

A stochastic model for Meningococcal Disease in the African Meningitis belt

Hanna Autio

June 2, 2017

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 made to a deterministic system governed by differential equations, and letting the populations vary continuously. By the law of large numbers and the central limit theorem, this is no issue 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 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.

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 understanding 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 policies, corresponding to several different policies 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 community themselves. For many diseases, including for example measles, there is a clear pattern of disease corresponding to population movements and population mixing[Spatial patterns of..]. 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 issues with distributing drugs and vaccines, meaning that any policy will be less efficient. This also increases the need for an early warning, as the time it takes to react to a given scenario increases. Furthermore, there can be issues considering the gathering of 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 project, 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.

Meningococcal disease are any diseases caused by *Neisseria Meningitidis*. The infection of *N. Meningitidis* can take many forms, of which two common (and severe) are meningococcal septicaemia and meningitis [source]. 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 [source: WHO]. 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, but possible consequences include amputation[source]. In rare cases, *N. Meningitidis* can cause arthritis and similar diseases[source].

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 [4]. 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 have an immune response that is putatively protective towards the disease [source: human immunity to the meningococcus]. Another bacterium frequently theorized as potentially leading to immunity to the disease is *Neisseria Lactamica*, often present in the nasopharynx of young children [source]. While immunity protects towards invasive disease, it has not been shown that there is any reduction of carriage caused by the immune response. [sources]

N. Meningitidis is genetically variable. It is most commonly classified based on the composition of its polysaccharide capsule 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, W135 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. [NM an overview..].

The African meningitis belt is a region in Sub-Saharan Africa with an increased rate of meningococcal disease as well as recurring epidemics. The belt was first described by Lapeyssonnie in 1968 [1]. This region is shown in ?? and has been characterized by semi-regular epidemics documented since the early 20th century. Its specific geographical area is defined by the WHO as the region in Africa between Senegal in the west and Ethiopia in the east [source], which coincides fairly well with the area described by Lapeyssonnie.

Meningitis dynamics in the region have, among others, these characteristics. 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. All of these properties are also affected by a strong seasonality, so that there are three distinguishable states for the disease in a population. There is the hypoendemic state, with a low incidence rate of about λ which occurs during the wet season. During the dry season, the system falls into a hyperendemic state with increased incidence rate, or occasionally an epidemic state. An epidemic state is defined as a rate of disease above λ . [3]

The reason for the specific dynamics in this region is often considered 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 [2]. 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* [sources on smoking, colds], 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 an area of priority by the WHO [source: research priorities].

The distribution of genetic variations of *N. Meningitidis* is different in the meningitis belt 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 population than in other regions. The temporal distribution of the bacterium is also unstable, leading to a clonal wave pattern, where one strain is dominant at a given time. [clonal waves]. 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.

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 [Epidemic meningitis...]. 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 group at most risk for disease are young children, 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, called polysaccharide vaccines. These vaccines 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. However, it is unclear whether this is caused by a decrease in carriage of the strain itself or only prevents the production of the capsule. [NM an overview]

Since 2010, a WHO 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 the specific strain, which could lead to herd immunity. These types of effects has been seen in other vaccination projects against *N. Meningitidis* [sources], 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 [recent] 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. The year-to-year variability will also be examined to a certain extent, as variations in the population distribution over time could lead to more long-term variability in risks for disease and epidemics.

This model is also the first, to my 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.

1.1 Main goals of this project

The primary goal of this project has been to create and implement stochastic model for the population and disease dynamics in the African Meningitis belt. The model should be evaluated, and its calculations should be compared to real-world data and other models, to determine whether there is an improvement in 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 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.

To further determine the seasonality of the disease, also the epidemic state should be considered. While epidemic outbreaks are connected to the introduction of a more invasive strain, the epidemics themselves also display a certain seasonality. Using the seasonal variations found to produce the most accurate results in the previous section, the results of introducing a more invasive strain during different seasons will be examined. If the introduction leads to an epidemic also during the wet season, or fails to lead to an epidemic in the dry season, the suggested explanation does not hold.

1.2 Terminology

2 Methods

The goal of this project is to develop methods to understand the dynamics characterizing meningococcal disease in the meningitis belt. The degree of urbanization is comparatively low in this region [source] and so any applicable method must be accurate also for smaller communities. As discussed in the introduction, traditional methods using deterministic dynamics and non-discrete populations become gradually more imprecise when dealing with decreasing population sizes [source], and so a different approach is used. We develop a dynamical model for integer-valued populations following a Markov-Jump process. The model is subsequently simulated

using the Feller-Kendall algorithm, and it's behaviour is validated by comparing to real-world data as well as to previous models (on an appropriate scale).

2.1 Model

We consider a population consisting of a discrete number of individuals. We are then interested in the distribution of, for example, disease in the population, and as such divide it into a set of n sub-populations based on disease status (and potentially other properties). This is done in such a way that no individual can be part of more than one subpopulation at a time. The state of the population can then be described by a vector containing the number of individuals in each subpopulation, $\mathbf{X} = [x_1 \dots x_n]$, where x_1 is the number of individuals in subpopulation one. The total number of individuals in the population is given by $\sum_{\eta=1}^n x_\eta$. We will use the vector $\mathbf{X} = \mathbf{X}(t)$ as the system state variable at time t . Note that this method does not allow for unique individuals in any way. Possible issues that may arise are discussed below.

Given the definition of the system state, a framework for dynamics like infection, transmission and recovery must be introduced. These dynamics are allowed to occur under an *event* framework, where each event corresponds to one specific type of dynamic. It is assumed that an event occurs quickly and can only increment or decrement a subpopulation by one individual. For example, the maturation of a child would be modelled as a instantaneous event that decreases the population of children by one and increases the population of adults by one, while the birth of a child would increment the population of children by one.

As mentioned in the introduction, it is intuitively clear that biological events such as, for example, contagion or infection are fundamentally random. This is modelled by taking each of these biological events and constructing a mathematical, stochastic event to match. These mathematical events affects the system state as described above, and their occurrence is assumed to follow a Poisson distribution. The rates governing each event is assumed to depend on the system state, and possibly time. Note that a rate that depends only on the system state will be constant between events, as the system state does not change if no event occurs.

This system state and the events as introduced form a Markov-Jump process, or alternatively a Markovian random walk in the n -dimensional population space. The model framework corresponds precisely to the one developed by Gillespie [source:a general method for numerically simulating] for chemical reactions. The algorithm constructed by Gillespie and Feller-Kendall then lets us construct realisations of the time-evolution of the system, allowing for a Monte-Carlo type set of simulations, giving an estimate of the time distribution of the system state.

The equation describing the probability density distribution of the system state over time is called the master equation, and is in general not analytically solvable. However, under the assumptions introduced above, the master equation follows the Kolmogorov-forward equation.

Depending on the problem one wishes to model, there can be events that might not obviously be considered random. Examples are aging, where the transition from a sub-population of individuals ≤ 4 years to a sub-population of ages 5 – 14 can not be considered random on the individual level. However, the individual level is not considered here. As such, we assume that the transition between the two groups above will occur about once per individual in the younger group every five years. This lets us formulate aging events as random, Poisson distributed events.

Another consequence of the lack of individual focus in this model is that it is in theory possible for very unlikely scenarios on the individual level to occur. For example, consider the case where there is no one current ill. Then an event occurs, incrementing the ill sub-population by one, which increases the rate of the recovery event to a non-zero value. Immediately after, the recovery event occurs. Here, we could have the case that a "specific" individual contracted the disease and recovered in an unreasonable short period of time. This is a potential scenario in this model, however not very likely.

Within the stochastic framework established, there is no definite way of implementing a minimum time of for example disease. There are however ways to emulate such a behaviour when constructing events and populations, leading to a more accurate model that incorporates periods of incubation and similar scenarios. A period of incubation is introduced to this model by constructing a new population of infected individuals who are neither carriers nor ill. This population increases when a susceptible individual contracts the bacterium, and the population decreases as individuals either become sick or become carriers.

A list of the events incorporated in the model is included in ?? together with the rates. A figure showing the flow of population between different groups of disease status is shown in ??.

2.1.1 Populations

As mentioned in the background, the disease and immunity show some dependencies of age, and so the population is divided into sub-populations based on age. The choice was made to divide the population into three age groups, infants (0-4 years), young (5-14 years), and adult (≥ 15 years). These age groups were chosen based on available data for the disease distribution and on population data.

In order to produce a model that can take into account the variations of the population over time, birth and death was also included into the model on a slightly simplified basis. It is assumed that only the oldest sub-population bears children, and it is also assumed that this group is the only one to die from non-disease related causes. This includes the infant mortality rate, and so the birth rate is calculated based on the number of children that survive the first year. These approximations leads to a small shift in the distribution of the age groups over time, but the shift is only shown after a time period greater than 10 years and there are no indications that this affects the results in any way.

2.1.2 Time dependency

As one of the main goals of this project is to examine the seasonality of the disease, some sort of time dependency must be included. So far, only probability rates that depend exclusively on the population has been considered. In this section, I will make some comments on probability rates that can vary with time.

We consider events for which the probability rate λ may depend on time.

To match the questions posed in the goals for this thesis, two types of time dependency was introduced and compared. They are intended to model hypothetical explanations for the seasonal variations in the disease pattern between dry and wet season, by changing one or more of the equations governing the event probability rates.

The dry and wet season are here modelled as discrete and deterministic periods. The dry season occurs between days 65 and 335 each year, lasting for about nine months. The rest is considered wet season.

As the modelling procedure requires the system state and probability rates to be constant between events, the changes brought by the seasonal shift is introduced immediately after the first event after the season change. The new probability rates are calculated based on the updated equations and the updated system state. Provided events are frequent enough, this method should not introduce significant errors. The uncertainty in the exact and discrete onset of the seasons should be a greater source of uncertainty and error.

The changes in the equations are different based on what type of time dependency is introduced. One is designed to model a change in the population behaviour, namely two separate groups coming in contact over the dry season. Practically, this is done by introducing events corresponding to the biological events introduced by this contact, i.e transmission from one population group to the other, and letting the probability rate from that event be zero during the wet season.

The second time dependent behaviour considered here is simply a change in the invasiveness of the disease itself, and is modelled by a change in one of the internal parameters.

2.2 Acquiring data and parameters

The implementation discussed in this project is designed to approximately match a theoretical population in northern Nigeria. The region was chosen as there is reliable data.

2.2.1 Population and climate data

The only climate data currently used is when the wet season occurs, which in this region is approximately from July though 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.

The distribution of the population was sourced from Gapminder, and the population growth resulting from a model disregarding the disease corresponds well to a 2% population growth, which is the current value in Nigeria.

2.2.2 Disease data

2.3 Implementation

The model was implemented in C++, and could be easily adapted to a different problem.

3 Results and analysis

4 Conclusions

References

- [1] Léon Lapeyssonnie. La méningite cérébro-spinale en afrique, 1963.
- [2] Anna M Molesworth, Madeleine C Thomson, Stephen J Connor, Mark P Cresswell, Andrew P Morse, Paul Shears, C Anthony Hart, and Luis E 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.
- [3] Judith E Mueller and Bradford D Gessner. A hypothetical explanatory model for meningococcal meningitis in the african meningitis belt. *International Journal of Infectious Diseases*, 14(7):e553–e559, 2010.
- [4] Muhamed-Kheir Taha, Ala-Eddine Deghmane, Aude Antignac, Maria Leticia Zarantonelli, Mireille Larribe, and Jean-Michel Alonso. The duality of virulence and transmissibility in neisseria meningitidis. *Trends in microbiology*, 10(8):376–382, 2002.