NTNU

Norwegian University of
Science and Technology

Epidemic Network Modeling for the Prediction of the Transmission of Gonorrhea in Norway

Sofie Grefsrud Solnes

Chemical Engineering and Biotechnology
Submission date: June 2018
Supervisor: Eivind Almaas, IBT
Co-supervisor: Christian Schulz, IBT

Norwegian University of Science and Technology
Department of Biotechnology and Food Science

Disclaimer: I don't have gonorrhea.

Acknowledgements

I would first like to thank my advisor Eivind Almaas. Thank you for always being so encouraging and positive, motivating me to work harder. I am glad that I have learned a lot from being in your systems biology group, and it has been super interesting to learn about epidemic network modeling. I would also very much like to thank my co-advisor Christian Schulz. Thank you for always being there when I needed to discuss over and over, and for always answering all my panicky messages. I am very grateful for the support and guidance you both have given me.

I would also like to thank Martina Hall for your help with ANOVA and making me understanding the analyses, and Emil Karlsen for your proofreading-skills and constructive comments until the very end.

Lastly I would like to thank my friends Elida and Karoline for all your love and support and for bearing with me these six months. I could not have wished for better friends!

Trondheim, June 2018

Sofie Grefsrud Solnes

Summary

The number of gonorrhea infections among men who have sex with men (MSM) in Norway continues to rise, and in 2017 the number of infections increased with more than 50 % relative to 2016. This has resulted in various campaigns to encourage people to always use prophylactics and to get tested more frequently. The use of stochastic epidemic modeling on networks to understand and predict present and future epidemics is increasingly important. The world is becoming smaller and more connected with easy and cheap travel, thus making infections spread faster on the global scale. This further exacerbated by the emerging global issue of antimicrobial resistance (AMR), spurred by the mis- and overuse of antibiotics. The growing AMR problem is also increasing the risk of prevention and treatment of infections, medical procedures and cancer treatment and increases costs of health care. The use of network epidemic models for the transmission of infections is getting more advanced and useful as spatial and geolocation data now is the norm.

The aim of this thesis was to create epidemic network models with specific parameters that were able to capture the temporal dynamics of the spread of gonorrhea in Norway in a sequence of years, using reports and statistics from Folkehelseinstituttet (FHI). Altogether the following ten SIS/SIR/SIERS epidemic network models were built: Four models using FHI reports, one model to predict the gonorrhea spread in 2018, two for the initial spread only seeded from Oslo and three models for the AMR transmission.

Of the ten models built, the four SIS epidemic network models using FHI reports and the 2018 model proved to yield the most interesting results, by capturing some key characteristics of the transmission of gonorrhea. With parameter combinations of the recovery rate (α) and the transmission rate (β) in the models, the total cumulative number of infected MSM in Norway was retrieved with an average discrepancy of 3.7 % from the reports for the respective years. The parameters were constant for all the models, with the exception of the probability of exposure to infection ($\beta_{2,G5}$). The parameter $\beta_{2,G5}$ was increased with an average of 1.2 % every year. The 2018 model was generated using the increase of $\beta_{2,G5}$ and predicted an increase of 14.7 % in the number of gonorrhea cases among MSM in Norway compared to 2017.

Sammendrag

Antall gonorésmittede menn som har sex med menn (MSM) har økt de siste årene i følge tall fra Folkehelseinstituttet. Økningen har vært på over 50 % det siste året. Dette har resultert i ulike kampanjer fra blant annet Helsedirektoratet hvor oppmerksomheten rettes mot viktigheten av å bruke kondom og å sjekke seg for kjønnssykdommer. Bruken av stokastiske epidemiologiske nettverksmodeller for å forstå og forutse fremtidige epidemier blir bare mer og mer viktig. I dagens samfunn er det en økende trend i reising fordi det er enkelt og billig å fly hvor som helst. På denne måten vil også sykdommer raskere kunne spre seg globalt med menneskene som reier. Den stadig forverrede situasjonen med antibiotikaresistente bakterier øker også globalt, og forekommer av overforbruk og feilforbruk av antibiotika. Den økende forekomsten av antibiotikaresistente bakterier fører også til komplikasjoner når det kommer til forebygging og behandling av infeksjoner, medisinske inngrep og kreftbehandling. Dette øker også kostnadene knyttet til helsetjenester. Med stadig mer avanserte og tilgjengelige epidemiologiske nettverksmodeller, kan disse brukes til å forstå hvordan sykdommer spres og hvordan forebygge økningen av sykdommer.

Målet med denne oppgaven var å lage en epidemiologisk modell og å se på effekten av spredningen av gonoré på et nettverk. Ved å bruke ulike parametre som reflekterte spredningsdynamikken i gonoré i Norge, ble ti ulike modeller laget. De ti epidemiologiske nettverksmodellene besto av syv SIS-modeller og tre SIR-/SIERS-modeller. De syv SIS-modellene skulle forutsi spredningsbildet i Norge fra 2014 til 2018 basert på data fra FHI, og to av dem så på spredningen av gonoré kun i Oslo for årene 2016 og 2017. Disse modellene skulle forutsi hvor mange MSM som ble smittet på et år, som var den maksimale simuleringsperioden i GLEaM. Modellene var basert på statistikk fra FHI og MSIS.

Av de ti modellene som ble skapt i denne oppgaven, var det de fire SIS-modellene med utgangspukt i FHI-rapportene som ga de mest interessante resultatene. Med kombinasjoner av spredningsraten (β) og forbedringsraten (α), simulerte modellene et kumulativt antall smittede MSM med et avvik på kun 3.7 % sammenlignet med rapportene fra FHI. Forbedringsraten var alltid konstant på 23 dager, og spredningsraten var konstant i de ulike aldersgruppene, med gruppe 5 som unntak. Sannsynligheten for å være seksuelt aktiv, eller sannsynligheten for å være eksponert for sykdommen $\beta_{2,G5}$ ble i gjennomsnitt økt med 1.2 % i de fire modellene som simulerte årene 2014-2017. Med denne justeringen av $\beta_{2,G5}$, genererte 2018-modellen 1039 smittede MSM, noe som gir en økning på 14.7 % blant MSM i Norge sammnelignet med 2017.

Table of Contents

List of Tables

List of Figures

Abbreviations and Symbols

α	=	Recovery rate
AMR	=	Antimicrobial Resistance
ANOVA	=	Analysis of variance
β	=	Transmission rate
Df	=	Degrees of freedom
E	=	Exposed
ECDC	=	European Centre for Disease Prevention and Control
FHI	=	Folkehelseinstituttet
G1	=	Age Group 1 (0-14 years)
G2	=	Age Group 2 (15-24 years)
G3	=	Age Group 3 (25-34 years)
G4	=	Age Group 4 (35-44 years)
G5	=	Age Group 5 (45+ years)
GLEaM	=	Global Epidemic and Mobility model
GUI	=	Graphical User Interface
H1N1	=	Hemagglutinin 1 Neurominidase 1
I	=	Infectious
IATA	=	International Air Transport Association
MSM	=	Men who have sex with men
MSW	=	Men who have sex with women
N. gonorrhoea	=	*Neisseria gonorrhoea*
NORM	=	Norwegian Surveillance System for Drug Resistance
OAG	=	Official Airline Guide
R	=	Recovered
R_0	=	Basic reproduction number
S	=	Susceptible
SARS	=	Severe Acute Respiratory Syndrome
SEDAC	=	Socioeconomic Data and Application Center
SI	=	Susceptible Infectious
SIR	=	Susceptible Infectious Recovered
SIS	=	Susceptible Infectious Susceptible
STD	=	Sexually transmitted disease
STI	=	Sexually transmitted infection
τ	=	Characteristic time
WHO	=	World Health Organization
WSM	=	Women who have sex with men
WSW	=	Women who have sex with women

Chapter 1

Introduction

Sexually transmitted infections (STI) are a globally growing problem in the public health sector today [11]. The enormous burden of morbidity and mortality due to STIs, impact directly on a patient's quality of life, sexual and reproductive health and child health [12]. Certain STIs such as syphilis, chancroid ulcers and genital herpes simplex virus ulcers greatly increase the risk of acquiring or transmitting human immunodeficiency virus (HIV) and in some cases account for upwards of 40 % HIV transmissions [12]. This indirect impact on the spread of HIV, as HIV often infects regions with activated immune cells where the body has been infected with an STI [13, 14]. STIs also impact on national and individual economies through working life years lost, and the treatment cost of these infections [11]. According to the World Health Organization (WHO), millions of individuals acquire an STI every day [11]. In 2012 the number of new curable STI cases worldwide was estimated to be 357 million, and this number includes 78 million cases of gonorrhea.

Gonorrhea is an exclusively human pathogen, is caused by the gram-negative bacteria *Neisseria Gonorrhoeae* and can affect all individuals who have unprotected sex with an infectious individual. Gonorrhoea is the second most common STI after chlamydia and is treated with antibiotics such as cephalosporins, quinolones or azithromycin [15, 11]. The gonorrheal infection is often asymptomatic. For cases of gonorrhea infections reported by Folkehelseinstituttet (FHI) in Norway from 2015 to 2017, only 44 % of men who have sex with men (MSM) reported a symptomatic infection [16]. In comparison, 50 % and 89 % of heterosexual women and men reported symptoms, respectively. The difference in the presentation of symptoms between MSM and heterosexual men is due to the location of infection. Rectal and oral infections are largely asymptomatic resulting in delays in seeking treatment, the increased risk of serious health problems such as pneumonia and not to mention the prolonged risk of transmitting the infection [17, 18, 19, 20]. For women gonorrheal infections can increase the risk of ectopic pregnancies, for both men and women untreated gonorrheal infections can lead infertility [17, 18, 21, 22]. Various symptoms as a consequence of the gonococcal infection are illustrated in Fig. 2.8.

Since the introduction of antibiotics in the 1930s and 1940s resistance has rapidly emerged [23]. Due to the rise in antimicrobial resistance (AMR) in strains of *N. gonorrhoeae*, it becomes increasingly difficult to treat the infection [24]. The WHO, supported by the United Nations' World Health Assembly, has set major goals on the global health sector. One of the most extensive targets is to reduce the occurrence of gonorrhea with 90 % by 2030 using 2018 as a baseline, due to the rising risk of untreatable infections, and co-infections with chlamydia and HIV [24, 25].

In 2017 the incidence of gonorrhea among MSM in Norway increased by 50 %, and with these high rates the infection is expected to increase in the next years [10]. This increase is also seen in other western countries, and gonorrhea among MSM has become a major public health concern [26]. The increases in reported STI and AMR cases could reflect the increased prevalence of behaviours that transmit the infections, and could be reduced with the use of condoms and a higher frequency of testing at clinics [27]. Another remark is the low incidence in symptomatic cases and lower partner notification among MSM compared to heterosexual individuals. These observations call for an intervention [27]. How does the infection spread in a population? Can the spreading patterns be understood to know when and where to intervene? How can the increasing occurrence of antimicrobial resistant *N. gonorrhoeae* be slowed down, or even stopped?

To further investigate the possible impact, a computational analysis using the GLEAM (Global Epidemic and Mobility) framework will be used to create a SIS epidemic model (susceptible - infectious - susceptible) to use for computational simulations of the spread of gonorrhea. A network perspective is used on the population to find out how to manage the spread of the infection itself and the spread of AMR in the best possible way. This approach belongs to the discipline of systems biology [28, 29]. Systems medicine originate in systems biology and includes components of health and disease. These systems are dynamic and very complex, and the study of their behaviour could state something about interacting components. Examples include genomics, bioinformatics or computational models used to predict the behaviour of a system [30, 31, 32].

The aim of this thesis is to computationally study the transmission of gonorrhea among MSM in Norway and the transmission of AMR *N. gonorrhoeae* using epidemic network models. The computational tool GLEAMviz will be used to predict the spread of the infection in real-time. The focus of the model simulations will be on Norway in order to identify parameter combinations with the potential to forecast near-future spreading patterns. The models will depend on parameters such as the transmission rate (β), the recovery rate (α) and the basic reproduction number (R_0). The model parameters and inputs will be based on gonorrhea reports from FHI, The Norwegian Surveillance System for Communicable Diseases (MSIS) and The European Centre for Disease Prevention and Control (ECDC).

Chapter 2

Theory

In the following sections epidemic spreading on networks and three computational disease network models will be presented. The epidemic models are the susceptible-infectious (SI) model, the susceptible-infectious-susceptible (SIS) model and the susceptible-infectious-recovered (SIR) model. Understanding these models may help understanding how epidemics spread on networks. The sections 2.2, 2.2.1, 2.2.2 and the 2.2.3 are based on the previous project "Network Analysis and Modeling of the Transmission of Pneumonic Plague from *Yersinia pestis*."

2.1 Epidemic Spreading on Networks

The use of networks has proven to be useful to understand the complexity of different systems [33]. Examples include the world wide web, the brain and biological systems. Yet the realization of how important it was to use a network perspective to understand the complex disease dynamics did not entirely come before the HIV/AIDS outbreak [34]. This has led to the development of mathematical models that can express universal network properties [33].

With good network data explaining the network topologies, information obtained from models could be tried out against real world networks. Understanding the behaviour of different epidemic models on the networks could prove very significant in understanding how an infections spreads across a network, and how the network influences the epidemic's dynamics. Since spreading is a ubiquitous process the knowledge of different spreading patterns and the system of social interactions is very important in realizing how to for example stop the spreading of an infection [35].

Many real world networks have a community structure [36, 33]. Some stylized random graph models with a community structure reveled that communities have an impact on how

an infection spreads across a network. However, this is to a great extent unexplained and it is of large interest to understand how this affects the network and the epidemic models and the spreading of infections. In GLEaM, the long range mobility layer, the airline transportation network, is the most developed network in the model. Large-scale networks explores the ubiquitous presence of the contact patterns and connectivities in large scale inhomogeneties [37]. GLEaM considers the complete world wide airline transportation infrastructure accompanied by data on census data for all populations. In the modern fast living world with cheap travels, it is important to consider this network as well as the commuting network and local contact patterns within census populations.

Understanding the structures of sexual networks is crucial for understanding the STI's spreading dynamics [34]. Standard models, like the GLEaMviz modeling tool, neglect the complex contact patterns, like intimate contacts. Having the possibility to implement social network analysis, the modeling of social interactions would provide a great potential of understanding the disease epidemics of STIs.

An example of the importance of understanding the modeling of spread on complex networks is the SARS (Severe Acute Respiratory Syndrome) epidemic outbreak in 2013. The aim with the stochastic model was to include actual travel and census data all over the world to be able to predict the spreading pattern of the epidemic [38]. The results from this modeling were in agreement with the empirical data of the SARS worldwide epidemic [38]. From this modeling it was learned that effective communication of insights would help to ensure that policymakers would involve modelers to in outbreaks of emerging epidemics [39].

Another example from utilizing the GLEaM model to understand how infections spread on networks is the 2009 H1N1 Pandemic. The influenza outbreak also known as the swine flu, was modeled on a complex network model with specific parameters that fit the disease [40]. The predictions made by GLEaM were compared with data from 48 countries. The findings were about the peak time of the pandemic and the impact of mass vaccination. The peak time is the week where the most individuals are infected in a chosen country during the outbreak [40]. The results showed an early peak compared to other influenza outbreaks in November and not January or February, and was important information for health officials when deciding the timing and quantity of the distribution of vaccines. This prediction from GLEaM turned out to be correct, confirming the predictive power of this modeling tool.

Understanding the transmission patterns of a sexually transmitted disease (STD) like HIV has been important in order to understand where and how to intervene to reduce risk, especially among MSM [41, 42, 43]. Increased attention has been given to the role that MSM networks play in HIV epidemiology [41]. This study on MSM networks revealed that members in the same social network often share similar norms, attitudes, and HIV risk behavior levels, and that network interventions are feasible and powerful for reducing unprotected sex and potentially for increasing HIV testing uptake among other findings [41]. This shows the importance of modeling and understanding how an STD or STI spreads on a network.

2.2 Epidemic Modeling

The framework used to model the spread of pathogens in epidemiology is divided into two fundamental hypotheses: The compartmentalization and the homogeneous mixing of a population [5]. In the following sections, the presentation of epidemic modeling will follow that of the book *Network Science* by Barabási [5].

In epidemic modeling all individuals are compartmentalized, meaning that they are classified and belong to a specific stage in the disease [5]. For the simplest models, there are three compartments or states: susceptible (S), infectious (I) and removed (R). The susceptible compartment includes individuals of the population that are healthy and have not been infected with the pathogen yet, but are receptive to it. The infectious individuals have caught the infection and can spread the disease to the susceptible in the population. Removed individuals have either recovered from the disease or died from it, and they are no longer infectious nor susceptible. For some diseases, additional compartments are needed. Examples are compartments for immune individuals who cannot be infected or a compartment for the latent individuals of a disease, who have been exposed for the pathogen but are not yet infectious. During the course of an outbreak, for example with an influenza virus, individuals in the population move between the compartments. Initially, everyone in the population is in the susceptible compartment. As individuals get infected they move to the infectious compartment. All infected individuals can infected susceptible individuals. When individuals recover, they are moved to the removed compartment and are in this case immune. Hence these individuals cannot spread the disease nor get infected again.

The homogeneous mixing hypothesis explains that each individual of a population has the same chance of encountering an infected individual. This means that in epidemic modeling it is not required to know the precise contact network on which the disease spreads [5]. In GLEaM the homogeneous mixing hypothesis is assumed for all the subpopulations. A subpopulation could be core groups that the population is divided into. For example based on gender, age or sexual orientation. This means that an expected number of infectious individuals I in a populations with S susceptible individuals will be $\beta\frac{SI}{N}$, where β is the transmission rate and N is the total number of individuals in the subpopulation [7].

2.2.1 The Susceptible-Infected Model

The simplest epidemic model built on the two hypotheses explained in Section 2.2 is the so called SI model, see Figure 2.1 [5]. In this model a disease is spread in a population of N individuals. $S(t)$ is the number of the susceptible individuals at time t, and $I(t)$ is the number of individuals infected at time t. The unit time, t is set to one day. Before the outbreak of a disease, the whole population is susceptible and no one is infected: $S(t_0) = N$ and $I(t_0) = 0$. Each individual in the population has $\langle k \rangle$ encounters with other individuals every day, and the probability of transmitting the disease from an infected to a susceptible in one day is β. For the disease to spread in the population, the number of infected individuals is set to one, $I(t_0) = 1$. According to the homogeneous mixing hypothesis, the chance of encountering a susceptible individual is $S(t)/N$, thus, an infected individual

will encounter $\langle k \rangle S(t)/N$ susceptible individuals every day [5]. The average number of newly infected individuals $I(t)$ in the time frame dt becomes

$$\frac{dI}{dt} = \beta \langle k \rangle \frac{S(t)I(t)}{N}, \tag{2.1}$$

when $I(t)$ equals the number of infected individuals transmitting the pathogen at rate β. For simplicity, $S(t)/N$ will be expressed as $s(t)$ and $I(t)/N$ as $i(t)$. Furthermore the (t) will be dropped in the following text, so $s(t)$ and $i(t)$ will be written as s and i.

$$s = s(t) = \frac{S(t)}{N}, \quad i = i(t) = \frac{I(t)}{N} \tag{2.2}$$

The product of $\beta \langle k \rangle$ is called the transmission rate. Rewriting Eq. (2.1) gives

$$\frac{di}{dt} = \beta \langle k \rangle si = \beta \langle k \rangle i(1 - i) \tag{2.3}$$

Further, solving for i and then integrating on both sides gives

$$\frac{di}{i} + \frac{di}{1 - i} = \beta \langle k \rangle dt \tag{2.4}$$

and Eq. (2.5) is obtained:

$$\ln i - \ln(1 - i) + C = \beta \langle k \rangle t \tag{2.5}$$

The initial condition $i_0 = i(t_0)$ gives $C = i_0/(1 - i_0)$, where C is the arbitrary constant of integration. Rewriting Eq. (2.5) the fraction of infected individuals will be given as Eq. (2.6):

$$i = \frac{i_0 e^{\beta \langle k \rangle t}}{1 - i_0 + i_0 e^{\beta \langle k \rangle t}} \tag{2.6}$$

When i from Eq. (2.6) is plotted as a function of time it is evident that the fraction of infected individuals is increasing exponentially, as shown in Fig. 2.2. This is because the fraction of susceptible individuals is much larger than that for the infected individuals in the beginning of the spreading of the disease, hence the probability for encountering a healthy person, is large. With time, the growth of $i(t)$ slows and as the epidemic ends, it is found that $i(t\infty) = 1$ and $s(t = \infty) = 0$.

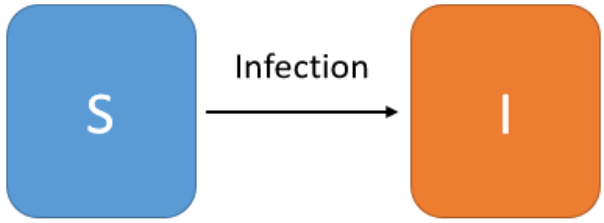


Figure 2.1: The SI model. Susceptible individuals in a population can get infected after being in contact with infectious individuals. When $t \to \infty$ all individuals will end up in the infected compartment.

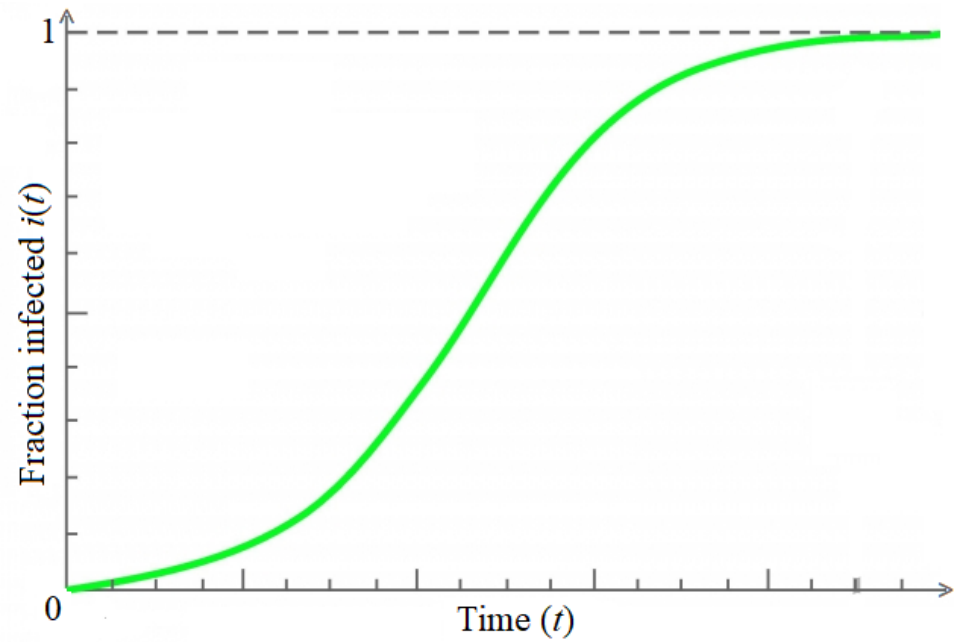


Figure 2.2: In this graph, the fraction of infected individuals as a function of time. In the SI model the individuals in a population are either in the susceptible or infected state. The susceptible individuals become infected in a rate $\beta\langle k\rangle$ if contact with an infected individual is made. Infected individuals can not recover, thus the fraction of infected individuals grows exponentially at early times. As all individuals in the population get infected, $i(\infty) = 1$. Retrieved from [5].

2.2.2 The Susceptible-Infected-Susceptible Model

In most cases, the pathogen infecting the body is defeated either by the immune system or cured by treatment and the patient recovers and can no longer spread the disease. In the SIS model, Figure 2.3, individuals recover at a rate α becoming susceptible again. The dynamics of the model is described by Eq. (2.7).

$$\frac{di}{dt} = \beta\langle k\rangle i(1-i) - \alpha i \tag{2.7}$$

The term αi explains the rate at which the individuals recover from the disease. The solution of Eq. (2.7) gives the fraction of infected individuals as a function of time,

$$i(t) = \Big(1 - \frac{\alpha}{\beta\langle k\rangle}\Big)\frac{Ce^{(\beta\langle k\rangle-\alpha)t}}{1+Ce^{(\beta\langle k\rangle-\alpha)t}}, \tag{2.8}$$

where the initial conditions for $i_0 = i(t=0)$ provides $C = i_0/(1-i_0-\alpha/(\beta\langle k\rangle))$. C is the arbitrary constant of integration. This is shown in Fig. 2.4.

The difference from the SIS model in Figure 2.4 compared to the SI model in Figure 2.2 on the preceding page, is that not everyone in the population gets infected at $t\to\infty$. The model provides two possible outcomes for the epidemic. The first possible outcome is the endemic state where $\alpha < \beta\langle k\rangle$ where i will follow a curve similar to the SI model curve [5]. This time not everyone gets infected, but i reaches a constant value of $i(\infty) < 1$. In the endemic state the number of newly infected individuals equals the number of individuals who recover from the disease. By setting di/dt to zero, the following expression for $i(\infty)$ is obtained:

$$i(\infty) = 1 - \frac{\alpha}{\beta\langle k\rangle} \tag{2.9}$$

The other possible outcome is the disease-free state where $\alpha > \beta\langle k\rangle$. When the recovery rate is sufficiently high, the exponent in Eq. (2.8) becomes negative, thus i decreases exponentially with time and the disease will die out due to the fact that the number of newly infected individuals will be inferior to the number of individuals cured per unit time. Hence, the pathogen will disappear from the population. The characteristic time, τ of a pathogen is expressed in Eq. (2.11) and Eq. (2.10) as

$$\tau = \frac{1}{\beta\langle k\rangle}, \tag{2.10}$$

or as

$$\tau = \frac{1}{\alpha(R_0-1)}. \tag{2.11}$$

The characteristic time explains how fast a process is, in Eq. (2.10) it says how fast the system will reach a fraction of $1/e$ (36 %) of the susceptible individuals. In Eq. (2.11) the expression is extended and R_0, the basic reproduction number is used. The expression for R_0 is

$$R_0 = \frac{\beta\langle k\rangle}{\alpha} \tag{2.12}$$

or

$$R_0 = c\beta D. \tag{2.13}$$

Where c is the number of potentially infectious contacts per individual, D is the mean duration (α) and β is the transmission rate [34]. The basic reproduction number R_0 is the average number of secondary infection cases expected from one infectious individual, and is the most common measurement of the transmission of infectious diseases in a population [44]. This is given when the whole population is set to be susceptible [45]. So "R_0 is the number of new infections each infected individual causes under ideal circumstances" [5]. In the endemic state, R_0 exceeds unity and τ is positive. If all the infected individuals infect more than one susceptible individual, the pathogen can spread and persist in the population. A high value for R_0 indicates a faster spreading process. In the disease-free state, where $R_0 < 1, \tau$ is negative and the epidemic dies out [46]. Hence, none of the infected individuals manage to spread the disease to more than one individual. R_0 is important and is usually the first parameter epidemiologists estimate during an outbreak of a new pathogen. An example of a reproductive number is for measles, an airborne and extremely contagious disease, where R_0 is about 12. This means that one infected individual on average infects twelve healthy individuals.

In modeling STIs the definition of R_0 could be extended to

$$R_0 = \rho_0(1 + \frac{\sigma^2}{\mu}) \tag{2.14}$$

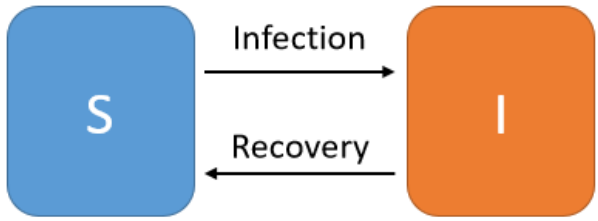


Figure 2.3: The SIS model. Susceptible individuals in a population can get infected after being in contact with infectious individuals from the infectious compartment. Infectious individuals can recover and move back to the susceptible compartment.

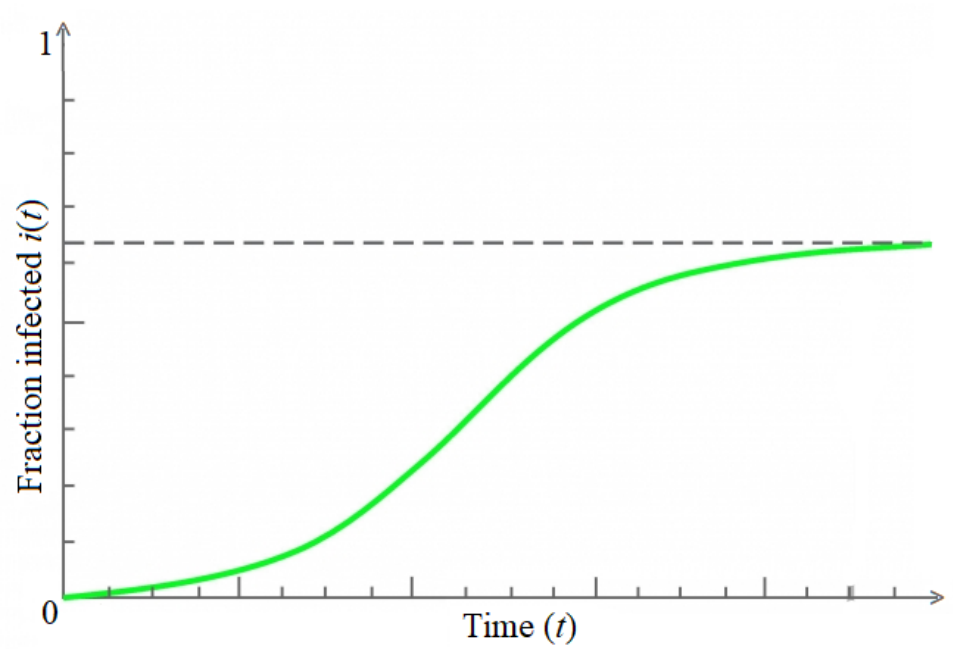


Figure 2.4: In the SIS model, the fraction of infected individuals is plotted against time. Since it is possible to recover, at large times, the system reaches an endemic state where the fraction of infectious individuals is constant $i = (t = \infty)$. If the recovery rate $\alpha > \beta$, the disease dies out because the number of infected individuals will decrease exponentially. This model suits the modeling of for example STIs or the common cold. Based on [5].

Where ρ_0 is is the average number of infections produced by one infected individual in a susceptible population, σ^2 is the variance in number of contacts per individual and μ is the average number of contacts in the subpopulation [34].

2.2.3 The Susceptible-Infected-Recovered Model

In many cases, individuals develop immunity for several types of influenza viruses or other pathogens. This means, that the individuals are no longer moved back into the susceptible compartment after recovery, but that they are moved into a new one: The recovered compartment, see Figure 2.5. The individuals contained in this compartment are removed from the population and cannot infect others or be infected. The properties of the SIR model is described in equations Eqs. (2.15), (2.16) and (2.17) for the susceptible, infected and recovered compartments, respectively. Here, β is the transmission rate and α is the recovery rate. The characteristics to the model are also shown in Fig. 2.6.

$$\frac{ds}{dt} = -\beta\langle k\rangle i[1 - r - i] \tag{2.15}$$

$$\frac{di}{dt} = -\alpha i + \beta\langle k\rangle i[1 - r - i] \tag{2.16}$$

$$\frac{dr}{dt} = \alpha i \tag{2.17}$$

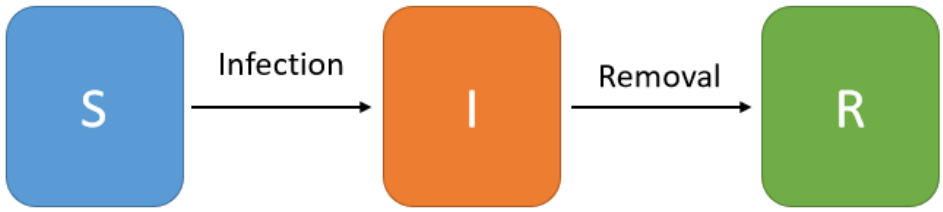


Figure 2.5: The SIR model. Susceptible individuals in a population can get infected after being in contact with infectious individuals. After a given recovery period, individuals recover and are transferred to the removed compartment. These individuals can not get re-infected, thus the whole population will be recovered as $t \to \infty$.

Fig. 2.7 shows the relationship between the three models explained in Sections 2.2.1, 2.2.2 and 2.2.3: the SI, SIS and the SIR model, and compares their characteristics. The start phase for the outbreak for all the three models are similar. This is because at the beginning of the epidemic a small amount of the population is infected, the rest is susceptible towards the disease. Thus the pathogen will spread quickly through the population, yielding an exponential curve. As the epidemic evolves in the population, the outcome differs. In the SI model the whole population gets infected. In the SIS model the population either reaches an endemic state where a finite number of the population is infected, or the disease reaches a disease-free state where the infection dies out. In the SIR model there will not be any infected individuals at the end of the outbreak.

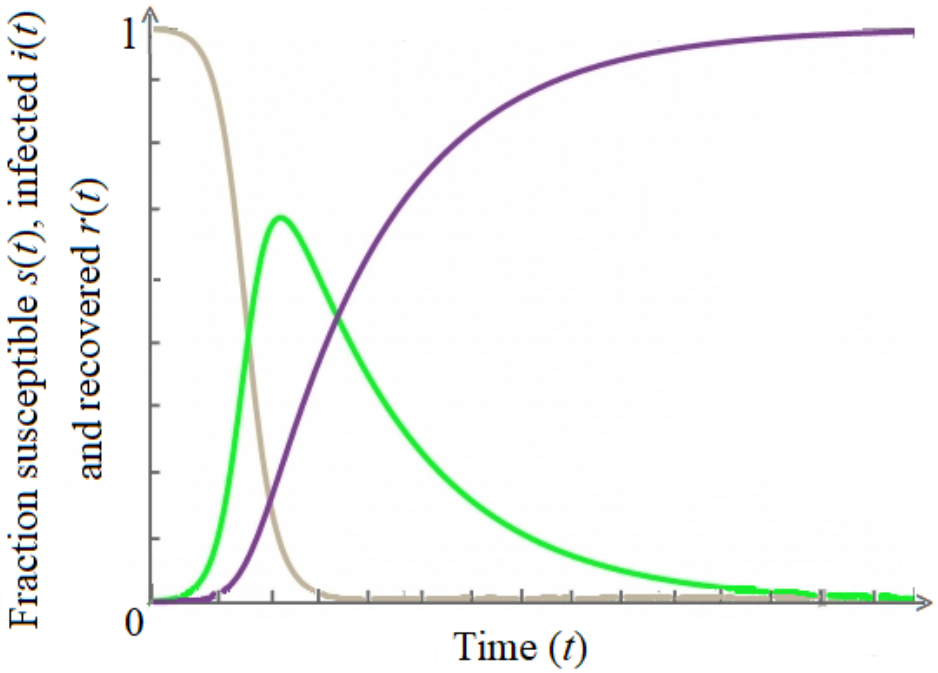


Figure 2.6: In this graph, the fraction of the susceptible (grey), infected (green) and recovered (purple) individuals are plotted against time. Individuals in the recovered state have developed immunity to the disease. This model suits the modeling of the spread of diseases such as SARS, influenza and plague diseases. [5].

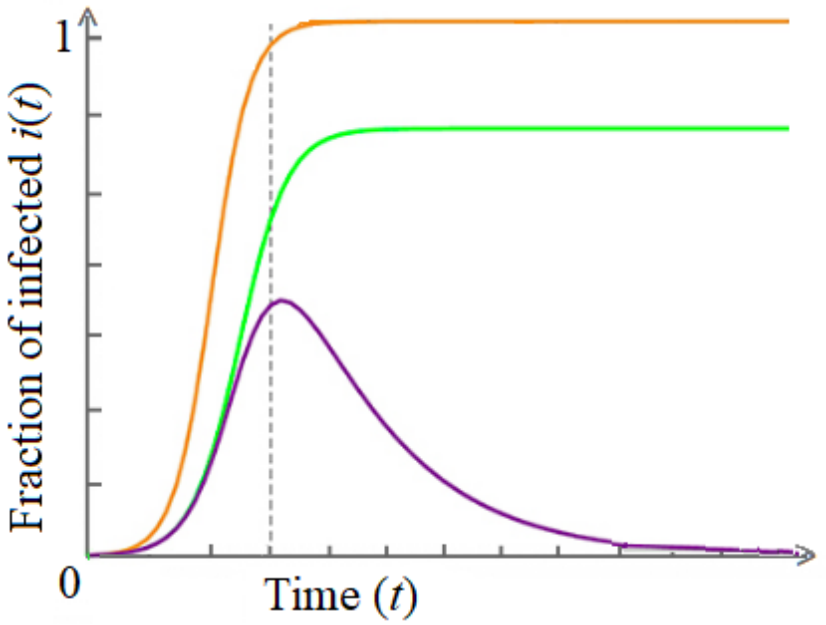


Figure 2.7: The figure shows a comparison between the three models, SI (orange line), SIS (green line) and SIR (purple line). In all the models, the early stage of the disease shows an exponential evolution. The SI model has the fastest growth with the same β for all the models. For the SIS and SIR models the growth is slower due to individuals recovering with time. As the recovery rate, α, is high enough there will be a disease free state both for the SIS and the SIR model where the number of infected decrease exponentially with time. In the final stage of the disease, the three models have different outcomes. All individuals get infected in the SI model, but only a finite fraction of the population gets infected in the SIS model, this is $i(t = \infty) < 1$. For the SIR models all the susceptible individuals get infected, all the infected recover, so $i(t = \infty) = 0$. Based on [5].

2.3 The Bacterium *Neisseria gonorrhoeae*

Gonorrhea caused by the gram negative bacteria *Neisseria gonorrhoeae*, an exclusive human pathogen, is the second most common STI after chlamydia [15, 11]. *N. gonorrhoea* could infect the mucous membranes of the reproductive tract, cervix and uterus in women and the urethra in women and men [18]. The pathogen could also infect mucous membranes of the eyes, mouth, throat and rectum [19, 20]. The STI is transmitted via sexual contact with an infected partner. The pathogen can also transmit from mother to infant during birth. There is an increased risk of acquiring HIV, especially for the MSM group [47]. The seasonal spread of gonorrhea, or the seasonal oscillations of the infection is very small, less than 10 % [48]. This is not the case for for example the annual influenza that often spreads in very high rates around mid October to mid November in Western Europe [48, 49]. The risk of infection is lowered by the consistent and correct use of condoms. A gonorrhea infection is treated with antibiotics, but treated individuals are not immune and can re-acquire the infection [50].

The infection is usually symptomatic in men, 90 % of the cases, and usually asymptomatic in women, 50 % [16, 51]. The usual symptoms are a urethral infection in men with white, yellow or green discharge, see Figure 2.8 [18]. Could be complicated by epididymitis. Symptoms usually occur one to 14 days after infection [18]. In women symptoms are often very mild and could be mistaken for a bladder or vaginal infection. Symptoms are increased vaginal discharge, vaginal bleeding and dysuria. Symptoms or not, women with gonorrhea are at risk of getting serious complications like causing pelvic inflammatory disease and cause infertility and the risk of ectopic pregnancy [18, 51, 15].

In the case of a rectal infection, present in both women and men, symptoms may include discharge, anal itching, bleeding or sore bowel movements. However this type of infection is usually asymptomatic [19, 18].

Gonorrhea is now most commonly diagnosed with the use of polymerase chain reaction, PCR [10]. This method provides very specific and accurate results in confirming the pathogen. The PCR method is preferred in detecting gonorrhea in the rectum or the throat, locations that is difficult to detect with other methods. In Norway, PCR has lowered the amount of false positive test results to a minimum [10].

2.4 Treatment and Antimicrobial Resistance

Gonorrhea is treated with antibiotics. The WHO, CDC and FHI among others recommend dual treatment of the infection to be able to respond to changing antimicrobial resistance (AMR) patterns of *N. gonorrhoeae* [11, 52, 10]. The bacteria have shown high-level resistance to the previously recommended quinolones and have shown decreased susceptibility to the third generation cephalosporines. The standard treatment is with an injection of ceftriaxone and azithromycin orally, both as a single dose [11, 10]. After two weeks all patients should take a control test to be sure the treatment was successful, due to the high

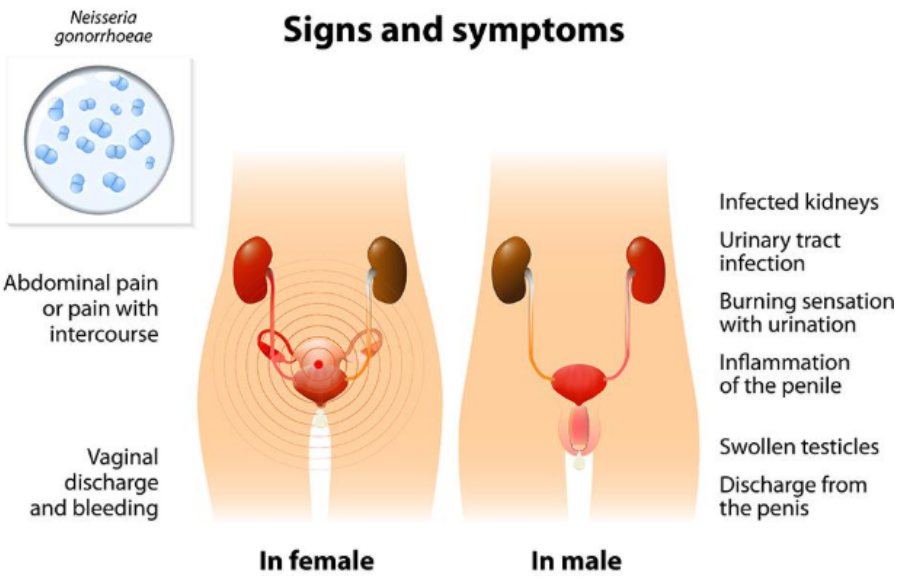


Figure 2.8: A schematic overview of symptoms for a gonorrhea infection in women and men. Figure retrieved from [6].

risk of antibiotic resistance [10]. See Table 2.1 for the percentage rate of how susceptible, intermediate and resistant gonorrhea is to the different antibiotics.

2.5 Analysis of Variance

ANalysis Of VAriance (ANOVA) is a statistical hypothesis testing method used for analysis of groups of experimental data. In ANOVA, the null hypothesis is that the groups arise from the same probability distribution or process, thus leaving no significant difference between the groups [53]. The test result is usually called statistically significant if there is only a small probability that it could occur through a relevant, random process. The statistical significance level, the p-value, is usually set to 0.05 in this thesis. If different, it is specifically stated. If the p-value from the statistical test is below the significance value, the null hypothesis (no difference) will be rejected.

The n-way ANOVA with three or more factors is used in this thesis. This model can be written as

$$y_{ijkr} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkr} \quad , \quad (2.18)$$

where y_{ijkr} is the observation of the response variable, μ is the general mean, α_i, β_j and γ_k are the deviations of groups in factors A, B and C, for example parameters in a model, respectively [54]. $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$ and $(\alpha\beta\gamma)_{ijk}$ are interactions between

Table 2.1: An overview of the situation on antimicrobial resistance in gonorrhea in Norway. The percentage rate of how susceptible (S), intermediate (I) or resistant (R) *N. gonorrhoea* is towards different antibiotics is listed [1].

		S [%]	I [%]	R [%]
Penicillin G	2.9	66.1	31	
Cefixime		97.6	0	2.4
Ceftriaxone		100	0	0
Azitromycin		64.1	24.4	11.5
Ciprofloxacin		51.4	0	48.6
Spectinomycin		100	0	0
Tetracykline		43.9	15.2	40.9
Beta-lactamase	76.5	0	23.6	

factors A and B, A and C, B and C and A and B and C, respectively. ϵ_{ijkr} are the random disturbances that are normally distributed and with constant variance.

In this thesis, many ANOVA tables will be used. Here comes a brief explanation of an example table, Table 2.2. Df is degrees of freedom, Sum Sq is sum of squares, Mean Sq is mean squares. The mean squares are found by dividing the sum of squares by the related degrees of freedom [55]. The number of stars, *, states the significance of the variable. One star means that the p-value is < 0.05, two stars means < 0.01 and three stars < 0.001.

Table 2.2: Example table of an ANOVA.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
var1	1	516778	516778	17.7429	0.0002519	*
var2	1	3708595	3708595	127.3301	1.000e-11	**
var3	1	37287	37287	1.2802	0.2678140	
var4	1	453121	453121	15.5574	0.0005128	***
res	27	786398	29126			

Chapter 3

Methods

Sections 3.1, 3.2, 3.3, 3.4 and the 3.5 are based on the previous project "Network Analysis and Modeling of the Transmission of Pneumonic Plague from *Yersinia pestis*."

3.1 The Global Epidemic and Mobility Model (GLEaM)

The modeling software used in this thesis is called GLEaM and was developed by Professor Vespignani and his research team. Since 2005 it has been under continuous development and improvement and been used in a lot of research including mobility in epidemic spread, analysis of the SARS outbreak in 2002-2003 and analysis of influenza outbreaks, specifically the H1N1 outbreak in 2009 [37, 56, 38, 57, 49, 7]. It is a model that integrates sociodemographic and population mobility data in a spatially structured stochastic disease approach to simulate the spread of epidemics at a worldwide scale. It simulates and visualizes the global spread of infectious diseases [58]. Parameters like β, α and R_0 can be estimated and implemented in GLEaM. With a network-based method, GLEaM relies on chronological data that obtain the global transmission of a pandemic, without using medical reports [49, 5]. In GLEaM the homogeneous mixing hypothesis is applied, as explained in subsection 2.1 Epidemic Modeling. The simulation engine contemplates discrete individuals and the transition processes are stochastic and discrete.

3.2 Top-level Architecture of the GLEaMviz Tool

The top-level architecture of the GLEaMviz tool is made up of three components: The GLEaMviz client application, the GLEaMviz proxy middle-ware, and the simulation engine [7]. These three components can be divided into the GLEaMviz client and the

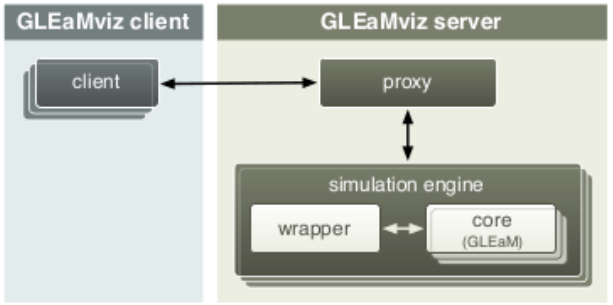


Figure 3.1: The GLEaMviz tool consists of clients that interact with the server. The server involves a proxy middle-ware component and many simulation engines. All the simulations done are performed by one GLEaMviz simulation engine. This launches a simulation core and performs analysis on the results [7].

GLEaMviz server, Figure 3.1. The GLEaMviz client application provides a graphical user interface (GUI) for the design and management of the simulations as well as visualizing the results. The GLEaMviz simulation engine is made up of a core and a wrapper that respectively executes and prepares the simulations based on information from the client by the proxy middle-ware. It is not the clients who are running the simulations, this is done by requesting the execution of a simulation to the GLEaMviz server's proxy middle-ware. By establishing a connection between these, multiple clients can use the server at the same time. When the proxy middle-ware is requested to run a simulation, the simulation engine instances that are needed to do the requests are started, and the status of the simulations is monitored. When they are completed, the GLEaMviz proxy middle-ware sends back all the data collected to the client. Results are written into files by the simulation engine and signals the middle-ware component. This is done rapidly, a large-scale simulation could take a couple of minutes on the best computers, because the core is written in C++ and the wrapper code is written in Python [7]. The architecture of the tool thus makes sure of high speed in high-volume setups and large-sale simulations, and it does not depend on the central processing unit-specific availability that is accessible by the user.

3.3 The Population, Mobility and Epidemic Layers

The structure of the model consists of three layers: Real-world data on the global population, real-world data of the mobility of this population and an individual stochastic mathematical model of the infection dynamics [7]. The three layers are illustrated in Figure 3.2. The population layer is made out of a grid, that divides the world into small square cells of 25·25 km^2, making in total over 250 000 cells. This is done with the help of satellite images and census sources that calculate the population density in each of the cells. Afterwards, it assigns each cell to the nearest transportation hub, an airport in this case. A subpopulation is defined by the cells having the same airport as the closest hub. In total, 3300 subpopulations are generated. The data are gathered from the websites of the

Gridded Population of the world and the Global Urban-Rural Mapping-project run by the Socioeconomic Data and Application Center of Columbia University [59].

The mobility layer is formed by the airport network and the commuting network [7]. The airport network in GLEaM consists of worldwide booking data sets from the Official Airline Guide (OAG) database and the International Air Transport Association (IATA) database [60, 61]. Altogether information and statistics about air-travel bookings, which airports that pairs connected flights and how many seats that are available per flight for more than four million connections for bookings between any two airports each month all over the world [62]. This covers information from more than 3200 airports in 232 countries in the world [7]. The network shows the significant variations in the number of passengers per connection and the number of destinations per airport. Some airports are big with a lot of connections and connecting flights, while other airports are small with a lower volume and fewer connections. This has a great impact on how infections spread around the world. The commuting network is not as thorough as the airport network, but covers 40 countries on five continents. The data set includes more than five million commuting connections between subpopulations, seizing the uneven structure of he network that influences the local diffusion of infections amidst neighbouring subpopulations.

The disease dynamics layer models the course of infection and simulates the infection dynamics according to the characteristics of the disease [7]. The characteristics could be the latent period of the infection, mortality rates, immunity and the proportion of infectious individuals. The characteristics are defined by a compartmental model, so that each individual in the model fits into one compartment at any given time. Compartments could be susceptible, exposed, infectious, recovered, resistant or any given compartment that fits into the model being used. The compartments are connected by paths that define how the individuals are moved from one to another and associated parameters determine the likelihood of these transitions. GLEaMviz utilizes stochastic algorithms or a stochastic meta population approach that are mathematically defined to visualize the unfolding of an epidemic and shape the spread of an infection [7]. From these *in silico* epidemics as they could be called, lies a lot of information to be congregated for each subpopulation every day of the simulation about morbidity, number of secondary cases and prevalence throughout the world to mention some.

3.4 Model Builder

The model builder shown in Figure 3.3 presents the visual modeling tool where arbitrary compartmental models can be created. Epidemic models are made as flow diagrams, with boxes representing the different compartments. It is possible to make any epidemic model, and adjustments can be made inside the model builder or by editing the input file for the model. The model builder makes it easy to model many kinds of infectious diseases that are transmitted from human to human [7].

When an epidemic model has been made in the model builder, it could be exported as an XML-file. This is a very simple way of making the model larger and with more details. For

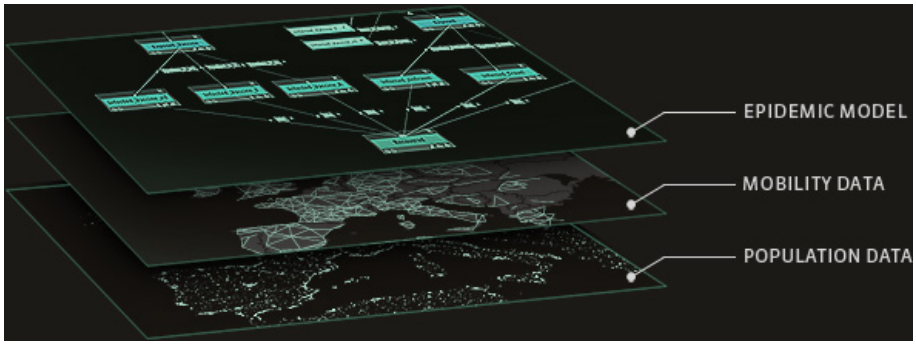


Figure 3.2: The structure of the GLEaM model consists of three layers. The first is the real-world data on the global population, the second is the real-world data of the mobility of the specific populations and the third is an individual stochastic mathematical model of the infection dynamics [8].

example, if many cities are chosen as hubs initially infected, and the model includes five compartments, it would take a huge amount of time to add all of these with city name and a number of infected. Instead, it is possible to make a script writing all the city number codes and a number of initial infected individuals for each compartment, and inserting these data in the XML-file. The data for which city are connected with which country and continent is found by exporting an already ran simulation.

The compartments can be labelled and assigned mobility constraints. This means that a compartment could be labelled with no flight or no commuting or both and individuals in that compartment will not be able to fly or commute. Whether the individuals in a specific compartment are clinical cases or not can also be defined. The three symbols on the bottom right of the compartments can be edited according to the infection status to the individuals in each compartment. This is shown in Figure 3.3, where the susceptible and the recovered compartments are non-clinical cases, visualized with a crossed out virus symbol. This means that they are not infectious and they do not have the infection latent in the body to spread the infection. The exposed compartment is a clinical case, but the individuals cannot spread the disease, hence only the virus symbol, while for the infectious compartment, the sign shows a thermometer adjacent to the virus sign meaning that the individuals in this compartment can spread the infection.

The color of the compartments can be changed and the variable names in the variable menu can be edited. In the variable menu a value for the variables has to be defined and the name of the variable have to correspond to the transition variable, see3.4. During the construction of the model, the client checks for errors and displays them in the inconsistencies table under the variable table.

The transitions can be made from the inverted minus sign (spontaneous) or the plus sign (infection). When the spontaneous transition is selected, a compartment is added as a transition target including a transition variable, see Figure 3.4. The transition variable can be modified in the model builder and a value for the variable is made by including this in the variable section seen in Figure 3.3. When adding an infection transition, one

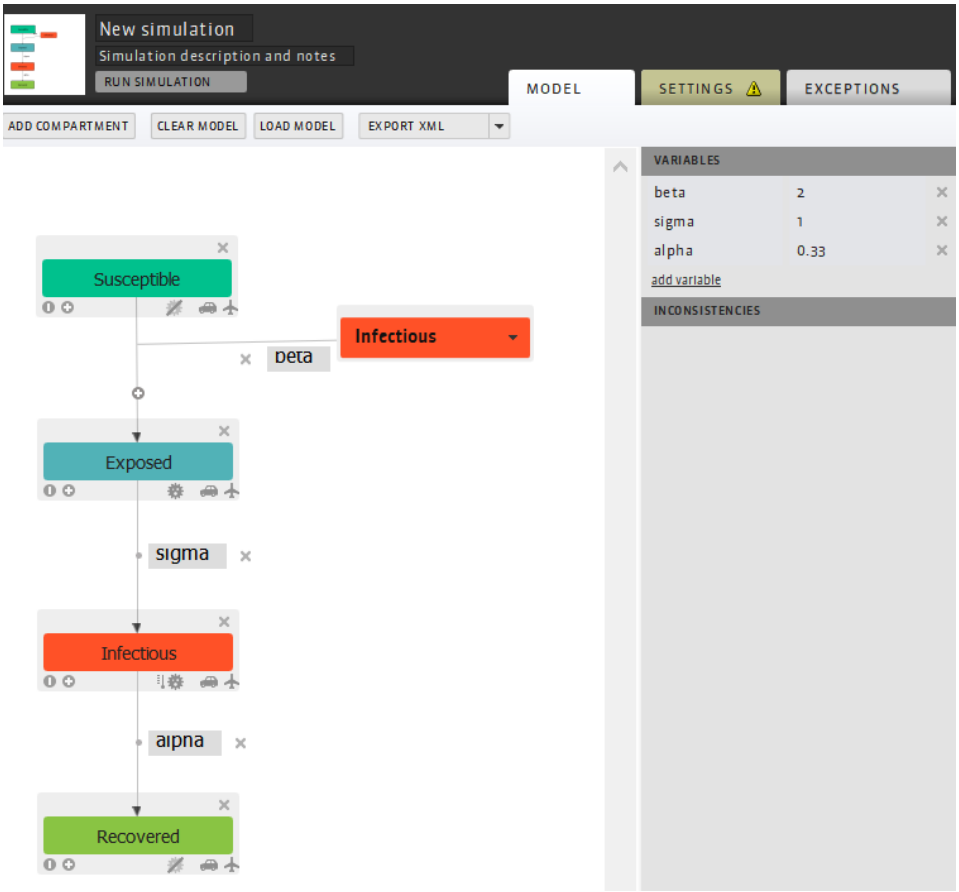


Figure 3.3: The graphical user interface of GLEaMviz is interactive and it is possible to construct any epidemic models. In the variable table it is possible to add, remove and edit the variables of the model and in the inconsistencies table inconsistencies are reported to simplify debugging.

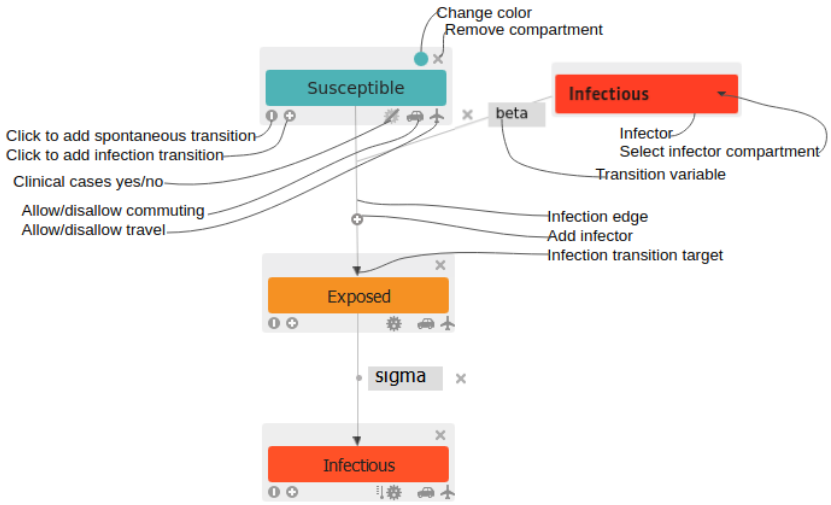


Figure 3.4: From a compartment, a new compartment can be created that is linked to it by either a spontaneous or an infection transition. In the figure an infection transition is selected. The infection transition includes an infector with a transition variable that shows the relationship between the first and the second compartment: The individual is moved from susceptible to exposed after the susceptible individual encountered an infectious individual in a rate β. In the compartments the following has to be chosen: Whether it is a clinical or a non-clinical case (the virus symbol), whether it can or cannot travel by airplane or by commuting.

infection transition target is made as before, but an additional compartment is included: The infector. The infector includes a transition variable. As shown in Figure 3.4, the infector is infectious and is connected by the susceptible compartment and the exposed compartment with a variable with rate β. This means, that if an infectious individual encounters a susceptible individual, it becomes exposed in a rate β. It is possible to add more infectors with the plus sign button on the infection edge between two compartments.

The settings for the model could be modified in the settings menu, see Figures A.1 on page 74, A.2 on page 75 and A.3 on page 76 in Appendix A. Here it is decided whether the simulation should be single- or multi-run. For single-run simulations the results for each day are cumulatively given to the client as the simulation runs. For the multi-run a number of runs for one simulation has to be set and the results are statistically analyzed when all the runs are finished. The results need to be retrieved by the client to be visualized. All the data retrieved from a multi-run simulation are shown as median values with a corresponding 95% confidence interval. The duration of the simulation must be chosen as a start-date with day, month and year for an outbreak. The duration for one simulation has a maximum of 365 days. Airline traffic has to be chosen, and at 100 % it runs the simulation with the most accurate flight data from the databases. Enabling seasonality means that the seasonality in the air traffic will be considered. The seasonality in GLEaM does not mean the same as for seasonality in gonorrhea, see 2.3. The commuting model is set to gravity and time spent on commuting destination meaning the transportation hubs

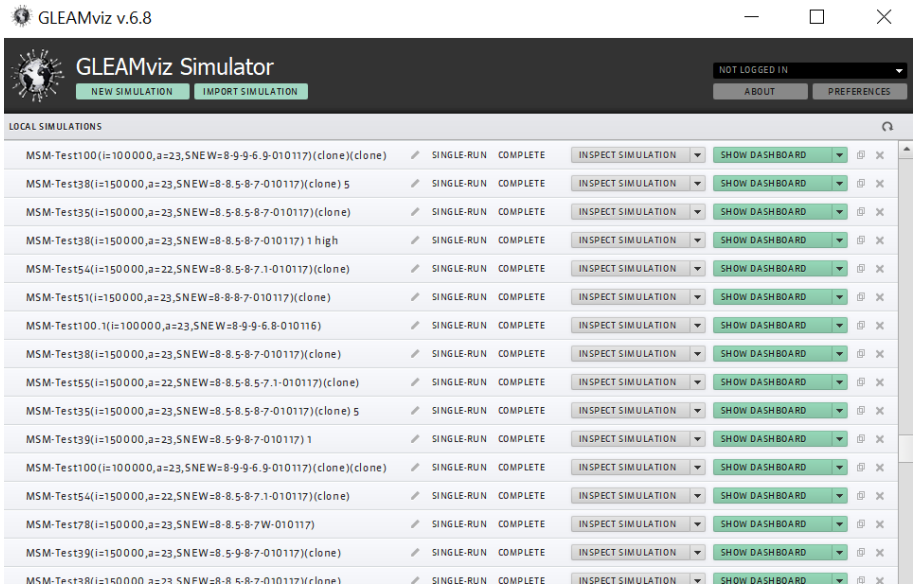


Figure 3.5: The dashboard includes all existing simulations. New simulations can be created and saved simulations can be imported. Each local simulation can be edited in the Inspect simulation, where the model builder and the settings are accessed. With the blue Run simulation, the visualization interface can be opened and the simulation starts. When the simulation is done, a green button with Show dashboard appears, this opens up the visualization interface where the simulation can be inspected in detail.

is set to 8 hours as default. The definition of an infected country and an epidemic is also chosen as a default to be 1 and 2 infected individuals, respectively.

Further, the distribution of the population in the model's compartments must be chosen, and has to include 100 % of the susceptible population. The initial geographic location of the epidemic also has to be chosen, this could include one or multiple initial locations depending on the type of epidemic. In the example in Figure A.1 on page 74, A.2 on page 75 and A.3 on page 76 the location is set to all the cities in Europe and the number of infectious individuals is generated from distributing 500.000 infectious women and men equally to all the cities. When a model is finished and the settings are set, it can be exported as an XML-file and the diagram representation can be exported as a PDF-file.

From the GLEaMviz dashboard seen in Figure 3.5 all existing simulations can be managed, new simulations can be created and saved simulations can be imported. Simulations that have been run can be exported and saved and all the information on the simulation can be retrieved. Each local simulation can be edited in the Inspect simulation, where the model builder and the settings are accessed. The blue Run simulation function opens the visualization interface and starts the simulation. The results from the simulation can be inspected as the simulation runs or it can be visualized like a movie after the simulation is done. Simulations can also be renamed from the dashboard.

3.5 Visualization Interface

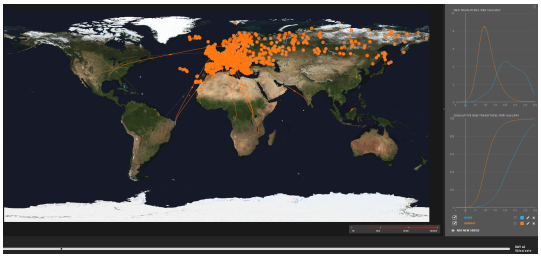
When a simulation is completed, the results can be visualized in the visualization dashboard. The epidemic's geo-temporal evolution can be interactively visualized as a movie, or by pausing the infection and clicking one day forward or backward. The visualization interface is shown in Figure 3.6 where three snapshots of a simulation movie are displayed. One from day 46, one from day 130 and one from day 365. In the figures orange dots are placed across Europe, these represent the initially infected individuals. In Fig 3.6b further into the simulation (day 130), many hubs in the world possess gray dots, this means that the area is infected. In the bottom right in the map, a scale bar is shown, this indicates how infected an area is. The two graphs on the right show the number of individuals infected every day and the cumulative number of infected individuals, per 1000 individuals as a function of time, respectively. Which areas to be visualized in the graphs could be chosen manually. In Figure 3.6a and 3.6b the infections via the airport network are shown as orange lines.

The visualization dashboard is a zoomable map, where the subpopulations or the different census areas are colored according to the quantity of new cases of the compartment to be displayed. For example the number of infectious individuals. The map is interactive, by hovering the mouse over the map it is possible to select a desired location, for example Trondheim. When a location is chosen data about that area will be shown, as in Figure 3.7. Information on the age structure in the subpopulation, how many individuals it contains, how many new individuals that are infected on a specific day (Day 130 in the figure) per 1000 individuals and the cumulative number of infected individuals per 1000 individuals are shown. In addition the number of hospital beds and number of physicians per 1000 individuals is displayed. The age structure for one country is the same in all the hubs in GLEaMviz.

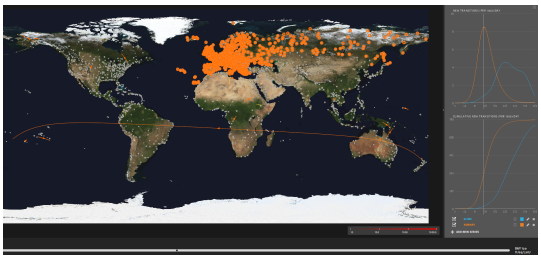
The analyzing tools can be found in the plus-sign menu. The *Map*, *Analyzer*, *Spato* and *Tree-Map* are already displayed when the visualization dashboard is opened. The *Map* is the main part of the dashboard, and *Analyzer* is shown to its right in Figure 3.6 as two graphs. These plots show the number of new transitions per 1000 per day as a function of time (days) and the number of cumulative new transitions per 1000 per day as a function of time (days), respectively.

The *Spato* and the *Tree Map* tools are found to the left of the Map tool. The *Spato* tool is based on the SPaTo Visual Explorer that is an interactive software tool where complex networks can be visualized and explored [63, 64]. What *Spato* does is that it reduces the network to the shortest-path tree of a chosen root too obtain a simpler view of the network. By changing the root node, the program allows to explore the networks from all perspectives, or transportation hubs, possible. The visualization can be switched between geographic mapping or concentric mapping of the hubs according to the shortest distance in the network in terms of transportation steps. The *Spato* tool is visualized in 3.8.

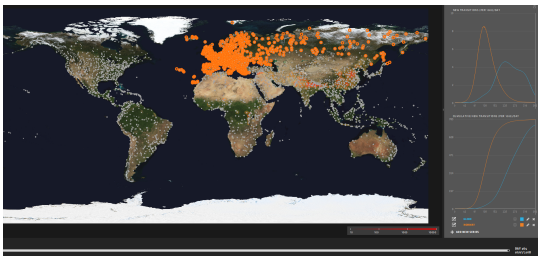
The *Treemap* tool visualizes the cumulative new transitions at the end of the simulation in all continents, the regions within a continent, the countries within a region or a city in a country, Fig 3.9 on page 28. The cumulative new transitions could be the number of



(a) Day 46



(b) Day 130



(c) Day 365

Figure 3.6: Example of a simulation in GLEaMviz. Snapshots from day 46, 130 and 365, respectively, are shown to demonstrate the evolution of the disease with time. The map shows the flights with infectious individual that could potentially spread the disease to the destination of the flight (orange lines). In Figure 3.6a orange dots are placed across Europe and these represent the initially infected individuals. In Fig 3.6b further into the simulation, many places/hubs in the world possess gray dots, this means that the area is infected. In the bottom right in the map, a scale bar is shown, this indicates how infected an area is. The two graphs on the right show the number of individuals infected every day and the cumulative number of infected individuals, per 1000 individuals as a function of time, respectively. Which areas to be visualized in the graphs could be chosen manually. Bigger versions of the figures can be found in Figures A.4 on page 78, A.5 on page 79 and A.6 on page 80.

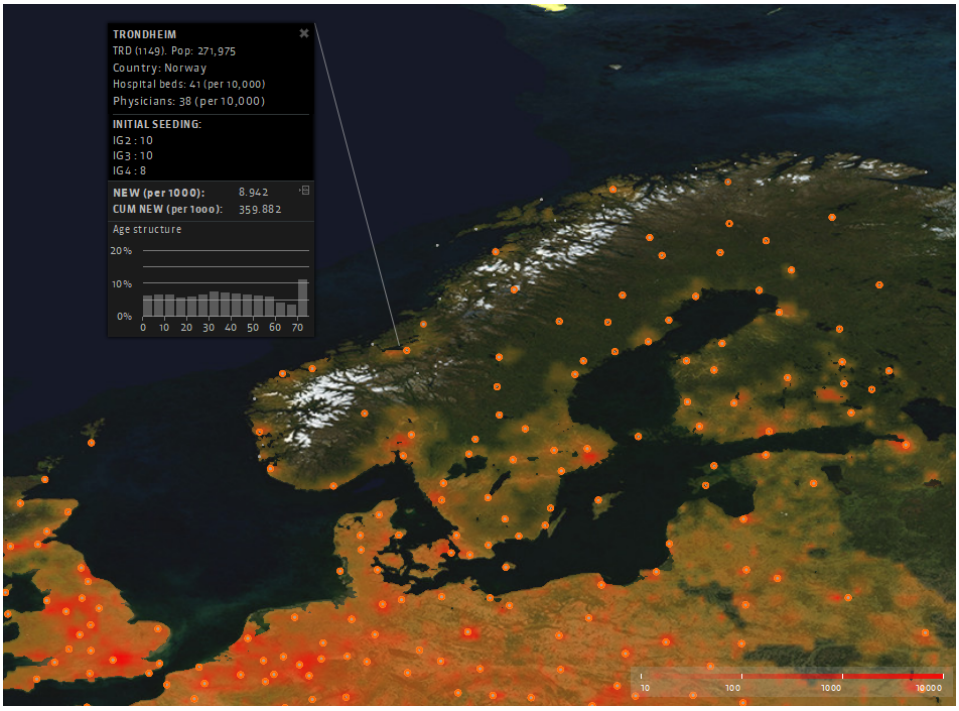


Figure 3.7: This figure shows a zoomed view on Europe and the northern countries. Details about the subpopulation/transportation hub Trondheim are shown. From day 130 of the simulation seen in Figure 3.6, the number of individuals, the country, number of hospital beds, number of physicians (the two last per 1000) are displayed. The age structure of the population can be seen together with the number of new infected people and cumulative number of infected individuals per 1000 on day 130. The data could be viewed for all subpopulation all over the world in GLEaMviz just by selecting the desired location. Orange dots are locations with initial infected individuals and the red color shows how many people that are infected in that specific area (see scale bar on the bottom right).

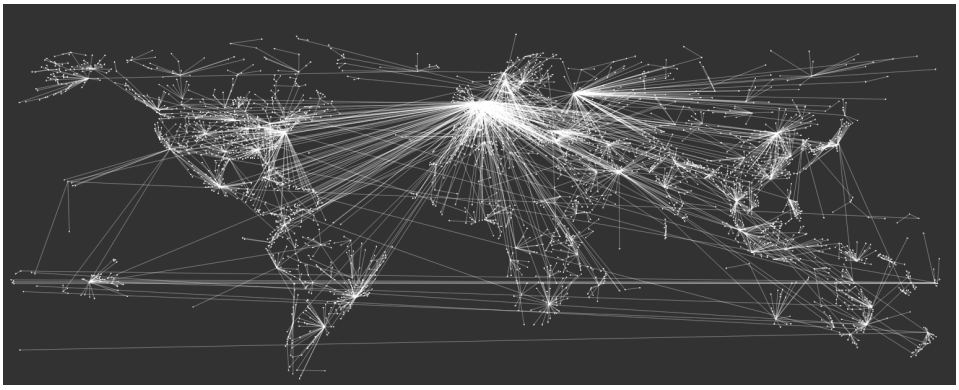


Figure 3.8: This represents a map view of the *Spato* tool. In this figure, Trondheim is the center of spread, and by hovering over any city in the world, the shortest possible path to the transportation hub Trondheim will be shown.

infected, susceptible or any other chosen compartment to easily retrieve information about the outcome of the simulation. The country with the largest fraction of infected individuals will be displayed with the largest color, in this case The United Kingdom.

The *Invasion tree* tool shows a representation of the the geographical diffusion of the pathogen with time from one location to the other [64]. Three different colors distinguish between locations being infected by flight or by commuting, or infection-free areas. In the control bar at the top of the tool, there is a slider, this can be adjusted for the time interval for which the invasion links are shown. It could last for one day or up to a year. The links shown are the links for all the infections happening up to the point of where the slider is set. For single-run simulations, the invasion tree could visualize the invasion tree at country level.The *Globe* tool visualizes the globe with an interactive spherical 3D representation.

3.6 Building Models in GLEaMviz

The simulations in this project were done using GLEaMviz. In the prediction models, the SIS model was used. In the models for the spread of antibiotic resistance, different versions of the SIR model were used where R represents being resistant and not removed as previously described.

All models were made with five compartments for susceptible and five for infectious individuals. Each of the five compartments represents one age group. The start date of all simulations was set to 01.01.xx, where xx represents the year of the respective simulation. All other settings were kept as default settings in GLEaMviz, Figure A.1 on page 74. In the models for the spread of AMR, five compartments for the resistant individuals were added, see Figures 4.13 on page 58, 4.14 on page 59, and 4.15 on page 61. All models were based on the homosexual male population exclusively, assumed to be 5 % of the

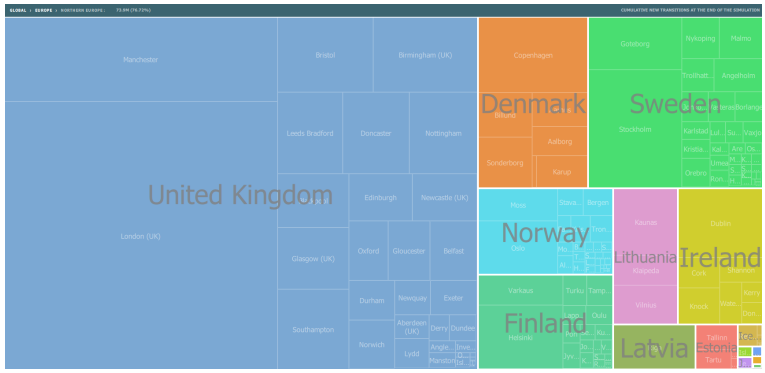


Figure 3.9: The *Treemap* tool visualizes the cumulative new transitions at the end of the simulation in all continents, the regions within a continent, the countries within a region or a city in a country, Fig 3.9. This snapshot from GLEaMviz represents the cumulative numbers in the Northern Europe. By choosing Europe in the top left corner, all the regions in Europe will be displayed, and by choosing globe, all the continents in the world will be shown. The country with the largest fraction of infected individuals will be displayed with the largest color, in this case The United Kingdom.

world population. Since the models only simulated spread in the MSM group, the total population in GLEaMviz was assumed to be MSM. Hence, when retrieving the results of the number of infected individuals in GLEaMviz, only 5 % of the total number was taken into account.

3.6.1 The Models

Epidemic network models for the years 2014, 2015, 2016, 2017 and 2018 were made. They included infection data from Tables B.4 and B.5, representing the respective years. They were used to predict the corresponding following year in Norway. For those models the total number of infected individuals was found from an average of 20 simulations for each year with a given set of parameters. The 2015-2017 models were also used to predict the disease outbreak in the year 2020. For those models, between ten and 20 simulations were performed with the same parameters as used for that particular year. Where one model was used to predict more than one year, input data from the average simulation was used instead of actual disease data given by FHI for that year. That way, a series of based-on-each-other simulations were performed.

3.7 World Population and How it is Divided

For all the models, the age structure (age distribution) of all the susceptible individuals was retrieved from the population pyramid web page [2]. The distribution of the population used in the models can be seen in Table 3.1. The pyramid displays the age structure in the

world from 2017. The population pyramid is divided into 42 age groups of five years each and equally split between men and women. In this thesis, only five age groups were used. For age group 1 (G1), the first three groups were merged. In age group 2 (G2), 3 (G3) and 4 (G4) the two next age groups were merged. For age group 5 (G5) the last twelve age groups were merged.

Table 3.1: The distribution of individuals based on age (years) and gender (men and women) for the five age groups (G1-G5) [2]. The last column shows the age structure of the MSM group, which was rounded due to integer requirement for GLEaM, and was used as input parameter in the models.

Gr.	Age	Men [%]	Women [%]	MSM [%]
G1	0-14	1.004.838.280 [26.5]	939.348.088 [25.2]	[27]
G2	15-24	612.479.218 [16.2]	574.103.002 [15.4]	[16]
G3	25-34	606.359.754 [16.0]	581.591.567 [15.6]	[16]
G4	35-44	504.450.984 [13.3]	493.316.347 [13.3]	[13]
G5	45+	1.062.936.986 [28.0]	1.135.859.927 [30.5]	[28]

The amount of initial seeds (infectious individuals) and their geographic location had to be determined. The initial spread of gonorrhea in the world was set to Norway and the background spread was set to Europe. To have a realistic model, gonorrhea would be transmitted everywhere. In order for the simulations to run without trouble, the spread was only set to Europe. The total number of infected individuals were counted in Norway only, and this is why the transmission in Europe is called the background spread. There were 100.000 infectious individuals in Europe averaged over all the 595 different locations, which is all the cities in Europe in GLEaM. Using the gonorrhea statistics for 2016 from ECDC, from 27 countries, the spread distribution of gonorrhea was given as 36 %, 43 % and 21 % for age groups 2, 3 and 4, respectively, in each hub in Europe in all the models [65]. G1 and G5 were only assigned one individual each per city in Europe. In Norway the infectious individuals was distributed as accurately as possible using the gonorrhea statistics from FHI for the specific year of the model, Tables B.4 on page 86 and B.5 on page 87, [66, 10, 67, 68, 69]. In cases of poor location information for the infected MSM, individuals were placed in the cities with the highest spread (Oslo, Bergen, Stavanger and Trondheim) [69, 68, 67, 66, 10]. The 2016 and the 2017 models were also used to model the spread of gonorrhea with infected individuals in Oslo only. For these models, ten simulations were performed and averaged to find the total number of infected individuals.

3.8 The Transmission Rate and the Basic Reproduction Number for *N. gonorrhoea*

The calculation of the basic reproduction number R_0 for gonorrhea was not as easy as for measles or pneumonic plague, see Section 2.2.2. The spread of gonorrhea could be explained with a number of factors given in Fig. 3.10, where the three main factors

were those used to calculate R_0. Additionally secondary factors like age, consistent condom use, treatment, other STIs and tertiary factors like symptoms, mobility, ethnicity and knowledge and opinions came into play. The secondary factors were the proximate determinants influencing all the components of R_0 and the tertiary factors were the underlying determinants [9]. In this thesis as many factors as possible have been included in the calculation of the R_0. Especially age, treatment, sexual orientation and presence or absence of symptoms. Other factors not included in the calculation will be discussed in Chapter 4.

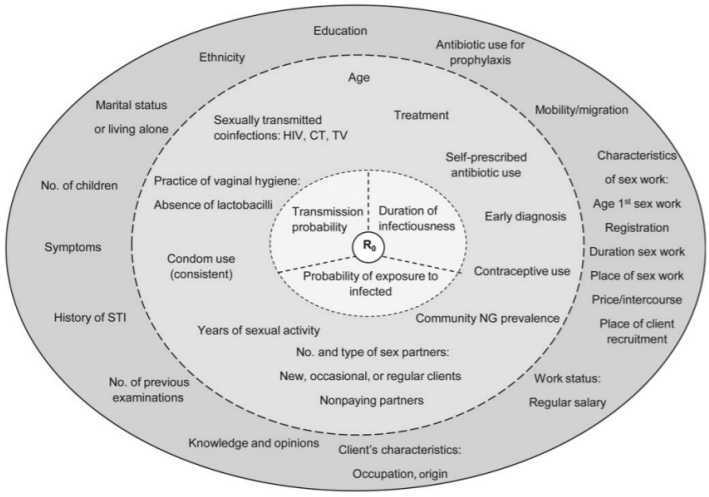


Figure 3.10: The basic reproduction number can be explained with many factors. The three factors in the inner circle are the transmission rate and the recovery period as in Eq. 2.13. The two outer circles represent factors that make out β and α. Figure retrieved from [9].

To determine R_0, β and α had to be calculated. The five most specific and measurable components are the interaction between the age groups (β_1), the probability of exposure to infected (β_2), the transmission probability (β_3), the duration of infectiousness (α) and the age groups. The age groups are included in the definitions of β_1 and β_2. In this thesis, β was composed of three primary factors from Figure 3.10.

The first factor considered the probability of exposure to infected individuals and the age groups. In the case of modeling for gonorrhea, this meant the probability of having intercourse at all during the year of the simulation. This is directly linked to age and sexual orientation. There is no research that states this exact number and it was based on and educated approximate guesses were done to decide on these probabilities, Table 3.2 on the facing page. The second factor was the probability of having intercourse with another individual from the same or different age group, see Table 4.3 on page 48. The third factor was the probability of getting the infection after having had intercourse. This rate was assumed to be 60 %. There are a lot of statistics presenting this probability. The FHI states that there is a chance between 50 - 70 % of getting gonorrhea for WSM (women who have sex with men) and 20 - 30 % for MSW (men who have sex with women) [66]. These data are for heterosexual transmission per intercourse, with vaginal, unprotected sex with an

infected individual [66]. The risk of getting gonorrhea increases to about 60 - 80 % after four or more exposures to the bacteria [70, 71]. For rectal intercourse the transmission rate is not reported, but appears to be as high as male to female vaginal intercourse, or even higher [71]. There also exists a perinatal transmission possibility, meaning transmission from mother to infant during vaginal delivery [66, 71]. The three factor making out β were implemented for all the interactions between the compartments, 25 in total, see for example Table B.1 on page 83.

Table 3.2: Matrix representation of the probability of having intercourse with same or different age groups for MSM. The rows and columns show β_1 of interactions between two individuals in any age group.

	$\beta_{1,G1}$	$\beta_{1,G2}$	$\beta_{1,G3}$	$\beta_{1,G4}$	$\beta_{1,G5}$
$\beta_{1,G1}$	0.996	0.001	0.001	0.001	0.001
$\beta_{1,G2}$	0.001	0.75	0.17	0.075	0.004
$\beta_{1,G3}$	0.001	0.17	0.6	0.18	0.049
$\beta_{1,G4}$	0.001	0.075	0.18	0.6	0.144
$\beta_{1,G5}$	0.001	0.004	0.049	0.144	0.802

3.9 The Recovery Rate for *N. gonorrhoea*

The recovery rate α is the inverse of how many days it takes to recover. The number of days tested for the recovery period was between 20 and 40 days, see tables in Section B.5. The recovery period found to be the best fit and that was used inn all the models, was 23 days.

3.10 Transportation Hubs in Norway

All the Norwegian airports in GLEaM are listed in Tables B.4 on page 86 and B.5 on page 87. The reports from FHI includes the number of infected individuals per county in Norway. Since some of the counties have more than one airport in GLEaMviz, one of the airports was chosen for the initial seeds: Alta Airport (Finnmark), Tromsø Airport (Troms), Bodø Airport (Nordland), Trondheim Airport (Sør-Trøndelag), Molde Airport and Ålesund Airport (Møre og Romsdal), Sogndal Airport (Sogn og Fjordane), Bergen Airport (Hordaland), Stavanger Airport (Rogaland), Oslo Airport (Akershus, Oslo, Hedmark, Buskerud, Vestfold and Telemark), Kristiansand Airport (Aust- and Vest-Agder). The remaining counties only had one transportation hub in GLEaM.

Chapter 4

Results and Discussion

In this chapter, a detailed discussion of the program, parameter data, the pick of individuals for the models, the models created and their generated data, and a comparison between the models regarding their predictability will be presented. Furthermore, a detailed explanation of the used parameters and an analysis of variance of these parameters will be given.

4.1 Why Use GLEaMviz?

The GLEaMviz computational modeling tool was used because it is a powerful modeling tool that can model infectious disease globally, integrating transportation networks based on flight and commute data. The tool makes it possible to seed an infection in any geographic census area, which makes it possible to model the STI as it is present anywhere in the world. The use of epidemic network models for the transmission of for example gonorrhea or HIV has been used in many studies [48, 72, 45, 73]. All of the transitions between the compartments are discrete and stochastic, which are suitable in the modeling of STIs [7, 74, 75]. GLEaM is currently not suitable for the simulation of diseases that are vector-borne, depend on local contact patterns such as STIs, or act on a time scale that would make demographic effects relevant [7]. A vector-borne disease is a disease that is carried by an agent that transmits a pathogen to another living organism [76]. Vectors are often parasites or microbes.

The tool, however, allows for the introduction of mitigation policies at the global level. Localized intervention in space or time can be implemented in the GLEaM model and their introduction in the GLEaMviz computational tool is planned for future releases [7]. Besides this, GLEaMviz uses compartmental models to represent the modeled variables with a network perspective, connecting the globe and making spreading realistic and possible anywhere.

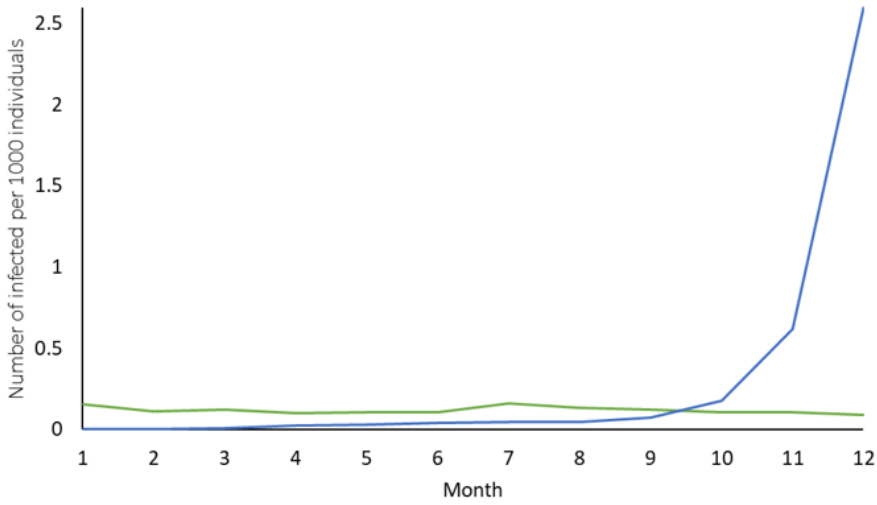


Figure 4.1: This graph represents the number of new infection transitions per day per 1000 individuals plotted against months. A simulation in GLEaM with the 2017 model yielding 888 infected in Norway (blue line) is compared with the real situation in Norway with 905 infected (green line) [3]. This trend was observed in the 2014-2016 models as well.

Modeling seasonality effects is an open problem in the transmission of influenza-like diseases [7]. A method for modeling with seasonality effects is included in GLEaMViz, yet the seasonality transmission effects in gonorrhea are minimal [7, 48]. Constant values for the parameters and the contact rates and patterns were used, which appeared to be good approximations for the gonorrhea spread [48]. However, analyzing the first simulation using the visualization interface the opposite was found. In Figure 4.1, and Table B.12 on page 92 the spreading happens very slowly in the beginning of the year, and peaks in the end (orange line). The actual spreading in Norway shows a linear, straight line (green line). From the reports in Norway it could be seen that there is no seasonal effect in the spread, only a little peak after the new year. That peak could be explained by the fact that many clinics are closed during the holidays, thus more individuals get tested after the holidays [48]. The incline in the spread in GLEaM could be because there are few initial seeds at the beginning of the simulations. After some time, more individuals are infected, thus resulting in a continuous increase in infected individuals throughout the year. This is one of the consequences of modeling in GLEaM. The program is built to model for influenza like diseases and not for STIs, because the local contact pattern network is yet to be further developed. Also because of the higher R_0 in the core group that are always present in the population, makes them difficult to model, see Table 4.4.

4.2 Why Modeling MSM?

Since 2010 MSM has been the largest risk group for gonorrhea infections in Norway [10]. FHI forecasts that the STI will be spreading like an epidemic in the following years in this group, thus it is important to understand how the infection spreads among MSM to find the best possible way to treat the disease and at the same time minimize the impact of AMR in *N. gonorrhoeae*. The infection fraction in the MSM group has increased by 400 % since 2011 and by 50 % from 2016 to 2017, see Figure 4.2 [10]. By comparison, the number of infections in the WSM and the MSW groups flattened out in 2017.

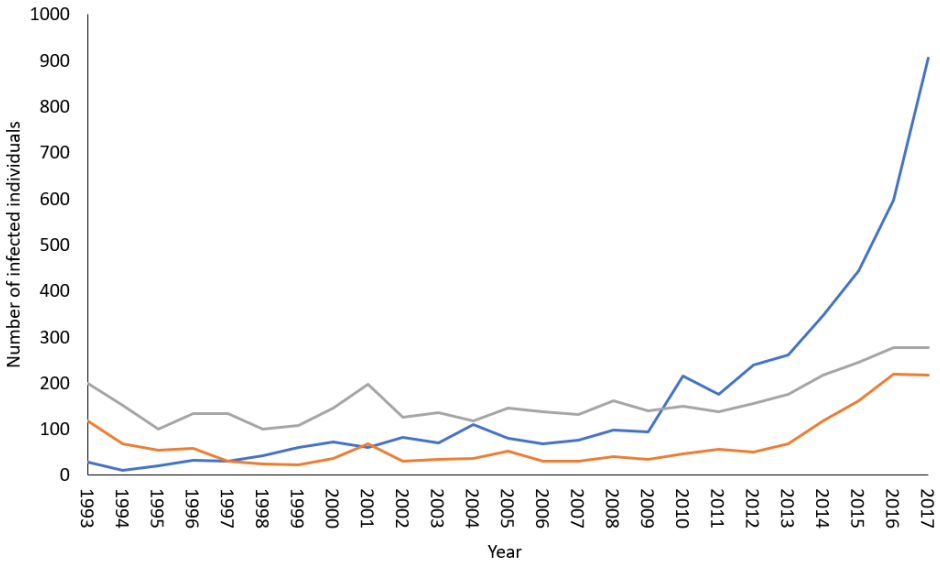


Figure 4.2: The number of gonorrhea cases for MSM (blue line), MSW (grey line) and WSM (red line) in Norway is plotted against years (1993-2017). Data gathered from MSIS and FHI [3, 10].

4.3 Models for Gonorrhea Transmission Among MSM

In this section, the distribution of MSM in GLEaM will be explained along with some observations of the information provided for gonorrhea among MSM, before the five SIS epidemic network prediction models for 2014-2018 and the two SIS epidemic network models with initially infected only in Oslo will be presented.

4.3.1 The Distribution of MSM

The overall distribution of homosexual male and female individuals is estimated to be 10 % on the global scale which has been discussed over the years [77, 78]. This claim is

controversial, but has become a general accepted fact [79]. In the models, the spread of gonorrhea is only considered for the MSM group, not WSW (women who have sex with women) because there is a negligible spread of gonorrhea among WSW [10]. 5 % is a very small fraction in this context, and modeling the spread of gonorrhea with 5 % MSM and 95 % MSW and WSM in GLEaM resulted in little or no spread, hence all the models constructed in this work only consisted of MSM. When simulations were performed the total amount of infectious individuals was multiplied by a factor 0.05 to represent the global 5 % estimate of the MSM group. Also, 75 % of individuals in the age between 15 and 24 years had shown sexual behaviour according to a US survey, and of those about 5 % were same sex experience for males [80]. In the age ranging from 25 to 44 around 95 % has had sexual behaviour, and of these 5.8 % same sex sexual behaviour, substantiating the 5 % assumption [80, 81].

4.3.2 Gonorrhea Statistics for MSM

Generating a model in GLEaM that would simulate the transmission of gonorrhea from one year to another requires solid data. According to the WHO around 78 million individuals distributed among six continents are affected by gonorrhea on a yearly basis, 4.7 million of them in Europe [82]. It was not possible to place 78 million initial seeds across the world into GLEaM, nor was it possible to insert 4.7 million seeds in Europe only. This was because of two main reasons: The first was because GLEaM resulted in error messages and stopped working when inserting too many initial seeds. The second reason was that the number of initial seeds can not exceed the number of inhabitants in a hub in GLEaM. For example, a hub like Røst in Norway has 297 inhabitants. This means that the highest amount of infectious individuals to be placed in Europe based on the hub Røst is 176,715, Eq. (4.1). The result retrieved using this equation accounts for the total amount of infected MSM that can be placed in GLEaM, given that Røst is the smallest hub compared to the hubs where there are placed initial seeds, for all the five age groups.

$$297\ inhabitants \cdot 595\ cities\ in\ Europe = 176,715\ individuals \tag{4.1}$$

Initial seeds for the spread in Europe were varied between 80,000 up to175,000. The seeds variable was varied to understand how it would affect the output result, but was found to have little impact on the number of infected individuals. This is further discussed in Section 4.7.2. A reason to only measure the spread of gonorrhea in Norway was that the quality of the data in Europe varies in various countries. Not all countries report their cases and very few report the transmission distribution based on sexual orientation and gender, which is important to have for the modeling of MSM [65]. Only 17 out of the 27 countries reported MSM cases in 2016, with the United Kingdom as the country with the highest amount of infected MSM (18,201) followed by the Netherlands (4,576) [65]. FHI provided comprehensive data, regardless of that it was a challenge to place all the initial seeds in Norway. Some counties do not have large airports, some have multiple. The decision where the seeds should be placed was done with discretion. Further analysis of the impact on where the seeds was placed can be found in Section 4.3.7.

4.3.3 The 2014 Model

The prediction of the year 2014 utilizing the 2014 model is found in Figure 4.3. The aim of building this model and run it in GLEaM was to explore whether it was possible to imitate the spreading of gonorrhea in Norway from 2013 to 2014, using the number of infected MSM from 2013 as initial seeds [69]. The statistics from FHI and MSIS were implemented to all transportation hubs in Norway affected by gonorrhea, see Table B.4 on page 86 [69, 3]. The number of infected individuals for the rest of Europe contributing to the background spreading was done according to Section 3.7 on page 28, with 100,000 infected individuals in total. The β parameters for this model is presented in Table 4.1 on the following page, and α^{-1} was set to 23 days on average. All parameters except for $\beta_{2,G5}$ were constant for all the following models in this thesis.

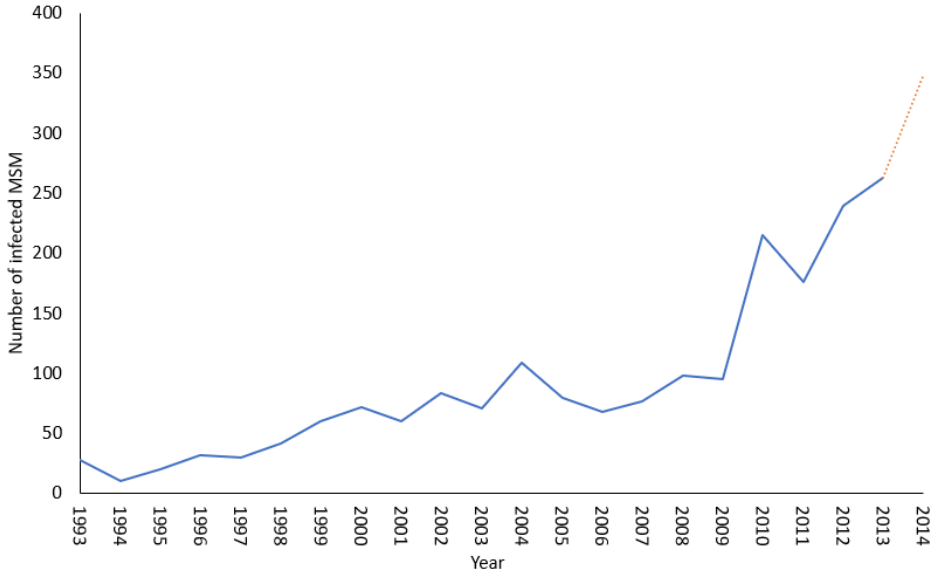


Figure 4.3: Number of infected individuals plotted against the years ranging from 1993 to 2013 (blue line) [3]. The orange dotted line shows the result from modeling transmission of gonorrhea for MSM in Norway from 2013 to 2014 with the 2014 model.

In Norway, FHI reported 263 infected MSM in 2013 and 347 in 2014. The goal with this model was simulate and retrieve 347 infected individuals starting with 263 infected. The number of infected individuals was counted only for the core group: age groups G2 to age group G4. Infected MSM in the groups G1 and G5 were not counted although they contributed to the spreading of gonorrhea, discussed further in Section 4.4.1. The average amount of infected MSM retrieved after running 20 simulations with the model was 350, a discrepancy of 0.86 % compared to the FHI report. In Figure 4.3 the blue line represents the number of infected individuals reported by FHI plotted against years. The orange dotted line shows the prediction with the 2014 model in GLEaM. The model was able to predict the situation in Norway in 2014 with the combination of parameters in Table 4.1,

α and initial seeds. Yet, looking at the spread throughout the year, Figure 4.1, the only similarity between the modeled result and the reported number by FHI is the cumulative number of infected MSM.

Table 4.1: Matrix representation of the transmission rates used in the 2014 model. The rows and columns show parameter β of interactions between any two age groups G1 to G5.

	β_{G1}	β_{G2}	β_{G3}	β_{G4}	β_{G5}
β_{G1}	0.0005976	0.0000006	0.0000006	0.0000006	0.0000006
β_{G2}	0.00048	0.36	0.0816	0.036	0.00192
β_{G3}	0.00054	0.0918	0.324	0.0972	0.02646
β_{G4}	0.00054	0.0405	0.0972	0.324	0.07776
β_{G5}	0.0003966	0.0015864	0.0194334	0.0571104	0.3180732

4.3.4 The 2015 Model

The model for 2014 was able to predict the spreading of gonorrhea in Norway deviating 0.86 % from the reported cases. Moreover, it was of interest to explore whether it was possible to make a model for 2015 as well.

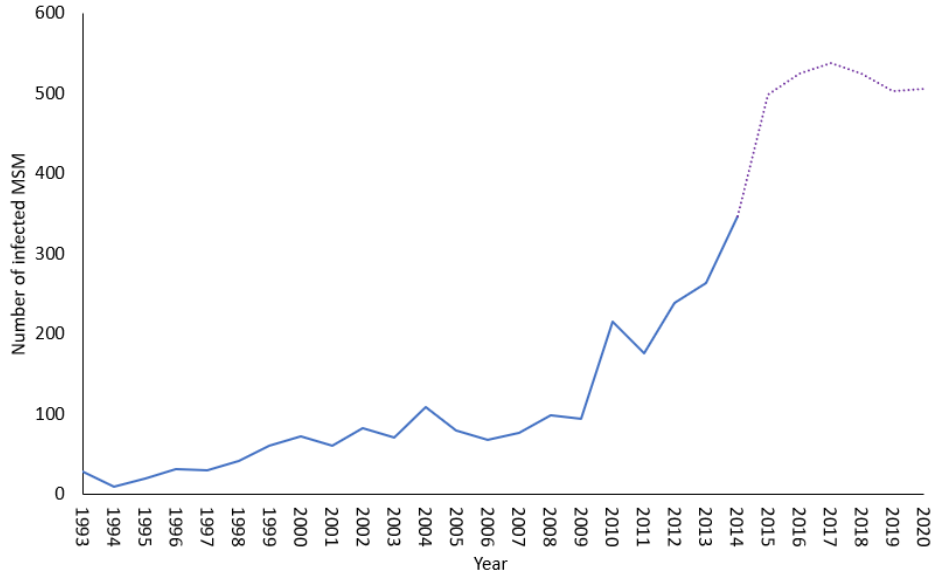


Figure 4.4: Number of infected individuals plotted against the years ranging from 1993 to 2014 indicated by the blue line [3]. The purple dotted line shows the result from modeling the transmission of gonorrhea for MSM in Norway from 2014 to 2020 with the 2015 model.

In Norway in 2014, FHI reported 347 infected MSM, and 444 in 2015 [68, 67]. The goal with the 2015 model was reproduce this spreading increase of gonorrhea from 347 infected

MSM to 444. Running 20 simulations with β as in Table B.1 on page 83 without changing the remaining parameters, the model produced on average 498 infected individuals. This number had a discrepancy of 12.16 % from the report cases for this year [67]. In Figure 4.4, the number of infected individuals as reported from FHI is plotted as a function of years (blue line). The purple line shows the results from modeling with the 2015 model. With only varying $\beta_{2,G5}$ from the 2014 model, this model was able to reproduce the real world data.

The model was further used to simulate 2016 to 2020. With constant parameters and a change in the initial seeds and simulation year, 10 simulations were performed and averaged. Results can be seen in Table 4.2. This method of only changing the initial seeds and not the parameters in the model, did not imitate the spreading of gonorrhea for more than one year. The stochastic model simulated approximately the same amount of infected individuals for all the years, regardless of the input data of the initial amount of infected individuals. A reason could be that the result from a simulation in GLEaM differed according to what was expected of the number of infected in the hubs in Norway. Most infected individuals are located in Oslo according to FHI, but in GLEaM the number of infected decreases in Oslo after one simulation year. For other smaller cities with initially low amounts of infected individuals, the number of infected increases. One explanation to this evolution of the transmission pattern could be that the spreading in GLEaM does not consider local contact patterns in the core group [7]. Hence the spreading will become equal to a great extent in small and large hubs, as in for example Alta or Molde compared to Oslo. Performing simulations with similar parameters, only changing the year of the simulation and the initial seeds, did not have much of an impact on the total number of infected MSM. The output results from these simulations are shown in Table 4.2.

Table 4.2: This table shows the different models (2014 to 2018) and the expected number of initial infected from reports of the respective years. The 2018 model's expected number was predicted based on Figure 4.8. The average (Avg) numbers display the output average from 20 simulations with the respective model. The following columns is the output from simulating the respective following years with one model and constant parameters.

Model	Expected	Avg	2015	2016	2017	2018	2019	2020
2014	347	350						
2015	444	498	506	538	525	531	505	508
2016	598	597		597	635	594	648	668
2017	905	891			888	841	859	808
2018	1063	1004				1039		

4.3.5 The 2016 Model

For 2015, the FHI reported 444 infected MSM in Norway. In 2016 the number had increased to 598 [66, 67]. The aim using the 2016 model was to successfully reproduce the amount of infected individuals. The result from the simulations for the 2016 model is presented in Figure 4.5 on the next page indicated by the yellow dotted line. The blue

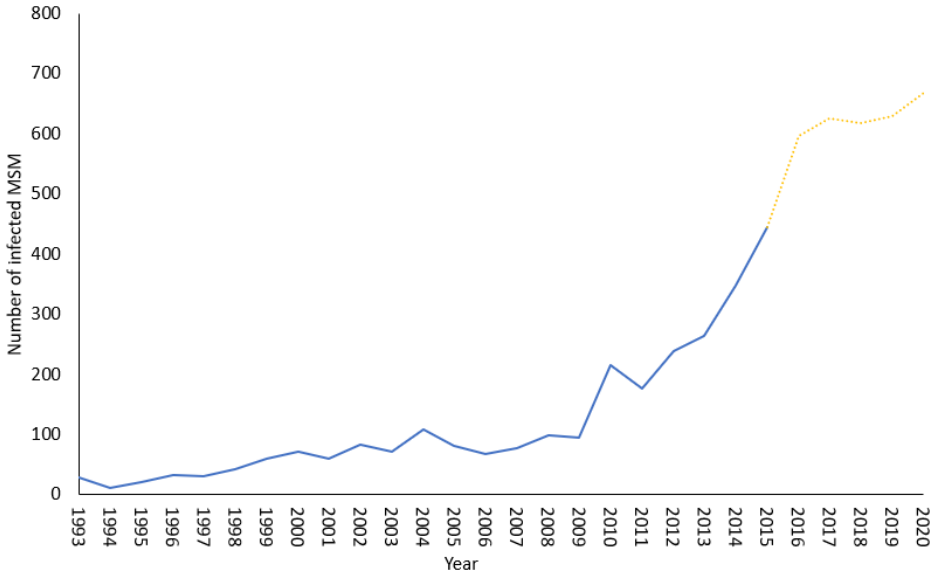


Figure 4.5: Number of infected individuals plotted against the years 1993 to 2015 represented with the blue line [3]. The yellow dotted line shows the result from modeling the transmission of gonorrhea for MSM in Norway from 2015 to 2020 with the 2016 model.

line is the number of infected individuals reported by FHI plotted against the years 1993 to 2015. Running 20 simulations with changing β according to Table B.2 on page 83, an average of 597 infected individuals could be retrieved, which differs 0.17 % from the reported number [66]. Similar to the 2015 model, the next years until 2020 were predicted; the results are shown in Table 4.2 on the previous page. Closely to the 2015 model, no trend in increase was found, although the trend in the 2016 model showed some increase.

4.3.6 The 2017 Model

The FHI reported 598 infected MSM in 2016 and 905 in 2017 [66, 10]. To reproduce this increase in infected persons, β was changed according to Table B.3 on page 84 and 20 simulations were performed to retrieve an average 891 infected individuals. This value differs by 1.55 % from the reported number (905). The 2017 model is represented in Figure 4.6 on the next page indicated by the red dotted line. The darker blue line shows the number of infected individuals reported by FHI plotted against the years 1993 to 2016.

Predicting the years until 2020 with the 2017 model as done for the 2015 model and the 2016 model, did not yield the expected increase as compared with the FHI reports. Results from the simulations can be found in Table 4.2 on the preceding page. For 2016 the 2015 model resulted in 538 models infected MSM and for 2017 525 infected MSM compared to 598 and 905 from the reports, respectively. No increase in the years to come is highly

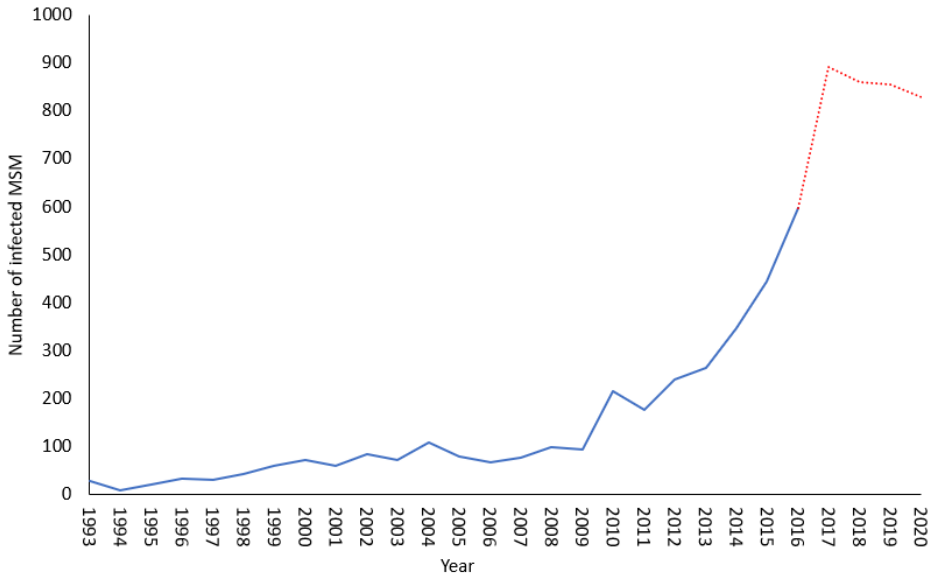


Figure 4.6: Number of infected individuals Norway plotted against the years, ranging from 1993 to 2016, indicated by the dark blue line [3]. The red dotted line represents the result from modeling transmission of gonorrhea for MSM in Norway from 2016 to 2020 with the 2017 model.

unlikely, as FHI reported that there will be a continued epidemic increase following the international trend in the Western countries including Norway.

Another approach to investigate the prediction capability of the model further, was to allocate all infected individuals into the Oslo transportation hub, since around 50 % to 60 % of all the infected individuals are located there.

4.3.7 The 2016 and 2017 Models - Placing all Infected Individuals in Oslo

The simulations inserting all the initially infected individuals in Norway in the Oslo hub were performed using both the 2016 model and the 2017 model. Using the same parameters as in the original 2016 model, Section 4.3.5, and initial seeds equal to 444, the simulations yielded 579 infected MSM on average after ten simulations, which is close to 598 (3.28 % deviation). This result is, however, inferior compared to placing initial seeds in their respective hubs. Modeling for 2017 with the 2016 model, changing the input seeds according to 597 in Oslo, a different result was expected: Since the input seeds were increased, an increase in the total number of infected MSM was expected. However, after ten simulations, the average was found to be 575 infected individuals. This is 0.05 % lower than for the original 2016 model and 36.46 % lower than what was expected for 2017. This observation is difficult to explain because it was expected that having more individu-

als initially infected in the largest city in Norway (accounting for 39 % of the population in Norway in GLEaM), would increase the total number of infected. This is because Oslo has a higher population density, higher contact pattern rates and a higher airplane traffic rate with both domestic and international flights.

After ten simulations using the 2017 model with an input of 598 initial seeds without further changing the original 2017 model, 799 infected individuals were retrieved. A result 11.7 % less than the reported number [10], and 10.3 % less than the result retrieved with the original 2017 model. This was because the spreading gonorrhea from just one hub in Norway in GLEaM, no matter if it is the largest hub, the spreading will first occur outside the hub. Then, when more individuals are infected, the infection will increase in the cities, but since GLEaM does not consider the local contact patterns, the size of the city and the population density will not have much impact on the total number of infected individuals after one year of simulation.

4.3.8 The 2018 Model

The goal with the 2018 was to forecast the spreading of gonorrhea for the MSM group in Norway based on the other models. The spreading with the 2018 model can be seen in Figure 4.7. Here, all results of the cumulative number of infected MSM from the models are plotted against the years ranging from 2014 to 2020. The blue solid line represents the data on infected MSM from FHI. Analyzing the increase in cases of infected MSM, the 2018 model manages to follow this trend. If the number of cases continue to increase in the same rate as the last years, the 2018 modeling could be thought to be a good estimate, only from looking at the trends in Figure 4.7.

In Figure 4.8, the reported number of individuals from FHI is multiplied with $\beta_{2,G5}$ and plotted against the year of the model. The linear regression in this plot makes it possible to estimate the number of gonorrhea infections among MSM for 2018. This holds if the spreading will continue to increase and varying $\beta_{2,G5}$ only. The average increase of $\beta_{2,G5}$ for the 2014 to 2017 models of 1.18 %, resulted in testing $\beta_{2,G5}$ for 2018 with values of 0.69, 0.695 and 0.7. From the linear regression,

$$y = 143.45x - 288,749, \tag{4.2}$$

changing to $y = \beta_{2,G5} \cdot z$,

$$z = \frac{143.45 \cdot 2018 - 288,749}{0.69} = 1063 \tag{4.3}$$

Thus, with $\beta_{2,G5} = 0.69$ a total of 1063 MSM would be expected to be infected in 2018. For a $\beta_{2,G5} = 0.695$, 1055 infected MSM could be expected and 1047 infected MSM for a $\beta_{2,G5} = 0.7$. Simulating models with these three parameters respectively, it was found that $\beta_{2,G5}$ of 0.69 was the best fit according to the linear regression of the Figure 4.8. The simulations performed retrieved on average 1004 infected MSM after ten runs, a deviation

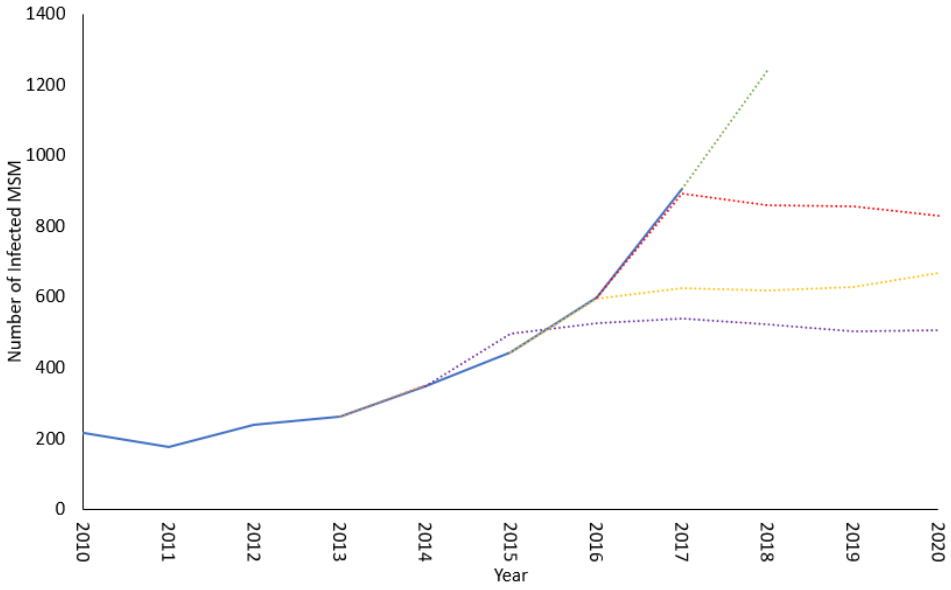


Figure 4.7: The number of infected individuals is plotted against years ranging from 2010 to 2020. The solid dark blue line represents the number of gonorrhea cases for MSM in Norway up to 2017 [3]. The dotted lines show the predicted spread among MSM in Norway using the different models: 2014 (orange), 2015 (purple), 2016 (yellow), 2017 (red) and 2018 (green).

of 5.55 % from the calculated number. These observations can not be confirmed yet, but comparing the spreading of gonorrhea in 2018 (January to May) to the same period in 2017, there is a decrease of 14.33 % [3]. These data are for both women and men and do not consider sexual orientation and the situation of the number of infected individuals could still increase and change. The number of infected MSM seems small comparing the increase in infections from 2017 to 2018, 14.7 %, and the increase of 50 % from 2016 to 2017 as well. If the spreading will continue to increase like FHI predicts, this may be of more than 14.7 % as predicted with the model.

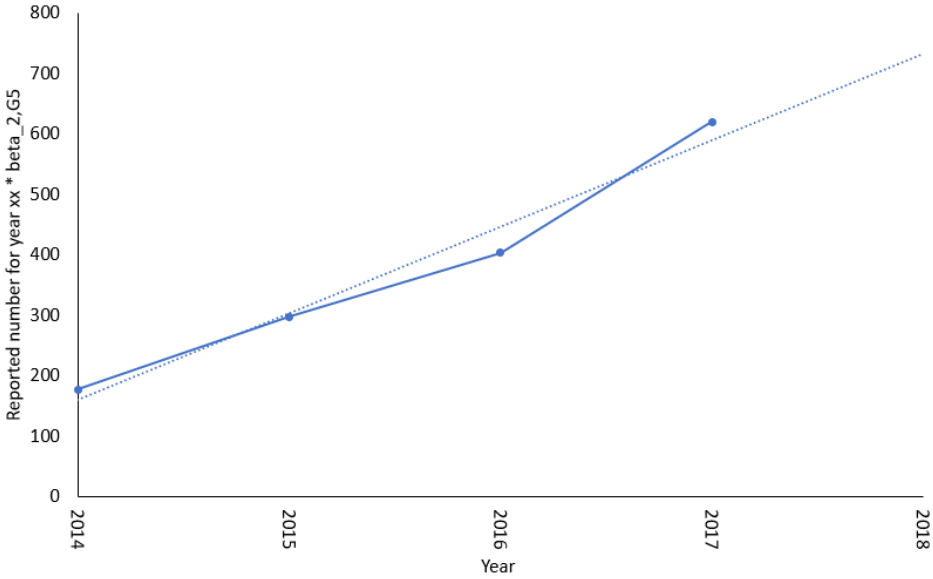


Figure 4.8: This graph represents the reported number of gonorrhea cases in year x from FHI multiplied by the parameter value $\beta_{2,G5}$ plotted against the year x of the report. Equation for the linear regression was given as $y = 143.45x - 288,749$.

4.4 The Transmission Rate

Figure 3.10 on page 30 shows relevant factors to be included when calculating the basic reproduction number, the transmission rate and the recovery rate. Especially to understand the transmission dynamics for an STI, it is important to expand the definition of R_0, especially in comparison to for example influenza like diseases [34]. Implementing all these factors in the model in GLEaM would be a challenge.

The five most specific and measurable components are (1) the interaction between the age groups (β_1), (2) the probability of exposure to infected (β_2), (3) the transmission probability (β_3), (4) the duration of infectiousness (α) and (5) the age groups. The age groups are included in the definitions of β_1 and β_2. This section will have its main focus on the listed parameters, while α will be discussed in Section 4.5.

4.4.1 The First Factor: The Interactions Between Age Groups

The probability of having intercourse with another age group or within the same age group is defined in Table 3.2 on page 31. This factor was included in the model as a part of β because it is important to understand how the compartments interact, since the transmission of STIs is not random [34]. In the models, all the age groups interact with each other to varying degrees. Because of this, the core group will have a symmetric association within the subpopulations [34].

Other possibilities when considering the transmission of STIs on a network are the assortative scenario and the disassortative scenario [34]. In the assortative the interactions will only occur within the age groups. The diassortative scenario takes interactions between different age groups into account. These scenarios generate spreading in two different ways: The assortative scenario generates a fast initial spread, but a small sized epidemic, while the disassortative scenario generates a slow initial spread of the infection but a large sized epidemic. The symmetric scenario is some place between these two scenarios. When analyzing the simulations, it could be seen that the spreading was initially very slow, and increased at the end. The total epidemic was not large when looking at the total population in Norway, similar to assortative, but the trend in the spreading evolution was similar to the disassortative scenario. Hence it could be argued that the literature is coherent with the experiments - the spreading had a symmetric association.

There is no available research stating the exact pattern of intercourse in between ages, but some educated approximate guesses were made, for example age gaps in MSM relationships. Age gap in heterosexual relationships where the woman is younger than the man, is well known [83, 84, 85]. The same trend could also be seen in homosexual relationships, and might be even more prevalent in same-sex relationships than in other-sex relationships [86, 87, 88]. Therefore the intercourse interactions between different age groups cannot be set to zero. In Table 3.2, the diagonal values are the highest, these are the interaction patterns within the same age groups. G1 has the highest probability of interacting with itself. This is because very few individuals in this group has intercourse [89, 90]. 5.6 % of males had experienced their first sexual intercourse before the age of 13 and 46 % of

high school students were sexually active (14-17 years) in a 2015 survey from the USA, substantiating the fact that G1 contain few sexually active individuals [91]. The interaction with G1 and the other age groups are looked upon as minimal. A case of gonorrhea in the G1 occurs mainly due to an infected mother giving birth to an infant, thus transmitting the infection [10, 3].

Examining the interactions between the age groups lead to a discussion on the size of the age groups and the extent to which they reflect the real world. The reason the whole population was divided into five age groups, was because the disease statistics are divided into age groups like these [65, 3]. The other reason for making five compartments in the model was because even though GLEaM provides age statistics, they are not possible to use. Meaning that after a simulation it would not be possible to know how many infected there are in each age group, just in total. Groups G1 and G5 are by far the largest, with 27 % and 28 %, respectively, making out 55 % of the total world population in GLEaM. Spreading of gonorrhea mainly takes place in between the core group G2-G4, not the G1 and G5 [3]. Thus, G1 and G5 are given one infected in each hub, not a percentage of the initial seeds like the other groups. As further work making a model without any initial seeds in these groups could be possible, to compare the number of infected MSM in the end of the year.

Considering the SIS model, Figure 2.4 on page 10, possible outcomes for G1 would be to reach a disease free state or the endemic state depending on whether α is bigger or smaller than β, Eq. 2.9 on page 8. This can be seen in Figure 2.7 on page 13, where the green line represents the endemic state, and the disease free state could be compared to the purple line in the SIR model. The transmission rate for G1 interacting with all other groups, $\alpha > \beta$, resulting in a disease free state when $t \rightarrow \infty$. Further discussion on these states in Section 4.6. However a disease free state is not seen from the simulations, see Figure 4.9. Since the maximum time for a simulation is one year, the number of new infected per 1000 throughout the year in G1 is not zero. If simulations could be ran with a longer time period, the result may be that the graph reaches a maximum before it decreases to zero. The trend in the cumulative number of infected MSM per 1000 individuals per month for the other groups were the same, only with more individuals infected per group, see Figure A.7 on page 81. The increase between October and November could be explained as the point where the infection reaches an exponential evolution before the breaking point where the infection will decrease exponentially if $\alpha > \beta$, or stabilize if not. Groups G2-G5 would however not have trends reaching zero if simulated for a longer time period, but flatten out when the endemic state peak is reached. This is because $R_0 > 1$ in all those groups, Table 4.4.

G5 includes individuals that are mostly sexually active, compared to G1. Considering the size of the group, this group will account for a very large amount of the spread of infection. In all the simulations, it was G5 that accounted for the largest, and highly unrealistic amount of spread. This was the reason why they were not counted when adding the total amount of infected individuals after one simulation. A possible solution for the G5 group size would be to divide it into smaller subgroups. In MSIS, statistics on gonorrhea are divided between age groups of ten years up to 80+, using these data would maybe have solved the issue [3]. However a critic to this possible method would be that data on

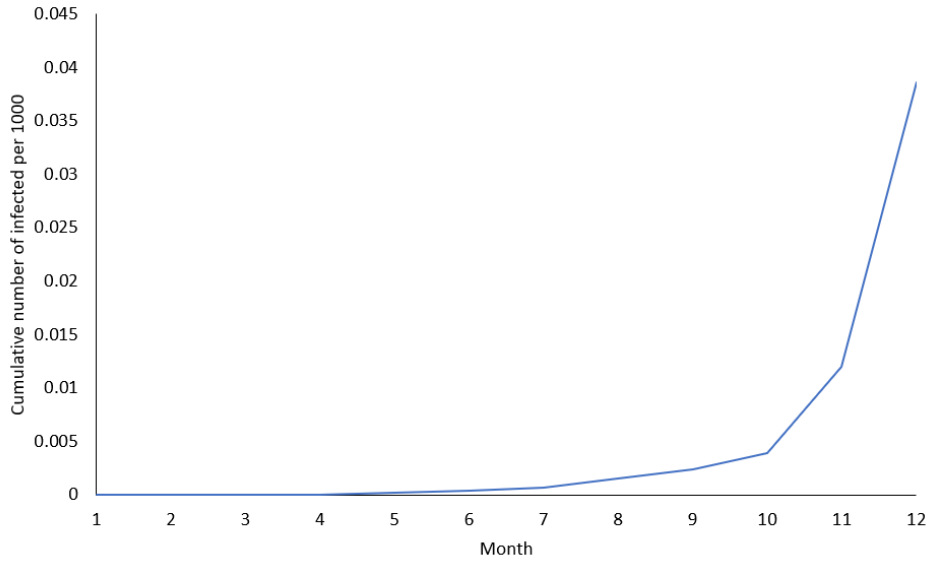


Figure 4.9: A graph representing the cumulative number of infected MSM per 1000 individuals plotted against months for G1. The simulation used to make this graph with the 2017 model yielded 930 infected in the core group (G2-G4). G1 had 9 infected MSM and in G5 7764 infected. Figure A.7 on page 81 includes all the age groups' cumulative number.

gonorrhea transmission in the new subgroups would be less accurate. If all the data from MSIS, FHI and ECDC had the same age group composition, it would be intriguing to build a model with these additional age groups for the G5 for even more accurate network modeling.

4.4.2 The Second Factor: The Probability of Exposure to Infected

The probability of having intercourse is called the probability of exposure to infected, because being sexually active means that an individual can be exposed to the infection. The interval estimate tested for this factor for MSM in a specific age group is presented in Table 4.3 on the next page. The parameters for the final models were all constant with the exception of $\beta_{2,G5}$ that was increased from year to year.

In a study from the WHO Europe, between 12 % and 38 % had their first sexual experience before the age of 15, depending on the country [92], substantiating that defining a rate for G1 is difficult. Statistics on sexual debut and how many that are sexually active can be found in National Health Statistics Reports by the CDC from 2011 [80]. 75 % of individuals in the age between 15 and 24 years had shown sexual behaviour. 66 % between 18-19 years and 85 % individuals between 20-24 years had shown sexual behaviour. In the models G2 had an interval estimate of 80 % - 95 %, Table 4.3, and according to the 2011 report, these numbers are a little too high. This was due to not finding sufficient data

or research on this field, and to adjust the model to fit the number of infected individuals in a year. Furthermore, a reasonable assumption due to demographic changes within the last years allowing for easier contact via social media and dating applications, the contact number nowadays might be even higher than in 2011.

In the age group ranging from 25 to 44 around 95 % has had sexual behaviour, and of these 5.8 % same sex sexual behaviour [80, 81]. There is not much research on this field, so this report was used when characterizing the percentage rate of having sexual intercourse being in an age group. For G1, this was assumed to be very small, discussed in Section 4.4.1. The G3 and G4 age groups had an average of around 90% being sexually active [80], these numbers fit well with the intervals tested for in the models, Table 4.3. The G5 group had 60 % to 70 % probability of having intercourse in the models, this is a lower rate compared to the report [80]. Around 80 % of male individuals reported sexual behaviour in the age ranging between 44 and 72, and among that. This percentage is lower in the interval estimate tested in the models, see Table 4.3 because all people up to an age of above 100 years were taken into account.

The definition of β_2 could have been changed to additionally include the probability of a susceptible individual having intercourse with an infected. This would have expanded the definition of β making it more realistic. Moreover, the definitions for β are already approximates, making β more detailed including more factors from Figure 3.10 would make the simulation results better. At the same time there is a risk of increasing the standard deviation and errors in the modeling with more factors without good data to support the different factors.

Table 4.3: This table shows the interval estimates for the probability of being sexually active for an age group x ($\beta_{2,Gx}$). The five models are for the years 2014 to 2018.

Gr.	Interval Estimate	2014	2015	2016	2017	2018
$\beta_{2,G1}$	0.001	0.001	0.001	0.001	0.001	0.001
$\beta_{2,G2}$	0.8-0.95	0.8	0.8	0.8	0.8	0.8
$\beta_{2,G3}$	0.8-0.95	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G4}$	0.7-0.9	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G5}$	0.5-0.75	0.661	0.67	0.675	0.685	0.69

4.4.3 The Third Factor - The Transmission Probability

The probability of getting gonorrhea after one intercourse has been defined as 60 % and 25 % for WSM and MSW through vaginal intercourse, respectively [66, 93]. The probability of contracting gonorrhea after more than four exposures to vaginal intercourse increases up till 60 % - 80 % for men [71]. The probability of contracting gonorrhea is larger for MSM than for MSW [94]. The difference in the presentation of symptoms between MSM and heterosexual men is due to the location of infection. Rectal and oral infections are largely asymptomatic resulting in delays in seeking treatment, the increased

risk of serious health problems such as pneumonia and not to mention the prolonged risk of transmitting the infection. MSM were in one study found to be 2.6 times as likely as WSM who had unprotected vaginal intercourse to report an STD diagnosis [95]. Hence it was assumed that the probability of contracting gonorrhea for MSM is at 60 %, the same probability as for WSM.

4.5 The Recovery Rate

The exact period of infectiousness is not given for gonorrhea, but this value could be estimated by making some assumptions. First, having a symptomatic or asymptomatic infection will determine whether or not an individual will seek medical help and get antibiotic treatment. The time it usually takes from an individual gets the infection until the individual is symptomatic, could be between three and 14 days [96, 71]. 24 hours after the started treatment the individual is no longer infectious [97]. Only 45 % of MSM experience symptoms. α in the models is a little higher than between three and 14 days due to many individuals experiencing an asymptomatic infection and will have the infection longer before testing voluntarily at a clinic or getting notified from a infected partner [16]. It was also assumed that the individuals no longer were sexually active during treatment. Including the asymptomatic individuals, who may or may not discover the infection throughout regularly STI testing within the simulation year, the recovery period was determined to 23 days.

Another study on a restricted network has used 55 days as α^{-1} for MSW and WSM [46, 48]. In this thesis MSM were modeled, and they were modeled with a network extending across the whole world, therefore the numbers from that study were not used. Moreover, using an α^{-1} that high for the models in this thesis would yield unrealistically high infection numbers, see simulation results done with higher α^{-1} in Appenidix B.5. The differentiation of α between symptomatic and asymptomatic women and men have been discussed [48]. The study states that α is very different for individuals with or without symptoms, and it is challenging to adjudicate a correct value for the parameter. Further it would be of interest to build a model with two values for the recovery rate and be able to predict the number of individuals who are without symptoms after one year. This number could be compared to the number of individuals who voluntarily go to the clinic to be tested without having symptoms [16]. If the simulated number is much higher than the reported, campaigns for more frequent testing, or other engaging work like lectures in schools or work places and commercials for condom use should be carefully followed up. In Norway now, Helsedirektoratet has a campaign encouraging all individuals to always use a condom based on the high number of gonorrhea cases in Norway, especially among MSM [98].

4.6 The Basic Reproduction Number

The basic reproduction number R_0, also called the number of secondary cases, has not been defined for the gonorrhea infections, and is generally difficult to define for STIs [34]. However, gonorrhea would be thought to have an $R_0 > 1$ because it always is present in the core group (G2-G4) [18, 82]. In epidemic network modeling it is most common to calculate R_0 considering various components of the infection [99, 48]. The basic reproduction number was in one study of gonorrhea calculated to be 1.01 for repeaters (individuals who got the infection more than once), and a little lower for non-repeaters [46]. The R_0 suggests that the spread is very low, but the outbreak will reach an endemic state [46]. However, that data was retrieved from a Manitoba community in Canada, a restricted subpopulation, or network. The models in this work consider traveling as one of the main infection sites, hence there will be more spread in the models, which results in higher R_0 values.

The factors considered in finding R_0 in this study were limited to the probability of interacting with other individuals considering the different age groups, the probability of having intercourse, the transmission probability and the recovery period. These made out β and α, and R_0 was found by applying Eq. 2.12. The values for R_0 from the 2017 model in Table 4.4 show, that for some age groups the infection outbreak will reach an endemic state (β>α), others however a disease free state (α>β. Using the complex network in GLEaM along with the age compartments, it is not only correct to state that the infection either will be endemic or disease free, the situation is much more complex.

G1 interacting with all other groups and the G2 - G4, G2 - G5 and G3 - G5 interactions, have an $\alpha > \beta$, thus $R_0 < 1$, resulting in a disease free state when $t \to \infty$. This observation reflects reality well, since spread of gonorrhea in G1 is highly uncommon. This also substantiates that β and α might be good parameter values chosen for the spread of gonorrhea in the core group. Because of the disease free state that G1 theoretically results in, it is not taken into consideration when counting for the total number of infected individuals in a simulation year.

For all other group interactions, $\alpha > \beta$ and mostly $R_0 > 1$ resulting in an endemic state. Looking at Table 4.4, in the core group, values range from about 1 to 8.3 secondary infections per infected individual. Although it is realistic that there will be higher reproduction numbers for the sexually active individuals in groups G2-G5, the credibility of $R_{0,G5}$ could be questioned. With the parameter combinations given in the 2017 model, an individual in G5 interacting with other G5 individuals will cause 7.6 additional infectious individuals. In a large group like this, the spreading will be enormous, which is reflected in the simulations. As seen in Figure A.7, the spreading in G5 is not realistic compared to gonorrhea statistics, and R_0 should have been lowered by decreasing β and at the same time divide G5 into smaller subgroups in future models to be built. Since the population in the other groups are regulated, R_0 being rather large does not have the same impact despite more infected individuals per hub. It is also important no notice, that the number of infected only is counted in Norway for G2-G4, where there are initial seeds. Moreover it should also be noticed that some of the spreading in G2-G4 is from G5. This is the reason that it was important to include that group in the model, to reflect reality because there are

interactions between the groups. Values for R_0 in the core group, Table 4.4 are considered to be good estimates as the models simulates number of infected MSM that reflect reality well.

A weakness in the modeling STIs with the SIS epidemic model on a network, is that interactions between individuals are assumed to be random [34]. Sex is not random, but depends on which age group, social class or ethnic group the sexual partner is in. It could also depend on the number of sexual partners an individual has had [34]. The β_1 parameter gives the interaction probability of having intercourse with any age group, but even within the age groups there are variations. Some have 1-4 sexual partners in their life time, others more than one hundred [34]. Some have partnerships lasting the whole life others always have shorter relationships. All these factors should be accounted for in β in order to optimize R_0 and fit the STI network modeling better.

Table 4.4: Basic reproduction number R_0 for gonorrhea in the MSM group based on the 2017 model. R_0 tells how many susceptible individuals one infectious individual will infect. The columns and rows represent the interactions between the different age groups. R_0 was calculated from Eq. 2.12, with β values from Table B.3 on page 84 and α^{-1} of 23 days.

	$R_{0,G1}$	$R_{0,G2}$	$R_{0,G3}$	$R_{0,G4}$	$R_{0,G5}$
$R_{0,G1}$	0.0137448	0.0000138	0.0000138	0.0000138	0.0000138
$R_{0,G2}$	0.01104	8.28	1.8768	0.828	0.04416
$R_{0,G3}$	0.01242	2.1114	7.452	2.2356	0.60858
$R_{0,G4}$	0.01242	0.9315	2.2356	7.452	1.78848
$R_{0,G5}$	0.009453	0.037812	0.463197	1.361232	7.581306

4.7 ANOVA

There were many variables in the models, and it was observed that changing them had a different impact on the cumulative number of infected MSM. As a consequence, ANOVA was done on the parameters to explore their significance in the models.

4.7.1 ANOVA for The Probability of Exposure to Infected

All the models had the same combination of the β parameter except for $\beta 2, G5$. An ANOVA was done to check the significance of the β_2 parameters compared against the response variable being the total number of infected MSM, Table 4.5. Raw data for the analysis are in Table B.10 on page 90. Other parameters were constant as explained in Section 4.3.3. The β_2 values for all groups except for $\beta_2, G1$ was compared. Since this variable never was changed from 0.001, and the spread in G1 was minimal, it was thought to not have a significant impact, in hindsight this should have been included in the analyses.

In the ANOVA, adding the coefficient $beta_{2,G2}$ gave a significantly better model fit (p-value 0.00517) compared to the simpler model with only the overall mean as covariate 4.5. Adding $\beta_{2,G3}$, then $\beta_{2,G4}$ and then $\beta_{2,G5}$ also gave significantly better model fits than the previous. Running simulations, the major change was observed when $\beta_{2,G5}$ was changed. This led to an ANOVA changing the order of the parameters. Adding $\beta_{2,G5}$ first gave a significantly better model fit (p-value 0.0001681) compared with the overall mean as covariate, see Table 4.6. None of the other coefficients yielded better model fits when added. This meant that the model was only significantly improved by adding $\beta_{2,G5}$. The same was seen by looking at the summary Table 4.7, overall it was only $\beta_{2,G5}$ that gave a significantly better model fit including the coefficient. The model fit in summary was not significantly improved when adding the other coefficients.

Table 4.5: The n-way ANOVA table was given as a result of investigating the effects of parameters $\beta_{2,G2}$-$\beta_{2,G5}$. Raw data for the analysis are in Table B.10.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	*
$\beta_{2,G2}$	1	163370	163370	18.373	0.005170	**
$\beta_{2,G3}$	1	116016	116016	13.047	0.011203	*
$\beta_{2,G4}$	1	116913	116913	13.148	0.011017	*
$\beta_{2,G5}$	1	223772	223772	25.166	0.002412	**
Residuals	6	53352	8892			

Table 4.6: The n-way ANOVA table was given as a result of investigating the effects of parameters $\beta_{2,G2}$-$\beta_{2,G5}$. Raw data for the analysis in Table B.10.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	*
$\beta_{2,G5}$	1	609481	609481	68.5426	0.0001681	***
$\beta_{2,G4}$	1	7008	7008	0.7882	0.4088320	
$\beta_{2,G3}$	1	1469	1469	0.1651	0.6985565	
$\beta_{2,G2}$	1	2112	2112	0.2376	0.6432641	
Residuals	6	53352	8892			

Table 4.7: The n-way ANOVA summary table was given as a result of investigating the effects of parameters $\beta_{2,G2}$-$\beta_{2,G5}$. Raw data for the analysis in Table B.10.

	Estimate	Std. Error	t value	Pr(>|t|)	*
(Intercept)	-28839.2	6705.0	-4.301	0.00509	**
$\beta_{2,G2}$	-527.8	1083.0	-0.487	0.64326	
$\beta_{2,G3}$	336.9	984.4	0.342	0.74384	
$\beta_{2,G4}$	1010.1	1058.4	0.954	0.37674	
$\beta_{2,G5}$	42225.6	8417.3	5.017	0.00241	**

4.7.2 ANOVA for the Probability of Exposure to Infected, the Recovery Rate, Initial Seeds and the Interactions Between Age Groups

ANOVA for the probability exposure to infected, β_2, the number of initial seeds in Europe and α were performed. Raw data for the analyses are in Table B.11. The ANOVA show that adding the variables $\beta_{2,G5}$ and α^{-1} gave significantly better model fits than without, see Tables 4.8 - B.9. Adding other variables first, did not significantly make the model fit better. In Table 4.8, the variables $\beta_{2,G2}$ and seeds are not yielding a significantly better model (p-values > 0.05), and it could be thought that these variables could be excluded from the model. $\beta_{2,G2}$ was not significant when added first, hence the overall mean as covariate was a better model fit in that case. To verify these observations, the order of adding the variables was changed, and in Table 4.9 the seeds variable was added first. This gave a better model fit compared to the overall mean (p-value 0.0002519), then $\beta_{2,G2}$ was added and gave a significantly better model compared to the seeds variable and the overall mean. Adding the rest of the variables also gave better model fits. In the summary Table B.6 on page 88 in Appendix B.3, however, $\beta_{2,G5}$ and α^{-1} are the only coefficients giving significantly better model fits with the variable included in the model than without. Removing days from the model, Table B.7, the ANOVA revealed that the only variable making the model significantly better was $\beta_{2,G5}$. Changing the order of the variables again, Tables 4.10, B.8 and B.9 all show that variables $\beta_{2,G2}$ and $\beta_{2,G4}$ never yields a significantly improved model. That observation could be thought to be a product of age groups G2 and G4 containing less infected individuals compared to G3, and less individuals in total compared to G5, thus making the impact on the response variable, the cumulative number of infected individuals smaller.

The recovery rate was always a significant variable in the model as seen in the ANOVA. This could also be illustrated as in Figure 4.10, where it is possible to see a linear relationship between the number of infected individuals and the α^{-1}. This linearity reflects the influence α^{-1} has on the cumulative number of infected MSM. The same could be said for $\beta_{2,G5}$, Figure 4.11, however to a lesser extent than for α^{-1}. The linear relationship means that the variables are significant. This could also be seen experimentally directly in the simulations. Changing these variables have a huge impact on the final number of infected MSM.

Table 4.8: The n-way ANOVA table was given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
$\beta_{2,G2}$	1	85421	85421	2.9328	0.098258	.
$\beta_{2,G3}$	1	1960969	1960969	67.3275	8.231e-09	***
$\beta_{2,G4}$	1	302108	302108	10.3725	0.003323	**
$\beta_{2,G5}$	1	4276013	4276013	146.8117	1.989e-12	***
α^{-1}	1	6418586	6418586	220.3743	1.654e-14	***
seeds	1	40400	40400	1.3871	0.249174	
Residuals	27	786398	29126			

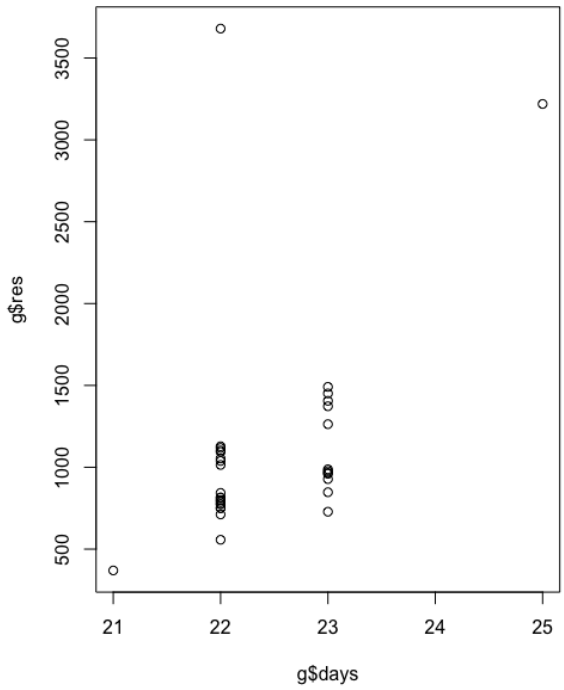


Figure 4.10: This figure shows the relationship between the cumulative number of infected MSM (g$res) plotted against α^{-1} (g$days).

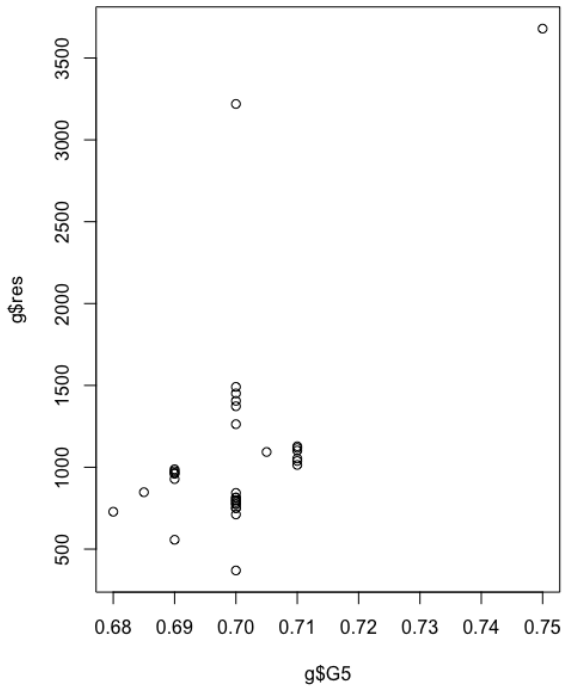


Figure 4.11: This figure shows the relationship between the cumulative number of infected MSM (g$res) plotted against $\beta_{2,G5}$ (g$G5).

Table 4.9: The n-way ANOVA table was given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
seeds	1	516778	516778	17.7429	0.0002519	***
$\beta_{2,G2}$	1	699231	699231	24.0073	3.988e-05	***
$\beta_{2,G3}$	1	952373	952373	32.6986	4.459e-06	***
$\beta_{2,G4}$	1	194108	194108	6.6645	0.0155848	*
$\beta_{2,G5}$	1	4412534	4412534	151.4990	1.385e-12	***
α^{-1}	1	6308474	6308474	216.5937	2.039e-14	***
Residuals	27	786398	29126			

Table 4.10: The n-way ANOVA table was given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
seeds	1	516778	516778	17.7429	0.0002519	***
α^{-1}	1	3708595	3708595	127.3301	1.000e-11	***
$\beta_{2,G2}$	1	85086	85086	2.9213	0.0988873	.
$\beta_{2,G3}$	1	411751	411751	14.1370	0.0008323	***
$\beta_{2,G4}$	1	9621	9621	0.3303	0.5702190	
$\beta_{2,G5}$	1	8351666	8351666	286.7443	6.607e-16	***
Residuals	27	786398	29126			

The β_{G1} parameter was also changed to experiment with the outcome of the simulations. Due to a not sufficient enough number of simulations, ANOVA could not be performed for this parameter. What was done, was plotting the change in β_{G1} from the original variable to the one that was used in all the simulations. In Figure 4.12, the x-axis represents the 25 different values for the interactions between the age groups, reading from G1-G1, G1-G2 etc. The y-axis represents the change in the value from the original to the one that is used in all the models in this thesis. The G2-G2, G2-G3, G5-G4 and G5-G5 are the values changed the most. The eight other values that are changed, have a change of between 0.005 and 0.03. The motivation for the change in interactions between the age groups was grounded in the fact that the G5 group stood for most of the spread. If the interactions with G5 and the other groups were to be reduced, would the total number of infected decrease? When many individuals in G5 are infected, an increased amount of individuals from G4 will be infected as a consequence of the interaction among those two groups. The result of the change did unfortunately not give any significant alteration in neither G4 nor G5, thus it was concluded that varying β_{G1} would not influence the total number of infected individuals in the groups G2-G4.

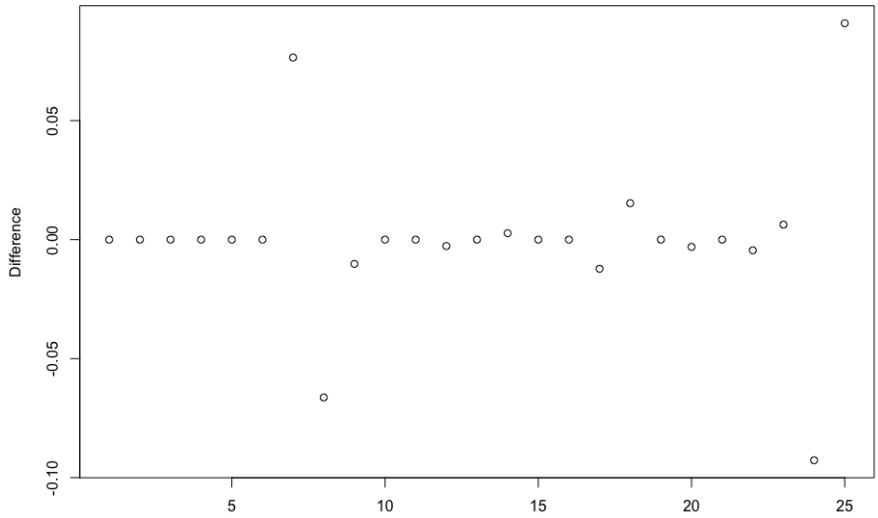


Figure 4.12: This figure shows the relationship between the difference in the variable β_1 plotted against all the different age group interactions.

4.8 The Transmission of Antibiotic Resistance in *N. gonorrhoeae*

For the modeling of AMR in *N. gonorrhoeae*, three models were generated. Two of them modeled for total resistance and one for partial resistance towards antibiotics in the bacteria. In the models, the same parameters as in the 2017 model were used, see Section 4.3.6.

4.8.1 The First AMR Model - Total Resistance One

Figure 4.13 on the following page shows the model simulating for total resistance. In this model, the SIS model is extended to a SIR model. That means, an additional node was introduced to the model, allowing for resistant individuals (R = resistant and not removed in these models). β between the compartments S and I and S and R is associated with probabilities. The associated probability between S and I had an interval estimate tested with 90 % - 99.999975 % and between S and I there was an associated probability of 0.000025 % - 10 %.

In one of the simulations, the probability of transferring from S to R was 10 % and S to I was 90 %. This resulted in the whole world getting an infection with resistant gonorrhea. When the probability of moving from S to R was set to 0.5 % (S to I was then 99.5 %), the

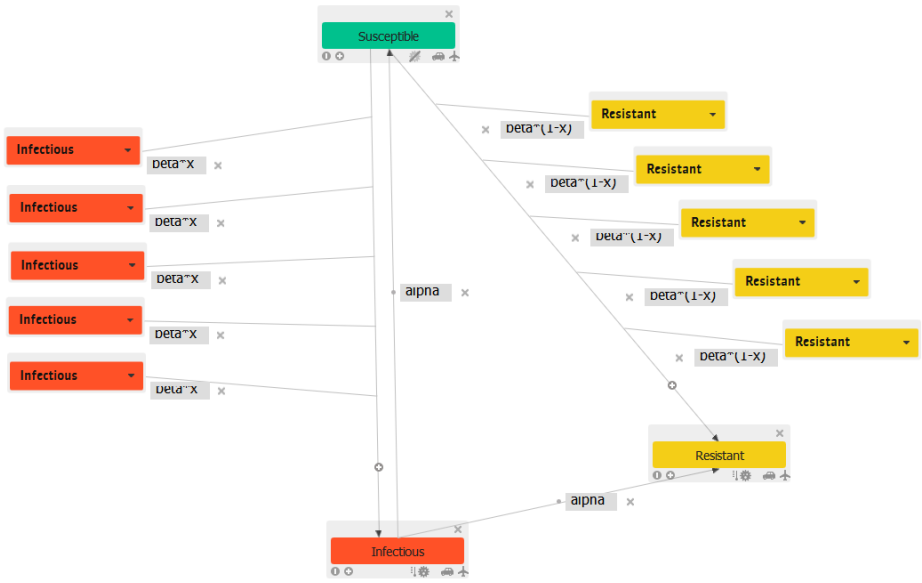


Figure 4.13: The first AMR model modeling for total resistance. The model is a SIR model where R is the resistant compartment. Infectious individuals of either the infectious or the resistant compartments can transmit the infection to susceptible individuals. Individuals can enter the susceptible or the resistant compartment from the infectious compartment in a rate α.

situation changed. Then the maximum cumulative number of infected MSM was almost 100. Considering that this is a model modeling for total resistance, it is unlikely that this number of infected MSM is realistic, since no cases of total resistant bacteria have been found in Norway yet [10, 1].

Because of the high number of cases with resistant bacteria, more simulations were performed with a lower probability associated between S and R of 0.25 %. This yielded the same number as with 0.5 %. Another attempt with a probability of 0.00025 % yielded the same result. It could thus be concluded that all resistant individuals were moved from I to R and not S to R and it could be argued that this model was not suitable for modeling resistance in gonorrhea in Norway. One possible solution could be to decrease α to lower the amount of infected individuals, with for example differentiating between symptomatic and asymptomatic individuals [48].

4.8.2 The Second AMR Model - Partly Resistance

The second model is represented in Figure 4.14 on the next page. In this model individuals could recover from having resistant gonorrhea, which means that there is no total resistant bacteria, only partly resistant bacteria. This is a more realistic view on the global situation. The difference from the first model in Figure 4.13 and this model, Figure 4.14 is α from

S to R, that was set to 23^{-1} + 14^{-1}. After 14 days, the duration of one antibiotic treatment for a gonorrhea infection, patients have to get re-tested. If patients are not completely healthy, another antibiotic cure has to be completed, in order to detect any remaining bacteria to prevent the development of AMR. Since few individuals experience failed treatment

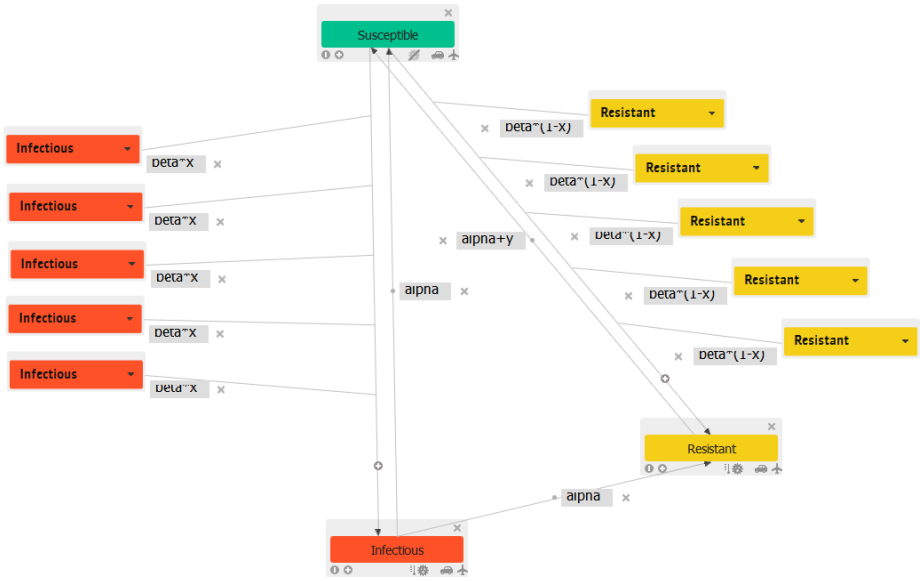


Figure 4.14: The second AMR model modeling for partial resistance. The model is a SIR model where R is the resistant compartment. Infectious individuals of either the infectious or the resistant compartments can transmit the infection to susceptible individuals. Individuals can enter the susceptible or the resistant compartment from the infectious compartment in a rate α and resistant individuals can enter the susceptible compartment in a rate $\alpha \cdot y$.

and develop resistant bacteria, the probability of infecting a susceptible individual is very small. Guidelines for which antibiotics to give to gonorrhea patients are carefully made to prevent the fast development of AMR [21, 25]. In Norway in 2013, ceftriaxone was the new standard antibiotic [69]. In 2016, The Norwegian Surveillance System for Antimicrobial Drug Resistance (NORM) reported reduced susceptibility in gonorrhea from 381 patients. The susceptibility for penicillin G had decreased with 97.1 %, and there was a 48.6 % increased resistance for ciprofloxacin to mention two examples [1].

In the models a probability of 0.25 % was used from S to R. The cumulative number of resistant individuals retrieved was then similar to the first AMR model, with 100 individuals infected. This increased a little bit when $\alpha + y$ was increased to 37 (y=14) days from 23 days. Considering different types of antibiotics and the amount of resistance related to it from Table 2.1, the model could be used to predict the number of gonorrhea cases in the near future. *N. gonorrhoeae* is more than 50 % resistant to ciprofloxacin [1]. Thus the expected number of resistant individuals in Norway would be 905x0.5 if ciprofloxacin was used as standard treatment modeling with the 2017 model. Predicting the future could be

done using the models from 4.3, and data from [1]. For ceftriaxone, which the gonorrhea bacteria presently is 100 % susceptible to, it would be difficult to measure the future probable resistant cases [10]. To further investigate this model, changing β, α and applying data on the development of AMR to explore the transmission in Norway could be possible further work.

4.8.3 The Third AMR Model - Total Resistance Two

The third model, Figure 4.15 on the facing page included an additional compartment compared to the other AMR models: the exposed compartment E, which makes it an SIERS model. In the exposed compartment, individuals cannot spread the infection nor get infectious before entering either the susceptible or the resistant compartment. β associated with S to I and S to R are equal to values in Table B.3 and the probabilities associated with them are 99.75 % and 0.25 %, respectively. In the latent phase, individuals recover after an average period of 23 days and are then moved to E. From there individuals are either moved to S or R compartments in rates of α equal to 1^{-1} or 7^{-1}, respectively. The latent period was implemented with the aim of reducing the number of individuals moved to the R compartment, as this was experienced at a rather high rate in the second AMR model, Figure 4.13. However, the amount of resistant cases was still found to be unrealistically high with a cumulative number of over 2000 individuals.

The attempt of making the exposed compartment to lead more individuals to the susceptible compartment than to the resistant failed. Running it with α_1=23^{-1}, α_2=24^{-1} and α_3=30^{-1} made an extreme difference to the outcome compared to the second AMR model. Around 2,500 were infected. The reason for this must be that there is no option for recovering after getting the resistant bacteria in the model. This model should definitely be developed to model for partial resistance, meaning that there should have been a fourth recovery rate from R to S.

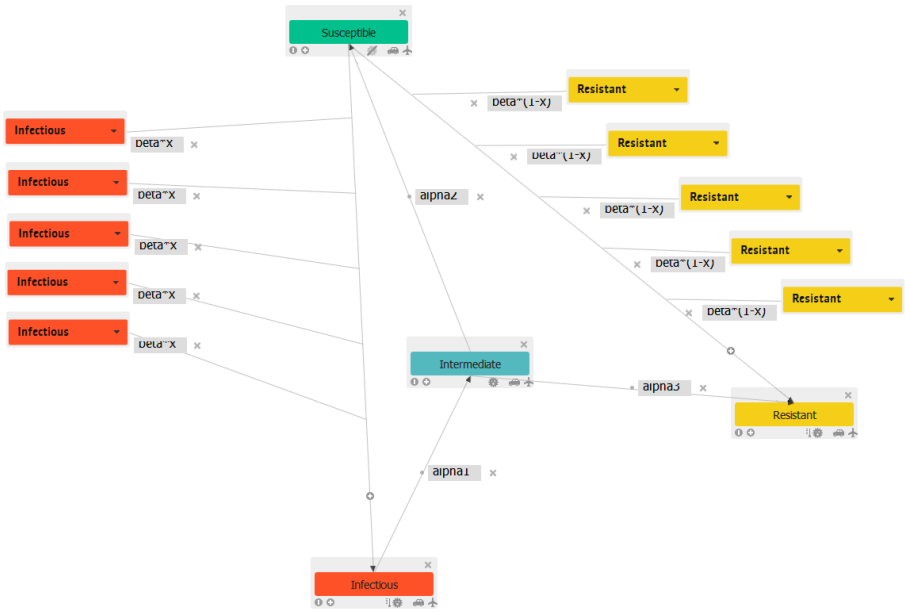


Figure 4.15: The third AMR model modeling for total resistance. The model is a SIERS model where R is the resistant compartment and E is the exposed compartment. Infectious individuals of either the infectious or the resistant compartments can transmit the infection to susceptible individuals. Individuals can enter the susceptible or the resistant compartment from the infectious compartment in a rate α.

Chapter 5

Conclusion and Further Work

In this work, the transmission of gonorrhea among MSM and the transmission of AMR in *N. gonorrhoeae* in Norway was modeled using epidemic network models with the GLEaMviz tool. Models for the prediction of the transmission of gonorrhea for MSM in Norway between 2014 and 2017 were successfully made. With the SIS epidemic network model, some key characteristics of the transmission of gonorrhea were captured. The models simulated the total cumulative number of infected MSM in Norway with an average discrepancy of 3.7 % from the reports for the respective years. The only parameter that was modified for all models was $\beta_{2,G5}$ with an average increase of 1.2 % over the years. It was using this information, the 2018 model was built and refined. The simulations performed using this model predicted on average 1039 infected, an increase of 14.7 % of the total amount of gonorrheal infections among MSM in Norway compared to 2017. If gonorrhea continues to increase like FHI predicts, the 2018 model could be used to predict the number of infected MSM in the core age group.

The parameters in the models were the transmission rate β, the recovery rate α and the number of seeds for the hubs with initially infected individuals in Norway and Europe. β was composed of three main factors in this thesis: The interactions between age groups, the probability of exposure to infected individuals and the transmission probability. The combination of parameters allowed for modeling the spread of gonorrhea. It was by changing these parameters that the optimal combination used in the models were found. The parameters were held constant in all the models, with an exception of the probability of exposure to infected MSM ($\beta_{2,G5}$) for age group five (G5). Performing an ANOVA on the variables, it was found that $\beta_{2,G5}$ made the model fit significantly better compared to the overall mean and to the other β_2 variables as covariate. Including initial seeds into the hubs in Europe was important to imitate the transmission of gonorrhea in Norway. STIs are always present in populations, and because of the air transportation and commuting networks that GLEAMviz provides, the simulations for spreading would be more realistic with initially infected in Europe. These results indicated that with the right use of param-

eters, it was possible to imitate the spreading of gonorrhea among MSM in Norway for several years.

For future work, it would be intriguing to investigate the possibility of making a sexual interaction network for the infected MSM in Norway. If FHI and MSIS provide sufficient data on infected individuals, it would be feasible to make a more detailed model. Understanding the sexual contact patterns in the MSM subpopulation might provide health authorities with better tools to help manage infection risk and treatment.

Additionally, future work could include even more factors in the transmission rate. For example the frequency of susceptible individuals having contact with infected individuals, the frequency of condom use or the duration of partnerships. The population could be divided into more groups, especially age group 5 (45+ years). This group could be split into for example six groups with ten years in each, for the transmission of gonorrhea to be more realistic in this segment. Another change could be to divide the infectious compartment into symptomatic and asymptomatic infection compartments. This would result in two more thorough recovery rates for further improvement of the model. Modeling the spread of AMR could have been optimized by merging the qualities of the partial resistance model and the second total resistance model together. This would result in a model where it was possible to recover from resistance, but the exposed compartment between the recovery from the infectious state would still be there. Due to time limitations, this was not performed, but it would be intriguing to build this model and see if this would make the transmission of AMR more realistic.

With this work, it was found that modeling an STI using the GLEaMviz tool is possible. With further development of the SIS epidemic model, adding more compartments, improving parameters, including more factors into the transmission rate and making two recovery rates, the transmission of gonorrhea and its network spreading dynamics could be further understood.

Bibliography

[1] NORM/NORM-VET. Norm/norm-vet 2016. usage of antimicrobial agents and occurrence of antimicrobial resistance in norway. tromsø / oslo 2017. ., 2016.

[2] https://www.populationpyramid.net/. Accessed: 09.03.18.

[3] The Norwegian Surveillance System for Communicable Diseases (MSIS). www.msis.no.

[4] World Health Organization. Who regional offices. http://www.who.int/about/regions/en/, 2018. Accessed: 06-03-2018.

[5] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge university press, 2016.

[6] Concept News Central. New research identifies possible new treatment for gonorrhea. http://conceptnewscentral.com/index.php/2017/06/27/new-research-identifies-possible-new-treatment-gonorrhea/, 2017.

[7] Wouter Van den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani. The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC infectious diseases*, 11(1):37, 2011.

[8] Figure of the three-layered model. http://www.gleamviz.org/model/. Accessed: 05.04.18.

[9] Gabriela B Gomez, Helen Ward, and Geoffrey P Garnett. Risk pathways for gonorrhea acquisition in sex workers: can we distinguish confounding from an exposure effect using a priori hypotheses? *The Journal of infectious diseases*, 210(suppl_2):S579–S585, 2014.

[10] Folkehelseinstituttet. Gonoré og syfilis i norge 2017. https://www.fhi.no/globalassets/dokumenterfiler/

seksuelt-overforbare-infeksjoner/gonore_syfilissituasjonen-2017.pdf, 2018. Accessed: 15-03-2018.

[11] World Health Organization. Who guidelines for the treatment of neisseria gonorrhoeae. http://apps.who.int/iris/bitstream/handle/10665/246114/9789241549691-eng.pdf?sequence=1, 2016.

[12] Manjula Lusti-Narasimhan, Francis Ndowa, and Susana Salgado Pires. Importance of sexually transmitted infections in funding for hiv within proposals to the global fund. *Sex Transm Infect*, 87(Suppl 2):ii19–ii22, 2011.

[13] Shannon R Galvin and Myron S Cohen. The role of sexually transmitted diseases in hiv transmission. *Nature Reviews Microbiology*, 2(1):33, 2004.

[14] James Wilton. Stis: What role do they play in hiv transmission? *Public Health*, 33(1,683):5–0, 2012.

[15] Jennifer L Edwards and Michael A Apicella. The molecular mechanisms used by neisseria gonorrhoeae to initiate infection differ between men and women. *Clinical microbiology reviews*, 17(4):965–981, 2004.

[16] Øivind. Folkehelseinstituttet Nilsen. Gonoré - epidemiologisk status i norge, olafiadagen 17. november 2017. , 2017.

[17] Karl E Miller. Diagnosis and treatment of neisseria gonorrhoeae infections. *American family physician*, 73(10), 2006.

[18] Centers for Disease Control and Prevention. Gonorrhea - cdc fact sheet (detailed version). https://www.cdc.gov/std/gonorrhea/stdfact-gonorrhea-detailed.htm, 2016.

[19] Eric J Klein, Larry S Fisher, Anthony W Chow, and Lucien B Guze. Anorectal gonococcal infection. *Annals of internal medicine*, 86(3):340–346, 1977.

[20] Paul J Wiesner, Evelyn Tronca, Paul Bonin, Alf HB Pedersen, and King K Holmes. Clinical spectrum of pharyngeal gonococcal infection. *New England Journal of Medicine*, 288(4):181–185, 1973.

[21] World Health Organization et al. Global action plan to control the spread and impact of antimicrobial resistance in neisseria gonorrhoeae. ., 2012.

[22] Robin. IFL Scienc.e Andrews. Man contracts worst known case of gonorrhea in the world. http://www.iflscience.com/health-and-medicine/man-contracts-worst-known-case-of-gonorrhea-in-the-world/, 2018.

[23] Rustam I Aminov. A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in microbiology*, 1:134, 2010.

[24] Teodora Wi, Monica M Lahra, Francis Ndowa, Manju Bala, Jo-Anne R Dillon, Pilar Ramon-Pardo, Sergey R Eremin, Gail Bolan, and Magnus Unemo. Antimicrobial

resistance in neisseria gonorrhoeae: Global surveillance and a call for international collaborative action. *PLoS medicine*, 14(7):e1002344, 2017.

[25] World Health Organization et al. Global health sector strategy on sexually transmitted infections 2016-2021: toward ending stis. ., 2016.

[26] Robert D Kirkcaldy, Akbar Zaidi, Edward W Hook, King K Holmes, Olusegun Soge, Carlos Del Rio, Geraldine Hall, John Papp, Gail Bolan, and Hillard S Weinstock. Neisseria gonorrhoeae antimicrobial resistance among men who have sex with men and men who have sex exclusively with women: the gonococcal isolate surveillance project, 2005–2010. *Annals of internal medicine*, 158(5_Part_1):321–328, 2013.

[27] Christopher K Fairley, Jane S Hocking, Lei Zhang, and Eric PF Chow. Frequent transmission of gonorrhea in men who have sex with men. *Emerging infectious diseases*, 23(1):102, 2017.

[28] Harvard. Department of Symptions Biology. `https://sysbio.med.harvard.edu/`.

[29] Howard J Federoff and Lawrence O Gostin. Evolving from reductionism to holism: is there a future for systems medicine? *Jama*, 302(9):994–996, 2009.

[30] Prahlad T Ram, John Mendelsohn, and Gordon B Mills. Bioinformatics and systems biology. *Molecular oncology*, 6(2):147–154, 2012.

[31] Olaf Wolkenhauer, Charles Auffray, Robert Jaster, Gustav Steinhoff, and Olaf Dammann. The road from systems biology to systems medicine. *Pediatric research*, 73(4-2):502, 2013.

[32] Duncan Ayers and Philip J Day. Systems medicine: the application of systems biology approaches for modern medical research and drug development. *Molecular biology international*, 2015, 2015.

[33] Clara Stegehuis, Remco van der Hofstad, and Johan SH van Leeuwaarden. Epidemic spreading on complex networks with community structures. *Scientific reports*, 6:29748, 2016.

[34] Fredrik Liljeros, Christofer R Edling, and Luis A Nunes Amaral. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and infection*, 5(2):189–196, 2003.

[35] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature physics*, 6(11):888, 2010.

[36] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[37] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global

epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, 2006.

[38] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC medicine*, 5(1):34, 2007.

[39] John W Glasser, Nathaniel Hupert, Mary M McCauley, and Richard Hatchett. Modeling and public health emergency responses: lessons from sars. *Epidemics*, 3(1):32–37, 2011.

[40] Paolo Bajardi, Chiara Poletto, Duygu Balcan, Hao Hu, Bruno Goncalves, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Modeling vaccination campaigns and the fall/winter 2009 activity of the new a (h1n1) influenza in the northern hemisphere. *Emerging Health Threats Journal*, 2(1):7093, 2009.

[41] Yuri A Amirkhanian. Social networks, sexual networks and hiv risk in men who have sex with men. *Current HIV/AIDS Reports*, 11(1):81–92, 2014.

[42] Richard B Rothenberg, John J Potterat, Donald E Woodhouse, Stephen Q Muth, William W Darrow, and Alden S Klovdahl. Social network dynamics and hiv transmission. *Aids*, 12(12):1529–1536, 1998.

[43] John J Potterat, L Phillips-Plummer, Stephen Q Muth, RB Rothenberg, DE Woodhouse, TS Maldonado-Long, HP Zimmerman, and JB Muth. Risk network structure in the early epidemic phase of hiv transmission in colorado springs. *Sexually transmitted infections*, 78(suppl 1):i159–i163, 2002.

[44] Hiroshi Nishiura. Correcting the actual reproduction number: a simple method to estimate r0 from early epidemic growth data. *International journal of environmental research and public health*, 7(1):291–302, 2010.

[45] P Van den Driessche and James Watmough. Further notes on the basic reproduction number. In *Mathematical Epidemiology*, pages 159–178. Springer, 2008.

[46] A M Jolly and J L Wylie. Gonorrhoea and chlamydia core groups and sexual networks in manitoba. *Sexually Transmitted Infections*, 78(suppl 1):i145–i151, 2002.

[47] Douglas T Fleming and Judith N Wasserheit. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of hiv infection. *Sexually transmitted infections*, 75(1):3–17, 1999.

[48] Wolfgang Alt and Gerhard Hoffmann. *Biological Motion: Proceedings of a Workshop Held in Königswinter, Germany, March 16–19, 1989*, volume 89. Springer Science & Business Media, 2013.

[49] Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Seasonal transmission potential and activity peaks of the new influenza a

(h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1):45, 2009.

[50] Annals of Iternal Medicine. High incidence of new sexually transmitted infections in the year following a sexually transmitted infection: A case for rescreening. http://annals.org/aim/fullarticle/729567/high-incidence-new-sexually-transmitted-infections-year-\following-sexually-transmitted, 2016.

[51] WilliamM McCormack, Kristine Johnson, RussellJ Stumacher, Allan Donner, and Rosalie Rychwalski. Clinical spectrum of gonococcal infection in women. *The Lancet*, 309(8023):1182–1185, 1977.

[52] Centers for Disease Control and Prevention. 2015 sexually transmitted diseases treatment guidelines. gonococcal infections. https://www.cdc.gov/std/tg2015/gonorrhea.htm, 2015. Accessed: 11.05.18.

[53] Statistics Solutions Advanced Through Clarity. Anova (analysis of variance). http://www.statisticssolutions.com/manova-analysis-anova/, 2013. Accessed: 10.05.18.

[54] Math Works. N-way anova. https://se.mathworks.com/help/stats/n-way-anova.html. Accessed: 11.05.18.

[55] Engineering Statistics Handbook. The anova table and tests of hypotheses about means. https://www.itl.nist.gov/div898/handbook/prc/section4/prc433.htm, 2012. Accessed: 06-03-2018.

[56] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. *Bulletin of mathematical biology*, 68(8):1893–1921, 2006.

[57] Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine*, 4(1):e13, 2007.

[58] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of computational science*, 1(3):132–145, 2010.

[59] http://sedac.ciesin.columbia.edu/data/collection/gpw-v4.

[60] Official Airline Guide. http://www.oag.com.

[61] International Air Transport Association. http://www.iata.org.

[62] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004.

[63] http://www.spato.net/about/.

[64] Gleamviz simulator client manual, version 6.6. wwww.gleamviz.org/simulator/client/, 2017. Accessed: 09.03.18.

[65] European Centre for Disease Prevention and Control. Disease data from ecdc surveillance atlas - gonorrhoea: Surveillance atlas of infectious diseases. https://ecdc.europa.eu/en/gonorrhoea/surveillance-and-disease-data/disease-data-atlas, 2017. Accessed: 15.02.18.

[66] Folkehelseinstituttet. Gonoré og syfilis i norge 2016. https://www.fhi.no/globalassets/dokumenterfiler/tema/gonore-syfilis-hiv-klamydia/gonore_syfilissituasjonen-2016.pdf, 2018. Accessed: 15-01-2017.

[67] Folkehelseinstituttet. Gonoré og syfilis i norge 2015. https://www.fhi.no/globalassets/dokumenterfiler/tema/gonore-syfilis-hiv-klamydia/gonore-og-syfilis-2015_2.pdf, 2018. Accessed: 07-02-2016.

[68] Folkehelseinstituttet. Gonoré og syfilis i norge 2014. https://www.fhi.no/globalassets/dokumenterfiler/tema/gonore-syfilis-hiv-klamydia/gonore-syfilis-2014.pdf, 2015. Accessed: 07-02-2018.

[69] Folkehelseinstituttet. Gonoré og syfilis i norge 2013. https://www.fhi.no/globalassets/dokumenterfiler/tema/gonore-syfilis-hiv-klamydia/gonore-og-syfilis-i-norge-i-2013.pdf, 2014. Accessed: 07-02-2018.

[70] Juey-Shin L Lin, S Patrick Donegan, Timothy C Heeren, Marilyn Greenberg, Elizabeth E Flaherty, Ruth Haivanis, Xiao-Hong Su, Deborah Dean, Wilbert J Newhall, Joan S Knapp, et al. Transmission of chlamydia trachomatis and neisseria gonorrhoeae among men with urethritis and their female sex partners. *The Journal of infectious diseases*, 178(6):1707–1712, 1998.

[71] Centers for Disease Control and Prevention. Ready-to-use: Std curriculum for clinical educators. program and training branch, division of std prevention, cdc. https://www.cdc.gov/std/ready-to-use/gonorrhea/gonorrhea-notes-April-2013.pdf, 2014. Accessed: 07-02-2018.

[72] Yonatan H Grad, Edward Goldstein, Marc Lipsitch, and Peter J White. Improving control of antibiotic-resistant gonorrhea by integrating research agendas across disciplines: key questions arising from mathematical modeling. *The Journal of infectious diseases*, 213(6):883–890, 2015.

[73] William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proc. R. Soc. Lond. A*, 138(834):55–83, 1932.

[74] Geoff P Garnett. An introduction to mathematical models in sexually transmitted disease epidemiology. *Sexually transmitted infections*, 78(1):7–12, 2002.

[75] RY Rubinstein. Simulation and the monte carlo method. new york, ny, usa: John wiley&sons, 1981.

[76] John M Last. A dictionary of epidemiology. *J Epidemiol Community Health*, 47(5):430, 1993.

[77] David Spiegelhalter. *Sex by numbers: what statistics can tell us about sexual behaviour*. Profile books, 2015.

[78] `https://www.theguardian.com/society/2015/apr/05/10-per-cent-population-gay-alfred-kinsey-statistics`. Accessed: 09.03.18.

[79] Bruce Voeller. Some uses and abuses of the kinsey scale. *Homosexuality/heterosexuality: Concepts of sexual orientation*, pages 32–38, 1990.

[80] Anjani Chandra, Casey E Copen, and William D Mosher. Sexual behavior, sexual attraction, and sexual identity in the united states: Data from the 2006–2010 national survey of family growth. In *International handbook on the demography of sexuality*, pages 45–66. Springer, 2013.

[81] Beth. Vitals Skwarecki. Here's how much sex everybody is having. `https://vitals.lifehacker.com/here-s-how-much-sex-everybody-is-having-1795561168`, 2017. Accessed: 14.05.18.

[82] World Health Organization et al. Antibiotic-resistant gonorrhoea on the rise, new drugs needed. *Saudi Medical Journal*, 38(8):878–879, 2017.

[83] Christine E Kaestle, Donald E Morisky, and Dorothy J Wiley. Sexual intercourse and the age difference between adolescent females and their romantic partners. *Perspectives on sexual and reproductive health*, pages 304–309, 2002.

[84] Jennifer Manlove, Kristin Moore, Janet Liechty, Erum Ikramullah, and Sarah Cottingham. Sex between young teens and older individuals: A demographic portrait. *Child Trends Research Brief*, 2005.

[85] Statistic Brain Research Institute. Relationship and dating statistics. `https://www.statisticbrain.com/dating-relationship-stats/`, 2018. Accessed: 01.05.18.

[86] Justin Lehmiller and Christopher Agnew. May-december paradoxes: An exploration of age-gap relationships in western society. ., 2011.

[87] Andrew F Hayes. Age preferences for same-and opposite-sex partners. *The Journal of Social Psychology*, 135(2):125–133, 1995.

[88] Monica Boyd and Anne Li. May–december: Canadians in age-discrepant relationships. *Canadian Social Trends*, 70(11):29–33, 2003.

[89] Patricia A Cavazos-Rehg, Melissa J Krauss, Edward L Spitznagel, Mario Schootman, Kathleen K Bucholz, Jeffrey F Peipert, Vetta Sanders-Thompson, Linda B Cottler, and Laura Jean Bierut. Age of sexual debut among us adolescents. *Contraception*, 80(2):158–162, 2009.

[90] Joyce C Abma and GM Martinez. Sexual activity and contraceptive use among teenagers in the united states, 2011-2015. *National health statistics reports*, .(104):1–23, 2017.

[91] L Kann, T McManus, WA Harris, SL Shanklin, KH Flint, J Hawkins, and S Zaza. Youth risk behavior surveillance-united states, 2015. mmwr surveillance summaries, 65 suppl 6, 1-174, 2016.

[92] Nuno Loureiro et al. A snapshot of the health of young people in europe, 2009.

[93] Joseph H Blount. A new approach for gonorrhea epidemiology. *American journal of public health*, 62(5):710–712, 1972.

[94] Samuel M Jenness, Elizabeth M Begier, Alan Neaigus, Christopher S Murrill, Travis Wendel, and Holly Hagan. Unprotected anal intercourse and sexually transmitted diseases in high-risk heterosexual women. *American journal of public health*, 101(4):745–750, 2011.

[95] Centers for Disease Control and Prevention. Anal sex and hiv risk. `https://www.cdc.gov/hiv/risk/analsex.html`, 2016. Accessed: 02-06-2018.

[96] Charles B Whitlow. Bacterial sexually transmitted diseases. *Clinics in colon and rectal surgery*, 17(4):209, 2004.

[97] National Health Service Choices. How long will i be infectious after starting antibiotics? `https://www.nhs.uk/chq/Pages/2561.aspx?CategoryID=73&SubCategoryID=108`, 2017. Accessed: 31.05.18.

[98] NRK Ertesvåg, Oda Ruggesæter. Helsedirektoratet med oppsiktsvekkende kondomkampanje. `https://www.nrk.no/norge/helsedirektoratet-med-oppsiktsvekkende-kondomkampanje-1.14053238`, 2012. Accessed: 27-05-2018.

[99] Klaus Dietz. The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, 2(1):23–41, 1993.

Appendix A

Figures

A.1 Settings

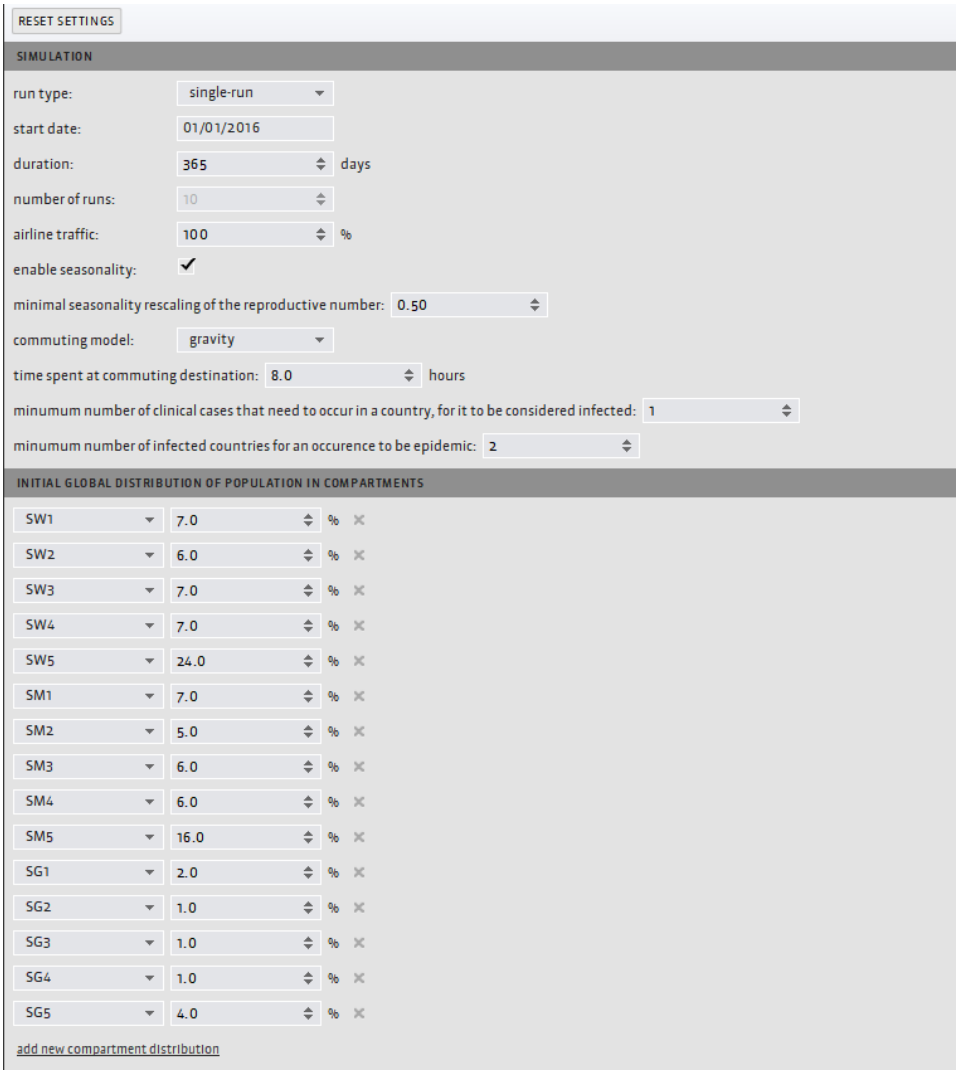


Figure A.1: There are multiple settings to define for each model or simulation. These includes the run-type, start-date for the outbreak, the duration of the outbreak, number of runs if multi-run is selected, airline traffic percentage, seasonality, type of commuting model, time spent on commuting destination, minimum number of clinical cases for a country to be considered infected and how many countries to be infected in order for it to be an epidemic. The distribution of 100 % susceptible individuals and the initial geographic location of the epidemic also have tp be decided. The result compartments for the investigation of the output results have to be chosen.

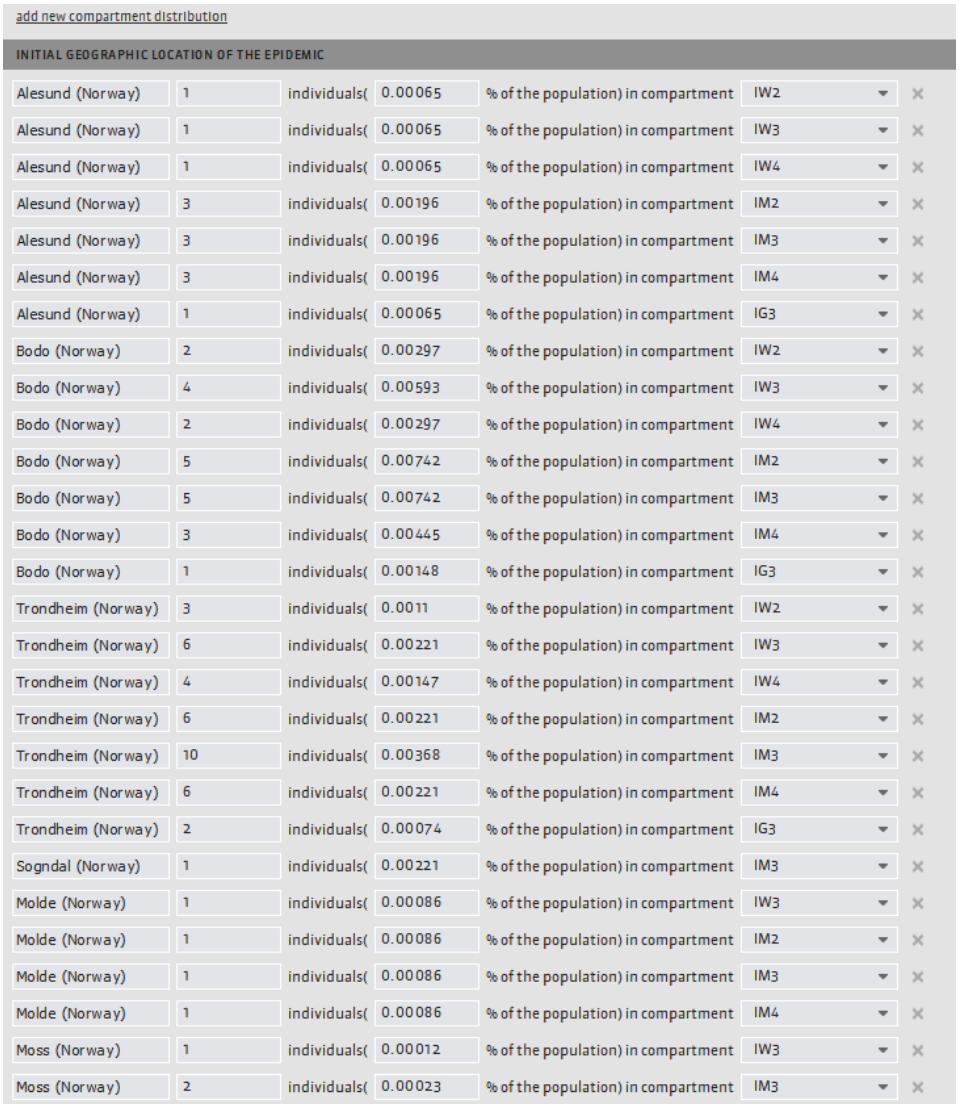


Figure A.2: There are multiple settings to define for each model or simulation. These includes the run-type, start-date for the outbreak, the duration of the outbreak, number of runs if multi-run is selected, airline traffic percentage, seasonality, type of commuting model, time spent on commuting destination, minimum number of clinical cases for a country to be considered infected and how many countries to be infected in order for it to be an epidemic. The distribution of 100 % susceptible individuals and the initial geographic location of the epidemic also have tp be decided. The result compartments for the investigation of the output results have to be chosen.

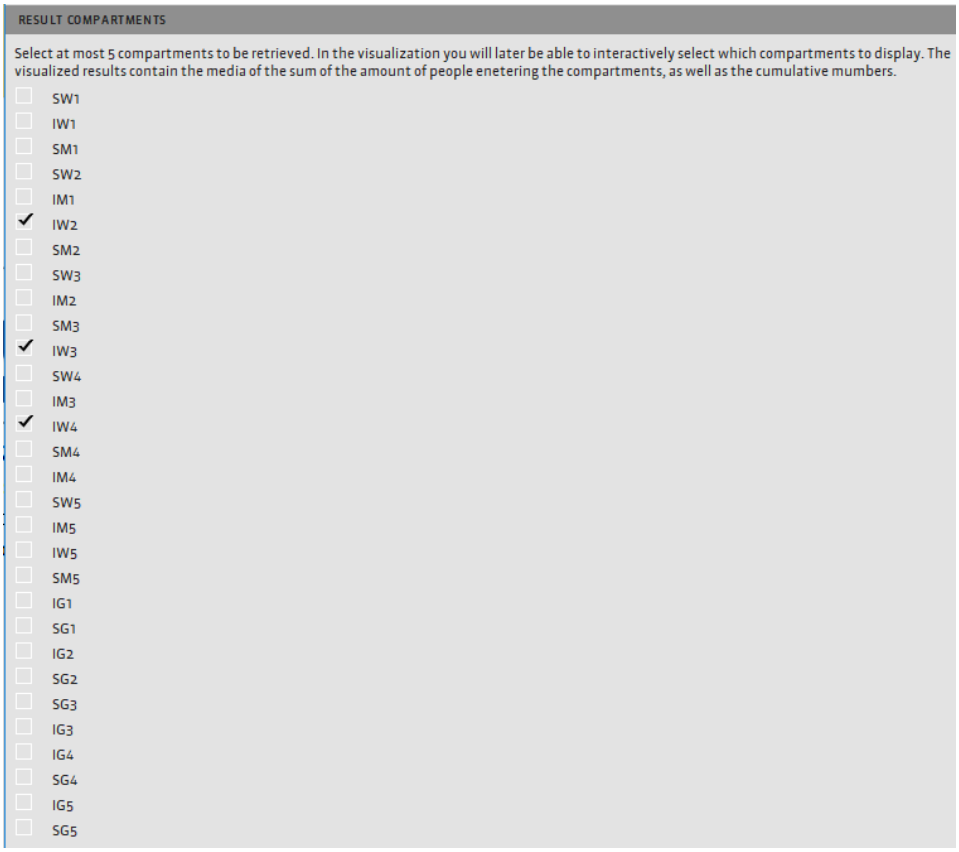


Figure A.3: There are multiple settings to define for each model or simulation. These includes the run-type, start-date for the outbreak, the duration of the outbreak, number of runs if multi-run is selected, airline traffic percentage, seasonality, type of commuting model, time spent on commuting destination, minimum number of clinical cases for a country to be considered infected and how many countries to be infected in order for it to be an epidemic. The distribution of 100 % susceptible individuals and the initial geographic location of the epidemic also have tp be decided. The result compartments for the investigation of the output results have to be chosen.

A.2 Visualization Interface

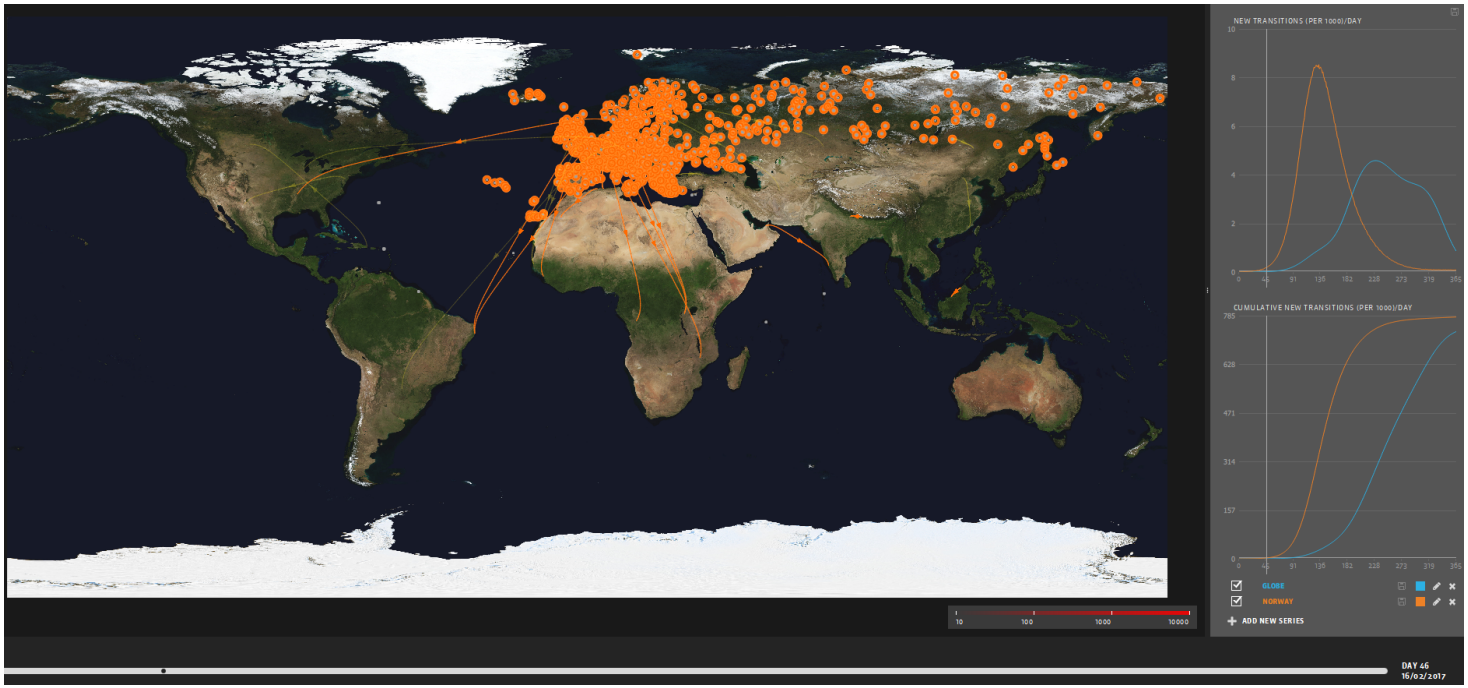


Figure A.4: The visualization interface in GLEaM. See Figure 3.5 for a thorough explanation.

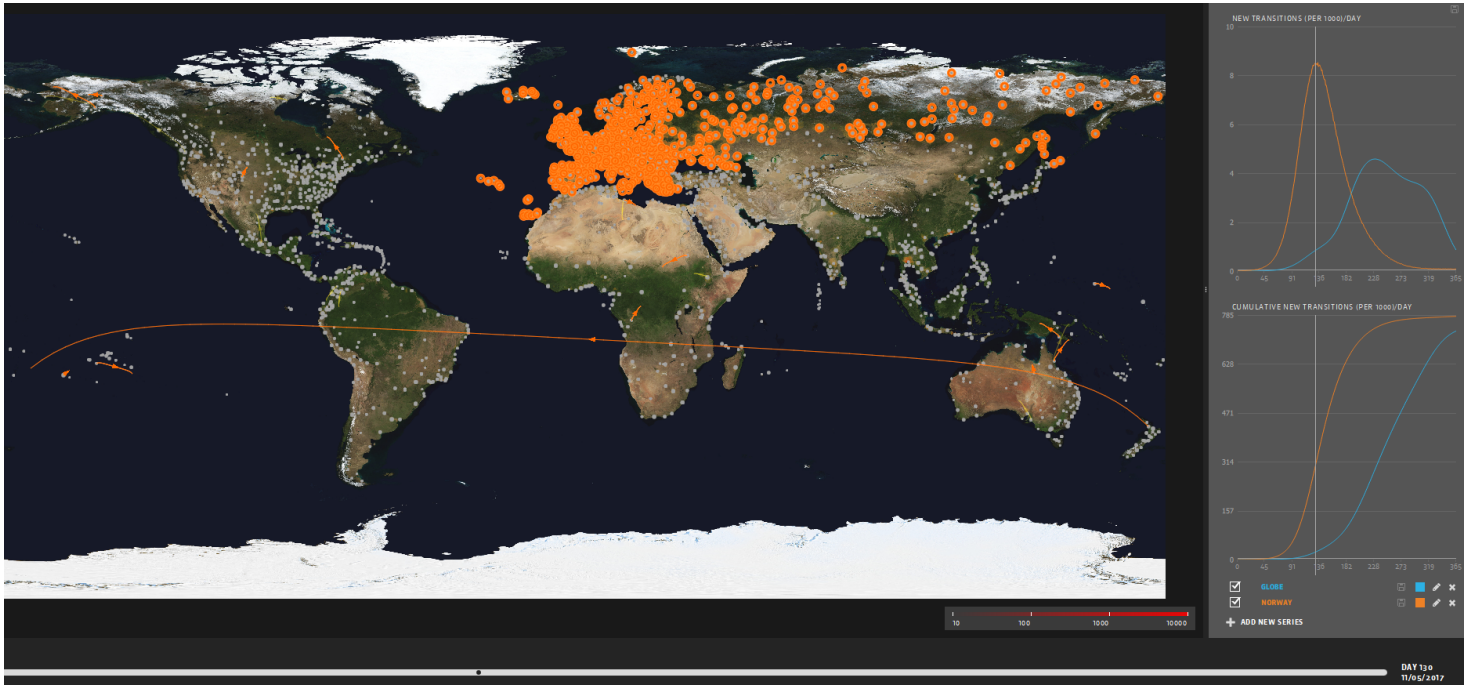


Figure A.5: The visualization interface in GLEaM. See Figure 3.5 for a thorough explanation.

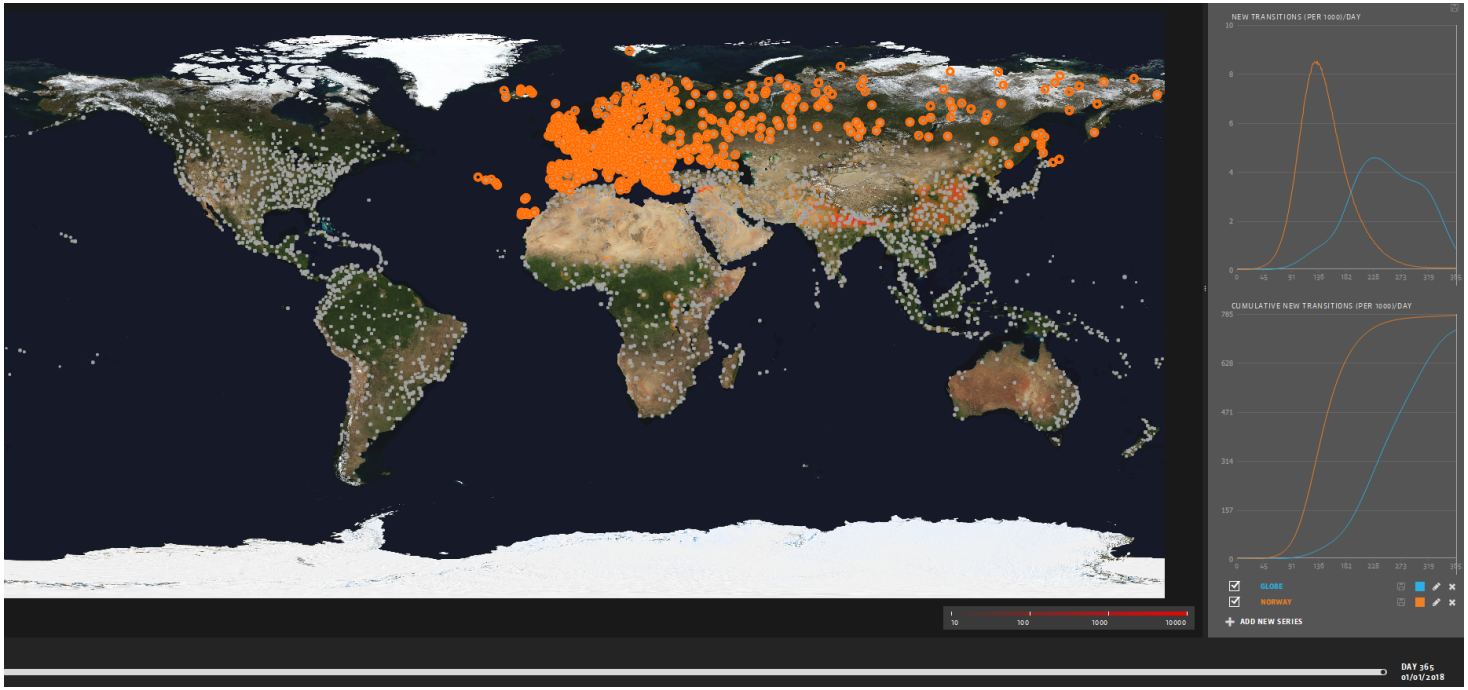


Figure A.6: The visualization interface in GLEaM. See Figure 3.5 for a thorough explanation.

A.3 Cumulative Number of Infected MSM per 1000 Individuals

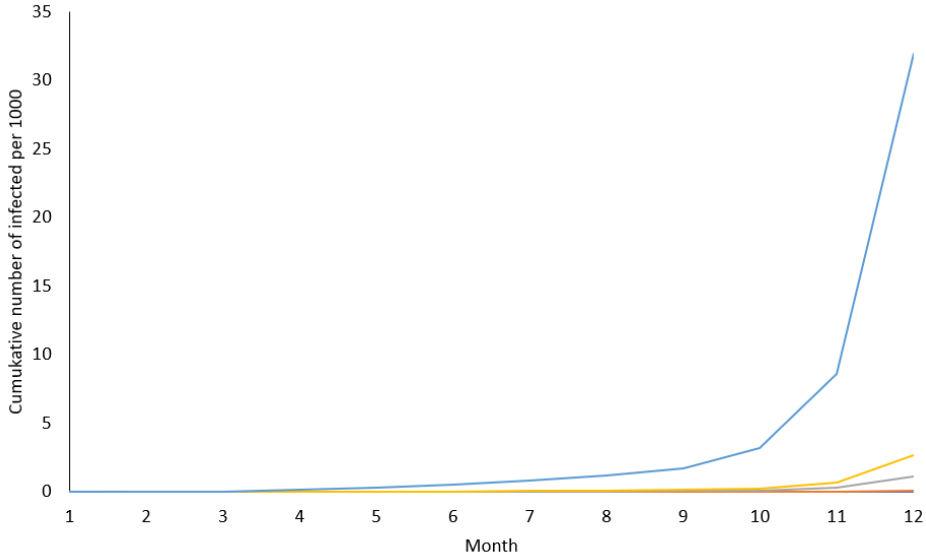


Figure A.7: A graph representing the cumulative number of infected MSM per 1000 individuals plotted against months for all the age groups G1-G5. The simulation used to make this graph with the 2017 model yielded 930 infected in the core group (G2-G4). G1 had 9 infected MSM and in G5 7764 infected. Figure A.7 includes all the age groups' cumulative number.

Appendix B

Tables

B.1 The Transmission Rate

Table B.1: Matrix representation of the transmission rate used in the 2015 model. The rows and columns show β of interactions between two individuals in any age group.

	β_{G1}	β_{G2}	β_{G3}	β_{G4}	β_{G5}
β_{G1}	0.0005976	0.0000006	0.0000006	0.0000006	0.0000006
β_{G2}	0.00048	0.36	0.0816	0.036	0.00192
β_{G3}	0.00054	0.0918	0.324	0.0972	0.02646
β_{G4}	0.00054	0.0405	0.0972	0.324	0.07776
β_{G5}	0.000402	0.001608	0.019698	0.057888	0.322404

Table B.2: Matrix representation of the transmission rate used in the 2016 model. The rows and columns show β of interactions between two individuals in any age group.

	β_{G1}	β_{G2}	β_{G3}	β_{G4}	β_{G5}
β_{G1}	0.0005976	0.0000006	0.0000006	0.0000006	0.0000006
β_{G2}	0.00048	0.36	0.0816	0.036	0.00192
β_{G3}	0.00054	0.0918	0.324	0.0972	0.02646
β_{G4}	0.00054	0.0405	0.0972	0.324	0.07776
β_{G5}	0.000405	0.00162	0.019845	0.05832	0.32481

Table B.3: Matrix representation of the transmission rate used in the 2017 model. The rows and columns show β of interactions between two individuals in any age group.

	β_{G1}	β_{G2}	β_{G3}	β_{G4}	β_{G5}
β_{G1}	0.0005976	0.0000006	0.0000006	0.0000006	0.0000006
β_{G2}	0.00048	0.36	0.0816	0.036	0.00192
β_{G3}	0.00054	0.0918	0.324	0.0972	0.02646
β_{G4}	0.00054	0.0405	0.0972	0.324	0.07776
β_{G5}	0.000411	0.001644	0.020139	0.059184	0.329622

B.2 Initial Seeds in Norway for All Models

Table B.4: An overview of where all individuals were placed in the 2013 model, the 2014 model and the 2015 model. The airports listed are all the transportation hubs in Norway in GLEAMviz.

	2013			2014			2015		
	G2	G3	G4	G2	G3	G4	G2	G3	G4
Stord (SRP)									
Hasvik (HAA)									
Brønnøysund (BNN)									
Haugesund (Hau)									
Kirkenes (KKN)									
Ålesund (AES)							1	2	1
Bodø (BOO)							1	2	1
Stavanger (SVG)	1	2		4	6	1	4	7	1
Lakselv (LKL)									
Trondheim (TRD)	2	2	1	5	6	2	5	6	3
Røros (RRS)									
Mosjøen (MJF)									
Sogndal (SOG)									
Røst (RET)									
Sørkjosen (SOJ)									
Vardø (VAW)									
Molde (MOL)									
Honningsvåg (HVG)									
Moss (RYG)	1	2		2	2		3	5	2
Leknes (LKN)									
Andenes (ANX)									
Kristiansand (KRS)					1		2	2	2
Hammerfest (HFT)									
Fagernes (VDB)							1	1	
Alta (ALF)									
Tromsø (TOS)				1	1		2	2	1
Mo i rana (MQN)									
Bergen (BGO)	1	2		4	7	1	8	8	6
Florø (FRO)									
Ørland (OLA)									
Harstad (EVE)									
Båtsfjord (BJF)									
Namsos (OSY)							1	1	
Oslo (OSL)	90	107	52	110	128	67	129	147	87
Bardufoss (BDU)									

Table B.5: An overview of where all individuals were placed in the 2016 model and the 2017 model. The airports listed are all the transportation hubs in Norway in GLEAMviz.

	2016			2017		
	G2	G3	G4	G2	G3	G4
Stord (SRP)						
Hasvik (HAA)						
Brønnøysund (BNN)						
Haugesund (Hau)						
Kirkenes (KKN)						
Ålesund (AES)	1	1		2	2	1
Bodø (BOO)		2		1	2	1
Stavanger (SVG)	5	7	3	7	8	5
Lakselv (LKL)						
Trondheim (TRD)	2	3	1	10	10	8
Røros (RRS)						
Mosjøen (MJF)						
Sogndal (SOG)						
Røst (RET)						
Sørkjosen (SOJ)						
Vardø (VAW)						
Molde (MOL)	1	1		2	2	1
Honningsvåg (HVG)						
Moss (RYG)	1	2		3	5	2
Leknes (LKN)						
Andenes (ANX)						
Kristiansand (KRS)	1	2		3	3	2
Hammerfest (HFT)						
Fagernes (VDB)		1		1	1	1
Alta (ALF)					1	
Tromsø (TOS)	1	2		1	2	1
Mo i rana (MQN)						
Bergen (BGO)	4	6	2	10	13	8
Florø (FRO)						
Ørland (OLA)						
Harstad (EVE)						
Båtsfjord (BJF)						
Namsos (OSY)		1			1	
Oslo (OSL)	174	211	163	261	301	223
Bardufoss (BDU)						

B.3 ANOVA Tables

Table B.6: The n-way ANOVA summary table is given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-5.151e+04	3.278e+03	-15.713	4.16e-15	***
$\beta_{2,G2}$	3.482e+02	8.708e+02	0.400	0.692	
$\beta_{2,G3}$	6.536e+01	1.034e+03	0.063	0.950	
$\beta_{2,G4}$	-5.607e+02	9.810e+02	-0.572	0.572	
$\beta_{2,G5}$	5.202e+04	3.072e+03	16.934	6.61e-16	***
α^{-1}	7.412e+02	5.036e+01	14.717	2.04e-14	***
seeds	-2.465e-03	2.093e-03	-1.178	0.249	

Table B.7: The n-way ANOVA summary table is given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.603e+04	6.552e+03	-2.447	0.020953	*
seeds	-4.745e-03	6.157e-03	-0.771	0.447331	
$\beta_{2,G2}$	-1.124e+03	2.551e+03	-0.441	0.662884	
$\beta_{2,G3}$	-3.327e+03	2.973e+03	-1.119	0.272672	
$\beta_{2,G4}$	-3.449e+03	2.835e+03	-1.217	0.233941	
$\beta_{2,G5}$	3.504e+04	8.398e+03	4.173	0.000264	***

Table B.8: The n-way ANOVA table is given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
seeds	1	516778	516778	17.7429	0.0002519	***
α^{-1}	1	3708595	3708595	127.3301	1.000e-11	***
$\beta_{2,G4}$	1	37287	37287	1.2802	0.2678140	
$\beta_{2,G3}$	1	453121	453121	15.5574	0.0005128	***
$\beta_{2,G2}$	1	16051	16051	0.5511	0.4642867	
$\beta_{2,G5}$	1	8351666	8351666	286.7443	6.607e-16	***
Residuals	27	786398	29126			

Table B.9: The n-way ANOVA table is given as a result of investigating the effects of parameters $\beta_{2,G2}$, $\beta_{2,G3}$, $\beta_{2,G4}$, $\beta_{2,G5}$, α^{-1} and the initial seeds. Raw data for the analysis in Table B.11.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
seeds	1	516778	516778	17.7429	0.0002519	***
α^{-1}	1	3708595	3708595	127.3301	1.000e-11	***
$\beta_{2,G4}$	1	37287	37287	1.2802	0.2678140	
$\beta_{2,G2}$	1	85333	85333	2.9298	0.0984239	.
$\beta_{2,G3}$	1	383839	383839	13.1786	0.0011668	**
$\beta_{2,G5}$	1	8351666	8351666	286.7443	6.607e-16	***
Residuals	27	786398	29126			

B.3.1 Raw Data for ANOVA

Table B.10: Table showing the raw material used in the ANOVAs displayed in Tables 4.5, 4.7 and 4.6. Int seeds are the initial seeds and Tot inf are the total number of infected individuals at the end of a simulation.

β_{2_G2}	β_{2_G3}	β_{2_G4}	β_{2_G5}	
0.85	0.85	0.8	0.7	1263.25
0.8	0.85	0.8	0.7	1405.88
0.85	0.9	0.8	0.7	1491.16
0.9	0.8	0.8	0.7	1317
0.8	0.8	0.8	0.7	1451.02
0.8	0.85	0.85	0.69	960.05
0.8	0.9	0.8	0.69	927.89
0.8	0.9	0.8	0.69	976.43
0.8	0.9	0.9	0.69	987.75
0.8	0.9	0.9	0.68	728.37
0.8	0.9	0.9	0.685	939.55

Table B.11: Table showing the raw material used in the ANOVAs displayed in Tables 4.8, B.6, 4.9, B.7, 4.10, B.8 and B.9. Int seeds are the initial seeds and Tot inf are the total number of infected individuals at the end of a simulation.

β_{2_G2}	β_{2_G3}	β_{2_G4}	β_{2_G5}	α^{-1}	Int seeds	Tot inf
0.85	0.9	0.85	0.75	31	200000	23643.45
0.85	0.9	0.8	0.7	28	200000	18863.35
0.85	0.85	0.8	0.7	28	200000	9172.05
0.85	0.85	0.8	0.7	28	150000	7366.5
0.85	0.85	0.8	0.7	21	150000	369.675
0.85	0.85	0.8	0.7	25	150000	3219.25
0.85	0.85	0.8	0.7	23	150000	1263.25
0.85	0.85	0.8	0.7	22	150000	774.34
0.8	0.9	0.9	0.69	23	100000	987.754
0.8	0.9	0.9	0.69	22	100000	557.2
0.8	0.9	0.9	0.7	22	100000	799.958
0.8	0.9	0.9	0.68	23	100000	728.37
0.8	0.9	0.9	0.71	22	100000	1038.45
0.8	0.9	0.9	0.685	23	100000	847.5445
0.8	0.9	0.9	0.705	22	100000	1093.46

B.4 Miscellaneous Tables

Table B.12: Table with data on how many cases of gonorrhoea that was reported in Norway every month in 2017 [3].

Month	# infected
January	909
February	725
March	769
April	800
May	785
June	763
July	891
August	1031
September	883
October	864
November	804
December	707

Table B.13: The regions listed below are all WHO regions. Which countries they include can be found on [4].

Region	# [million]
Western Pacific	32.5
South-East Asia	11.4
Africa	11.4
The Americas	11
Europe	4.7
Eastern Mediterranean	4.5

B.5 Experimental Data for all the Models

B.5.1 2014 Model - Experimental Data

Table B.14: Parameter values used in the modeling of the 2014 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6	Sim7
$\beta_{2,G2}$	0.8	0.8	0.8	0.8	0.8	0.8	0.8
$\beta_{2,G3}$	0.9	0.9	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G4}$	0.9	0.9	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G5}$	0.667	0.665	0.664	0.663	0.662	0.6619	0.66175
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	22	22	23	23	23	23	23
seeds	100000	100000	100000	100000	100000	100000	100000
	421.05	258.6	437.85	351.9	374.2	421.3	348.4
	557.1	387.9		590.4	478.05	342.55	356.55
		457.85		425.95	450.75		
		466.65		338	360.15		
		557.65		270.8			
				344.45			
		448.25					
Avg	489.08	429.48	437.85	386.92	415.79	381.93	352.48
StDev	96.20	99.90		111.21	57.52	55.68	5.76

Table B.15: Parameter values used in the modeling of the 2014 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim8	Sim9
$\beta_{2,G2}$	0.8	0.8
$\beta_{2,G3}$	0.9	0.9
$\beta_{2,G4}$	0.9	0.9
$\beta_{2,G5}$	0.661	0.66
β_3	0.6	0.6
α^{-1}	22	22
seeds	100000	100000
	296.8	296.9
	342.4	463.8
	257.45	266.3
	353.45	418.2
	317.9	308.65
	442.3	434.5
	442.9	352.7
	322.8	407.8
	421.4	358.15
	299.7	378.15
Avg	349.71	368.52
StDev	65.00	64.12

B.5.2 2015 Model - Experimental Data

Table B.16: Parameter values used in the modeling of the 2015 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim10	Sim11	Sim12
$\beta_{2,G2}$	0.8	0.8	0.8
$\beta_{2,G3}$	0.9	0.9	0.9
$\beta_{2,G4}$	0.9	0.9	0.9
$\beta_{2,G5}$	0.68	0.69	0.67
β_3	0.6	0.6	0.6
α^{-1}	22	22	23
seeds	100000	100000	100000
	344.6	370.95	418.4
	462.2	465.3	483.4
	395.45	440.9	537
	333.35	358	506.35
	413.65	325.7	438.45
			436.7
			506.75
			543.7
			394.55
			513.7
			559.35
			542.85
			489.55
			445.4
			661.4
			310.45
			524.3
			470.65
			689
			489.25
Avg	389.85	392.17	498.06
StDev	52.61	58.65	84.93

B.5.3 2016 Model - Experimental Data

Table B.17: Parameter values used in the modeling of the 2016 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim13	Sim14	Sim15	Sim16	Sim17	Sim18	Sim19
$\beta_{2,G2}$	0.85	0.9	0.8	0.8	0.8	0.8	0.8
$\beta_{2,G3}$	0.9	0.95	0.85	0.85	0.9	0.9	0.9
$\beta_{2,G4}$	0.8	0.85	0.85	0.85	0.9	0.9	0.9
$\beta_{2,G5}$	0.7	0.7	0.69	0.69	0.68	0.69	0.67
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	22	22	23	23	23	22	23
seeds	150000	150000	150000	100000	100000	100000	100000
	917.95	803.2	1055.15	767.25	833.25	763.45	542.75
	635.95	740.4	1022.2	862.8	847.55	1131.55	468.9
	907.95	702.35	812.65	951	967.85	855.9	513.85
	875.6	981.2	1053.65	1117.65	1279.6	866.2	469.15
	594.5	870.2	816.7	998.75	960	959.5	551.2
							459.45
							541.25
							430.8
							530.95
							408.1
Avg	786.39	819.47	952.07	939.49	977.65	915.32	491.64
StDev	157.72	110.64	126.12	133.17	179.83	139.39	50.99

Table B.18: Parameter values used in the modeling of the 2016 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim20	Sim21	Sim22	Sim23	Sim24
$\beta_{2,G2}$	0.8	0.8	0.8	0.8	0.8
$\beta_{2,G3}$	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G4}$	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G5}$	0.67	0.67	0.68	0.675	0.675
β_3	0.6	0.6	0.6	0.6	0.6
α^{-1}	22	24	22	23	23
seeds	100000	100000	100000	100000	100000
	233.65	1214.95	375.8	779.5	658.45
	239.7		327.4	502.05	577.7
	220		341.25	690.95	725.3
				535.4	678.95
				536.55	810.9
				420.75	443.15
				663.65	579.15
				761.35	415.35
				381	597.4
				601.4	516
				455	536.65
				488.3	588.55
				565.15	641.9
				520.45	464.85
				588.2	716.9
				374.55	476.25
				521.5	633.7
				634.9	632.7
				598	645.05
				649.1	597
Avg	231.12	1214.95	348.15	563.39	596.80
StDev	10.09		24.93	113.20	100.97

B.5.4 2017 Model - Experimental Data

Table B.19: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim25	Sim26	Sim27	Sim28	Sim29	Sim30	Sim31
$\beta_{2,G2}$	0.85	0.85	0.85	0.85	0.85	0.9	0.85
$\beta_{2,G3}$	0.9	0.9	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G4}$	0.85	0.85	0.9	0.85	0.85	0.9	0.85
$\beta_{2,G5}$	0.7	0.75	0.7	0.75	0.75	0.75	0.75
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	35	30	35	32	31	30	31
seeds	200000	200000	200000	200000	200000	200000	200000
	906.7	678.4	689.05	1417.85	960.15	818	23643.45
	816.15	762.75	630.5	1048.85	784.2	962.35	
	618.5	708.6	970.6	1835.4	857.1	550.75	
	759.95	760.45	915.7	1725.96	1028.3	600.3	
	652.6	757.1	1336	1518.7	814.85	613.1	
			1062.2		668.5	711.1	
			781.45		1015.25	723.85	
			711.9		944.6	628.8	
			942		1028.4	659.55	
			653.45		885.9	709.65	
			682.25				
			1090.7				
			772.75				
			1048.65				
			990.65				
Avg	750.78	733.46	885.19	1509.352	898.73	697.75	23643.45
StDev	118.12	38.06	203.15	305.67	119.31	120.17	

Table B.20: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim32	Sim33	Sim34	Sim35	Sim36	Sim37	Sim38
$\beta_{2,G2}$	0.85	0.85	0.9	0.85	0.85	0.85	0.85
$\beta_{2,G3}$	0.9	0.85	0.9	0.85	0.85	0.85	0.85
$\beta_{2,G4}$	0.8	0.8	0.9	0.8	0.8	0.8	0.8
$\beta_{2,G5}$	0.7	0.7	0.75	0.7	0.7	0.7	0.7
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	28	28	30	28	21	25	23
seeds	200000	200000	200000	150000	150000	150000	150000
	18863.35	9172.05	818	7366.5	377.05	3219.25	1200.85
			962.35		362.3		1210.15
			550.75				1374.55
			600.3				1347.45
			613.1				1183.25
			711.1				
			723.85				
			628.8				
			659.55				
			709.65				
Avg	18863.35	9172.05	697.75	7366.50	369.68	3219.25	1263.25
StDev			120.17		10.43		90.26

Table B.21: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim39	Sim40	Sim41	Sim42	Sim43	Sim44	Sim45
$\beta_{2,G2}$	0.85	0.85	0.8	0.85	0.85	0.9	0.9
$\beta_{2,G3}$	0.85	0.85	0.85	0.9	0.9	0.9	0.8
$\beta_{2,G4}$	0.8	0.8	0.8	0.8	0.8	0.8	0.8
$\beta_{2,G5}$	0.7	0.75	0.7	0.7	0.7	0.7	0.7
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	22	22	23	23	22	22	23
seeds	150000	150000	150000	150000	150000	150000	150000
	773.3	3679.5	1492.6	1414.65	735.6	1152.85	1766.55
	770		1364.55	1226.65	833.85	776.1	1236.65
	760.4		1445.4	1941.15	899.35	722.9	1265.6
	802.7		1362.4	1362.4	636.4	838.75	1277.45
	765.3		1364.45	1510.95	858.8	726.25	1317
					704.75		
					996.3		
					897.5		
					721		
					786.45		
					651.55		
					798.25		
					697.1		
					766.9		
					817.3		
Avg	774.34	3679.50	1405.88	1491.16	786.74	843.37	1372.65
StDev	16.59		59.99	271.70	100.11	179.27	222.08

Table B.22: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim46	Sim47	Sim48	Sim49	Sim50	Sim51	Sim52
$\beta_{2,G2}$	0.87	0.9	0.95	0.95	0.95	0.9	0.95
$\beta_{2,G3}$	0.9	0.95	0.9	0.95	0.9	0.95	0.95
$\beta_{2,G4}$	0.8	0.8	0.8	0.85	0.85	0.85	0.8
$\beta_{2,G5}$	0.7	0.7	0.7	0.7	0.7	0.7	0.7
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	22	22	22	22	22	22	22
seeds	150000	150000	150000	150000	150000	150000	150000
	770.2	887.9	859.8	835.25	729.05	1104.2	734.35
	754.85	818.7	558.8	703.55	667.3	810.85	871.65
	864.95	706.35	852.4	532.5	770.2	875.7	913.9
	788.25	968.4	774.15	787.55	861.4	670.85	865.05
	897.25	799.7	713.05	699.9	719.5	657.65	688.75
		805.1				638.45	
		782.8				878.75	
		643.8				634.85	
		690.05				698.25	
		867.6				782.95	
Avg	815.1	797.04	751.64	711.75	749.49	775.25	814.74
StDev	62.45	98.03	123.53	115.47	72.50	148.86	97.39

Table B.23: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim53	Sim54	Sim55	Sim56	Sim57	Sim58	Sim59
$\beta_{2,G2}$	0.9	0.8	0.9	0.85	0.8	0.8	0.8
$\beta_{2,G3}$	0.95	0.8	0.9	0.85	0.85	0.85	0.85
$\beta_{2,G4}$	0.85	0.8	0.85	0.85	0.8	0.85	0.85
$\beta_{2,G5}$	0.71	0.7	0.71	0.71	0.71	0.71	0.69
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	22	23	22	22	22	22	23
seeds	150000	150000	150000	150000	150000	150000	150000
	1035.1	1126.85	1150.3	1005.9	1127.7	894.35	771.8
	1175.95	1286.3	1102.85	971.75	934.65	1229.9	1042.8
	1291.6	1603.9	1316.7	1169.4	1014.95	1086.25	830.95
	1070.3	1585.2	897.5	918.15	1245.3	974.9	810.3
	1067.75	1652.85	1121.85	1158.3	1014.25	886.1	1136.8
				1027.4	977.65		1070
				914.1	1247.1		1213
				1136.9	1320.35		777
				1264.9	934.15		1138.8
				971.2	1205		809.05
Avg	1128.14	1451.02	1117.84	1053.8	1102.11	1014.3	960.05
StDev	105.66	231.48	149.49	120.28	144.32	144.94	175.42

Table B.24: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim60	Sim61	Sim62	Sim63	Sim64	Sim65	Sim66
$\beta_{2,G2}$	0.8	0.8	0.8	0.8	0.8	0.8	0.8
$\beta_{2,G3}$	0.9	0.9	0.9	0.9	0.9	0.9	0.9
$\beta_{2,G4}$	0.8	0.8	0.8	0.9	0.9	0.9	0.9
$\beta_{2,G5}$	0.69	0.69	0.69	0.69	0.7	0.68	0.71
β_3	0.6	0.6	0.6	0.6	0.6	0.6	0.6
α^{-1}	23	23	23	23	22	23	22
seeds	150000	100000	80000	100000	100000	100000	100000
	907.65	897.7	886.7	921.3	912.2	721.65	897.55
	974.2	831.65	981.75	921.2	839.4	666.9	901.6
	1104.25	1153.3	890.45	803.25	821.45	682.3	969.25
	669.25	1258.85	1036.9	1293.45	676.3	686.8	936.1
	1080.1	1023.15	1034.4	921.75	889.9	884.2	1153.7
	1101.7	961.35		918.75	1015.4		1254.65
	881.65	913.45		858.35	809.5		1145.15
	757.95	855.2		1018.95	901.8		864.55
	863.05	841.6		1032.75	767.45		1190.6
	939.05	1028.05		751.15	753.8		1071.35
				1132.5	784.6		
				1103.7	1128.45		
				951.3	936.2		
				1079.7	704.05		
				1064.65	726.5		
				1035.3	563.3		
				979.45	576.7		
				893.7	1008.2		
				1007.65	831		
				1015.1	767.85		
				801.4	745.55		
				833.25	742.3		
				988.95	757.2		
				1223	691.6		
				1143.3	648.25		
Avg	927.89	976.43	966.04	987.75	799.96	728.37	1038.45
StDev	145.02	141.47	74.08	132.56	134.39	89.38	141.31

Table B.25: Parameter values used in the modeling of the 2017 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim67	Sim68
$\beta_{2,G2}$	0.8	0.8
$\beta_{2,G3}$	0.9	0.9
$\beta_{2,G4}$	0.9	0.9
$\beta_{2,G5}$	0.705	0.685
β_3	0.6	0.6
α^{-1}	22	23
seeds	100000	100000
	1363.6	1181.45
	911.6	857.7
	1058.9	938.45
	1071.15	750.9
	1062.05	852.5
		817.7
		887.8
		756
		659.8
		999.1
		875.45
		781.35
		721.35
		796.15
		869.35
		920.95
		1000.6
		843.7
		922.1
		1383.65
Avg	1093.46	890.80
StDev	164.87	163.34

B.5.5 2018 Model - Experimental Data

Table B.26: Parameter values used in the modeling of the 2018 model. β is the transmission rate for the different age groups G2-G5 and α^{-1} is the recovery period, seeds the number of infectious individuals initially in the model. A number of between 0 and 20 simulations were done for each parameter combination. Avg is the average number of infected MSM and StDev is the standard deviation from the simulations.

	Sim69	Sim70	Sim71
$\beta_{2,G2}$	0.8	0.8	0.8
$\beta_{2,G3}$	0.9	0.9	0.9
$\beta_{2,G4}$	0.9	0.9	0.9
$\beta_{2,G5}$	0.695	0.69	0.7
β_3	0.6	0.6	0.6
α^{-1}	22	22	23
seeds	100000	100000	100000
	989.7	1118.55	2344.45
	1002.95	1160.5	2564.76
	1428.85	1038.65	1389.32
	1130.25	963.55	1956.43
	1427.6	1164.95	3109.23
	970	883.45	
	1442.55	934.15	
	1321.1	873.3	
	1497.3	807.55	
	1111.8	1091.3	
	1293.35		
Avg	1237.77	1003.60	2272.84
StDev	201.91	128.70	646.26