M)} = \sum_{k=1}^{M} R^{(j)}_{(i_1,\ldots,i_M)}(k)$, where $R^{(j)}_{(i_1,\ldots,i_M)}(k)$ is the number of infections directly caused by an infective individual at compartmental level j, among individuals at compartmental level k. In this way, random variables $R^{(j)}_{(i_1,\ldots,i_M)}(k)$, for $1 \leq j, k \leq M$, allow one to assess the role played by the different potential routes of infection during a nosocomial outbreak in a hospital ward, in our numerical results in Section 3. We note that the global variable $R^{(j)}_{(i_1,\ldots,i_M)}$ measures the infectiousness of an infective individual in compartmental level j, until this individual stops spreading the infection (he/she is removed) or until the outbreak is detected and declared (so that control strategies such as antibiotic prescription, isolation of infected individuals, patient cohorting, or environmental cleaning, can be implemented, impacting on the infection spread dynamics). These summary statistics can be studied from the solution of systems of linear equations, by implementing first-step arguments. In the Supplementary Material, we explain the corresponding algorithmic procedures designed for solving these systems in a matrix-oriented fashion.

2.2 Reproduction number for an individual at compartmental level j, among individuals at compartmental level k

For a given compartmental level j and a given initial state $(i_1, \ldots, i_M)$, we can define the random variable $R^{(j)}_{(i_1,\ldots,i_M)}$, which amounts to the total number of infections directly caused by a marked infective individual at compartmental level j until he/she is removed, or until the outbreak is declared. We note that since quantity $R^{(j)}_{(i_1,\ldots,i_M)}$ refers to an infective individual at compartmental level j, it is only properly defined for initial states $(i_1, \ldots, i_M)$ with $i_j > 0$. In case studies 1-5 in Section 3, we focus on initial states of the form

$$(0, \ldots, 0, \underbrace{1}_{j}, 0, \ldots, 0),$$

representing that the infective individual under study is the one at compartmental level j starting the outbreak. For this initial state, the mean value $E[R^{(j)}_{(0,\ldots,0,1,0,\ldots,0)}]$ directly relates to the *basic reproduction number* (measuring the average number of individuals this individual directly infects until he/she is removed –or, in this case, until the outbreak is detected–, for an initially fully susceptible population).

We note that $R^{(j)}_{(i_1,\ldots,i_M)}$ is in fact the sum of several contributions,

$$R^{(j)}_{(i_1,\ldots,i_M)} \;=\; \sum_{k=1}^{M} R^{(j)}_{(i_1,\ldots,i_M)}(k),$$

where $R^{(j)}_{(i_1,\ldots,i_M)}(k)$ represents the number of infections caused, by this individual who is at compartmental level j, only *among* individuals at compartmental level k. The analysis of each variable $R^{(j)}_{(i_1,\ldots,i_M)}(k)$ helps to measure not only how infectious an individual that belongs to compartmental level j is, but also how much of a risk he/she is for individuals at a given compartmental level k. This allows us in Section 3 to explore the role played by the different potential transmission routes during a nosocomial outbreak.

The probability distribution of each random variable $R^{(j)}_{(i_1,\ldots,i_M)}(k)$ is given in terms of probabilities

$$\nu^{(j)}_{(i_1,\ldots,i_M)}(k;n) \;=\; \mathbb{P}(R^{(j)}_{(i_1,\ldots,i_M)}(k)=n), \quad n\geq 0.$$

Since these probabilities refer to a particular infected individual, it is necessary to specify the contribution that each infective individual has in the global infection rates $\lambda_j(i_1,\ldots,i_M)$, as well as the rate at which this particular individual is removed. Thus, we analyse quantities $R^{(j)}_{(i_1,\ldots,i_M)}(k)$ and $R^{(j)}_{(i_1,\ldots,i_M)}$ for the following family of infection and removal rates:

$$\begin{aligned}\mu_j(i_1,i_2,\ldots,i_M) &= \mu_j i_j,\\ \lambda_j(i_1,i_2,\ldots,i_M) &= \left(\lambda_j+\sum_{k=1}^{M}\lambda_{kj}i_k\right)(N_j-i_j),\end{aligned}$$

for $1\leq j\leq M$, and any outbreak detection and declaration rate $\delta(i_1,\ldots,i_M)$. This specification of rates is based on the following general assumptions:

- Each infective individual at compartmental level j is removed independently at rate μ_j;
- Each susceptible individual at compartmental level j can be infected due to an external source of infection, with rate λ_j, or by an infective individual at compartmental level k, with rate λ_{kj}.

We note that these functions have been defined in this way so that they can be used in case studies 1-5 for the spread of nosocomial pathogens in hospital wards, where events related to rates μ_j, λ_j and λ_{kj} have specific meanings in each case study in Section 3, according to different scenarios and hypotheses considered in Refs. [18, 20, 21, 22, 23].

We follow here a first-step argument conditioning on the next event to occur in the process. In particular, for the initial state $\mathbf{i}=(i_1,\ldots,i_M)$, we have

$$\begin{aligned}\mathbb{P}(R^{(j)}_{\mathbf{i}}(k)=n) \;=\; & \mathbb{P}(R^{(j)}_{\mathbf{i}}(k)=n \mid \mathbf{i}\to\Delta)\mathbb{P}(\mathbf{i}\to\Delta)\\ &+\sum_{k=1}^{M}\mathbb{P}(R^{(j)}_{\mathbf{i}}(k)=n \mid \mathbf{i}\to(i_1,\ldots,i_k-1,\ldots,i_M))\\ &\qquad\times\mathbb{P}(\mathbf{i}\to(i_1,\ldots,i_k-1,\ldots,i_M))\\ &+\sum_{k=1}^{M}\mathbb{P}(R^{(j)}_{\mathbf{i}}(k)=n \mid \mathbf{i}\to(i_1,\ldots,i_k+1,\ldots,i_M))\\ &\qquad\times\mathbb{P}(\mathbf{i}\to(i_1,\ldots,i_k+1,\ldots,i_M)).\end{aligned} \tag{1}$$

Notation $\mathbf{i}\to(i_1,\ldots,i_k-1,\ldots,i_M)$ represents the event that, if the process is at state $\mathbf{i}$ at present time, the next event that occurs in the process is the transition to state $(i_1,\ldots,i_k-1,\ldots,i_M)$ (*i.e.,* a removal occurs at compartmental level k). The equation above, if we use notation

$$\begin{aligned}\mathbf{i} &= (i_1,\ldots,i_M),\\ \mathbf{i}^+(s) &= (i_1,\ldots,i_s+1,\ldots,i_M),\\ \mathbf{i}^-(s) &= (i_1,\ldots,i_s-1,\ldots,i_M),\end{aligned}$$

leads to the system of equations

$$\begin{aligned}\theta_{\mathbf{i}}\nu^{(j)}_{\mathbf{i}}(k;n) = & \underbrace{(\mu_j+\delta(\mathbf{i}))1_{n=0}}_{\substack{\text{Removal of the marked individual,}\\ \text{or outbreak declaration}}} + \underbrace{\mu_j(i_j-1)\nu^{(j)}_{\mathbf{i}^-(j)}(k;n)}_{\substack{\text{Removal of an individual,}\\ \text{–not the marked one–}\\ \text{at compartmental level (CL) } j}}\\ &+\underbrace{1_{n>0}\lambda_{jk}(N_k-i_k)\nu^{(j)}_{\mathbf{i}^+(k)}(k;n-1)}_{\substack{\text{Infection of an individual at CL } k,\\ \text{caused by the marked individual}}}+\sum_{p=1,\ p\neq j}^{M}\underbrace{\mu_p i_p\nu^{(j)}_{\mathbf{i}^-(p)}(k;n)}_{\substack{\text{Removal of an individual}\\ \text{at CL } p\neq j}}\\ &+\sum_{p=1,\ p\neq k}^{M}\underbrace{(N_p-i_p)\left(\lambda_p+\sum_{l=1}^{M}\lambda_{lp}i_l\right)\nu^{(j)}_{\mathbf{i}^+(p)}(k;n)}_{\text{Infection of an individual at CL } p\neq k}\\ &+\underbrace{(N_k-i_k)\left(\lambda_k+\lambda_{jk}(i_j-1)+\sum_{l=1,\ l\neq j}^{M}\lambda_{lk}i_l\right)\nu^{(j)}_{\mathbf{i}^+(k)}(k;n)}_{\text{Infection of an individual at CL } k\text{, not caused by the marked individual}}\end{aligned} \tag{2}$$

for $n\geq 0$ and $(i_1,\ldots,i_M)\in\mathcal{C}$, with $i_j>0$. $1_{\mathcal{A}}$ above is a function equal to 1 if $\mathcal{A}$ is satisfied, and 0 otherwise, and

$$\theta_{\mathbf{i}} \;=\; \delta(\mathbf{i})+\sum_{k=1}^{M}\left(\mu_k i_k+(N_k-i_k)\left(\lambda_k+\sum_{l=1}^{M}\lambda_{lk}i_l\right)\right).$$

We note that Eq. (2) is obtained by following arguments in Eq. (1), and conditioning on the next event that can potentially occur in the process. For example, let us assume that process is at state $\mathbf{i}=(i_1,\ldots,i_M)$ at present time, and we are computing probability $\nu^{(j)}_{\mathbf{i}}(k;n)=\mathbb{P}(R^{(j)}_{\mathbf{i}}(k)=n)$, which relates to the reproduction number $R^{(j)}_{\mathbf{i}}(k)$ for a marked infective individual at compartmental level j, among individuals at compartmental level k. A potential event which can occur is the recovery of an individual –different to the marked one– at compartmental level j, which by the theory of Markov processes occurs with probability $\frac{\mu_j(i_j-1)}{\theta_{(i_1,\ldots,i_M)}}$, moving the process to the new state $\mathbf{i}^-(j)=(i_1,\ldots,i_j-1,\ldots,i_M)$. This leads to the addend $\mu_j(i_j-1)\nu^{(j)}_{\mathbf{i}^-(j)}(k;n)$ in Eq. (2), and similar arguments can be applied for the rest of potential possible events that can occur. Finally, we point out that the system of equations given by Eq. (2) can be represented in matrix form, and solved by starting with $n=0$, and then sequentially solving the system of equations for any value $n\geq 1$ by using previously computed probabilities for $n-1$, in an iterative fashion; see the Supplementary Material.

It is clear that, since

$$R^{(j)}_{(i_1,\ldots,i_M)} \;=\; \sum_{k=1}^{M}R^{(j)}_{(i_1,\ldots,i_M)}(k),$$

we can also focus on computing probabilities

$$\nu^{(j)}_{(i_1,\ldots,i_M)}(n) \quad = \quad \mathbb{P}(R^{(j)}_{(i_1,\ldots,i_M)} = n), \quad n \geq 0,$$

for any initial state $(i_1, \ldots, i_M) \in \mathcal{C}$ with $i_j > 0$. Probabilities $\nu^{(j)}_{(i_1,\ldots,i_M)}(n)$ satisfy

$$\begin{aligned}\theta_{\mathbf{i}}\nu^{(j)}_{\mathbf{i}}(n) = &\sum_{k=1,\ k\neq j}^{M} \mu_k i_k \nu^{(j)}_{\mathbf{i}^-(k)}(n) + \mu_j(i_j-1)\nu^{(j)}_{\mathbf{i}^-(j)}(n) \\ &+ (\mu_j + \delta(\mathbf{i}))1_{n=0} + \sum_{k=1}^{M}(N_k - i_k)\lambda_{jk}\nu^{(j)}_{\mathbf{i}^+(k)}(n-1)1_{n>0} \\ &+ \sum_{k=1}^{M}(N_k - i_k)\left(\lambda_k + \sum_{l=1,\ l\neq j}^{M}\lambda_{lk}i_l + \lambda_{jk}(i_j-1)\right)\nu^{(j)}_{\mathbf{i}^+(k)}(n),\end{aligned} \tag{3}$$

for $n \geq 0$ and for any $(i_1, \ldots, i_M) \in \mathcal{C}$, with $i_j > 0$. This system is expressed in matrix form, and solved in an iterative fashion, in the Supplementary Material.

3 Case studies

In this Section, we focus on five different representative existing models in the literature for the spread of nosocomial infections. Our aim is to show how these models can be seen as particular cases of the unified stochastic modelling framework presented in Section 2, so that the methodology in subsection 2.1 can be directly applied, and the infectiousness of each agent in the hospital ward can appropriately be quantified. In particular, case studies 1-5 can be represented into our framework by specifying the number M of compartmental levels and their meaning, as well as the meaning of the infection and removal events occurring at each compartmental level, and the specifications of rates μ_j, λ_j, λ_{jk} and $\delta(i_1, \ldots, i_M)$. These rates are general enough in Section 2 in order to account for all hypotheses usually considered when modelling nosocomial infections (such as those considered in Refs. [18, 20, 21, 22, 23] related to case studies 1-5), and also allow one to consider different hospital surveillance policies for outbreak detection and declaration [24, 25]. A summary of these rates for each case study studied in this Section can be found in Table S6 in the Supplementary Material.

3.1 Modelling spread among patients and health-care workers

We focus here on the model by Artalejo (2014) [20], for a nosocomial outbreak in a hospital ward with N_p patients and N_{HCW} HCWs. Patients can be colonized or non-colonized at any given time, and are discharged at rate μ, regardless of their colonization status. HCWs can have their hands contaminated or uncontaminated, and they wash their hands at rate μ'. Each colonized patient *contaminates* (the hands of) each uncontaminated HCW at rate β', while each contaminated HCW colonizes each non-colonized patient at rate β. Admission of new patients occurs immediately after discharge, and newly admitted patients can be colonized with probability σ. It is assumed in Ref. [20] that each colonized patient is detected at rate γ, which can be incorporated here by setting $\delta(i_1, i_2) = \gamma i_1$ (*i.e.,* outbreak declaration occurs upon detection of the first colonized patient); see Figure 2.

We note that the outbreak detection and declaration rate $\delta(i_1, \ldots, i_M)$ can be set to account for different hypotheses regarding hospital surveillance and screening. By setting $\delta(i_1, i_2) = \gamma i_1$ as above, one can represent random screening being in place as the surveillance policy in the hospital ward, where each patient is screened at an average time γ^{-1}; see Ref. [24] where this screening policy is identified as one of the most efficient ones for the control of nosocomial outbreaks. We also note that outbreak declaration rate $\delta(i_1, i_2) = \gamma i_1$ can also be used to represent the scenario where outbreak is declared after the first colonized patient showing some symptoms, each colonized patient showing symptoms at rate γ (*e.g.,* norovirus outbreaks are declared upon detection of suspected cases, consisting of patients showing symptoms such as diarrhoea and vomiting). On the other hand, if a colonized patient is admitted into a hospital ward, and detection occurs by screening upon admission where laboratory results take an average time δ^{-1} to arrive, one could represent this by setting $\delta(i_1, i_2) = \delta$ and with time $t = 0$ representing the admission of the colonized patient into the ward.

In Figure 2, we show how this model can be represented into our framework, by setting $M = 2$, $N = N_p + N_{HCW}$, where compartmental level $j = 1$ amounts to colonized/non-colonized patients and $j = 2$ amounts to uncontaminated/contaminated HCWs. In order to incorporate the hypotheses above, rate functions $\lambda_j(i_1, i_2)$, $\mu_j(i_1, i_2)$ and $\delta(i_1, i_2)$ are defined as in Figure 2, and summarised in Table S6 in the Supplementary Material. Moreover, summary statistics analysed in Section 2 have specific meanings in this particular case study, as described in Table 1. We note here that an alternative existing approach in the literature, such as the model in Ref. [3], is to consider only colonized/non-colonized patients explicitly in the model, where the role played by contaminated HCWs is only implicitly incorporated via a transmission rate β. Model in Ref. [3] could be represented into our framework by setting $M = 1$ (colonized/non-colonized patients) and appropriately setting rates $\mu_1(i_1)$, $\lambda_1(i_1)$ and $\delta(i_1)$, which is omitted here for the sake of brevity.

We use here parameter values considered in Ref. [20], for the spread of MRSA in an hypothetical intensive care unit, which are reported in Table S1 in the Supplementary Material. When analysing the infectiousness of colonized patients and contaminated HCWs, we can focus on computing the reproduction number of these individuals, as described in Section 2; see Table 1. While the reproduction number can be computed, for a contaminated HCW ($R^{(2)}_{(0,1)}$), by direct application of Eq. (3), a slight modification needs to be considered when analysing the reproduction number of a colonized patient; that is, when computing probabilities $\nu^{(1)}_{(i_1,i_2)}(n) = \mathbb{P}(R^{(1)}_{(i_1,i_2)} = n)$. In particular, Eq. (3) for

$R^{(1)}_{(1,0)} = R^{(1)}_{(1,0)}(2)$	Reproduction number of a colonized patient starting the outbreak (among HCWs)
$R^{(2)}_{(0,1)} = R^{(2)}_{(0,1)}(1)$	Reproduction number of a contaminated HCW starting the outbreak (among patients)

Table 1: Meaning of our summary statistics for model in Figure 2. Case study 1

model and rate functions in Figure 2 leads to

$$\begin{aligned}\theta_{(i_1,i_2)}\nu^{(1)}_{(i_1,i_2)}(n) &= \mu' i_2 \nu^{(1)}_{(i_1,i_2-1)}(n) + 1_{n=0}((1-\sigma)\mu + \gamma i_1) \\ &+ (i_1-1)\left((1-\sigma)\mu\nu^{(1)}_{(i_1-1,i_2)}(n) + (N_2-i_2)\beta'\nu^{(1)}_{(i_1,i_2+1)}(n)\right) \\ &+ (N_1-i_1)(\sigma\mu+\beta i_2)\nu^{(1)}_{(i_1+1,i_2)}(n) \\ &+ 1_{n>0}(N_2-i_2)\beta'\nu^{(1)}_{(i_1,i_2+1)}(n-1) \qquad (4)\end{aligned}$$

with $\theta_{(i_1,i_2)} = \mu' i_2 + (1-\sigma)\mu(i_1-1) + (N_1-i_1)(\sigma\mu + \beta i_2) + (N_2-i_2)\beta' i_1 + (1-\sigma)\mu + \gamma i_1$. However, we note that $R^{(1)}_{(1,0)}$ should amount to the number of infections (*i.e.,* in this case, HCW hands contaminations) directly caused by a given colonized patient starting the outbreak until this patient is discharged or the outbreak is detected, regardless of the newly admitted patient being or not colonized. This means that terms $1_{n=0}(1-\sigma)\mu$ in Eq. (4) and $(1-\sigma)\mu$ in $\theta_{(i_1,i_2)}$ need to be replaced by $1_{n=0}\mu$ and μ, respectively, and the same applies when analysing the reproduction number of a colonized patient in case studies 2-4.

In Figure 3, we plot the probability mass functions of the reproduction number of a colonized patient ($R^{(1)}_{(1,0)}$) and of a contaminated HCW ($R^{(2)}_{(0,1)}$) starting the outbreak. While the average outbreak declaration time is crucial for limiting the reproduction number of a colonized patient, this is not the case when looking at the reproduction number of a contaminated HCW. This is related to the fact that the main limiting factor for the infectiousness of a HCW is his/her hand-washing rate, which is something that we explore in more depth in the following case studies.

3.2 Considering different HCW types

We focus here on the model by Wang et al. (2011) [22], which incorporates volunteers working at the hospital ward. Authors in Ref. [22] consider the spread of MRSA in the Respiratory Intensive Care Unit (RICU) at Beijing Tongren Hospital, which is formed by N_p patients, N_{HCW} HCWs and N_V volunteers. As assumed in Ref. [22], patients are admitted at rate λ, who can already be colonized upon admission with probability φ, and discharged at rate δ_C (if colonized) or δ_U (if non-colonized). HCW-patient transmission rate $\frac{(1-\eta)}{N_P}\beta_{PH}$ consists of two contributions: the hygienic level $\eta \in (0,1)$ during each HCW-patient contact, which is encoded in a probability $(1-\eta)$ of transmission per contact, and a contact rate β_{PH}, and similar comments apply to volunteer-patient transmission rate $\frac{(1-\xi)}{N_P}\beta_{PV}$; see details in [22, Page 3] and related equations in [22, Appendix]. In Figure 4, we depict how this model is represented into our framework, in the asymptotic situation where immediate arrival of patients is assumed after discharge (*i.e.,* $\lambda \to \infty$), which is a reasonable approximation for hospital wards under high demand [3, 23]. Since no detection is considered in Ref. [22], where the interest is in the long-term dynamics of the nosocomial spread and in analysing the infectiousness of each individual in the ward, we set $\delta(i_1,\ldots,i_M) = 0$.

For parameter values in Table S2 in the Supplementary Material, we plot in Figures 5-6 the mean reproduction numbers of the different agents in this ward, for varying values of model parameters. We compute in Figure 5 the mean reproduction number of a colonized patient starting the outbreak, among HCWs ($E[R^{(1)}_{(1,0,0)}(2)]$) and volunteers ($E[R^{(1)}_{(1,0,0)}(3)]$), versus (δ_C^{-1}, η) and (δ_C^{-1}, ξ), respectively. Our results suggest that transmission from patients to HCWs played a significant role in this outbreak, where a given colonized patient contaminates $E[R^{(1)}_{(1,0,0)}(2)] = 10.05$ HCWs during his/her stay in the ward. On the other hand, our model suggests little transmission from colonized patients to volunteers, with $E[R^{(1)}_{(1,0,0)}(3)] = 0.65$. This remains true even though the low hygienic level during patient-volunteer contacts ($\xi = 0.23$ for volunteers vs $\eta = 0.46$ for HCWs), and seems to be related to the low intensity of these contacts ($\beta_{PV} = 0.2$ for volunteers vs $\beta_{PH} = 0.72$ for HCWs). Stochastic variability of the reproduction numbers $E[R^{(1)}_{(1,0,0)}(2)] = 10.05$ and $E[R^{(1)}_{(1,0,0)}(3)] = 0.65$ can also be assessed by our methodology in Section 2, in terms of standard deviations $SD[R^{(1)}_{(1,0,0)}(2)] = 10.50$ and $SD[R^{(1)}_{(1,0,0)}(3)] = 0.94$. These are readily obtained from the probability distributions computed from Eq. (2).

When looking at possible control strategies, it seems clear that the reproduction number of a colonized patient among HCWs can be significantly reduced by improving the hygienic level of each HCW-patient contact, while reducing the length of stay of each patient does not significantly reduce the infectiousness (*i.e.,* contamination ability) of this patient, and similar comments apply to patient-volunteer contacts.

In Figure 6, the mean reproduction number of a contaminated HCW or volunteer is computed for varying values of the hygienic levels during each contact, as well as of the hand-washing rates. The fact that HCWs wash their hands an average of 24 times per day in this ward keeps the reproduction number of these agents low, and only under significantly low hand-washing compliance levels ($\gamma_H < 5$) a substantial increase for this reproduction number is predicted. Thus, for a particular HCW with low hand-washing compliance level, hygienic level during each HCW-patient contact becomes the most important factor determining the infection spread, and similar comments apply to volunteers.

3.3 Assessing environmental contamination

The important role played by environmental contamination in nosocomial spread has been discussed in recent

works in the field [8, 9], since pathogens such as MRSA and *Vancomycin-resistant Enterococci* (VRE) are able to survive on dry surfaces for weeks [27]. We consider here the model by Wolkewitz et al. (2008) [23], which incorporates contaminated/non-contaminated surfaces. Authors in Ref. [23] consider N_p patients, N_s HCWs and N_e surfaces for analysing an VRE outbreak in the onco-haematological unit at the University Medical Center Freiburg in Germany. Colonized patients are discharged at rate γ', while non-colonized patients are discharged at rate γ. Discharged patients are immediately replaced by newly admitted patients, who can be colonized with probability ϕ. HCWs wash their hands at rate μ, while surfaces are decontaminated at rate κ. Transmission between patients, HCWs and surfaces occur at rates $(\beta_{sp}, \beta_{se}, \beta_{ps}, \beta_{pe}, \beta_{es}, \beta_{ep})$, where s stands for staff (HCWs), p for patients and e for environment (surfaces). In Figure 7, we show how this model can be represented into our framework, with the corresponding definition of the function rates. Since no outbreak detection is considered in Ref. [23], we set $\delta(i_1, i_2, i_3) = 0$.

In Figures 8-10 we compute the mean reproduction number of all the agents (*i.e.,* patients, HCWs and surfaces) in this hospital ward, for parameter values in Table S3 in the Supplementary Material which are the ones considered in Ref. [23] for the VRE outbreak in the onco-haematological unit, and carry out a sensitivity analysis for several model parameters. In particular, we plot in Figure 8 the mean reproduction number of a colonized patient among HCWs and among surfaces, versus the patient-to-HCW (respectively, patient-to-surface) transmission rate β_{ps} (β_{pe}), and the average length of stay γ'^{-1} of any given colonized patient. For the VRE outbreak considered in Ref. [23], an average number of $E[R^{(1)}_{(1,0,0)}(2)] = 9.09$ HCWs and $E[R^{(1)}_{(1,0,0)}(3)] = 96.83$ surfaces are contaminated by a colonized patient during his/her stay in the ward, these results suggesting that environmental contamination might be playing a significant role in the infection spread, as suspected by authors in Ref. [23]. Stochastic variability of these summary statistics can be represented in terms of the standard deviations $SD[R^{(1)}_{(1,0,0)}(2)] = 9.40$ and $SD[R^{(1)}_{(1,0,0)}(3)] = 73.75$, these large quantities suggesting that the corresponding infection processes are highly stochastic. We note that for a colonized patient staying in the ward for an average of 20 *days*, and an environmental cleaning rate of $\kappa = 1$ *time/day*, the same surface can be contaminated several times by this patient during his/her stay. According to results in Figure 8, both reducing the average length of stay of patients, and decreasing contact rates (*i.e.,* avoiding when possible patient-surface contacts, or improving the hygienic level during each patient-HCW contact) can help to reduce these mean reproduction numbers.

Once a HCW is contaminated, his/her infectious potential can be measured by means of his/her mean reproduction number, which is analysed in Figure 9. It seems clear from results in Figure 9 that the hand-washing rate $\mu = 24$ *times/day* allows to keep this mean reproduction number, for a contaminated HCW, low among patients, although it can be still significant (above 1) among surfaces. Results in Figure 9 also suggest that HCWs with significantly low hand-hygiene compliance levels ($\mu < 10$) could lead to reproduction numbers above 1.75 (among patients) and above 30 (among surfaces), so that our results support the fact that a single HCW with relatively low hand-hygiene compliance level could play a significant infectious role by means of contaminating a large amount of surfaces, and colonizing several patients, until he/she washes his/her hands.

In Figure 10, we plot analogous values for a contaminated surface. Although for parameters considered in Ref. [23] the reproduction numbers of any given contaminated surface (among HCWs and patients) are relatively low, given the substantial number of surfaces that can be contaminated by a colonized patient (Figure 8) or a contaminated HCW with a low hand-hygiene compliance level (Figure 9), these numbers should still not be neglected. It seems clear from Figure 10 that decontamination rate $\kappa = 1$ *time/day* can not be considered as optimal during the course of a nosocomial outbreak, since just by increasing this up to $\kappa = 2$ *times/day* a significant reduction in the reproduction number of any contaminated surface could be achieved. This seems to support existing control policies such as the ones recommended within the *national guidelines on the management of outbreaks of norovirus infection in healthcare settings* [26] issued by the National Disease Surveillance Centre in Ireland, which involve cleaning affected areas of the ward twice daily during norovirus outbreaks. Results in Figure 10 also suggest that, if $\kappa = 1$ *time/day* had to be maintained for any reason, then recommendations among HCWs and patients on reducing as much as possible *infectious* contacts with surfaces during an outbreak could still have a significant impact in reducing the infectivity of any given contaminated surface, specially among patients.

3.4 Incorporating space through room configuration of the ward

The model by López-García (2016) [18] incorporates room configuration into the nosocomial infection dynamics, where the main hypothesis is that for some nosocomial pathogens, the transmission rate between patients in the same room would be higher than the transmission rate for patients in different rooms (this might be the case, for example, when considering airborne transmission [10], if patients in the same room are treated by the same common HCW [21], or when considering isolation rooms where specific control protocols are followed [18]). Since the infection dynamics in Ref. [18] are model for an intensive care unit with four rooms, by a simple SIR epidemic model, where no discharge and arrival of patients is considered, we analyse a more realistic scenario here where patients are discharged at rate ν, and immediately replaced by newly admitted patients, who can be colonized with probability p_C. A transmission rate β_{SR} is considered for patients in the same room, while β_{DR} is the transmission rate for patients in different rooms, and HCWs are not explicitly included into the model. A spontaneous colonization rate λ is also considered in Ref. [18], and no outbreak detection and declaration is assumed so that we set $\delta(i_1, i_2, i_3, i_4) = 0$; see Figure 11 for the representation into our framework.

For parameter values considered in Ref. [18], reported in Table S4 in the Supplementary Material, we compute in Figure 12 the reproduction number of a colonized patient starting the outbreak at Room 1 (*left*) and 2 (*right*), versus transmission rates (β_{DR}, β_{SR}). We note that Rooms 3 and 4 are *equivalent* to Room 2, and are thus not analysed. It is interesting to note that for parameter values considered in Ref. [18] the reproduction number of a patient at Room 1 is $E[R^{(1)}_{(1,0,0,0)}] = 1.62$, while it is $E[R^{(2)}_{(0,1,0,0)}] = 1.54$ for a patient at Room 2. Stochastic variability of these summary statistics can be represented in terms of the standard deviations $SD[R^{(1)}_{(1,0,0,0)}] = 1.73$ and $SD[R^{(2)}_{(0,1,0,0)}] = 1.67$. A threshold behavior can be observed in both plots in Figure 12, where reducing the contact rate between patients in the same room does not seem to have a significant effect on the reproduction number of a patient starting the outbreak at Room 2. For this room, it is the transmission rate between different rooms β_{DR} which has a significant impact. This seems to support the idea of implementing patient cohorting as an infection control strategy, where a given HCW treating patients in the same room would avoid, when possible, to treat patients in a different room during the course of a nosocomial outbreak. On the other hand, a parameter threshold can also be observed for a patient starting the outbreak at Room 1, but this threshold depends on a non-linear combination of the values (β_{SR}, β_{DR}). In particular, both reducing the contact rate between patients in the same room and between patients in different rooms can move the value of the reproduction number near or below 1.

3.5 Modelling HCW-patient contact network with different HCW infection risk profiles

Finally, we focus here on the model by Temime et al. (2009) [21], where the potential of some HCWs in a hospital ward to act as super-spreaders during a nosocomial outbreak is assessed. Temime et al. [21] consider an hypothetical hospital ward with three types of HCWs: AP1 (a profile involving frequent contacts with a limited number of patients, typically a nurse), AP2 (a profile involving fewer contacts with more patients, typically a physician), and a *peripatetic* HCW (involving a single daily contact with all patients, for instance a therapist or a radiologist). These different HCW profiles lead to different transmission risks, where AP1-patient contacts can be considered as high risk, AP2-patient contacts have moderate risk, and peripatetic-patient contacts have low risk; see [21, Figure 1]. This is encoded here by considering transmission rates $\beta_{AP1} > \beta_{AP2} > \beta_{Peri}$. Authors in Ref. [21] consider an hypothetical hospital ward with 18 beds, that all HCWs wash their hands at rate μ, and that all patients are discharged at rate γ, being immediately replaced by new non-colonized admitted patients. By means of agent-based stochastic simulations, authors simulate the spread of a nosocomial pathogen (using data for MRSA and VRE) in this ward while incorporating details such as the duration of each HCW-patient contact, the probability of pathogen transmission during a $20min$ HCW-patient contact, or the existence of day/night HCW shifts.

In Figure 13 we represent a simplified version of this model into our framework, for a smaller hospital ward with 8 patients, 4 AP1 HCWs, 2 AP2 HCWs and 1 peripatetic HCW, but when considering the same contact network structure than the one studied in [21, Figure 1]. Transmission rates β_{AP1}, β_{AP2} and β_{Peri} in Table S5 in the Supplementary Material are obtained by taking into account the duration of each HCW-patient contact type, as well as the probability of pathogen transmission during each contact, by using values in [21, Table 1] and following the arguments in [21, Supplementary Material I]. Since no outbreak detection is considered in Ref. [21], we set $\delta(i_1, \ldots, i_{11}) = 0$ and

$$\begin{aligned}
\lambda_j(i_1,\ldots,i_{11}) &= (\beta_{AP1} i_{4+j} + \beta_{AP2} i_9 + \beta_{Peri} i_{11})(N_j - i_j), \\
&\qquad\qquad 1 \le j \le 2, \\
\lambda_j(i_1,\ldots,i_{11}) &= (\beta_{AP1} i_{4+j} + \beta_{AP2} i_{10} + \beta_{Peri} i_{11})(N_j - i_j), \\
&\qquad\qquad 3 \le j \le 4, \\
\lambda_j(i_1,\ldots,i_{11}) &= \beta_{AP1} i_{j-4}(N_j - i_j), \quad 5 \le j \le 8,
\end{aligned}$$

$$\begin{aligned}
\lambda_9(i_1,\ldots,i_{11}) &= \beta_{AP2}(i_1 + i_2)(N_9 - i_9), \\
\lambda_{10}(i_1,\ldots,i_{11}) &= \beta_{AP2}(i_3 + i_4)(N_{10} - i_{10}), \\
\lambda_{11}(i_1,\ldots,i_{11}) &= \beta_{Peri}(i_1 + i_2 + i_3 + i_4)(N_{11} - i_{11}), \\
\mu_j(i_1,\ldots,i_{11}) &= \gamma i_j, \quad 1 \le j \le 4, \\
\mu_j(i_1,\ldots,i_{11}) &= \mu i_j, \quad 5 \le j \le 11.
\end{aligned}$$

Given the complexity of this model, we report in Table 2 the meanings of our summary statistics in Section 2. In Figure 14 we plot the mean reproduction number of a representative colonized patient (*e.g.,* $P_{1,a}$) starting the outbreak, among those HCWs that treat him/her ($AP1_1$, $AP2_1$ and peripatetic). These values are mainly dominated by β_{AP1} and γ^{-1}; that is, by the contact rate for high transmission risk contacts and the length of stay of the patient in the ward. For parameters in Table S5 in the Supplementary Material a colonized patient contaminates around $E[\sum_{j\in\{5,9,11\}} R^{(1)}_{(1,0,\ldots,0)}(j)] = 5.3$ HCWs during his/her stay, with $SD[\sum_{j\in\{5,9,11\}} R^{(1)}_{(1,0,\ldots,0)}(j)] = 5.78$. By analysing values of $E[R^{(1)}_{(1,0,\ldots,0)}(5)]$, $E[R^{(1)}_{(1,0,\ldots,0)}(9)]$ and $E[R^{(1)}_{(1,0,\ldots,0)}(11)]$ separately, one can decipher that this corresponds to $E[R^{(1)}_{(1,0,\ldots,0)}(5)] = 3.42$ contamination events to the $AP1_1$, $E[R^{(1)}_{(1,0,\ldots,0)}(9)] = 1.19$ to the $AP2_1$ and $E[R^{(1)}_{(1,0,\ldots,0)}(11)] = 0.69$ to the peripatetic HCW. However, we note that since $AP1_1$ only treats two patients, while the peripatetic treats eight patients, the peripatetic HCW might have his/her hands contaminated for longer periods during a nosocomial outbreak.

In Figure 15, we plot the mean reproduction number of the $AP1_1$ ($E[R^{(5)}_{(0,0,0,0,1,0,\ldots,0)}(1)]$), the $AP2_1$ ($E[R^{(9)}_{(0,\ldots,0,1,0,0)}(1) + R^{(9)}_{(0,\ldots,0,1,0,0)}(2)]$) and the peripatetic ($E[\sum_{j=1}^{4} R^{(11)}_{(0,\ldots,0,1)}(j)]$) HCW starting the outbreak. Larger values are found for the peripatetic HCW, even though its low transmission risk per contact ($\beta_{Peri} < \beta_{AP2} < \beta_{AP1}$),

$R^{(1)}_{(1,0,\ldots,0)} = R^{(1)}_{(1,0,\ldots,0)}(5) + R^{(1)}_{(1,0,\ldots,0)}(9) + R^{(1)}_{(1,0,\ldots,0)}(11)$	Reproduction number of patient $P_{1,a}$
$R^{(5)}_{(0,0,0,0,1,0,\ldots,0)} = R^{(5)}_{(0,0,0,0,1,0,\ldots,0)}(1)$	Reproduction number of the $AP1_1$ HCW
$R^{(9)}_{(0,\ldots,0,1,0,0)} = R^{(9)}_{(0,\ldots,0,1,0,0)}(1) + R^{(9)}_{(0,\ldots,0,1,0,0)}(2)$	Reproduction number of the $AP2_1$ HCW
$R^{(11)}_{(0,\ldots,0,1)} = \sum_{j=1}^{4} R^{(11)}_{(0,\ldots,0,1)}(j)$	Reproduction number of the *peripatetic* HCW

Table 2: Meaning of our summary statistics for model in Figure 13. Case study 5

which is directly related to the large number of patients this peripatetic HCW treats. Larger mean reproduction numbers found for $AP1_1$ than for $AP2_1$ suggest however that there exists a trade-off between the transmission risk profile of each contact (encoded by rates β_{AP2} and β_{AP1}) and the number of patients that each HCW treats (*i.e.,* the particular contact network within the hospital ward). The potential for the peripatetic HCW to act as a super-spreader can be noticed from a combination of results in Figures 14-15. In particular, we note that the infectious potential of the peripatetic HCW is enhanced by the fact that this HCW might have his/her hands contaminated for long periods, since each of the eight patients treated by this HCW, who might be colonized, contaminates peripatetic HCW hands an average of 0.69 times during their stay. Moreover, it is clear from our results that low hygiene levels during peripatetic-patient contacts (*i.e.*, increasing values of β_{Peri}) might significantly increase the number of patients that this HCW colonizes until washing his/her hands, and results in Figure 15 suggest that the same applies for his/her hand-washing compliance level, which could enhance his/her role as a super-spreader during a nosocomial outbreak.

4 Discussion

In this work we present a unified stochastic modelling framework for the analysis of the spread of nosocomial infections. This unified model allows one to move from more compartment-based models for highly homogeneous scenarios ($M \approx 1$), to agent-based type models when dealing with highly heterogeneous settings ($M \approx N$). We note that when considering the asymptotic case $M = N$, with $N_j = 1$ for all $1 \leq j \leq M$, the resulting space of states $\mathcal{C}$ contains $\prod_{j=1}^{M}(N_j + 1) = 2^N$ states, since in this case one is in fact analysing the SIS epidemic model on a network; see Refs. [17, 18]. Our unified framework allows for considering different hypotheses related to the detection and declaration of the nosocomial outbreak, or to analyse the long-term infection spread when this detection is not relevant. The versatility of this model allows one to consider a wide range of agents involved in the nosocomial outbreak, to account for hand-washing compliance levels, environmental cleaning, patients arrival/discharge, spatial components such as the hospital ward room configuration, different types of HCWs corresponding to different pathogen transmission risks, as well as specific patient-staff contact network topologies.

Our methodology within this unified framework allows for the exact analysis of the probability distribution of the exact reproduction number of each agent in the ward. Moreover, this summary statistic can be split into several ones accounting for the infections caused by a given individual among individuals of a particular type. This translates into analysing the infectiousness of patients, HCWs, volunteers or surfaces among individuals of each of these groups, so that the role played by each potential contact transmission route can be assessed for nosocomial outbreaks corresponding to different health-care facilities and pathogens. To the best of our knowledge, this is the first time that this analytical approach, which has been usually neglected when analysing infection spread among individuals in populations of moderate-to-large sizes –due to computational constraints–, is applied in the area of nosocomial infections where populations are usually small and heterogeneous, making its implementation feasible. We note that, although the focus here has been on studying the reproduction number of each individual, alternative summary statistics of interest allowing for first-step analysis (such as the length or the final size of the outbreak, see Refs. [17, 18]) could be analysed in the same way by means of this unified framework and our methodology in Section 2.

Our unified framework, together with the analytical approach in Section 2, allows one to exactly compute the corresponding reproduction numbers, and to use these to assess the role played by the different routes of infection during a nosocomial outbreak. At the same time, the fact that all scenarios in Section 3 -and potentially others- can be represented into our unified framework, means that computer codes developed for solving Eqs. (2)-(3) for the general model in Figure 1 can be readily applied in all these scenarios, just by specifying the corresponding $\mu_j(i_1, \ldots, i_M)$, $\lambda_j(i_1, \ldots, i_M)$ and $\delta(i_1, \ldots, i_M)$ rates. On the other hand, we acknowledge that this unifed stochastic framework represented by the diagram in Figure 1 entails several simplifying assumptions and limitations. The constant size assumed for each compartmental level means that the total number of agents of each type (patients, HCWs, surfaces, volunteers,...) remains constant during the course of the nosocomial outbreak. When focusing on patients, this is only appropriate under high demand situations, where the time during which any given bed is empty is short enough and can be neglected in the corresponding model. Under moderate demand, and if one needs to incorporate empty beds explicitly in the model, the stochastic process in Section 2 could be modified so that $S_1(t)$ (if $j = 1$ represents the compartmental level corresponding to patients) is incorporated as an additional variable into the continuous-time Markov chain $\mathcal{X}$, so that $S_1(t) + I_1(t)$ is not necessarily constant throughout time. Moreover, more complex situations such as nosocomial outbreaks occurring across several hospital wards, with patient movement between wards, or competitive scenarios where several bacterial strains (*e.g.,* antibiotic-sensitive vs antibiotic-susceptible

[28]) are spreading simultaneously within the same hospital ward, cannot be directly represented into our framework by just specifying rates $\mu_j(i_1,\ldots,i_M)$, $\lambda_j(i_1,\ldots,i_M)$ and $\delta(i_1,\ldots,i_M)$. Instead, alternative diagramatic representations to that in Figure 1 should be explored, potentially including movement of agents between different compartmental levels.

We also note that our methodology directly depends on the fact that the model proposed is a continuous-time Markov chain, so that events are Markovian and inter-event times are assumed to be exponentially distributed. While this is a typical assumption in the literature when analysing nosocomial outbreaks from a stochastic perspective, we acknowledge that the exponential distribution might not be appropriate for some particular events in these processes, such as patients' lengths of stay. Although relaxing the Markovian assumption in these models is out of the scope of this paper, it is worth to point out here that some attempts have already been made in this area, some of them based on the use of phase-type distributions for incorporating these non-Markovian events [29, 30].

Finally, we acknowledge here that additional limitations of our approach are of computational nature, related to solving systems of around $\#\mathcal{C}=\prod_{k=1}^{M}(N_k+1)$ linear equations. However, populations usually involved in nosocomial outbreaks are small enough for this methodology to be efficiently implemented, where specific procedures for dealing with systems of equations involving highly sparse matrices can be specially useful. We also note that while $N=20+5+100=125$ individuals in case study 3 (patients, HCWs and surfaces) lead to analysing a stochastic process with $\#\mathcal{C}=12726$ states, only $N=2+2+2+2+1+1+1+1+1+1+1=15$ individuals in case study 5 (patients, AP1, AP2 and peripatetic HCWs) lead to $\#\mathcal{C}=10368$ states, which is directly related to the high level of individual heterogeneity introduced into this model (encoded by the number of compartmental levels $M=3$ versus $M=11$). These comments suggest that while agent-based simulation approaches should prevail under highly heterogeneous scenarios, such as the complete model by Temime et al. (2009) [21], more homogeneous or low-to-moderate heterogeneous settings allow for this exact approach to be implemented.

Acknowledgments

We thank the editor and three anonymous referees for their constructive feedback, which helped to improve the manuscript.

Software and reproducibility

Computer codes (in *Python*) to reproduce our results in Section 3 are available in an online repository [31].

Author contributions statement

MLG and TK conceived the idea. MLG developed the analysis in Section 2, and the computer codes used in Section 3. Both authors designed the five case studies, reviewed the literature, and contributed to writing and revising the manuscript.

Funding

M. López-García would like to acknowledge the support of the Medical Research Council, through a Skills Development Fellowship (reference number MR/N014855/1). This work was also supported by the Spanish Ministry of Economy, Industry and Competitiveness (MTM2014-58091-P).

Competing interests

The authors declare no competing interests.

References

[1] Harbarth S, Sax H, Gastmeier P (2013) *The preventable proportion of nosocomial infections: an overview of published reports.* J Hosp Infect, 54: 258-66.

[2] van Kleef E, Robotham JV, Jit M, Deeny SR, Edmunds WJ (2013) *Modelling the transmission of healthcare associated infections: a systematic review.* BMC Infect Dis, 13: 294.

[3] Pelupessy I, Bonten MJ, Diekmann O (2002) *How to assess the relative importance of different colonization routes of pathogens within hospital settings.* P Natl Acad Sci, 99: 5601-5.

[4] Taubes G (2008) *The bacteria fight back.* Science, 321: 356-61.

[5] Chamchod F, Ruan S (2012) *Modeling the spread of Methicillin-resistant Staphylococcus aureus in nursing homes for eldery.* PLoS ONE, 7: e29757 (9 pages).

[6] D'Agata EMC, Webb GF, Horn MA, Moellering RC, Ruan S (2009) *Modeling the invasion of community-acquired Methicillin-resistant Staphylococcus aureus into hospitals.* Clin Infect Dis, 48: 274-84.

[7] Sax H, Allegranzi B, Chrati M-N, Boyce J, Larson E, Pittet D (2009) *The World Health Organization hand hygiene observation method.* Am J Infect Control, 37: 827-34.

[8] Doan TN, Kong DCM, Kirkpatrick CMJ, McBryde ES (2014) *Optimizing hospital infection control: the role of mathematical modelling.* Infect Cont Hosp Ep, 35: 1521-30.

[9] King MF, Noakes CJ, Sleigh PA (2015) *Modeling environmental contamination in hospital single- and four-bed rooms.* Indoor Air, 25: 694-707.

[10] Noakes CJ, Sleigh PA (2009) *Mathematical models for assessing the role of airflow on the risk of airborne infection in hospitals.* J R Soc Interface, rsif20090305.

[11] Bonten MJM (2002) *Infection in the intensive care unit: prevention strategies*. Curr Opin Infect Dis, 15: 401-5.

[12] Sébille V, Chevret S, Valleron AJ (1997) *Modeling the spread of resistant nosocomial pathogens in an intensive-care unit*. Infect Cont Hosp Ep, 18: 84-92.

[13] Grundmann H, Hellriegel B (2006) *Mathematical modelling: a tool for hospital infection control*. Lancet Infect Dis, 6: 39-45.

[14] Assab R, Nekkab N, Crépey P, Astagneau P, Guillemot D, Opatowski L, Temime L (2017) Mathematical models of infection transmission in healthcare settings: recent advances from the use of network structured data. Current Opinion in Infectious Diseases, 30(4), 410-418.

[15] Heffernan JM, Smith RJ, Wahl LM (2005) *Perspectives on the basic reproductive ratio*. J R Soc Interface, 2: 281-93.

[16] Keeling MJ, Ross JV (2008) *On methods for studying stochastic disease dynamics*. J R Soc Interface, 5:171-81.

[17] Economou A, Gómez-Corral A, López-García M (2015) *A stochastic SIS epidemic model with heterogeneous contacts*. Physica A, 421: 78-97.

[18] López-García M (2016) *Stochastic descriptors in an SIR epidemic model for heterogeneous individuals in small networks*. Math Biosci, 271: 42-61.

[19] Ross JV (2011) *Invasion of infectious diseases in finite homogeneous populations*. J Theor Biol, 289: 83-89.

[20] Artalejo, JR (2014) *On the Markovian approach for modeling the dynamics of nosocomial infections*. Acta Biotheor, 62: 15-34.

[21] Temime L, Opatowski L, Pannet Y, Brun-Buisson C, Boëlle PY, Guillemot D (2009) *Peripatetic health-care workers as potential superspreaders*. P Natl Acad Sci, 106: 18420-25.

[22] Wang J, Wang L, Magal P, Wang Y, Zhuo J, Lu X, Ruan S (2011) *Modelling the transmission dynamics of meticillin-resistant Staphylococcus aureus in Beijing Tongren hospital*. J Hosp Infect, 79: 302-8.

[23] Wolkewitz M, Dettenkofer M, Bertz H, Schumacher M, Huebner J (2008) *Environmental contamination as an important route for the transmission of the hospital pathogen VRE: modelling and prediction of classical interventions*. Infect Dis: Research and Treat, 1.

[24] Robotham JV, Jenkins DR, Medley GF (2007) *Screening strategies in surveillance and control of methicillin-resistant Staphylococcus aureus (MRSA)*. Epidemiol Infect, 135: 328-42.

[25] Trapman P, Bootsma MC (2009) *A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection*. Math Biosci, 219: 15-22.

[26] National Disease Surveillance Centre (2003) *National guidelines on the management of outbreaks of norovirus infection in health-care settings*. ISBN: 0-9540177-4-9.

[27] Boyce JM (2007) *Environmental contamination makes an important contribution to hospital infection*. J Hosp Infect, 65: 50-54.

[28] Gómez-Corral A, López-García M (2018) *Perturbation analysis in finite LD-QBD processes and applications to epidemic models*. Numer Linear Algebra Appl, e2160.

[29] Marshall AH, McClean SI (2003) *Conditional phase-type distributions for modelling patient length of stay in hospital*. International Transactions in Operational Research, 10.6: 565-576.

[30] Fackrell M (2009) *Modelling healthcare systems with phase-type distributions*. Health care management science, 12(1): 11.

[31] Computer codes are being uploaded to the University of Leeds repository, where they will be allocated a DOI. The corresponding reference will be included here during the editorial publication process

Figure captions

Figure 1: Diagram representing the epidemic dynamics among M different compartmental levels.

Figure 2: Model by Artalejo (2014) [20] and its corresponding representation in our framework. Our representation leads to the same stochastic process to that in Ref. [20]. Case study 1

Figure 3: Probability mass functions of the reproduction number of a colonized patient ($R^{(1)}_{(1,0)}$, *left*) and of a contaminated HCW ($R^{(2)}_{(0,1)}$, *right*) starting the outbreak. Average detection time of each patient $\gamma^{-1} \in \{1, 2, 3, 4\}$ *days*. Case study 1

Figure 4: Model by Wang et al. (2011) [22] and its corresponding representation in our framework. Our representation leads to the same stochastic process to that in Ref. [22], when $\lambda \to \infty$. Case study 2

Figure 5: Mean reproduction number of a colonized patient starting the outbreak, among HCWs ($E[R^{(1)}_{(1,0,0)}(2)]$, *left*) and volunteers ($E[R^{(1)}_{(1,0,0)}(3)]$, *right*), versus δ_C^{-1}, η and ξ. Blue dot corresponds to parameter values $(\eta, \xi, \delta_C^{-1}) = (0.46, 0.23, 13.0)$ in Table S2 in Supplementary Material, leading to values $E[R^{(1)}_{(1,0,0)}(2)] = 10.05$ and $E[R^{(1)}_{(1,0,0)}(3)] = 0.65$. Case study 2

Figure 6: Mean reproduction number of a HCW ($E[R^{(2)}_{(0,1,0)}]$, *left*) and a volunteer ($E[R^{(3)}_{(0,0,1)}]$, *right*), versus γ_H, η, γ_V, and ξ. Blue dot corresponds to parameter values $(\gamma_H, \eta, \gamma_V, \xi) = (24.0, 0.46, 12.0, 0.23)$ in Table S2 in Supplementary Material, leading to values $E[R^{(2)}_{(0,1,0)}] = 0.02$ and $E[R^{(3)}_{(0,0,1)}] = 0.01$. Case study 2

Figure 7: Model by Wolkewitz et al. (2008) [23] and its corresponding representation in our framework. Our representation leads to the same stochastic process to that in Ref. [23]. Case study 3

Figure 8: Mean reproduction number of a colonized patient among HCWs ($E[R^{(1)}_{(1,0,0)}(2)]$, *left*) and among surfaces ($E[R^{(1)}_{(1,0,0)}(3)]$, *right*), versus γ'^{-1}, β_{ps} and β_{pe}. Blue dot corresponds to parameter values $(\beta_{ps}, \beta_{pe}, \gamma'^{-1}) = (2.0, 2.0, 20.0)$ in Table S3 in Supplementary Material, leading to values $E[R^{(1)}_{(1,0,0)}(2)] = 9.09$ and $E[R^{(1)}_{(1,0,0)}(3)] = 96.83$. Case study 3

Figure 9: Mean reproduction number of a HCW among patients ($E[R^{(2)}_{(0,1,0)}(1)]$, *left*) and among surfaces ($E[R^{(2)}_{(0,1,0)}(3)]$, *right*), versus μ, β_{se} and β_{sp}. Blue dot corresponds to parameter values $(\beta_{sp}, \beta_{se}, \mu) = (0.3, 2.0, 24.0)$ in Table S3 in Supplementary Material, leading to values $E[R^{(2)}_{(0,1,0)}(1)] = 0.05$ and $E[R^{(2)}_{(0,1,0)}(3)] = 1.64$. Case study 3

Figure 10: Mean reproduction number of a surface among patients ($E[R^{(3)}_{(0,0,1)}(1)]$, *left*) and among HCWs ($E[R^{(3)}_{(0,0,1)}(2)]$, *right*), versus κ, β_{es} and β_{ep}. Blue dot corresponds to parameter values $(\beta_{es}, \beta_{ep}, \kappa) = (2.0, 0.3, 1.0)$ in Table S3 in Supplementary Material, leading to values $E[R^{(3)}_{(0,0,1)}(1)] = 0.06$ and $E[R^{(3)}_{(0,0,1)}(2)] = 0.10$. Case study 3

Figure 11: Hospital ward room configuration from López-García (2016) [18] and its representation in our framework. Our representation leads to an arguably more realistic stochastic process to that in Ref. [18], where patients arrival and discharge are incorporated. Case study 4

Figure 12: Mean reproduction number of a colonized patient at Room 1 ($E[R^{(1)}_{(1,0,0,0)}]$, *left*) and at Room 2 ($E[R^{(2)}_{(0,1,0,0)}]$, *right*) starting the outbreak, versus (β_{SR}, β_{DR}). Blue dot corresponds to parameter values $(\beta_{SR}, \beta_{DR}) = (0.0366, 0.0238)$ in Table S4 in Supplementary Material, leading to values $E[R^{(1)}_{(1,0,0,0)}] = 1.62$ and $E[R^{(2)}_{(0,1,0,0)}] = 1.54$. Case study 4

Figure 13: Staff-patient contact network from Temime et al. (2009) [21] and representation in our framework. Our representation leads to a simplified version of the stochastic process in Ref. [21], for a reduced version of the hospital ward represented in Ref. [21, Figure 1]. Case study 5

Figure 14: Mean reproduction number of patient P_{1a} among all HCWs treating him/her ($E[\sum_{j\in\{5,9,11\}} R^{(1)}_{(1,0,\ldots,0)}(j)]$), versus γ^{-1}, β_{AP1}, β_{AP2} and β_{Peri}, for $\mu = 24$ *times/day*. Blue dot corresponds to parameter values $(\beta_{AP1}, \beta_{AP2}, \beta_{Peri}) = (0.35, 0.12, 0.07)$ in Table S5 in Supplementary Material, leading to value $E[\sum_{j\in\{5,9,11\}} R^{(1)}_{(1,0,\ldots,0)}(j)] = 5.3$. Case study 5

Figure 15: Mean reproduction number of an AP1 ($E[R^{(5)}_{(0,0,0,0,1,0,\ldots,0)}(1)]$, *left*), an AP2 ($E[R^{(9)}_{(0,\ldots,0,1,0,0)}(1) + R^{(9)}_{(0,\ldots,0,1,0,0)}(2)]$, *middle*) and the peripatetic ($E[\sum_{j=1}^{4} R^{(11)}_{(0,\ldots,0,1)}(j)]$, *right*) HCW starting the outbreak, among the patients that they treat, versus μ, β_{AP1}, β_{AP2} and β_{Peri}. Blue line corresponds to parameter values explored in Table S5 in Supplementary Material. Case study 5