**MSc Thesis APPLIED MATHEMATICS**

**Specialisation Computational Science and Engineering**

# "Individual based models for the spread of MRSA clonal complexes in a population"


Eva van Tegelen
4725557

**Delft University of Technology**

**Faculty of Electrical Engineering, Mathematics & Computer Science**

**Numerical analysis group, Mathematics**


**Supervisors**

Prof.dr.ir. C. Vuik (TU Delft)

Prof.dr. S.J. de Vlas (Erasmus MC)

**Other members of the committee**

Dr.ir. M. Keijzer (TU Delft)

July 7, 2022
Delft

# ABSTRACT

A big proportion of hospital-associated infections caused by *Staphylococcus aureus* can be attributed to Methicillin-resistant *S. aureus* (MRSA). Many countries take interventions to try and minimise the spread of MRSA. Interventions, such as a search-and-destroy policy and restrictive antibiotics use, have proven to be effective. Different strains of MRSA are grouped into clonal complexes (CCs) by their similarity to a central allelic profile. Separate MRSA CCs have evolved independently over time. In most countries a limited number of CCs is responsible for the prevalence in the population. In each country different CCs are present and the exact reasons why these CCs are successful in the specific countries is unknown. The aim of this project was therefore to study two different models, one implemented in R and one implemented in Java, that both simulate the spread of multiple MRSA CCs in a population. The two individual based models (IBMs) considered in this research produce similar results when setting them side-by-side using simple model set-ups. The R model turned out to be computationally expensive and very restrictive, where on the other hand, the Java model was much faster and more extensive. Consequently, the Java model was used to simulate the spread of MRSA CCs in a more advanced setting. Although the model set-up and population parameters were not yet realistic, it demonstrated some interesting findings. The general observation was that CCs with higher antibiotic resistance contribute most of the MRSA infections. The model showed that intervention by means of a search-and-destroy policy can lower the overall prevalence in a population significantly and create more variation between the CCs present. Since the model set-ups adopted during this research most likely do not completely agree with the biological processes, populations and interventions in the real world, future research should determine whether the obtained exploratory results also hold true in more representative populations.

# PREFACE

This work focuses on modelling the spread of MRSA clonal complexes in a human population. Two different individual based models are intensively studied and compared. One of the models was furthermore used to study the spread of MRSA in a fictional population. The results of this research illustrate which aspects influence the spread of MRSA clonal complexes and how a search-and-destroy policy might effect this spread.

This project was done in collaboration with the Department of Public Health of Erasmus MC. I want to thank the Infectious disease group for the opportunity to further explore my interest in epidemiological modelling. I enjoyed the weekly section meetings very much and I learned a lot about what goes on behind the scenes in controlling infectious diseases.

Above all, I would like to thank my two supervisors, Kees Vuik and Sake de Vlas, for their guidance and advice throughout this project. I want to thank Kees for the weekly meetings and the structure he provided during the entirety of this research project. He was always willing to give feedback on my written work and was very invested in the project. I want to thank Sake for the many things he taught me about infectious disease modelling. During this project he challenged me to step out of my comfort zone and made me a better researcher for it.

Lastly I would like to thank the most important people in my life. I want to thank my family and boyfriend for supporting me throughout this project and my studies. Without you I would never be where I am today.

Eva van Tegelen
Delft, june 2022

# CONTENTS

# List of Symbols

| Symbol | Description |
|---|---|
| $S$ | Susceptibles |
| $I$ | Infected |
| $N$ | Number of household members<br>(in deterministic model the size of the population) |
| $c_{GP,k}$ | Contacts within general population for individual $k$ |
| $c_{HH,k}$ | Contacts within household for individual $k$ |
| $c_{NH,k}$ | Contacts within nursing home population for individual $k$ |
| $c_{HO,k}$ | Contacts within the hospital population for individual $k$ |
| $s_j$ | Susceptibility of individual $j$ |
| $i_j$ | Infection status or infectivity of individual $j$ |
| $i_{j,AB}$ | Infectivity of individual $j$ while taking antibiotics |
| $P_{CC}$ | Transmission probability of a CC |
| $P_{\text{clear } CC}$ | Probability of clearance when antibiotics are taken |
| $\alpha$ | Tuning parameter for saturation of contact rate |
| $\beta$ | Transmission number deterministic model |
| $\gamma$ | Recovery rate deterministic model (1/infection duration) |
| $\lambda$ | Rate parameter of the exponential distribution |
| $I_{\text{antibiotics}}$ | Adjusted infectivity parameter during antibiotic treatment |
| $S_{\text{antibiotics}}$ | Adjusted susceptibility parameter during antibiotic treatment |
| $M$ | Parameter enhanced effectiveness targeted treatment |

# List of Abbreviations

| | |
|---|---|
| MRSA | Methicillin-resistant *S. aureus* |
| CC | Clonal complex |
| GP | General population |
| HH | Household |
| NH | Nursing home |
| HO | Hospital |
| Exp | Exponential distribution |

# 1 INTRODUCTION

In many countries *Staphylococcus aureus* is the number one cause of hospital-associated infections, and a big proportion of these are caused by Methicillin-resistant *S. aureus* (MRSA)[1, 2]. Methicillin-susceptible *S. aureus* (MSSA) develops into MRSA upon acquisition of SCC*mec*, which is a mobile genetic element that carries methicillin resistance genes[3, 4]. Healthy people that are carriers of the bacteria have a low risk of getting an MRSA-infection, however in hospitals and nursing homes, patients are often more susceptible to infections due to chronic diseases, drug use, operations and other invasive procedures[5]. In most cases MRSA causes skin infections, however also more severe problems such as bloodstream infections, pneumonia and surgical site infections can occur[6].

In 1961 MRSA infections were first detected in hospitals (healthcare-associated (HA) MRSA)[7]. Until the mid 1990s, these infections were limited to hospitals, however, in recent years infections in the community are emerging[2]. These infections were caused by the rise of new distinct strains of MRSA, now referred to as community-associated (CA) MRSA[4, 8]. CA-MRSA strains were traditionally limited to populations outside the healthcare setting. However, the prevalence of CA-MRSA strains has increased and numerous reports describe its invasion of healthcare institutions[9]. Therefore, more recently, the distinction between CA-MRSA and HA-MRSA is fading as they start to overlap. Additionally, from the early 2000s new strains of MRSA have been found that are associated with farming of livestock (LA-MRSA)[10, 11].

The hospitals in the Netherlands apply a search-and-destroy policy[12]. This entails that patients and employees are assigned to different risk categories. Hospital guidelines [13] state that people with a high risk of being a carrier of MRSA are tested and if determined positive, are placed in isolation to prevent further spread within the hospital. Depending on the condition of the infected patient, choices have to be made about the necessity of treatment for MRSA. Employees that carry the MRSA bacteria are treated to eradicate carriage in order to prevent further spread in the healthcare system. For healthy individuals for whom the risk of developing infections from the MRSA bacteria is low and who are not likely to contribute to further spread in the healthcare system, it is often not deemed necessary to apply treatment.

During the past years this search-and-destroy policy in combination with restrictive antibiotic use has proven to be successful[14–17]. The percentage of *S. aureus* isolates that are MRSA in the Dutch population at hospital admission is among the lowest in the world. The highest percentages of MRSA measured in *S. aureus* isolates (>50%) are found in North America, South America and Asia[2, 18–21]. MRSA is endemic in many healthcare facilities throughout the world. Because of its resistance to often multiple classes of antibiotics, treatment options for MRSA are limited and therefore it has become a focus of infection control efforts all over the world[21].

Different strains of MRSA are grouped into clonal complexes (CCs) by their similarity to a central allelic profile. Separate MRSA CCs have evolved independently over time[22]. Some of the major MRSA CCs are CC5, CC8, CC22, CC30 and CC45[9]. In most countries a limited number of CCs dominates the prevalence in the population [8] and the composition of these CCs is different in each country. One study found for instance that in the Netherlands, Belgium and Denmark a specific CC, often associated with livestock, was present in the population, but could not be found in other countries in Europe[11]. The exact reasons why these CCs are successful in some countries and not in others are still unknown[15].

In order to limit the spread of different MRSA CCs it is critical to understand the transmission patterns of the infectious bacteria. Mathematical epidemiology can be a useful tool to study spread within a population, but also to simulate the outcome of possible measures against an infectious disease, such as MRSA. Different modelling studies into the spread of MRSA have already been completed. Multiple studies have used deterministic models to simulate the spread of MRSA[22–26]. These projects mainly focus on the transmission of MRSA within hospitals. Although in the models an external force from the community is included, the spread outside of hospitals is not taken into account. Even though MRSA bacteria mostly impact healthcare facilities, it is likely that transmission of MRSA within the population also influences the overall dynamics in the hospitals[9].

Deterministic models are often a relatively simple and comprehensible method to simulate the spread of a disease in a homogeneous population. However, due to the heterogeneity of infectious disease spread, individual-based models (IBMs) are becoming increasingly popular[27]. IBMs simulate populations by tracking individual agents and their properties[28]. Probability distributions are used to create heterogeneity between the different individuals. Unlike deterministic models, that use a limited set of differential equations, IBMs can describe very complex heterogeneous populations. However, this complexity is also the main criticism against IBMs[27]. The complexity of the model creates a black-box idea, which makes the process less apprehensible and transparent.

Being that the transmission of infectious diseases can be very complex, IBMs are powerful tools to study their spread. IBM has been applied to study multiple infectious diseases such as HIV and COVID-19[29–31]. Since MRSA is an infectious disease with a high degree of heterogeneity, different researchers have applied IBM to model the spread of MRSA[32–34]. In contrast to the deterministic models developed for MRSA, the IBMs often include not only hospitals, but also the entire population. The existing MRSA IBMs mostly explore in which settings transmission occurs and how in different healthcare facilities, such as hospitals and nursing homes, MRSA can become endemic. Unlike some of the deterministic models that included multiple strains of MRSA in their model[22, 25], IBMs predominantly model the spread of one type of MRSA.

In conclusion, a considerable amount of research has been done into the spread of MRSA and most of these studies focus on hospital settings. However, the increasing prevalence of CA-MRSA suggests that it is also important to include the general population in simulations. As mentioned before, a limited number of CCs dominates the MRSA populations of different countries. The exact reason for the successes of distinct CCs on a country level is still unknown and mathematical models have already shown to be an excellent tool to study competition between CCs[22]. However, much more research still has to be done in understanding the influence of different interventions, such as the search-and-destroy policy, on the prevalence of certain CCs in the general population and healthcare facilities.

The aim of this project will therefore be to study two different models that both simulate the spread of MRSA CCs in a population. Since MRSA is an infectious disease with a high degree of heterogeneity in infection dynamics, an IBM is a suitable tool to study the epidemiological dynamics of different CCs. Prior to this study the two IBMs were implemented in different programming languages: R and Java. The main goal of the research is to demonstrate the differences between the models and determine how each can be used to model the spread of multiple MRSA CCs. During this project both implementations have been carefully studied and the basic workings are compared. The Java model is furthermore used to do some exploratory research into the spread of multiple CCs in a generic population, in order to get a better understanding of some of the factors that contribute to their success.

The implementation of the R model and Java model will first be discussed in some details in Chapter 2 and 3 respectively. In Chapter 4 the IBMs will be put side by side for a comparison of the basic dynamics. Multiple simple model set-ups were developed to analyse different features that are included in both models. In Chapter 5 a more complicated and extensive model set-up was created for the Java model. The model will be applied to two settings: a population where no interventions are being taken (5.2.1) and one where a search-and-destroy policy is employed (5.2.2). In the discussion in Chapter 6 the performance and limitations of both IBMs and the results of the exploratory research will be debated.

# 2 IMPLEMENTATION R MODEL

The first MRSA model that will be discussed was implemented in the programming language R. The model was created by Anneke de Vos and is a time stepping IBM. At each time step the model uses information from the previous time step to determine, among other things, new infections and hospitalisations. This chapter explains the methods that were used to implement the different elements of the model. During this project some adjustments were made to the original model that was created by de Vos. Section 2.9 shortly discusses these alterations.

## 2.1 DATA STRUCTURE

During a simulation a population of humans is tracked. Births and deaths are not taken into account in this model and therefore the population size does not change during a simulation. For each individual different characteristics are stored inside a matrix-structured data frame which is updated every time step. Each individual in the population is assigned a row in the data frame in which their characteristics are collected. The R model stores multiple characteristics for each human: the household they belong to; if they are nursing home inhabitants; their assigned hospital; their hospitalisation rate; the amount of contacts within their household and within the general population; their infection status and their clearance type. It is important to note that during this research the definition of an infected individual is someone that is a carrier of the bacteria and can transmit MRSA to other individuals. Later on, the different characteristics of the individuals will be explained in greater detail.

Some of the characteristics stored in the data frame do not change over the course of the simulation. Household ID, assigned hospital, hospitalisation rate and clearance type are constant over time for all individuals. On the other hand, infection status, hospital status and the number of contacts individuals have, can vary over time. Considering that at each time step the data frame updates, one can imagine that if the population becomes larger, also the time it takes for the data frame to update during each time step increases. For very large populations, accessing and updating the data frame becomes computationally intense.

## 2.2 TRANSMISSION

The MRSA transmission within the R model is modelled by calculating for all susceptibles a force of infection for each of the CCs using the infection statuses stored in the data frame. This force of infection depends on the number of infected individuals within a subpopulation and the number of contacts an individual has. In this model, individuals are either susceptible or infectious. Every time step, within the general population (GP), households (HH), nursing homes (NH) and hospitals (HO) an infection force is calculated for each individual $k$ using the formulas presented in Equation 2.1.

$$\begin{aligned} \text{Infection force GP}_k &= c_{GP,k}\frac{\sum_j i_j c_{GP,j}}{\sum_j c_{GP,j}}, & \text{Infection force HH}_k &= c_{HH,k}\frac{\sum_l i_l c_{HH,l}}{\sum_l c_{HH,l}}, \\ \text{Infection force NH}_k &= c_{NH,k}\frac{\sum_m i_m c_{NH,m}}{\sum_m c_{NH,m}}, & \text{Infection force HO}_k &= c_{HO,k}\frac{\sum_n i_n c_{HO,n}}{\sum_n c_{HO,n}}. \end{aligned} \tag{2.1}$$

$$\text{Infection force}_k = \text{Infection force GP}_k + \text{HH}_k + \text{NH}_k + \text{HO}_k. \tag{2.2}$$

$c_{GP,k}$ = contacts within general population for individual $k$;
$c_{HH,k}$ = contacts within households for individual $k$;
$c_{NH,k}$ = contacts within nursing homes for individual $k$;
$c_{HO,k}$ = contacts within hospital for individual $k$;
$i_k$ = infection status of individual $k$.

The number of contacts each individual has within their household, nursing home, hospital and the general population are all drawn from separate user-defined distributions. By supplying non-constant probability distributions, heterogeneity can be inserted between the number of contacts each individual has.

The fractions in Equations 2.1 calculate the proportion of all the experienced contacts that are with an infected individual. By multiplying these fractions with the number of contacts of the susceptible individual, the infection force is created and essentially reflects the number of contacts the susceptible has with infected individuals. The transmission probability for the CC ($P_{CC}$) is the probability of transmission or infection when a susceptible and infected individual come into contact. This probability is then used to calculate the likeliness that the individual will become infected given their total infection force of that CC. This can be considered the complementary event to the situation where no infections would take place. This results in the calculation of the probability of infection presented in Equation 2.3.

$$\text{Infection probability}_k = 1 - (1 - P_{CC})^{\text{Infection force}_k}. \tag{2.3}$$

Equation 2.3 is often referred to as the Reed-Frost formula[35]. The infection force experienced by the susceptibles in the population is similar to the force of infection defined in the deterministic SIR model. Within the SIR model the force of infection is multiplied by the transmission probability of the CC and denoted by $\beta I/N$ in the differential equations. At each time step for all susceptibles the infection probability, denoted in Equation 2.3, is calculated. A random number is drawn from a uniform distribution and if this number is smaller than the infection probability the individual will become infected.

When an infection occurs, the infection status of that newly infected individual is updated and they get assigned a time to clearance. This time to clearance is dependent on the clearance type of the individual. The different clearance types are supplied by the user and indicate how long clearance of the MRSA bacteria takes on average for that specific clearance type. The decision to include the option to supply different clearance types in the model, is based on several studies that show that humans essentially can be subdivided in multiple groups of MRSA carriers: persistent carriers (~20% of humans), intermittent carriers (~30% of humans), and non-carriers (~50% of humans)[36–38]. At the time of infection an initial indication for the infection duration is drawn from a user-supplied distribution and this is multiplied by the clearance type of the individual to obtain their final time to clearance. As long as the time to clearance of an individual is not equal to zero, they are assumed to be infectious and therefore contribute to the number of infectious contacts. After each time step this time to clearance is decreased by 1 until the individual is no longer infectious. The individual is then returned to a susceptible state and is again vulnerable for new infections.

For each individual at every time step the infection force is calculated for all CCs. It is assumed that individuals can only be infected by one CC at a time. Therefore a check is build in that if an individual is infected by multiple CCs during the same time step, one of these CCs will be selected and the other infections will be cleared.

## 2.3 NURSING HOMES

The number of nursing homes and the sizes of these homes is supplied by the user. At the initialisation of the model the specified number of individuals are randomly selected and assigned to nursing homes. Within the data frame they are labelled as being nursing home residents. In this model, it is assumed that during the entirety of the simulation the same humans will stay inhabitants of the respective nursing homes. The user can also supply some additional characteristics to inhabitants of nursing homes. An additional rate for hospitalisation and antibiotic use can be defined specifically for the nursing home members. A distribution for the number of contacts the inhabitants have within their nursing home can also be supplied by the user.

## 2.4 HOSPITALS

Each human has an individual hospitalisation rate. Similar to the clearing types, multiple groups can be defined with each a different hospitalisation rate. The user supplies both the proportions of the different groups and the corresponding hospital rates. This gives the opportunity to model, for example, groups that never go, rarely go or frequently go to the hospital. This hospitalisation rate is the probability that an individual is hospitalised per time step. As mentioned before, nursing home inhabitants can be given an increased probability of hospitalisation, since generally they have a higher risk of becoming hospitalised than the general population. The hospital rates of the humans are stored inside the data frame and are used in the simulation to sample individuals from the population to be hospitalised.

At each time step the situation in the hospitals is updated. For the already hospitalised individuals the remaining length of stay is decreased by one. The individuals that reach the end of their stay are removed from the hospital and returned to the general population. From the general population the individuals that are hospitalised are determined based on their hospitalisation rate. The length of their hospital stay is drawn from a user-supplied distribution. Furthermore a contact weight for contacts in the hospital is drawn from another user-supplied distribution.

## 2.5 ANTIBIOTICS

Since one of the important properties of MRSA is resistance to certain types of antibiotics, it is important to incorporate this feature in the model. The user can supply the probabilities of getting antibiotics in the general population, in nursing homes and in the hospital. At each time step, individuals have a probability of getting antibiotics depending on their antibiotic use rate. The susceptibility of the different CCs to antibiotics, in other words the chance an infection of that CC is cleared when antibiotics are taken, can also be varied. The use of antibiotics is assumed to be one-time only and is not occurring over a period of multiple time steps. Once an infected individual is selected for antibiotics, a random number is drawn from a uniform distribution. In the event that this number is smaller than the susceptibility of that CC to antibiotics, the infection of the individual will be cleared.

## 2.6 EXTERNAL FORCE OF INFECTION

An external force of infection can be included in the model. For each of the CCs the average amount of yearly external infection events has to be supplied by the user. The points in time at which these infection events happen are randomly sampled and used to infect an uninfected individual with that CC at the selected time. The user has the possibility to supply if all individuals can be infected by an external force, or that only hospitalised individuals are exposed to external infections.

## 2.7 SEARCH-AND-DESTROY

The possibility to apply a search-and-destroy policy in hospitals is also implemented. The user can supply both the probability of a newly hospitalised patient and a patient already in hospital to be tested for MRSA. If an individual is selected for testing and has an MRSA infection, the patient is put in isolation. The effectiveness of isolation can be adjusted by the user and determines the amount of remaining contacts. Patients that are placed in isolation, remain in isolation until the end of their hospital stay or until their infection has cleared.

## 2.8 OUTPUT

After every time step for each of the CCs the number of infected individuals inside the general population, the hospitals and nursing homes are stored in an output file. Furthermore the number of people in the hospital and the number of people in isolation are collected. These outputs can be used to study the MRSA prevalence over time. It is not possible to look at the history of infection for specific individuals or households, because storing all individual data would be both intensely time and memory consuming.

## 2.9 ADJUSTMENTS

In the course of this research project multiple changes have been made to the original R model:

- The first alteration that was made concerned the order of the different steps of the algorithm. A mistake had been made, causing that at the end of each time step new individuals were infected, however at the beginning of the next time step their time to clearance was already decreased by one. Therefore the infection duration was actually a day shorter than anticipated. Furthermore the model did calculations of the infection force with the wrong number of susceptible and infected individuals. The number of calculated new infections at each time step was likely too low, as an infectious individual probably infected less individuals than when the order had been correct.

- The next alteration to the implementation actually did not change the results of the model, only improved the computation time. By profiling the code, it had become clear that the function responsible for calculating the infection force for each individual was very time consuming. This function consisted of multiple calculations of the weighted mean and had to be performed every time step for multiple CCs. As the programming language C++ has a higher computational speed when it comes to for loops and consequently for calculating a weighted mean, this specific part of the code was reprogrammed in C++ and incorporated in the R model. This alteration to the code made the model almost 10 times faster.

- An option that was added to the model, focused on the calculation of the general population infection force. The initial model assumed that the infection force $GP_k$ was not influenced by contacts with household members. This was later altered, because it is also possible to come in contact with household members in a general population setting. Because the infection force of a single household does not impact the general population infection force very much, it is not likely that this greatly impacted the overall prevalence.

- Another option that was added to the model is the possibility to include indirect contacts with oneself in the model. Instead of creating a custom mean function that excluded contacts with oneself, the option for a more straightforward weighted mean was included.

# 3 IMPLEMENTATION JAVA MODEL

The second model that will be discussed is again an IBM, but was now implemented in Java. Like the R model, it was created by Anneke de Vos with assistance of Roel Bakker and Rinke Hoekstra. The Java model is a lot more extensive than the R model and was developed to create a faster and more complete model to simulate the spread of MRSA. Since Java is an object oriented programming language the model consisted of many different classes with multiple methods. The Java implementation turned out significantly faster than the R model. Runs performed with the Java model were often more than 30 times faster.

A big difference with the R model, is that the Java model does not use a time stepping algorithm, but is event-based. Instead of updating the population after a certain time step, the population is only modified when an event occurs. Events can, for example, be a hospitalisation of individuals or an infection within a certain household. One event can trigger another event and this event can in turn schedule more events. This chain reaction is what is used to move the model through time. Event-based modelling can be a lot quicker than a time stepping algorithm, because the population is only updated when something changes and is therefore very efficient. When the population, however, becomes really large also the number of events can drastically increase. It is important to find a balance between updating the population regularly, yet making sure the computation time does not become too long.

## 3.1 DATA STRUCTURE

The Java model uses augmented B-trees for storage. The use of these trees has proven to be a technique that can enhance the performance of IBMs and can simplify the development[39]. The infection status of the different CCs is recorded in separate trees. Augmented B-trees allow for fast collection of transmission rates in the overall population, but also in smaller groups within the population. It can also significantly speed up weighted random selection of individuals for a next transmission event. This efficient data structure contributes to keeping the computing time of the model manageable despite possibly very large population sizes.

## 3.2 TRANSMISSION

In the Java model transmission is modelled differently than in the R model. In the R model for each susceptible a force of infection was calculated that not only consisted of the infection force experienced within the household, but also within the general population. Which route of infection is in the end responsible for the transmission can not be distinguished. In the Java model the infections within the general population, households, nursing homes and hospitals are modelled as separate events, which makes it possible to store the source of infection for each infected individual. The methods used to model the transmission in the different settings is also distinct. We will shortly discuss how transmission is modelled and which additional choices the user has for each route of infection.

### 3.2.1 GENERAL POPULATION TRANSMISSION

Infections that occur in the general population are modelled by calculating a transmission rate. This entails that depending on the number of infected and susceptible individuals an infection force is calculated which determines how many people will become infected in the general population. The transmission rate is calculated using the formula in Equation 3.1.

$$\text{Transmission rate GP} = \frac{\sum_j \left(c_{GP,j}\, s_n\right) \sum_j \left(c_{GP,j}\, i_m\right) P_{CC}}{\sum_j c_{GP,j}}. \tag{3.1}$$

$c_{GP,i}$ = general population contact rate of individual $i$,
$i_i$ = infectivity of individual $i$,
$s_i$ = susceptibility of individual $i$,
$j$ = individuals that are part of the general population.

In Equation 3.1 the parameter $P_{CC}$ is again the infection probability of a CC. Contrary to the R model no separate force of infection is calculated for each of the susceptible individuals. The Java model calculates a general transmission rate within the population. Linking this to a deterministic SIR model, the transmission rate would correspond to the term denoted by $\beta SI/N$ in the system of differential equations. Based on this transmission rate a population infection event is scheduled. When the transmission rate is higher, infection events in the population are scheduled closer together and occur more frequently. When the transmission rate is low, infection events are scheduled farther apart.

When an infection in the general population occurs a new infection is scheduled by drawing a time until the next event from an exponential distribution with the rate parameter being the transmission rate calculated in Equation 3.1.

$$dt \longleftarrow \text{Exp}(\lambda = \text{Transmission rate GP}).$$

A risk of these self repeating events is that the time between events could be so long that the circumstances within the population change a lot. Therefore the user can supply a waiting period. If the time step $dt$ is larger than the waiting period, the situation within the population is examined again after the waiting period. A new transmission rate is calculated and a new time step is drawn from the exponential distribution, which is possible as it is a memoryless distribution. This prevents that the situation in the population changes too much in between scheduled events. At the scheduled time of a population infection event a random individual is selected from the susceptible individuals, weighted by their susceptibility.

### 3.2.2 HOUSEHOLD TRANSMISSION

Within households transmission is modelled differently. Every time an individual is infected, either via population, hospital or household transmission, their household transmission rate is calculated. Depending on this household transmission rate a new infection event is scheduled within the household. In the Java model three contact rate models are implemented that can be employed to calculate the transmission rate. These are shortly described below:

- SUBPOPULATION: When this contact rate model is used it is assumed that all individuals within the household have the same contact rate. The formula in Equation 3.2 is then used to determine the raw transmission rate in the household, which also includes indirect contact with oneself.

$$\text{raw transmission rate} = P_{CC} \cdot c \cdot \sum_{j=1}^{N} i_j \cdot \sum_{j=1}^{N} s_j, \tag{3.2}$$

  where $c$ is the contact rate for each of the household members, $P_{CC}$ the transmission probability for the CC, $N$ the number of household members and $i_j$ and $s_j$ the infectivity and susceptibility of individual $j$ respectively.

- SUBPOPULATIONMEMBERS: When this contact rate model is used, the individuals in the household have their own contact rate. This contact rate also includes indirect contact with oneself. The formula in Equation 3.3 is then used to determine the raw transmission rate in the household.

$$\text{raw transmission rate} = N\frac{P_{CC}\sum_{j=1}^{N} c_j i_j \cdot \sum_{j=1}^{N} c_j s_j}{\sum_{j=1}^{N} c_j}, \tag{3.3}$$

  where the parameters are defined as in the previous setting, except for $c_j$, which is defined as the household contact rate of individual $j$.

- SUBPOPULATIONMEMBERSSC: Similar to the previous option, when this contact rate model is used each of the household members has its own individual contact rate. However, using this contact model indirect contacts with oneself are excluded by using the formula in Equation 3.4 to calculate the raw transmission rate.

$$\text{raw transmission rate} = (N-1)P_{CC}\sum_{j=1}^{N}\frac{s_j c_j \sum_{k=1}^{N} c_k i_k}{\left(\sum_{l=1}^{N} c_l - c_j\right)}, \tag{3.4}$$

  where the parameters are defined the same as in the previous setting.

In the algorithm these raw transmission rates are then transformed to the household transmission rate using a selected subpopulation contact structure. For this structure the model also has multiple options:

- LINEAR: When this setting is used the average number of contacts per person scales with the number of household members. This means that it is assumed that individuals with larger sized households have more contacts than individuals living in small households. Someone in a household with four members is assumed to have twice as much contacts as somebody living in a two-person household. In that case the household transmission rate is the same as the raw transmission rate.

$$\text{Transmission rate HH} = \text{raw transmission rate}. \tag{3.5}$$

- CONSTANT: By setting the contact structure to CONSTANT the average number of contacts per person is constant or in other words independent of the number of household members. This is similar to the frequency dependent transmission model that is used for infections in the general population. Someone in a household with four members is assumed to have the same amount of contacts as they would have in a two-person household. Equation 3.6 shows that when this setting is in place, the raw transmission rate is divided by the number of household members. Note that if the third calculation of the raw transmission rate (3.4) is used, it has to be divided by $N-1$.

$$\text{Transmission rate HH} = \frac{\text{raw transmission rate}}{N}. \tag{3.6}$$

- SATURATING: The last setting, SATURATING, is a middle way between the first two settings. When this setting is in place the average number of contacts per person saturates with the number of household members. Equation 3.7 is used to calculate the transmission rate.

$$\text{Transmission rate HH} = \frac{\text{raw transmission rate} \cdot \alpha \cdot \log(N)}{N}, \tag{3.7}$$

  where $\alpha$ is a tuning parameter that impacts how much the average contact rate saturates. It determines for each household size how much the number of household members scales with the number of contacts experienced by each individual.

The household transmission rate is then used to determine at what time the next infection event within that household will be scheduled. The time step to the next infection event is again drawn from an exponential distribution (3.8).

$$dt \longleftarrow \text{Exp}(\lambda = \text{Transmission rate HH}). \tag{3.8}$$

Infection events can be cancelled when something changes within the household. This can include household members being hospitalised or an infectious individual being cured, either naturally or by antibiotics. Depending on the household transmission rate that is recalculated, a new infection event can be scheduled.

### 3.2.3 Nursing home and hospital transmission

Transmission within nursing homes and hospitals is modelled using the same implementation as within households. When a nursing home inhabitant or a patient in the hospital becomes infected, the transmission rate within that hospital or nursing home is calculated. Depending on this transmission rate, a new infection event is scheduled after a certain amount of time, again drawn from an exponential distribution. Similarly as to the household infections, events can be cancelled when the situation chances in the nursing home or hospital.

Unlike the household setting, for hospitals and nursing homes not all contact structures that were described previously make sense. Within nursing homes the three contact rate models described in Section 3.2.2 can all be implemented. This means that the individuals in nursing homes can either all have the same nursing home contact rate or an individual contact rate. When considering plausible contact structures within nursing homes, the LINEAR setting seems unrealistic. This would mean that the number of contacts the nursing home residents have, would scale with the size of the nursing home. Therefore the contact structure within nursing homes is assumed to be CONSTANT.

Also within hospitals, the contact structure is assumed to be CONSTANT. Moreover, in hospitals it is also assumed that the patients have individual contact rates. This means that in the model only the second and thirth contact rate models can be used for the transmission within hospitals.

## 3.3 Nursing homes

Just like the R model, the Java model allows the user to supply the number of nursing homes and their sizes. At the initialisation of the model the specified number of individuals is randomly selected and assigned to nursing homes. Also for this model it is assumed that during the entirety of the simulation the same humans will stay residents in their nursing home.

The user can also provide some additional characteristics to inhabitants of nursing homes. Similar to the R model an additional hospitalisation rate and antibiotic use rate can be defined specifically for the nursing home members. The distribution that is used to determine the number of contacts the residents have within their nursing home, can be supplied by the user and is adopted to calculate the transmission rate within the home as mentioned in 3.2.3. On top of the features that the R model also has, the Java model holds additional options. The user can also choose to include personnel in the nursing homes. At simulation start, random individuals are assigned as nursing home personnel and will be part of the nursing home population, as well as of their own household. The distribution type for the number of contacts personnel has within the nursing home is assumed the same as for the inhabitants, but the average contact rate and if relevant the shape, can be set separately for personnel.

## 3.4 HOSPITALS

In the model a number of hospitals can be included in the population. At the beginning of the simulation no individuals will start in the hospital. Since all individuals are assigned a hospitalisation rate, at the outset the hospital population will grow until a balance is reached between patients entering and leaving. A distribution can be provided for the hospitalisation rate to create heterogeneity in how often individuals are hospitalised. This is different from the R model, where instead of one distribution, multiple groups are created and where each member of such a group has the same constant hospitalisation rate.

Hospitalisation events are scheduled as self-repeating events. The hospitalisation rates of all individuals are summed up and using this summed rate, the time until a next hospitalisation event is scheduled. An exponential distribution is used to determine the time step until the next event as stated in Equation 3.9.

$$dt \longleftarrow \text{Exp}(\lambda = \sum_j \text{Hospitalisation rate individual j}), \tag{3.9}$$

where the parameter $j$ displays all individuals that can potentially be hospitalised. At the time of the scheduled hospitalisation event, weighted random selection is used to pick an individual that is hospitalised. When a patient enters the hospital the length of their stay is drawn from a user-supplied distribution. After that time, an event is scheduled that regulates the end of hospitalisation.

Transmission within the hospital is modelled as described in 3.2.3. A distribution can be provided for the number of contacts patients have when they are in the hospital. The Java model has the option to either draw a new contact rate at each hospitalisation event, or to give every individual a hospital contact rate that stays the same over multiple hospitalisations. Furthermore, also the remaining fraction of household and population contacts can be adjusted. Similar as in the R model, the antibiotic use in the hospital can be provided separately from general population antibiotic rates.

A component that the Java model has, but was not incorporated in the original R model, is the possibility to divide the hospital into multiple wards. The user can supply for each hospital which fraction of a patients contacts is within their own ward and the remaining contacts are assumed to be random over the entire hospital. Similar as in the nursing homes, also an option to include staff within the hospital population is at hand. The user can provide the number of personnel for each patient and the average number of contacts this personnel has.

## 3.5 ANTIBIOTICS

Selecting individuals for antibiotic treatment is done using self-repeating events, which is similar to the selection of new hospital patients. The user can supply a distribution of individual tendencies to use antibiotics for individuals in the general population, nursing homes and hospitals. Note that this is different from the circumstances in the R model, where it was only possible to supply one constant antibiotic rate for each of these groups. Similar as to the management of hospitalisations, all antibiotic rates are summed together and the time step until a next antibiotic event is drawn from an exponential distribution (3.10).

$$dt \longleftarrow \text{Exp}\left(\lambda = \sum_j \text{Antibiotics rate individual j}\right), \tag{3.10}$$

where $j$ covers all individuals that can potentially be prescribed antibiotics. At the time of the next antibiotic event an individual is chosen through weighted random selection. Unlike the R model,

where it was assumed that getting antibiotics is an instantaneous event, in the Java model it is possible to adjust the length of antibiotic therapy. The probability that antibiotic treatment will have cleared an infection at the end of the antibiotic therapy is dependent on the susceptibility of the responsible CC to antibiotics. During treatment with antibiotics, due to a lowering of bacterial load, an individual's infectiousness may be lowered. The infectivity of infected individual $j$ while taking antibiotics, $i_{j,AB}$, is then altered according to Equation 3.11.

$$i_{j,AB} = i_j(1 - P_{\text{clear } CC})^{I_{\text{antibiotics}}}, \tag{3.11}$$

where $i_j$ is the infectivity of individual $j$ before antibiotics were prescribed and $P\text{clear } CC$ the probability of clearing the infection for that specific CC when antibiotics are prescribed. The constant $I_{\text{antibiotics}}$ can be supplied by the user and determines how much the infectivity is adjusted during treatment with antibiotics. The lowering of bacterial load can possibly also have an effect on the susceptibility of individuals. The susceptibility of an individual $j$ while taking antibiotics, $s_{j,AB}$, can also be altered using the formula presented in Equation 3.12.

$$s_{j,AB} = s_j(1 - P_{\text{clear } CC})^{S_{\text{antibiotics}}}, \tag{3.12}$$

where $s_j$ is the susceptibility of individual $j$ before antibiotics were prescribed and $P\text{clear } CC$ again the probability of clearing the infection when antibiotics are prescribed. The constant $S_{\text{antibiotics}}$ can be supplied by the user and determines how much the susceptibility of an individual is adjusted throughout antibiotic therapy. As competitor bacteria can also be cleared by antibiotics, overall susceptibility to any CC may actually be enhanced by a certain factor for a period of time after antibiotics use. The user can supply the length of the period for which an individual has an adjusted susceptibility after taking antibiotics and also with what constant multiplication factor the susceptibility is adjusted.

## 3.6 External force of infection

In addition, the model includes an external source of infection. These external infection forces come from outside the population and can represent infections from abroad, or possibly farmers being in contact with livestock. The user can supply the fraction of individuals that is to be exposed to this external infection force. The rate of exposure can be set for each CC separately as the number of external infections per exposed individual per day. Immediately upon an external infection event a new external infection event is scheduled, depending on the number of exposed individuals at that moment in time and the average number of external infections per exposed individual per day.

## 3.7 Search-and-destroy

Comparatively to the R model, there is also an option for a search-and-destroy policy in the Java model. However, the definition and implementation used in the Java model are very different. At hospital entrance only people who are assumed to have a high risk for MRSA carriage are isolated and then tested. This risk group consists of the fraction of people that is exposed to external infections and individuals that have been tested positive during a previous visit. Additionally, or alternatively, it is also an option that a random fraction of those already in hospital is tested for MRSA each day. All individuals that test positive for infections by MRSA CCs are placed in isolation. The effectiveness of the isolation can also be altered by the user.

Another advanced feature of the Java model, that was not included in the R model, is that a result delay for the MRSA tests can be provided by the user in the form of a distribution. Until a patient gets the results of a test, they will be put in isolation out of precaution. If an individual is tested positive, they will stay in isolation, otherwise the isolation will be relieved.

The strength of a model that includes both the hospital and the general population, is that interventions in both groups can be considered. To do this, a feature was added to the Java model that focuses on stopping the spread of MRSA by targeting the general population. Optionally, a user provided number of days after testing positive in the hospital, an individual and his or her household members may be tested. Those found positive via family testing may receive targeted antibiotic treatment, which can also be given to those tested positive in hospital. The targeted course of antibiotics is the same as the standard course, only the probability for clearance is increased by a factor. The probability of clearing MRSA becomes $1-(1-P_{\text{clear }CC})/M$, where $M$ scales the enhanced effectiveness of targeted treatment.

## 3.8 Output

The user can supply how often they want output to be generated. For each CC the cumulative number of infection events up to that moment, the cumulative number of infections for each possible infection route (general population, household, hospital, nursing home and external infections) and the cumulative number of clearing events is given. Also the number of infected and uninfected individuals at the time of the output moment, optionally for each hospital and nursing home, are collected. The entire output will be stored in a csv file and can be used to study the prevalence over time in the population. Additionally a text file with the used input parameters is also outputted. Optionally, much more detailed individual data can be outputted in additional files, for those in a chosen subset of the households. In these files the exact time and cause of clearing and infection are listed for the selected individuals.

Where the R model could only show general information, such as the number of infections for each CC and the number of people in hospital, the Java model has many options for more detailed information. Not only is it possible to output the source of infection for all infections, it is also possible to focus on specific households.

## 3.9 Adjustments

Overall, not many adjustments were made to the Java model in the course of this project. However, what is important to note for future users, is that the input of distributions is slightly different from standard parameterisations. During this project this led to multiple confusions and errors. It would be advised to take special care in making sure the right parameters are supplied, in order to obtain the desired distributions. During this research project two small changes have been made to the original Java model.

- The first small alteration to the model was made in the step from the raw transmission rate to the final transmission rate. For the CONSTANT and SATURATING setting the raw transmission rate was originally divided by $N-1$, but as the model may include indirect contacts with oneself this should be $N$ (except when the SUBPOPULATIONMEMBERSSC setting is used).
- Another adjustment done to the model was not made in the algorithm, but in managing different runs of the model. Originally already an option was included to provide the number of repeated runs that has to be performed. Due to an error the seed did not change during these runs, so all runs were exactly the same. A small alteration was made such that the seed for the random number generator is increased by one each run.

# 4 COMPARISON OF THE R MODEL AND JAVA MODEL

In this section the R and Java implementations will be compared. Different model set-ups have been created to observe the results of the two models side by side. The model set-ups that will be discussed in this section are basic and not complex. By stripping down to simpler models, comparison of the basic dynamics is made more accessible. Each of the simple models that will be discussed, was created to study a specific aspect of the implementations. In the subsections below the different model set-ups that were used, are shortly discussed and the results of their comparison are presented. Note that in the remainder of the research the terms model and model set-up might be used in a somewhat overlapping fashion. From the context it should be clear if the expressions refer to the general Java and R model or a specific model set-up that was used in the implementations.

All model set-ups simulate a population of 10.000 humans, which is equivalent to around 4.000 households. At the start of every simulation for each of the CCs 100 humans are assumed to be infected. Unfortunately, not all distributions that can be used in the R implementation, can also be used in the Java implementation. To prevent potential differences and errors that can be made when introducing complicated distributions, only constant distributions have been used in this comparison. Parameters such as the number of household contacts, general population contacts, infection duration and length of hospital stay were assumed to be constant and the same for all individuals. In Table 4.1 at the end of this chapter a summary of the parameters that were used in the different model set-ups is denoted. The number of individuals (10.000) and the infection duration (8 days) are the same for all model set-ups. The other parameters stated in the table, can differ between the different models. Since the R and Java model are IBMs and therefore probability dependent, no two runs will be exactly the same. For each model set-up at least 5 runs are performed for both implementations. This was done to get a better understanding of the general behaviour of the runs and to make sure that the presented results are an accurate representation of the possible spread in each scenario.

### MODEL 1

The first model set-up that was created, Model 1, aims at comparing the spread of a CC via general population contacts for the R and Java model. The population considered does not include any nursing homes or hospitals. Since we want to specifically focus on the spread of MRSA by means of general population contacts, this was assumed to be the only route of infection. No household infections were included, which was done by setting the number of household contacts to zero. All individuals in the population were assumed to have the same amount of daily general population contacts, which for this model set-up was fixed on 10.

This simple model set-up can be linked to a compartmental deterministic model. By avoiding household infections and only considering general population infections, essentially two compartments are created. One compartment consists of susceptibles $S$ and the other compartment of infected individuals $I$. Every individual in the susceptible group can be infected by someone from the infected group and move to that compartment. Each infected individual can move in turn to the susceptible group again if their infection is cleared. These interactions coincide with the differential equations of the deterministic SI-model shown in Equations 4.1 and 4.2.

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma I, \tag{4.1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I. \tag{4.2}$$

The constant $\beta$ is assumed to be the average number of contacts per individual per time, multiplied by the probability of infection given an infectious contact. The duration of the infection is captured in the constant $\gamma = 1/\textit{Infection duration}$. The constants in this deterministic model were imposed such that they coincide with the parameters of the IBMs. Using the Euler forward method the solution of this deterministic model can be set side by side to the results of the two individual based implementations.

In Figure 4.1 the results of the first model set-up are shown for the different implementations. Three different infection probabilities were used to study the spread within the population over a period of 5 years. For each infection probability 5 simulations were performed with both the R and Java implementations. The 5 runs that correspond to a specific infection probability overlap and are all plotted with the same colour in all figures. As the IBMs are stochastic processes, this was done to give a more general overview of how the prevalence in the population progresses.

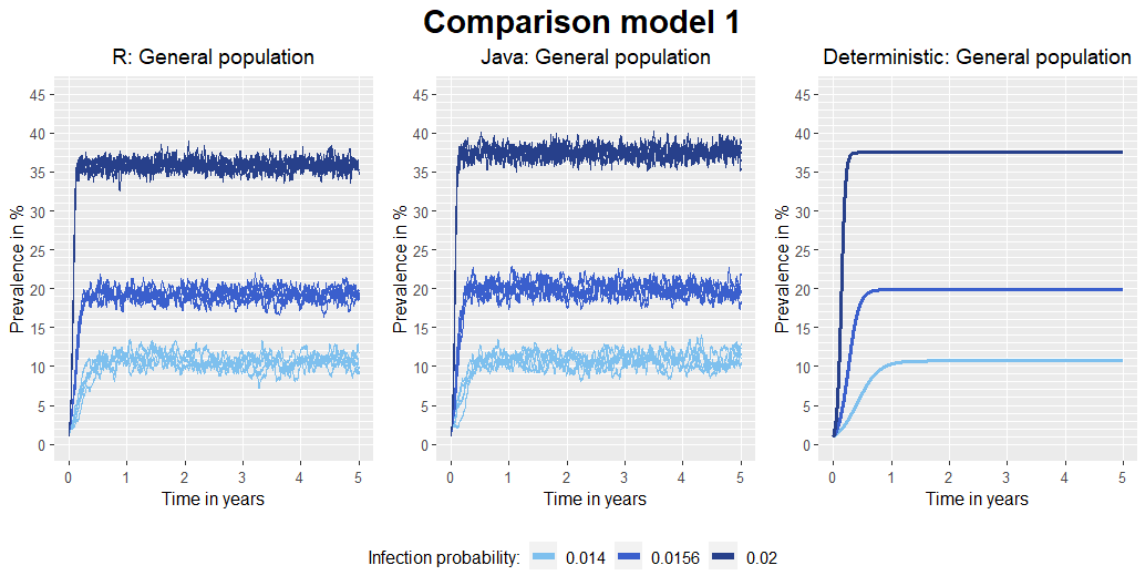


Figure 4.1: **Model 1.** Includes: A single CC, general population infections. Excludes: household infections, hospitals, antibiotics. Three different infection probabilities were used in the simulations and for each infection probability 5 runs are plotted for both the R model (left) and Java model (middle). The IBMs are put side-by-side with the corresponding deterministic model (right).

The results of the two IBMs shown in Figure 4.1 exhibit very similar behaviour as the solution of the deterministic model on the right. For the deterministic model it can be seen that when starting with 100 initially infected individuals, over time the prevalence within the population will increase to an equilibrium. At this equilibrium the number of newly infected individuals is equal to the number of infected individuals that gets cured every time step. It can be observed, that the height of this equilibrium prevalence is dependent on the infection probability of the CC. With an infection probability of 0.014, it can be seen that the equilibrium will be around 10%. By increasing the infection probability and therefore the infectiousness of the CC, the equilibrium prevalence will also become higher.

For the R implementation, each of the 5 runs for which an infection probability of 0.014 was used, show the same behaviour of moving to an equilibrium prevalence of around 10%. Because the R model is probability dependent, the prevalence in the population will not be steady and constant, but will fluctuate around the equilibrium. Around this equilibrium the number of new infections

is approximately the same as the number of individuals that is cleared from their infection. Also for the higher infection probabilities, 0.0156 and 0.02, it can be seen that the runs of the R implementation will approach the deterministic equilibria closely. The same can be said for the results of the Java model. For each of the infection probabilities, the 5 runs move quickly towards the deterministic equilibrium. An interesting result is that the speed by which the prevalence reaches the equilibrium, seems to be slightly higher in the IBMs than in the deterministic model.

The similarity between the deterministic model and the two IBMs gives confidence that the different implementations exhibit the same behaviour regarding general population contacts and infections. Although the methods that model the spread of a CC within the population are slightly different, the R and Java implementations show similar results, that are comparable to the equivalent deterministic model.

### MODEL 2

Model 2 was created to examine whether adding household infections to Model 1, would still result in similar behaviour of the two implementations. By incorporating households into the population, essentially a lot of small subpopulations are created, which all have their own susceptible and infected compartment. Translating this to a deterministic model would correspond to an extremely elaborate compartmental model with a lot of differential equations. Consequently, for this model set-up the R and Java model will no longer be put alongside a deterministic model for verification.

Although possibly not realistic, in the R implementation it is rooted that each individual has a certain amount of household contacts, which does not scale with the number of household members. To mirror this, in the Java implementation the CONSTANT setting (explained in section 3.2.2) was used. For simplicity it is assumed that all individuals have the same amount of household contacts and that some of these contacts include indirect contacts with oneself. This requires the SUBPOPULATION setting in the Java model and a constant distribution in the R implementation. For each individual the number of daily general population contacts was set to 5.0 and the daily number of household contacts to 10.0. Again the results were simulated for multiple values of the infection probability in order to study different speeds of transmission. In figure 4.2 the results are shown for the R and Java model.

Again it can be observed that for both IBMs all runs move towards and then fluctuate around a specific equilibrium. For both the R model and the Java model, an infection probability of 0.015 will result in an equilibrium prevalence of around 12%. The pace at which this happens seems to be comparable. Also for the higher infection probabilities, 0.017 and 0.019, the equilibria of the two implementations are much the same. The results of Model 1 already showed that the spread of an MRSA CC via general population contacts is similar for the R and Java model. By adding household interactions and consequentially household infections to the population, Model 2 was created. The results obtained with this second model set-up show again that all runs of the IBMs move towards equivalent equilibria. This indicates that the two implementations exhibit also comparable behaviour in connection to household infections.

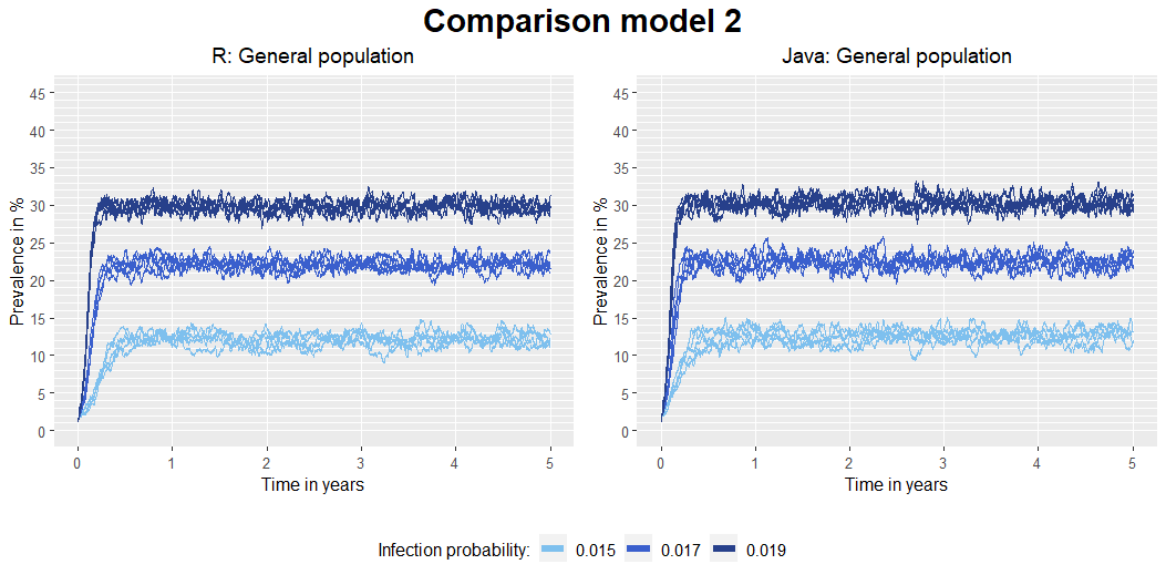


Figure 4.2: **Model 2.** Includes: A single CC, general population infections, household infections. Excludes: hospitals, antibiotics. Three different infection probabilities were used in the simulations and for each infection probability 5 runs are plotted for both the R model (left) and Java model (right).

### MODEL 3

Model 3 was designed to research the addition of hospitals to the population and the consequences of this to the prevalence of the MRSA CC. The same model set-up was used as in Model 2, but additionally one hospital was added to the population. It is assumed that all individuals have the same probability to be hospitalised. In the set-up of the R model the daily hospitalisation rate was set to 0.002, which translates into a yearly hospitalisation rate of 0.73 within the Java implementation. All hospitalised individuals were assumed to have the same amount of hospital contacts, which was set to 15.0 contacts per day. It was assumed that the number of household contacts of the hospitalised individuals is halved during the entirety of their hospital stay. At the time of hospitalisation the length of stay in the hospital is set to 12 days.

In Figure 4.3 the prevalence in the general population and the hospital are shown for both implementations. Again for three different infection probabilities 5 runs were performed with both the R and Java model. Note that in the figures that demonstrate the prevalence in the hospital, the darker blue lines overlap the other lighter coloured lines and therefore some information may be lost in the figure. The first thing to note is that for both implementations the prevalence of the CC in the hospital is higher than in the general population. It was assumed that the number of hospital contacts is relatively high, namely 15.0, which explains why the CC is able to spread so quickly within the hospital. Both implementations show that also in the hospitals the prevalence will move towards an equilibrium. The equilibrium around which the hospital prevalence fluctuates, appears to be comparable for the R model and the Java model for each of the infection probabilities.

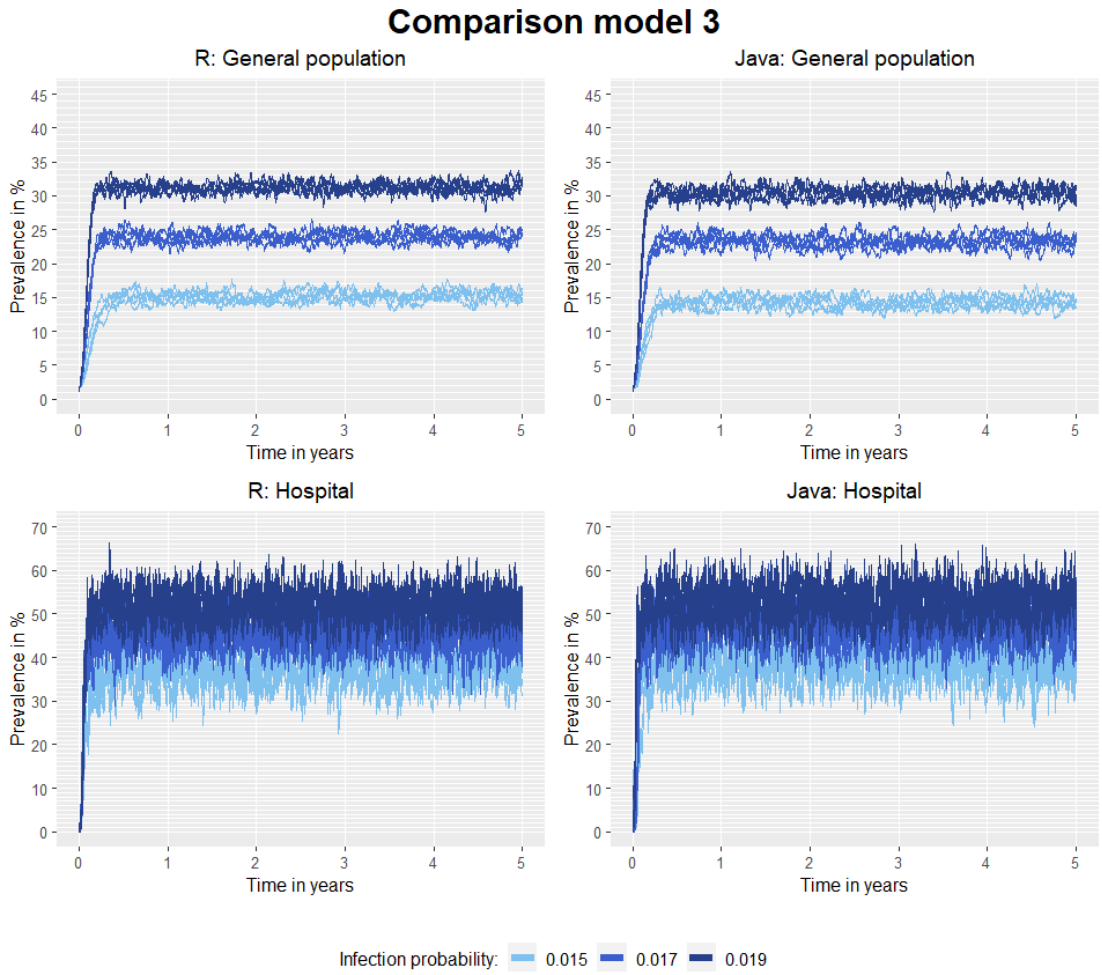


Figure 4.3: **Model 3.** Includes: A single CC, general population infections, household infections, 1 hospital. Excludes: antibiotics. Three different infection probabilities were used in the simulations and for each infection probability 5 runs are plotted for both the R model (left) and Java model (right). The top figures present the prevalence in the general population and the bottom figures in the hospital.

Comparing the prevalence in the general population in Figure 4.3 to that in Figure 4.2, especially for the smallest infection probability (0.015), it can be seen that due to the additional spread within the hospital, also the prevalence within the general population increases. Since hospitalised individuals, who become infected during their stay in the hospital, return to the general population when discharged, they can boost the general population and household infections. For all three infection probabilities it is apparent that the addition of hospitals has increased the overall prevalence in the general population. The increase that is caused by the addition of hospitals, turns out to be of similar size for the R and Java model.

In conclusion, adding hospitals to the population resulted in similar behaviour for the R and Java model regarding hospital prevalence and prevalence in the general population. Although there are some differences between the R and Java implementation of hospitalisation and hospitalised individuals, the comparable results evoke the idea that the incorporation of hospital infections has the same effect on the population prevalence in both models.

### MODEL 4

The fourth model that was studied focuses on including the use of antibiotics in the population. Again the set-up of Model 2 is reused, but with addition of antibiotics. It is assumed that all individuals in the population are equally likely to get antibiotics. A yearly antibiotic use rate of 10.0 is used. Given that an individual is selected, they receive a one-time dose of antibiotics. In case an infected individual is the receiver of antibiotics, the chance of clearing their infection was set to 0.3. In Figure 4.4 the results of multiple runs are shown for three different infection probabilities.

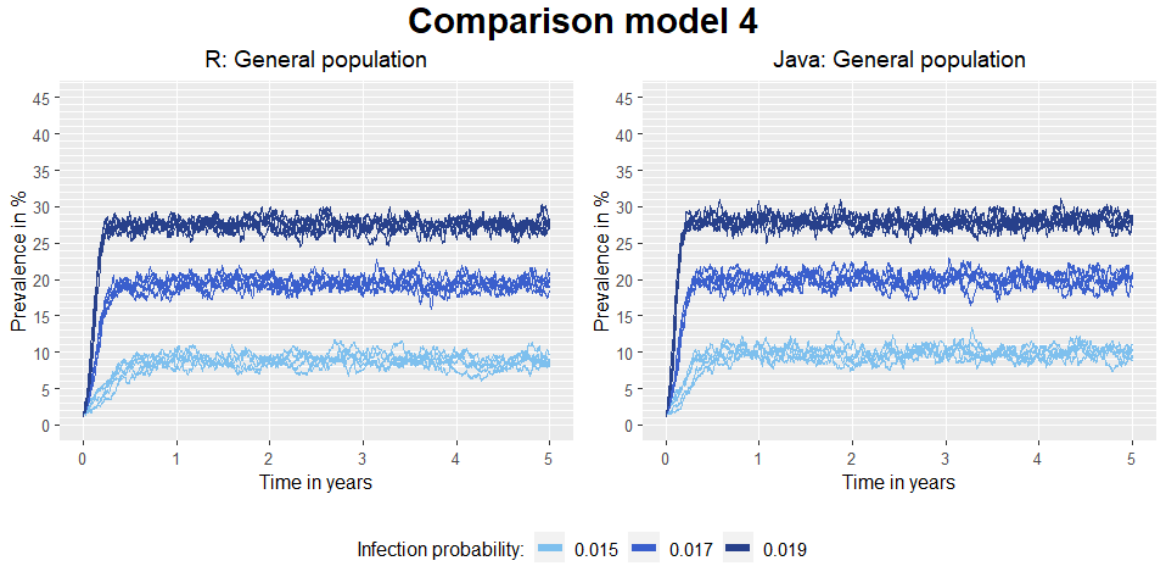


Figure 4.4: **Model 4.** Includes: A single CC, general population infections, household infections, antibiotics. Excludes: hospitals. Three different infection probabilities were used in the simulations and for each infection probability 5 runs are plotted for both the R model (left) and Java model (right).

In Figure 4.4 once more it is shown that all 5 runs of the one infection probability go to the same equilibrium. The equilibria of the R and Java implementations again appear to be similar. If the height of the equilibria is compared to those of Model 2 in Figure 4.2, it can be seen that the general population prevalence was shifted down. For example, for the infection probability of 0.015 the equilibrium prevalence in Model 2 was just above 10% and by including antibiotics in the population this has dropped below 10% percent. Likewise, for the other two infection probabilities, this drop in prevalence is the case for both the R and Java implementation. This clearly coincides with what one would expect. By including antibiotic use in the population, there now is a chance that infected individuals are cleared before their infection would have cleared naturally. When this "early" clearance happens, they can no longer infect other individuals and therefore this will negatively impact the spread of the CC within the population.

The similarities between the effects that the use of antibiotics has on the prevalence in the population for the two implementations, suggests that likely the different implementations of antibiotic use in the R and Java model have the same effect on the spread of the CC within the population.

### MODEL 5

For Model 5 household infections were omitted. The same set-up was used as in Model 1, but an extra CC was included. By studying Model 5 the interactions between two CCs can be analysed and how they possibly differ for the R and Java model. Similar to Model 1, by avoiding household infections and only considering general population infections, multiple compartments are created, between which interactions can take place. Once again there is the susceptible group, but moreover, since individuals can become infected by two different CCs there are two infected compartments. From the susceptible group each individual can move to either one of the infected groups. From each infection group, they will move back to the susceptible group after the infection duration. These different groups and interactions coincide with the differential equations of the deterministic SII-model presented in Equations 4.3-4.5.

$$\frac{dS}{dt} = -\frac{\beta_1 S I_1}{N} - \frac{\beta_2 S I_2}{N} + \gamma(I_1 + I_2), \tag{4.3}$$

$$\frac{dI_1}{dt} = \frac{\beta_1 S I_1}{N} - \gamma I_1, \tag{4.4}$$

$$\frac{dI_2}{dt} = \frac{\beta_2 S I_2}{N} - \gamma I_2. \tag{4.5}$$

The constants $\beta_1$ and $\beta_2$ are assumed to be the average number of contacts per individual per time, multiplied by the infection probability of CC1 and CC2 respectively. The duration of the infection is captured in the recovery rate $\gamma = 1/\textit{Infection duration}$, and is assumed to be the same for the CCs. By recurrently using the Euler forward method the solution of this deterministic model can be placed side by side to the results of the two IBMs.

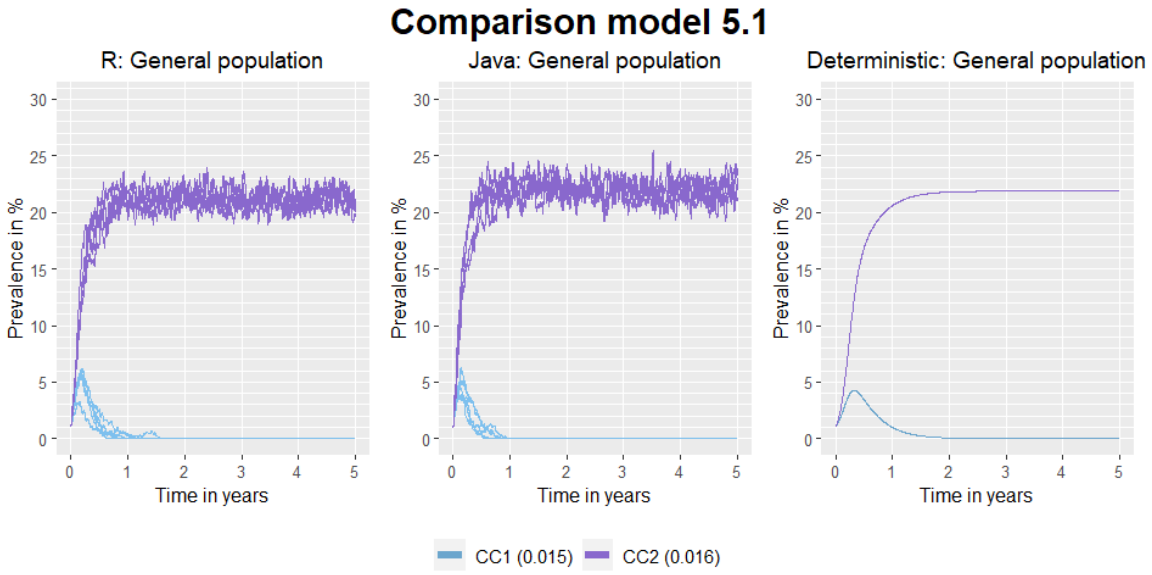


Figure 4.5: **Model 5.1**. Includes: Two CCs, general population infections. Excludes: household infections, hospitals, antibiotics. The CCs are assumed to have different infection probabilities and 5 runs both the R model (left) and Java model (middle) are presented. The IBMs are put side-by-side with the corresponding deterministic model (right).

In Figure 4.5 the results of 5 runs are shown for one set of infection probabilities. The scenario that is considered in the figure is where the two CCs have a different infection probability. CC1 has an infection probability of 0.015 and CC2 a probability of 0.016. This gives a slight advantage to CC2, as it is likely to spread faster. This is exactly what can be observed in the solution of the deterministic model presented in the right graph in Figure 4.5. At first both CCs increase in prevalence as

there are plenty susceptibles to become infected. Already at the beginning it can be seen that the spread of CC2 is faster. At a certain point the number of susceptibles and infected has reached a balance and at that point in time the number of infections caused by the more infectious CC2 starts to take over. Over time a smaller portion of infections is caused by CC1, which makes the number of new infections less and consequently the CC will spread even slower. In the end it can be seen that due to the competition between the CCs, CC1 will eventually leave the population and all infections will be caused by CC2. The disappearance of CC1 is also present in the runs of the IBMs. In most runs, after 1 year CC1 has completely disappeared from the population. From that time on, all infections are caused by CC2 and move around an equilibrium similar to that of the deterministic model.

Both the R model and Java model show a similar small peak in the prevalence of CC1 in the first year. It is interesting to note that in most IBM runs this peak seems to be a little bit higher and less wide than in the corresponding deterministic model. This could possibly be explained by the fact that if by chance there is a time at which the prevalence of CC1 is low and of CC2 is high, this effect is possibly amplified during the next time step or event, which does not happen in the deterministic model. In the case where one CC has an advantage over the other we see that both the deterministic model and IBMs show that the CC with the advantage is likely to dominate the infections and that the other CC will disappear from the population.

By appointing a higher infection probability to one of the CCs a clear advantage is created. Another interesting topic of study is the behaviour of the models if both CCs have the same characteristics and therefore no CC has an initial advantage over the other. In Figure 4.6 the equilibria are shown for the deterministic model where both CCs have an infection probability of 0.016.

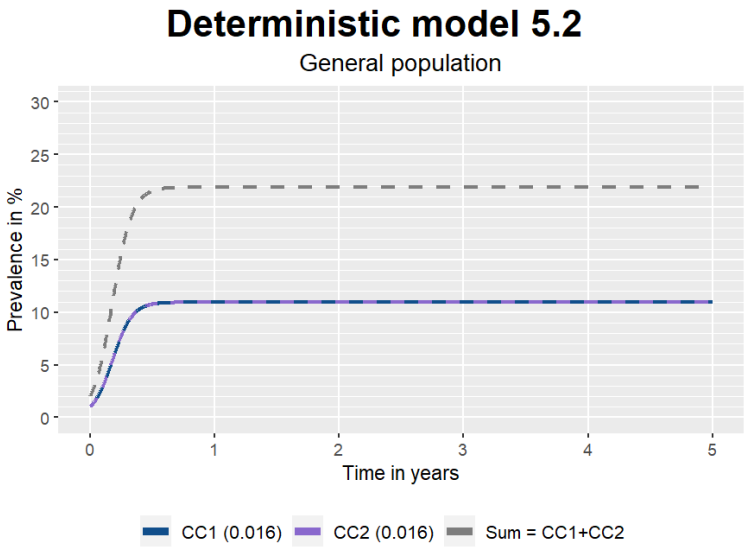


Figure 4.6: **Model 5.2**. Deterministic SII-model. The two CCs have the same infection probability. The dotted line is the sum of the prevalence of CC1 and CC2.

In Figure 4.6 it can be seen that in the solution of the deterministic model, since no CC has an advantage, both prevalences will go to the same equilibrium. The combined prevalence in the population is presented by the dashed line. This total prevalence is divided equally over the two CCs. Figure 4.7 illustrates that this is not necessarily the case for the IBMs.

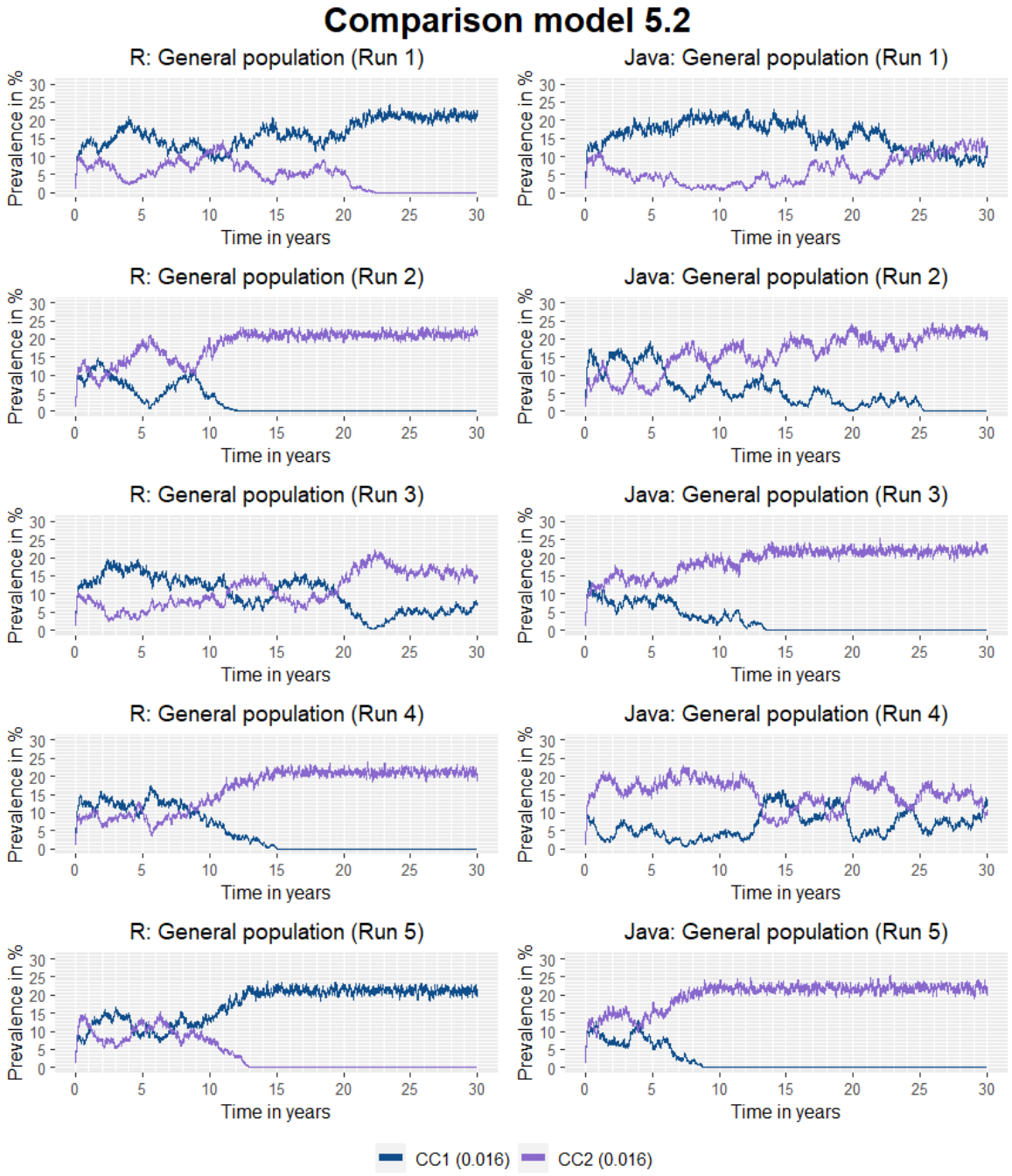


Figure 4.7: **Model 5.2**. Includes: Two CCs, general population infections. Excludes: household infections, hospitals, antibiotics. The CCs are prescribed the same infection probability. 5 runs were done for both the R model (left) and Java model (middle) and are presented as individual graphs.

Figure 4.7 shows 5 runs for both the R model on the left and the Java model on the right. These 5 runs were randomly selected. The first thing that can be noted, is that there is quite a lot of variation between the different runs. Unlike the model set-ups presented so far in this section, the behaviour of the runs is never the same. This can be attributed to the fact that the success of a CC is now solely dependant on the stochasticity of the IBM. Because both CCs have the same infection probability and consequently the same characteristics, the only way one CC can be ahead of the other is by chance.

In all runs, initially the general population prevalence of both CCs will increase for a short amount of time. However, where in the deterministic model the prevalences will remain the same and eventually go to identical equilibria, for the IBMs a difference in prevalence appears between the CCs. For example, for run 1 of the R model and run 1 of the Java model the prevalence of CC1 is in the beginning higher than that of CC2. When the number of infections for one CC is relatively low, the new number of infections caused by that CC is likely to be lower than that of the more prevalent CC. For the prevalent CC there are more infected individuals that can contribute to its spread. However, if due to stochasticity the more prevalent CC has a small decrease in prevalence, there are more susceptibles available for the less prevalent CC. It is possible for the prevalence of a CC to increase again. This can be seen in run 1 of the Java model, where CC2 appears to make a comeback after some time. Not only can the prevalence increase, but also the role of most prevalent CC can be reversed, as can for example be seen in run 3 of the R model and run 2 of the Java model.

Over time the difference in prevalence can become larger between the CCs. In run 1, 2, 4 and 5 of the R model and run 2, 3 and 5 of the Java model this difference becomes so large, that one of the CCs is actually eliminated from the population. From that point on all infections will be caused by the remaining CC. Which CC will be the dominant CC is up to chance. In run 1 and 5 of the R model it can be seen that CC1 is dominant, but in run 2 and 4 CC2 is dominant. It can be expected that over a very large time span, the simulation will always reach a point were the difference between the prevalent CCs will be so large that one of the CCs will disappear from the population.

Because of the variation between the different runs for both the R and Java model it is a little more complicated to compare the results. However, between the runs of the two models multiple similarities can be spotted. For example in run 2, 4, 5 of the R model and run 3 and 5 of the Java model it can be seen that in a period of less than 15 years one of the CCs has disappeared from the population and the other CC is the only one left. For run 1 of the R model and run 2 of the Java model this happens after a longer period of time. In both these runs it can be seen that the prevalence of the CCs sometimes goes down, but then later increases again. This we can also see in run 3 of the R model and run 1 and 2 of the Java model. However, these simulations have not reached the point where the prevalence of one of the CCs becomes so low that it leaves the population.

Although for the IBMs it appears that the two CCs will never move to the same the equilibrium, the total amount of MRSA infections (CC1+CC2) is at almost all times close to the sum of the two prevalences in the deterministic model. In Figure 4.6 it can be observed from the dashed line that the sum of the prevalence of the two CCs for the deterministic model is around 22%. This is very close to the prevalence of the remaining CC in run 1, 2, 4 and 5 of the R model and Run 2, 3 and 5 of the Java model. Also for the runs were both CCs are still prevalent in the population after 30 years, such as run 3 of the R model and run 1 and 4 of the Java model, the sum of the prevalences is at all times very close to the 22%. This prevalence seems to be the balance between the number of susceptibles and infected individuals. It is the total MRSA prevalence at which the amount of new infections is roughly the same as the number of infections that is cleared.

Summarising, the R and Java implementation exhibit the same behaviour regarding multiple CCs in a population, when only general population infections are considered. When one of the CCs has an advantage, such as a higher infection probability, it will become the dominant CC and the other CC will disappear from the population. When the CCs have equally advantageous properties it depends on chance which CC will be more prevalent. The fact that the simple R and Java model set-ups show similar behaviour when multiple CCs are included, suggests that the different implementations of the interactions between multiple CCs have comparable effects on the results.

### MODEL 6

Model 6 was created to examine whether the spread of the two CCs is still similar for both implementations, if household infections are included in the population dynamics. Within the Java model there is a difference between the methods that model general population infections and household infections. Model 5 already showed that interactions between two CCs were similar for the different IBMs if only general population infections are considered, but it has to be checked that this is also the case when additionally household infections are incorporated. The same parameters were used as in Model 2, but with the addition of an extra CC. In Figure 4.8 5 runs are presented for both the R and Java model.

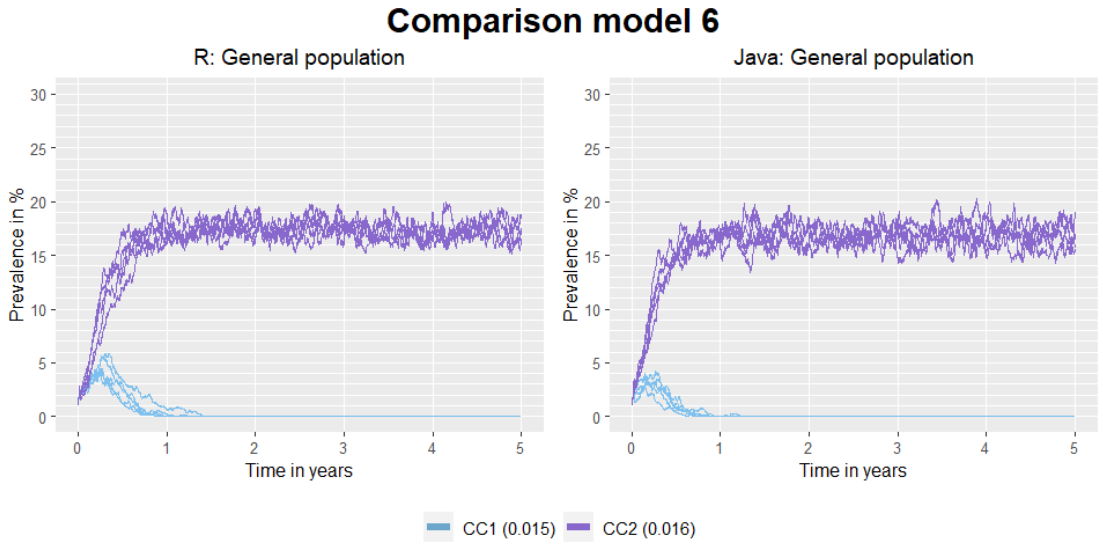


Figure 4.8: **Model 6.** Includes: Two CCs, general population infections, household infections. Excludes: hospitals, antibiotics. The CCs are assumed to have different infection probabilities and 5 runs both the R model (left) and Java model (right) are presented.

Also in the case where household infections are added to the population, it can be seen in Figure 4.8 that when the infection probabilities of the CCs are distinct, the CC with the highest infection probability will become dominant. Just as in Figure 4.5 the other CC will disappear from the population after some time. When considering the runs shown in this figure the equilibrium towards which the more infectious CC moves is comparable for the R and Java model. The 5 runs presented in these graphs do show a slight variation when considering the prevalence of CC1. For some runs of the R model the light blue peak appears to be slightly higher than for the Java model. This is possibly explained by the fact that, by coincidence, in these runs the number of infections of CC2 does not increase as fast as for the other runs. This corresponds to the slightly lower purple line in the beginning. This leaves more susceptibles for also the less infectious CC to infect and could therefore explain the slightly higher peak.

Although the height of the peak of the less infectious CC might differ a little bit between simulations, the runs for the R and Java model show similar behaviour. Just as in Model 5.1, it can be seen that over time, one CC becomes dominant and the other one disappears from the population.

Overall, the R and Java model show similar behaviour in all the simple model set-ups that were studied. More complex models would make the comparison challenging, as not all features are comparable or even included in both models. In the next chapter a more intricate model set-

up will be considered. The computational and modelling limits of the R model resulted in the decision to only consider the Java model in the second part of the project.

Table 4.1: Parameters of the different comparison models studied in Section 4. All distributions were assumed to be constant.

| | **Model 1** | **Model 2** | **Model 3** | **Model 4** | **Model 5** | **Model 6** |
|---|---|---|---|---|---|---|
| Population size | 10.000 | | | | | |
| Infection duration | 8 days | | | | | |
| Initial infections | 100 for each CC | | | | | |
| Number of CC | 1 | 1 | 1 | 1 | 2 | 2 |
| Contacts general population | 10 | 5 | 5 | 5 | 10 | 5 |
| Contacts household | - | 10 | 10 | 10 | - | 10 |
| Contacts Hospital | - | - | 15 | - | - | - |
| Hospitalisation rate (yearly) | - | - | 0.73 | - | - | - |
| Length hospital stay | - | - | 12 days | - | - | - |
| Antibiotics usage rate (yearly) | - | - | - | 10 | - | - |
| Antibiotics clearance probability | - | - | - | 0.3 | - | - |

# 5 COMPLEX SPREAD OF MRSA CCS USING THE JAVA MODEL

The previous chapter presented a comparison of the basic dynamics of the R and Java model and the implementations showed similar behaviour in straightforward settings. In this chapter the performance and results of the Java model will be considered in a more complicated setting. This was done as a first attempt to show how the IBM can give more insight in the spread of MRSA CCs within a population. Similar simulations were originally done with an older version of the R model. This original R model contained certain mistakes that have most likely influenced the results and because the computation time covered multiple days, the decision was made not to redo the simulations. The findings of the original simulations are presented in Appendix A. In this chapter first the model set-up will be discussed in Section 5.1. In the second part of the chapter the results of the model are presented by considering two different settings: no interventions (5.2.1) and a search-and-destroy policy (5.2.2).

## 5.1 MODEL SET-UP

In this section the parameters and distributions used to obtain the results, that will be discussed later, are explained. A summary of the parameters and distributions can be found in Appendix B. It is important to note that the parameters chosen in this section of the research are not necessarily realistic. Because the aim of this part of the study is to find out how various factors influence the spread of multiple MRSA CCs, a fictional population was created. To restrict the computational cost, a limited population size has been chosen. To be able to study the influence of certain factors, such as spread in the hospital, some parameters are not chosen as realistic values, but rather as suitable for the fictional population that is modelled. These choices will be presented and explained in the paragraphs below.

For each run a population of 10.000 individuals is modelled. Although the Java model is not as computationally expensive as the R model, larger populations would make the run time significantly longer. As this part of the study is still exploratory research, the decision was made to focus on a smaller population and thereby limit the computation time. In the population four nursing homes are included, each with 100 residents. Furthermore two hospitals have been incorporated, where halve of the population is assigned to one hospital and halve to the other. For the distribution of household sizes the distribution shown in Figure 5.1 is used, which is very similar to the distribution of the Netherlands[40]. Each run, household sizes are drawn from this distribution at initialisation and for 10.000 individuals this results in around 4000 households.

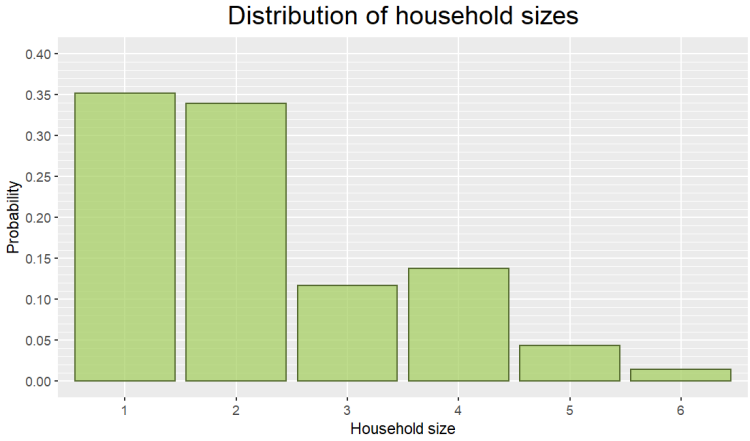


Figure 5.1: Distribution for the household sizes in the complex model set-up. The household sizes are based on the distribution of the Netherlands[40].

The hospitalisation rate is set the same for all individuals, with an additional rate for nursing home residents. It is assumed that all individuals in the population have a yearly hospitalisation rate of 1.5. The additional hospitalisation rate of nursing home inhabitants is set to 3.0. This makes them three times more likely to be hospitalised compared to individuals in regular households. Whenever an individual enters hospital they are assigned a length of stay which is at least one day, with an additional time span drawn from a Negative Binomial distribution with an average stay of 5 days and shape parameter of 0.8. This gives the distribution shown in Figure 5.2. This Negative Binomial distribution was chosen, because its long tail results in a distribution with high probabilities of being in the hospital for a short amount of time, but also smaller probabilities of being there for a longer period of time.

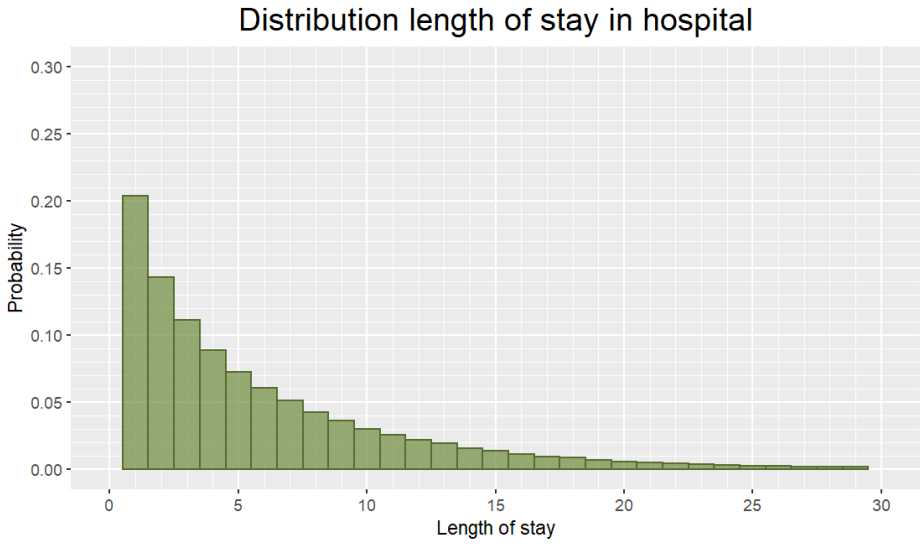


Figure 5.2: Distribution for the length of stay in the hospital. At hospitalisation each individual is assigned a length of stay drawn from the presented distribution.

With the parameters described above, the number of patients in each hospital varies roughly between 110 and 150. This means that more than 1% of the population is in hospital at each point in time. In realistic situations this percentage of hospitalised individuals is much lower. However, because we want to include and study the spread of MRSA within hospitals, we need a relatively large hospital population with respect to the size of the overall population.

At initialisation all individuals are assigned to a clearance type: 50% has clearance type I, 30% clearance type II and 20% clearance type III, as can be seen in Figure 5.3. The choice for these different clearance types was based on previously mentioned studies that showed that humans essentially can be subdivided in multiple groups: persistent carriers (~20% of humans), intermittent carriers (~30% of humans) and non-carriers (~50% of humans)[36–38]. When an individual becomes infected an indication for the infection duration is drawn from a gamma distribution with a mean of 1.0 and shape parameter of 3.0. This number is then multiplied by a clearance type factor which is different for each clearance type. For clearance type I the multiplication factor is 1, for type II it is 8 and for type III it is 40. This results in different distributions for the infection duration for each of the clearance types, which are shown in Figure 5.4.

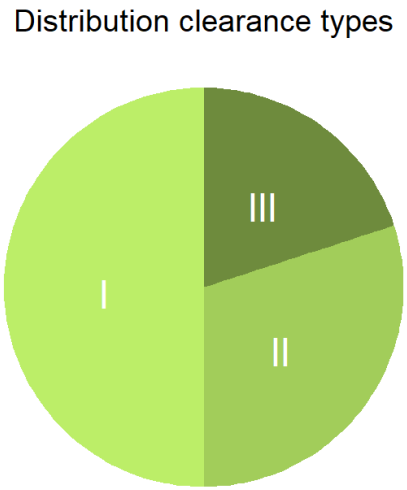


Figure 5.3: Distribution of different clearance types in the population. 50% of the population has clearance type I, 30% clearance type II and 20% clearance type III.

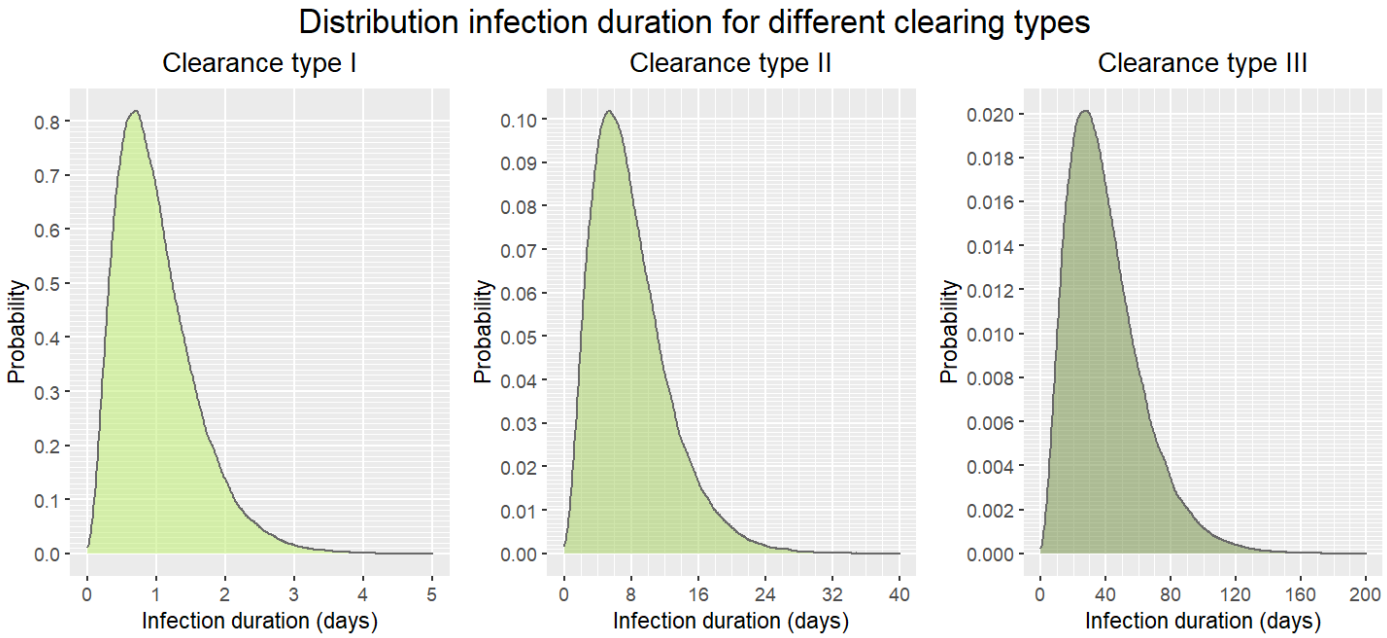


Figure 5.4: Distribution of the infection duration for each of the different clearing types. The mean time to clearance is 1 for type I, for type II it is 8 and for type III 40.

In the model 6 CCs are included. One of these (CC0) is an MSSA CC; the other ones are CCs of MRSA (CC1-CC5). For CC0 it is assumed that the probability of infection per infected contact is 0.02. For the other CCs the probability is set to 0.0185. Each CC is assigned a certain resistance to antibiotics. The probability of clearance when antibiotics are taken for each of the CCs can be found in Table 5.1. CC0 is the MSSA CC and is therefore most susceptible to antibiotics. CC5 has the smallest chance of clearance and therefore is of all CCs most resistant to antibiotics. It was assumed that everyone in the population is at risk for external infections. The external infection force was set to 0.000008 per exposed individual per day for all MRSA CCs. This resulted in around 25 external infections each year for CC1-CC5.

Table 5.1: Susceptibility of the CCs to antibiotics. Probability of clearance when antibiotics are taken for each of the CCs.

| CC | clearance probability | | |
|---|---|---|---|
| CC0 | 0.95 | CC3 | 0.12 |
| CC1 | 0.20 | CC4 | 0.08 |
| CC2 | 0.16 | CC5 | 0.04 |

A mean number of contacts is assumed in each of the different infection settings. Within the general population a mean of 5.0 contacts is assumed and within households and nursing homes 10.0. In the hospital it is assumed that patients have 15.0 potentially infectious contacts. At initialisation each individual is assigned a contact rate both for their household contacts (or nursing home contacts) and their general population contacts. These contact rates are drawn from a gamma distribution with the mean number of contacts as described above and with a shape parameter of 2.0, which results in the distributions shown in Figure 5.5. The hospital contact rate is not drawn at the beginning of the simulation, but is drawn at every hospitalisation event. The contact rate is drawn from a gamma distribution with shape 2.0 and mean 15.0, which is shown as the bottom right density in Figure 5.5. At a next hospitalisation event a new hospital contact rate is drawn.

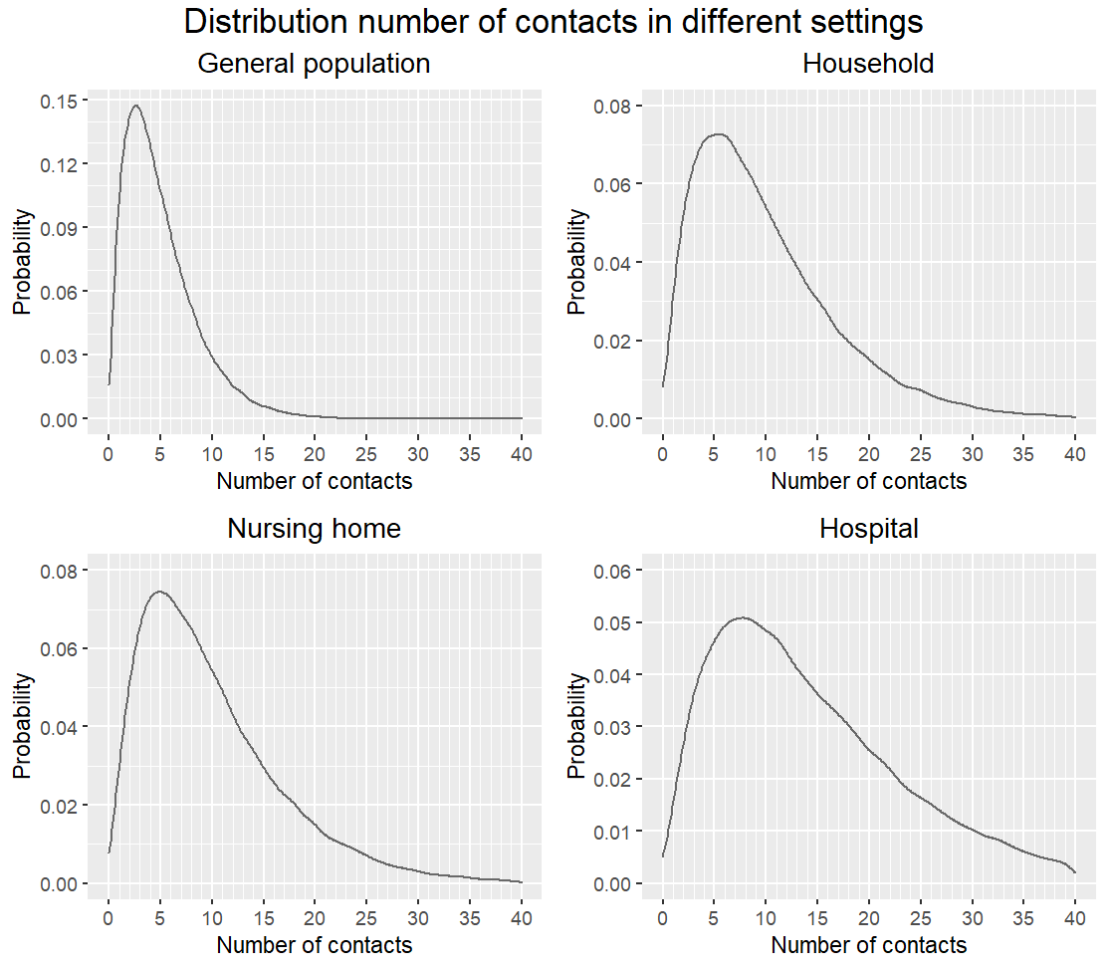


Figure 5.5: Distributions for the number of contacts in the different infection settings. Top-left: general population; top-right: household; bottom-left: nursing home; bottom-right: hospital.

Lastly there is a possibility for isolation of hospital patients that are carriers of MRSA. In the second part of the runs a search-and-destroy policy was included to see how interventions can influence the spread of MRSA within a population. The fraction of hospital patients that was tested for MRSA each day was set to 0.25. It was assumed that the isolation was 99% effective, so only 0.01 of all contacts remained when an individual was placed in isolation.

## 5.2 Results

The model set-up described in the previous section was used to obtain the results that will be discussed in this section. Two different scenarios will be presented: one where no interventions are taken and one where a search-and-destroy policy is employed. In total for each scenario 100 runs were performed and both the individual behaviour of the runs and general patterns will be discussed.

### 5.2.1 No intervention

The first scenario that will be focused on is where no interventions are included to stop the spread of MRSA CCs within the population. In Figure 5.6 the prevalence of MRSA CCs in the general population is shown for 9 different runs. The figures display the prevalence in the general population over a period of 30 years and the prevalence of each CC is indicated by a different coloured line, where CC1 is least resistant to antibiotics and CC5 most resistant. At the start of the simulation no MRSA CCs are present in the population. The MSSA CC, CC0, is present at the start of the simulation and for most runs remains at a prevalence of around 20%. To make it easier to distinct the prevalence of the MRSA CCs, the prevalence of the MSSA CC was mostly excluded from the figures as a result of the choice of vertical axis. The introduction of an MRSA CC in the population is caused by external infections. In the runs presented in Figure 5.6 it can be seen that all CCs make an introduction in the population, although not always successful.

Each of the runs in Figure 5.6 show different peaks in the general population prevalence for multiple CCs. Although the peaks in run 1, 2, 4, 5 and 7 never move above the 2% prevalence, run 3, 6, 8 and 9 exhibit relatively large peaks in the number of MRSA infections. In most of the runs, there does not seem to be one CC that is dominant over the entirety of the 30 year period. However, in run 3, 6 and 8 it can be seen that a lot of the prevalence in the population can be attributed to CC5. Run 6, especially, demonstrates a spread of CC5 that is very different from the other runs.

In Figure 5.7 the prevalences of the CCs in run 6 is presented, but now using a larger scale for the vertical axis. The blue line illustrates the prevalence of the susceptible CC over time and it can be seen that for the first 20 years this MSSA CC is the dominant CC in the population. The prevalence of CC0 moves stably around the 18%. However, after around 15 years the prevalence of CC5 starts increasing a lot. This appears to go hand in hand with a decrease of the until then most prevalent CC0. At one point the increase of CC5 becomes so large that it completely takes over the dominant position from CC0. It can be seen that the MSSA after some time even completely disappears from the population. This specific scenario where MRSA becomes more prevalent than MSSA does not occur in a lot of runs, but one can imagine that in reality this would be a very problematic situation.

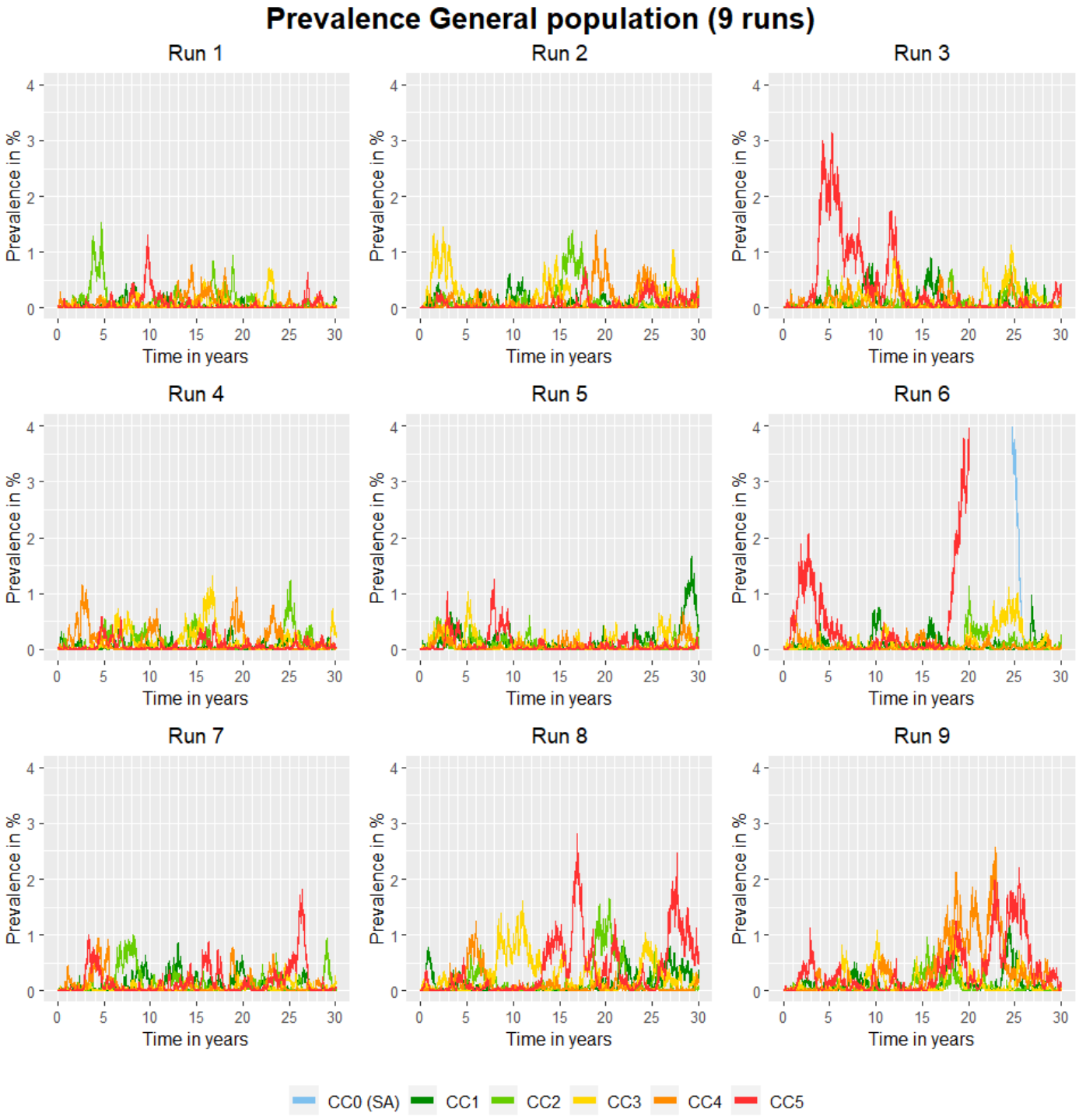


Figure 5.6: Prevalence of MRSA CCs in the general population for 9 randomly selected runs, when no interventions are employed. The prevalence is presented over a period of 30 years.

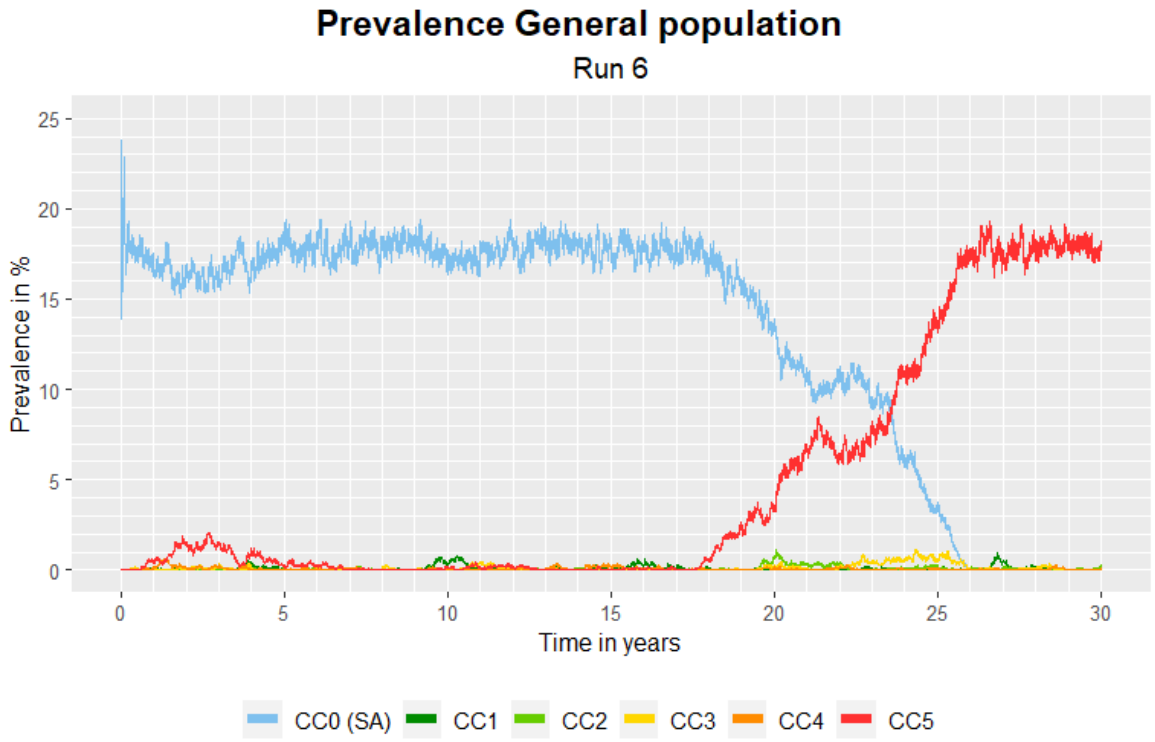


Figure 5.7: Prevalence of CCs in the general population for run 6. The prevalence is presented over a period of 30 years.

In Figure 5.8 the mean prevalence of the MRSA CCs within the general population is shown over the period from 20 to 30 years for the same 9 runs presented in Figure 5.6. This specific period was considered to make sure that the number of people in hospital was stabilised and to give the CCs enough time to spread within the population and become dominant. The diagrams show the contribution of each of the CCs to the MRSA prevalence within the general population. The percentages displayed in the diagrams refer to the mean total prevalence of MRSA within the general population over the studied period. In run 6, 7, 8 and 9 it can be seen that during this period, CC5 was the CC that contributed most to the MRSA prevalence in the population. In run 1, 2 and 3 this was either CC3 or CC4. The least resistant CCs, CC1 and CC2, dominate the number of infections only in run 4 and 5. Figure 5.7 illustrated that in run 6 CC5 became the most prevalent CC, even more prevalent than the MSSA CC (CC0). This can also be recognised in the pie diagram, where it takes over a very large proportion of all MRSA infections and with an overall very high total MRSA prevalence in the population.

The diagrams from Figure 5.8 further contribute to the idea that a higher resistance to antibiotics increases the possibility of being the dominant CC, since CC5 in most runs has the highest prevalence. It is interesting to note that in runs where the dominant CC was one with a high resistance, overall the MRSA prevalence was relatively high. This can for example be seen if we consider run 6, 7, 8 and 9, where CC5 was most prevalent. These 4 runs have the highest total MRSA prevalences of all runs. This possibly suggests that when a CC with relatively low resistance is the most successful CC, the total prevalence is lower than when a CC with high resistance is most successful. This possible relation will be researched in more detail.

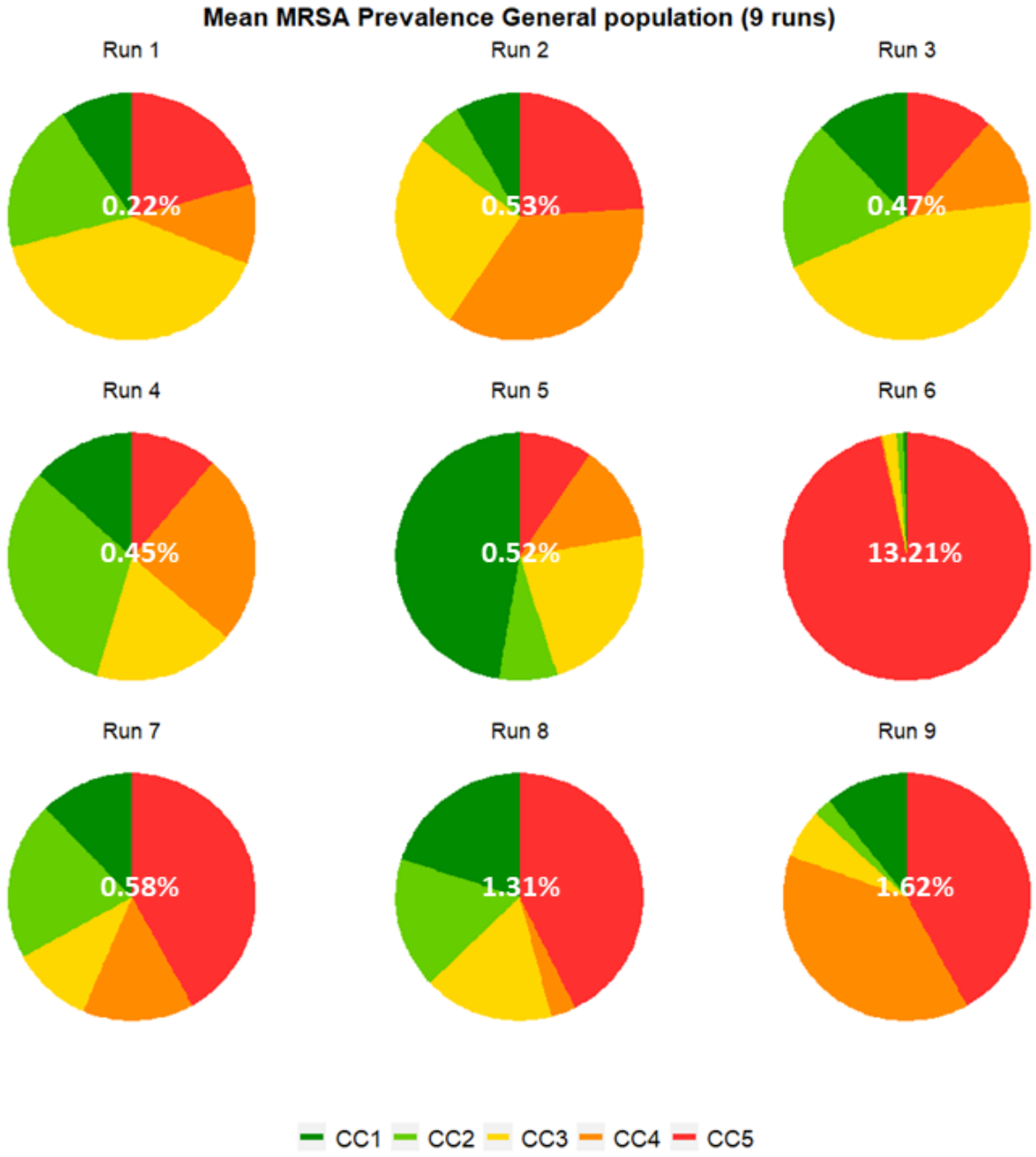


Figure 5.8: Mean prevalence of the MRSA CCs within the general population over a period from 20 to 30 years for 9 randomly selected runs, when no interventions are employed. Each fraction illustrates the proportion of MRSA infections that can be attributed to that specific CC. The percentages displayed in the diagrams refer to the mean total prevalence of MRSA within the general population over the studied period.

Figure 5.9 shows the percentage of the winning CC against the total prevalence of MRSA in the general population for 100 runs. The runs where an MRSA CC takes over the dominant role from the MSSA CC, are not presented in this figure, as they fall outside the range chosen for the vertical axis. The winning CC is the CC that contributes most to the prevalence of MRSA in the general population over the period from 20 to 30 years. The first thing that can be noted is that the red and orange dots make up the majority of all dots. This means that overall, the CCs with a higher resistance are more likely to be the winning CC. The second thing that can be noted is that the green dots are located mostly in the bottom left of the figure. This means that if CC1 or CC2 is the winning CC, they are not likely to dominate the prevalence by a lot. Moreover, the low positioning

in the vertical direction indicates that the event where a lower resistant CC is the winner often coincides with a low overall MRSA prevalence.

One might suggest that a positive (increasing) relation can be seen in Figure 5.9. This indicates that the events where the percentage of a winner are high often coincides with an overall high prevalence in the general population. More research has to be done to see if this relation holds up in other circumstances and to study the possible cause for the linking of these events.

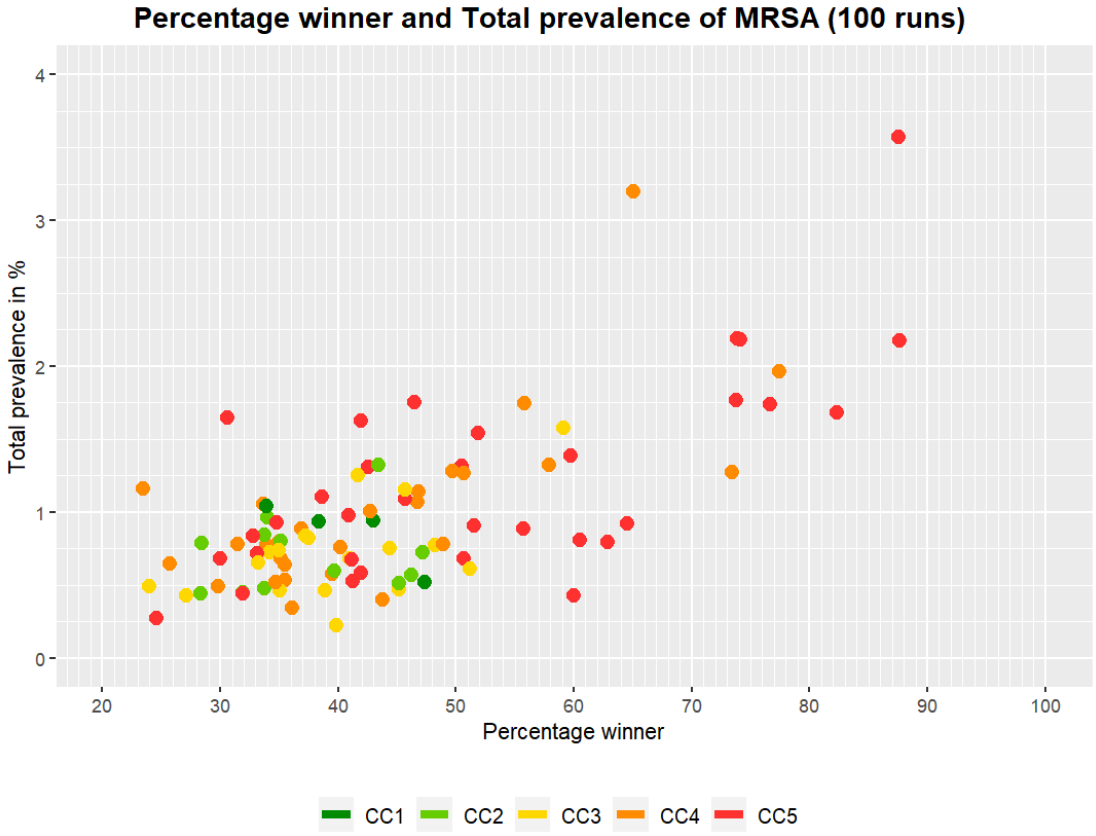


Figure 5.9: Relation between the percentage of the winner (horizontal direction) and the total prevalence of MRSA (vertical direction) for 100 runs, when no interventions are employed. The colour of each dot represents the winner of that run.

So far only the prevalence in the general population has been presented. The diagrams in Figure 5.10 show the contribution of each of the CCs to the MRSA prevalence averaged over 100 runs in different settings. Again, the percentages displayed in the diagrams refer to the mean total prevalence of MRSA averaged over 100 runs for the studied period. The diagrams show the distribution of CCs in different parts of the population: the general population, the hospital and nursing homes. The first thing that can be seen, is that averaged distribution of the CCs is similar for all three settings. Over all runs, CC5 contributes most to the prevalence in each of the settings, where the more susceptible CCs contribute less.

When observing the prevalences in the different settings it can be seen that the prevalence within the hospitals and nursing homes is higher than in the general population. Because the number of contacts in the hospital is relatively high, CCs can spread easily in that setting. A newly hospitalised patient can import a specific CC into the hospital and from there on it can spread fast due to the many contacts, which explains the high prevalence in the hospital setting. In the model

set-up described in Section 5.1 the additional hospitalisation rate of nursing home residents was discussed. Since inhabitants of nursing home are more frequently hospitalised than humans in the regular population, they are also more likely to become infected inside the hospital during their stay. After their hospitalisation, nursing home residents can further spread that CC in their respective nursing homes. Their more frequent exposure to the higher prevalence setting of the hospital makes them more exposed to MRSA infections than the general population.

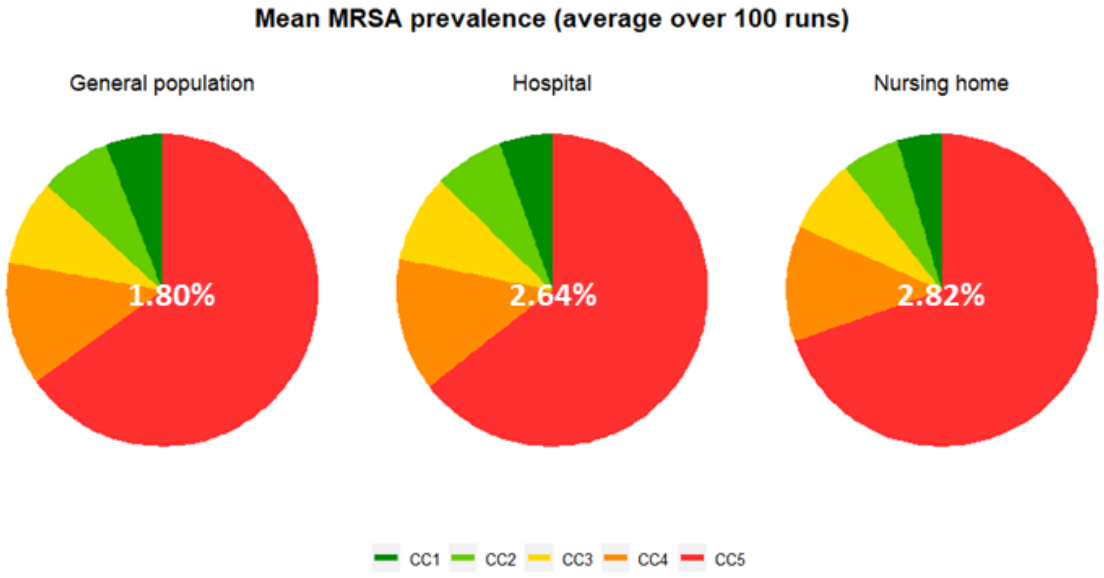


Figure 5.10: Mean prevalence of MRSA CCs averaged over 100 runs in three groups: general population, hospitals and nursing homes, when no interventions are taken. Each fraction illustrates the proportion of MRSA infections that can be attributed to that specific CC. The percentages displayed in the diagrams refer to the mean total prevalence of MRSA averaged over the 100 runs.

Since Figure 5.10 only presents the average over all the runs, additionally a table was created that shows the amount of successes of the different CCs over the 100 runs. Table 5.2 demonstrates for each of the MRSA CCs the mean prevalence in the general population and how often they came out of a run as the winning CC. The table also presents how many times that CC was an absolute winner. An absolute winner is defined as a CC that contributed to more than 75% of all MRSA prevalence in that run. One can immediately note that from CC1 to CC5 the averaged mean general population prevalence moves upwards. This further supports the believe that higher resistance positively contributes to the prevalence of a CC in the population.

This can also be seen when looking at the number of times the different CCs are the winner of a run. The integers presented in Table 5.2 are the number of runs that a specific CC turned out as the winner. The table shows that in general, the higher the antibiotic resistance of the CC, the more likely it is to be the winner of a run. CC5 ended up as the winner more than a thirth of the 100 runs. The two most resistant MRSA CCs were determined as the winning CC in more than 60% of the 100 runs.

The last entry shown in Table 5.2 is the number of times a CC turned out as the absolute winner of a run. Similar as to being the winner, it can be seen that a higher resistance makes it more likely that a CC will be an absolute winner. It is even the case that during these 100 runs only CC4 and CC5 ever turned out to be an absolute winner. CC5 was an absolute winner in 10% of the runs and CC4 only once. This suggests that a higher resistance not only increases the change of being most successful, but also increases the proportion of its success.

Table 5.2: Properties of CCs for 100 performed runs without interventions. The table includes the mean average prevalence, the number of times a CC ends up as the winner of a run and the number of times that CC is an absolute winner. *In this percentage also the runs where a MRSA CC becomes more prevalent than the MSSA CC are included. If these runs were excluded the mean average prevalence would be 0.40%. ** If the runs are excluded where an MRSA CC becomes more dominant than the MSSA CC the number of absolute wins would be 5.

| | CC1 | CC2 | CC3 | CC4 | CC5 |
|---|---|---|---|---|---|
| Mean average prevalence | 0.11% | 0.13% | 0.16% | 0.23% | 1.17%* |
| Winner | 4 | 12 | 18 | 27 | 39 |
| Absolute winner (>75%) | 0 | 0 | 0 | 1 | 10** |

### 5.2.2 SEARCH-AND-DESTROY

The second model set-up that will be considered includes a search-and-destroy policy. The same parameters are used as for the runs without interventions. Additionally, 0.25 of all hospital patients is tested each day and carriers of MRSA can be placed in isolation. In Figure 5.11 the prevalence of MRSA within the general population is shown for 9 different runs which include the search-and-destroy policy over a period of 30 years. The same scale was chosen for the axes as in the runs without interventions (Figure 5.6 in order to make comparison easier.

From Figure 5.11 one can immediately notice that overall the general population prevalence, when incorporating the search-and-destroy policy in the hospital, is significantly lower than when no measures are taken. In contrast to the results without interventions in Figure 5.6 in most runs no specially successful and dominating CC can be distinguished. Another difference that is important to note, is that the peaks caused by the MRSA CCs are a lot less high and harder to distinguish when the search-and-destroy policy is employed.

In Figure 5.12 the pie diagrams for the mean prevalence of the MRSA CCs within the general population are presented over the period from 20 to 30 years for the same 9 runs shown in Figure 5.11. The diagrams show the contribution of each of the CCs to the MRSA prevalence in the general population when a search-and-destroy policy is employed. The percentages displayed in the diagrams again refer to the mean total prevalence of MRSA within the general population over the studied period. When comparing it to Figure 5.6, the first thing that can be seen in Figure 5.12 is that average total prevalence of MRSA is much lower with a search-and-destroy policy than without any interventions being taken.

From the diagrams shown in Figure 5.12 it furthermore can be observed, that in most runs often a large variety of CCs contributes to the prevalence of MRSA in the general population. Not only CCs with a high resistance, such as CC9 in run 9, but also CCs with a low resistance, such as CC1 in run 7 and CC2 in run 1 and 5, are able to attribute large parts of the infections to the total MRSA prevalence. These contributions coincide with the small peaks that were displayed in Figure 5.11. Multiple CCs have smaller peaks, which results in multiple CCs contributing to the MRSA prevalence.

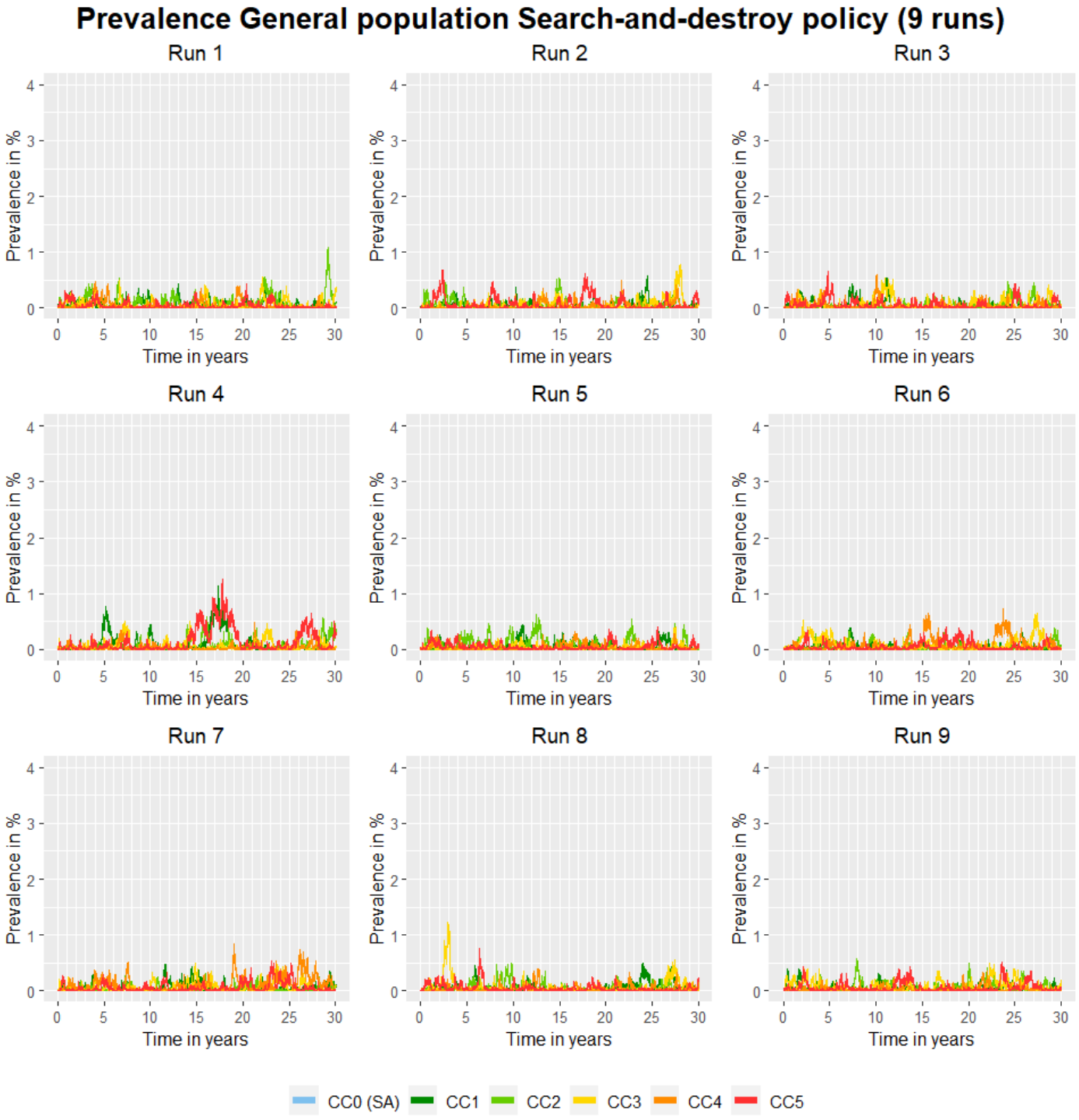


Figure 5.11: Prevalence of MRSA CCs in the general population for 9 randomly selected runs, when a search-and-destroy policy is employed. The prevalence is presented over a period of 30 years.

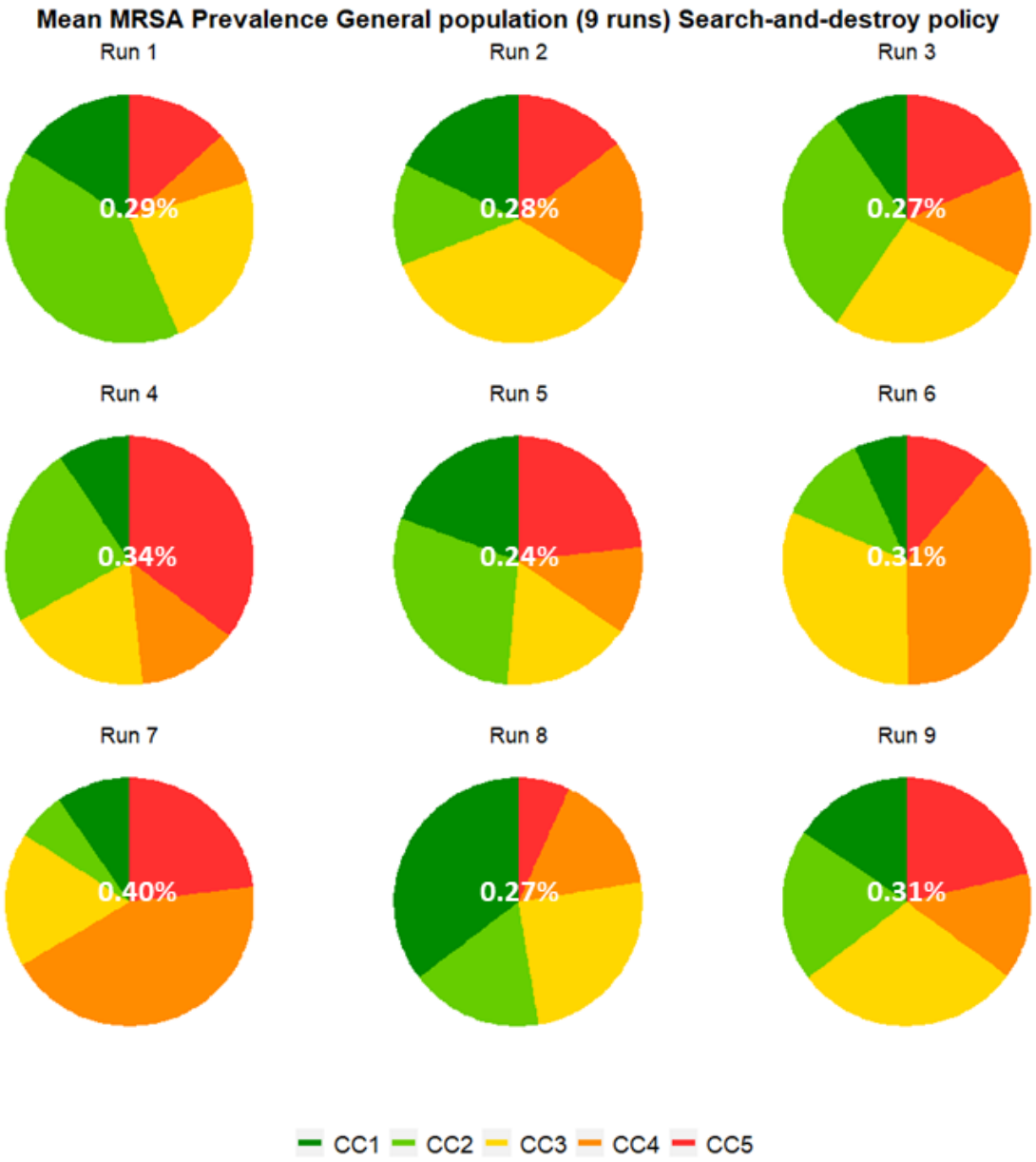


Figure 5.12: Mean prevalence of the MRSA CCs within the general population over a period from 20 to 30 years for 9 randomly selected runs, when a search-and-destroy policy is employed. Each fraction illustrates the proportion of MRSA infections that can be attributed to that specific CC. The percentages displayed in the diagrams refer to the mean total prevalence of MRSA within the general population over the studied period.

The fact that multiple CCs contribute to the MRSA prevalence, can also be established when considering the relation between the percentage of the winning CC and the total MRSA prevalence shown in Figure 5.13. Note that the same range is used for the horizontal axis as in Figure 5.9, but the vertical axis is adjusted. In the diagrams from Figure 5.12 it could be seen that each run a large variety of CCs contributed to the prevalence in the general population. This can also be concluded by analysing the location of the coloured dots shown in this figure. It can be seen that for almost all 100 runs the prevalence percentage that can be attributed to the winner is below 50%. In the same figure but without interventions (Figure 5.9), the percentage of a winner ranges from around 25% to almost 90%. This suggests that implementation of the search-and-destroy policy creates more variation between the CCs present in the population.

What additionally can be noted is that, where in Figure 5.9 the largest part of the presented dots were red or orange, in Figure 5.13 more green dots can be spotted. This leads to believe that the implementation of a search-and-destroy policy increases the changes for also less resistant CCs, such as CC1 and CC2, to become the most prevalent MRSA CC. Without interventions it could be observed that the dots representing the lower resisting CCs are always located on the left part of the horizontal axis and the higher antibiotic resistant CCs often had a higher winning percentage. In Figure 5.13 this relation can no longer be seen.

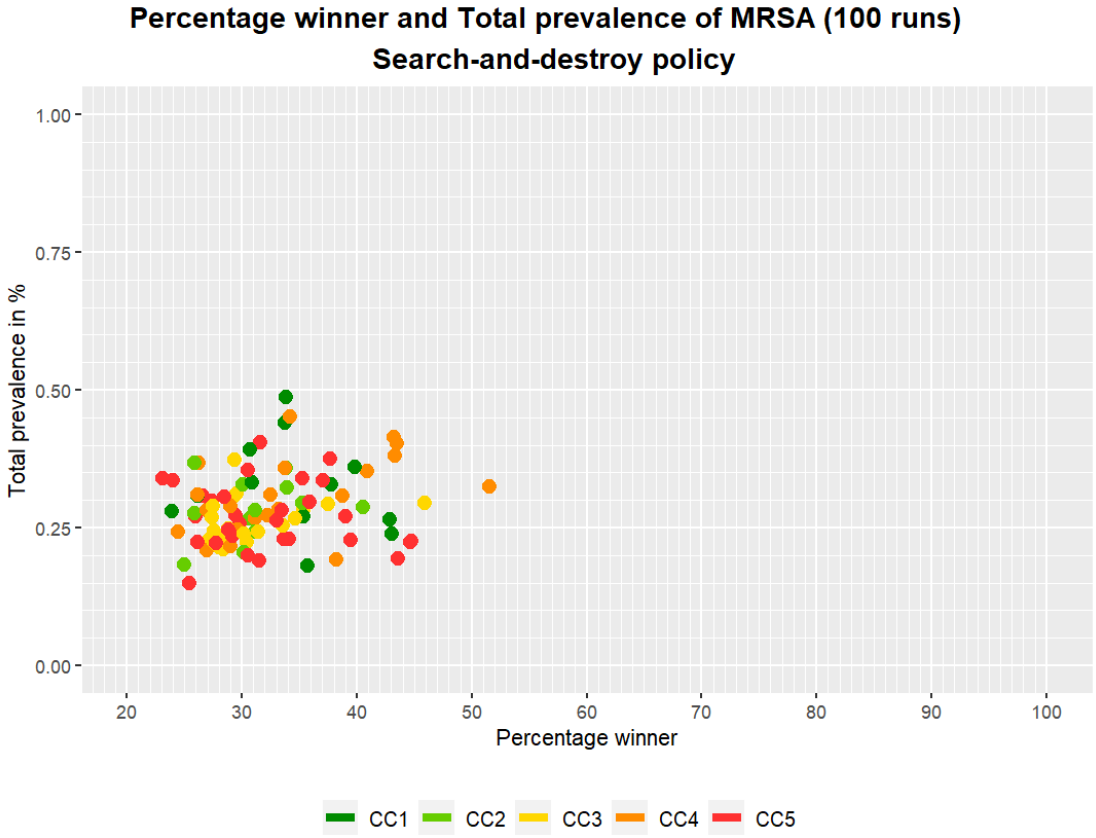


Figure 5.13: Relation between the percentage of the winner (horizontal direction) and the total prevalence of MRSA (vertical direction) for 100 runs, when a search-and-destroy policy is employed. The colour of each dot represents the winner of that run.

From the percentages presented in Figure 5.12 it was apparent that the search-and-destroy policy resulted in a relatively low total MRSA prevalence in these 9 runs. If we now focus on the vertical location of the dots in Figure 5.13, it can similarly be seen that the total MRSA prevalence for the different runs is much lower when applying a search-and-destroy policy. The same can also be recognised when we consider the averaged MRSA prevalence over the 100 runs. The diagrams in Figure 5.14 show the contribution of each of the CCs to the MRSA prevalence within the general population, hospitals and nursing homes. Again, the percentages displayed in the diagrams refer to the mean total prevalence of MRSA averaged over 100 runs for the studied period.

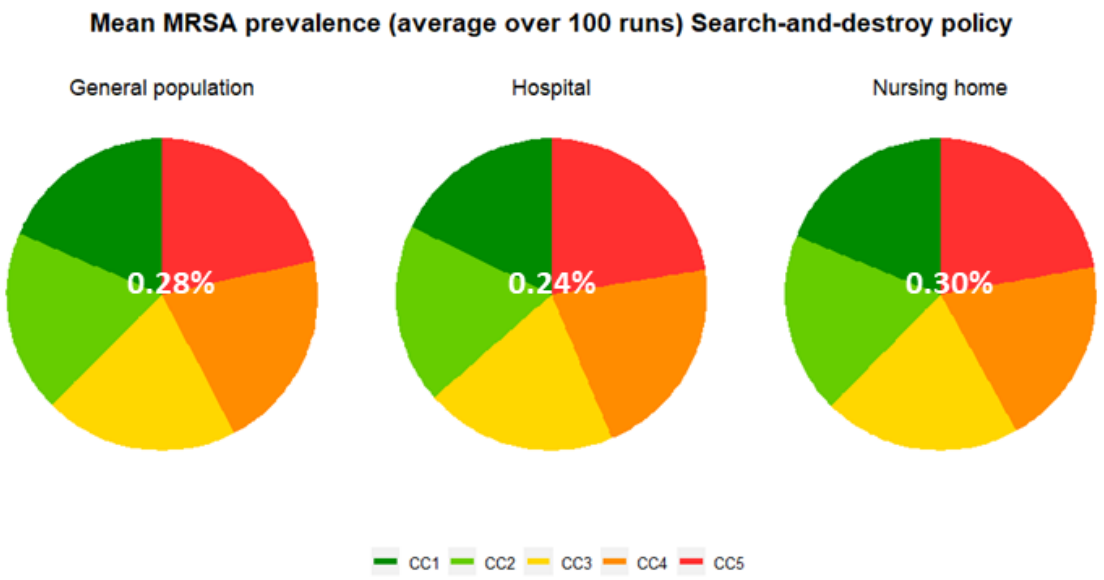


Figure 5.14: Mean prevalence of MRSA CCs averaged over 100 runs in three groups: general population, hospitals and nursing homes, when a search-and-destroy policy is employed. Each fraction illustrates the proportion of MRSA infections that can be attributed to that specific CC. The percentages displayed in the diagrams refer to the mean total prevalence of MRSA averaged over the 100 runs.

If we compare the percentages in Figure 5.14 to those of Figure 5.10 it can be seen that the overall prevalence in each of the settings was drastically decreased by the implementation of a search-and-destroy policy. The highest MRSA prevalence can still be found in the nursing home, but it can be seen that due to the strict search-and-destroy policy in the hospitals, the prevalence in said hospitals is even lower than in the general population.

If we look at the averaged contribution of the CCs in Figure 5.14 to the total MRSA prevalence in each of the settings, it can be seen that this is distributed relatively even. Where Figure 5.10 demonstrated that higher resistant CCs are more likely to contribute to the number of infections, the implementation of a search-and-destroy policy seems to diminish this relation. Although the contributions of CC4 and CC5 are still slightly bigger than those of the lesser resistant CC1 and CC2, the different fractions have become more similar. The figure shows that in general, applying a search-and-destroy policy increases the variation between the CCs in the population.

Since Figure 5.14 only shows the average over 100 runs, again a table was created that shows the distribution of successes of the different CCs over the 100 runs. Table 5.3 demonstrates for each of the CCs their mean prevalence in the general population and how often they came out of a run as the winning CC. The table also shows how many times that CC was an absolute winner. Similar as in Table 5.2, one can note that from CC1 to CC5 the mean prevalence moves upwards. However, the implementation of the search-and-destroy policy has not only significantly decreased this mean prevalence, but also reduced the differences in prevalences between the different CCs.

The same consequences of the search-and-destroy policy can also be observed when considering the number of times the different CCs are the winner of a given run. Table 5.3 shows that in general, the higher the resistance of the CC, the more likely it is to be the winner of a run. Although, the two most resistant MRSA CCs were determined as the winner in more than 50% of the 100 runs, more than 25% of the wins can be attributed to the two least resistant CCs. The implementation of the search-and-destroy policy has resulted in a more even spread of the winnings over the different CCs.

The last entry shown in Table 5.3 is the number of times a CC turned out as the absolute winner of a run. When no measures are taken to prevent the spreaf of MRSA, CC5 was an absolute winner in 10% of the runs. Including the search-and-destroy policy in the model, resulted in the fact that no CC came up as an absolute winner in any of the runs. As was concluded from Figure 5.13, this suggests that implementation of the search-and-destroy policy creates more variation between the CCs present in the population.

Table 5.3: Properties of CCs for 100 performed runs when a search-and-destroy policy is employed. The table includes the mean average prevalence, the number of times a CC ends up as the winner of a run and the number of times that CC is an absolute winner.

| | CC1 | CC2 | CC3 | CC4 | CC5 |
|---|---|---|---|---|---|
| Mean average prevalence | 0.053% | 0.054% | 0.057% | 0.059% | 0.061% |
| Winner | 13 | 13 | 20 | 23 | 31 |
| Absolute winner (>75%) | 0 | 0 | 0 | 0 | 0 |

# 6 DISCUSSION

In the course of this project the R implementation for the spread of MRSA CCs illustrated many restrictions. The biggest limitation was the computation time of the model. Due to the very large and inefficient data structures that were used, the model simulations often took many hours, even after the optimisation discussed in Section 2.9. As a result of the large computation time, it was extremely time consuming to perform multiple runs and modelling large populations was not feasible. Furthermore, because the model uses a time stepping algorithm with little room to incorporate additional mechanisms, such as adjusted susceptibility after infection, the model is very limited in the amount of options for the model set-up. Another disadvantage of the R model is the limited output. Since the total infection force experienced by an individual is made up of the infection force in different settings, it is not possible to see which route is the cause of an infection.

Almost all limitations of the R model were solved with the implementation of the Java model. Java implementations of mathematical models are inherently fast and the event-based approach and the use of efficient storage, drastically improved the computation time with respect to the R model. For a model set-up that existed of 5 CCs in a relatively complex population, the R model took 45 minutes for one run, where the Java model only needed around 90 seconds. Additionally, in the implementation of the Java model much more options for the user are incorporated. The model includes different settings, such as multiple household transmission structures, hospital personal and susceptibility changes over time, that could not be incorporated in the R model. Where the R model could only output general information, the Java model has the possibility for more detailed information. Not only is it possible to output for each infection its source, it is also possible to examine specific households and individuals. One of the disadvantages, however, of this complicated and not so comprehensible model is that it can become a black box. During multiple stages of this project it was unclear how different settings impacted the model. It is advisable to construct future models via a step-by-step approach and to pay a lot of attention to the parameters and distributions that are introduced.

For the comparison of the two models very simple model set-ups were used. All distributions were taken as constant to prevent any mistakes caused by differing distributions between the two models. For two of the model set-ups, Model 1 and Model 5, the results of the R and Java model are presented next to a deterministic model, by only considering general population contacts. Three infection probabilities were studied and the prevalence in the runs of the IBMs appear to go to the same equilibria as the deterministic model. The similarity between the deterministic model and the two IBMs gives confidence that the different implementations exhibit the same behaviour regarding general population contacts and infections. Although the methods to model the spread of one or multiple CCs within the population are slightly different, the R and Java implementations show similar results and are comparable to the equivalent deterministic model.

Other aspects of the model that have been studied in the basic model set-ups were household infections, hospitals and antibiotics. These different features were each covered by a separate model set-up and for all model set-ups the R and Java model exhibited similar behaviour. Both models showed that incorporating hospitals in the population can increase the prevalence in both the hospital, as well as the general population. Incorporating antibiotic use in the general population resulted in a decreased prevalence. Although the methods used to implement these aspects are not the same in the R and Java model, the behaviour of the model set-ups is comparable.

The model set-ups that included the spread of multiple CCs in the population revealed that a CC with an advantage, such as a high infection probability, will become the dominant CC in the population. In the model is it assumed that individuals can only carry one CC at a time and be-

cause of this competition, the less infectious CC will eventually disappear from the population. When the CCs have identical properties, it depends on chance which CC will be more prevalent. Where in a deterministic setting the prevalence of the two CCs will move towards the same equilibrium, the stochasticity of the IBMs can create a situation where one CC is more successful than the other. It could be observed that in most of the runs eventually one of the CCs will disappear entirely from the population due to competition. It can be expected that on a very large time scale for each run this will eventually happen.

The simplified versions of both the R and Java model showed very similar behaviour in all settings. Although the methods of implementation were slightly different, the results of the basic dynamics seemed to be the same. Due to the differences in many features of the model, it is not really possible to compare them in more complicated situations. The Java model has many more features and flexibility, and will therefore be a more viable approach for the modelling of MRSA in realistic settings.

For the second part of this study the Java model was used to simulate the spread of MRSA CCs in a more complex setting. Although the model set-up and the population are not yet realistic, some interesting findings were made, which could possibly also hold for more realistic and complex models. For the first complex model set-up, which does not include any interventions, individual runs demonstrate quite a bit of variation between them. One of the strengths of IBMs is that they can provide more than one possible outcome for the given model set-up. The different runs illustrate that it is possible for every CC to be the most successful, however, the more resistant CCs are more likely to. Plotting the percentage of the winner against the total prevalence in the general population, demonstrated that in the runs where CCs with low resistance are the most dominant CC, the total MRSA prevalence is relatively low. A possible explanation for this could be that if by chance CCs with a high resistance are not successful, there is more room for the less resistant CCs to spread. However for these CCs it is tougher to spread fast and compete with the MSSA in the population and therefore the total prevalence will be lower. It is likely that over a longer time scale a more resistant CC will become dominant.

The average over 100 runs revealed that the prevalence in the nursing homes and hospitals is generally higher than in the general population. Although the antibiotic use in the hospital is frequent, CCs can spread easily in that setting, because the number of contacts in the hospital is relatively high. Additionally, nursing home residents are often hospitalised and they can potentially transmit a CC from the hospital to their nursing home. Since in nursing homes less antibiotics are used, it has the potential to spread even faster than in the hospital. Although indeed it is true that the MRSA prevalence is very high among the elderly and therefore in nursing homes[41], it can be questioned if it is realistic that the prevalence in the nursing homes is higher than in hospitals.

The results of 100 runs showed us furthermore that higher resistant CCs supplied most of the MRSA infections. CC5 was the winning CC in almost 40% of the runs and even sometimes became more prevalent than the MSSA. Although this happened only in 6 out of 100 runs, this of course would be the worst-case scenario. Because the probabilities of the CCs were not yet realistically set and might be too favourable towards the MRSA CC, it is hard to say whether this will actually be possible in more realistic settings.

The second complex model set-up that was considered included a search-and-destroy policy. This search-and-destroy policy consisted of a fraction of hospitalised patients being tested each day. The most apparent result of the search-and-destroy policy is that the overall prevalence in the general populations is significantly lower than when no measures are taken. This actually corresponds with the real world, where countries with a search-and-destroy policy display among

the smallest prevalences of MRSA in the world[14–17]. The pie diagrams presented, illustrate that in the runs often a large variety of CCs contributes to the prevalence of MRSA within the general population. Not only CCs with a high resistance, but also CCs with a low resistance, can be responsible for a large part of the infections. The percentage of the winning CC was never more than 50% when the search-and-destroy policy was applied. The average over the 100 runs furthermore demonstrated that implementation of a search-and-destroy policy creates more variation between the CCs, as all CCs almost evenly contributed to the averaged mean total MRSA prevalence.

In order to correctly model the spread of CCs, the biological progress of MRSA infections has to be accurately represented in the model. The model set-up used to calculate the exploratory results of Chapter 5, did not include any adjustments in susceptibility after an infection or antibiotic treatment. However, different studies have shown that after exposure to an antibiotic treatment a delayed regrowth of surviving bacteria can be observed, which is referred to as the post-antibiotic effect (PAE)[42, 43]. Antibiotics are capable of having many different effects on surviving bacteria, which can still be detectable after antibiotic treatment is completed. Not only is it likely that the susceptibility of an individual is altered by antibiotics, it can also be expected that due to a decreased bacterial load during treatment, the infectivity of an individual may be decreased. Including a change in susceptibility and infectivity for an individual during and after antibiotic treatment can have significant effects on the spread of MRSA. If for example an individual has a decreased susceptibility after antibiotic treatment, they are less likely to be infected by household members for some time, which could halt the spread of MRSA. A closer study of literature should be performed to understand the effects of antibiotic treatment on the susceptibility and infectivity of individuals. This should then be incorporated in the model to correctly represent the biological process of antibiotic treatment.

Correctly representing the biological processes of MRSA should be the first step in setting up an accurate model for the spread of MRSA CCs. To get close to a realistic situation, the second step would be to model more representational populations. The population sizes of the simulations in this research were relatively small, which resulted in the use of very high and unrealistic hospitalisation rates, in order to be able to study the spread of MRSA within hospitals. A larger population gives the opportunity to more accurately study the role that hospitals play in the spread of MRSA. During this study it was also assumed that all individuals experience an external force of infection. Reality would be better matched if only a fraction of the population would experience this external force. This fraction can represent people travelling abroad or individuals that come in frequent contact with livestock. Incorporating this is in the model will likely create an initial spread more focused around the individuals at risk.

When a more realistic model set-up is used for the population, also the effect of certain interventions can be studied more accurately. The search-and-destroy policy applied in this project consisted of a fraction of hospitalised patients being tested each day. In reality, testing for MRSA only happens for people that have a high risk of being carriers[12, 13]. Although the Java model also includes this option, the decision was made to initially apply this more straightforward approach to select patients for testing. Focusing the search-and-destroy policy on a specific risk group can result in a much more realistic search-and-destroy policy. Although it is likely that by only testing this risk group not all infections will be intercepted, probably a lot less testing has to be done. Incorporating more practical interventions can give an improved perception of the effect of such interventions and how the spread of MRSA CCs within a population can be halted. The flexibility of the Java model makes it possible to further improve the model, when more knowledge is obtained about the spread of MRSA CCs.

# 7 CONCLUSION

The two IBMs studied and presented in this research produce similar results when setting them side-by-side and comparing the basic dynamics. The R model turned out to be computationally expensive and very restrictive. The Java model, on the other hand, was much faster and was more extensive. The Java model is therefore much more suitable for further studies into the spread of MRSA CCs.

The Java model that was used to simulate the spread of MRSA in a more advanced setting, although the model set-up and the population were not yet realistic, demonstrated some interesting findings, that can possibly also hold for more complex models and in real life settings. The general observation was that CCs with a higher antibiotic resistance are responsible for most of the MRSA infections in the population. In the few cases that CCs with a low resistance were the largest CC, the total MRSA prevalence was relatively low.

The model showed that applying a search-and-destroy policy can lower the overall prevalence in the general population significantly, compared to a set-up where no interventions are included. When a search-and-destroy policy is in place, not only CCs with a high resistance, but also CCs with a low resistance, are shown to cause a large fraction of the MRSA infections. The results support the idea that implementation of a search-and-destroy policy creates more variation between the CCs present in the population.

A study of literature should be performed to get an understanding of the effects of antibiotic treatment on the susceptibility and infectivity of individuals. These effects should be incorporated into the model to correctly represent the biological processes that play a role in the spread of MRSA. More realistic settings can be studied by considering larger population sizes and incorporating different risk groups in the population. This more realistic model set-up could be used to study the effect of certain interventions on the prevalence of MRSA CCs more accurately.

Although the model set-ups adopted during this research most likely not realistically represent the biological processes, populations and interventions that play a role in the spread of MRSA CCs, the exploratory results can be used as guiding observations for further studies that include more realistic model set-ups. The results of this research show the potential of the Java model as a promising tool to study the spread of MRSA CCs.

# REFERENCES


[1] Diep BA and Otto M. "The role of virulence determinants in community-associated MRSA pathogenesis". In: *Trends in microbiology* 16.8 (2008), pp. 361–369.

[2] Stefani S, Chung DR, Lindsay JA, Friedrich AW, Kearns AM, Westh H, and MacKenzie FM. "Meticillin-resistant Staphylococcus aureus (MRSA): global epidemiology and harmonisation of typing methods". In: *International journal of antimicrobial agents* 39.4 (2012), pp. 273–282.

[3] Ito T, Kuwahara-Arai K, Katayama Y, Uehara Y, Han X, et al. "Staphylococcal cassette chromosome mec (SCCmec) analysis of MRSA". In: *Methicillin-Resistant Staphylococcus Aureus (MRSA) Protocols*. Springer, 2014, pp. 131–148.

[4] Otto M. "Community-associated MRSA: what makes them special?" In: *International Journal of Medical Microbiology* 303.6-7 (2013), pp. 324–330.

[5] Rijksinstituut voor Volksgezondheid en Milieu. *MRSA*. URL: `https://www.rivm.nl/mrsa`. (accessed on 02.12.2021).

[6] Centers for Disease Control and Prevention. *Methicillin-resistant Staphylococcus aureus (MRSA)*. URL: `https://www.cdc.gov/mrsa/index.html`. (accessed on 02.12.2021).

[7] MP Jevons. ""Celbenin"-resistant staphylococci". In: *British medical journal* 1.5219 (1961), p. 124.

[8] Deurenberg RH and Stobberingh EE. "The evolution of Staphylococcus aureus". In: *Infection, genetics and evolution* 8.6 (2008), pp. 747–763.

[9] Lakhundi S and Zhang K. "Methicillin-resistant Staphylococcus aureus: molecular characterization, evolution, and epidemiology". In: *Clinical microbiology reviews* 31.4 (2018), e00020–18.

[10] van Den Broek IVF, van Cleef BAGL, Haenen A, Broens EM, van Der Wolf PJ, et al. "Methicillin-resistant Staphylococcus aureus in people living and working in pig farms". In: *Epidemiology & Infection* 137.5 (2009), pp. 700–708.

[11] van Cleef BAGL, Monnet DL, Voss A, Krziwanek K, Allerberger F, et al. "Livestock-associated methicillin-resistant Staphylococcus aureus in humans, Europe". In: *Emerging infectious diseases* 17.3 (2011), p. 502.

[12] Vandenbroucke-Grauls C. "Optimaliseren van het antibioticabeleid in Nederland. XII. SWAB-richtlijn voor de behandeling van MRSA-dragerschap". In: *Nederlands Tijdschrift voor Geneeskunde* 152 (2008), pp. 2667–71.

[13] Werkgroep Infectiepreventie RIVM. "Meticilline-resistente Staphylococcus aureus (MRSA)". In: (2017).

[14] Bode LGM, Wertheim HFL, Kluytmans JAJW, Bogaers-Hofman D, Vandenbroucke-Grauls CMJE, et al. "Sustained low prevalence of meticillin-resistant Staphylococcus aureus upon admission to hospital in The Netherlands". In: *Journal of Hospital Infection* 79.3 (2011), pp. 198–201.

[15] Johnson AP. "Methicillin-resistant Staphylococcus aureus: the European landscape". In: *Journal of antimicrobial chemotherapy* 66.4 (2011), pp. iv43–iv48.

[16] Wertheim HFL, Vos MC, Boelens HAM, Voss A, Vandenbroucke-Grauls CMJE, et al. "Low prevalence of methicillin-resistant Staphylococcus aureus (MRSA) at hospital admission in the Netherlands: the value of search and destroy and restrictive antibiotic use". In: *Journal of Hospital Infection* 56.4 (2004), pp. 321–325.

[17] Souverein D, Houtman P, Euser SM, Herpers BL, Kluytmans J, and Den Boer JW. “Costs and benefits associated with the MRSA search and destroy policy in a hospital in the region Kennemerland, The Netherlands”. In: *PLoS One* 11.2 (2016), e0148175.

[18] Song J, Hsueh P, Chung D, Ko K, Kang C, et al. “Spread of methicillin-resistant Staphylococcus aureus between the community and the hospitals in Asian countries: an ANSORP study”. In: *Journal of antimicrobial chemotherapy* 66.5 (2011), pp. 1061–1069.

[19] Mejıa C, Zurita J, and Guzmán-Blanco M. “Epidemiology and surveillance of methicillin-resistant Staphylococcus aureus in Latin America”. In: *Brazilian Journal of Infectious Diseases* 14 (2010), pp. 79–86.

[20] Economics Policy Center for Disease Dynamics. *ResistanceMap*. URL: `https://resistancemap.cddep.org/index.php`. (accessed on 12.06.2022).

[21] Lee AS, De Lencastre H, Garau J, Kluytmans J, Malhotra-Kumar S, et al. “Methicillin-resistant Staphylococcus aureus”. In: *Nature reviews Disease primers* 4.1 (2018), pp. 1–23.

[22] de Vos AS, de Vlas SJ, Lindsay JA, Kretzschmar MEE, and Knight GM. “Understanding MRSA clonal competition within a UK hospital; the possible importance of density dependence”. In: *Epidemics* 37 (2021), p. 100511.

[23] Skov RL and Jensen KS. “Community-associated meticillin-resistant Staphylococcus aureus as a cause of hospital-acquired infections”. In: *Journal of Hospital Infection* 73.4 (2009), pp. 364–370.

[24] Cooper BS, Medley GF, Stone SP, Kibbler CC, Cookson BD, et al. “Methicillin-resistant Staphylococcus aureus in hospitals and the community: stealth dynamics and control catastrophes”. In: *Proceedings of the National Academy of Sciences* 101.27 (2004), pp. 10223–10228.

[25] D’Agata EMC, Webb GF, Horn M, Moellering RC, and Ruan S. “Modeling the invasion of community-acquired methicillin-resistant Staphylococcus aureus into hospitals”. In: *Clinical Infectious Diseases* 48.3 (2009), pp. 274–284.

[26] Wang L and Ruan S. “Modeling nosocomial infections of methicillin-resistant Staphylococcus aureus with environment contamination”. In: *Scientific reports* 7.1 (2017), pp. 1–12.

[27] Cornell SJ, Suprunenko YF, Finkelshtein D, Somervuo P, and Ovaskainen O. “A unified framework for analysis of individual-based models in ecology and beyond”. In: *Nature communications* 10.1 (2019), pp. 1–14.

[28] DeAngelis DL and Grimm V. “Individual-based models in ecology after four decades”. In: *F1000prime reports* 6 (2014).

[29] Hontelez JAC, Lurie MN, Bärnighausen T, Bakker R, Baltussen R, Tanser F, et al. “Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study”. In: *PLoS medicine* 10.10 (2013), e1001534.

[30] de Vlas SJ and Coffeng LE. “Achieving herd immunity against COVID-19 at the country level by the exit strategy of a phased lift of control”. In: *Scientific reports* 11.1 (2021), pp. 1–7.

[31] Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, et al. “Covasim: an agent-based model of COVID-19 dynamics and interventions”. In: *PLOS Computational Biology* 17.7 (2021), e1009149.

[32] Di Ruscio F, Guzzetta G, Bjørnholt JV, Leegaard TM, Moen AEF, et al. “Quantifying the transmission dynamics of MRSA in the community and healthcare settings in a low-prevalence country”. In: *Proceedings of the National Academy of Sciences* 116.29 (2019), pp. 14599–14605.

[33] Kardaś-Słoma L, Boëlle PY, Opatowski L, Brun-Buisson C, Guillemot D, and Temime L. "Impact of antibiotic exposure patterns on selection of community-associated methicillin-resistant Staphylococcus aureus in hospital settings". In: *Antimicrobial agents and chemotherapy* 55.10 (2011), pp. 4888–4895.

[34] Macal CM, North MJ, Collier N, Dukic VM, Wegener DT, et al. "Modeling the transmission of community-associated methicillin-resistant Staphylococcus aureus: a dynamic agent-based simulation". In: *Journal of translational medicine* 12.1 (2014), pp. 1–12.

[35] Vynnycky E and White R. *An introduction to infectious disease modelling*. Oxford university press, 2010.

[36] Eriksen NHR, Espersen F, Rosdahl VT, and Jensen K. "Carriage of Staphylococcus aureus among 104 healthy persons during a 19-month period". In: *Epidemiology & Infection* 115.1 (1995), pp. 51–60.

[37] Hu L, Umeda A, Kondo S, and Amako K. "Typing of Staphylococcus aureus colonising human nasal carriers by pulsed-field gel electrophoresis". In: *Journal of medical microbiology* 42.2 (1995), pp. 127–132.

[38] Wertheim HFL, Melles DC, MC Vos, van Leeuwen W, van Belkum A, et al. "The role of nasal carriage in Staphylococcus aureus infections". In: *The Lancet infectious diseases* 5.12 (2005), pp. 751–762.

[39] Bakker R, Busker T, White RG, and Choenni S. "Rapid weighted random selection in agent-based models of infectious disease dynamics using augmented B-trees". In: *SIMUL 2013 : The Fifth International Conference on Advances in System Simulation* (2013).

[40] Centraal Bureau Statistiek (CBS). *Huishoudens; grootte, samenstelling, positie in het huishouden, 1 januari*. URL: `https://www.cbs.nl/nl-nl/cijfers/detail/82905NED?dl=453BC`. (accessed on 19.06.2022).

[41] Cuervo G, Gasch O, Shaw E, Camoez M, Domınguez MA, Padilla B, et al. "Clinical characteristics, treatment and outcomes of MRSA bacteraemia in the elderly". In: *Journal of Infection* 72.3 (2016), pp. 309–316.

[42] MacKenzie FM and Gould IM. "The post-antibiotic effect". In: *Journal of Antimicrobial Chemotherapy* 32.4 (1993), pp. 519–537.

[43] Craig WA and Vogelman B. "The postantibiotic effect". In: *Annals of Internal Medicine* 106.6 (1987), pp. 900–902.

# A COMPLEX SPREAD OF MRSA CCS USING THE R MODEL

Initially, similar simulations as those presented in Chapter 5 were executed using the R model. However, since then a mistake has been found in the code and the model has been adjusted. Unfortunately, because the R model is very computationally expensive the choice was made to not redo the simulations. However, because a lot of effort (and time) has been put into these results they are presented in this appendix. In section A.1 the model set-up will be shortly explained and the obtained results are described in section A.2. Note that the results of the R model are not as elaborate as those of the Java model applied to a complex model set-up.

The mistake in the R model was found in the order of different steps of the algorithm. At the end of each time step, new individuals were infected, however, at the beginning of the new time step their time to clearance was already decreased by one. Therefore the infection duration was actually a day shorter than anticipated and an individual likely infected less people than when the order had been correct. Furthermore the model did calculations of the infection force with the wrong number of susceptible and infected individuals. Therefore the number of calculated new infections was incorrect at each time step. This probably also impacted the interactions between the different CCs. It is essential that the reader keeps in mind that the results presented in this section are flawed, because of this mistake. Although the mistake impacted the results of the model, the basic structure still performed comparable and therefore the results in this sections may be useful in supporting further research.

## A.1 MODEL SET-UP

In this section we will discuss the parameters and distributions chosen to create the obtained results that will be discussed later. It is important to note that the parameters chosen in this part of the study are not necessarily realistic. Because the goal is to find out how various factors influence the spread of different MRSA CCs, a fictional population was created. To limit the computational cost, a relatively small population size has been chosen. To understand the influence of certain factors, such as spread in the hospital, some parameters are chosen different from reality. These choices will be mentioned and explained in the paragraphs below.

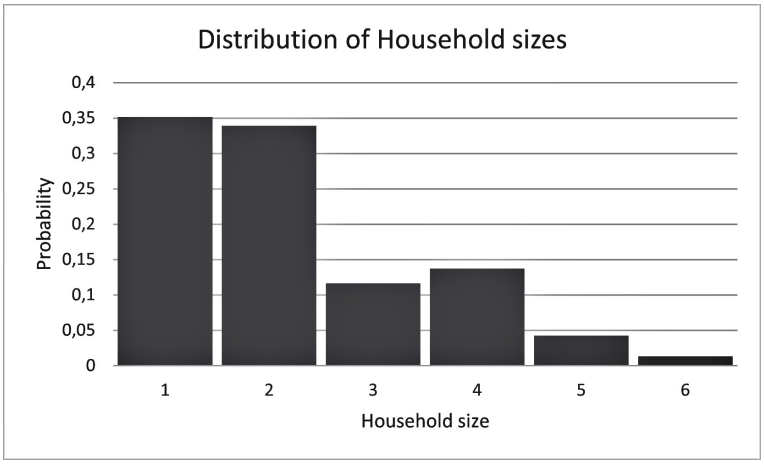


Figure A.1: Distribution for the household sizes in the complex model set-up for the R model. The household sizes are based on the distribution of the Netherlands[40].

For each run a population of 10.000 individuals is modelled. The R model is computationally expensive and larger populations would make the run time unmanageable. Inside this population six nursing homes are included, each with 100 inhabitants. Furthermore, two hospitals have been incorporated, where halve of the population is assigned to one hospital and halve to the other.

For the distribution of household sizes the distribution shown in Figure A.1 is used, which is very similar to the distribution of the Netherlands [40]. Household sizes are drawn from this distribution at the beginning of each run. The use of this distribution results in around 4000 households.

Individuals are divided in three different groups, each with a specific probability of hospitalisation: 1/2 of the population never visits the hospital, 1/3 is hospitalised with a probability of 0.002 each day and 1/6 of the population has a probability of 0.008. Nursing home residents have an additional probability of hospitalisation each day of 0.004. The resulting distributions for nursing home inhabitants and other individuals in the population are shown in Figure A.2.

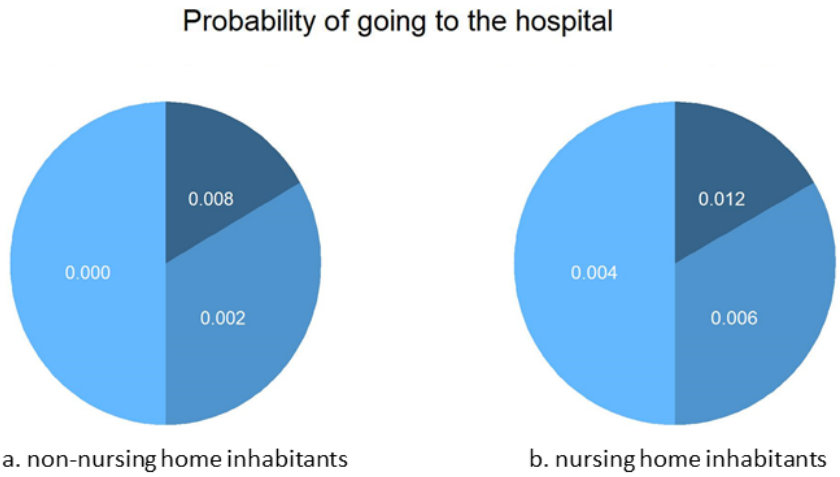


Figure A.2: Distribution of different hospitalisation rate groups. (a) Non-nursing home inhabitants; (b) Nursing home inhabitants.

Whenever an individual enters hospital they are assigned a length of stay which is at least one day with an additional time span drawn from a log-normal distribution with a log mean of 1.0 and a log standard deviation of 1.7. This gives the distribution shown in Figure A.3. With the parameters described above, the number of patients in each hospital varies roughly between 100 and 150. This means that more than 1% of the population is in hospital at each point in time. In realistic situations this number of hospitalised individuals is much lower. However, because we want to include and study the spread of MRSA within hospitals, we need a relatively large hospital population with respect to the overall population size.

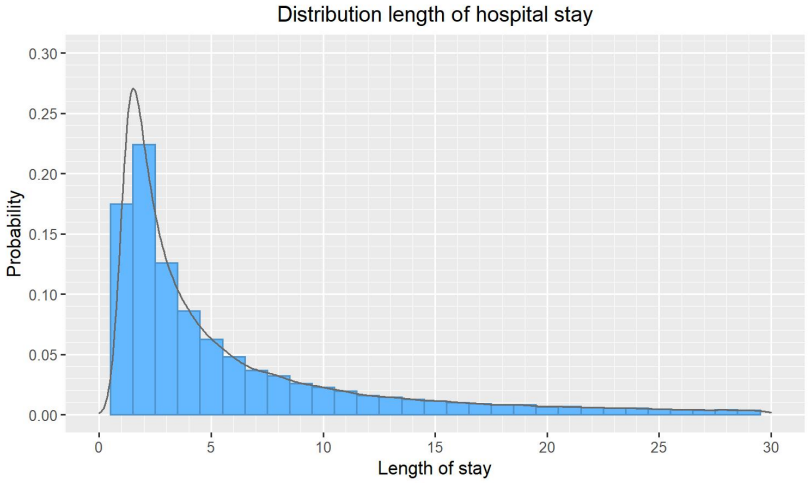


Figure A.3: Distribution length of stay in the hospital

At initialisation all individuals are assigned a clearance type: 50% has clearance type I, 30% clearance type II, 20% clearance type, which can also be seen Figure A.4a. When an individual gets infected an indication for the time to clearance is drawn from a log-normal distribution with a logarithmic mean of 0 and logarithmic standard deviation of 0.5. This number is then multiplied by a clearance type factor which is different for each clearance type. For clearance type I the multiplication factor is 1, for type II it is 8 and for type III it is 40. This results in different distributions for the assigned time to clearance for each of the clearance types, which are shown in Figure A.4.

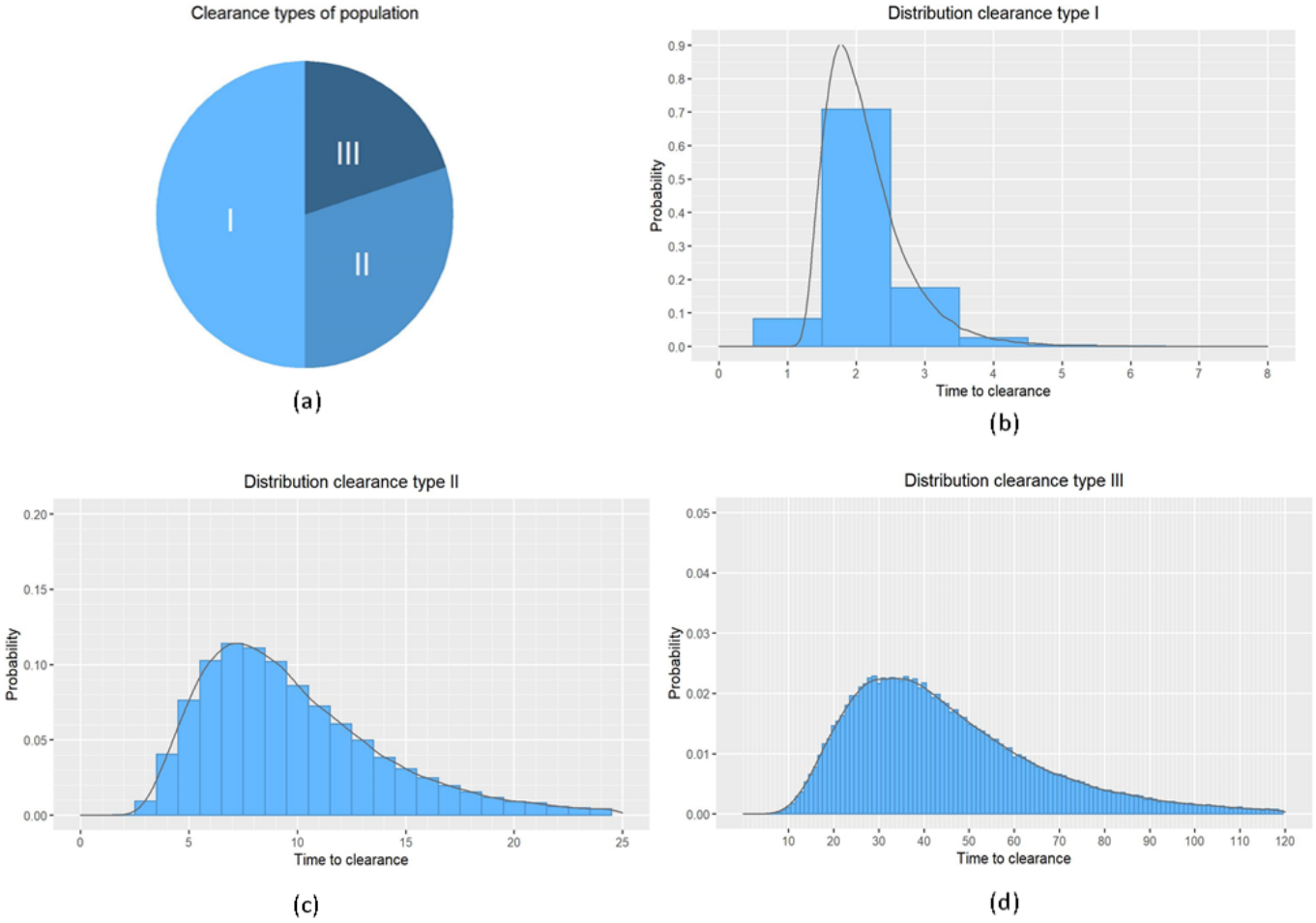


Figure A.4: Infection duration for different clearance types in the population. (a) Distribution of clearance types in the population; (b) Distribution infection duration for clearance type I; (c) Distribution infection duration for clearance type II; (d) Distribution infection duration for clearance type III.

In the model 10 CCs are included. One of these (CC0) is an MSSA CC; the other ones are CCs of MRSA (CC1-CC9). For CC0 it is assumed that the probability of infection per infected contact is 0.017. For the other CCs the probability is set to 0.0146. Each CC is assigned a certain susceptibility to antibiotics. The probability of clearance when antibiotics is taken for each of the CCs can be found in Table A.1. CC0 is the MSSA CC, so it is least resistant to antibiotics. CC9 has the smallest chance of clearance and therefore is most resistant to antibiotics.

Table A.1: Probability of clearance when antibiotics is given for each of the CCs.

| CC | clearance probability | | |
|---|---|---|---|
| CC0 | 0.95 | CC5 | 0.12 |
| CC1 | 0.16 | CC6 | 0.11 |
| CC2 | 0.15 | CC7 | 0.10 |
| CC3 | 0.14 | CC8 | 0.09 |
| CC4 | 0.13 | CC9 | 0.08 |

A mean number of possible infectious contacts is assumed in different settings. Inside the general population a mean of 5.0 contacts is assumed and within households and nursing homes 10.0. In the hospital it is assumed that patients have 15.0 potentially infectious contacts. At initialisation each individual is linked to two contact factors, for their household contacts and their general population contacts respectively. These contact factors are drawn from a gamma distribution with rate and shape both 2.0. This contact factor indicates how many contacts they have with respect to the mean. In the hospital patients do not have contact factors, as it is assumed that each patient has the same number of contacts. Using these distributions for the contact factors and the mean number of contacts, the distributions in Figure A.5 is created for the number of contacts in the different settings.

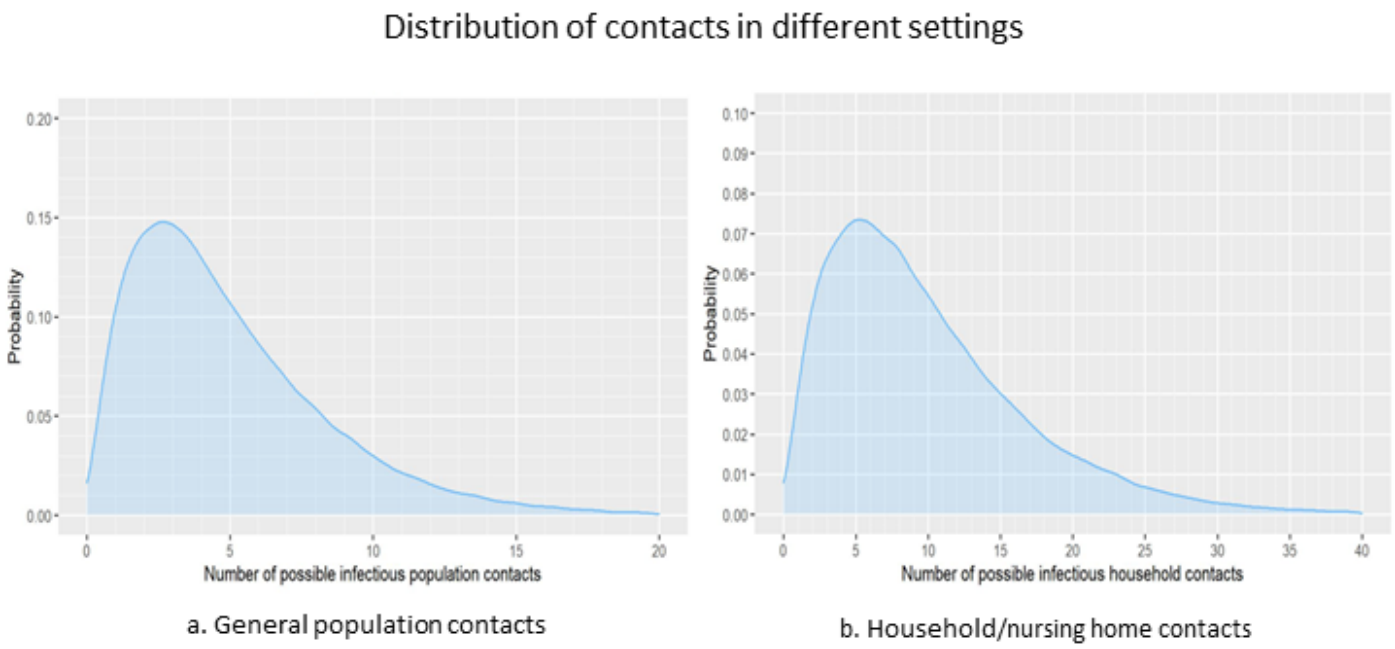


Figure A.5: Distributions for the number of contacts in different groups. (a) contacts within general population; (b) contacts within household/nursing home.

Lastly there is a possibility for isolation of patients that enter hospital and are carriers of MRSA. For the runs where the search-and-destroy policy was included, the probability of isolation of somebody entering the hospital with MRSA was set to 0.74. The effectiveness of the isolation can also be adjusted. For the runs done in this section the fraction of remaining contacts for somebody in isolation was set to 0.01. This is equivalent to saying that isolation prevents 99% of all possibly infectious contacts. In Table A.2 the parameters and distributions of this model set-up are summarised.

Table A.2: Definition of parameters and their values for the runs performed in this appendix.

| Parameter | Value |
|---|---|
| Population size | 10.000 |
| Number of hospitals | 2 |
| Number of nursing homes | 6 |
| Size of each nursing home | 100 |
| Household sizes | See Figure A.1 |
| Hospital visits | 3/6 of individuals never go to the hospital<br>2/6 of individuals go to the hospital with a daily rate of 0.002<br>1/6 of individuals go to the hospital with a daily rate of 0.008<br>Nursing home residents have an additional rate of 0.004 |
| Length of hospital stay | Log-normal($\mu = 1$ , $\sigma^2 = 1.7$)+1 |
| Number of CCs | 10 |
| Infection probability per infected contact | CC1: 0.017,<br>CC2-CC10: 0.0146 |
| Probability of clearance with antibiotics | CC0: 0.95, CC3: 0.14, CC6: 0.11, CC9: 0.08.<br>CC1: 0.16, CC4: 0.13, CC7: 0.10,<br>CC2: 0.15, CC5: 0.12, CC8: 0.09, |
| Clearance type factors | 0.5 of individuals has clearance multiplication factor of 1 day.<br>0.3 of individuals has clearance multiplication factor of 8 days.<br>0.2 of individuals has clearance multiplication factor of 40 days. |
| Distribution time to clearance | Log-normal($\mu = 0$ , $\sigma^2 = 0.5$) |
| Distribution number of contacts | General population: Gamma(mean=5.0,shape=2.0)<br>Household: Gamma(mean=10.0,shape=2.0)<br>Nursing homes: Gamma(mean=10.0,shape=2.0)<br>Hospital: Gamma(mean=15.0,shape=2.0) |
| Fraction household contacts remaining in hospital | 0.5 |
| Initial infections | CC0: 1000,<br>CC1-CC9: 0 |
| Initial infections in hospital | CC0: 50,<br>CC1-CC9: 0 |
| Number of external infections each year | CC0: 0,<br>CC1-CC9: 6 |
| Probability of isolation if infected | 0.74 |
| Isolation effectiveness | 0.01 |

## A.2 RESULTS

In this section the results of the R model will be discussed. The parameters in these runs are as described in Section A.1. The modelled time span is 30 years and time steps of 1 day are used. First, a situation without interventions will be studied. Second, a search-and-destroy policy at hospital entrance was incorporated in the model. For both situations 400 runs were performed. The results described in this section will mainly focus on the prevalence of the different MRSA CCs in the hospital. The hospital data will be discussed for a small number of individual runs. However, also the general behaviour that can be concluded from the 400 runs was studied.

### A.2.1 NO INTERVENTIONS

The first scenario that will be considered will not include any interventions taken to prevent the spread of MRSA CCs in the population. In Figure A.6 the prevalence of MRSA within the hospitals is shown for 9 different runs that were performed. The figures display the prevalences (in %) over a period of 30 years. The prevalence of each CC is indicated by a different coloured line. At the start of the simulation no MRSA CCs are present in the population. The introduction of an MRSA CC into the population is caused by external infections. In the runs presented in Figure A.6 it can be seen, that all CCs make an introduction in the population, however they are not always successful. In runs 1, 4, 5, and 9 the most successful CC appears to be CC9. For run 3 and 8 the dominant CCs seem to be CC8 and CC7 respectively. For the other runs (2, 6 and 7) it is not immediately apparent from the graphs which CC is most successful.

The MRSA CCs that appear most successful in Figure A.6 are CC7, CC8 and CC9. It is interesting to note that these CCs have been assigned the highest antibiotic resistance. This possibly implies that a better resistance to antibiotics has a significant impact on the possibility of being the most successful MRSA CC. However, it can also observed that it is not always the case that the CC with highest resistance is the dominant or most successful CC.

Additionally, it is interesting to note that for the runs where no CC was the clear winner (runs 2, 6 and 7) the average prevalence seems relatively low. In contrast, for the runs where one CC is clearly dominating, for example run 4, the average prevalence of MRSA appears very high. This indicates a possible relation between the mean MRSA prevalence and the success ratio of the CC.

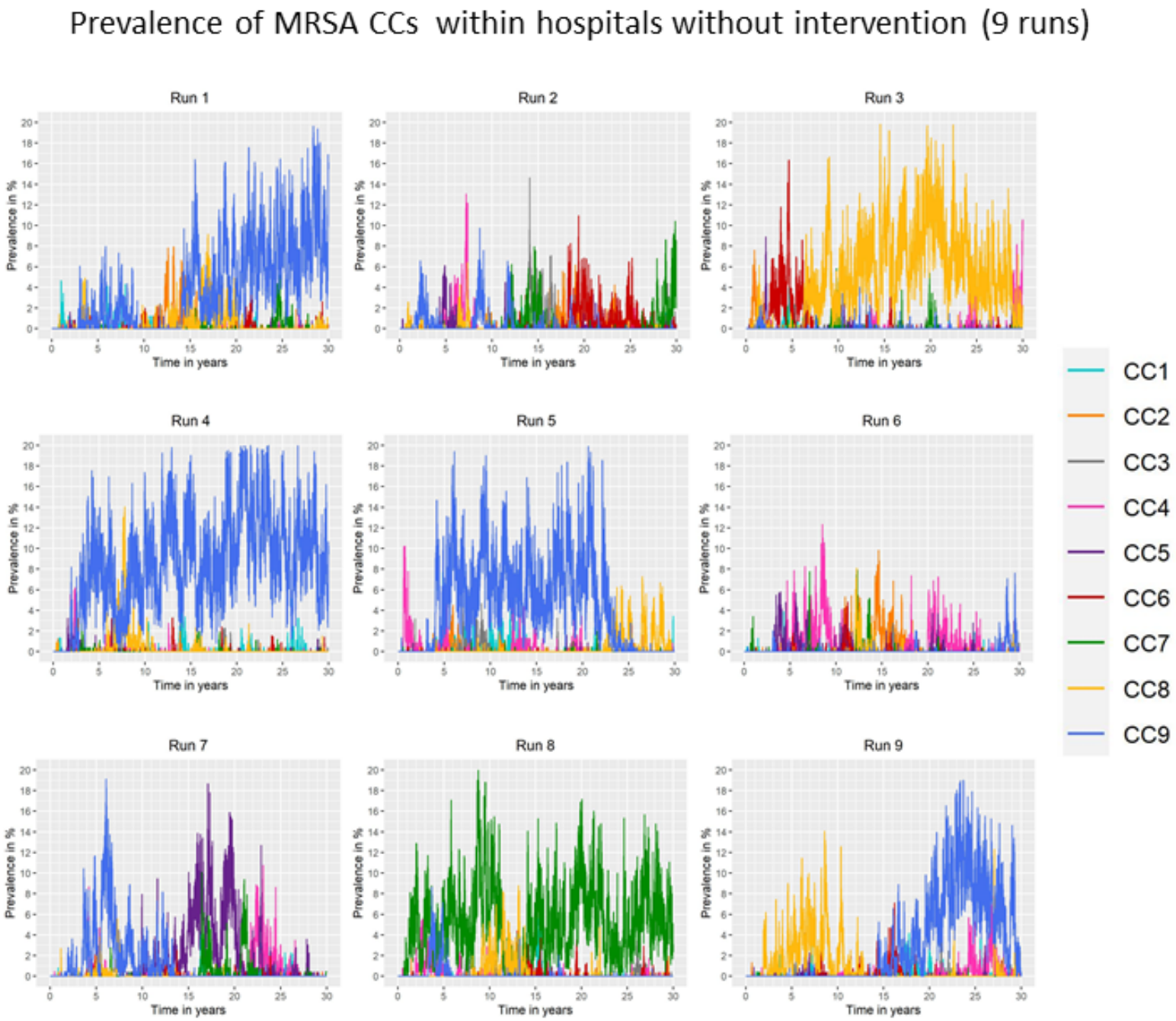


Figure A.6: Prevalence of MRSA CCs in the general population for 9 randomly selected runs, when no interventions are in place. The prevalence is presented over a period of 30 years.

In Figure A.7 the mean prevalence of the MRSA CCs within the hospitals is shown over the period from 20 to 30 years for the same 9 runs. This specific period was considered to make sure that the number of people in hospital was stabilised and to give the CCs some time to spread inside the population and become dominant. The diagrams show the contribution of each of the CCs to the MRSA prevalence. The percentages displayed in the diagrams refer to the mean of the total MRSA prevalence in the hospitals over the studied period.

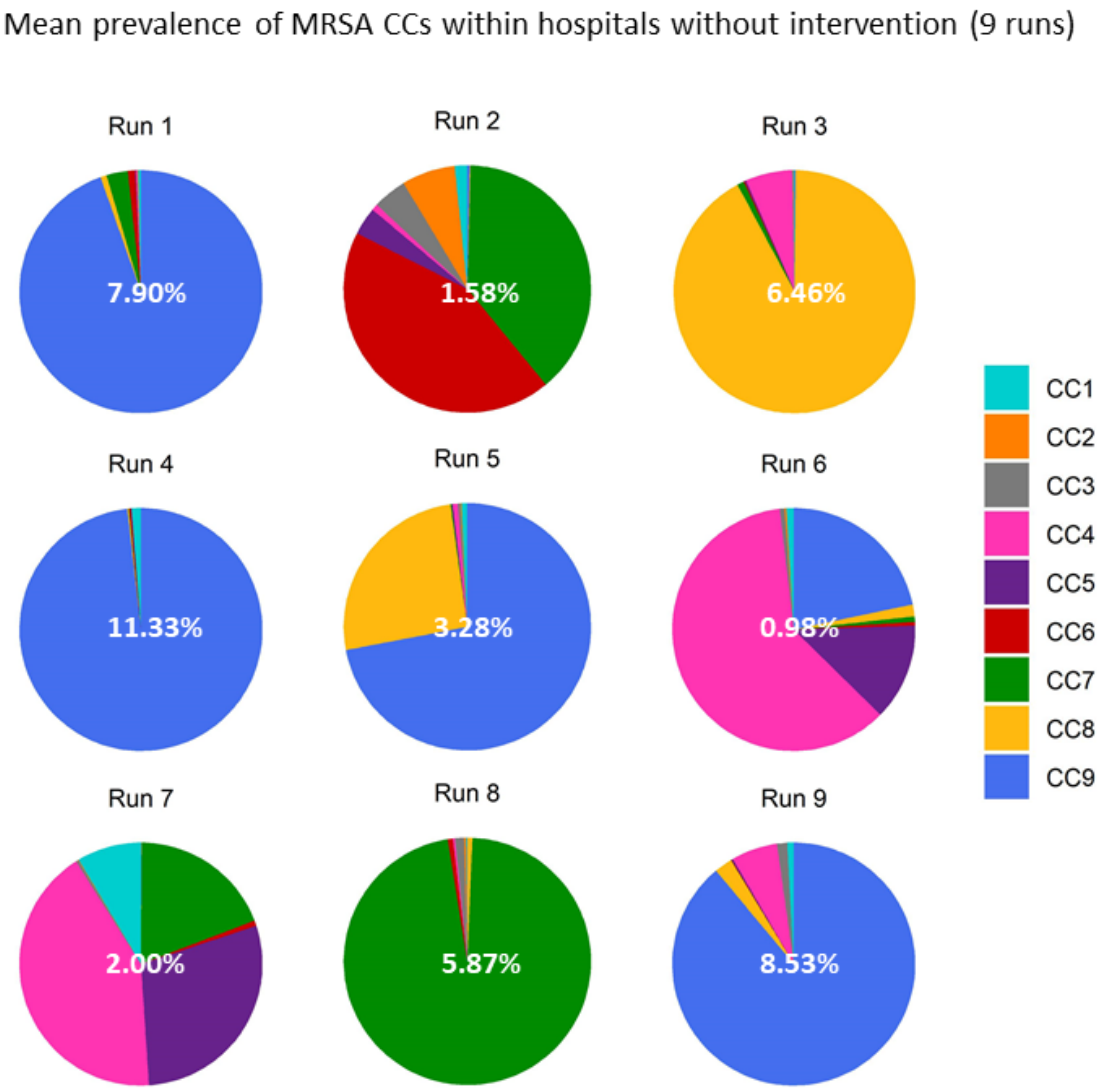


Figure A.7: Mean prevalence and distribution of MRSA CCs within hospitals over the period 20-30 years for 9 different runs without interventions.

As was already observed in A.6, for runs 1, 4, 5, and 9 from Figure A.7 it can be concluded that the most contributing CC to the total prevalence is CC9. For run 3 and 8 the dominant CCs seem to be CC7 and CC8, respectively. For run 2 and 7 again no clear winner can be concluded from the diagrams. From Figure A.6 no obvious CC could be distinguished as most successful in run 6. However, from Figure A.7 we can see that more than halve of the total prevalence of MRSA can be attributed to CC4.

The diagrams from Figure A.7 further attribute to the idea that a higher antibiotic resistance increases the possibility of being the dominant CC. In run 1, 3, 4, 5, 8 and 9 the winners were the three CCs with highest resistance. Dissimilar, in run 6 the most successful CC was CC4, which has a relatively low resistance. It is interesting to note that in that specific run the mean total prevalence was a lot lower than in the other simulations with a clear dominant CC. This possibly suggests that when a CC with relatively low resistance is the most successful CC, the total prevalence is lower than when a CC with high resistance is most successful.

To get a more general understanding of the success of different CCs, the data of 400 runs was combined. All 400 runs use the same model set-up described in Section A.1. In Table A.3 some properties of the 400 runs are shown. For each of the CCs the first entry in the table is the mean prevalence averaged over the 400 runs of each of the CCs. One can immediately note that from CC1 to CC9 the mean prevalence moves upwards. This further supports the believe that higher resistance positively attributes to the prevalence of a CC in hospitals.

This can also be seen when looking at the number of times the different CCs are the winner of a given run. A CC is the winner of a run if it has contributed the most number of MRSA infections to the prevalence. The value presented in Table A.3 is the number of times that specific CC turned out as the winner of one of the 400 runs. The table shows that in general, the higher the resistance of the CC, the more likely it is to be the winner of a run. CC9 ended up as the winner of the runs more than a thirth of the time. The three most resistant MRSA CCs were established as the winner in more than 75% of the 400 runs.

The last entry shown in Table A.3 is the number of times a CC turned out as an absolute winner of a run. A CC is referred to as the absolute winner if it has attributed to the prevalence more than %75 of MRSA infections. Similar as to being the winner, it can be seen that a higher resistance also increases the chance of being the absolute winner. What furthermore is noteworthy, is that the higher the resistance, the larger part of the wins were absolute wins. This suggests that a higher resistance not only increases the change of being the most successful, but also increases the proportion with which it is most successful.

Table A.3: Properties of CCs over 400 runs without interventions. The table includes the averaged mean prevalence, the number of times a CC is the winner of a run, and the number of times that CC is an absolute winner.

| | CC1 | CC2 | CC3 | CC4 | CC5 | CC6 | CC7 | CC8 | CC9 |
|---|---|---|---|---|---|---|---|---|---|
| Mean average prevalence | 0.06% | 0.08% | 0.09% | 0.12% | 0.19% | 0.37% | 0.62% | 1.26% | 2.41% |
| Winner | 6 | 9 | 8 | 13 | 17 | 39 | 50 | 95 | 163 |
| Absolute winner (>75%) | 1 | 1 | 0 | 2 | 4 | 14 | 29 | 60 | 116 |

### A.2.2 SEARCH-AND-DESTROY POLICY

The second model set-up that will be considered includes a search-and-destroy policy. The same parameters are used as for the runs without interventions. However, additionally, at hospital entrance individuals that are carriers of MRSA have a probability of 0.74 to be placed in isolation. The same scale was adopted for the axes as in the runs without interventions, in order to make comparison easier. In Figure A.8 the prevalence of MRSA within hospitals is shown for 9 different runs which include the search-and-destroy policy over a period of 30 years. One can immediately see that the prevalence is a lot lower, when incorporating the search-and-destroy policy. In contrast to the results without interventions in Figure A.6, in none of the runs a dominating CC can be distinguished. Another difference that is important to note, is that the time span over which a certain CC causes a peak in the prevalence of MRSA is a lot shorter when a search-and-destroy policy is implemented. In the presented runs there are no CCs that appear dominant over the entire time span.

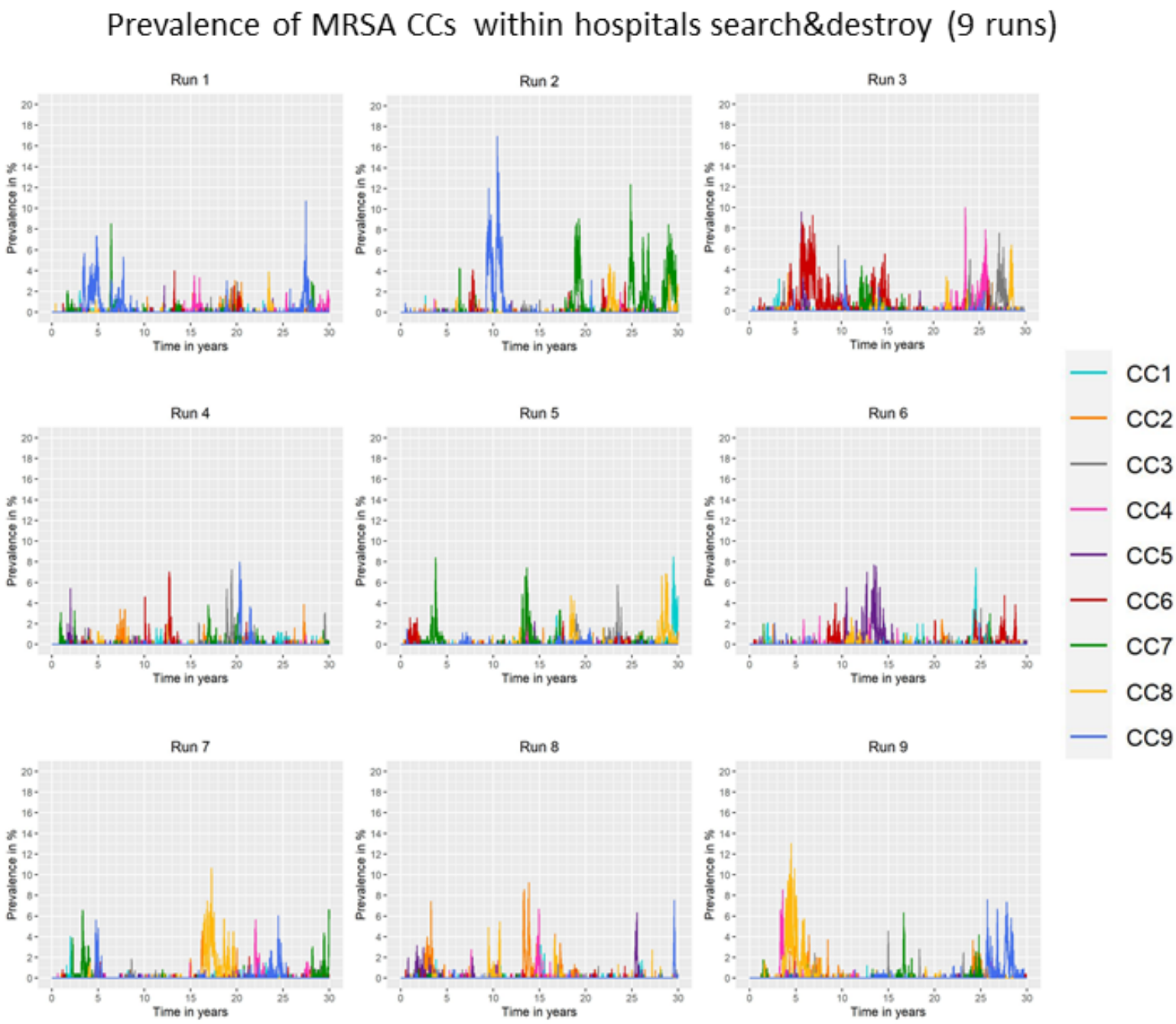


Figure A.8: Prevalence in the general population of MRSA CCs for 9 randomly selected runs, when a search-and-destroy policy is in place. The prevalence is presented over a period of 30 years.

In Figure A.9 the mean prevalence of the MRSA CCs in hospitals is shown over the period from 20 to 30 years for the same 9 runs. The diagrams show the contribution of each of the CCs to the MRSA prevalence, when a search-and-destroy policy is employed. The percentages displayed in the diagrams refer to the mean total prevalence of MRSA in the hospitals over the studied period. It can be seen in Figure A.9 that the mean MRSA prevalence is much lower than when no interventions were made.

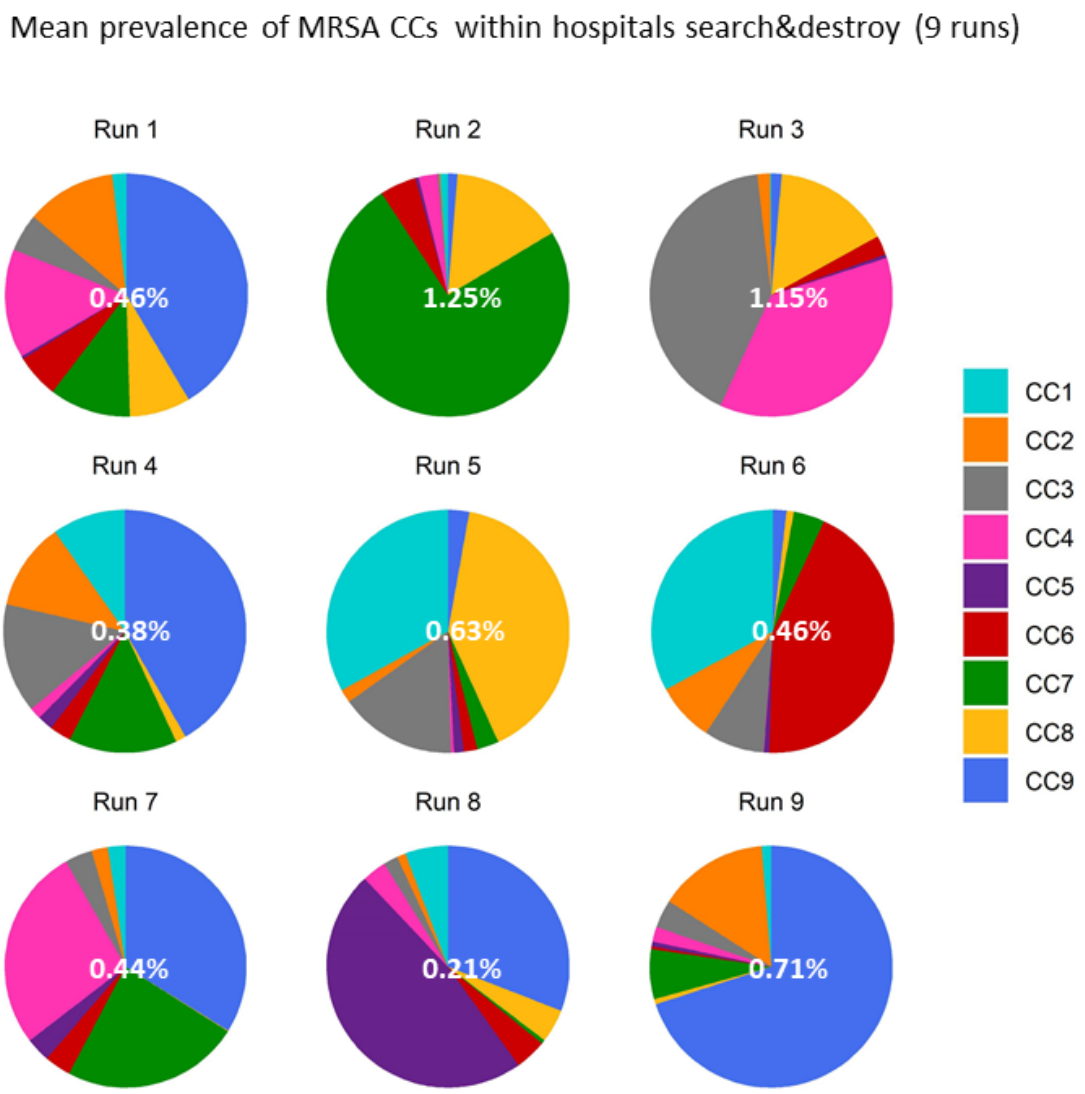


Figure A.9: Mean prevalence and distribution of MRSA CCs within hospitals over the period 20-30 years for 9 different runs where a search-and-destroy policy was included.

Furthermore it can be noted that only in two runs (2 and 9) there are CCs that attribute to more than halve of the prevalence. The successful CCs in these cases are CC7 an CC9 which have a relatively high resistance. This might imply that, also when a search-and-destroy policy is applied, a higher resistance increases the chances of a CC to be the most successful. Furthermore it can also be seen that in these runs the mean MRSA prevalence is higher than in most other runs, except for run 3. This suggests a positive relation between the success-percentage of the winner and the average prevalence.

In the other diagrams shown in Figure A.9 we see that in most cases a large variety of CCs contributes to the prevalence of MRSA in hospitals. Not only CCs with a high resistance, such as CC9 in run 4, but also CCs with a low resistance, such as CC1 in run 5, can contribute a large part of the infections to the mean prevalence. These contributions coincide with the small peaks that could be seen in Figure A.8. Multiple CCs have smaller peaks, which leads to multiple CCs contributing to the MRSA prevalence.

To even better showcase the difference between no interventions and the implementation of a search-and-destroy policy, Table A.4 was created that, among other things, shows the mean prevalence averaged over the 400 runs for each of the CCs. One can again note that from CC1 to CC9 the averaged mean prevalence moves upwards, as it did without interventions. This supports the idea that higher resistance positively attributes to the prevalence of a CC within hospitals. As we already saw in A.9 the mean MRSA prevalence is a lot lower when a search-and-destroy policy is included. The biggest decrease in mean prevalence is observed for the CCs with the highest prevalence. As a result, the difference in mean prevalence for the lower and higher resistance CCs is reduced.

Interesting to see in Table A.4, is that the number of wins for CCs with a relatively low resistance is increased, when compared to the situation without interventions. Although a higher resistance still positively impacts the number of wins, the implementation of a search-and-destroy policy creates the opportunity for also lower resistant CCs to attribute to the prevalence in hospitals. However, if we consider the number of times a CC ended up as an absolute winner we see that this does not often happen for CCs with a low resistance. This suggests that a higher resistance not only increases the change of being most successful, but also increases the proportion with which it is most successful.

Table A.4: Properties of CCs over 400 runs where a search-and-destroy policy included in the simulation. The table includes the averaged mean prevalence, the number of times a CC is the winner of a run, and the number of times that CC is an absolute winner.

| | CC1 | CC2 | CC3 | CC4 | CC5 | CC6 | CC7 | CC8 | CC9 |
|---|---|---|---|---|---|---|---|---|---|
| Mean average prevalence | 0.03% | 0.04% | 0.04% | 0.05% | 0.05% | 0.06% | 0.08% | 0.14% | 0.18% |
| Winner | 26 | 27 | 22 | 37 | 32 | 39 | 66 | 69 | 82 |
| Absolute winner (>75%) | 0 | 2 | 2 | 4 | 4 | 5 | 3 | 18 | 23 |

Figure A.10 previews two diagrams that show the averaged mean prevalence in the hospitals and the distributions of MRSA CCs. The left diagram shows the prevalence in hospitals when no interventions are taken and the right diagram the prevalence if a search-and-destroy policy is included in the runs. The white percentage in the diagram is the averaged mean total MRSA prevalence. The first thing that can be noted is that the averaged mean total MRSA prevalence is a lot higher if no search-and-destroy policy is applied. As this intervention is meant to limit the spread of MRSA, this is a desired result. Another difference that can be spotted is the distribution of the mean average prevalence over the MRSA CCs. When no interventions are performed, the CCs with the highest resistance contribute most to the prevalence. When a search-and-destroy policy is employed, a lot of the prevalence can also be attributed to CCs with a low resistance. Overall, more variation can be found between the CCs present in the hospital.

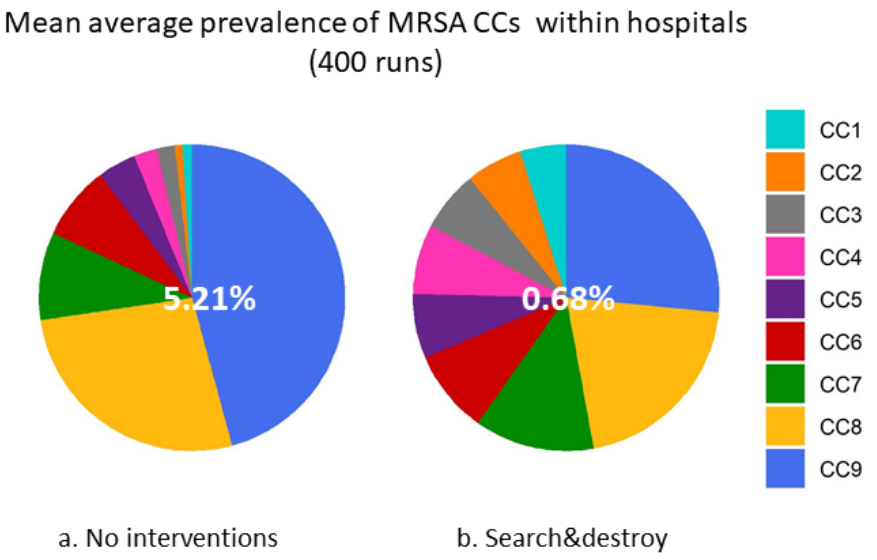


Figure A.10: Averaged mean prevalence and distribution of MRSA CCs in hospitals from the period 20-30 years over 400 runs: (a) No interventions; (b) Search-and-destroy policy.

# B SUMMARY PARAMETERS USED IN JAVA MODEL

Table B.1: Definition of parameters and their values for the runs performed in Chapter 5.

| Parameter | Value |
|---|---|
| Population size | 10.000 |
| Number of hospitals | 2 |
| Number of nursing homes | 4 |
| Size of each nursing home | 100 |
| Household sizes | See Figure 5.1 |
| Hospitalisation rate | 0.002<br>Nursing home residents have an additional rate of 0.004 |
| Length of hospital stay | NegBin(mean= 5.0 , $k = 0.8$)+1 |
| Number of clonal complexes | 6 |
| Infection probability per infected contact | CC0: 0.020,<br>CC1-CC5: 0.0185 |
| Probability of clearance with antibiotics | CC0: 0.95, CC3: 0.12,<br>CC1: 0.20, CC4: 0.08,<br>CC2: 0.16, CC5: 0.04 |
| Clearance type factors | 0.5 of individuals has clearance type I: average of 1 day<br>0.3 of individuals has clearance type II: average of 8 days<br>0.2 of individuals has clearance type III average of 40 days |
| Distribution time to clearance | Gamma(mean=1.0 , shape= 3.0)×clearance type |
| Distribution number of contacts | General population: Gamma(mean=5.0,shape=2.0)<br>Household: Gamma(mean=10.0,shape=2.0)<br>Nursing homes: Gamma(mean=10.0,shape=2.0)<br>Hospital: Gamma(mean=15.0,shape=2.0) |
| Fraction remaining household contacts in hospital | 0.5 |
| Initial infections | CC0: 2000,<br>CC1-CC5: 0 |
| External infection force per exposed individual per day | CC0: 0,<br>CC1-CC5: 0.000008 |
| Fraction of patients tested for MRSA per day | 0.25 |
| Isolation effectiveness | 0.01 |