

# CTCmodeler: an agent-based framework to simulate pathogen transmission along an inter-individual contact network in a hospital

Audrey Duval <sup>1</sup>[\[ORCID\]](mailto:0000-0002-1376-4101), David Smith<sup>1</sup>, Didier Guillemot<sup>1</sup>, Lulla Opatsowski<sup>1</sup> and Laura Temime<sup>2,3</sup>

<sup>1</sup> Equipe PheMI, unité B2PHI, Inserm, Université de Versailles Saint Quentin, Institut Pasteur, 25 rue du docteur roux 75015 Paris, France

<sup>2</sup> Laboratoire MESuRS, Conservatoire national des Arts et Métiers, 292 rue Saint Martin 75003 Paris, France

<sup>3</sup> Institut Pasteur, Cnam, unité PACRI, 25 rue du docteur roux 75015 Paris, France  
audrey.duval@ens.uvsq.fr

**Abstract.** Over the last decade, computational modeling has proved a useful tool to simulate the transmission dynamics of nosocomial pathogens and can be used to predict optimal control measures in healthcare settings. Nosocomial infections are a major public health issue especially since the worldwide increase of antimicrobial resistance worldwide. Here, we present CTCmodeler, a framework that incorporates an agent-based model to simulate pathogen transmission through inter-individual contact in a hospital setting. CTCmodeler uses real admission, swab and contact data to deduce its own parameters, simulate inter-individual pathogen transmission across hospital wards and produce weekly incidence estimates. Most previous hospital models have not accounted for individual heterogeneity of contact patterns. By contrast, CTCmodeler explicitly captures temporal heterogeneous individual contact dynamics by modelling close-proximity interactions over time. Here, we illustrate the use of CTCmodeler to simulate methicillin-resistant *Staphylococcus aureus* dissemination in a French long-term care hospital, using longitudinal data on sensor-recorded contacts and weekly swabs from the i-Bird study.

**Keywords:** agent based model, hospital, multiresistant bacteria.

## 1 Introduction

Healthcare-associated infections are an important public health threat in the hospital setting, in particular as a result of increasing selection for bacterial pathogens that are multi-resistant to antibiotics [20, 21]. Effective control strategies are needed to limit the global expansion of multi-resistant pathogens, yet better understanding of the transmission dynamics of nosocomial pathogens is necessary to design and implement optimal strategies.

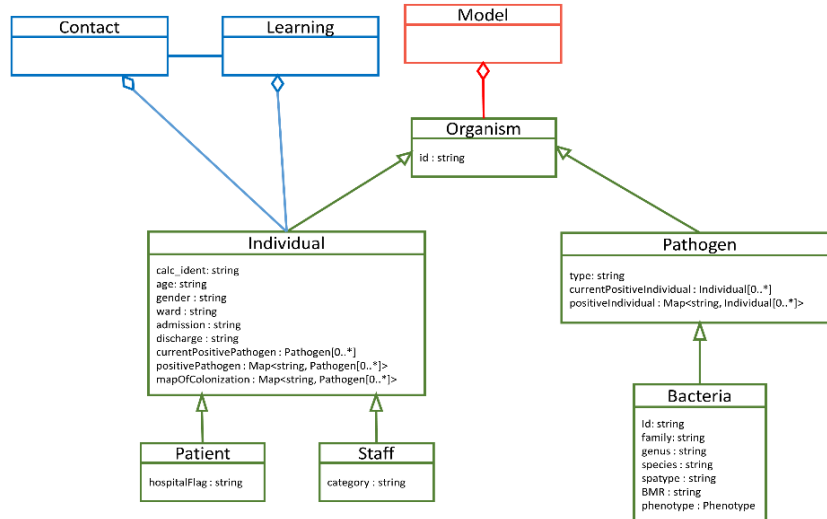
Mathematical models are useful tools to understand and predict the dynamics of pathogen spread in hospitals [1, 8, 10]. Such models have two main objectives: to describe the epidemiology of pathogen transmission and to assess the impact of control strategies such as hand hygiene, on endemic situations at the hospital scale [4, 12, 16, 17, 19]. The majority of published epidemiological mathematical models are so-called “compartmental” models, which group individuals into compartments according to their epidemic status (*e.g.* infectious, susceptible or recovered). However, because hospital settings typically host small populations with highly heterogeneous characteristics and behaviors, agent-based models (ABMs) may be more realistic, allowing for stochastic simulation of dynamics at the individual level [2, 3, 6, 9, 11, 13–15, 18].

In this paper, we present a new stochastic ABM set within a framework called CTCmodeler (Clinical Transmission and Contact modeler), which simulates the dissemination of pathogens at the hospital scale. The originality of our model is twofold. First, it explicitly accounts for networks of inter-individual within-hospital contacts. Second, it includes a module that computes model parameters from observed data. Here, we describe the ABM and illustrate the output of CTCmodeler using the example of methicillin-resistant *Staphylococcus aureus* (MRSA) spread in a French long-term stay hospital. For this illustration, the model is informed by detailed data on contacts and MRSA colonization collected over 6 months during the i-Bird study.

## 2 Methods

### 2.1 General framework description

CTCmodeler consists of three modules: learning, modeling and contact simulation. The learning module estimates parameters from observed data, including pathogen transmission rates, durations of colonization, probabilities of colonization at admission and daily swab frequencies. The modeling module is the ABM, detailed further below. Finally, the contact module simulates contact networks for future predictions. To simulate realistic contacts, this last module computes hourly probabilities of contact between different individual categories from observed contact data provided by the user. The whole framework was programmed in C++ with the repast HPC library 2.2.0 and is illustrated in a class diagram in Figure 1.



**Fig. 1.** Simplified class diagram for CTCmodeler. Patient and Staff objects inherit from Individual, which inherits all its common variables from Organism. Bacteria inherits from Pathogen, which also inherits from Organism.

## 2.2 Overview of the ABM modeling module

The modeling module of CTCmodeler has two options. The first option runs the ABM using user-provided data to reproduce and visualize real output. The second option simulates the transmission of a pathogen along the contact network using three sources of input described in further detail below: (i) input parameters that may be estimated from the learning module or directly provided as a csv file by the user, (ii) a contact file that may either be user-provided (based on observed data) or generated by the contact simulation module, and (iii) a patient admission file that can either be user-provided or obtained from the learning module based on real admission data. In the following, we describe the ABM following the ODD protocol of Grimm et al. [7]. R software (version 3.4.2) was used for graphics.

**Purpose.** The CTCmodeler ABM module simulates nosocomial pathogen transmission between humans, using admission, contact and swab data from any hospital. The main goals are to trace the possible routes of transmission at the individual scale and to describe interaction dynamics between individuals. In the future, potential control measures can be implemented in the ABM.

**Entities, State Variables and Scales.** Figure 1 summarizes all agent variables in a simplified class diagram. The ABM has a main abstract class, the Organism agent, which is split into the Individual and Pathogen classes. Individual class has two children, Patient and Staff classes, which inherit some common variables from Individual,

while Pathogen class is composed of a Bacteria class only. Each class has its own set of variables. The unique id variable characterizes each object during the simulation. Common variables for individuals are hospital anonymous number, age, gender, status (patient or staff), admission date, discharge date, allocated ward, a map of colonization, current positive pathogens and positive pathogens. Current positive pathogens lists all pathogens carried by the individual at a given point in time. Positive pathogens is a container with dates of swabs undergone by the individual and a list of pathogens carried at each swab date. The map of colonization contains decolonization dates for each pathogen carried by the individual. Patient class also includes a hospital flag variable, which describes the reason for hospitalization, while the staff class also includes a category variable, which describes that individual's occupation (nurse, hospital porter, physician, etc.).

Common variables for Pathogen class are the type (bacteria), positive individuals and current positive individuals. Current positive individuals lists all individuals who carry the pathogen at a given point in time. Positive individuals is a container with the last individual date of carriage and the corresponding individual. Bacteria class-specific variables are id, family, genus, species, spatype, MRB (type of multiresistance) and phenotypes. The learning module of CTCmodeler computes environmental parameters such as the transmission rate, the colonization at admission rate, the colonization duration and the patient and staff daily probabilities of getting swabbed.

**Scales.** ABM timescales are set by the user, including simulation dates and time step size. At each time step, the model builds a contact between any two individuals. The computer code for CTCmodeler is fully parallelized, with separate processes for each hospital ward. Hence, a distinct computer process is associated with each ward. For example, if the hospital has three wards, the program will have three processes in which individuals will be scattered. At the beginning of the model, corresponding ward-processes receive newly admitted individuals. The simulator also adds pathogens carried by the newly added individual's process if she/he acquires new ones. At each time step, an individual has contacts with another individual through a network space based on the contact file.

**Process Overview and Scheduling.** Simulations are run on three time levels: weeks, days and time steps.

*Weeks.* Each week, the simulation chooses which individuals to swab. Selected individuals only have a single swab day. This is based on the learning module of CTCmodeler, which computes the mean and standard deviation from the swab distribution of each swab day from provided data. On each simulation day, the model chooses a set of individuals thanks to a normal distribution deduced from mean and standard deviation of the corresponding day of the week.

*Days.* The model performs four actions on each day: add, remove, swab and decolonize individuals. Individuals are added or removed based on daily admissions and discharges

listed in the admission file. Readmissions are possible during the simulation for both patients and staff. When individuals leave the simulation, they go into a transitory container to keep track of their information (e.g. colonization status). The model simulates swabs for each individual listed among a set determined weekly. Finally, colonization is cleared (“decolonization”) for each pathogen carried by a given individual if the current simulation date matches the patient’s decolonization date (stored in the map of colonization for all pathogens stored inside current positive pathogens).

*Time step.* At each time step, the model simulates contacts and possible pathogen transmission events between individuals. Users choose the time step that corresponds to the contact frequency recorded in the contact file. The model builds an edge if two individuals are in contact. Colonized individuals can transmit all the pathogens they carry (i.e. all pathogens listed in current positive pathogen) with the transmission rate either (i) previously determined by the learning module of the framework or (ii) user-defined. The transmission rate is different for each couple (patient-to-patient, staff-to-patient, patient-to-staff and staff-to-staff). When a contact results in a new colonization, the duration of this colonization is randomly drawn from a Gamma distribution. The shape and scale parameters of this distribution are deduced from the mean and standard deviation calculated from real data or user-defined in the learning module of the framework.

### 2.3 Design concept

**Interaction.** Two types of interactions occur during the simulation.

*Inter-individual contacts.* The ABM uses the contact file to link individuals during the simulation. At each time step, the model reads the file and adds an edge to the network.

*Inter-individual transmission.* Colonized individuals can transmit their pathogens to a susceptible individual through the contact network. The transmission rate may either be user-defined or calculated from the learning module of CTCmodeler. In the latter case, it is computed from the observed data as,

$$T_{c1-c2} = \sum_{s=1}^S \frac{A_s}{\sum_{n=1}^N C_{c1-c2,s-n}} \quad (1)$$

Where  $T_{c1-c2}$  represents the transmission rate between an individual from category 1 (e.g. patient) and another individual from category 2.  $S$  is the total number of weeks,  $A_s$  is the number of acquisitions observed during week  $s$  and  $C_{c1-c2,s-n}$  is the cumulative duration of contacts between colonized individuals from category 1 and non-colonized individuals from category 2 during weeks  $s-n$ .  $N$  is the number of previous weeks that are taken into account.

For each interaction, the transmission rate is multiplied by the duration of the contact. If desired, the model can incorporate transmission rate saturation for longer contacts. This option (named duration threshold) is done by using  $1+\log(\sum_{n=1}^N C_{c1-c2,s-n})$  as the denominator in  $T_{c1-c2}$ .

**Stochasticity.** Colonization duration, transmission rate, swab day of week probabilities and colonization at admission are randomly drawn from probability distributions during the simulation. Moreover, admission and discharge rates, as well as contacts, are randomly simulated when the user does not choose to read them directly from observed data.

**Observations.** The model predicts weekly incidence of pathogen acquisition, calculated as the number of new acquisitions at week  $s$  divided by the number of negatively swabbed individuals during week  $s-1$ . To account for possible lack of swab sensitivity, the user may also choose to define a new acquisition as a positive swab following two negative swabs. In this case, the number of new acquisitions at week  $s$  is divided by the number of negatively swabbed individuals during both weeks  $s-1$  and  $s-2$ .

## 2.4 Details

**Initialization.** Before the first time step, the initial number of colonized individuals is determined. When a swab file is available, the model observes the number of colonized swabbed individuals at time 0 and estimates the number of colonized non-swabbed individuals based on proportions of colonized swabbed patients and staff. When no swab data is available, the number of colonized individuals is computed from a user-defined number. Initially colonized individuals of each category are chosen at random.

**Input.** The three input files needed to run the ABM (admission, contact and parameter files) can either be user-provided or computed by the framework from the learning and contact modules. The admission file lists dates of hospital arrival and departure for all individuals included in a simulation. When not user-provided, admission files are randomly simulated based on daily admission and discharge rates for each category of individuals. The contact file lists all unsymmetrized contacts that occur between patients and staff over time during simulations. When not user-provided, contact files are randomly simulated using contact probabilities between individual categories. These contact probabilities may change depending on the hour of the day and on the type of day (weekday vs. weekend). The parameter file gathers all parameter values that the ABM needs. When not user-provided, these parameter values are computed from observed contact, admission and swab data.

## 3 Results

### 3.1 Data

Data from the i-Bird study of nosocomial pathogen transmission were used in this illustrative example of CTCmodeler. This study took place at Berck-sur-Mer in a French long-term care facility (LTCF) from the beginning of May to the end of October 2009, with the first two months as a pilot phase.

During July to October, 95% of patients and hospital staff carried an RFID sensor. Sensors recorded all close proximity (<1.5 meters) interactions occurring in the 5 wards of the hospital every 30 seconds over this 4-month period. The 5 wards were specialized in neurology rehabilitation, geriatric rehabilitation, nutrition and post-operation rehabilitation. The staff population was heterogeneous and included nurses, auxiliary nurses, hospital porters, physicians, administration and animation staff, etc. This diversity entailed idiosyncratic contact dynamics, with many patient-to-patient, patient-to-physician and patient-to-hospital porters' contacts, as described in an earlier paper [5].

Moreover, over the entire study period, dedicated nurses swabbed patients and hospital staff weekly for MRSA colonization.

For this illustration, we used the iBird data to simulate MRSA spread in the hospital from July 6, 2009 to the September 28, 2009. During this period, sensors recorded 1,147,005 contacts and dedicated nurses swabbed 505 individuals (264 of whom were found to carry MRSA at some point during the study).

### 3.2 Input and parameters

CTCmodeler was run using the i-Bird user-contact data file as the contact file to create the contact network. Individual admissions and discharges reproduced the observed data stored in the i-Bird user-admission file. The learning module calculated transmission rates, means and standard deviations of the colonization duration distribution and daily swab probabilities, based on i-Bird user-admission and i-Bird user-swab files. Finally, for these simple simulations, admission and contact file were user-provided and parameters file was deduced from the learning module. Table 1 summarizes all rates, probabilities and parameters deduced by the learning part (included in the parameters file) and used by the modeling part in this illustrative example. For this example, we assumed that an acquisition episode occurred when the individual had two negative swabs followed by a positive one. We also chose to focus on first acquisitions only (without possible recolonization). Second, the ABM used the computed patient-to-patient, patient-to-staff, staff-to-patient and staff-to-staff transmission rates to simulate pathogen transmission. We used a threshold to saturate patient-to-staff and staff-to-staff transmission rates with time. The i-Bird data is composed of 5 wards and hence 5 processes were used for simulations. Thanks to the Repast HPC library, processes were run in parallel and computationally efficient (approximate runtime of 10 minutes).

### 3.3 Output

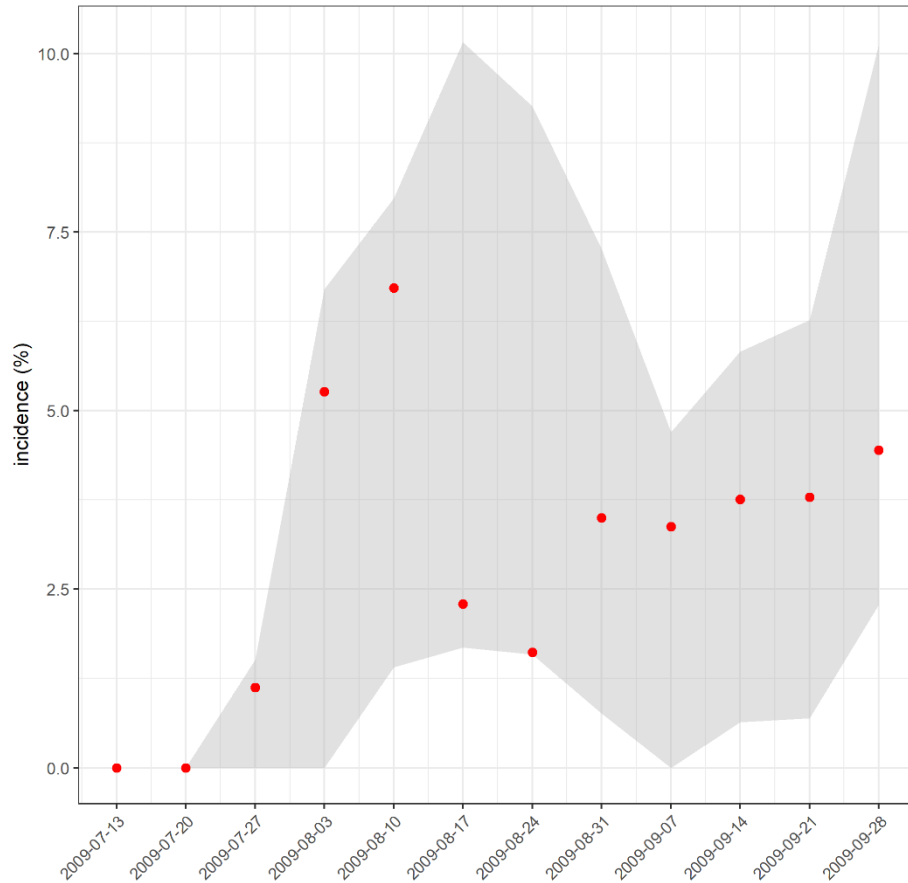
Simulations started with 387 individuals, 151 patients and 236 hospital staff. Among them, 84 swabbed individuals (40 patients and 44 staff) were colonized with MRSA on the first day. Figure 2 depicts the predicted incidence of MRSA colonization over time within the entire hospital, compared with observed data collected during the i-Bird study. We computed 95% prediction bands, i.e. bands that cover 95% of model output

obtained over 500 stochastic model simulation. These 95% prediction bands include all the observed data points (red points), thus validating the model predictions.

**Table 1.** Model parameters

<b>Variable</b>	<b>Value</b>
Transmission rates	
Patient-to-patient	0.000009
Patient-to-staff	0.133615
Staff-to-patient	0.000509
Staff-to-staff	0.136793
Colonized at admission rates	
Patient	0.21
Staff	0.33
Colonization duration (days)	
Patient mean (sd)	32 (771)
Staff mean (sd)	27 (561)
Patient daily swab (%)	
Monday mean (sd)	60.7 (15.1)
Tuesday mean (sd)	34.9 (10.7)
Wednesday mean (sd)	14.5 (10.4)
Thursday mean (sd)	4.1 (2.8)
Staff daily swab (%)	
Monday mean (sd)	24.2 (10.1)
Tuesday mean (sd)	12.0 (4.1)
Wednesday mean (sd)	7.4 (4.3)
Thursday mean (sd)	10.2 (2.1)





**Fig. 2.** Weekly incidence of MRSA colonization at the hospital scale. Red points represent the observed incidence in the i-Bird data. Grey bands correspond to the 95% prediction bands obtained with 500 ABM simulations. Incidence is in percentage of individuals found negative over the two previous weeks. Consequently, the first two data points are at 0.

## 4 Discussion

In this paper, we introduce CTCmodeler, a framework that simulates individual interactions and pathogen transmission in a hospital setting using an agent-based modeling approach. The main originality of this framework is to explicitly account for the inter-individual contact network within the hospital. This allows for tracing of transmission routes and makes the model realistic in terms of interaction mixing, as compared to most previously published models of nosocomial pathogen spread. Multiple options make the model flexible, such as the option for users to enter their own parameter values

to input real data, allowing CTCmodeler to calculate relevant hospital parameters and run corresponding simulations.

The MRSA example we present illustrates how the model can be used to simulate the dissemination of a resistant bacteria through inter-individual contacts in a 5-ward long-term care facility. The framework calculated all needed parameters from real data and then simulated acquisition incidence over a 4-month period. The results obtained show that model predictions are consistent with the observed data. Indeed, the first two points are at zero because of the definition of acquisition chosen. The 500-simulations 95% prediction bands include all incidence points showing that the model can predict MRSA acquisition.

The current version of CTCmodeler has several limitations.

The first limitation is the assumption that pathogens spread exclusively through inter-individual contacts. This allows simulating the contact-mediated dissemination of pathogens such as MRSA or influenza, but would not work for pathogens spread through the environment for instance. Future extensions of the ABM should include the possibility of environmental contamination.

In addition, the model may be used to simulate multiple pathogens simultaneously, but without allowing for interactions between these pathogens (e.g. exchange of a resistance plasmid between two bacteria).

Lastly, although its ability to acquire parameters directly from observed data is among our model's strengths, it also makes it dependent on the precision of these data. When the ABM is used with parameters directly computed from real data files, any incorrect data entry will automatically affect simulations. Moreover, observed "contact" data often reflects close interactions, rather than actual skin-to-skin contacts, as was the case in the i-Bird study. In that case, these "contacts" may not necessarily facilitate pathogen transmission.

In conclusion, CTCmodeler has the potential to be a useful tool to predict pathogen transmission dynamics in a realistic, dynamic hospital setting. Future versions of the framework will include a wide array of control measures, such as hand hygiene, staff cohorting or antibiotic exposure and stewardship, allowing CTCmodeler to be used for public health decision making.

## References

1. Assab, R. et al.: Mathematical models of infection transmission in healthcare settings: recent advances from the use of network structured data. *Curr. Opin. Infect. Dis.* 30, 4, 410–418 (2017).
2. Barnes, S.L. et al.: The Impact of Reducing Antibiotics on the Transmission of Multidrug-Resistant Organisms. *Infect Control Hosp Epidemiol.* 38, 6, 663–669 (2017).
3. Caudill, L., Lawson, B.: A unified inter-host and in-host model of antibiotic resistance and infection spread in a hospital ward. *Journal of Theoretical Biology.* 421, 112–126 (2017).
4. DalBen, M. de F. et al.: A Model-Based Strategy to Control the Spread of Carbapenem-Resistant Enterobacteriaceae: Simulate and Implement. *Infect Control Hosp Epidemiol.* 37, 11, 1315–1322 (2016).

5. Duval, A. et al.: Measuring dynamic social contacts in a rehabilitation hospital: effect of wards, patient and staff characteristics. *Scientific Reports*. 8, 1, 1686 (2018).
6. Ferrer, J. et al.: Nosolink: An Agent-based Approach to Link Patient Flows and Staff Organization with the Circulation of Nosocomial Pathogens in an Intensive Care Unit. *Procedia Computer Science*. 18, 1485–1494 (2013).
7. Grimm, V. et al.: A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*. 198, 1, 115–126 (2006).
8. Grundmann, H., Hellriegel, B.: Mathematical modelling: a tool for hospital infection control. *Lancet Infect Dis*. 6, 1, 39–45 (2006).
9. Hotchkiss, J.R. et al.: An agent-based and spatially explicit model of pathogen dissemination in the intensive care unit. *Crit. Care Med*. 33, 1, 168–176; discussion 253-254 (2005).
10. van Kleef, E. et al.: Modelling the transmission of healthcare associated infections: a systematic review. *BMC Infect. Dis*. 13, 294 (2013).
11. Laskowski, M. et al.: Agent-based modeling of the spread of influenza-like illness in an emergency department: a simulation study. *IEEE Trans Inf Technol Biomed*. 15, 6, 877–889 (2011).
12. McBryde, E.S. et al.: A stochastic mathematical model of methicillin resistant *Staphylococcus aureus* transmission in an intensive care unit: predicting the impact of interventions. *J. Theor. Biol*. 245, 3, 470–481 (2007).
13. Meng, Y. et al.: An application of agent-based simulation to the management of hospital-acquired infection. *J Simulation*. 4, 1, 60–67 (2010).
14. Milazzo, L. et al.: Modelling of healthcare-associated infections: a study on the dynamics of pathogen transmission by using an individual-based approach. *Comput Methods Programs Biomed*. 104, 2, 260–265 (2011).
15. Ong, B.S. et al.: An Individual-Based Model of Influenza in Nosocomial Environments. In: Bubak, M. et al. (eds.) *Computational Science – ICCS 2008*. pp. 590–599 Springer Berlin Heidelberg (2008).
16. Pittet, D. et al.: Evidence-based model for hand transmission during patient care and the role of improved practices. *Lancet Infect Dis*. 6, 10, 641–652 (2006).
17. Temime, L. et al.: Peripatetic health-care workers as potential superspreaders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 43, 18420–18425 (2009).
18. Triola, M.M., Holzman, R.S.: Agent-based simulation of nosocomial transmission in the medical intensive care unit. In: *16th IEEE Symposium Computer-Based Medical Systems, 2003*. Proceedings. pp. 284–288 (2003).
19. Ueno, T., Masuda, N.: Controlling nosocomial infection based on structure of hospital social networks. *J. Theor. Biol*. 254, 3, 655–666 (2008).
20. Antimicrobial resistance: global report on surveillance 2014. WHO (2014).
21. Surveillance of antimicrobial resistance in Europe 2017. ECDC (2018).

**Acknowledgments.** i-Bird Study Group : Anne Sophie Alvarez (AP-HP, Paris, France), Audrey Baraffe (AP-HP, Paris, France), Mariano Beiró (Universidad de Buenos Aires, Buenos Aires, Argentina), Inga Bertucci (AP-HP, Paris, France), Pierre-Yves Boëlle (Univ. Pierre et Marie Curie, Paris, France), Camille Cyncynatus (AbAg, Chilly-Mazarin, France), Florence Dannet (AP-HP, Paris, France), Marie Laure Delaby (AP-HP, Paris, France), Pierre Denys (AP-HP, Paris, France), Matthieu Domenech de Cellès (Univ. Pierre et Marie Curie, Paris, France), Eric Fleury (ENS Lyon, Lyon, France), Antoine Fraboulet (Insa, Lyon, France), Jean-Louis Gaillard (AbAg, Chilly-

Mazarin, France), Boris Labrador (AP-HP, Paris, France), Jennifer Lasley (Inserm, Paris, France), Christine Lawrence (AP-HP, Paris, France), Judith Legrand (Univ. Paris Sud, Orsay, France), Odile Le Minor (Institut Pasteur, Paris, France), Caroline Ligier (Institut Pasteur, Paris, France), Lucie Martinet (Inria, Lyon, France), Karine Mignon (AbAg, Chilly-Mazarin, France), Catherine Sacleux (AP-HP, Paris, France), Jérôme Salomon (Cnam, Paris, France), Thomas Obadia (Univ. Pierre et Marie Curie, Paris, France), Marie Perard (AP-HP, Paris, France), Laure Petit (Institut Pasteur, Paris, France) Laeticia Remy (AP-HP, Paris, France), Anne Thiebaut (Inserm, Paris, France), Damien Thomas (AbAg, Chilly-Mazarin, France), Philippe Tronchet (AP-HP, Paris, France), Isabelle Villain (AP-HP, Paris, France).