Master of Public Health

Master de Santé Publique

Analysis of carbapenem-resistant Enterobacteriaceae outbreaks in the French healthcare network using mathematical modeling

Clément MASSONNAUD

Master of Public Health
2018 – 2019

Location of the practicum:
METIS department, EHESP

Professional advisor:
Pascal CREPEY
METIS department, EHESP

Academic advisor:
Judith MUELLER
METIS department, EHESP

À mon père.

Acknowledgements

I'd like to thank my family, who has always been there for me, during all these years of med school and beyond. I couldn't have done it without you. I love you.

I'd like to thank Fanny, who has supported me throught this year, and has always been there for me too. Thank you for the wise and precious advices on the writing of this thesis.

I'd like to thank all the great people I met during this amazing year. It was a wonderful experience, and a great pleasure to share this with you. Special thanks to those who invited me into their homes. Special thanks to Juwel for the food, the laughs, and the dancing. I'd also like to thank Martine and all the EHESP staff for their efforts into making this Master a pleasant and satisfying experience. Special thanks to Stephanie for her help with the manuscript.

Of course, I'd like to thanks my supervisor, Pascal Crépey. Who, although defeated at Mario Kart, didn't held a grudge against me, and supported me and taught me a lot during these few months. Special thanks to Jonathan, with whom I worked closely on this project. It was a pleasure working with you, and I wish you the best for the future. Thanks to all the METIS departement, for making it such a nice working environment. Special thanks to Mathilde for the shikakai.

I also wish to thank Professor Merle and Doctor Marini for their help and precious expertise.

I wish to thank my dear fried and collegue André Gillibert. Thank you for having welcomed me into your home this year. Thank you for always being there to answer my many (and sometimes stupid) questions. You are the smartest person I have ever met, and I hope that one day you'll achieve your amazing goals. It is an honor to know you and a priviledge to work with you.

I also wish to thank my dear friend and collegue Mikael Dusenne. When I first met you I barely knew how to turn on a computer. You've taught me everything I know (about computer science, let's not get carried away). You're the living proof that geek is the new sexy, and that one day nerds will rule the world.

*

Finally, I'd like to have a word for all the nice bacteria out there. In this troubled time, when humanity is faced against the evil forces of killing bacteria, it is easy to forget that not all of you are hostile to human kind, and that we wouldn't exist without you. We owe you, we love you, we are you.

*

Escherichia coli will return

Note: this report was generated with R markdown (2019). An online version of this report is available at: https://bookdown.org/clement_massonnaud/report/

Contents

1 Abstract **1**

2 Introduction **2**

2.1 Carbapenem-resistant Enterobacteriaceae . . . 2

2.2 Modeling outbreaks in the metropolitan French healthcare facilities network . . . 5

2.3 Objective . . . 5

3 Methods **6**

3.1 Building the network . . . 6

3.2 Analyzing the network . . . 7

3.3 Developing the model . . . 9

3.4 Simulations . . . 14

4 Results **14**

4.1 Networks' characteristics . . . 14

4.2 Model simulations . . . 19

5 Discussion **23**

6 Conclusion **25**

7 References **26**

8 Appendices **32**

8.1 Appendix A . . . 32

8.2 Appendix B . . . 33

8.3 Appendix C . . . 35

8.4 Appendix D . . . 36

9 Summary in French **38**

Acronyms

- **ABR** Antibacterial resistance
- **AMR** Antimicrobial resistance
- **CDC** Centers for Disease Control and Prevention
- **CPE** Carbapenemase-producing Enterobacteriaceae
- **CRE** Carbapenem-resistant Enterobacteriaceae
- **EARS-Net** European Antimicrobial Resistance Surveillance Network
- **ECDC** European Center for Disease Prevention and Control
- **ESBL** Extended-spectrum beta-lactamase
- **HAD** Hospitalisation à domicile
- **HAI** Health-associated infection
- **HCF** Healthcare facility
- **MCO** Médecine-Chirurgie-Obstétrique-Odontologie
- **MDR** Multidrug-resistant
- **MRSA** methicillin-resistant Staphylococcus aureus
- **PDR** Pandrug-resistant
- **PMSI** Programme de médicalisation des systèmes d'information
- **RIM-P** Recueil d'information médicalisée en psychiatrie
- **SPF** Santé publique France
- **SSR** Soins de suite et de réadaptation
- **WHO** World Health Organization
- **XDR** Extensively drug-resistant

1 Abstract

Introduction. Carbapenem-resistant Enterobacteriaceae (CRE) are spreading at an alarming rate, and threaten health systems and patient safety worldwide. However, data on the dynamics of CRE outbreaks is lacking, especially in France. This study aims at developing a mathematical model reproducing CRE outbreaks on the French network of healthcare facilities (HCF).

Methods. We constructed the network of HCF of metropolitan France, over the years 2014-2016, using the national discharge database. We then developed a stochastic, hospital-based, susceptible-colonized-infected model to reproduce the dynamics of CRE outbreaks across the 2015 network. It takes into accout, intra-hospital dissemation, spreading by transfers, and importation from the community. We fitted the model on the 2015 surveillance data and performed simulations

Results. The network included 2,433 HCF for a maximum of 1,285,991 transfers recorded (2016). It was stable in its main characteristics over the three years and showed a high level of clustering. We estimated that the risk of infection after colonization by CRE in a healthcare setting was between 3.5% and 8.5%. We estimated a current level of detection of CRE episodes between 100% and 40% (1,209 undetected episodes). Assuming a baseline level of detection of 40%, we estimated that raising the level of detection up to 100% would only reduce the total number of episodes from 2,207 to 1,751 (20.7% reduction).

Conclusions. This model suggests that patient transfer between HCF could play a critical role in the dynamics of CRE outbreaks. To our knowledge, this model is the first to study spread of pathogens in HCF on such a large scale. It could be a valuable tool for further research, and to help stakeholders in the management of the increasingly important issue of anti-microbial resistance.

2 Introduction

2.1 Carbapenem-resistant Enterobacteriaceae

2.1.1 Antibacterial resistance: definitions

Bacteria were one the first life form to emerge on Earth and were first observed in 1676 by Antoine van Leeuwenhoek.[1] They are microorganisms present almost everywhere in our environment, including the human body itself. Most of the time, bacteria cohabit harmoniously with human cells, but can sometimes cause infections, ranging from benign diseases to more serious afflictions or even death. Although many ancient cultures knew about antibacterial properties of specific plants and used them to treat wounds,[2] the era of modern antibacterial treatment by chemotherapy began in the early 20th century with the work of Paul Ehrlich and Alfred Bertheim on synthesized arsenical derivatives.[3] Since then, antibacterial drugs have been increasingly developped and used, and saved millions of people thanks to their ability to treat infections that were previously incurable. However, besides bacteria that can be naturally resistant to antibacterial drugs, bacteria that were usually sensitive to one or multiple drugs can acquire resistance mechanisms, making some antibacterial agents ineffective. In the litterature, antibacterial resistance (ABR) most often denotes specifically these bacteria for which the resistance was not naturally present but was subsequently acquired.

Developping resistance to an antibacterial agent is a natural phenomenon, but can be accelerated by a process of selective pressure. An extensive use, and more importantly misuse, of antibacterial drugs will select resistant bacteria strains and result in increased levels of ABR over time. Antibacterial drugs can be categorized in different ways, based on either: the source, the type of action, the spectrum of activity, the chemical structure or the function.[4] Most of the common bacteria strains are naturally susceptible to multiple categories of antibacterial drugs. Consenquently, different levels of resistance can be defined according to the number of antibacterial agents that remain effective on the bacteria. In 2011, an international group of experts from the Centers for Disease Control and Prevention (CDC) and the European Center for Disease Prevention and Control (ECDC) proposed a standardized terminoly with which to describe resistance profiles of selected bacteria.[5] They identified three levels of resistance, defined as follows: (i) multidrug-resistant (MDR) if the the isolate is non-susceptible to at least one agent in three or more antibacterial categories, (ii) extensively drug-resistant (XDR) if the isolate remains susceptible to only one or two categories and (iii) pandrug-resistant (PDR) if the isolate is non-susceptible to all agents in all antibacterial categories. Progression in resistance levels consenquently limits therapeutic options, which can lead to complex situations or even therapeutic dead ends. In April 2014, the World Health Organization (WHO) released a global report that revealed alarming levels of resistance worldwide, and declared ABR to be a major global public health concern, calling for urgent action.[6]

2.1.2 The threat of antibacterial resistant Enterobacteriaceae

Enterobacteriaceae are a family of Gram-negative bacteria. They are pathogens of great concern for human health for three main reasons: (i) they are responsible for the largest proportion of HAI,[7,8] (ii) they are ubiquitous in the environment and are easily transmitted between humans (by hand carriage, contaminated food, water, etc.),[9] and (iii) they can rapidly acquire resistance mechanisms and transmit them by genetic material transfer.[10,11]

Two of the most important members of this family are *Escherichia coli* and *Klebsiella pneumoniae*. *E. coli* is commonly found in the gastrointestinal tract of humans and other warm-blooded animals and is spread through fecal-oral transmission.[9] Although most of the strains are innocuous, some are pathogenic and can cause infections ranging from simple urinary tract

infections to intra-abdominal infections, neonatal meningitis, hemolytic-uremic syndrome or septicemia.[12] *K. pneumoniae* is commonly found in humans in the gastrointestinal tract, the mouth or skin, and is mainly transmitted through person-to-person contact.[9] *K. pneumoniae* is an oportunistic pathogen, that is, infections occur primarily in people with weakened immune system. In consequence, they are often health-associated infections (HAI): hospital-acquired pneumonia, ventilator-associated pneumonia, catheter-associated infections, bloodstream infections, complicated intra-abdominal or urinary tract infections.[12] But *K. pneumoniae* infections, such as pneumonia, can also be seen in the community.

Enterobacteriaceae can usually be treated with β-lactam antibiotics. β-lactam antibiotics can be categorized into four groups: penicillins, monobactams, cephalosporins and carbapenems.[13] However, the last three decades have seen the emergence of MDR Enterobacteriaceae, mainly through the developement of β-lactamases, enzymes that can inactivate several β-lactam antibiotics.[14,15] A specific group of β-lactamases, the extended-spectrum β-lactamases (ESBL), can inactivate a wider range of important β-lactam antibiotics such as third-generation cephalosporins and monobactams.[16] ESBL-producing Enterobacteriaceae are therefore resistant to the majority of, if not all, first line antibiotics.[17] Even though the evidence is scarce, carbapenems are considered the main therapeutic option for ESBL infections.[18–22] Some studies suggested that other antibiotics - or associations of antibiotics - might be used, such as cefepim, cephamycins, fosfomycin or combinations of β-lactam antibiotics with β-lactamases inhibitors (piperacillin/tazobactam, amoxicillin/clavulanate, cefoperazone/sulbatacm), their clinical efficacy remains controversial, especially in more severe infections.[15,23–26] With the spread of ESBL-producing Enterobacteriaceae over the years,[27,28] the medical community made extensive use of carbapenems, thus creating a selective pressure that progressively selected carbapenem resistant organisms.

The first carbapenem-resistant Enterobacteriaceae (CRE) was identified in 1996 in North Carolina, USA, from a *K. Pneumoniae* isolate.[29] Since then, identification of CRE have been increasingly reported all around the world.[30–35] Enterobacteriaceae can acquire carbapenem resistance through different mechanisms, the main one being the production of enzymes that can inactivate carbapenems (i.e. carbapenemases).[13] Such organisms are often referred to as carbapenemase-producing Enterobacteriaceae (CPE). The term CRE encompasses organisms resistant to carbapenems, regardless of the underlying mechanism.

CRE have become one of the main concerns in the management of infectious diseases in the 21st century, and now represents a major threat to modern healthcare, worldwide. Cassini et al. performed a population-level modeling analysis to estimate attributable deaths and disability-adjusted life-years (DALY) caused by infections with ABR bacteria in Europe in 2015.[36] They estimated a number of deaths attributable to infections by carbapenem-resistant *K. pneumoniae* and *E. coli* between 1,914 and 2,638. Although they represented 2.8% of the total number of infections by ABR bacteria, they accounted for 6.8% of the deaths. More alarming than these absolute values, is the rapidly increasing trend. The proportion of DALYs due to carbapenem resistant *K. pneumoniae* and *E. coli* went from 4.3% in 2007 to 8.79% in 2015. As CRE are often also resistant to many other antibacterial drugs,[37] therapeutic options are very limited and evidence on their effectiveness is lacking.[38–41] In 2017, WHO released a priority list of bacteria for which new antibiotics are needed. CRE were among the top three priority bacteria (along with carbapenem-resistant *Acinetobacter baumannii* and *Pseudomonas aeruginosa*), with critical need for new antibiotics.[42] However, the development of new antibacterial drugs is a long and complex process and very few novel antibacterial drugs are expected to be available in a near future.[13] In any case, the management of these emergent extensively drug-resistant bacteria cannot rely on novel antibiotics. Considering the worrisome epidemiology of ESBL-producing Enterobacteriaceae, further spread of CRE would have serious consequences on modern medicine, which could be faced with pandrug-resistant bacteria. There is an urgent need for comprehensive, effective and realistic measures to limit the spread of CRE.

2.1.3 Carbapenem-resistant Enterobacteriaceae: the situation in France

In France, the first episode of CRE infections occurred in 2004 in an university hospital involving a *K. pneumoniae* strain. The index case was a patient transferred from Greece.[43] Since then, the number of episodes has not ceased to increase (Figure 1). Every year, Public Health France (Santé Publique France) releases a status report on the number of notified CPE episodes.[44] A total of 3,604 episodes affecting 5,541 patients have been identified from 2004 to 2016. Of those cases, 998 (18%) were clinical infections and 4543 (82%) were colonisations (carriage without clinical infection). The two strains most frequently involved were *K. pneumoniae* (54% of episodes) and *E. coli* (38%).

These reports raise serious concerns. First, the rapidly increasing trend is particularly worrisome. According to the French national center for antibiotic resistance (Centre National de Référence de la Résistance aux Antibiotiques), of the isolates recieved for suspicion of cabarpenemase production in 2017, 60.6% were CPE while they were only 23.1% in 2012.[45] Second, from 2004 to 2010, the proportion of episodes that were not associated with cross-border transfer was 20%,[46] whereas it was 56% in 2016, suggesting an important autochthonous circulation of CPE in France. Finally, a non negligible proportion of episodes involved secondary cases, sometimes in multiple healthcare institutions. The management of such episodes is more complicated and resource intensive. A 2010 episode involving 13 cases led to the screening of 280 contact cases in fifteen different healthcare facilities. In 2012, an episode involved 200 cases, and a 2013 episode involved 143 cases.[44]

Although the number of CRE episodes in France remains relatively limited, it is increasing every year. In July 2015, the French ministry of health released a national HAI prevention program (Programme national d'actions de Prévention des Infections Associées aux Soins, Propias).[47] Although the effect of such a program might not be seen immediately, it has not yet succeeded to stop the increasing trend. Efforts must be carried on, and new tools must be developped to gain insight on this emergent problem.

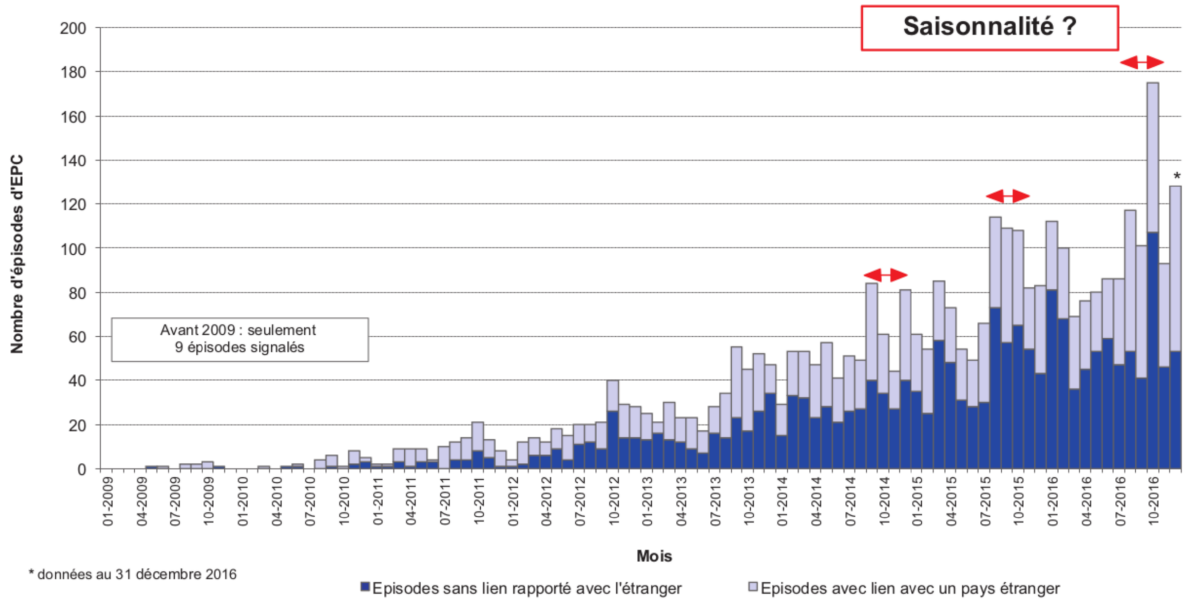

Figure 1: Monthly incidence of CPE episodes in France, 2009 - 2016: figure from Vaux et al.

2.2 Modeling outbreaks in the metropolitan French healthcare facilities network

2.2.1 Modeling the spread of pathogens in healthcare facilities: the SPHINx study

Mathematical modeling is a powerful tool that can help study a wide range of phenomena. When studying the dissemination of infectious diseases and evaluating the potential impact of control measures, observational data might not allow to answer some important questions. On the other hand, traditional experimental designs are often impractical and unethical. In such cases, modeling remain the only option to answer these questions.

The spread of pathogens in healthcare institutions is the result of processes that happen at different scales. A pathogen can be transmitted (i) within a healthcare facility (HCF), through contact between patients, contact between patients and healthcare workers or transfer of patients between wards; (ii) between HCFs through patient transfers and (iii) between a HCF and the community.

The use of modeling to investigate HAI spread is relatively recent,[48] and these previous studies were limited to describing HAI spread at a single scale. However, ignoring the dependance of the three distinct scales in HAI spread limits the ability of models to provide accurate and realistic insights into this phenomenon.

The purpose of the SPHINx project (Spread of Pathogens on Healthcare Institutions Networks) is to develop a unified simulation framework in which each scale will be taken into account. Three models (one for each scale) will be develop independently by different research teams and then will be integrated together in a meta-model to investigate the intricated mechanics of these distinct processes in HAI spread. The goal is to devise and evaluate innovative, realistic and cost-effective multi-scale strategies to control HAI incidence.

2.2.2 Modeling pathogen spread between healthcare facilities

One task of the SPHINx project is to study the role of patient transfers between HCFs in the spread of HAI. Recent studies have underlined the necessity to take into account the links between different institutional settings to plan efficient HAI control strategies.[48] However, most of them did not use observed data on HAI circulation.

The PMSI database (French acronym for “Programme de Médicalisation des Systèmes d’Information”) is the French national database containing patients discharge data. In particular, it includes information on all patient transfers between hospitals, and from hospitals to post-acute care and rehabilitation centers. The national discharge data can be used to construct the French healthcare institutions network and develop models of pathogen spread between these institutions.

2.3 Objective

The objective of this work was, first, to use the French national discharge database (PMSI) to construct the networks of healthcare facilities in metropolitan France, over the years 2014-2016. Then, to develop a mathematical model that could reproduce the dynamics of carbapenem-resistant Enterobacteriaceae outbreaks across the 2015 network of HCF, which could then be used to gain insights on the CRE epidemic in metropolitan France.

3 Methods

3.1 Building the network

In order to study the spread of carbapenem-resistant Enterobacteriaceae (CRE) across the network of healthcare facilities (HCF), the first task was to construct the network. Formally, a network is composed of *nodes* which may or may not be connected by *edges* (the study of networks is based on the mathematical field called *graph theory*). Depending on the context, nodes can be people, locations or web pages, so their connectedness would be expressed by different aspects of their relations. In our case, nodes are HCF and their connectedness is represented by the number of patients transferred between them in a defined time interval (typically one year or one month). The greater the number of patients transferred between two HCF, the stronger their connection. The "level of connectedness" can be quantified in many different ways, and the choice of the most appropriate indicator is dependant on the context, and is often subject to debate (network analysis is discussed in the next section). One important aspect of the network of HCF is that the edges are oriented: we know in which direction the patients move between facilities. Thus, the network is said to be *directed*.

In terms of data structure, a common way of representing a network is with a simple $n * n$ matrix. The rows and columns contain the nodes (which appear once in each), and each cell contains the information on whether or not the two nodes are connected. The matrix of a network is called the *adjacency matrix* (two nodes connected by an edge are called adjacent). For example, Table 1 is the adjacency matrix of a network of five HCF. By convention, the rows contain the facilities of origin, and the columns contain the target facilities. Each cell contains the number of patients transferred.

Table 1: Adjacency matrix of a 5 nodes network. The rows contain the facilities of origin, and the columns contain the target facilities. Each cell contains the number of patients transferred.

	A	B	C	D	E
A	0	687	373	296	0
B	0	0	1294	263	598
C	602	0	0	0	0
D	0	0	718	0	0
E	339	0	86	35	0

Therefore, to build the network of French HCF we needed to obtain the number of patients transferred between each pair of HCF. This information was retrieved using the French national hospital discharge database (PMSI). The PMSI contains all discharge data from public and private HCF in France, but is actually comprised of four distinct databases:

1. hospitalization in medicine, surgery, obstetrics or odontology departments (**MCO** - Médecine, Chirurgie, Obstétrique, Odontologie)
2. hospitalization in post-acute care and rehabilitation facilities (**SSR** - Soins de suite et de réadaptation)
3. hospitalization in psychiatric departments (**RIM-P** - Recueil d'information médicalisée en psychiatrie)
4. home care (**HAD** - Hospitalisation à domicile)

We built the network of HCF using only the MCO and SSR databases. We did not use the HAD database as we are only modeling the spread of pathogens across HCF. We also chose not to include the RIM-P database because it would significantly increase the difficulty of linking these databases as they all have different structures.

Among others, the MCO and SSR databases contain the following variables: unique patient and unique facility identifiers, dates of admission and discharge, entry mode (domicile, mutation, or transfer) and exit mode (domicile, mutation, transfer, or death) . We can therefore retrace patients' movements and then compute the number of patients exchanged between each pair of HCF (i.e. the adjacency matrix) with good reliability.

3.2 Analyzing the network

The next step after having built the network, is to explore its characteristics. The study of networks, particularly in biomedical contexts, is a new and rapidly growing discipline. With the help of graph theory advancements, many tools are available - and are continuously proposed - to explore networks. As networks can have different structures and properties, not all analysis tools can be applied to all kind of networks. We discuss here the tools that are most widely used, and that are relevant to the specific network of healthcare facilities (HCF).

3.2.1 Centrality of a node

One essential information about the network, is the importance of each node. Characterizing the importance of a node helps understand to what extent it influences the spreading process. The question is how to quantify the importance of a node? Many indicators have been devised, often called *centralities* (i.e. indicating how central a node is in the network). Some popular centralities that are applicable to the network of HCF are: the *degree*, the *closeness centrality*, the *betweenness centrality*, and the *hub* and *authority* scores.

3.2.1.1 Degree of a node

Two nodes u and v are *adjacent* if they are connected by an edge e. The edge e is said to be *incident* with the nodes u and v. The degree of a node u is the number of edges incident with u. The degree is a simple measure that counts the number of nodes to which u is *directly* connected (i.e. there is no intermediary node on the path). In the case of the network of HCF, the degree of a node can be thought of in two manners. We can choose to consider that if a facility A has sent at least one patient to a facility B, A and B are connected by an edge (in that case the edge will have a weight attribute equivalent to the actual number of patients transferred). Or, we can choose to draw one edge for each patient transferred, thus allowing multiple edges between A and B. Moreover, since the network is directed, we can actually define three degree measures for a facility A:

- the **in-degree**: the number of incoming incident edges (i.e. either the number of facilities from which A as received patients or the total number of patients received by A)
- the **out-degree**: the number of outgoing incident edges (i.e. either the number of facilities to which A has transferred patients or the total number of patients transferred by A)
- the **total degree**: the number of incident edges, outgoing or incoming, (i.e. the number of facilities to which A is connected or the sum of received and transferred patients)

For example, Figure 2 is the graph of the adjacency matrix 1, but with only one edge drawn if there is at least one patient transferred. We can see that node A has a total degree of 5, an in-degree of 2 and an out-degree of 3 ; node B has a total degree of 4, an in-degree of 1 and an out-degree of 3 ; etc. If we choose to draw one edge per patient transferred, A would have a total degree of 2297, an in-degree of 941 and an out-degree of 1356.

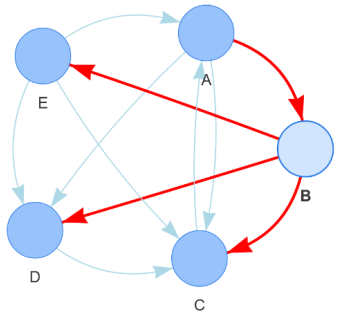

Figure 2: Graph of a 5 nodes network

3.2.1.2 Betweenness and closeness centralities

Bavelas,[49] Shimbel,[50] Shaw,[51] and Marriott and Cohn[52] had a common idea that a node is central if it has the potential to control the communication processes between other nodes. If a path connecting two nodes i and j passes through the node k, it is natural enough to say that k is *between* i and j. But the idea behind the betweenness centrality measures proposed by Anthonisse[53] and Freeman[54] is that k can have different levels of *betweenness* based on its potential to control the communication between i and j. If the path through k is the only one from i to j, then k controls completely the communication. If it is not the only one, but the *shortest path* (i.e. the path with the minimum number of edges), k would still have a high betweenness because it has a strategic location.

Although closeness and betweenness centralities both aim at identifying the most central nodes, they are somewhat conceptually opposite. Where the betweenness is the extent to which a node can control communication between other nodes, the closeness is the extent to which a node can "avoid the control potential of others" (Freeman).[55] A node will have a high closeness centrality if it can reach many other nodes without "intermediaries" or "relayers": it is *close* to many other nodes. On the other hand, a node is considered isolated if it depends on other nodes to transmit the information.

The betweenness and closeness centrality scores of the HCF network's nodes were computed using the R package `igraph`.[56] For more details on how these scores are computed, see Freeman[55] and Brandes.[57]

3.2.1.3 Hub and authority scores

Hub and authority scores are other popular centrality measures. They were first developed as ranking measures of webpages. Hubs are nodes that tends to have many outgoing connections, whereas authorities are nodes that tends to have many incoming connections. For a precise definition of these scores, see the seminal paper by Jon Kleinberg.[58] They were computed using the R package `igraph`.[56]

3.2.2 Communities (clusters)

To further explore a network, and understand how its nodes interact with each other, it is often interesting to look for communities. Essentially, a community is a subset of nodes that tends to have more interactions with each other, than with nodes of other communities. The question of finding communities in a network is fundamentally a question of identifying sub-networks within the network (also called subgraphs). Many methods and algorithms have been devised, which can define communities in different ways. We identified clusters in the healthcare facilities networks using two popular algorithms: the Greedy algorithm and the Map equation algorithm. For more details on these algorithms, see Clauset et al.[59] and Rosvall et al.[60,61] These clustering algorithms were applied using the R package `igraph`.[56]

3.3 Developing the model

3.3.1 Definitions and main considerations

The goal was to develop a complete model that could reproduce the spreading dynamics of Carbapenem-resistant Enterobacteriaceae (CRE) across healthcare institutions over time. The model should be able, from a defined initial state, to reproduce the observed number of CRE episodes reported by Public Health France (SPF). SPF defines CRE episodes as follows:

- a CRE **episode** is defined as one or multiple cases of CRE colonizations or infections linked by an epidemiological chain of transmission.
- a case of CRE **infection** is defined as a patient who developed a clinical infection (as defined by the clinical guidelines) caused by a CRE.
- a case of CRE **colonization** is defined as a patient detected for carrying a CRE, but with no clinical infection.
- for each episode, it is recorded whether the presumed index case has traveled, and/or been hospitalized in a foreign country in the past twelve months. If so, the episode is categorized as **imported**.

The most important fact to consider when developing this model, is that the episodes reported by SPF, are the episodes that were observed, that is, episodes that were **actually detected**. We must take into account the possibility that a certain number of episodes are not detected. However, we can assume that all CRE infection cases are detected. It is reasonable to expect that if a patient is infected by a CRE, it will lead to a clinical investigation, and the detection of the case. Thus, we assume that only episodes with no infected patients can escape detection. Because they are not detected, such episodes are likely to play an important role in the spreading dynamics. These considerations must be implemented in the model for it to replicate the proper dynamic. Consequently, the model will not only produce the number of episodes reported by SPF, but also a number of undetected episodes. Table 2 summarizes the main outputs of the model.

Table 2: Main outputs of the model.

Output	Details
infected detected	number of episodes that included at least one infected case, and were therefore detected.
colonized detected	number of episodes that included no infected cases (only colonized cases), but that were still detected
colonized undetected	number of episodes that included no infected cases (only colonized cases), and that remained undetected

3.3.2 Fitting the model

The model was fitted using two of the three outputs presented in the previous section and in table 2: infected detected and colonized detected. Each of these two outputs was actually splited into two values, depending on whether the episode is imported or not. In the end, the model was fitted on four time series:

- **episodes infected imported (I_i)**: the number of episodes that included at least one infected case, and were therefore detected, for which the index case was an imported case.
- **episodes infected not imported (I_{ni})**: the number of episodes that included at least one infected case, and were therefore detected, for which the index case was not an imported case.
- **episodes colonized detected imported (Cd_i)**: the number of episodes that included no infected cases (only colonized cases), that were detected, and for which the index case was an imported case.
- **episodes colonized detected not imported (Cd_{ni})**: the number of episodes that included no infected cases (only colonized cases), that were detected, and for which the index case was not an imported case.

The model was fitted using data obtained from SPF. We used 2015 data since complete data on 2016 and 2017 was not available at the time of the study. Figure 3 shows the four values for 2015. A total of 984 episodes were observed, 191 of which included at least one infected case (19.4%). Of the episodes with at least one infected case, 33.5% were imported, against 42.2% for episodes with only colonized cases. As shown in figure 3, we assumed that the number of episodes with infected cases is close to the true number of episodes (reality), whereas for episodes with only colonized cases we assumed that the true number of episodes was unknown. The model was fitted to obtain a number of episodes (episodes with infected cases or detected episodes with only colonized cases) in a 10% range around the reported data (± 5%).

Figure 3: Repartition of the number of episodes and the relationships between what is observed, the reality and what is modeled.

3.3.3 Implementation of the model

The main input of the model is the transfer matrix (adjacency matrix) of the 2015 network of healthcare facilities (HCF), that is the number of patients transferred between each pair of HCF in 2015. The base unit of the model is the HCF. An episode is limited to the scope of one HCF: one episode occurs in one HCF. An episode occurs in a HCF if a colonized or infected patient is admitted in the HCF. The patient can either come from the community, or from another HCF by transfer (figure 4, "contact"). Another main assumption we made, is that if a colonized or infected patient comes from the community, it is an imported case (the index case has traveled, and/or been hospitalized in a foreign country in the past twelve month). It means that we are not considering the possibility for a case to be discharged from the HCF, and then being admitted to another HCF later on, still colonized, thus creating a new episode.

Once a colonized or infected patient is admitted in a HCF, it can spread the pathogen to other patients (figure 4, "within hospital dissemination"). Then, patients can either be discharged back to the community, or be transferred to another facility. We assume that once patients are detected (either infected or colonized), they no longer play a role in the spreading dynamic since they are identified and control measures are applied. It means that, even if detected patients are transferred, they will not create a new episode in the target HCF. Thus, the spreading dynamic between HCF is mainly driven by undetected patients (who can only be colonized patients, since infected ones are always detected).

The model is almost entirely stochastic. It means that almost all events that occur are the results of random draws from a particular probability distribution. The time unit of the model is the day. Each day is modeled by a round of multiple computations that successively produce or terminate episodes in HCF. For computational purposes, we made implementation choices that must be specified. Several attributes are drawn *a priori*, that is at the time of arrival of the index case in the HCF (figure 4, "outbreak definition"):

- first, the **episode size**: we defined the episode size as the number of cases detected during the investigation that followed the detection of the first case. It is drawn from a distribution derived from the observed episodes' sizes in 2015.
- second, the presence of an **infected patient**: from the episode size drawn previously, we draw the number of patients that will be infected, ranging from zero to the episode size (all cases of the episode are infected). It is implemented as a binomial draw, where the parameter is p_{inf}, the probability of a patient developing an infection.
- third, whether the episode will be **detected or not**: implemented as a binomial draw of size one (one if the episode will be detected, zero otherwise). The parameter is the probability of detecting the episode, p_{det}, which is one if at least an infected case as been drawn, and which will be estimated otherwise (probability of detecting an episode without any infected cases).

These three values can be viewed as constituting the final state of an episode. Then, with each round (i.e. each day) the state of the HCF will be updated until the final state is reached. Each round, new colonized cases are generated (patients can become colonized), but are still undetected. At the same time, colonized cases that were previously generated can be "removed", either due to a natural spontaneous decolonization, or simply by leaving the hospital by transfer. If infected cases must occur (initial draw), they are generated once all colonized patients have been generated. Therefore, the time between the arrival of the index case and the completion of the final state is dependent on the competition between the two stochastic processes of generation and removal (figure 4, "within hospital dissemination"). The number of colonized cases generated

each round is drawn from a *Poisson* distribution, with a rate estimated from Cuzon et al.[62] The cases removed by spontaneous decolonization are drawn from a binomial distribution, with a probability α estimated from Davido et al.[63]

The number of colonized patients transferred each round is computed using the transfer matrix. The matrix is first transformed so that each cell no longer represents the annual absolute number of patients transferred between the two HCF, but the probability of any patient of the origin HCF being transferred to the target HCF on any given day (which is equivalent to the daily proportion of transfers from the origin HCF that occur towards the target HCF). It means that each pair of HCF of the network is attributed a specific probability. Then, this matrix of probabilities is multiplied by the main spreading parameter β, which will be estimated when fitting the model. Finally, a binomial draw is performed between each HCF that has undetected colonized patients, and each of their potential target HCF, where the size is the number of colonized patients present, and the probability is the corresponding cell in the matrix of probabilities. Thus, another matrix is created, where each cell is the number of colonized patients that are transferred from the origin HCF to the target HCF.

Once the final state is reached, one of two scenarios occurs depending on whether the episode is detected or not. If it is not detected, nothing happens, and cases stay in the HCF and can be transferred as before, can be spontaneously decolonized, or can be discharged (*Poisson* distribution with a rate drawn from the mean length of stay of patients). If the episode is detected, then an investigation starts, and cases are progressively detected. Based on experts opinions, we assumed a detection rate of 0.5, meaning all cases are detected in two days on average. Even if there is an investigation occuring, we allow the possibility for cases to be discharged before they are detected. Once all cases have been "removed", that is detected by investigation, or having left the HCF, the episode is over and the HCF returns to a null state (Figure 4, "investigation").

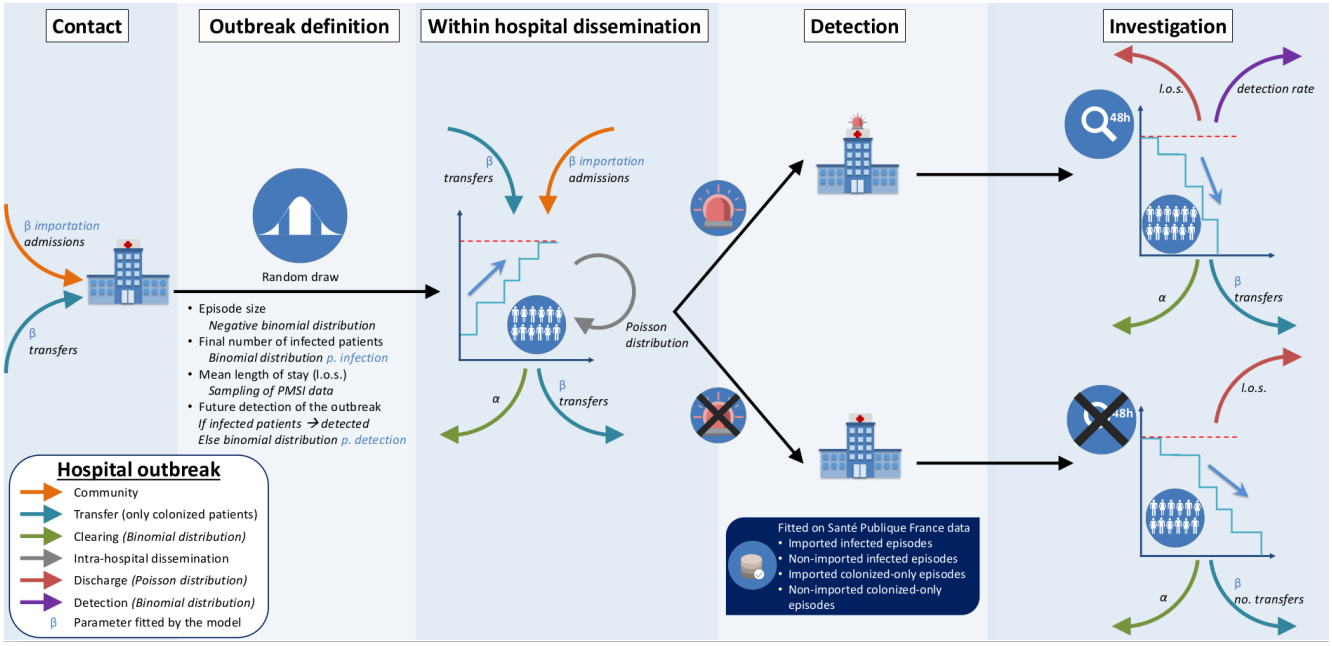

Figure 4: Representation of the model flow.

These implementations choices present two main advantages. First, it is more efficient from a computational perspective, since it reduces the number of random draws. Second, it creates the underlying dynamic of undetected cases. Indeed, the cases actually detected during the episodes reported by SPF, although thoroughly investigated, might not represent all the cases that were really "generated" in the HCF. And, the presumed index case might not be the true index case. Here, besides the number of cases observed once an episode is detected, we model the total number of cases that were generated from the true index case.

Table 3 summarizes the main assumptions of the model. Table 4 lists the parameters used (estimated and fitted) in the model. The model implementation can be summarized as follows:

1. a HCF receives a colonized or infected patient (index case), either from the community or from another HCF. The probability of a case arriving by transfer is related to the transfer matrix and the parameter β. The probability of case arriving from the community is related to the rate of admissions of the HCF (national data) and a parameter β_i. Those parameters will be estimated when fitting the model.

2. a scenario is drawn *a priori*: the episode's size (negative binomial), the number of infected cases (binomial, parameter p_{inf}), and whether the episode will be detected (binomial, parameter p_{det}). This is the *final state* of the episode.

3. the HCF's status is updated each round. New colonized patients can be generated (*Poisson*). Previously colonized patients can be removed by spontaneous decolonization (binomial), or by transfer (transfer matrix and β).

4. check if the final state is reached:

- if the final state is not reached, proceed to the next round.
- if the final state is reached, check if the episode is detected:
  - if it is detected, an investigation starts, and cases are progressively removed (binomial), or are discharged before the investigation detected them (*Poisson*).
  - if it is not detected, cases can leave the HCF by discharge (*Poisson*), or by transfer, or they can be removed by spontaneous decolonization (binomial, parameter α).

5. if all cases have left the HCF, the episodes terminates, and the HCF returns to a null state.

Table 3: Main assumptions of the model

Assumptions
1. Some episodes might not be detected
2. All episodes with at least one infected case are detected
3. Episodes not imported originate only from transfers
4. Once patients are detected, they can no longer spread the pathogen

Table 4: Parameters of the model, their role, and their source

Parameter	Role	Source
β	transfer of cases	fitted
β_i	admission of cases from community	fitted
p_{inf}	number of infected cases in an episode	fitted
p_{det}	whether the episode is detected or not	fitted
transfer matrix	transfer of cases	PMSI[1] database
size	episode's size	SPF[2] data
λ	generation of new cases during an episode	Cuzon et al. (2011)
α	spontaneous decolonization	Davido et al. (2018)
detection rate	speed at which cases are detected	experts opinions
mean length of stay	discharge of cases	PMSI[1] database

[1] PMSI: programme de médicalisation des systèmes d'information.

[2] SPF: santé publique France (public health France).

3.4 Simulations

To limit the random fluctuations that occur due to the high degree of stochasticity of the model, it was initialized each time at the same state. That is, the same facilities were initially attributed one colonized patient. To limit bias, these facilities were the one where an episode was recorded in the course of December of the previous year. Then the model was run over two years. Only the second year of each run is used to compute the results that are presented. The first year of a run is used as a "burning period", which limits the bias of initialization. For each run, we computed the total number of episodes that occured, the number of them that included infected patients, the number of them that included only colonized patients, and if they were detected or not. We also computed the number of cases per episode (infected and colonized), the proportion of cases that were actually identified when the episode was detected, and the duration of the episodes. We then analysed epidemic chains, defined as episodes in several facilities linked by the transfer of colonized patients. We computed the number of chains that occured, the length of the chains (defined as the number of transfers that occured within the chain), and the number of facilities involved in the chain. The model is currently implemented using the R language for statistical computing, version 3.6.0 "Planting of a Tree".[64]

4 Results

4.1 Networks' characteristics

4.1.1 Structure

The characteristics of the healthcare facilities (HCF) networks of metropolitan France are stable over the three years studied (2014, 2015, 2016). The networks were comprised of 2433, 2426, and 2427 MCO/SSR facilities, for a total of 1,233,309; 1,263,844; and 1,285,991 patients transfers recorded in 2014, 2015, and 2016, respectively (Table 5). The median edge weight (number of patients transfered between two connected hospitals) was equal for the three years. In both 2014 and 2015, the maximum number of transfers recorded between two hospitals was between university hospitals Charles Nicolle

and Bois-Guillaume in Rouen, Normandie (6,822 and 5,825, respectively). In 2016, it was between university hospitals Salengro and Claude Huriez in Lille, Nord (5333). The hospitals having received the most patients from transfers were university hospital Claude Huriez of Lille in 2014 and 2016 (6,443 and 6,764 patients recieved, respectively) and university hospital Arnaud de Villeneuve of Montpellier, Hérault in 2015 (6,678). The hospital having sent the most patients was the same the three years: university hospital Salengro of Lille, with 15,904 patients sent in 2014, 16,551 in 2015, and 17,041 in 2016. In terms of connectedness, the median total degree of an hospital (i.e. the number of distinct hospitals to/from which it has sent or received at least five patients) was 15 the three years. The most connected hospital was the university hospital Pitié-Salpêtrière of the Assistance Publique - Hôpitaux de Paris (AP-HP), Paris, in 2014, 2015, and 2016, having exchanged patients with 431, 427, and 454 different healthcare facilities, respectively. The three years, hospital Necker enfants malades of AP-HP, Paris, was the hospital with both the highest betweenness and closeness measures. We ranked all facilities for each of the computed centrality measures. We saw that the ranking of a facility can vary greatly depending on the measure we consider (Appendix A).

Table 5: Main characteristics of the 2014, 2015, and 2016 MCO/SSR hospitals networks.

Characteristics	2014	2015	2016
Facilities	2433	2426	2427
Transfers	1233309	1263844	1285991
Structure (min-max)			
Median edge weight[1]	20 (5-6822)	20 (5-5825)	20 (5-5333)
Median total degree[2]	15 (1-431)	15 (1-427)	15 (1-454)
Median in-degree[2]	8 (0-201)	8 (0-208)	8 (0-213)
Median out-degree[2]	6 (0-230)	6 (0-219)	6 (0-241)
Communities by Greedy algorithm			
Number of clusters	23	24	24
Median cluster size (min-max)	88 (34-342)	88 (18-344)	92 (19-346)
Communities by Map Equation algorithm			
Number of clusters	115	118	118
Median cluster size (min-max)	15 (1-188)	14 (1-155)	14 (2-139)

[1] Number of patients transfered between two connected hospitals

[2] Number of distinct hospitals from/to which one hospital has received ('in') or transfered ('out') patients, or both ('total')

The distributions of all centrality measrues are highly skewed. Figure 5 shows the distribution of the total degree of healthcare facilities in the 2015 network. 75% of HCF had a total degree of 28 or less (i.e. they were connected to 28 other facilities or less), and 95% were connected to 77 other facilities or less. 121 HCF had a total degree between 78 and 287, and one hospital had the highest degree of 427. These few highly connected HCF can be referred to as "hubs" or "authorities". The ditributions for the other years were similar, with the 75th and 95th quantiles respectively at 29 and 77 degrees in 2014, and 28 and 78 degrees in 2016 (Appendix B). Regarding patients transfers, the distributions are also highly skewed, but are somewhat different for the number of patients sent and the number of patients received. In 2015 (Figure 6), the 95th quantiles for patients received and sent were 1,510 and 2,346, respectively. 126 HCF (5%) recieved more than 1,500 patients but 222 HCF (9%) sent more than 1,500 patients. The maximum number of patients received by a HCF was 6,678, while 13 HCF sent over 6,678 patients, the maximum being 16,551. The distributions for the other years are similar (Appendix B).

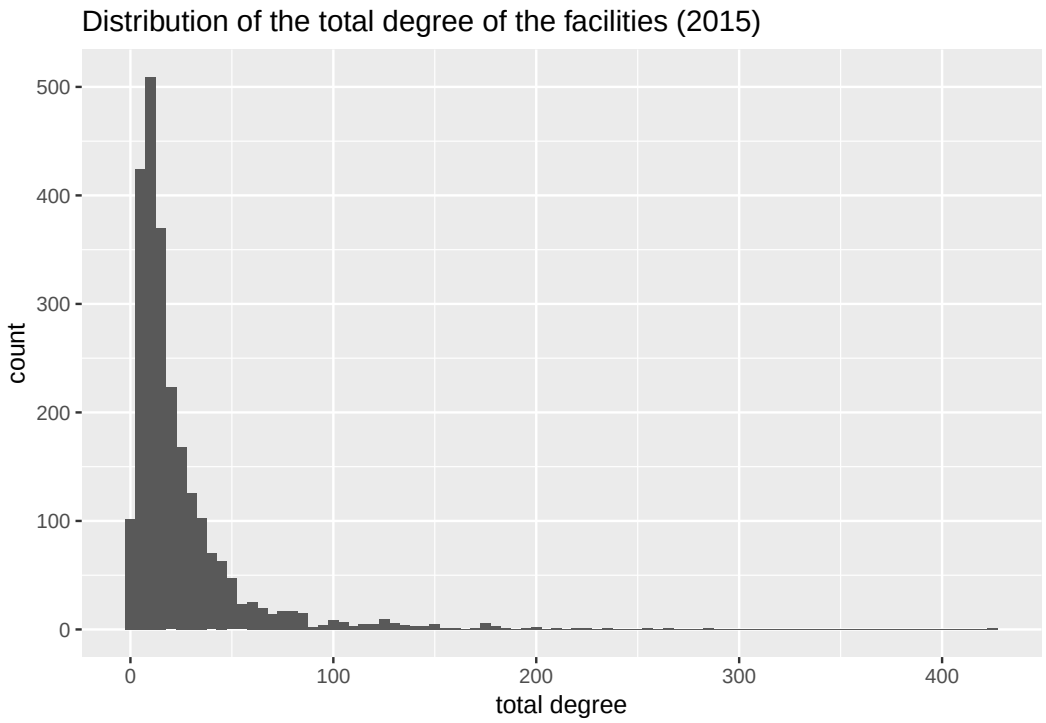

Figure 5: Distribution of the total degree of healthcare facilities in the 2015 network.

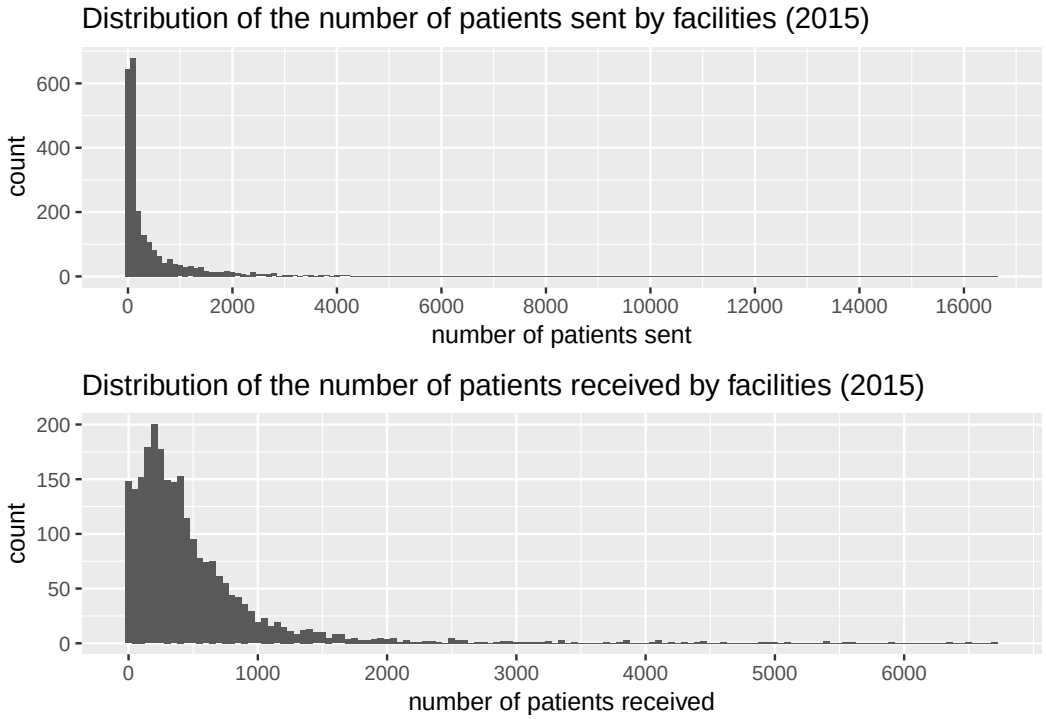

Figure 6: Distribution of the number of patients sent and received by healthcare facilities in the 2015 network.

4.1.2 Communities

We identified community structures in the HCF networks using two clustering algorithms: the Greedy and Map Equation algorithms. The Greedy algorithm systematically identified less (and therefore larger) clusters of HCF (Table 5). In 2014, Greedy identified 23 clusters, while Map Equation identified 116 of them. In 2015 and 2016, the number of clusters identified were 24 for Greedy and 118 for Map Equation. The median cluster size was consequently greater for Greedy than for Map equation (respectively 88 vs. 15 in 2014 and 2015, and 92 vs. 14 in 2016). The larger cluster was the Parisian one with both algorithms, the three years, with 342 (2014), 344 (2015), and 346 (2016) HCF with Greedy, and 139 (2014), 153 (2015), and 139 (2016) HCF with Map Equation. Hub and authority scores were computed for each HCF, by cluster. Table 6 shows the list of clusters identified by the Greedy algorithm on the 2015 network, with the main city of the cluster, the cluster's size and the HCF which had the greatest hub and authority scores of the cluster. The data for the other years is in Appendix C. All healthcare facilities in the networks were geocoded and mapped dynamically, which allowed to explore geographical characteristics of the networks. Figure 7 displays the geographic repartition of the clusters of HCF identified by the Greedy and the Map equation algorithms in the 2015 network. Each cluster is represented by a different color and delimited by a polygon. Although the algorithms identify the clusters using solely the transfer matrices, and no geographic data, we see that transfer clusters are also geographic clusters. We also see that the clusters identified by the Greedy algorithm correspond approximatively to the former regions of metropolitan France.

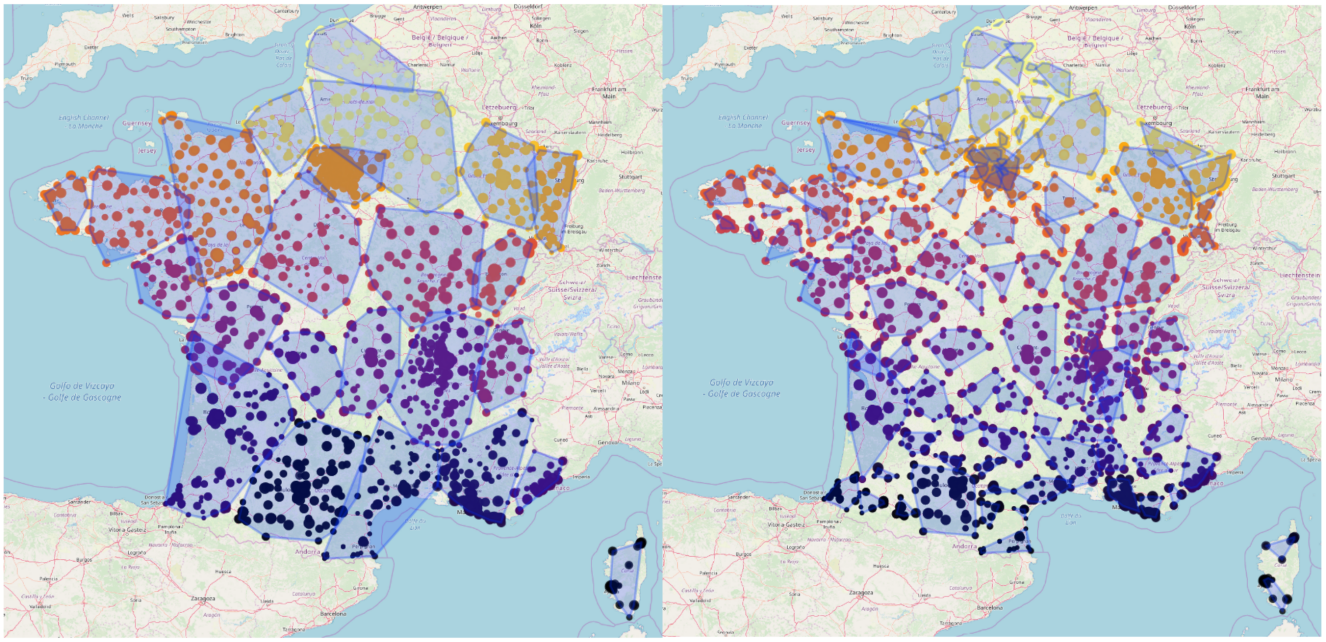

Figure 7: Geographic repartition of the clusters of healthcare facilities identified by the Greedy (left) and Map equation (right) algorithms in the 2015 network. Each cluster is represented by a different color and delimited by a polygon.

Table 6: List of clusters identified by the Greedy algorithm in the 2015 network, with the main city of the cluster, the cluster's size, and the HCF which had the greatest hub and authority scores of the cluster

Main city	Cluster size	Hub	Authority
Paris-Créteil	344	GPE HOSP HENRI MONDOR-ALBERT CHENEVIER	CHARLES-FOIX - JEAN-ROSTAND GHU EST
Lyon	200	HOPITAL EDOUARD HERRIOT	HOPITAL CARDIO-VASCULAIRE ET PNEUMOLOG
Marseille	170	HOPITAL LA TIMONE ADULTES	HOPITAL DE LA CONCEPTION
Bordeaux	153	CHU PELLEGRIN	HOPITAL DE HAUT LEVEQUE
Toulouse	132	HOPITAL DE PURPAN CHU TOULOUSE	HOPITAL DE RANGUEIL CHU TOULOUSE
Lille	130	HOP SALENGRO - HOPITAL B CHR LILLE	HOP CLAUDE HURIEZ CHR LILLE
Angers	125	CHU D' ANGERS :SITE LARREY	CHU D' ANGERS:CENTRE DE SSR
Montpellier	123	HOPITAL LAPEYRONIE CHU MONTPELLIER	HOPITAL ARNAUD DE VILLENEUVE CHU MPT
Amiens	119	HOPITAL SUD CHU AMIENS	HOPITAL NORD CHU AMIENS
Nancy	90	HOPITAL CENTRAL CHU NANCY	HOPITAUX DE BRABOIS CHU NANCY
Tours	90	C.H.R.U. -TROUSSEAU-	C.H.R.U. BRETONNEAU
Rennes	88	C.H.R. PONTCHAILLOU-RENNES	POLE GERIATRIQUE RENNAIS
Strasbourg	87	CHU STRASBOURG / HOP HAUTEPIERRE	CHU DE STRASBOURG /NOUVEL HOPITAL CIVIL
Dijon	81	HÔPITAL LE BOCAGE CHU DIJON	MÉDECINE À CHAMPMAILLOT CHU DIJON
Nantes	69	C.H.U. NANTES HÔTEL-DIEU ET HME	CHU DE NANTES HOPITAL G. R. LAENNEC
Nice	67	HOPITAL SAINT ROCH DU CHU DE NICE	HOPITAL PASTEUR DU CHU DE NICE
Rouen	66	HOPITAL CHARLES NICOLLE CHU ROUEN	HOPITAL DE BOIS-GUILLAUME CHU ROUEN
Grenoble	62	HOPITAL NORD (GRENOBLE)	HOPITAL SUD
La Rochelle	57	GROUPE HOSP. LA ROCHELLE-RE-AUNIS	C.H. ST HONORE - ST-MARTIN-DE-RE
Clermont-Ferrand	47	CHU G. MONTPIED	CHU ESTAING
Vesoul-Besançon	38	CH VESOUL	CHU JEAN MINJOZ
Limoges	37	C H U DUPUYTREN LIMOGES	HOPITAL JEAN REBEYROL LIMOGES
Brest	33	CHRU HOPITAL CAVALE BLANCHE	CENTRE DE CURE MED.& CONVALESC.
Corse	18	CH ND LA MISERICORDE	C.R.F. ET MAISON DE REPOS DU FINOSELLO

To further explore the clustering characteristics of the networks, we grouped the transfer matrices by cluster identified by the Greedy algorithm. It allowed us to compute the number of transfers recorded between facilities of a same cluster (transfers within clusters), and betweenn facilities of different clusters (transfers outside clusters). Table 7 shows, for each cluster identified by Greedy in the 2015 network, the number of transfers recorded within the cluster, and outside the cluster. Thus, the level of clustering can be simply computed as the proportion of all the transfers recorded by HCF of the cluster that were within the cluster. The levels of clustering were high for all clusters, for the three years. In 2014, it ranged from 91.6% (Angers) to 99.2% (Lille), in 2015 from 90.1% (Corse) to 99.2% (Lille), and in 2016 from 89.9% (Corse) to 99.0% (Lille). To explore the relations between them, the clusters can be thought of as forming their own network, where each node represents a cluster, and each edge that connects two nodes corresponds to patient transfers between HCF of the two clusters. Figure 8 is a graph visualization of such a network for the year 2015. The node's diameter is proportional to the cluster's size, and the edge's width is proportional to the number of patients transfered. The nodes' positions are determined dynamically by an algorithm, and are based on their connectedness: the more two nodes are connected, the closer they are. An interactive version of this graph, and additional data, can be found in the online version of this report.

Table 7: For each cluster identified by the Greedy algorithm, the number of transfers recorded within the cluster, outside the cluster, and the proportion of all the transfers recorded by HCF of the cluster that were within the cluster (2015).

		Transfers within cluster	Transfers outside cluster
Cluster	Size	n (%)	n (%)
Lille	130	83021 (99.2)	705 (0.8)
Strasbourg	87	42248 (98.1)	827 (1.9)
Marseille	170	84551 (97.7)	2014 (2.3)
Bordeaux	153	71551 (97.6)	1754 (2.4)
Toulouse	132	65389 (97.3)	1836 (2.7)
Nice	67	35446 (97.2)	1027 (2.8)
Montpellier	123	62349 (97.1)	1859 (2.9)
Paris-Créteil	344	209202 (97)	6384 (3)
Nancy	90	44741 (96.4)	1668 (3.6)
Lyon	200	96289 (96.1)	3909 (3.9)
Brest	33	18343 (96)	755 (4)
Nantes	69	32880 (95.9)	1400 (4.1)
Clermont-Ferrand	47	22703 (95.5)	1066 (4.5)
Rennes	88	41160 (95.5)	1960 (4.5)
Rouen	66	33681 (95.5)	1583 (4.5)
Angers	125	56037 (95)	2954 (5)
Tours	90	44285 (93.6)	3043 (6.4)
Vesoul-Besançon	38	16238 (92.7)	1270 (7.3)
Amiens	119	48797 (92.6)	3886 (7.4)
Grenoble	62	30348 (92.5)	2462 (7.5)
Limoges	37	15491 (92.4)	1279 (7.6)
La Rochelle	57	23283 (91.7)	2097 (8.3)
Dijon	81	31725 (90.2)	3441 (9.8)
Corse	18	4422 (90.1)	485 (9.9)

4.2 Model simulations

It was not possible to find a unique solution (unique set of parameter) that could fit the data. In fact, we have explored the parameters space to find limits, thus estimating a range of plausible values. This range of plausible values for the parameters has been split into 13 sets (Table 8). The estimation of the parameter p_{det} (overall probability of detecting an episode if it includes no infected cases) ranged from 40% to 100%. The estimation of the parameter p_{inf} (probability of an infection occuring in a colonized patient) ranged from 3.5% to 8.5%. There was an almost perfect positive correlation between the two parameters (Pearson's correlation coefficient: 0.99). Table 9 shows the results of the simulations, for one year. One hundred runs of the model were performed for each of the thirteen sets of parameters. The mean number of episodes with at least one infected case ranged from 178 to 201 (observed value from surveillance data: 191 [-6.8%; +5.2%]). The mean number of detected episodes with only colonized patients ranged from 755 to 836 (observed value: 793 [-4.8%; +5.4%]). Depending on the level of detection (i.e. p_{det}), the mean number of undetected episodes ranged from 0 (100% detection level) up to 1,209 (40% detection level).

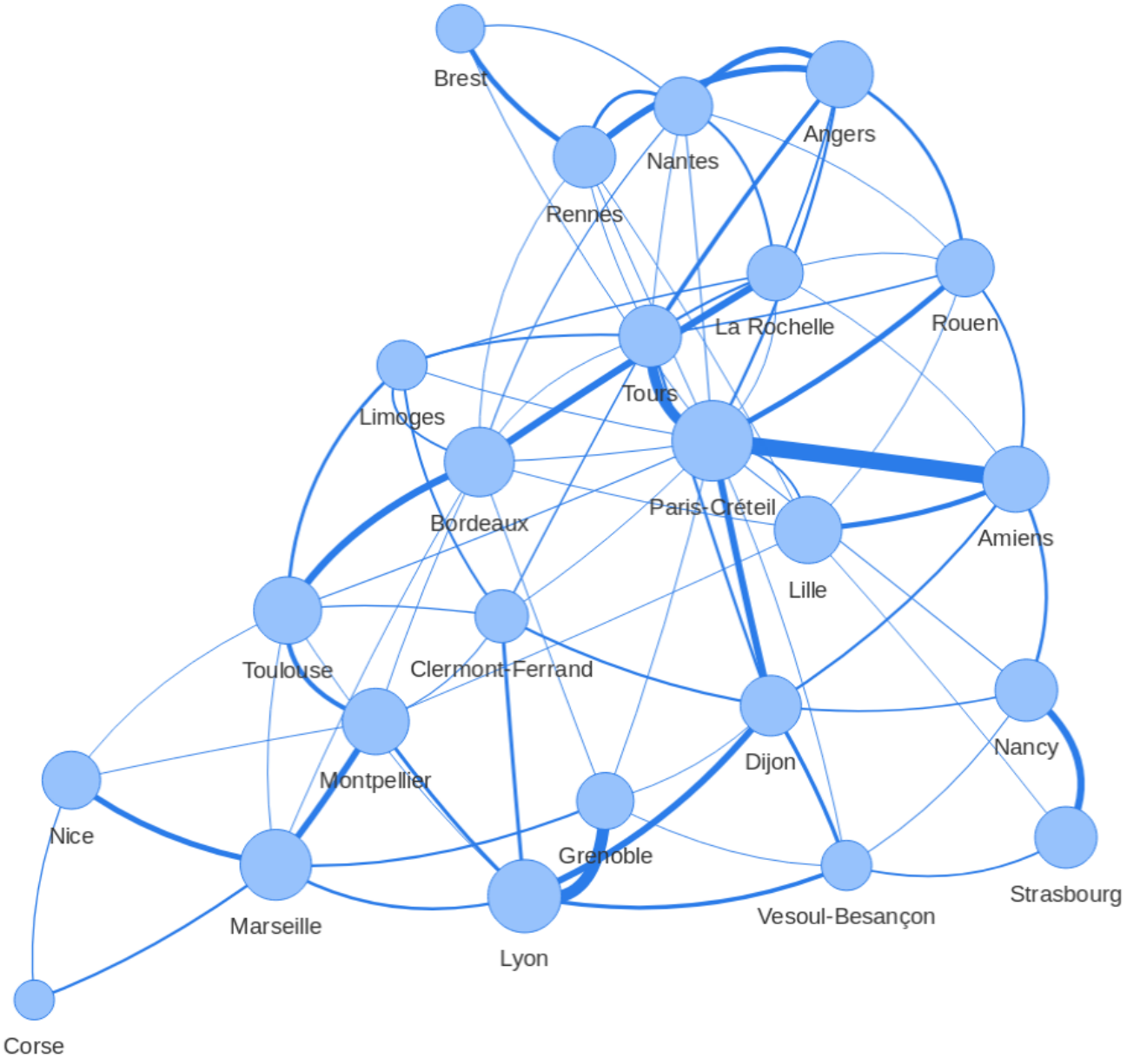

Figure 8: Graph of the 2015 network of clusters. Each node represents a cluster and each edge represents a connection between the two clusters (patients transfered). The node's diameter is proportional to the cluster's size, and the edge's width is proportional to the number of patients transfered.

The mean number of cases per episode was between 5.7 and 8.2. When an episode was detected, the proportion of all the cases generated during the episode that were actually identified by the investigation was, on average, between 75.4% and 76.7%. The median duration of an episode was 8 or 9 days when the episode was detected, whereas it ranged from 18 to 62.5 days when the episode was not detected. The mean number of epidemic chains (episodes in multiple facilities linked by the transfer of colonized patients) was estimated to be between 89 and 258. On average, the length of the chain (number of transfers) was between 4.3 and 4.9, and the number of facilities involved ranged from 3.7 to 3.9.

We performed another set of simulations to explore the potential impact of raising the level of detection. We explored the scenario where the current level of detection is at 0.40 (parameter set M). We performed multiple sets of simulations (one hundred runs each time), fixing the parameters on parameter set M, and then raising progressively the level of detection from 0.4 to 1, by 0.05 increments. The total number of episodes estimated by the model in the scenario of a 0.40 detection level (parameter set M) was 2,207. When the level of detection was raised to 1 (100% detection), the total number of episodes estimated was 1,751 (complete data in Appendix D).

Table 8: List of the thirteen sets of parameters that can fit the model to the observed data

Set	p_{det}	p_{inf}	β_i	β
A	1.00	0.085	0.40	0.0350
B	0.95	0.083	0.43	0.0350
C	0.90	0.080	0.45	0.0350
D	0.85	0.074	0.48	0.0325
E	0.80	0.069	0.50	0.0325
F	0.75	0.065	0.53	0.0325
G	0.70	0.062	0.55	0.0315
H	0.65	0.058	0.60	0.0300
I	0.60	0.053	0.65	0.0300
J	0.55	0.049	0.70	0.0280
K	0.50	0.045	0.80	0.0280
L	0.45	0.040	0.85	0.0270
M	0.40	0.035	0.99	0.0270

Table 9: Results of a one year course of the spreading model for each of the thirteen parameter sets (one hundred runs each time)

Parameter set (p_{det})	A (1)	B (0.95)	C (0.9)	D (0.85)	E (0.8)	F (0.75)	G (0.7)	H (0.65)	I (0.6)	J (0.55)	K (0.5)	L (0.45)	M (0.4)
Number of episodes (mean (sd))													
total	938 (130)	1041 (135)	1099 (129)	1100 (117)	1163 (125)	1314 (138)	1317 (121)	1385 (136)	1542 (136)	1573 (124)	1802 (147)	1861 (131)	2207 (134)
With infected cases													
all	183 (28)	196 (29)	201 (28)	186 (23)	186 (24)	199 (23)	192 (23)	190 (25)	192 (21)	186 (20)	193 (21)	178 (17)	187 (18)
imported	74 (8)	78 (8)	78 (9)	78 (8)	77 (8)	77 (9)	75 (9)	76 (9)	75 (8)	76 (9)	78 (8)	73 (8)	73 (8)
not imported	109 (27)	118 (27)	123 (26)	108 (22)	109 (22)	121 (22)	116 (21)	114 (22)	116 (21)	110 (19)	115 (20)	105 (16)	113 (15)
With colonized only, detected													
all	755 (105)	803 (103)	807 (95)	777 (84)	784 (84)	836 (94)	789 (74)	778 (77)	814 (74)	762 (61)	805 (70)	759 (56)	811 (54)
imported	317 (17)	324 (17)	322 (17)	332 (17)	326 (19)	324 (19)	315 (17)	323 (17)	326 (19)	319 (19)	330 (18)	317 (17)	328 (17)
not imported	438 (103)	479 (104)	485 (94)	445 (82)	458 (81)	512 (88)	475 (72)	455 (77)	488 (70)	444 (60)	475 (68)	441 (53)	483 (50)
With colonized only, undetected													
all	0 (0)	42 (9)	90 (15)	137 (19)	193 (25)	279 (32)	337 (33)	417 (44)	536 (51)	625 (56)	804 (66)	925 (71)	1209 (79)
imported	0 (0)	17 (4)	36 (7)	57 (8)	80 (9)	108 (10)	136 (10)	170 (14)	214 (14)	260 (15)	330 (18)	384 (18)	488 (19)
not imported	0 (0)	25 (8)	54 (13)	80 (16)	113 (23)	171 (32)	201 (30)	247 (40)	322 (47)	365 (53)	474 (63)	540 (68)	721 (77)
Patients per episode (mean (sd))													
all cases	8.2 (19.4)	7.6 (17.8)	7.6 (17.9)	6.9 (16.2)	7.6 (18.3)	6.9 (16.7)	6.9 (16.7)	6.8 (16.7)	6.3 (15.3)	5.8 (13.6)	6.1 (15.3)	6 (14.8)	5.7 (14.6)
infected*	1.1 (0.4)	1.1 (0.4)	1.1 (0.4)	1.1 (0.4)	1.1 (0.4)	1.1 (0.3)	1.1 (0.4)	1.1 (0.3)	1.1 (0.3)	1.1 (0.3)	1.1 (0.3)	1.1 (0.3)	1 (0.2)
% detected†	75.4 (31)	76.2 (30.5)	75.9 (30.7)	76.6 (30.1)	76.4 (30.3)	76.5 (30.4)	76.7 (29.9)	76.7 (30.3)	76.7 (30.2)	76.2 (30.4)	76.7 (30.2)	76.5 (30.2)	76.6 (30.2)
Duration of the episode, in days (median (IQR))													
if episode detected	9 (2 - 54)	8 (2 - 44)	8 (2 - 44)	9 (2 - 46)	9 (2 - 46)	9 (2 - 47)	9 (2 - 43)	9 (2 - 42)	8 (2 - 38)	9 (2 - 48)	9 (2 - 42)	9 (2 - 43)	9 (2 - 40)
if episode not detected	NA (NA - NA)	62.5 (25 - 102)	65 (19 - 126)	47.5 (13 - 113)	39 (12 - 124)	30 (11 - 124)	27 (10 - 124)	25 (10 - 121)	22 (9 - 108)	21 (9 - 94)	20 (9 - 66)	19 (8 - 53)	18 (8 - 44)
Epidemic chains (mean (sd))													
number of chains	89 (8)	96 (9)	106 (11)	111 (10)	120 (10)	128 (10)	139 (9)	150 (11)	172 (13)	183 (13)	208 (16)	220 (14)	258 (16)
length of a chain‡	4.8 (6)	4.6 (5.5)	4.9 (6.1)	4.5 (5.4)	4.8 (5.9)	4.8 (5.9)	4.6 (5.3)	4.5 (5.1)	4.5 (5.4)	4.4 (4.9)	4.4 (4.9)	4.3 (4.7)	4.5 (5.1)
facilities involved in a chain	3.8 (3.5)	3.8 (3.2)	3.9 (3.5)	3.8 (3.3)	3.9 (3.6)	3.9 (3.5)	3.8 (3.3)	3.7 (3.2)	3.8 (3.3)	3.7 (3)	3.7 (3.2)	3.7 (3)	3.7 (3.2)

* When at least one infected case occured during the episode

† Proportion of all the cases generated during the episode actually identified by the investigation, when the episode was detected.

‡ Mean number of transfers that occured within an epidemic chain

5 Discussion

This works sets a precedent in France by studying the spread of carbapenem-resistant Enterobacteriaceae outbreaks on the national network of healthcare facilities. We developed a stochastic, susceptible-colonized-infected spreading model, fitted on the 2015 national surveillance data. To our knowledge, this model is the first to study spread of pathogens in HCF on such a large scale. With 2,433 facilities, and up to 1,285,991 transfers recorded each year, it constitutes, to this day, the largest network used to model pathogen spread across healthcare facilities. In 2010, Donker et al.[65] reconstituted the Dutch healthcare network to study if different referral patterns between hospitals can influence rates of hospital-acquired infections like methicillin-resistant Staphylococcus aureus (MRSA). However, this network included only 98 facilities. In 2012, Donker et al.[66] studied the dispersal of hospital-acquired pathogens due to patient transfer on the United Kingdom-National Health Services network, which included only 146 facilities. Lee et al.[67,68] modeled the spread of MRSA and CRE outbreaks throughout Orange County, California, USA, which represents 32 hospitals.

After having tried different approaches and methods, we came to the conclusion that is was not possible to find a unique solution of parameters that could fit the data. This can be explained by the fact that we are lacking data for several of the main parameters. We do not know the true proportion of undetected episodes (by definition), and data on the risk of infection is scarce, and likely to be biased. Since both these parameters affect the spreading dynamics, a unique solution could have been found only by knowing, or hypothesizing one of them. Instead, we estimated a range of possible values.

We estimated that the risk of a patient developing a clinical infection once colonized by a CRE, in a healthcare setting, is between 3.5% and 8.5%. Previous work estimated this risk between 10% (Tamma et al.)[69] and 16.5% (Tischendorf et al.).[70] These differences are mainly due to the fact these previous estimations are solely based on episodes that were actually detected. If we assume that not all episodes are detected, the risk of infection is overestimated, since undetected episodes are more likely to not include any infected patient but only colonized ones. By modeling undetected episodes, we are able to propose a slightly less biased estimation of the risk of infection. The uncertainty in the estimation of this risk resides mainly in its dependence towards the level of detection, as those two parameters are positively correlated. We estimated that the overall probability of detecting an episode, if it includes no infected cases, can be no less than 40%. In this worst case scenario, the estimated number of undetected episodes was 1,209, that is 54.8% of all episodes (2.207). Assuming a baseline level of detection of 40%, we estimated that raising the level of detection up to 100% would only reduce the total number of episodes to 1,751 (20.7% reduction). This confirms, as it could have been expected, that the time before detection is more critical in the spreading dynamics than the level of detection itself. Moreover, the model showed that an important number of epidemic chains occur during outbreaks. The estimated number of epidemic chains occuring in one year ranged from 89 to 258. This suggests that the transfer of patients between HCF could play a critical role in the dynamics of CRE outbreaks.

However this work has several limitations. Some of them are inherent to the concept of modeling. Models, even though thoroughly designed, are always simplifications of real-world processes. Although many factors have been considered and implemented in this model, it cannot reproduce exactly the real spreading dynamics, and one must be careful when interpreting its results. Several important assumptions had to be made, either for simplification purposes, or for lack of better alternatives; and they must be addressed. The first main assumption is that if an episode includes an infected patient, it will always be detected. This assumption is needed in order to calibrate the model. If we assumed that the level of detection of episodes with infected cases was less than 100%, we would have to fit this level as another parameter, which would make the calibration significantly more difficult, if not impossible. We believe that this assumption is reasonable, as a patient with an infection would undergo several examinations to identify the causing agent, and even more so since CRE infections are often severe.[12] Second, we assumed that once a case is detected, it can no longer spread the pathogen, that is, it no longer participates in

the generation of new cases, both within and outside the HCF. Even though this might not be the case in reality, as control measures might be insufficient, or not applied with due diligence, this assumption was needed. Indeed, if we were to assume that cases, once detected, can still participate in the spreading dynamics, we would need to estimate to what extent, and this is likely to vary greatly depending on the healthcare facility. Third, we made the simplification that episodes originating from the community are always imported (the index case has traveled and/or been hospitalized in a foreign country in the past twelve month). This is linked to the fact that we are considering only direct transfers between HCF, and we are not, yet, considering the possibility for a case to be discharged from the HCF, and then being admitted to another HCF later on, still colonized, thus creating a new episode. Another limitation of the model is that we are estimating a global detection level (single p_{det}) for both imported and non-imported episodes. Although the estimated overall number of episodes detected was in an acceptable range (-6.8%; +5.4%), the number of imported episodes with infected cases estimated by the model was always slightly superior to the actual data, while the number of imported episodes with only colonized cases (detected) was slightly inferior to the actual data. This suggests that (while it 100% for episodes with infected cases), the level of detection must be different for imported and non-imported episodes when there are only colonized cases. It is likely that the level of detection of an episode with no infected cases is higher if the episode is imported, since the guidelines recommend systematic screening of patients having traveled and/or been hospitalized in a foreign country in the past twelve month.[47] Finally, one critical feature of the model that drives its dynamics, is the rate at which new cases are generated within the HCF ("within-hospital dissemination"), as it impacts directly the time before detection. This parameter was estimated from Cuzon et al.[62] and Tamma et al.,[69] but data on the matter is scarce, and further research is needed to better understand this phenomenon.

Despite these limitations, this work has many strengths. While being reasonably parsimonious, the model takes into account many factors. Especially, its ability to model pathogen propagation at different levels. Besides modeling intra-hospital propagation, it uses the national network of healthcare facilities to explore the role of patient transfers on CRE outbreaks. Although at a preliminary stage of its development, this model provides precious insights on the characteristics of such outbreaks. It confirms that cooperation between institutions is crucial, and, with further research, this tool could permit to devise new control strategies, at different scales, and test their potential impact and feasibility. By performing more numerous simulations, we could study patterns in outbreaks and epidemic chains, and identify sensitive facilities in the network, which we could target for specific interventions. It could help decision makers and stakeholders in the management of the CRE epidemic in France, at local, regional, or national levels. To this end, an interactive web application is currently being developed. It serves as an interactive visualization tool of the model. It would allow us to visualize and explore the network of healthcare facilities, to select parameters to perform, and visualize in real time, various simulations of CRE outbreaks, thus allowing to test different potential control measures and assess their impact. Finally, the scale at which the model operates is major asset. As the problem of anti-microbial resistance is growing worldwide, tools to model outbreaks across healthcare networks are increasingly popular. Our ability to reconstruct large networks of hospitals, with good computational efficiency, will drive European-level cooperations.

Although already useful, there are perspectives to further develop the model. The first one would be to have a more sophisticated "within-hospital dissemination" model. Such models are currently being developed by other teams in the SPHINx study, and those models could be merged in the near future to create a meta-model. Another perspective of amelioration would be to take into account not only direct transfers, but also patients that are discharged to the community, who can then be readmitted later on. However, doing so would imply the need to estimate the probability of a case still being colonized when readmitted (as done by Donker et al.).[65] A third possibility for improvement would be to incorporate trends. This could mean performing simulations over multiple years. Finally, another possible amelioration would be regarding the implementation of the detection processes, for example by allowing different levels of detection in different scenarios.

6 Conclusion

We constructed the network of healthcare facilities of metropolitan France over the years 2014-2016. The network was stable in its main characteristics and showed a high level of clustering. We then developed a stochastic, hospital-based, susceptible-colonized-infected model to reproduce the dynamics of carbapenem-resistant Enterobacteriaceae outbreaks across the 2015 network of HCF. Using this model, we estimated that the risk of patient colonized with CRE becoming infected, in a healthcare setting, was between 3.5% and 8.5%. We estimated that the current level of detection could be as low as 40%, which implies that the number of undetected episodes might be as high as 1,209. Assuming a baseline level of detection of 40%, we estimated that raising the level of detection up to 100% would only reduce the total number of episodes from 2,207 to 1,751 (20.7% reduction). The model showed that an important number of epidemic chains occur during outbreaks, suggesting that the transfer of patients between HCF could play a critical role in the dynamics of CRE outbreaks. To our knowledge, this model is the first to study spread of pathogens in HCF on such a large scale. It could be a valuable tool for further research, and for helping stakeholders in the management of the increasingly important issue of anti-microbial resistance.

7 References

1. Porter, J. R. Antony van leeuwenhoek: Tercentenary of his discovery of bacteria. *Bacteriol Rev* **40**, 260–269 (1976).

2. Forrest, R. D. Early history of wound treatment. *J R Soc Med* **75**, 198–205 (1982).

3. Williams, K. J. The introduction of 'chemotherapy' using arsphenamine - the first magic bullet. *J R Soc Med* **102**, 343–348 (2009).

4. Ullah, H. & Ali, S. Classification of antibacterial agents and their functions. in *Antibacterial agents* (ed. Kumavath, R. N.) (InTech, 2017). doi:10.5772/intechopen.68695

5. Magiorakos, A.-P. *et al.* Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection* **18**, 268–281 (2012).

6. World Health Organization. *Antimicrobial resistance: Global report on surveillance*. (World Health Organization, 2014).

7. World Health Organization. *Report on the burden of endemic health care-associated infection worldwide.* (World Health Organization, 2011).

8. Santé Publique France. *Enquête nationale de prévalence des infections nosocomiales et des traitements anti-infectieux en établissements de santé, france, mai-juin 2017.* (2018).

9. *The gammaproteobacteria*. (Springer, 2005).

10. Partridge, S. Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiol. Rev.* **35**, 820–855 (2011).

11. Stokes, H. & Gillings, M. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* **35**, 790–819 (2011).

12. Pilly, E. & Collège des universitaires de maladies infectieuses et tropicales (France). *Maladies infectieuses et tropicales*. (Alinéa plus, 2018).

13. Nordmann, P., Dortet, L. & Poirel, L. Carbapenem resistance in Enterobacteriaceae: Here is the storm! *Trends Mol Med* **18**, 263–272 (2012).

14. Pitout, J. D. & Laupland, K. B. Extended-spectrum beta-lactamase-producing enterobacteriaceae: An emerging public-health concern. *Lancet Infect Dis* **8**, 159–166 (2008).

15. Sheu, C.-C. *et al.* Management of infections caused by extended-spectrum β-lactamase-producing enterobacteriaceae: Current evidence and future prospects. *Expert Rev Anti Infect Ther* **16**, 205–218 (2018).

16. Bradford, P. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clin. Microbiol. Rev.* **14**, 933–951 (2001).

17. Ghafourian, S., Sadeghifard, N., Soheili, S. & Sekawi, Z. Extended spectrum beta-lactamases: Definition, classification and epidemiology. *Curr Issues Mol Biol* **17**, 11–21 (2015).

18. Du, B. *et al.* Extended-spectrum beta-lactamase-producing Escherichia coli and Klebsiella pneumoniae bloodstream infection: risk factors and clinical outcome. *Intensive Care Med* **28**, 1718–1723 (2002).

19. Paterson, D. L. *et al.* Antibiotic therapy for Klebsiella pneumoniae bacteremia: implications of production of extended-spectrum beta-lactamases. *Clin. Infect. Dis.* **39**, 31–37 (2004).

20. Ramphal, R. & Ambrose, P. Extended-spectrum beta-lactamases and clinical outcomes: current data. *Clin. Infect. Dis.*

42 Suppl 4, S164–172 (2006).

21. Collins, V. L. *et al.* Efficacy of ertapenem for treatment of bloodstream infections caused by extended-spectrum beta-lactamase-producing Enterobacteriaceae. *Antimicrob. Agents Chemother.* **56**, 2173–2177 (2012).

22. Vardakas, K. Z., Tansarli, G. S., Rafailidis, P. I. & Falagas, M. E. Carbapenems versus alternative antibiotics for the treatment of bacteraemia due to Enterobacteriaceae producing extended-spectrum beta-lactamases: a systematic review and meta-analysis. *J. Antimicrob. Chemother.* **67**, 2793–2803 (2012).

23. Tamma, P. D. *et al.* Carbapenem therapy is associated with improved survival compared with piperacillin-tazobactam for patients with extended-spectrum beta-lactamase bacteremia. *Clin. Infect. Dis.* **60**, 1319–1325 (2015).

24. D'Angelo, R. G., Johnson, J. K., Bork, J. T. & Heil, E. L. Treatment options for extended-spectrum beta-lactamase (ESBL) and AmpC-producing bacteria. *Expert Opin Pharmacother* **17**, 953–967 (2016).

25. Tamma, P. D. & Rodriguez-Bano, J. The Use of Noncarbapenem beta-Lactams for the Treatment of Extended-Spectrum beta-Lactamase Infections. *Clin. Infect. Dis.* **64**, 972–980 (2017).

26. Harris, P. N., Tambyah, P. A. & Paterson, D. L. Beta-lactam and beta-lactamase inhibitor combinations in the treatment of extended-spectrum beta-lactamase producing Enterobacteriaceae: time for a reappraisal in the era of few antibiotic options? *Lancet Infect Dis* **15**, 475–485 (2015).

27. Paterson, D. & Bonomo, R. Extended-spectrum beta-lactamases: a clinical update. *Clin. Microbiol. Rev.* **18**, 657–686 (2005).

28. Malloy, A. & Campos, J. Extended-spectrum beta-lactamases: a brief clinical update. *Pediatr. Infect. Dis. J.* **30**, 1092–1093 (2011).

29. Yigit, H. *et al.* Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a carbapenem-resistant strain of Klebsiella pneumoniae. *Antimicrob. Agents Chemother.* **45**, 1151–1161 (2001).

30. Nordmann, P., Cuzon, G. & Naas, T. The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. *Lancet Infect Dis* **9**, 228–236 (2009).

31. Magiorakos, A.-P., Suetens, C., Monnet, D. L., Gagliotti, C. & Heuer, O. E. The rise of carbapenem resistance in europe: Just the tip of the iceberg? *Antimicrob Resist Infect Control* **2**, 6 (2013).

32. Boyle, D. & Zembower, T. Epidemiology and Management of Emerging Drug-Resistant Gram-Negative Bacteria: Extended-Spectrum Beta-Lactamases and Beyond. *Urol. Clin. North Am.* **42**, 493–505 (2015).

33. Albiger, B. *et al.* Carbapenemase-producing enterobacteriaceae in europe: Assessment by national experts from 38 countries, may 2015. *Eurosurveillance* **20**, 30062 (2015).

34. European Centre for Disease Prevention and Control. *Annual report of the european antimicrobial resistance surveillance network (ears-net) 2017.* (2018).

35. Grundmann, H. *et al.* Occurrence of carbapenemase-producing Klebsiella pneumoniae and Escherichia coli in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. *Lancet Infect Dis* **17**, 153–163 (2017).

36. Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* **19**, 56–66 (2019).

37. Logan, L. K. & Weinstein, R. A. The epidemiology of carbapenem-resistant enterobacteriaceae: The impact and evolution

of a global menace. *J Infect Dis* **215**, S28–S36 (2017).

38. Trecarichi, E. & Tumbarello, M. Therapeutic options for carbapenem-resistant Enterobacteriaceae infections. *Virulence* **8**, 470–484 (2017).

39. Rodriguez-Bano, J., Gutierrez-Gutierrez, B., Machuca, I. & Pascual, A. Treatment of Infections Caused by Extended-Spectrum-Beta-Lactamase-, AmpC-, and Carbapenemase-Producing Enterobacteriaceae. *Clin. Microbiol. Rev.* **31**, (2018).

40. Papst, L. *et al.* Antibiotic treatment of infections caused by carbapenem-resistant Gram-negative bacilli: an international ESCMID cross-sectional survey among infectious diseases specialists practicing in large hospitals. *Clin. Microbiol. Infect.* **24**, 1070–1076 (2018).

41. Sheu, C.-C., Chang, Y.-T., Lin, S.-Y., Chen, Y.-H. & Hsueh, P.-R. Infections caused by carbapenem-resistant enterobacteriaceae: An update on therapeutic options. *Front Microbiol* **10**, (2019).

42. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* **18**, 318–327 (2018).

43. Kassis-Chikhani, N. *et al.* First outbreak of multidrug-resistant Klebsiella pneumoniae carrying blaVIM-1 and blaSHV-5 in a French university hospital. *J. Antimicrob. Chemother.* **57**, 142–145 (2006).

44. Pontiès, V. *et al. Surveillance des epc en france : Bilan 2004-2016*. (2017).

45. Plésiat, P., Cattoir, V., Bonnet, R. & Naas, T. *Centre national de référence de la résistance aux antibiotiques : Rapport d'activité 2017*. (2018).

46. Vaux, S. *et al.* Emergence of carbapenemase-producing enterobacteriaceae in france, 2004 to 2011. *Eurosurveillance* **16**, (2011).

47. Ministère des affaires sociales, de la santé et des droits des femmes. *Propias. programme national d'actions de prévention des infections associées aux soins*. (2015).

48. Assab, R. *et al.* Mathematical models of infection transmission in healthcare settings: Recent advances from the use of network structured data. *Current Opinion in Infectious Diseases* **30**, 410–418 (2017).

49. Bavelas, A. A mathematical model for group structures. *Human Organization* **7**, 16–30 (1948).

50. Shimbel, A. Structural parameters of communication networks. *The Bulletin of Mathematical Biophysics* **15**, 501–507 (1953).

51. Shaw, M. E. Group structure and the behavior of individuals in small groups. *The Journal of Psychology* **38**, 139–149 (1954).

52. Marriott, M. & Cohn, B. S. Networks and centres in the integration of indian civilization. *Journal of Social Research* **1**, 1–9 (1958).

53. Anthonisse, J. M. The rush in a graph. *Amsterdam: Mathematische Centrum* (1971).

54. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* 35–41 (1977).

55. Freeman, L. C. Centrality in social networks conceptual clarification. *Social networks* **1**, 215–239 (1978).

56. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**,

1695 (2006).

57. Brandes, U. A faster algorithm for betweenness centrality. *Journal of mathematical sociology* **25**, 163–177 (2001).

58. Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* **46**, 604–632 (1999).

59. Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Physical review E* **70**, 066111 (2004).

60. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* **105**, 1118–1123 (2008).

61. Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *The European Physical Journal Special Topics* **178**, 13–23 (2009).

62. Cuzon, G., Ouanich, J., Gondret, R., Naas, T. & Nordmann, P. Outbreak of OXA-48-positive carbapenem-resistant Klebsiella pneumoniae isolates in France. *Antimicrob. Agents Chemother.* **55**, 2420–2423 (2011).

63. Davido, B. *et al.* Germs of thrones - spontaneous decolonization of Carbapenem-Resistant Enterobacteriaceae (CRE) and Vancomycin-Resistant Enterococci (VRE) in Western Europe: is this myth or reality? *Antimicrob Resist Infect Control* **7**, 100 (2018).

64. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2019).

65. Donker, T., Wallinga, J. & Grundmann, H. Patient referral patterns and the spread of hospital-acquired infections through national health care networks. *PLoS computational biology* **6**, e1000715 (2010).

66. Donker, T., Wallinga, J., Slack, R. & Grundmann, H. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. *PLoS One* **7**, e35002 (2012).

67. Lee, B. Y. *et al.* Modeling the spread of methicillin-resistant staphylococcus aureus (mrsa) outbreaks throughout the hospitals in orange county, california. *Infection Control & Hospital Epidemiology* **32**, 562–572 (2011).

68. Lee, B. Y. *et al.* The potential trajectory of carbapenem-resistant enterobacteriaceae, an emerging threat to health-care facilities, and the impact of the centers for disease control and prevention toolkit. *American journal of epidemiology* **183**, 471–479 (2016).

69. Tamma, P. D. *et al.* The likelihood of developing a carbapenem-resistant enterobacteriaceae infection during the hospital stay. *Antimicrobial agents and chemotherapy* AAC–00757 (2019).

70. Tischendorf, J., Avila, R. A. de & Safdar, N. Risk of infection following colonization with carbapenem-resistant enterobactericeae: A systematic review. *American journal of infection control* **44**, 539–543 (2016).

List of Figures

1 Monthly incidence of CPE episodes in France, 2009 - 2016: figure from Vaux et al. . . . 4
2 Graph of a 5 nodes network . . . 8
3 Repartition of the number of episodes and the relationships between what is observed, the reality and what is modeled. . . . 10
4 Representation of the model flow. . . . 12
5 Distribution of the total degree of healthcare facilities in the 2015 network. . . . 16
6 Distribution of the number of patients sent and received by healthcare facilities in the 2015 network. . . . 16
7 Geographic repartition of the clusters of healthcare facilities identified by the Greedy (left) and Map equation (right) algorithms in the 2015 network. Each cluster is represented by a different color and delimited by a polygon. . . . 17
8 Graph of the 2015 network of clusters. Each node represents a cluster and each edge represents a connection between the two clusters (patients transfered). The node's diameter is proportional to the cluster's size, and the edge's width is proportional to the number of patients transfered. . . . 20
9 Distribution of the total degree of healthcare facilities in the 2014 network. . . . 33
10 Distribution of the total degree of healthcare facilities in the 2016 network. . . . 33
11 Distribution of the number of patients sent and received by healthcare facilities in the 2014 network. . . . 34
12 Distribution of the number of patients sent and received by healthcare facilities in the 2016 network. . . . 34

List of Tables

1 Adjacency matrix of a 5 nodes network. The rows contain the facilities of origin, and the columns contain the target facilities. Each cell contains the number of patients transferred. . . . 6
2 Main outputs of the model. . . . 9
3 Main assumptions of the model . . . 13
4 Parameters of the model, their role, and their source . . . 14
5 Main characteristics of the 2014, 2015, and 2016 MCO/SSR hospitals networks. . . . 15
6 List of clusters identified by the Greedy algorithm in the 2015 network, with the main city of the cluster, the cluster's size, and the HCF which had the greatest hub and authority scores of the cluster . . . 18
7 For each cluster identified by the Greedy algorithm, the number of transfers recorded within the cluster, outside the cluster, and the proportion of all the transfers recorded by HCF of the cluster that were within the cluster (2015). . . . 19
8 List of the thirteen sets of parameters that can fit the model to the observed data . . . 21
9 Results of a one year course of the spreading model for each of the thirteen parameter sets (one hundred runs each time) . . . 22
10 Fifty hospitals with the greatest total degree in 2015, and their respective ranking in others centrality measures 32
11 List of clusters identified by the Greedy algorithm in the 2014 network, with the main city of the cluster, the cluster's size, and the HCF which had the greatest hub and authority scores of the cluster . . . 35
12 List of clusters identified by the Greedy algorithm in the 2016 network, with the main city of the cluster, the cluster's size, and the HCF which had the greatest hub and authority scores of the cluster . . . 36

13 Potential impact of raising the level of detection of episodes, assuming a current level of detection of 40% (one year course of simulation, one hundred simulations each time) 37

8 Appendices

8.1 Appendix A

Table 10: Fifty hospitals with the greatest total degree in 2015, and their respective ranking in others centrality measures

Facility	Total degree	Betweenness	Closeness	Patients received	Patients sent
GROUPE HOSP. PITIE-SALPETRIERE (AP-HP)	1	4	2	10	7
GPE HOSP COCHIN-SAINT VINCENT DE PAUL	2	86	82	44	26
GPE HOSP BROUSSAIS-HEGP	3	15	18	51	25
G.I.H. BICHAT / CLAUDE BERNARD (AP-HP)	4	56	54	49	21
HOPITAL BICETRE (AP-HP)	5	9	6	42	31
HOPITAL LA TIMONE ADULTES	6	8	373	13	4
GPE HOSP HENRI MONDOR-ALBERT CHENEVIER	7	41	32	41	22
GPE HOSP LARIBOISIERE-FERNAND WIDAL	8	129	39	94	34
HOPITAL DE PURPAN CHU TOULOUSE	9	34	742	21	8
HOPITAL SAINT-ANTOINE (AP-HP)	10	110	34	65	35
HOPITAL NECKER ENFANTS MALADES (AP-HP)	11	1	1	56	154
CHU PELLEGRIN	12	5	20	5	11
HOPITAL NORD (MARSEILLE)	13	79	1237	33	32
HOPITAL DE RANGUEIL CHU TOULOUSE	14	38	449	15	16
HOPITAL TENON (AP-HP)	15	46	13	112	74
HOPITAL SAINT-LOUIS (AP-HP)	16	63	72	93	128
HOPITAL FOCH	17	22	87	79	45
HOPITAL BEAUJON (AP-HP)	18	45	21	177	75
HOPITAL EDOUARD HERRIOT	19	28	120	35	12
HOPITAL DE LA CROIX-ROUSSE	20	32	215	32	37
INSTITUT MUTUALISTE MONTSOURIS	21	343	145	114	104
CH LYON SUD	22	53	123	40	33
HOP SALENGRO - HOPITAL B CHR LILLE	23	73	727	8	1
HÔPITAL PRIVÉ DES PEUPLIERS	24	214	107	39	124
C.H.U. NANTES HÔTEL-DIEU ET HME	25	19	79	27	6
C.H.R. PONTCHAILLOU-RENNES	26	27	440	22	5
HOPITAUX DE BRABOIS CHU NANCY	27	6	33	3	40
GPE HOSP SAINT-JOSEPH	28	246	138	146	38
GHU PARIS-ILE-DE-FRANCE OUEST SITE AMBROISE PARE	29	156	110	174	50
HOPITAL ST JOSEPH MARSEILLE	30	281	718	92	56
HOPITAL PRIVE CLAIRVAL	31	108	799	108	88
HOPITAL DE HAUT LEVEQUE	32	37	279	11	41
CH VERSAILLES ANDRE MIGNOT	33	177	113	156	18
HÔPITAL LE BOCAGE CHU DIJON	34	11	11	6	13
GROUPE HOSPITALIER HOPITAUX UNIVERSITAIRES PARIS-ILE-DE-FRANCE OUEST	35	162	90	181	303
HOPITAL ANTOINE-BECLERE (AP-HP)	36	321	73	172	61
SA CLINIQUE PASTEUR	37	54	560	62	95
HOPITAL CHARLES NICOLLE CHU ROUEN	38	30	45	7	3
INSTITUT GUSTAVE ROUSSY	39	39	37	139	135
CH SUD FRANCILIEN SITE JEAN JAURES	40	417	175	66	55
HOPITAL DE LA CONCEPTION	41	175	1490	34	134
CHU COTE DE NACRE - CAEN	42	18	65	16	14
HOPITAL NORD (ST ETIENNE)	43	58	308	19	17
HOPITAL NORD (GRENOBLE)	44	2	5	26	28
HOPITAL PRIVE JEAN MERMOZ	45	109	430	173	115
CLINIQUE CHIRURG ALLERAY-LABROUSTE	46	164	27	38	244
HOPITAL PRIVE JACQUES CARTIER	47	33	8	222	122
CLINIQUE DU TONKIN	48	24	398	225	158
HOPITAL SUD CHU AMIENS	49	29	209	20	30
HOPITAL LAPEYRONIE CHU MONTPELLIER	50	104	890	54	2

8.2 Appendix B

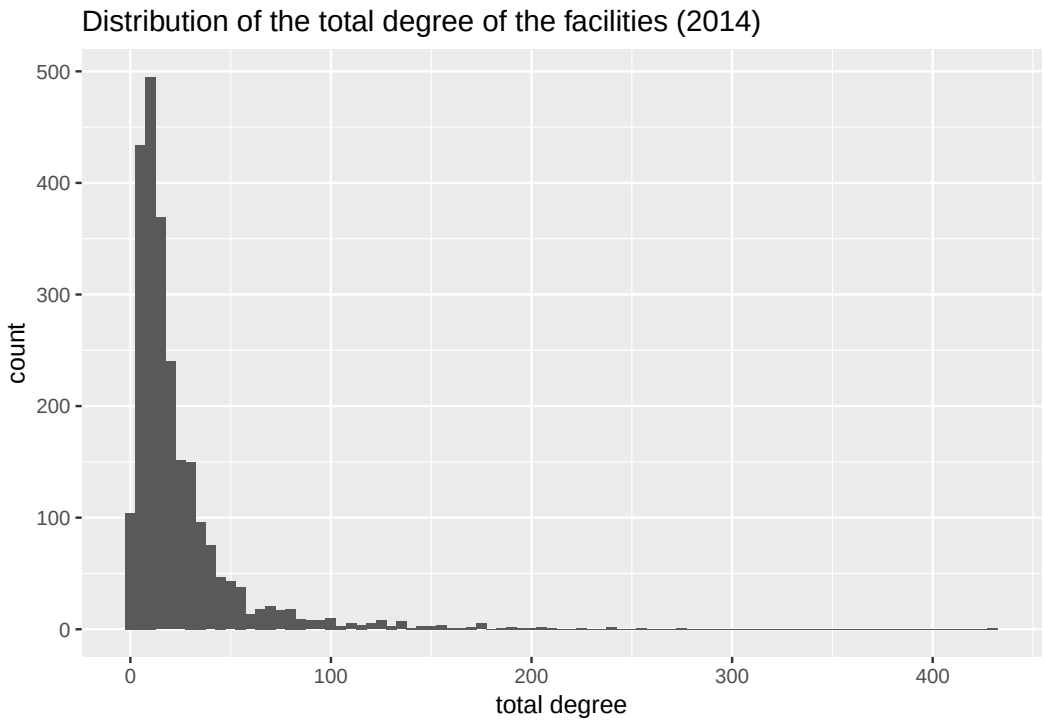

Figure 9: Distribution of the total degree of healthcare facilities in the 2014 network.

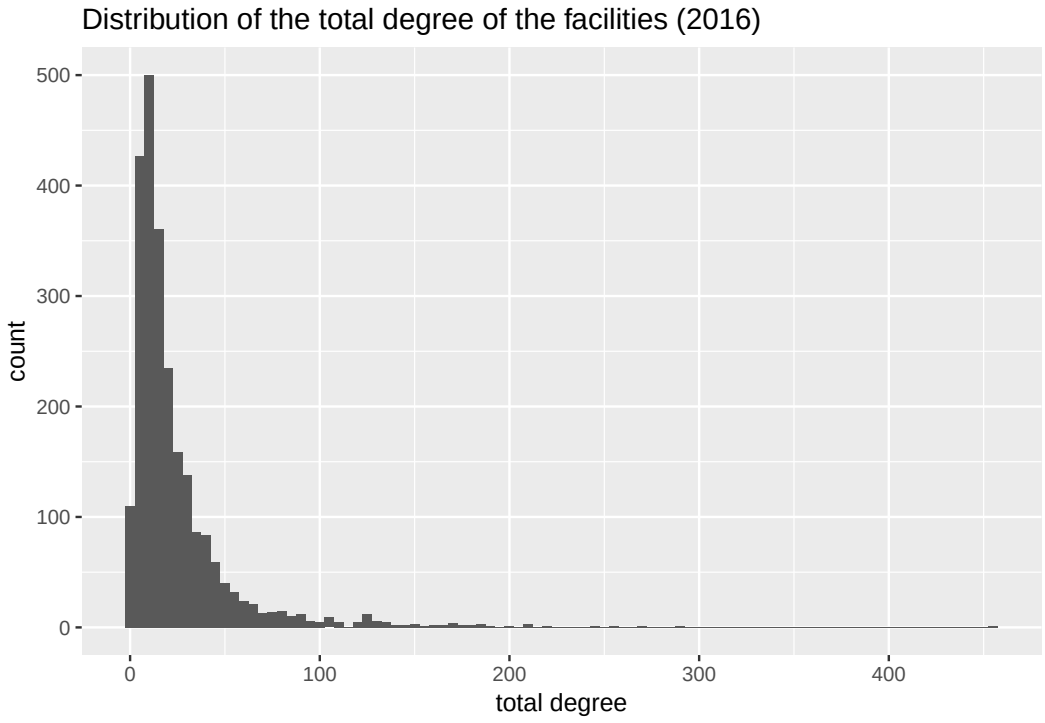

Figure 10: Distribution of the total degree of healthcare facilities in the 2016 network.

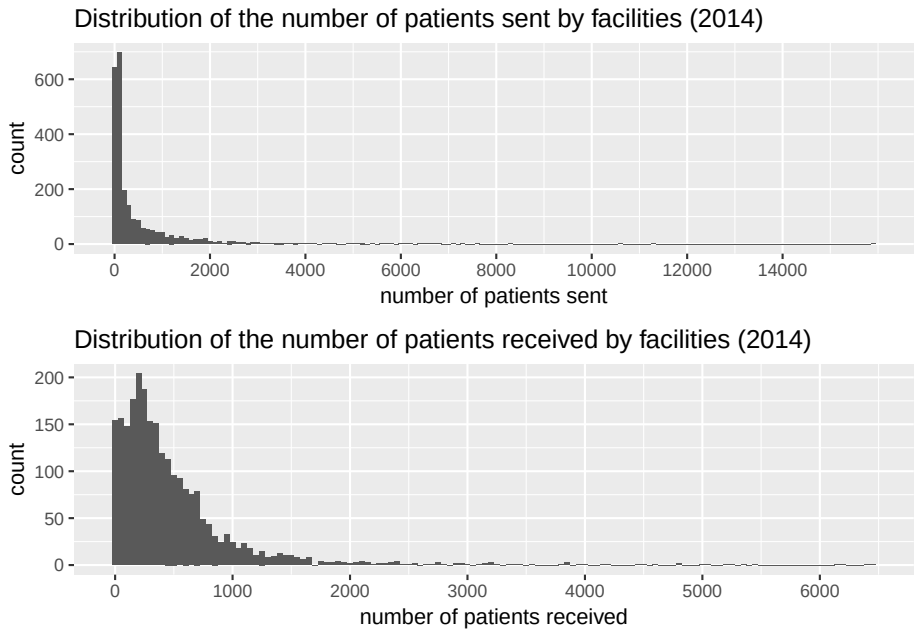

Figure 11: Distribution of the number of patients sent and received by healthcare facilities in the 2014 network.

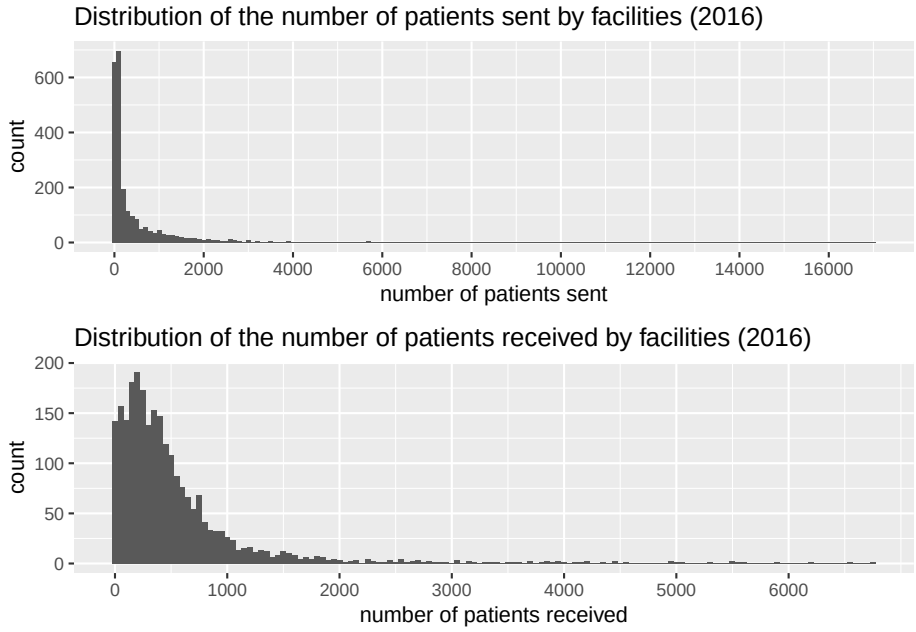

Figure 12: Distribution of the number of patients sent and received by healthcare facilities in the 2016 network.

8.3 Appendix C

Table 11: List of clusters identified by the Greedy algorithm in the 2014 network, with the main city of the cluster, the cluster's size, and the HCF which had the greatest hub and authority scores of the cluster

Main city	Cluster size	Hub	Authority
Paris	342	GPE HOSP BROUSSAIS-HEGP	HOPITAL CORENTIN CELTON (AP-HP)
Lyon	202	HOPITAL EDOUARD HERRIOT	HOPITAL DE LA CROIX-ROUSSE
Marseille	193	HOPITAL LA TIMONE ADULTES	HOPITAL LA TIMONE ADULTES
Bordeaux	147	CHU PELLEGRIN	HOPITAL DE HAUT LEVEQUE
Lille	132	HOP SALENGRO - HOPITAL B CHR LILLE	HOP CLAUDE HURIEZ CHR LILLE
Toulouse	132	HOPITAL DE RANGUEIL CHU TOULOUSE	HOPITAL DE PURPAN CHU TOULOUSE
Montpellier	120	HOPITAL LAPEYRONIE CHU MONTPELLIER	HOPITAL ARNAUD DE VILLENEUVE CHU MPT
Dijon	119	HÔPITAL LE BOCAGE CHU DIJON	MÉDECINE À CHAMPMAILLOT CHU DIJON
Reims	116	HOPITAL MAISON BLANCHE CHR REIMS	HOPITAL ROBERT DEBRE CHR REIMS
Tours	94	C.H.R.U. -TROUSSEAU-	C.H.R.U. BRETONNEAU
Nancy	92	HOPITAL CENTRAL CHU NANCY	HOPITAUX DE BRABOIS CHU NANCY
Rennes	88	C.H.R. PONTCHAILLOU-RENNES	POLE GERIATRIQUE RENNAIS
Strasbourg	87	CHU STRASBOURG / HOP HAUTEPIERRE	CHU DE STRASBOURG /NOUVEL HOPITAL CIVIL
Nantes	71	C.H.U. NANTES HÔTEL-DIEU ET HME	CHU DE NANTES HOPITAL G. R. LAENNEC
Angers	67	CHU D' ANGERS :SITE LARREY	CHU D' ANGERS:CENTRE DE SSR
Nice	66	HOPITAL SAINT ROCH DU CHU DE NICE	HOPITAL PASTEUR DU CHU DE NICE
Rouen	65	HOPITAL CHARLES NICOLLE CHU ROUEN	HOPITAL DE BOIS-GUILLAUME CHU ROUEN
Grenoble	64	HOPITAL NORD (GRENOBLE)	HOPITAL SUD
Caen	61	CHU COTE DE NACRE - CAEN	CHR GEORGES CLEMENCEAU - CAEN
Poitiers	58	GROUPE HOSP. LA ROCHELLE-RE-AUNIS	CHU LA MILETRIE
Clermont-Ferrand	47	CHU G. MONTPIED	CHU ESTAING
Limoges	36	C H U DUPUYTREN LIMOGES	HOPITAL JEAN REBEYROL LIMOGES
Brest	34	CHRU HOPITAL CAVALE BLANCHE	CENTRE DE CURE MED.& CONVALESC.

Table 12: List of clusters identified by the Greedy algorithm in the 2016 network, with the main city of the cluster, the cluster's size, and the HCF which had the greatest hub and authority scores of the cluster

Main city	Cluster size	Hub	Authority
Paris-Créteil	346	GROUPE HOSP. PITIE-SALPETRIERE (AP-HP)	CHARLES-FOIX - JEAN-ROSTAND GHU EST
Lyon	189	HOPITAL EDOUARD HERRIOT	CLINIQUE MUTUALISTE DE LYON
Marseille	166	HOPITAL LA TIMONE ADULTES	HOPITAL DE LA CONCEPTION
Bordeaux	148	CHU PELLEGRIN	HOPITAL DE HAUT LEVEQUE
Lille	132	HOP SALENGRO - HOPITAL B CHR LILLE	HOP CLAUDE HURIEZ CHR LILLE
Toulouse	132	HOPITAL DE PURPAN CHU TOULOUSE	HOPITAL DE RANGUEIL CHU TOULOUSE
Montpellier	126	HOPITAL LAPEYRONIE CHU MONTPELLIER	HOPITAL ARNAUD DE VILLENEUVE CHU MPT
Reims	120	HOPITAL MAISON BLANCHE CHR REIMS	HOPITAL ROBERT DEBRE CHR REIMS
Caen	115	CHU COTE DE NACRE - CAEN	KORIAN COTE NORMANDE
Rennes	94	C.H.R. PONTCHAILLOU-RENNES	POLE GERIATRIQUE RENNAIS
Dijon	92	HÔPITAL LE BOCAGE CHU DIJON	MÉDECINE À CHAMPMAILLOT CHU DIJON
Tours	92	C.H.R.U. -TROUSSEAU-	C.H.R.U. BRETONNEAU
Nancy	92	HOPITAL CENTRAL CHU NANCY	HOPITAUX DE BRABOIS CHU NANCY
Strasbourg	88	CHU STRASBOURG / HOP HAUTEPIERRE	CHU DE STRASBOURG /NOUVEL HOPITAL CIVIL
Nantes	69	C.H.U. NANTES HÔTEL-DIEU ET HME	CHU DE NANTES HOPITAL G. R. LAENNEC
Nice	69	HOPITAL SAINT ROCH DU CHU DE NICE	HOPITAL PASTEUR DU CHU DE NICE
Rouen	65	HOPITAL CHARLES NICOLLE CHU ROUEN	HOPITAL DE BOIS-GUILLAUME CHU ROUEN
Grenoble	60	HOPITAL NORD (GRENOBLE)	HOPITAL SUD
Poitiers	57	GROUPE HOSP. LA ROCHELLE-RE-AUNIS	CHU LA MILETRIE
Clermont-Ferrand	47	CHU G. MONTPIED	CHU ESTAING
Vesoul-Besançon	40	CH VESOUL	CHU JEAN MINJOZ
Limoges	36	C H U DUPUYTREN LIMOGES	HOPITAL JEAN REBEYROL LIMOGES
Brest	33	CHRU HOPITAL CAVALE BLANCHE	CENTRE DE CURE MED.& CONVALESC.
Corse	19	CH ND LA MISERICORDE	C.R.F. ET MAISON DE REPOS DU FINOSELLO

8.4 Appendix D

Table 13: Potential impact of raising the level of detection of episodes, assuming a current level of detection of 40% (one year course of simulation, one hundred simulations each time)

Detection level	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
Number of episodes (mean (sd))												
total	2176 (151)	2101 (139)	2073 (146)	2042 (142)	2020 (132)	1955 (115)	1913 (123)	1883 (121)	1834 (129)	1821 (117)	1777 (103)	1751 (112)
With infected cases												
all	183 (17)	176 (15)	175 (17)	171 (20)	171 (18)	165 (15)	159 (18)	158 (15)	155 (16)	152 (16)	146 (16)	146 (13)
imported	74 (8)	73 (8)	74 (9)	73 (8)	75 (9)	77 (9)	75 (10)	77 (9)	76 (8)	77 (8)	76 (9)	77 (8)
not imported	109 (15)	102 (14)	101 (15)	97 (17)	96 (15)	88 (13)	84 (14)	81 (14)	79 (14)	76 (13)	70 (12)	69 (11)
With colonized only, detected												
all	896 (67)	967 (66)	1047 (76)	1124 (76)	1207 (77)	1254 (76)	1313 (86)	1382 (91)	1427 (103)	1503 (97)	1549 (90)	1605 (103)
imported	366 (19)	410 (21)	455 (22)	498 (22)	543 (21)	586 (25)	629 (26)	678 (25)	720 (28)	771 (30)	813 (25)	856 (28)
not imported	529 (63)	557 (65)	592 (74)	626 (78)	664 (79)	667 (76)	685 (82)	704 (87)	707 (97)	732 (97)	736 (91)	749 (107)
With colonized only, undetected												
all	1097 (80)	958 (69)	851 (65)	747 (57)	642 (49)	537 (39)	440 (32)	343 (28)	252 (23)	166 (14)	81 (10)	0 (0)
imported	450 (18)	409 (21)	368 (18)	330 (20)	291 (14)	249 (15)	210 (14)	168 (12)	126 (12)	84 (8)	42 (6)	0 (0)
not imported	647 (80)	549 (66)	483 (65)	417 (52)	351 (46)	287 (38)	230 (30)	175 (25)	126 (19)	81 (12)	39 (7)	0 (0)
Patients per episode (mean (sd))												
all cases	5.8 (14.6)	5.5 (13.1)	6 (15.5)	6 (15.2)	5.6 (13.7)	5.8 (14.2)	6.2 (15.9)	6 (15)	6.2 (15.4)	6.1 (15.3)	6.1 (15.4)	6.2 (15.5)
infected*	1.1 (0.3)	1.1 (0.3)	1.1 (0.4)	1 (0.2)	1 (0.2)	1 (0.2)	1 (0.2)	1.1 (0.3)	1 (0.2)	1 (0.2)	1.1 (0.3)	1.1 (0.3)
% detected†	77.6 (29.6)	78.1 (29.3)	78.5 (29.1)	78.5 (29.1)	79.2 (28.4)	79.7 (28.3)	79.5 (28.4)	79.9 (28.1)	80.2 (27.9)	80.6 (27.5)	80.4 (27.8)	80.3 (27.9)
Duration of the episode (median (IQR))												
if episode detected	8 (2 - 38)	8 (2 - 33)	8 (2 - 31)	8 (2 - 30)	8 (2 - 27)	8 (2 - 27)	8 (2 - 27)	8 (2 - 27)	7 (2 - 25)	8 (2 - 25)	7 (2 - 24)	7 (2 - 24)
if episode not detected	18 (8 - 49)	19 (8 - 51)	20 (9 - 64)	20 (9 - 65)	21 (9 - 86)	22 (9 - 103)	25 (10 - 120)	27 (10 - 118)	31 (11 - 120)	47.5 (13 - 126)	63 (19 - 112)	NA (NA - NA)

* When at least one infected case occured during the episode

† Proportion of all the cases generated during the episode actually identified by the investigation, when the episode was detected.

9 Summary in French

Analyse des épidémies d'entérobactéries résistantes aux carbapenems sur le réseau des établissements de santé de France métropolitaine.

Introduction. La propagation des entérobactéries résistantes aux carbapenems (ERC) s'accélère à une vitesse alarmante, et menance d'ores et déjà les systèmes de santé du monde. En France, la tendance rapidement croissante du nombre d'épisodes d'ERC inquiète. L'objectif de cette étude est de proposer un modèle mathématique de simulation d'épidémies d'ERC sur le réseau français des établissements de santé (EDS).

Methodes. À partir de la base de données PMSI (Programme de médicalisation des systèmes d'information), nous avons construit le réseau des EDS de France métropolitaine pour les années 2014 à 2016. Nous avons ensuite développé un modèle stochastique susceptible-colonisé-infecté pour reproduire la dynamique des épidemies d'ERC sur le réseau de 2015. Ce modèle, dont l'unité est l'EDS, modélise à la fois, la propagation intra-établissement, inter-établissements (via transferts de patients), et entre établissements et communauté. Le modèle a été calibré en utilisant les données de surveillance Santé Publique France de 2015.

Résultats. Le réseau comprenait 2,433 EDS pour un maximum de 1 285 991 transferts enregistrés en 2016. Ses caractéristiques étaient stable sur les trois années, la principale étant son découpage en grappes d'EDS relativement indépendants. Nous avons estimé que le risque d'infection après colonisation à ERC était compris entre 3.5% et 8.5%. Nous avons estimé que le niveau de détection actuel des épisodes était compris entre 100% et 40% (soit potentiellement 1 209 épisodes non détectés). En faisant l'hypothèse que le niveau de détection actuel était de 40%, nous avons estimé qu'une amélioration vers un niveau de détection optimal (100%) ne réduirait le nombre total d'épisodes que de 20.7% (2 207 à 1 751).

Conclusions. Ce model suggère que le transfert de patients colonisés entre EDS pourrait avoir un rôle majeur dans la dynamique des épidémies d'ERC. À notre connaissance, il s'agit du premier modèle capable d'étudier la propagation de pathogènes au sein d'EDS à une telle échelle. Il peut constituer un outil de choix pour approfondir notre connaissance de ces phénomènes, et assister les pouvoirs publics face aux enjeux majeurs de l'antibio-résistance.

The end.