A Stochastic Model for Meningococcal Disease

Hanna Autio

Master's thesis
2017:E48

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Numerical Analysis

Centrum Scientiarum Mathematicarum

Abstract

This paper presents a stochastic model for Meningococcal disease. Mathematical models for diseases are an important tool in population dynamics, and their results carry real-world implications as they guide policies and vaccination strategies. While the systems being modelled are generally stochastic in nature, deterministic models are often used, possibly to the detriment of accuracy. In this project, a stochastic model is developed for Meningococcal disease in the African Meningitis belt, a region plagued by recurring epidemics, a high incidence rate of disease and significant seasonal variations. The model uses the Feller-Kendall algorithm, and is able to accurately emulate some of the specificities present in the region. Simulation results suggest that the seasonality cannot be exclusively explained by behavioural variations, but does not dismiss its influence entirely. Furthermore, there are signs that the system in a non-epidemic state can be modelled using a deterministic framework.

Contents

1 Introduction

The mathematical description of populations and disease dynamics is a field which can be extremely complex. It is a subject of much research, relevant to policy makers and medical researchers. An accurate model that can predict future dynamics can be of immense value, both to prepare for and deal with epidemics, and to evaluate different vaccination policies. However, accuracy in itself is not enough. The model must also be precise enough, as well as capable of evaluating several possible scenarios within a reasonable time frame. This leads to a need for a balance between accuracy, which means several different interactions must be considered, and swiftness, which calls for approximations and numerical efficiency.

A common practise for population dynamics is to approximate a discrete and random system, according to the central limit theorem and the law of large numbers, by a continuous and deterministic system. It is obvious that in a given population, the number of individuals is a discrete number. Furthermore, it is intuitively clear that events such as contagion and infection are fundamentally random. However, models that incorporate this randomness into their structure are in general significantly slower to use, as a single evaluation will take longer and more evaluations are needed to increase precision and accuracy (in correspondence to the Monte Carlo framework). Consequently, an approximation is often made to a deterministic system governed by differential equations. In order for this to work, populations must vary continuously and the discreteness of individuals is lost. By the law of large numbers and the central limit theorem, this method is sound if the population is large enough and a continuous population will, while not entirely accurate, be at least as specific as a realisation of a discrete random system. With this new framework standard numerical procedures can be used, and as the model is deterministic only one evaluation is needed. However, the law of large numbers only holds as long as there is indeed a large number of individuals; whenever a population (or subpopulation) faces extinction the rule no longer holds. These and other factors must all be considered when determining what framework and what model is the most appropriate for the system at hand, with respect to accuracy, speed and likely system states.

The complexity in modelling population dynamics arises from several factors. Firstly, the modelling of a population in itself, disregarding external factors, can be made increasingly complex by considering each individual and the exact changes they are subject to. Secondly, external factors should be considered so as to make the modelling relevant for policy makers. What external factors to include and how precisely they should be modelled, as well as their precise effects on the population, introduces a whole new layer of complexity. For example, one external factor that may be interesting to consider is weather, and currently there is no reasonable way to reliably model the weather. It can even be argued that to accurately model a population, the whole system that is the earth itself must be considered. Consequently, some approximations and assumptions will be made, even when choosing the stochastic approach to modelling. As the stochastic approach is in general more complex, it can be argued that more of these assumptions and approximations will be made than in a corresponding deterministic model, to compensate for the increased complexity otherwise. This should also be considered when evaluating different options for modelling.

Most classical approaches to population dynamics focus on the system as it is described by the current population numbers, but there are limitations to this method and alternatives should be considered. One example is the frequently used SIR-framework, where a given population is

compartmentalised by disease status, and the ratios of the population in each compartment are used to describe the system state at the given time. However, in some cases this will not provide sufficient information. Consider for example the case where a model is used to estimate what capacity a maternity ward should have some years into the future. It is then not sufficient to consider only the population growth over time, but the raw number of births is the relevant factor. In this case, the population ratios do not provide the desired information and a model more focused on the relevant dynamics should be constructed.
The goals of modelling diseases can be many. There is obviously an interest in understanding and explaining the dynamics of the disease itself, to increase knowledge about the disease and potentially understand the underlying biological mechanisms. It can also be used as a validation tool to confirm or dismiss theories about a disease, if the model and simulation method is proven sound beforehand. One of the more relevant and obvious goals of these models is to help policy makers determine what the consequences of different actions can be in different circumstances. This will help showing whether a vaccination policy is efficient, or what the best cause of action during an epidemic may be. Furthermore, as there are frequently issues in gathering real-world data, a reliable model can provide more and higher-quality data that can later be used to identify dynamics that could, for example, be used as a predictor for future epidemics.
For the reasons mentioned above, it is important that a given model is flexible. The framework must not only support different scenarios in the disease dynamics itself, but it should also support variations in external factors and official policies, corresponding to different decisions but also to variations in the resources controlled by the policy makers. This is especially important when modelling populations in areas of the world with poor infrastructure, as this can affect the data available as well as the possible responses from medical personnel and governments.
A flexible and reliable model should also be able to adapt to changes within the communities themselves. For many diseases, including for example measles, there is a clear pattern of disease corresponding to population movements and population mixing [1]. These mechanisms can lead to an increased rate of transmission, as populations grow, or the introduction of new disease variations to groups with lacking immunity or poor preparations in other ways.
Where the infrastructure is poor, a plethora of new issues arise. There are often troubles with distributing drugs and vaccines, meaning that any policy will be less efficient. This also increases the need for a warning system with early detection, as the time it takes to react to a given scenario increases. Furthermore, there can be issues gathering information itself. If sick individuals can not reach a medical institution in time, or if the information from the medical institutions take too long to spread, the true situation will be unknown. These are factors that should be considered when using models as an aid for political decisions, but it should also be considered when constructing the model itself.
In this report, a stochastic and discrete model for the dynamics of Meningococcal disease in the African Meningitis belt is constructed and evaluated. The model is then used to examine dynamics of the disease, as well as to examine some possible explanations for the dynamics in question.

1.1 Meningococcal disease

Meningococcal disease is any disease caused by *Neisseria Meningitidis*. The infection of N. Meningitidis can take many forms, of which two common (and severe) are meningococcal

septicaemia and meningitis [6]. Meningitis is an infection of the thin lining surrounding the brain and the spinal cord, and is a serious condition that untreated leads to death in about 50% of cases. Early diagnosis and adequate treatment reduces the fatality rate to about 5-10%, with permanent disabilities in about 10-20% of cases. The incubation time varies from 2-10 days, but is on average 4 days [26]. Meningococcal septicaemia is the infection of the bloodstream, and is often even more severe than meningitis. The infection damages the walls of blood vessels and leads to bleeding into the skin and other organs. The treatment is similar to that for meningococcal meningitis. In rare cases, N. Meningitidis can cause arthritis and similar diseases.

Humans are the only known reservoir for N. Meningitidis. At a given time, about 2-50% of the population are likely to be carriers of the bacterium, and it is spread by salivary droplets. Carriage can last both for a very short period of time as well as for several months, during which the bacterium is present in the nasopharynx of the carrier [24]. In general, carriage does not lead to invasive disease, but it has been linked to a subsequent immune response to the bacterium. There are several mechanisms that can participate in the immune response, and research has shown that a significant part of the adult population has an immune response that is putatively protective towards the disease [9]. Another bacterium frequently theorized as potentially leading to immunity to the disease is *Neisseria Lactamica*, often present in the nasopharynx of young children [25]. While immunity protects towards invasive disease, it has not been shown that there is any reduction of carriage caused by the immune response. [27]

N. Meningitidis is genetically variable and different strains are not always easily classified. The most common method of classification is based on the composition of its polysaccharide capsule, by which it is divided into 13 different serogroups. While it is rather common, world-wide, for carriage strains to not have a capsule (around 50% of carriage cases), and thus be non-serogroupable, disease is primarily caused by encapsulated strains. Serogroups A, B, C, W-135 and X account for more than 90% of cases of invasive disease world wide. The production of a capsule can be switched on and off with high frequency, and also non-serogroupable carriage strains can be pathogenic [27]. In this report, serogroups A and W-135 are modelled due to available data and their relevance to the geographic area of interest.

There is a clear relationship between age and the distribution of carriage and disease caused by N. Meningitidis. Globally, the carriage rates are around 10% in an endemic state, but the rate is significantly lower ($< 3\%$) for young children, and significantly higher for teenagers and young adults, especially for military recruits and university students ($> 24\%$). Other factors that correlate with increased carriage are smoking and crowding, for example at bars or during the Hajj [23]. The significant increase in carriage in closed communities such as the military, the Hajj or to some extent new university students suggest that population mixing may be related to the transmission of disease.

The relationship between age and disease is also clear, but follows a different pattern. The group at most risk for disease are young children above 6 months of age, in connection with waning maternal immunity. There is also an increased risk in teenagers and young adults, similar to the increase in carriage rates for these groups.

Vaccines against the bacterium have historically been focused primarily on antigens found in the capsule, and are called polysaccharide vaccines. These vaccines can be, and have been, used to prevent invasive disease in individuals older than infants, but does not induce an immunological memory. As such, preventative vaccine policies can not efficiently be carried out using these types

of vaccines. Recently, polysaccharide conjugate vaccines have been introduced, which are capable of inducing immunological memory as well as immunizing infants. These conjugate vaccines currently exist for serogroups A and C. There have been indications that carriage of the targeted strain decreases following a conjugate vaccine intervention [18]. However, it is unclear whether this is caused by a decrease in carriage of the strain itself or is only a result of decreased production of the capsule [27].

1.2 The African Meningitis belt

The African Meningitis belt is a region in Sub-Saharan Africa with a comparatively (to the rest of the world) high rate of meningococcal disease as well as recurring and severe epidemics. The belt was first described by Lapeyssonnie in 1968 [16]. This region is shown in fig. 1 and has been characterized by semi-regular epidemics documented since the early 20th century. Its specific geographical area is defined be the WHO as the region in Africa between Senegal in the west and Ethiopia in the east [26], which coincides fairly well with the area described by Lapeyssonnie. There are several characteristics that distinguish the dynamics of meningococcal disease in this region from elsewhere in the world. Firstly, there are recurring epidemics with intervals of about 5-15 years. Secondly, the incidence rates of disease is higher than in other regions of the world. Thirdly, the disease is most commonly caused by a serogroup which does not usually cause epidemics in the rest of the world. Furthermore, all of these properties are affected by a strong seasonality. This leads to three distinguishable states for the disease in the region. There is the (hypo-)endemic state, with a low weekly incidence rate of 0-1 cases per 100 000, which occurs during the wet season. During the dry season, the system falls into a hyperendemic state with an increased weekly incidence rate (about 1-20 cases per 100 000), or occasionally an epidemic state. An epidemic state on a district level is defined as a weekly incidence rate of disease at or above 10 per 100 000, but community level incidence rates can be significantly higher during epidemics. This is because epidemics are often very localized, and some health centres can have incidence rates of up towards 100 per 100 000 [20].

The reason for the specific dynamics in this region is not clearly established, but often thought to be related to climatic factors. One such factor frequently considered is the total precipitation. The region coincides fairly well with the area between the 300mm and 1100mm isohyets [19], meaning that the area receives between 300mm and 1100mm precipitation yearly. Another common climate factor for the area is that it is affected by the Harmattan, a dry wind from the Sahara desert. As there are links between factors that irritate the airways and subsequent infection by N. Meningitidis [23], a link between the Harmattan and the hyperendemic meningitis season would seem plausible. It is an area of active research, and finding any climatic factors that could be driving the epidemics is outlined as a priority [10].

In order to establish more possible explanatory models for the seasonal variations, the factors that distinguish the region should be identified. However, this region of Africa is heterogeneous, and the infrastructure is somewhat lacking. As such, it can be difficult to estimate what factors are common for this region and what characterises it compared to other regions. It is however clear that there is a significant pattern of migration in the region [5], and there are some suggestions that an alternative explanation to the seasonal variations can be found in behavioural variations [1]. The distribution of genetic variations of N. Meningitidis is different in the Meningitis belt

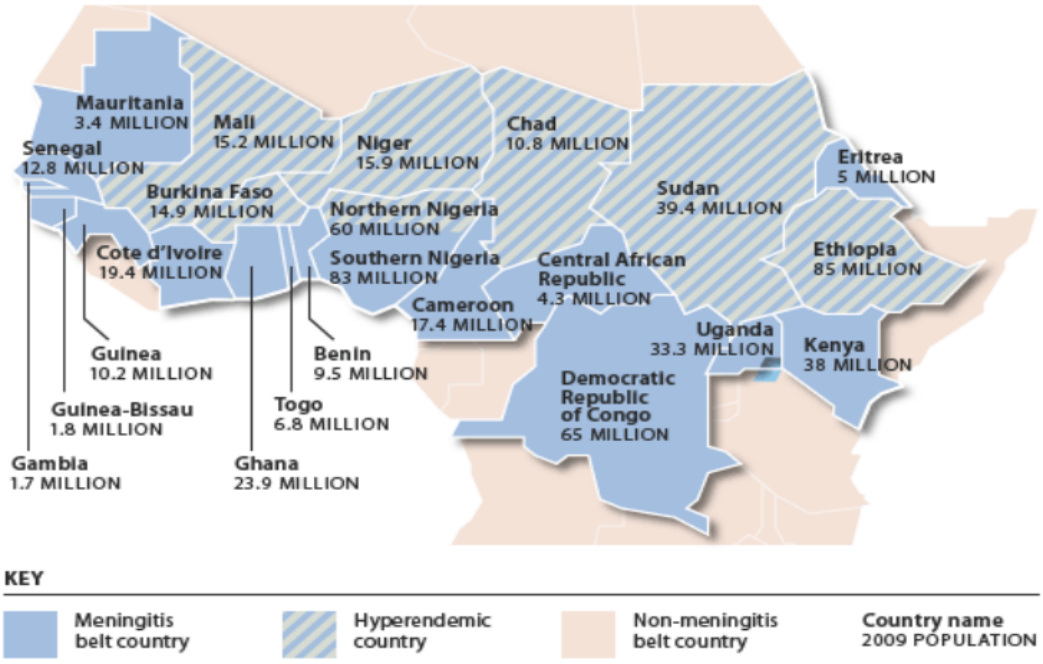


Figure 1: A map showing the region commonly considered the Meningitis belt. Hyperendemic countries show some of the characteristics common for the countries in the Meningitis belt, but not necessarily all. Source: `www.globalhealthhub.org`, [22].

compared to the rest of the world. There is a higher ratio of encapsulated virulent strains carried in the population, most notably of serogroup A. This serogroup has also been the cause of most epidemics in the region. Furthermore, there is significantly less genetic variation in the bacterial population than in other regions, and its temporal distribution is unstable. This leads to a wave pattern of clonal variations, where one strain is dominant at a given time [17]. It has been theorized that this lack of genetic variation can contribute to the epidemic patterns, as an epidemic strain can quickly become dominant and out-compete less virulent strains. The reasons for the lack of stability and lack of variation is unclear.

Since 2010, a WHO mass-vaccination project has been undertaken to combat epidemic meningitis in the region. A polysaccharide-conjugate vaccine effective against serogroup A has been introduced, and since its introduction, no new epidemics has been recorded from that specific strain. There are also indications that the vaccine affects carriage rates of this strain, which could lead to herd immunity. These types of effects has been seen in other vaccination projects against N. Meningitidis [18], but the effects on the specific dynamics in the Meningitis belt is yet unclear. One concern is whether the eradication of serogroup A could lead to an increase in other serogroups, specifically those with a higher virality.

A conference on the Meningitis belt outlined a few priorities for future research, where it was determined that attempts to construct mathematical models should include factors that may affect the spatial, seasonal and year-to-year variability in risk for meningococcal disease. The model constructed in this project focuses on the seasonal variability, and examines two different explanatory models for the variations.

This model is also the first, to our knowledge, stochastic model for Meningococcal meningitis in the African Meningitis belt. As much of the region has a low urbanization rate, a model suitable for smaller populations is appropriate. Furthermore, as the incidence rates during the endemic season can be very low there are clearly situations where the law of large numbers may not hold. This model could be used to verify previous modelling attempts, as well as evaluate current theories on the underlying causes of epidemiological dynamics.

1.3 Main goals of this project

The primary goal of this project has been to create and implement a stochastic model for the population and disease dynamics in the African Meningitis belt. The model should be evaluated, and its results should be compared to real-world data to examine the accuracy and specificity of the model.
If the model and its implementation seems sound, some further questions are to be considered. The primary consideration of this thesis is to attempt to determine whether the seasonal difference between the endemic and the hyperendemic states of the disease can be explained by behavioural changes or whether the explanation must be found somewhere else, for example in biochemical factors of the bacterium itself.

2 Methods

Stochastic models for population dynamics aim at determining the probability for the system to be in a given state at a given time. Typically, there will not be a closed-form analytical solution to this problem, and the Monte Carlo framework is often used to produce viable estimates. This is also the basic method used in this project. This means that a model is developed for the general, stochastic behaviour of the system, and a multitude of realisations are simulated. The aggregate of these realisations should then be a viable approximation of the probabilistic behaviour of the system.
For this problem, it was determined that a model should focus not only on the population distribution over time but also on the dynamics of the system, so that for example the number of times any individual was infected would be recorded. There are several reasons for this; One is that the rate of permanent sequelae is significant, meaning that while an individual may recover from the disease itself the history of their disease is still relevant. Another reason is that the information used to determine when an epidemic is, is the number of cases that has occurred and not the specific number of sick individuals at a given time. These factors leads us to construct a model that focuses on the dynamics of the system.
In this section, we first explain the model constructed for this problem and the theory behind the algorithm used to simulate realisations. Then follows a section on the implementation, both from a mathematical standpoint and a computational, and finally a section discussing the real-world data and parameters used in the project.

2.1 Model

According to the comments made above, a stochastic model focusing on the relevant dynamics was created. We begun by defining what dynamics should be considered, and construct classes of events corresponding to each type of dynamic. For example, we are interested in dynamics such as transmission and recovery and so we defined one event class for each of these. Using this framework, we can record the number of times an event of each class occurs and take it as the system state variable. Given m event classes, the system state variable at time t is the vector $\boldsymbol{N}(t) = [N_1(t), \ldots, N_m(t)]$, each $N_\mu \in \mathbb{N}$. The ultimate goal is to determine the probability $P\left(\boldsymbol{N}(t) = \boldsymbol{n}|X_0\right)$ or: The probability that precisely $\boldsymbol{n} = [n_1, \ldots, n_m]$ events of each class has occurred at time t, given an initial population space X_0.
The population space was constructed by choosing a particular population and compartmentalising it into k sub-populations, similarly to the methodology of classical SIR-models. The compartmentalisation should be such that no individual can be included in more than one sub-population at a time, and no event should increase or decrease a given sub-population by more than one. The compartmentalisation used in this project is shown in figs. 2 to 3 and further discussed in section 2.2.1. The population space at time t, X_t, is a k-dimensional stochastic variable. Later we will discuss its dependence on the event counts $N(t)$. The event counts $N(t)$ form an m-dimensional stochastic process, which under the following assumptions is a density dependent Markov Jump process:

Assumption 1 *For a sufficiently small time interval h, we assume that:*

(i) *Event occurrences in disjoint time-intervals are independent*

(ii) *The Chapman-Kolmogorov equation [14, 4] holds.*

(iii) $P\left(N_\mu(t+h) - N_\mu(t) = 1|X_t\right) = W_\mu(X_t)\,h + o(h)$.

(iv) $P\left(N_\mu(t+h) = N_\mu(t)\,|X_t\right) = 1 - W_\mu(X_t)\,h + o(h)$.

(v) $P\left(N_\mu(t+h) - N_\mu(t) = k|X_t\right) = o(h)$, *for all* $k \geq 2$.

Note that items (iii) to (v) are assumptions on the marginal distributions of events, and not statements about the full system state. Simply put, the assumptions imply that the probability for a specific event increases more or less linearly with time (up to a deviation of $o(h)$). Further, we assume that each rate W_μ is a smooth function of the population. Assumption item (v) ensures that no more than one event can occur simultaneously. Note also that conditioning the probabilities on the population X_t (as we have done) is sufficient, although conditioning based on the event counts $N(t)$ and the initial population state would also hold in this context.
Item (ii) refers to the Chapman-Kolmogorov equation, which is as follows: Consider a stochastic process $\xi_1, \ldots, \xi_k$, with a joint probability density function $p_{\xi_1 \ldots \xi_k}(x_1, \ldots, x_k)$. Then the probability density

$$p_{\xi_1,\ldots,\xi_{k-1}}(x_1, \ldots, x_{k-1}) = \int_{-\infty}^{\infty} p_{\xi_1,\ldots,\xi_k}(x_1, \ldots, x_k)\,\mathrm{d}x_k,$$

corresponding to a straight-forward marginalisation of the joint probability density function. In the specific case of a Markov process, this equation can be written as follows:

$$p_{\xi_3|\xi_1}\left(x_3|x_1\right) = \int_{-\infty}^{\infty} p_{\xi_3|\xi_2}\left(x_3|x_2\right) p_{\xi_2|\xi_1}\left(x_2|x_1\right) \mathrm{d}x_2,$$

or: "The probability that $\xi_3 = x_3$ given that $\xi_1 = x_1$ can be found by considering the probability that the system should go from $\xi_1 = x_1$ to $\xi_2 = x_2$ and then to $\xi_3 = x_3$ for all possible intermediate states x_2, and summarising."
This equation, together with the remaining assumptions, lead to the following result for sufficiently short time intervals:

Lemma 1 *The waiting time H to the next event is exponentially distributed with a rate $W = \sum_{\mu=1}^{m} W_\mu$* □

PROOF The waiting time to the next event is given by the time that the system remains in the same state, so we consider the time h and the probability $P\left(\boldsymbol{N}\left(t+h\right) = \boldsymbol{N}\left(t\right)|X_t\right)$. In order to use the Chapman-Kolmogorov equation, we should consider all possible intermediate states $\boldsymbol{n}$, but as all N_μ are non-decreasing the only possible intermediate state is $\boldsymbol{N}\left(t\right)$. The Chapman-Kolmogorov equation is then:

$$P\left(\boldsymbol{N}\left(t+h\right) = \boldsymbol{n}|X_t\right) = P\left(\boldsymbol{N}\left(t\right) = \boldsymbol{n}|X_t\right)\left(1 - \sum_{\mu=1}^{m} W_\mu\left(X_t\right) h\right) + o\left(h\right) \Leftrightarrow$$

$$\frac{P\left(\boldsymbol{N}\left(t+h\right) = \boldsymbol{n}|X_t\right) - P\left(\boldsymbol{N}\left(t\right) = \boldsymbol{n}|X_t\right)}{h} = -P\left(\boldsymbol{N}\left(t\right) = \boldsymbol{n}|X_t\right) \sum_{\mu=1}^{m} W_\mu\left(X_t\right) + \frac{o\left(h\right)}{h}$$

Letting $h \to 0$, we get:

$$\dot{P}\left(\boldsymbol{N}\left(t\right) = \boldsymbol{n}|X_t\right) = -P\left(\boldsymbol{N}\left(t\right) = \boldsymbol{n}|X_t\right) \sum_{\mu=1}^{m} W_\mu\left(X_t\right)$$

According to standard methods of solving differential equations, we arrive at:

$$P\left(\boldsymbol{N}\left(t+\tau\right) - \boldsymbol{N}\left(t\right) = \boldsymbol{0}|X_t\right) = C \exp\left(-\tau \sum_{\mu} W_\mu\left(X_t\right)\right)$$

As this is a probability density function, its integral must be zero. This is used to determine the constant C, and the solution becomes

$$P\left(\boldsymbol{N}\left(t+\tau\right) - \boldsymbol{N}\left(t\right) = \boldsymbol{0}|X_t\right) = \left(\sum_{\mu} W_\mu\left(X_t\right)\right) \exp\left(-\tau \sum_{\mu} W_\mu\left(X_t\right)\right)$$ ■

The expression above is precisely the equation describing an exponential probability density function with rate $\sum_\mu W_\mu(X_t)$. This result will be used later.
Now, we consider the dependence of the population space on the event space. For each event class μ we define a vector δ^μ, where $\delta^\mu_\varkappa$ describes the effects of event μ on population $\varkappa$. The event counts can then easily be mapped to the population space by constructing a matrix $\boldsymbol{\delta} = \left[\delta^1, \ldots, \delta^m\right]^T$ as $X_t = X_0 + \boldsymbol{\delta}\boldsymbol{N}(t)$, but note that some information is lost in this projection. Using this description, there are no changes in the population outside of the events, which matches the initial idea that all relevant dynamics should be modelled using the event framework.
Earlier we mentioned that the event process is a density dependent Markov jump process. The jump aspect of the process is in the discrete nature of the event counts, meaning that the system state jumps from one state to another with the instantaneous occurrence of an event. The density dependence is introduced via the probability rates $W_\mu(X_t)$, as they describe the future evolution of the process based on the current population and the current system state. Then only the Markov property should receive some further consideration. Again, we look to the W_μ. They depend on the population state (and consequently the system state) X at time t, but not on any earlier state. This is precisely the Markov property.
From assumption 1, we are able to make the following proposition:

Theorem 1 *Under the previous assumptions, the dynamics in the event space obeys the Kolmogorov Forward equation.*

$$\frac{d}{dt}P\left(N(t+s) = \boldsymbol{n} | X_t\right) = \sum_{\mu=1}^{m} P\left(N(t+s) = \left[n_1 \ldots n_\mu - 1 \ldots n_m\right]\right) W_\mu(X_{t+s} - \delta^\mu) - \quad (1)$$

$$- \sum_{\mu=1}^{m} P\left(N(t+s) = \boldsymbol{n}\right) W_\mu(X_{t+s}) \quad (2)$$

□

Here we use X_{t+s} to mean the population state given the event counts $\boldsymbol{n}$. The interpretation of the equation is as follows: The probability that a system should be in a given state at a given time increases as the probability that it should enter the state increases, and decreases as the probability that it should leave that state increases. The first term corresponds to the probability that it should enter the given state, as it is the probability that it should be in an immediately preceding state and enter the given state. The second term is interpreted similarly.
Theorem 1 provides us with a differential equation describing the probability distribution evolution for the event counts, but in general there are no simple solutions to it. Instead we looked for an approximate solution using the Monte Carlo-framework, meaning we constructed many artificial "samples" of the stochastic variable $\boldsymbol{N}(t)$, and assumed that the distribution of the samples can be used as an approximation for the distribution of the system. This required us to find an effective method of simulating a system as the one we have described, and one such method is the Feller-Kendall algorithm, also known as the Gillespie algorithm [13, 8].
To construct a realisation of the system, we worked iteratively and used the fact that the system state is constant between events. This means that the only times we were interested in were the

times where an event occurs. Furthermore, we were only interested in the very next event, as that event may have changed the probability distributions for future events. Consequently, in order to move the realisation forward in time, the stochastic variables we needed to sample were the time of the next event, H, and what event it is that occurs, M. An initial idea for sampling these variables together was the following:

Algorithm 1 *Let the system be in state N at time t, with populations X.*

(i) *Calculate the probability rates $W_\mu(X)$*

(ii) *Sample times H_μ for the time of the next event of each event class*

(iii) *Choose the event class M so that $H_M = \min\{H_1, \ldots, H_m\}$*

(iv) *Increase the event count of M by one:*

$$N := [N_1(t), \ldots, N_M(t) + 1, \ldots, N_m]$$

(v) *Update the time and the population:*

$$X := X + \delta^M$$
$$t := t + H_M$$

(vi) *Repeat*

This algorithm would work, but is not very efficient. In each step, m random numbers are generated, of which $m - 1$ are discarded. If the event count is high, this could lead to a significant increase in simulation time on each step. Instead, we used the following result (lemma and proof from [3]):

Lemma 2 *M and H are independent.* □

In order to prove this, we use the definition of independence: $P(A \cap B) = P(A) \cdot P(B)$. We will determine the joint probability density function of (H, M) and show that this is the product of the marginal distributions.

PROOF First, we consider the marginal distributions. In lemma 1, we showed that $H \sim \exp\left(\sum_{\mu=1}^{m} W_\mu\right)$. We now consider the distribution of M, by first studying the case where there are two exponentially distributed independent variables and generalising. Let $S \sim \exp(\lambda_S)$, $T \sim \exp(\lambda_T)$ be independent. We will determine the probability that S occurs before T, or $P(S < T)$. Using independence and the exponential distribution, we get:

$$P(S < T) = \int_0^\infty P(S = s)P(T > s)\mathrm{d}s = \int_0^\infty \lambda_S \mathrm{e}^{-\lambda_S s}\mathrm{e}^{-\lambda_T s}\mathrm{d}s = \tag{3}$$

$$= \frac{\lambda_S}{\lambda_S + \lambda_T}\int_0^\infty (\lambda_S + \lambda_T)\mathrm{e}^{-(\lambda_S+\lambda_T)s}\mathrm{d}s = \frac{\lambda_S}{\lambda_S + \lambda_T}, \tag{4}$$

showing that the probability that it should be a particular event is proportional to its rate. Now for the generalization: Consider a set of variables $T_j \sim \exp\left(\lambda_j\right), j = 1, \ldots, I$. Let $S = T_i$ and U the minimum of the remaining variables. By the result from lemma 1, $U \sim \exp(\sum_{j\neq i}\lambda_j)$. Combining this with eq. (4), we have

$$P\left(T_i = \min\{T_1, \ldots, T_I\}\right) = P(S < U) = \frac{\lambda_i}{\sum_{j=1}^I \lambda_j}, \tag{5}$$

implying that the marginal distribution for M is proportional to the individual rates W_μ.
Now to compute the joint probability distribution for (H, M).

$$P\left(H = h \cap M = \mu\right) = P\left(H_\mu = h, H_n > h \text{ for } n \neq \mu\right) = W_\mu \mathrm{e}^{-W_\mu h}\prod_{n\neq\mu}\mathrm{e}^{-W_\mu h} =$$

$$= W_\mu \mathrm{e}^{-h\sum_{n=1}^m W_n} = \frac{W_\mu}{\sum_{n=1}^m W_n}\left(\sum_{n=1}^m W_n\right)\mathrm{e}^{-h\sum W_n},$$

which is precisely the product of the probability density and distribution of H, M. Hence they are independent. ∎

This result allowed us to sample H and M separately from the marginal distributions defined above. The new algorithm was:

Algorithm 2 *Let the system be in state N at time t, with populations X:*

(i) *Calculate the probability rates $W_\mu(X)$ and use this to determine the distributions of H and M*

(ii) *Sample H and M*

(iii) *Increase the event count for class M by one,*

$$N := [N_1(t) \ldots N_M(t) + 1 \ldots N_m(t)]$$

(iv) *Update the time and the population:*

$$t := t + H$$
$$X := X + \delta_M$$

(v) *Repeat*

This method only requires two random variables to be sampled in each iteration, and is therefore significantly faster than the option presented in algorithm 1. This is the basic Feller-Kendall algorithm, and also the one used in this project.

2.2 Implementation

The first part of this section focuses on how the model described above was applied to the specific problem discussed in this paper. The population compartmentalisation is explained, and an example of an equation for calculating the rates is given. Furthermore, the introduction of a time dependency is discussed. The second part focuses on the work to implement a computer program suitable for generating realisations of the system.

2.2.1 Events and populations

When constructing a compartmentalisation suitable for this problem, we started by considering what dynamics we wanted to model as events. We then constructed the sub-populations in such a way that each event would increase one sub-population by one, decrease a sub-population by one, or both (moving a single individual from one population to another).
The types of dynamics that we wanted to model in this project can roughly be divided into two categories, dynamics relating to meningococcal disease and dynamics present in the healthy population. Due to the rather long intervals between the large-scale epidemics, it was determined that changes in the population distribution, as well as ageing individuals, could affect the model. This is the reason for including some dynamics of a healthy population, and finding a suitable approximation for the large-scale population variation over time was the first step.
The dynamics concerning a healthy population includes death by causes other than meningococcal disease, ageing and birth. The compartmentalisation and the events for a healthy population is shown in fig. 2. The population is divided into three age groups (infants (0-4 years), young (5-14 years) and adults ($\geq$ 15 years)), based on available data for the population and the disease. Death and birth was included in a slightly simplified manner. It was assumed that death only occurs in the oldest age group, and that only the oldest age group bear children. In order to compensate for the lack of infant mortality, the birth rate was based on the number of children that survive the first year.
Modelling meningococcal disease is the primary goal of this project, and as such much care must be put in determining what events and populations to include. The compartmentalisation and events relating to disease (for a single age group) are shown in fig. 3. A central part of the disease dynamics in this case is the presence of non-symptomatic carriers, and how this relates to transmission and immunity. Here, we chose to model the acquisition of carriage (on a susceptible individual) as a two-step process. Consider the event of transmission to a susceptible individual. This moves one individual from the sub-population ”Susceptible” into a sub-population we call ”Infected”. Subsequently there is a second event, either one where the individual becomes sick or one where they become a carrier. This event occurs on average 10 days after the transmission, corresponding to the incubation time for meningococcal disease. Carriers can then recover in one of two ways, either becoming immune to invasive disease or becoming susceptible. When it comes to transmission, we assume that carriers can transmit the bacterium to other individuals. We consider immunity to be protective against invasive disease, but not against carriage, so immune

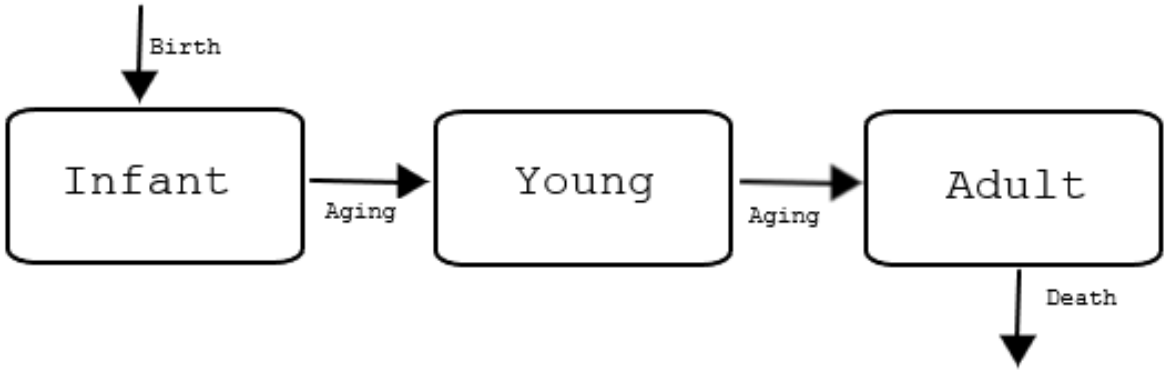


Figure 2: Schematic figure of the healthy population and the events that affect it. Each arrow symbolises the effects of a specific event, and the type of dynamic it is intended to mimic is specified next to it. Arrows between subpopulations indicate that a single individual is transferred between them, and arrows that only connect to one population only affects that one population.

individuals can become carriers. In this case they never pass through the infected sub-population but are instead immediately moved to the carrier population.
The two systems shown in fig. 2 and fig. 3 were combined, leading to a more complex system. However, while the system is more complex, these graphs should be sufficient to reconstruct the compartmentalisation, and all events used are more extensively defined in section A. We only comment on the choices made regarding the ageing events when performing the combination of the system. Ageing is a comparatively infrequent event per population, and it is unlikely that individuals should age in the short period of time when they are sick, or in the time between transmission and disease or carriage. On an individual basis, each individual should be in that state for a very short time, and on a population basis very few individuals should be in those groups at any given time. Based on these assumptions, ageing is only considered for carriers, immune and susceptible individuals. Similar assumptions are made when calculating the birth and death rates.
When we constructed the events and the populations, the equations governing the events were also created. We assumed that the rates follow the law of mass-action, so that the rate is proportional to the concentration of the relevant population. This also means we assumed the population is perfectly mixed at all times, or (equivalently) that the population is evenly and randomly distributed in space. The equations for each event rate can be found in section A, but we will explain the basic idea behind them here using a (simplified) example. We will consider the transmission event to a susceptible individual, but for now we will not care about the different age groups. The generalisation is straight-forward.
Consider an infective individual. In order for this individual to transmit disease, they must encounter a susceptible individual, and this encounter must also be sufficient to transmit bacteria. We call the expected number of contacts that lead to transmission per time unit and infective individual β. However, some of these contacts will be with non-susceptible individuals. As we expect the population to be evenly distributed, we can assume that the chance that a given individual is susceptible is equal to the relative number of susceptible individuals in the population, $\frac{S}{N}$ where S is the number of susceptible individuals and N the full population. This holds for each infective individual, of which there are I, which gives us the rate W:

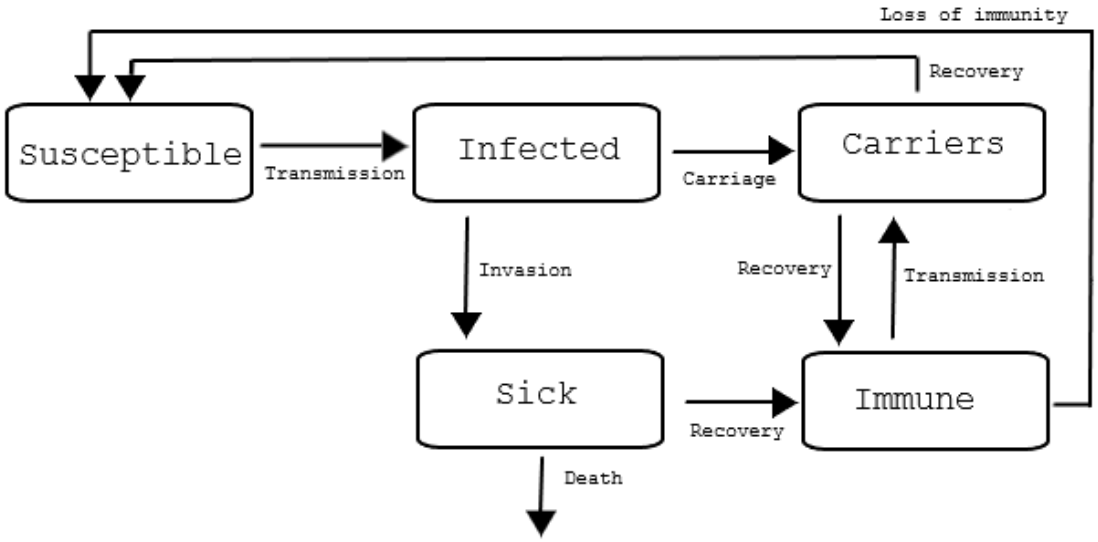


Figure 3: Schematic showing the subpopulations and events related to disease status. Each arrow symbolises the effects from one of the events on the population. An arrow between populations indicate that a single individual is transferred in the shown direction. Arrows that only connect to one of the subpopulations indicate that only that subpopulation is affected by the event.

$$W = I\beta\frac{S}{N}. \tag{6}$$

2.2.2 Time dependency

One of the main goals of this thesis was to examine possible explanations for the seasonal difference between the endemic and the hyperendemic state. In order to evaluate the effects of different mechanics, we introduced events with a time dependency, that were intended to model different mechanics for different time periods. We let the probability rate for these events be zero during the season where they are not active. For this project, the activation and de-activation of these events were deterministic. The combination of these methods lead to discrete and deterministic seasonal shifts. The reason behind this choice is predominantly simplicity, but it also results in a system where the changes introduced by the shift in mechanics will be very clear and easy to compare.

Due to the nature of the algorithm we use (algorithm 2), deterministic changes can be slightly complex to introduce. Time moves forward in discrete and random steps, meaning that the time of introduction of the new event will never explicitly occur. We solved this by introducing the new events immediately following the first event after the predetermined time. Provided events are frequent enough, this should not introduce any errors of more significance than what is introduced by approximating the seasonality as an entirely deterministic process.

Two types of seasonal events were introduced, one intended to model a behavioural change in the population and one intended to model an external change in the disease itself. For the event concerning behavioural change, we considered the specific dynamic where several, typically separate populations, are mixed due to migration or lack of resources. This was modelled by constructing a realisation of two separate populations in parallel and having a seasonal event corresponding to cross-transmission, where infective individuals in one population can transmit bacteria to individuals in the other population.
When we considered a change in the disease itself, instead of introducing more events during the dry season, we had one set of events during the dry season and another during the wet season. In general, the only difference between these sets was in (some of) the parameters used in the equations governing the probability rates (see β in eq. (6)). Predominantly we considered a seasonal variation in the invasiveness parameter, which affects the probability that a susceptible individual that is infected by the bacterium becomes sick rather than becoming a carrier.

2.2.3 Computer implementation

The model was implemented in `C++`, and the program could easily be adapted to a different problem. All analysis was performed in `Python`, using the `numPy` library.
The main parts of the program consist of a `Population` class, a `Simulation` class, an interpreter class for parameters and constants and a library of `Event` classes. The `Population` class contains an array of integers keeping track of the number of individuals in each subpopulation. All changes to any of the populations are performed through this class.
The `Event` library can be pre-compiled, and consists of a derived `Event` class for each event defined in the model. These calculate the rates for each given event and contain a pointer to the `Population` instance they affect. This pointer is used to execute events as they occur, changing the population accordingly.
A `Simulation` is used to construct one realisation of the system. It contains an array of pointers to the `Event` base class, each pointing to an instantiation of the derived `Event` class specific for each event. The main function of this class is to perform the algorithm used to advance the realisations of the system (algorithm 2). It also keeps track of time and uses this to record the process. When the time of the next event is the following day, the current system state (event counts) and the population is saved as a series of snapshots. This data is saved for analysis.
In order to perform several simulation runs with different configurations, a class responsible for interpreting parameters was created. This class reads from a `json`-file at runtime and passes parameters to relevant classes. This method allows for changing the configuration without recompiling the whole program.
The data is saved as `txt`-files in a folder specific for each run. These folders also contains the `json`-file with the constants for that specific run. Population- and event data is saved as a 2-dimensional array of integers, each line describing the values at a given time. Population indices are found in the constants file, and event indices are found in a separate file with short descriptions for each event.
Each model configuration was run 100 times. This was estimated to be enough to examine the general behaviour of the system, and to give an estimate of the expected behaviour. Furthermore, in cases where there are no notable deviations between realisations, we also assume the variations

from realisation to realisation to be Gaussian distributed. This allows us to use standard methods for calculating confidence intervals.

2.2.4 Data processing and analysis

Scripts for analysing the data were written in `Python`, using libraries `NumPy` and `StatsModels`. Most calculations are straight-forward estimates, using the arithmetic mean as an estimate of the expected value and the measurement variance as an estimate of the variance. Mean values was calculated using the `NumPy` mean function.
When calculating incidence rates, data on the events was used. Initially it was differentiated to find the daily events data. In order to account for transient behaviours and only consider a steady state solution, the data from the first 2000 days of each realisation was discarded. Again, the arithmetic mean for all realisations was considered. These steps yield an estimation of the average daily event counts. Using this, we calculate the weekly incidences for each day (except the first six) as the sum of the the last seven days, including the current. Using this method, rather than dividing the sequence into weekly segments and summing over these, leads to more data points, but also means that the data points are not independent. This must be considered when estimating the statistical properties of the resulting values. In this report it is primarily relevant when calculating the variance of the mean estimate. We also note that enough data points are produced that the aggregate will be approximately Gaussian distributed. Using tools from time series statistics [12], the variance of the mean as an estimate of the expected value, for large enough data sets, is:

$$\mathrm{V}(\hat{m}) \approx \frac{1}{N} \sum_{k=-\infty}^{\infty} r_x(k)$$

where k are the lags, r_x is the auto-covariance function and N is the number of available data points. The standard biased estimator for the auto-covariance function supplied by `StatsModels` was used to find an estimate of r_x:

$$\hat{r}_x(k) = \frac{1}{N} \sum_{t=k+1}^{N} (x_t - m)(x_{t-k} - m).$$

This estimator can be somewhat reliably used for lags k up to about $\frac{N}{4}$. For larger lags, the number of data points is too small. Based on the properties of the system, as well as the assumptions on events in disjoint time intervals being independent (item (i) in assumption 1), we let $\hat{r}_x(k) = 0$ for all $k > \frac{N}{4}$.

2.3 Acquiring data and parameters

This implementation is designed to approximately match a theoretical population in northern Nigeria. The region was chosen as there is reliable population and disease data.

Parameter	Value
Crude birth rate per 1000 population	39.1
Infant mortality rate per 1000 population	69.4
Life expectancy	65 years
Ratio in ages $0-4$	0.16
Ratio in ages $5-14$	0.26

Table 1: The parameters governing the initial population distribution and the events relating to a healthy population.

2.3.1 Population and climate data

The only climate data used is when the wet season occurs, which in this region is approximately from July through September. The year start is set to the first of August, to ensure a simulation start and end with as few interesting events as possible. We use a year of 365 days, with the dry season occurring between days 65 and 335.
The population distribution for the healthy population was sourced from Gapminder [7], and the chosen number of individuals for a base population is 50000. Out of this number, 58% are adults, 26% are 5-14 years old and the remaining 16% are younger. The population size was chosen as being large enough to provide a reasonable estimate of an actual community, while still being small enough that simulation time was reasonable. Larger populations could also decrease the value of the stochastic methodology.
In order to accurately model the behavioural variations of the population, some data for these are required. However, within the scope of this project, this data could not be acquired in a reasonable manner. It is clear that migration is frequent in this region [5], and it is furthermore clear that it affects seasonal variations in for example measles [1]. In this project, only a qualitative estimate of how migration works is considered, and the resulting values should not be used as numerical approximations but as qualitative assessments of whether a factor could be potentially influential. We consider the migration patterns by simultaneously modelling two equal sets of the base population and introducing cross-transmission during the dry season.
Some relevant parameters for the events related to the healthy population are presented in table 1. All values are from Gapminder.

2.3.2 Disease data

Several articles were considered when acquiring parameters for the modelling effort. Parameters were determined from numbers estimated based on the aggregate of articles. In table 2, numbers used to calculate the rates are presented, together with sources that support them. Two virulent serogroups were considered, A and W-135. In the cases where the parameters differ between them, it is clearly marked in the table. The same is true for cases where the parameters differ between age groups.
Note that these numbers need further mathematical operations to derive the rates used in the implementation. For example, if we consider the rate at which a sick individual dies, we would expect about $\frac{1}{10}$ to die within one week (the duration of disease), leading to the rate $\frac{1}{70}\left(\text{day}\right)^{-1}$.
The equation for each parameter can be found in section B

Parameter	Value	Source
Half life of carriage	90 days	[2]
Duration of disease	7 days	[11]
Risk of death given disease	0.1	[26]
Risk of invasive disease, given infection	0.01	[17]
	0.002	Estimated from case-carrier ratios
Time from transmission to disease or carriage	10 days	
Chance of acquiring immunity after carriage (A)	0.71	[21]
Chance of acquiring immunity after carriage (W-135)	0	[21]
Risk of lost immunity after 4 months (A, $<$ 14 years)	0.15	[21]
Risk of lost immunity after 4 months (A, $>$ 14 years)	0.06	[21]
Risk of lost immunity after 4 months (W-135, $<$ 14 years)	0.57	[21]
Risk of lost immunity after 4 months (W-135, $>$ 14 years)	0.5	[21]
R_0	1.03	Estimated by carriage rates

Table 2: The values used to estimate the parameters needed to model the system. The equations relating each value above to the relevant parameters can be found in section B

The transmission rates depend heavily on the model and the parametrization, and are in general impossible to measure. In this project, the estimate for the transmission rates are based on estimate of the basic reproductive number, R_0. This number is the expected number of individuals that disease will be transmitted to, by one single individual carrying the disease in an entirely susceptible population. This number is assumed to be equal for carriers and sick. As carriage lasts so much longer compared to disease, this leads to a much higher transmission rate from sick individuals than from carriers, which, due to the nature of infections, is reasonable. The base assumption, however, is not based on real data. Rather, it is used as it provides a simple way to reconstruct the model and its parameters. The transmission rate from both carriers and sick individuals is largely unknown, and very difficult to estimate and measure.
With R_0 and the expected duration of disease and carriage, the transmission rates are fairly easy to calculate. We also note that the transmission rates to immune and to susceptible are equal, the only difference being that immune individuals will not be subject to invasive disease. This choice was made as there was no data suggesting that the naturally occurring immunity would lead to a reduction of carriage. There are some data that suggests that the modern conjugate vaccines induce immunity sufficient to prevent carriage, but these vaccines are not modelled in this project. [18]
Estimation of the basic reproductive number was made using by trying a range of different values and choosing the one producing the carriage rate closest to a target value of 5% [20, 21, 15]. The estimated value is $R_0 \approx 1.03$
One way to approximate the risk of invasive disease given infection is by studying case-carrier ratios, how many disease cases there are for each carrier. This value is what is presented in [17], and is also what was used in the first iterations of the model. However, this resulted in an unexpectedly high number of cases, both compared to the population as a whole and compared to the number of carriers. Instead, for the data presented in this report, the second value in table 2 is

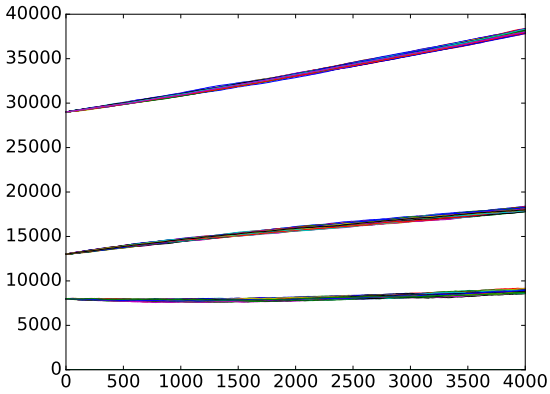
used. This was estimated by varying the invasiveness parameter studying the resulting case-carrier ratios, and matching them to the value presented in [17]. Possible reasons for the first method not working are considered in the discussion.
The initial population values are expected to be of limited importance, as the system stabilizes into something of a steady-state solution, and not a lot of effort was put into making these values optimal. Estimates of population immunity ranges widely, and differs between what measures were used [9, 21]. For this project, we use an initial value of 22.5% of the population in each age group.
One hypothesis on the increase in disease cases over the dry season is a change in the risk of infection given disease. In this report, an increase of $25\times$ is used, and this increase is sourced from [20], that suggests a value between 7-67 would be reasonable.

3 Results and analysis

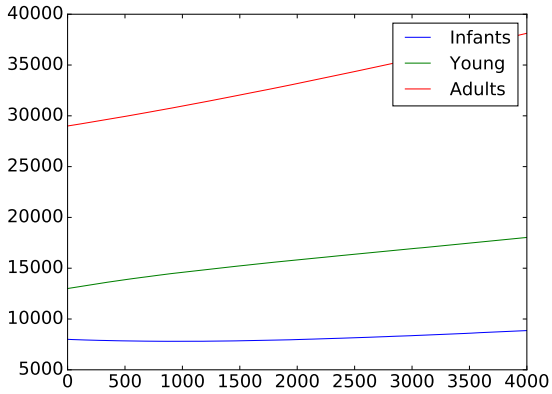
In this section, some results from the different simulations are shown and explained. All data can be found online at `www.github.com/chuzzle`, and some relevant examples are shown in this section. First the validity of the model is examined, and subsequently the results from trials for different seasonal models is presented.
In general, the arithmetic mean of several realisations is presented. The choice to do so is based on the fact that there are no significant qualitative difference between realisations, implying that the mean represents an adequate estimate of the probable behaviour of the system. In the case of realisations deviating significantly from the mean, further comments will be made.
The disease rates are generally too small to be clearly seen in the graphs presented (real-world data suggest they should be $\leq 0.00001\%$ of the population in the endemic state), and we choose to present incidence rates instead, where applicable. Incidence rates are measured as the number of disease cases per 100 000 individuals during a week. These results are presented in tables, and the reader is advised to keep the scale in mind when comparing results.

3.1 Validation results

A graph showing the average of the runs for a system with no disease is presented in fig. 4. This configuration is intended to determine whether the baseline population model is sound, as presented in section 2.2.1. The interesting aspects to consider are the population distribution over time, as well as the population growth. We find that the population distribution is fairly constant over time, while the population growth is at $\sim 2.6\%$.
The results in fig. 4 should also be used to estimate the validity of the choice to primarily present the mean of the results. Studying the raw data in fig. 4a shows no significant variations between realisations, while still presenting enough variations to ensure that the simulations are in fact random. Additionally, comparing the raw data to the arithmetic mean presented in fig. 4b supports the theory that presenting the mean is sufficient. The results presented further ahead in this report are, unless otherwise stated, shown in the mean.
We also consider the validity of the model after the introduction of endemic disease for two different serogroups. The results are shown in fig. 5. The primary goal of this simulation is to examine whether the model could mimic the endemic behaviour of the disease, with the appropriate number of disease cases and carriers. The results from these runs are quite similar to



(a) Raw data from a set of realisations on a system without disease. Note how all runs seem to be fairly similar.



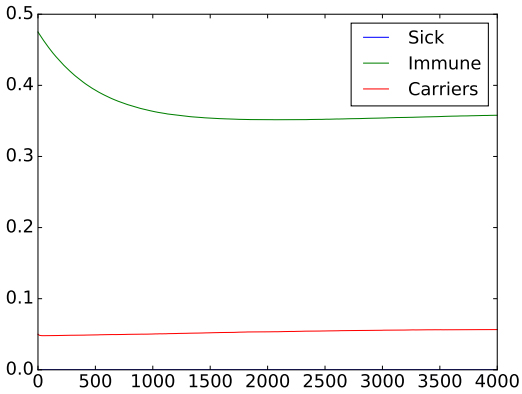
(b) Average of the runs shown in fig. 4a.

Figure 4: Basic data sets for a set of realisations of a system without disease. Right is an average of the simulations on the left.

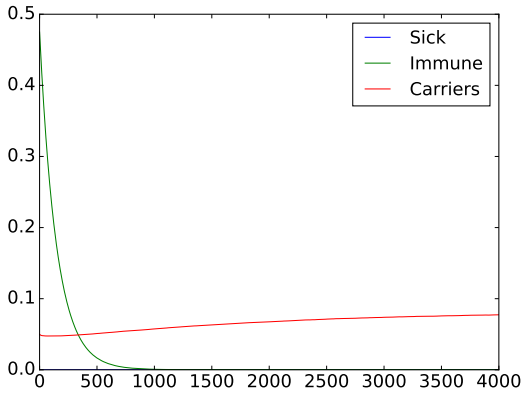
the available results from field research for both serogroups. Also note that there is no significant ratio of immune individuals of any age group for the system with serogroup W-135, as carriage of that strain does not induce immunity in the model.
Considering the model with disease included, it is important to note that the parameter governing the transmission of disease was used to have the system match the data, due to lack of available research. A common estimate of the parameter R_0 was produced using a set of simulations, and the value producing the best correspondence to real data was chosen. As such, these results are not as reliable as would be desired. While the results support the theory that this model can be used to mimic the dynamics, it is not confirmed that the model is accurate.
Real-world data suggest that the weekly incidence rate during the endemic period is below 1 per 100000 individuals. The average incidence rates provided by the model matches these values quite well, as would be expected when several of the parameters are estimated in order to better show correspondence to data.

3.2 Age dependence

As mentioned in the introduction, there are some variations between age groups concerning carriage and disease rates. These differences can be observed in the simulations corresponding to serogroup A, and plots over the age-wise disease distribution for this serogroup are presented in fig. 6. There are no clear variations by age for serogroup W-135, and as such no new plots for that disease is shown. Consider instead fig. 5b, which corresponds to the results for each age group.
When comparing the graphs in fig. 6, there are a few interesting aspects to comment on. The ratio of immune varies significantly by age. The increase in immunity from age groups 5-14 to adults can partially be explained by the increase in expected duration of immunity for adults. However,



(a) Average of one set of runs on a system with parameters matching disease of serogroup A in an endemic state. No seasonal variations were introduced in these simulations. The ratio of disease is too small to be seen in this graph, and is presented in table 3.



(b) Average of one set of simulations on a system with parameters matching disease of serogroup W-135 in an endemic state. No seasonal variations were introduced in these simulations. After transient behaviours, no significant population immunity can be seen. Disease rates are also too small to be seen in this graph, and is presented in table 3.

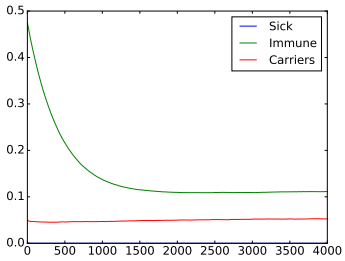
Figure 5: Figures showing the results from simulation for a system with parameters matching two different serogroups of disease. Overall, the results are similar except in the subpopulations of immune individuals. Immunity is lacking from all age groups for the system corresponding to serogroup W-135, after transient behaviour. Note that disease incidence is too small to be seen in these graphs, and presented in table 3.

there is no such parametrical explanation for the increase between ages 0-4 and ages 5-14. This increase must be due to the dynamics included in the model, and cases of repeated carriage leading to immunity.

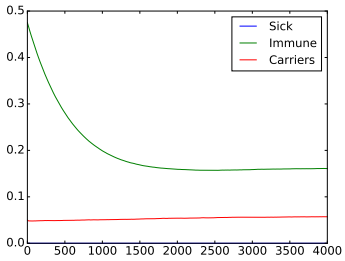
While there are some clear variations in immunity ratios between the ages, there are no clear variations in carriage rates. In real-world data, variations in the carriage rates between age groups are frequently observed. The fact that no such variations are shown in the results here suggest that the mechanics that leads to such an increase are not included. Possible examples include variations in transmission rates by age.

The relative incidence rates for serogroup A (shown in table 3) show some of the variations expected, in that disease is more common for the younger individuals of the population. This is not true for serogroup W-135. Likely explanations for the presence of this dynamic for serogroup A is that there is a pattern of increased immunity by age. As there is no significant population immunity for W-135, those dynamics can not be seen. When considering incidence rates, we should also note that the expected peak for teenagers is not clearly present for either serogroup, similar to what was seen in the carriage rates.

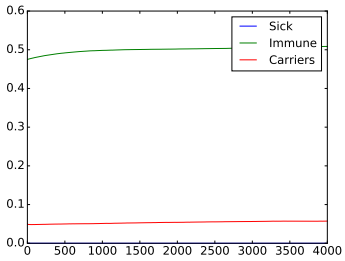
For serogroup W-135 the ratio of immunity seems insignificant for each age group, and carriage seems similar between ages. As such, there should be no clear variations by age in the incidence rates either. However, there seem to be some variations outside the confidence bounds. These



(a) Graphs showing the distribution of carriage, immunity and disease in ages 0-4. The ratio of disease is too small to be seen in this graph, and is presented in table 3.



(b) Graphs showing the distribution of carriage, immunity and disease in ages 5-14. The ratio of disease is too small to be seen in this graph, and is presented in table 3.



(c) Graph showing the distribution of carriage, immunity and disease in ages 15 and up. The ratio of disease is too small to be seen in this graph, and is presented in table 3.

Figure 6: Graphs showing the distribution of disease in the different age groups modelled in this project, younger to older left to right, for serogroup A. As disease rates are too small to be seen in these graphs, they are presented elsewhere in a more appropriate format and a more appropriate scale, see table 3. In these graphs, the ratio of immunity is particularly interesting.

variations are either the result of complicated and interesting interactions within the system, or suggest that the estimates used for the confidence intervals are too small.

3.3 Seasonal variations

The seasonal variations in number of disease cases was postulated to have two possible explanations, and both of these are examined in this section. First we consider an explanation from behavioural changes, and then an explanation from the biology of either the bacterium itself or the humans it infects. It is worth commenting that in the second case, the parameter can be controlled so that the data matches reality, as the parametrization is dependent on the model itself. Only a negative result would be significant here, as it means that the model can not be applied to this case. Validation of this theory must come from some other source.
As these simulation sets were run for both of the invasiveness parameters presented in table 2, one additional interesting result was provided: The increase in incidence rate was very similar with both invasiveness parameters. This is not surprising when it concerns the parametrical variation, but it might have implications in the behavioural model.

3.3.1 Behavioural changes

A graph showing the seasonal variations in disease cases for this type of run is shown in fig. 7, and incidence rates are presented in table 4. Disease rates are too small, even in the hyperendemic state, to be seen in the graphs, but some patterns appear in both sets of results. It is clear from the graphs that there are significant differences between the different seasons, and the weekly incidence rates reflects these variations, showing a significant difference between the dry and the wet season. While the increase in incidence rates is not as large as suggested by real-world data,

Event type	Weekly incidence rate per 10^5 individuals (95% CI)	Relative incidence rate (95% CI)
Invasion, any age, A	0.425 (0.419, 0.431)	-
Invasion of infant, A	0.078 (0.076, 0.081)	0.489 (0.475, 0.506)
Invasion of young, A	0.175 (0.171, 0.178)	0.672 (0.658, 0.685)
Invasion of adult, A	0.172 (0.168, 0.177)	0.297 (0.290, 0.305)
Invasion, any age, W-135	0.829 (0.806, 0.853)	-
Invasion of infant, W-135	0.116 (0.113, 0.118)	0.722 (0.706, 0.738)
Invasion of young, W-135	0.235 (0.229, 0.240)	0.903 (0.881, 0.923)
Invasion of adult, W-135	0.479 (0.456, 0.502)	0.826 (0.786, 0.866)

Table 3: Table displaying some results for a system modelling the endemic state of disease. The weekly incidence rate is calculated as the number of times the given event occurs, and the relative rate is found by adjusting for the relative portion (assuming adherence to initial values, see table 1) of the population that are in the given age group.

there is a clear seasonal effect. It is possible that further increases in the population density will lead to further increases in the incidence rates, but this is not explicitly tried in this project.
Another interesting result for these runs are the carrier rates, which are about four times larger than what is produced for a run without seasonal changes. This difference can be seen both during the wet season and the dry season. The increase during the dry season is explained by the increased rate of disease spread during this time, and as the expected duration of carriage is so long this will also affect the ratios during the wet season.
According to [20], the case-carrier ratios increases as the system transitions from endemic to hyperendemic incidence. This is self-evident in the case of the parametrical variation, but what is more interesting is the fact that this seems to occur also when behaviour is the cause of the transition. The increase in case-carrier ratios is very similar to the increase in incidence rates, suggesting that the case-carrier ratios may be more useful in studying variations in the incidence. This would be the expected case for a relatively constant carriage rate. The results suggest that, while the seasonal variations in carriage rate can be clearly seen in fig. 7, they may not be very influential.
These results do not differ significantly between the serogroups examined. The results that do differ is again the immunity rates, where there is no significant population immunity against serogroup W-135. As the remaining results are so similar, it is reasonable to assume that population immunity, at least on these scales, does not affect the difference between the endemic and hyperendemic states.

3.3.2 Parametrical changes

Compared to the behavioural model, the seasonal variations introduced via parameter changes affect most ratios significantly less (see fig. 8). While some seasonal variations can still be seen, the largest change is seen in the incidence rates of disease (table 4). This is to be expected, as the parameter changed is the one governing the incidence rate. In some way, it is more surprising that there are in fact changes in the other populations as well. This could be caused by the increase in

Model	Season	Weekly incidence rate per 10^5 individuals (95% CI)	Weekly case-carrier ratio $\times 10000$ (95%CI)
Behavioural model, A	Dry	1.561 (1.549, 1.573)	0.475 (0.475, 0.475)
	Wet	1.147 (1.122, 1.172)	0.351 (0.351, 0.351)
Behavioural model, W-135	Dry	4.044 (4.020, 4.069)	1.169 (1.169, 1.169)
	Wet	3.118 (3.071, 3.166)	0.869 (0.860, 0.878)
Parametrical change, A	Dry	9.393 (9.334, 9.453)	17.335 (17.180, 17.490)
	Wet	0.965 (0.821, 1.109)	1.780 (1.511, 2.049)
Parametrical change, W-135	Dry	20.734 (20.399, 21.070)	28.414 (28.300, 28.528)
	Wet	2.059 (1.704, 2.413)	2.815 (2.33, 3.302)

Table 4: Table displaying weekly incidence rates of disease for systems incorporating different models of seasonal variations. The incidence rates increase by about 1.3 times in the dry season in the behavioural model, and about 10 times in the parametrical model. These ratios between the case-carrier ratios are similar. Also note that the case-carrier ratios are lower than real-world data for the behavioural models.

disease rates, as people subject to invasive disease are modelled as more infective per time unit. Sick individuals and carriers will transmit the disease to as many individuals (on average), but over different time periods. If there are many sick people at the same time, this will lead to a momentary increase in the transmission rate for the whole population, which will in itself affect the number of carriers. The increase in immune is directly caused by the increase in disease, as recovery leads to immunity.

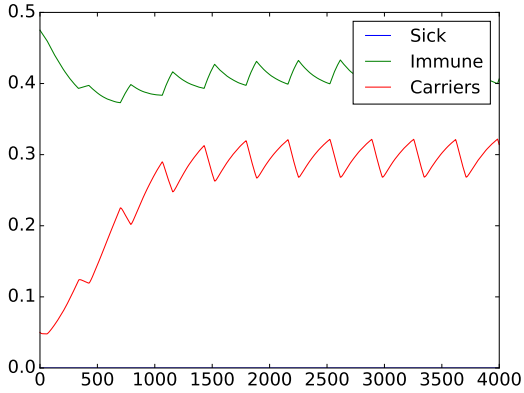
The increase in incidence rate between dry and wet season is on the low end of what is expected (an increase of about 10 times, with an expected increase of 10 - 100 times), but this is reasonable as the chosen parameter change is on the low end of what is suggested to match the change from endemic to hyperendemic. Furthermore, while it is true that the increase is smaller than expected, the incidence rates themselves are high enough to match what would be considered an epidemic incidence rate during the dry season and hyperendemic during the wet season. Some possible explanations are discussed further down.

One thing that sets this specific iteration of the model apart is the presence of immunity in a system affected by serogroup W-135. This is caused by the large number of disease cases in this system, as people that recover from invasive disease are always modelled as immune. Whether this level of immunity is reasonable or not is difficult to determine, but it is a significant result.

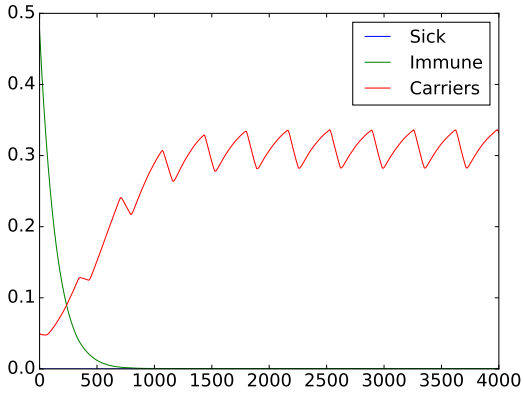
4 Discussion

The goal of this project has been to create a stochastic and discrete model for the dynamics of Meningococcal disease in the African Meningitis belt. In large, the project has been successful, in that a model has been created and implemented, and some interesting results have been produced. The model can emulate the endemic and the hyperendemic incidence rates measured in the region, and seems flexible enough to model different setups.

A further goal has been to examine possible explanations for the seasonal variations, and the results in this area have been inconclusive. Using only parametrical changes to model the transition from



(a) Graph showing some data for a system corresponding to a population with seasonal behavioural variations, subject to disease of serogroup A. Displayed are the ratios of the population in all age groups that are immune, carriers or sick. The seasonal variations in the ratio of immune and carriers are very clear. The ratio of sick is too small to be seen in this plot, and incidence rates are presented elsewhere. The carriage rates for this system are significantly higher than in the system without seasonal variations.

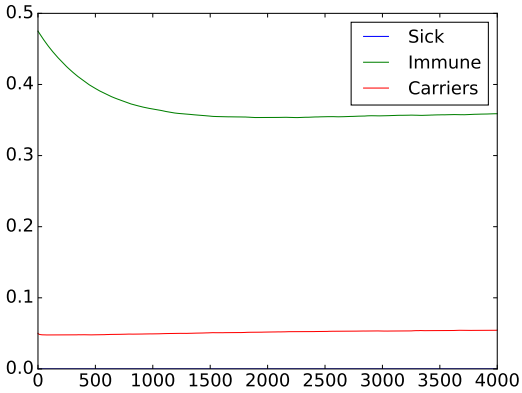


(b) Graph showing some data for a system corresponding to a population with seasonal behavioural variations, subject to disease of serogroup W-135. Displayed are the ratios of the population in all age groups that are immune, carriers or sick. The seasonal variations in the ratio of carriers are very clear, but the ratio of immune and sick individuals are not significant enough to be seen in this graph (after transient behaviour). The carriage rate is significantly higher in this system than in the corresponding system without seasonal variations.

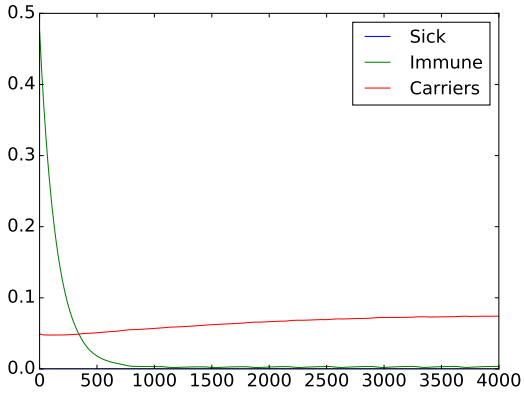
Figure 7: Figures showing some data for systems corresponding to seasonal variations caused by behavioural changes, for the two serogroups considered in this project. The two systems are similar other than in the ratio of immunity. Also note the significantly higher carriage rates overall compared to fig. 5

endemic to hyperendemic state works, but this should always be the case for a sound model and does not offer any explanation for the variation in itself. A possible explanation explicitly examined in this project is that behavioural and migratory patterns are solely responsible for the transition from endemic to hyperendemic state. The tests of this model suggest that these mechanics, on the scale examined, are not sufficient. With this in mind, it should still be noted that there are measurable effects on the system based on these seasonal variations. The explanation can not be dismissed entirely, especially based on two main points: 1) The increase in incidence rate seems robust to at least one other parameter (the invasiveness), which could suggest that the population density has an influence on a set size and 2) The case-carrier ratios are changed by these seasonal mechanics. The second point is relevant, as it is the argument used in [20] to suggest that only biological explanations affect the transition from endemic to hyperendemic. Furthermore, it should be noted that while this model is adept at describing the system, there seems to be a sort of steady state solution towards which the realisations tend. This suggests two things, firstly that epidemics require some external triggering factor, and secondly that a deterministic modelling procedure should be sufficient to model the endemic and the hyperendemic states.

In the initial modelling effort, the invasiveness (the risk of disease given infection) was estimated



(a) Graph showing the average result of the simulations using parametrical changes to model the seasonality of the disease, for a system matching serogroup A. There are some small seasonal variations visible in some populations, but in general the system looks visually similar to what is shown in fig. 5a. The relevant seasonal changes are in the number of disease cases, presented in table 4



(b) Graph showing the average result of the simulations using parametrical changes to model the seasonality of the disease, for a system matching serogroup W-135. Some seasonal variations can be seen in the graph, especially in the ratio of immune. Comparing to the result in fig. 5b, the presence of a visible population immunity is itself significant. This is a result of the higher rate of disease (see table 4) leading to more immune individuals.

Figure 8: Graphs showing the average distribution of carriage, disease and immunity within a population implementing seasonal variations through parametrical changes. The seasonal variations compared to fig. 7 are much smaller, which is reasonable as the largest difference should be in the number of sick which is not properly shown in this graph. Interestingly, the realisations of this particular model suggests there is some level of population immunity towards serogroup W-135 (right).

using case-carrier ratios. However, this proved inaccurate. This method provided consistently too high incidence rates of disease, and the case-carrier ratios measured in the realisations did not match what had been used as an initial value. Furthermore, as presented in table 4, the case-carrier ratios are not constant even as the invasiveness parameter is kept constant (as it is in the case of behavioural changes). The reason for the discrepancy between the case-carrier ratio and the invasiveness can be found in the discrepancy between the duration of carriage and the duration of disease, at least when carriage can not transition to disease. In all, this should be sufficient to show that case-carrier ratios are, while possibly useful, not in general a good estimate of the invasiveness of a disease. The usefulness depends on the properties of the disease itself.

One of the most central numbers in this model, and in population dynamics in general is the basic reproductive number. The value used in this model was, as mentioned, estimated specifically for the base population, trying to mimic the carrier rates measured in real world data. However, when the same number was used in a population of a different size, as in the behavioural model, the carriage rates increased significantly. This can be explained by the increased population density when the populations in the behavioural model meet, but it accentuates the fact that the basic reproductive number is population dependent. Estimating it for a different base population would likely lead to a different value, and it is possible that this should have been taken into

consideration when setting up the parameters for the behavioural model. In fact, it is possible that this would lead to a more appropriate increase in the incidence rates, as it would likely lead, at least, to a decrease in the endemic incidence rates.

Cross-immunity is an important factor in the human immune response to N. Meningitidis, but is not modelled in this project. Cross-immunity can be induced both by carriage of non-invasive strains and by carriage of for example *Neisseria Lactamica*. These mechanisms are not modelled in this project, and could potentially lead to an underestimate of the ratio of immune individuals, especially concerning the system with W-135.

In fact, only including one serogroup in the population at the time is not realistic. While there are results suggesting that one serogroup will be dominant at a given time, this is not sufficient to assume that other serogroups are just not present at all. Rather, it would be interesting to attempt to determine what factors lead to one serogroup being more viable than another. This could include factors like invasiveness, transmission rate, how sensitive the bacterium is to immunity and the average duration of carriage.

Including more serogroups and genetic variations of N. Meningitidis could also be used in an effort to mimic the epidemic dynamics, which unfortunately could not be done within the scope of this project. This would mean including more invasive strains and studying incidence rates as well as the prevalence of the different serogroups, and attempting to determine what setup would lead to an epidemic. This could also be used as a tool to evaluate possible explanatory models for the seasonality. As epidemics are known to occur only during the dry season, the explanation that is most accurate for the system should also be able to account for the occasional transition to epidemic state (under the correct circumstances). Examples could be if it is shown that epidemics can only occur when the population density is above a certain threshold (supporting the behavioural explanation), or that the increase in invasiveness that corresponds to an adequate increase in incidence rates is sufficient to have the system transition to an epidemic state.

Age-wise variations in the disease distribution are a well-established phenomenon for meningococcal disease, and it is possible that some important aspects in the disease dynamics can be explained by these mechanics. While the model in this paper seem to accurately show some of these variations, a significant amount of them are missing. Further efforts should strive to introduce events or other mechanics that more accurately model the variations by age. This includes efforts to include cross-immunity, as for example N. Lactamica affects primarily young children and should affect immunity in those (and older) age groups.

The age-dependent variations noted in this project are immune ratios for serogroup A and incidence rates in the endemic state for both serogroups. When discussing the variations in incidence rates, some care should be taken. The variations can be seen in the relative incidence rates, but these could be subject to error. When estimating the relative incidence rates, only the initial values of the population in each age group was considered, but these could be subject to change. In fact, when constructing the healthy population, it was assumed that only adults would die from non-disease related causes. While it was estimated that this would not lead to any significant changes in the population distribution over time, cannot be dismissed that it has lead to variations that could be large enough to affect the relative incidence rates.

One possible issue with the model as it is constructed is that individuals are non-specific. In cases where there are very small populations, for example concerning sick individuals, there is a risk that events occur in such a way that the duration of disease is unrealistically short. However, these

behaviours are quite unlikely, and for the events that are stochastic in nature, there could be a non-zero probability that such a series of events could in fact occur. A more significant issue would be in cases where reality would not behave stochastically, for example concerning ageing populations. In this particular project, the age groups are large enough that treating ageing as a stochastic event is allowed. These factors are, however, one of the reasons to not introduce an event allowing for ageing of sick individuals. It also accentuates the need to choose a model appropriate for the problem at hand.

When an immune individual is infected by the bacterium, they can not be subject to invasive disease, and they are immediately regarded as a carrier. However, this can in some sense lead to individuals losing immunity. This would depend on whether the parameter governing the rate at which carriers lose carriage and acquires immunity actually describes the rate at which carriage leads to immunity or whether it describes the rate at which people are immune after being carriers. These two may sound very similar, but they are not identical. The article used to estimate this parameter was intended to examine the rate at which carriage leads to immunity, and as such might not provide accurate estimates. This could lead to an under-estimate in the ratio of immunity in the population, which in turn could explain the slightly high incidence rates.

The largest issue in the implementation of this model are the uncertainties in the parameters used. The parameters needed to model a system such as this are in general quite difficult to measure in reality, partially due to their stochastic nature, partially due to the sheer magnitude of the system and partially due to lack of reliable testing methods. An example concerns measures of immunity, where the complexity of the human immune system becomes a factor. Cut-off points as to what entails immunity, what substances elicit an immune response and even uncertainty in the mechanisms governing the interaction between bacterium and human tissue all lead to difficulties in forming a reliable estimate. Furthermore, in the regions considered in this project, the infrastructure needed to perform comprehensive population testing is not even present. In all, a lot of the parameters used in this report are quite uncertain, and the results are of course affected by this. While the general behaviour of the system seem sound and realistic, it is important to consider the results more from a qualitative viewpoint than a quantitative. The mechanisms that are displayed in this system are interesting and should be considered when further examining the problem, but the numbers themselves are not what should be the focus.

5 Conclusions

Using a discrete and stochastic framework to model the dynamics of Meningococcal disease in the African Meningitis belt seems to work well, but there are difficulties in estimating the necessary parameters from real-world data. Furthermore, the endemic and hyperendemic states seem to behave in a way that would allow for a deterministic modelling method, which would allow for more efficient simulations.

Further modelling efforts should focus on the effects of including variations of the bacterium, as well as refining the framework for the age dependencies. Attempts at modelling epidemics should ideally include all of these factors, and great care should be taken when choosing whether to construct a stochastic or deterministic model. The results in this report do not make any statement on the validity of deterministic models for the epidemic system.

Case-carrier ratios are a useful tool to examine the general properties of a given disease, but in general they are not appropriate estimates of the risk of disease given invasion. Variations in case-carrier ratios depend on the population behaviour as well as on the biological properties of host and bacterium.

The transition from endemic to hyperendemic state does not seem to exclusively be the result of an increase in population density, but its effects cannot be dismissed. If the incidence rates (or even the epidemics) are facilitated by migration and crowding, infrastructure investments to decrease these types of displacements could be efficient in counteracting the disease.

A Complete list of events and the equations governing their rates

This appendix provides a complete list of the events used in the implementation of the model, together with the equations used to estimate the rate corresponding to each event. Parameters are marked by a hat, and a full list of parameters can be found in section B. $P_{\{\bullet\bullet(E)\}}$ are used for the number of individuals in a certain subpopulation. The first indexing letter gives the age group and can be A (adult), Y (young) or I (infant). The second letter gives the disease status and can be any of S (susceptible), C (carrier), I (ill), R (immune), or L (infected by the bacterium, but not yet ill nor a carrier). The E is used in cases where more than one base population is involved in the equation, and is then used to signify a subpopulation that belongs to a different base population than the one affected by the event.

Type of event	Affected populations		Equation
	Decreased population	Increased population	
Susceptible youth ageing	Young susceptible	Adult susceptible	$W = P_{YS} \cdot \hat{\Gamma}$
Susceptible infant ageing	Infant susceptible	Young susceptible	$W = P_{IS} \cdot \hat{\gamma}$
Immune youth ageing	Young immune	Adult immune	$W = P_{YR} \cdot \hat{\Gamma}$
Immune infant ageing	Infant immune	Young immune	$W = P_{IR} \cdot \hat{\gamma}$
Young carrier ageing	Young carriers	Adult carriers	$W = P_{YC} \cdot \hat{\Gamma}$
Infant carrier ageing	Infant carriers	Young carriers	$W = P_{IC} \cdot \hat{\gamma}$
Birth	-	Infant susceptible	$W = (P_{AS} + P_{AC} + P_{AR}) \cdot \hat{\eta}$
Death of adult carrier	Adult carriers	-	$W = P_{AC} \cdot \hat{\alpha}$
Death of adult immune	Adult immune	-	$W = P_{AR} \cdot \hat{\alpha}$
Death of adult susceptible	Adult susceptible	-	$W = P_{AS} \cdot \hat{\alpha}$
Death of sick adult	Adult sick	-	$W = P_{AI} \cdot \hat{\omega}$
Death of sick youth	Young sick	-	$W = P_{YI} \cdot \hat{\omega}$
Death of sick infant	Infant sick	-	$W = P_{II} \cdot \hat{\omega}$
Recovery of sick adult	Adult sick	Adult immune	$W = P_{AI} \cdot \hat{\rho}$
Recovery of sick youth	Young sick	Young immune	$W = P_{YI} \cdot \hat{\rho}$
Recovery of sick infant	Infant sick	Infant immune	$W = P_{II} \cdot \hat{\rho}$
Transmission to susceptible adult	Adult susceptible	Adult infected	$W = \left((P_{AI} + P_{YI} + P_{II}) \hat{\beta} + (P_{AC} + P_{YC} + P_{IC}) \hat{\varphi} \right) \cdot \frac{P_{AS}}{\sum_{i,j} P_{ij}}$

Transmission to susceptible young	Young susceptible	Young infected	$W = \left((P_{AI} + P_{YI} + P_{II}) \hat{\beta} + (P_{AC} + P_{YC} + P_{IC}) \hat{\varphi} \right) \cdot \frac{P_{YS}}{\sum_{i,j} P_{ij}}$
Transmission to susceptible infant	Infant susceptible	Infant infected	$W = \left((P_{AI} + P_{YI} + P_{II}) \hat{\beta} + (P_{AC} + P_{YC} + P_{IC}) \hat{\varphi} \right) \cdot \frac{P_{IS}}{\sum_{i,j} P_{ij}}$
Invasive disease of infected adult	Adult infected	Adult sick	$W = P_{AL} \hat{\vartheta}$
Invasive disease of infected youth	Young infected	Young sick	$W = P_{YL} \hat{\vartheta}$
Invasive disease of infected infant	Infant infected	Infant sick	$W = P_{IL} \hat{\vartheta}$
Infected adult becoming carrier	Adult infected	Adult carriers	$W = P_{AL} \hat{\Theta}$
Infected youth becoming carrier	Young infected	Young carriers	$W = P_{YL} \hat{\Theta}$
Infected infant becoming carrier	Infant infected	Infant carriers	$W = P_{IL} \hat{\Theta}$
Transmission to immune adult	Adult immune	Adult carriers	$W = \left((P_{AI} + P_{YI} + P_{II}) \hat{\beta} + (P_{AC} + P_{YC} + P_{IC}) \hat{\varphi} \right) \cdot \frac{P_{AR}}{\sum_{i,j} P_{ij}}$
Transmission to immune youth	Young immune	Young carriers	$W = \left((P_{AI} + P_{YI} + P_{II}) \hat{\beta} + (P_{AC} + P_{YC} + P_{IC}) \hat{\varphi} \right) \cdot \frac{P_{YR}}{\sum_{i,j} P_{ij}}$
Transmission to immune infant	Infant immune	Infant carriers	$W = \left((P_{AI} + P_{YI} + P_{II}) \hat{\beta} + (P_{AC} + P_{YC} + P_{IC}) \hat{\varphi} \right) \cdot \frac{P_{IR}}{\sum_{i,j} P_{ij}}$
Adult carrier recovering and gaining immunity	Adult carriers	Adult immune	$W = P_{AC} \hat{\psi}_A$
Young carrier recovering and gaining immunity	Young carriers	Young immune	$W = P_{YC} \hat{\psi}_Y$
Infant carrier recovering and gaining immunity	Infant carriers	Infant immune	$W = P_{IC} \hat{\psi}_I$
Adult carrier recovering and not gaining immunity	Adult carriers	Adult susceptible	$W = P_{AC} \hat{\Psi}_A$
Young carrier recovering and not gaining immunity	Young carriers	Young susceptible	$W = P_{YC} \hat{\Psi}_Y$

Infant carrier recovering and not gaining immunity	Infant carriers	Infant susceptible	$W = P_{IC}\hat{\Psi}_I$
Adult immune losing immunity	Adult immune	Adult susceptible	$W = P_{AR}\hat{\xi}_A$
Young immune losing immunity	Young immune	Young susceptible	$W = P_{YR}\hat{\xi}_Y$
Infant immune losing immunity	Infant immune	Infant susceptible	$W = P_{IR}\hat{\xi}_I$
Special events only used in certain simulations			
Transmission to susceptible adult from external population	Adult susceptible	Adult infected	$W = \left((P_{AIE} + P_{YIE} + P_{IIE})\hat{\beta} + (P_{ACE} + P_{YCE} + P_{ICE})\hat{\varphi} \right) \cdot \frac{P_{AS}}{\sum_{i,j} P_{ij} + \sum_{k,l} P_{klE}}$
Transmission to susceptible youth from external population	Young susceptible	Young infected	$W = \left((P_{AIE} + P_{YIE} + P_{IIE})\hat{\beta} + (P_{ACE} + P_{YCE} + P_{ICE})\hat{\varphi} \right) \cdot \frac{P_{YS}}{\sum_{i,j} P_{ij} + \sum_{k,l} P_{klE}}$
Transmission to susceptible infant from external population	Infant susceptible	Infant infected	$W = \left((P_{AIE} + P_{YIE} + P_{IIE})\hat{\beta} + (P_{ACE} + P_{YCE} + P_{ICE})\hat{\varphi} \right) \cdot \frac{P_{IS}}{\sum_{i,j} P_{ij} + \sum_{k,l} P_{klE}}$
Transmission to immune adult from external population	Adult immune	Adult carriers	$W = \left((P_{AIE} + P_{YIE} + P_{IIE})\hat{\beta} + (P_{ACE} + P_{YCE} + P_{ICE})\hat{\varphi} \right) \cdot \frac{P_{AR}}{\sum_{i,j} P_{ij} + \sum_{k,l} P_{klE}}$
Transmission to immune youth from external population	Young immune	Young carriers	$W = \left((P_{AIE} + P_{YIE} + P_{IIE})\hat{\beta} + (P_{ACE} + P_{YCE} + P_{ICE})\hat{\varphi} \right) \cdot \frac{P_{YR}}{\sum_{i,j} P_{ij} + \sum_{k,l} P_{klE}}$
Transmission to immune infant from external population	Infant immune	Infant carriers	$W = \left((P_{AIE} + P_{YIE} + P_{IIE})\hat{\beta} + (P_{ACE} + P_{YCE} + P_{ICE})\hat{\varphi} \right) \cdot \frac{P_{IR}}{\sum_{i,j} P_{ij} + \sum_{k,l} P_{klE}}$
Invasive disease of infected adult with seasonal parameters	Adult infected	Adult sick	$W = P_{AL}\hat{\vartheta}_{\text{season}}$
Invasive disease of infected youth with seasonal parameters	Young infected	Young sick	$W = P_{YL}\hat{\vartheta}_{\text{season}}$

Invasive disease of infected infant with seasonal parameters	Infant infected	Infant sick	$W = P_{IL}\hat{\vartheta}_{\text{season}}$
Infected adult becoming carrier, seasonal parameters	Adult infected	Adult carriers	$W = P_{AL}\hat{\Theta}_{\text{season}}$
Infected youth becoming carrier, seasonal parameters	Young infected	Young carriers	$W = P_{YL}\hat{\Theta}_{\text{season}}$
Infected infant becoming carrier, seasonal parameters	Infant infected	Infant carriers	$W = P_{IL}\hat{\Theta}_{\text{season}}$

If the model includes another strain of disease with different parameters, the relevant events are duplicated for the other strain. This includes transmission events and events relating to carriage, so that the population groups that carry different strains are kept separate.

B Parameters and equations to calculate them

This appendix contains a table with all parameters needed in order to calculate the rates given in section A. The values needed for the calculations are presented with their respective sources in table 2.

Parameter	Description	Derivation	Explanation
β	Rate of transmission from a sick individual.	$\beta = \frac{R_0}{7}$	About R_0 individuals in an otherwise healthy population should become infected during the expected duration of disease, 7 days.
φ	Average rate of transmission from a carrier	$\varphi = \frac{R_0}{130}$	About R_0 individuals in an otherwise healthy population should become infected during the expected duration of carriage, 130 days.
ϑ	Rate at which infected individuals become sick	$\vartheta = \frac{0.02}{10}$	About 2% of infected individuals become sick at some point during the incubation period, estimated as 10 days.

Θ	Rate at which infected individuals become carriers	$\Theta = \frac{1-0.02}{10}$	The 98% of infected individuals that do not become sick at some point during the 10 day incubation period become carriers in the same time.
ξ_{age}	Rate at which immune individuals loses immunity	$\xi_{\text{age}} = \frac{\ln D_{\text{age}}}{120}$	The solution to the equation $\exp\left(-\xi 120\right) = D_{\text{age}}$. Here, D_{age} is the ratio of the (age specific) population that retains their immunity after 120 days, with ξ giving the rate at which immunity is lost.
ψ_{age}	Rate at which a carrier loses carriage and becomes immune	$\psi_{\text{age}} = \lambda_{\text{age}} \frac{\ln 2}{90}$	λ_{age} is the portion of the (age specific) population to gain immunity after carriage. The rate at which carriage is lost is $\frac{\ln 2}{90}$, found by considering the half life of carriage, 90 days.
Ψ_{age}	Rate at which a carrier loses carriage without gaining immunity.	$\Psi_{\text{age}} = \left(1 - \lambda_{\text{age}}\right) \frac{\ln 2}{90}$	Compare with above. $1 - \lambda_{\text{age}}$ gives the portion of carriers that do not gain immunity, and the remaining gives the rate at which carriers recover.
γ	Rate at which infants become young	$\gamma = \frac{1}{1460}$	An individual is expected to leave the infant sub-population once, and they are expected to remain within it for four years (ages 0-4).

Γ	Rate at which young become adults.	$\Gamma = \frac{1}{3650}$	An individual is expected to leave the young sub-population once, and they are expected to remain within it for ten years (5-14).
α	Rate at which adults die.	$\alpha = \frac{1}{16425}$	An adult is expected to die (from non-disease related causes) once in their adult life ($\geq$ 15), and life expectancy is 16 years.
ω	Rate at which sick individuals die.	$\omega = \frac{0.1}{7}$	0.1, or 10% of sick individuals are expected to die. Disease lasts for about 1 week.
ρ	Rate at which sick individual recover.	$\rho = \frac{0.9}{7}$	90% of sick individuals are expected to recover, and disease lasts for about 1 week.
η	Rate at which a new individual is born, per adult	$\eta = \frac{41.1a}{0.58 \cdot 1000 \cdot 365}$	41.1 is the crude birth rate per 1000 individuals and year, and a is the survival rate of individuals surviving their first year. This is divided by the expected number of adults in those 1000 people ($0.58 \cdot 1000$) and the time period (365 days) to find the rate over time and per individual.

References

[1] N. Bharti, H. Broutin, R. Grais, M. Ferrari, A. Djibo, A. Tatem, and B. Grenfell. Spatial dynamics of meningococcal meningitis in niger: observed patterns in comparison with measles. *Epidemiology & Infection*, 140(8):1356–1365, 2012.

[2] I. Blakebrough, B. Greenwood, H. Whittle, A. Bradley, and H. Gilles. The epidemiology of infections due to neisseria meningitidis and neisseria lactamica in a northern nigerian community. *Journal of Infectious Diseases*, 146(5):626–637, 1982.

[3] R. Durrett. *Essentials of stochastic processes*, volume 1. Springer, 1999.

[4] W. Feller. On the integro-differential equations of purely discontinuous markoff processes. In *Selected Papers I*, pages 539–566. Springer, 2015.

[5] S. Findley. Does drought increase migration? a study of migration from rural mali during the 1983-1985 drought. *International Migration Review*, pages 539–553, 1994.

[6] C. for disease control and prevention. Meningococcal disease. https://www.cdc.gov/meningococcal/index.html, 2017. Accessed: 2017-08-09.

[7] Gapminder. Gapminder bubble chart. http://www.gapminder.org/tools/#_locale_id=en;&chart-type=bubbles, 2017. Accessed: 2017-08-10.

[8] D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[9] I. Goldschneider, E. Gotschlich, and M. Artenstein. Human immunity to the meningococcus. *Journal of Experimental Medicine*, 129(6):1307–1326, 1969.

[10] B. Greenwood. Priorities for research on meningococcal disease and the impact of serogroup a vaccination in the african meningitis belt. *Vaccine*, 31(11):1453–1457, 2013.

[11] T. Irving, K. Blyuss, C. Colijn, and C. Trotter. Modelling meningococcal meningitis in the african meningitis belt. *Epidemiology & Infection*, 140(5):897–905, 2012.

[12] A. Jakobsson. *An Introduction to Time Series Modeling*. Studentlitteratur, 2015.

[13] D. Kendall. An artificial realization of a simple "birth-and-death" process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):116–119, 1950.

[14] A. Kolmogoroff. Über die analytischen methoden in der wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104(1):415–458, 1931.

[15] T. Koutangni, H. Maïnassara, and J. Mueller. Incidence, carriage and case-carrier ratios for meningococcal meningitis in the african meningitis belt: a systematic review and meta-analysis. *PloS one*, 10(2):e0116725, 2015.

[16] L. Lapeyssonnie. La méningite cérébro-spinale en afrique, 1963.

[17] J. Leimkugel, A. Hodgson, A. Forgor, V. Pflüger, J. Dangy, T. Smith, M. Achtman, S. Gagneux, and G. Pluschke. Clonal waves of neisseria colonisation and disease in the african meningitis belt: eight-year longitudinal study in northern ghana. *PLoS medicine*, 4(3):e101, 2007.

[18] M. Maiden, A. Ibarz-Pavón, R. Urwin, S. Gray, N. Andrews, S. Clarke, A. Walker, M. Evans, J. Kroll, K. Neal, et al. Impact of meningococcal serogroup c conjugate vaccines on carriage and herd immunity. *The Journal of infectious diseases*, 197(5):737–743, 2008.

[19] A. Molesworth, M. Thomson, S. Connor, M. Cresswell, A. Morse, P. Shears, C. Hart, and L. Cuevas. Where is the meningitis belt? defining an area at risk of epidemic meningitis in africa. *Transactions of the royal society of tropical medicine and hygiene*, 96(3):242–249, 2002.

[20] J. Mueller and B. Gessner. A hypothetical explanatory model for meningococcal meningitis in the african meningitis belt. *International Journal of Infectious Diseases*, 14(7):e553–e559, 2010.

[21] J. Mueller, S. Yaro, Y. Traoré, L. Sangaré, Z. Tarnagda, B. Njanpop-Lafourcade, R. Borrow, and B. Gessner. Neisseria meningitidis serogroups a and w-135: carriage and immunity in burkina faso, 2003. *The Journal of infectious diseases*, 193(6):812–820, 2006.

[22] PATH. Meningitis belt in sub-saharan africa. http://www.globalhealthhub.org/wp-content/uploads/2010/11/extra-menigitisbelt-map2-300x215.gif, 2010. Accessed: 2017-08-10.

[23] D. Stephens, B. Greenwood, and P. Brandtzaeg. Epidemic meningitis, meningococcaemia, and neisseria meningitidis. *The Lancet*, 369(9580):2196–2210, 2007.

[24] M. Taha, A. Deghmane, A. Antignac, M. Zarantonelli, M. Larribe, and J. Alonso. The duality of virulence and transmissibility in neisseria meningitidis. *Trends in microbiology*, 10(8):376–382, 2002.

[25] G. Troncoso, S. Sanchez, M. Criado, and C. Ferreiros. Analysis of neisseria lactamica antigens putatively implicated in acquisition of natural immunity to neisseria meningitidis. *FEMS Immunology & Medical Microbiology*, 34(1):9–15, 2002.

[26] WHO. Meningococcal meningitis. http://www.who.int/mediacentre/factsheets/fs141/en/, 2017. Accessed: 2017-08-09.

[27] S. Yazdankhah and D. Caugant. Neisseria meningitidis: an overview of the carriage state. *Journal of medical microbiology*, 53(9):821–832, 2004.

Master's Theses in Mathematical Sciences 2017:E48
ISSN 1404-6342

LUTFNA-3041-2017

Numerical Analysis
Centre for Mathematical Sciences