Life History Trade-offs in *Streptococcus pneumoniae*: Antibiotic Sensitivity and Vaccination

Martin Emons

Theoretical Biology Group
Institute of Integrative Biology
ETH Zürich
martin.emons@inf.ethz.ch

April 2021

Abstract

Antibiotic resistance in *Streptococcus pneumoniae* is a growing reason of concern in public health. In this paper, we investigate the dynamics of antibiotic sensitivity of *S. pneumoniae* in response to vaccination. In a SIS model, we find that a change in availability of hosts leads to transient domination of sensitive over resistant strains. As an extension to this framework, we set up a SISV model which incorporates vaccination explicitly. The SISV model supports our prediction of transient dominance of sensitive strains. This domination of sensitive strains can be seen as a transient shift towards higher infectiousness. Data from the Massachusetts pneumococcal vaccination campaign does not show the same trends as we predict with our models. This means that i) the data resolution is too poor and we couldn't see the trends we expect to see or ii) the model misses important aspects of vaccination.

1 Introduction

Streptococcus pneumoniae is a gram-positive bacterium that can colonise the upper respiratory tract asymptomatically. It can lead to infections with diseases such as bacterial pneumoniae, sepsis, and meningitis. The most severe infections are collectively called invasive pneumococcal disease (IPD)[1, 2]. Estimates on the worldwide death toll for children from 2009 range from 700'000 to 1'000'000 each year [3]. Especially high were fatalities in children aged 1-59 months prior to vaccination with an estimated 8-12% of all death cases. IPD is as well a problem for adults with an estimated death toll of $\geq 50'000$ deaths in the US each year [1, 2].

There are currently around 100 different serotypes known of *S. pneumoniae*. Because the manufacturing of conjugate vaccines is quite difficult, only a small fraction of all strains (7-13) are included in a vaccine [2]. This means we can distinguish the strains of *S. pneumoniae* in terms of being included in the vaccine (vaccine-type strains) or not being included in the vaccine (non-vaccine-type strains).

A growing concern for public health is antibiotic resistance in *S. pneumoniae*. The economic costs induced by multi-drug resistance are estimated to be around $ 100 trillion in the year 2050 in a worst-case estimate overall [4]. When trying to understand resistance, one tries to understand the trade-offs resistance implies. On the one hand, a resistant cell has a selective advantage since it

is not affected by antibiotic treatment. On the other hand, it has a selective disadvantage through any costs the resistance incurs. Costs of resistance affect fitness. Such a cost could be a decreased per capita growth rate [5].

The trade-off between reduced antibiotic-induced mortality and fitness cost can be seen as a variant of the duration of infection vs. infectiousness trade-off. Infectiousness and duration of infection are two life-history traits. These two traits are empirically negatively correlated, which can lead to the formulation as a trade-off. If e.g. the duration of the infection increases, the infectiousness decreases and vice versa [6]. This concept can be applied to antibiotic sensitivity. A strain sensitive to antibiotics is more infectious, as it does not have to invest resources in resistance. This lack of resistance on the other hand means that a sensitive strain can be more easily treated with antibiotics, which results in a decreased duration of infection. The resistant strains on the other hand can increase their duration of infection by evading antibiotic treatment. By investing resources in the resistance mechanism, they reduce their infectious potential [7, 8]. This means that the frequency of sensitive vs. resistant strains is a type of infectiousness vs. duration of infection trade-off. Which process is more beneficial in this trade-off is context-dependent [5].

One context in which the optimum of the trade-offs might change is the availability of hosts for the bacteria to infect. Vaccination is a process that changes the availability of hosts. This means that the trade-offs are expected to change as a result of vaccination. In 2000 the pneumococcal conjugate vaccine (PCV7) including 7 strains was approved by the FDA [9]. For some types (vaccine-types) the availability in host decreased whereas for other types it increased (non-vaccine-types). This will result in a decrease of vaccine-type strains and an increase in non-vaccine-type strains. This process is called strain replacement [10].

In the context of strain replacement, it is of interest to investigate the dynamics of antibiotic sensitivity. This leads to the main question of this report. How does the frequency in antibiotic sensitivity of *S. pneumoniae* change upon introduction of a vaccine? We address this question by modelling antibiotic resistance and vaccination. These models are verified with data from the Massachusetts pneumococcal vaccination campaign.

2 Results

2.1 SIS Model of Antibiotic Resistance

To understand the implications of the availability of hosts and the selection for resistance, we begin by analysing a SIS epidemiological model with two strains (sensitive and resistant) as introduced by Lehtinen et al. [4]. SIS stands for Susceptible-Infected-Susceptible. This model is suitable for modelling a lot of bacterial infections, including *S. pneumoniae* infections. A SIS model assumes that infected individuals are again susceptible after recovery and that there is no immunity. Furthermore, there are no births or deaths, as these two forces are equalising. The infected population can be subdivided into a compartment that is infected with resistant bacteria I_r and a compartment infected with sensitive bacteria I_s. The infections with sensitive bacteria can be treated with antibiotics at a rate τ, which increases their clearance rate. An assumption is that antibiotics lead to direct clearance of the infection.

Since antibiotic resistance confers an advantage in clearance but a disadvantage in transmission (see Introduction), we assume that $\beta_s > \beta_r$ and $\gamma_r > \gamma_s$. This is encoded via a cost term c_β resulting in $\beta_r = \frac{\beta}{c_\beta}$ and $\beta_s = \beta$. For clearance, this is modelled with a cost c_γ and we get $\gamma_r = c_\gamma \gamma$ and $\gamma_s = \gamma$. This leads to the following system of ordinary differential equations (ODEs) that describe

the dynamics of antibiotic resistance as set up by Lehtinen et al. [4].

$$\begin{aligned}
\frac{dS}{dt} &= (\gamma+\tau)I_s - \beta S I_s + c_\gamma \gamma I_r - \frac{\beta}{c_\beta} S I_r \\
\frac{dI_s}{dt} &= \beta S I_s - (\gamma+\tau) I_s \\
\frac{dI_r}{dt} &= \frac{\beta}{c_\beta} S I_r - c_\gamma \gamma I_r
\end{aligned} \tag{1}$$

The detailed description of the parameters can be found in table 1.

Parameter	Description	Simulation rate [month^{-1}]
β	transmission rate	0.002
γ	clearance rate	1
τ	antibiotic consumption rate	∼0.21
c_β	transmission cost of antibiotic resistance	1.2
c_γ	clearance cost of antibiotic resistance	1.0

Table 1: The parameters of the SIS model and the values used in the simulation are described in this table as a reference.

Competitive exclusion of strains at equilibrium The resistant and the sensitive strain will exclude each other competitively (appendix A). There is a condition when coexistence is possible:

$$S = \frac{\gamma+\tau}{\beta} = \frac{c_\gamma c_\beta \gamma}{\beta} \tag{2}$$

If we take the inverse of S it follows that $\frac{\beta}{\gamma+\tau} = \frac{\beta}{c_\gamma c_\beta \gamma}$. This is the definition of R_0 meaning that we get the relationship $R_{0,s} = R_{0,r}$ at equilibrium in a coexistence scenario. [4]. If $R_{0,s} > R_{0,r}$ we will get the above mentioned competitive exclusion of resistant types over sensitives and vice versa if $R_{0,r} > R_{0,s}$ (appendix B) [4]. This means the trade-off which strain is more competitive at equilibrium depends on the R_0 values.

Transient dynamics show a dominance of sensitive strains Numerical integration of system 1 leads to the trajectories in figure 1. We see a drop in susceptibles early in the disease progression. In this phase, we observe a transient increase in sensitive infections and later on a domination by resistant infections with the chosen parameter values. At which parameter values we expect to see a faster increase of sensitives can be quantified by the following inequality (appendix C):

$$\beta S_{df}\left(1 - \frac{1}{c_\beta}\right) + \gamma(c_\gamma - 1) - \tau > 0 \tag{3}$$

If this inequality holds we observe a faster increase of sensitive strains relative to resistant strains. This inequality tells us as well, that availability of susceptibles plays a role in transmission, meaning that if $c_\beta > 1$ we expect an influence of the number of susceptibles on the transient dominance of sensitives. Availability of susceptibles has though no influence on clearance, as the cost of clearance c_γ is not affected by S_{df}.
If the number of susceptibles drops, as is the case in the numerical simulation 1, the inequality does not hold anymore and the resistant strains grow faster after the drop. It is apparent that the trade-off between resistant and sensitive strains in this transient phase depends on the availability of host and changes with the choice of parameter.

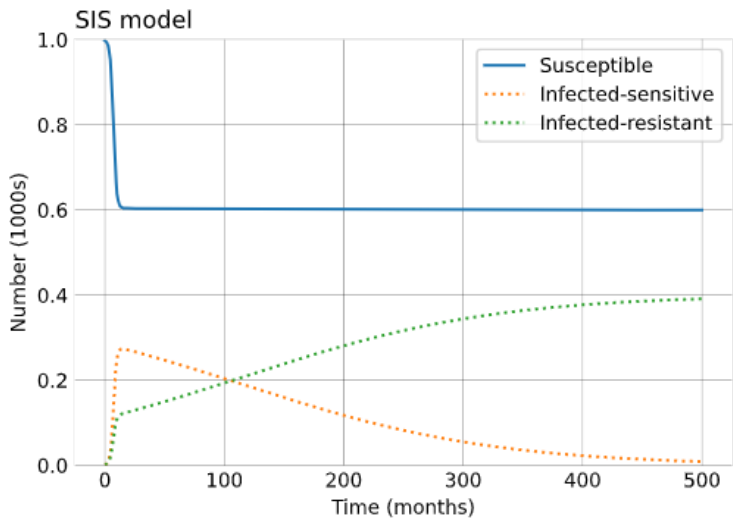

Figure 1: The trajectories of the SIS ODE system are plotted. On the y-axis is the number of individuals per compartment in 1000s. The x-axis is the time measured in months. We see a sharp drop in susceptibles in the first 10 months and an increase in both infected compartments. After approx. 10 months the infected-sensitive cases go back again and the infected-resistant cases make up 40% at equilibrium. The parameter values [rates per month] are $c_\beta = 1.2$, $c_\gamma = 1.0$, $\beta = 0.002$, $\gamma = 1$ and $\tau = 0.21$

Initial conditions matter for the transient dynamics A phase plane analysis of the SIS system in figure 2 shows the behaviour of the sensitive and resistant strains for different starting conditions. In the long run, we see the domination of the resistant strains no matter where we start from. Where the transient peak (if present) in sensitives lies is, however, highly dependent on the initial condition. The higher the initial value of I_s is, the higher will be the number of infections at the peak. This shows that the trade-off between sensitive and resistant strains depends on the initial frequencies.

Through the SIS antibiotic resistance model in equations 1, we get an idea about how the sensitive and resistant strains change in frequency subject to a decrease in the available hosts (figure 1). Vaccination is a process that leads to a decrease in the availability of hosts, so we expect a transient increase of sensitives in a vaccination setting followed by domination of resistance types at equilibrium. Via the phase plane analysis in figure 2, we understand that initial conditions matter a lot for transient phenomena. The flaw in the modelling up until now is, that we do not account for vaccination explicitly. This leads to the next, more complex SISV model.

2.2 SISV Model of Antibiotic Resistance and Vaccination

In the SIS model equation 1 we see what a drop in susceptibles means in terms of resistance. To verify the robustness of the insight we construct a model of vaccination. The main idea of this model was to take the SIS model from equation 1 and integrate the vaccination process. This gives rise to a SISV model, which includes a vaccinated compartment V. Since a vaccine against *S. pneumoniae* only includes certain strains and not others, we had to distinguish between infections with a vaccine-type I^v (meaning a strain that is included in the vaccine) and a non-vaccine-type I^{nv}(not included in the vaccine). The dynamics of resistant and sensitive strains happen in the same way for non-vaccine-type and vaccine-type as described in the SIS model (equation 1). Vaccination

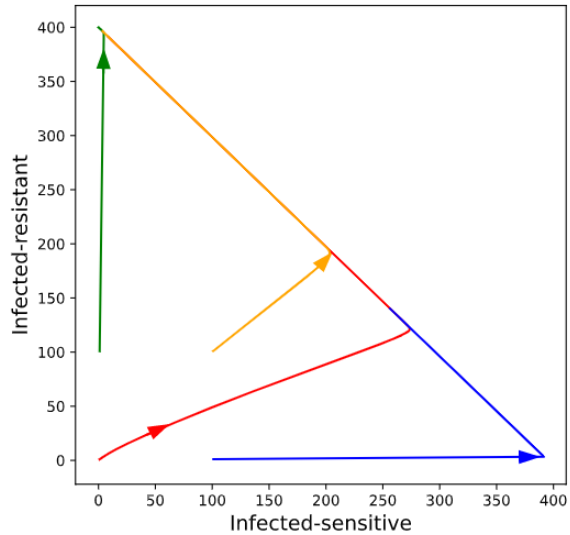

Figure 2: This graph shows a phase plane analysis of infected subtypes of the SIS system described in equations 1. The total population size N was kept constant so the number of susceptibles changed according to the initial values of I_s and I_r. According to different starting conditions (the points in the lower-left quadrant where the trajectories start), we see different transient dynamics. In the long run, we see, that the resistant types will outcompete the sensitives

of individuals happens at a rate ν. Individuals that were vaccinated can not be infected again with a vaccine-type bacterium. To keep track of the non-vaccine-type infections, a subcompartment was created e.g. $I^{nv,S}$ and $I^{nv,V}$ for non-vaccinated individuals and vaccinated individuals respectively. This gave rise to the following system of ODEs described in figure 3 and equations 4.

$$\begin{aligned}
\frac{dS}{dt} &= (\gamma+\tau)(I_s^{nv,S}+I_s^{v,S}) - \beta S(I_s^{nv,S}+I_s^{nv,V}+I_s^{v,S}) + c_\gamma\gamma(I_r^{nv,S}+I_r^{v,S}) \\
&\quad - \frac{\beta}{c_\beta}S(I_r^{nv,S}+I_r^{nv,V}+I_r^{v,S}) - \nu S \\
\frac{dI_s^{nv,S}}{dt} &= \beta S(I_s^{nv,S}+I_s^{nv,V}) - (\gamma+\tau)I_s^{nv,S} - \nu I_s^{nv,S} \\
\frac{dI_r^{nv,S}}{dt} &= \frac{\beta}{c_\beta}S(I_r^{nv,S}+I_r^{nv,V}) - c_\gamma\gamma I_r^{nv,S} - \nu I_r^{nv,S} \\
\frac{dI_s^{nv,V}}{dt} &= \beta V(I_s^{nv,V}+I_s^{nv,S}) - (\gamma+\tau)I_s^{nv,V} + \nu I_s^{nv,S} \\
\frac{dI_r^{nv,V}}{dt} &= \frac{\beta}{c_\beta}V(I_r^{nv,V}+I_r^{nv,S}) - c_\gamma\gamma I_r^{nv,V} + \nu I_r^{nv,S} \\
\frac{dI_s^{v,S}}{dt} &= \beta S I_s^{v,S} - (\gamma+\tau)I_s^{v,S} - \nu I_s^{v,S} \\
\frac{dI_r^{v,S}}{dt} &= \frac{\beta}{c_\beta}S I_r^{v,S} - c_\gamma\gamma I_r^{v,S} - \nu I_r^{v,S} \\
\frac{dV}{dt} &= \nu S + (\gamma+\tau)I_s^{nv,V} - \beta V I_s^{nv,V} + c_\gamma\gamma I_r^{nv,V} - \frac{\beta}{c_\beta}V I_r^{nv,V} + \nu(I_s^{v,S}+I_r^{v,S})
\end{aligned} \tag{4}$$

The variables of equation 4 are described in table 2 and the parameters in table 3.

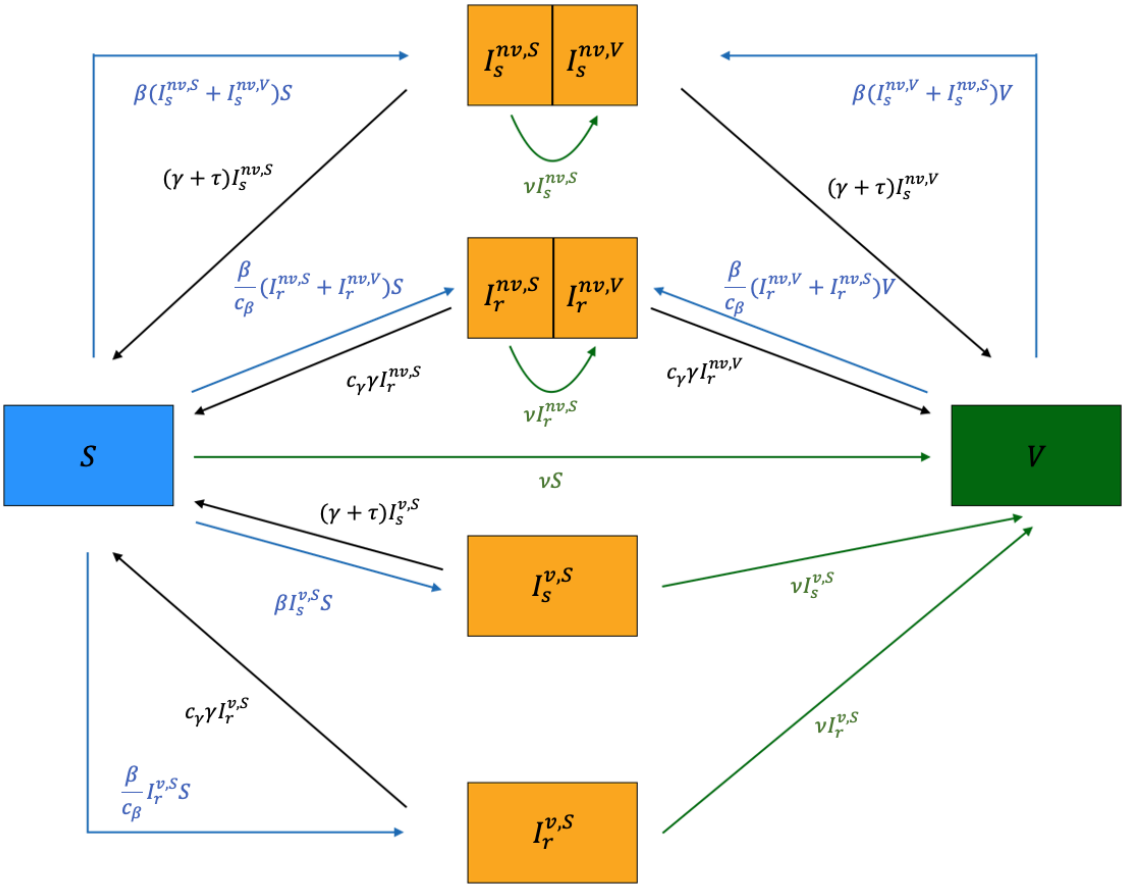

Figure 3: Flowchart describing the system of equations 4. The variables are described in table 2 and the parameters in table 3. In blue are all transmission terms, in black all clearance terms and in green all vaccination terms.

Variable	Description
S	Susceptibles
$I_s^{nv,S}$	Unvaccinated, infected with a sensitive strain of the non-vaccine-type
$I_r^{nv,S}$	Unvaccinated, infected with a resistant strain of the non-vaccine-type
$I_s^{nv,V}$	Vaccinated, infected with a sensitive strain of the non-vaccine-type
$I_r^{nv,V}$	Vaccinated, infected with a resistant strain of the non-vaccine-type
$I_s^{v,S}$	Unvaccinated, infected with a resistant strain of the vaccine-type
$I_r^{v,S}$	Unvaccinated, infected with a sensitive strain of the vaccine-type
V	Vaccinated

Table 2: The variables of the SISV model are described as a reference in this table.

Parameter	Description	Simulation rate [month^{-1}]
β	transmission rate	0.002
γ	clearance rate	1
τ	antibiotic consumption rate	~ 0.21
c_β	transmission cost of antibiotic resistance	1.2
c_γ	clearance cost of antibiotic resistance	1.0
ν	vaccination rate	0.05

Table 3: The parameters and the values used in the simulation of the SISV model are described as a reference in this table.

If we want to introduce a rollout of the vaccine after reaching equilibrium, we need to have sensitive and resistant strains present in the population. To do this we will enforce coexistence. We saw that the R_0 values for the sensitives and the resistants need to be the same as we have derived in equation 2. This leads to a constraint on the antibiotic consumption rate as derived by Lehtinen et al. [4]:

$$\tau = \gamma(c_\gamma c_\beta - 1) \tag{5}$$

We parameterise the model in such a way that resistance is outcompeting sensitivity very slowly achieving a transient coexistence of the two strains. This is done by taking the τ from equation 5 and subtracting a very small difference $\Delta = 0.01$.

Competitive exclusion at equilibrium This system can be analysed numerically and we obtain the trajectories visible in figure 4 (A). The result from the SIS model in section 2.1 is supported by the simulation in figure 4 (A). We distinguish vaccine-types and non-vaccine-types. After the introduction of the vaccine ($t = 50$), the vaccine types decrease strongly which we assume to happen since the vaccine works explicitly against these types. The non-vaccine-types evolve in coexistence and after transient dynamics, we find a competitive exclusion from $\sim t = 100$ on. With our parameterisation, the resistant strains will outcompete the sensitive strains in the long run. This shows again that the parameterisation matters as well in the SISV model. The trade-off of resistant strains vs. sensitive strains is in favor of resistant types at equilibrium given our parameterisation.

Transient dynamics show a dominance of sensitives In the SIS model trajectories in figure 1 we saw a transient dominance of sensitive types of resistant types before going to equilibrium. We see a similar behaviour in figure 4 (A). The parameterisation in both cases was the same. The difference in the SISV model is the explicit vaccination. The vaccination starts at $t = 50$ and is characterised

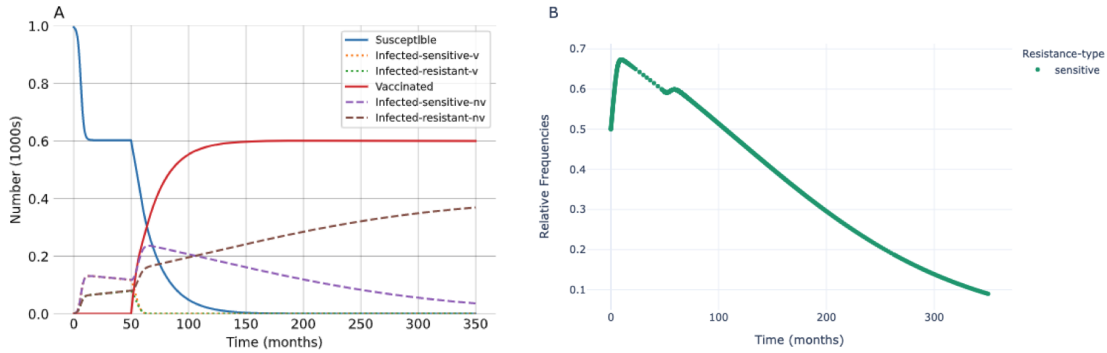

Figure 4: (A)Plot of the dynamics of the SISV model over time. Vaccination after timepoint $t = 50$. On the y-axis are the number of individuals per compartment in 1000s. The x-axis is the time measured in months. The parameter values [rates per month] are $c_\beta = 1.2$, $c_\gamma = 1.0$, $\beta = 0.002$, $\gamma = 1$, $\tau = 0.21$ and $\nu = 0.05$. In this plot, the subcompartments are added together, so that the behaviour of vaccine-types vs non-vaccine-types in terms of susceptibility can be analysed more easily. The initial conditions were $I_s^{nv,S} = 1, I_r^{nv,S} = 1, I_s^{nv,V} = 0, I_r^{nv,V} = 0, I_s^v = 1, I_r^v = 1$ and $V = 0$. (B) Plot of the relative frequency of sensitive strains in the non-vaccine-types. The x-axis is the time in months and the y-axis the frequency of sensitives relative to the total amount of infected individuals(sensitives and resistant strains)

by a drop in susceptibles and a steep rise in vaccinated individuals. We observe an overall increase in the non-vaccine-types. In a transient phase, the sensitive non-vaccine-types increase faster than the resistant types and dominate the resistant types over ~ 50 months. This is again a form of trade-off as the transient decrease in susceptibles and the increase in vaccinated individuals change the context of competition. This leads to a transient domination of sensitives.

Relative frequencies show transient dominance at different timepoints In figure 4 (B) the relative frequencies of non-vaccine-types are plotted. In the beginning, the relative frequency of sensitive strains increases and shortly after that decreases slightly until the vaccination starts. Then we see a small peak which is the transient dominance which is resolved after $t = 100$ when the resistant strains become dominant and outcompete the sensitive strains. This illustration facilitates the observation of dominance. There are multiple steps in this trajectory. In the beginning, the relative frequencies rise until a value of ~ 0.68. After that, they decrease until the vaccination starts at $t = 50$. Then follows a transient increase in the relative frequency which last a bit over $t = 80$. After that the resistant types increase faster and after $t = 100$ the resistant types dominate.

The simulation gives rise to the hypothesis that a high level of hosts followed by a decrease, as is the case during an ongoing vaccination campaign, will lead to a transient increase of sensitive frequencies. This increase is faster than the increase in resistant strains. To verify this claim of transient dominance, we look at data from a pneumococcal vaccination campaign.

2.3 Data Analysis

First, we look at the development of vaccine-types vs. non-vaccine-types over time. This is visible in figure 5 (A). The dataset is from Massachusetts hospitals, collected in the form of nasal swaps from children. The vaccination campaign with PCV7 started in 2000. Therefore, we assume that the

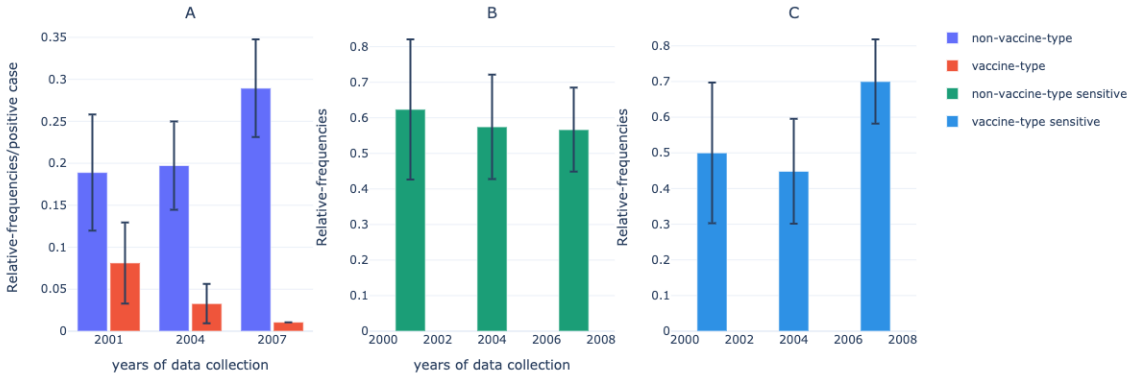

Figure 5: (A) Plotted here are the relative frequencies of vaccine-types vs. non-vaccine-types over the course of the collection period. The relative frequency was corrected for different test positivity over the years (27%, 23%, and 30%). (B) Plotted here are the relative frequencies of sensitive strains in the non-vaccine-types over the course of the collection period. (C) Plotted here are the relative frequencies of sensitive strains in the vaccine-types over the collection period. The binomial confidence interval is indicated to show the uncertainty in the data.

year 2001 constitutes as a pre-vaccination data point. This pre-vaccination data point tells us that the non-vaccine-types are already more common than the vaccine-types in 2001. We can manually fit our model in 4 (A) to this and set the parameter values as indicated. This gives us a similar relative frequency in 4 (B) at the start of the vaccination campaign. The question that arises from this is, whether these dynamics are only visible for the chosen parameter values. For a broader sensitivity analysis of the parameters refer to appendix D. Next, it is visible in figure 5 (A) that first, the vaccine-type frequencies decline before the non-vaccine-type frequencies rise. In the simulation figure 4 (A) we observe that the vaccine-types are replaced by non-vaccine-types but with no lag.

The interest in the transition for the non-vaccine-types as seen in the simulation (figure 4) meant that in the following analysis we focused on non-vaccine-types. This was due to the hypothesis on the transient dominance of sensitive strains in the non-vaccine types. The overall amount of positive *S. pneumoniae* cases increased over the years in figure 5 (A). To compare the years, the relative frequencies of sensitive strains in the non-vaccine-type were computed in figure 5 (B). In 2001 there are already more sensitive than resistant strains. There is a trend of equilibration, but this trend is not statistically significant. A chi-squared test on the absolute counts of the non-vaccine-types gave a p-value of 0.62. When investigating the vaccine-type frequencies in figure 5 (C) we find that there is a trend of increase in the sensitives. This trend, however, was again not significant in a chi-squared test (p-value = 0.39)

3 Discussion

Using a SIS model of antibiotic sensitivity we find that next to known effects such as competitive exclusion of sensitive types we see a transient dominance of sensitive strains over resistance strains. These transient dynamics are found to be highly dependent on the initial conditions. Via an explicit SISV model of vaccination, we can make these claims about transient dynamics more robust and get a better understanding of the underlying processes. In the SISV model, we find

next to the known process of strain replacement (vaccine-types decline and non-vaccine-types surge) a transient dominance of non-vaccine-type sensitive strains over resistant strains. Data from the Massachusetts pneumococcal vaccination campaign shows that the concept of strain replacement is observed, whereas the transient dominance of sensitive types is not visible.

SIS model shows a transient shift in trade-off optimum In the results section on the SIS section we have seen that the drop in susceptibles early in the simulation leads to interesting transient dynamics. This drop in susceptibles is what we expect to happen via vaccination as well. The vaccination campaign will lead to a drop in fully-susceptible individuals and an increase in vaccinated individuals. This means we expect similar processes to happen in vaccination as in the SIS model after the drop of susceptibles.

We see in the numerical analysis figure 1 the transient dynamics before going to equilibrium. Since we have a shift in the context of our trade-off (the availability of hosts), we can interpret the transient dominance of sensitive types as being an effect of this shift. The trade-off optimum changes transiently towards the sensitive types. In the framework of the infectiousness vs. duration of infection trade-off, we find that sensitivity corresponds to infectiousness (refer to section 1). This means as a consequence that because of the change of going from a high host density to a low host density the trade-off between infectiousness and duration of infection shifts transiently towards higher infectiousness.

This finding is visible in equation 3 as well. This equation poses a constraint on when we expect to see a faster increase in sensitive strains relative to resistant ones. What we can deduce from this equation is, that there is a dependence on the number of susceptibles S. If S is high, I_s will grow faster than if S is low. This gives us a more intuitive understanding of the findings in figure 1. In the beginning the number of susceptibles is high and I_s grows faster than I_r. When S decreases though, the inequality of equation 3 does not hold anymore and I_r will grow faster than I_s. This shows that the shift in trade-off optimum in response to the availability in hosts can be found analytically in equation 3 and visually in figure 1.

Equation 3 shows yet another interesting fact. The dependence on the availability of hosts is only found in the transmission term. If the cost of transmission $c_\beta = 1$ then the dependence on the host falls away. In other terms, if the cost of resistance only manifests itself in clearance ($c_\gamma > 1$) and not in transmission ($c_\beta = 1$), then we do not expect the trade-off optimum shift as a function of host availability as explained in the paragraph above.

To make the arguments about the different combinations of costs more robust, we performed a parameter sensitivity analysis that can be found in appendix D. We expect to find the peak in sensitives for various combinations of c_γ and c_β. The constraint when we find a transient dominance is derived in equation 3. The other parameters are chosen according to literature values to approximate *S. pneumoniae*. The values for β, τ and γ were taken from Colijn et al. [11] and Lehtinen et al. [4]. Colijn et al. describe in detail which parameter values make sense in which context.

In summary, the SIS model is a good approximation for vaccination and we find a shift in the trade-off optimum towards transiently higher infectiousness.

SISV model supports claims of the SIS model The main idea of the SISV model was to make the claims we set up from the SIS model more robust. To do so, we extended the SIS model with vaccine functionality. In figure 4 we explicitly see a drop in susceptibles and a rise in vaccinated

individuals. There are again transient dynamics in the non-vaccine-types that show a transient dominance of sensitive strains over resistant ones. Therefore, these findings support our claim from the SIS model that vaccination poses a shift in trade-off optimum towards transiently higher infectiousness. The parameterisation of the SISV model was the same as in the SIS model, so the discussion of parameters for the latter model works here as well.

Summarising, we saw that a model of explicit vaccination supports the finding that vaccination poses a shift towards transiently higher infectiousness.

Prior Decline in vaccine-types not reproduced in the simulation In figure 4 (A) we note a decline in vaccine-types and a rise in non-vaccine-types. This is what is referred to as strain replacement. This replacement of vaccine-types can be found in the data in figure 5 (A). The data in figure 5 (A) predicts a decline of vaccine-types before the increase of non-vaccine-types. In the model 4 (A) the decline and increase happen at the same time. One problem with the interpretation of the data in figure 5 (A) is that the uncertainty is very high. Given the binomial confidence intervals, we can not be sure whether or not the trend we observe is really what is to be found in the data. This explanation is supported by other publications that show no lag but a simultaneous strain replacement [12, 13]. If the trends are, however, correct then this means that our model does not capture this lag phase in the strain replacement process. One possible explanation might be, that not all carriages of non-vaccine-type infections are asymptomatic. Individuals with symptomatic infections might not get vaccinated and thus the vaccination rate from the non-vaccine-type compartments might be lower than expected. This could lead to a lag phase because of different vaccination rates according to which compartment the individual is in.

Strain replacement is visible both in the simulation figure 4 (A) and the data figure 5 (A). The difference lies in a lag phase in the data that is not observed in the simulation.

Transient increase of sensitives is not found in the data The transient increase as found in the simulation figure 4 (B) is not found in the data figure 5 (B). In the beginning, there are more sensitive strains than resistant ones (which we accounted for in the simulation with our parameter conditions). We see no increase in sensitives but rather a decrease and an increase in resistant types in the data figure 5 (B). This is not in line with our hypothesis. There are multiple ways to explain this mismatch, i) the predictions of the model are correct but the data is too poor to show the effect or ii) the predictions of the model are incorrect.

i) Firstly, the period between the data points is quite large (4-year intervals or 48 months). In the simulation, the vaccination campaign starts at $t = 50$, and the plateau after which we expect a decline in sensitives again is at $\sim t = 100$. This means in the simulation, we expect the transient behaviour to show within approximately 50 months. Since the data was only collected 48 months after the first time point, it could be that we simply miss the transient increase phase. This should be investigated in more detail with a more fine-grained dataset, where we ideally would have a datapoint each year.

ii) The other explanation might be, that our model in equation 4 was too simplified. One aspect is, that we had introduced coexistence by setting the treatment rate to equation 2. A more complex model of coexistence would allow the treatment rate τ to be set to a lower value as in the literature and maybe the transient dynamics would change in response. This means a more sophisticated model of coexistence taking into account aspects such as balancing selection could be an extension to the theoretical framework [4]. Another way how our model could be wrong is that we saw from equation 3 that the effect of the host dependence is only visible if there is a transmission cost. If resistance only shows a clearance cost c_γ the host dependence in the trade-off shift would not be

visible. This might mean, that the data could only show a clearance cost for resistance and therefore no host dependence is expected in the transient phase.

4 Conclusion

Our modelling gives rise to the idea that a transient shift towards higher infectiousness is expected upon the introduction of a vaccine against *S. pneumoniae*. These trends are robust to parameter changes and can be found in a simple model of antibiotic resistance (SIS) and are supported by a model of explicit vaccination (SISV). Data from the Massachusetts pneumococcal vaccination campaign do not show this transient higher infectiousness. This might be due to either a wrong model or the data being too poor to observe these transient effects. Further work should investigate a more fine-grained dataset of pneumococcal vaccination that would allow for a more robust understanding of the real-world setting. Furthermore, models including different approaches to coexistence and different cost choices should be considered for getting a broader picture of the expected processes.

5 Acknowledgments

I thank Sonja Lehtinen for the kind supervision, it has been a very pleasant working atmosphere and a stimulating research topic. I thank Sebastian Bonhoeffer for enabling the lab rotation in the theoretical biology group.

6 Material & Methods

6.1 Datasets and Data Availability

The data is made available on the GitHub page of the project `https://github.com/mjemons/IDD_TB` in the subfolder `data`. The data used was the Massachusetts dataset described in Croucher et al.[14]. It contains information on nasal swaps of 616 children collected between 2001 and 2007.

6.2 Model Implementation and Code Availability

The models and data analysis was implemented using python 3.7.9. All the packages and their corresponding versions are documented in the `requirements.txt` on the following GitHub page `https://github.com/mjemons/IDD_TB` in the subfolder `scripts`. All code that was used in this project can be found on the same GitHub page.

6.3 Parameterisation

The model parameters chosen were taken from Lehtinen et al. [4] and Colijn et. al. [11]. They are given as rates per month (see as well discussion). The initial conditions were chosen to be always 1 for the infected individuals.

References

[1] Paul N. Zivich, John D. Grabenstein, Sylvia I. Becker-Dreps, and David J. Weber. Streptococcus pneumoniae outbreaks and implications for transmission and control: a systematic review. *Pneumonia*, 10(1), 2018.

[2] Alessandra Løchen, Nicholas J. Croucher, and Roy M. Anderson. Divergent serotype replacement trends and increasing diversity in pneumococcal disease in high income settings reduce the benefit of expanding vaccine valency. *Scientific Reports*, 10(1):1–17, 2020.

[3] Katherine L. O'Brien, Lara J. Wolfson, James P. Watt, Emily Henkle, Maria Deloria-Knoll, Natalie McCall, Ellen Lee, Kim Mulholland, Orin S. Levine, and Thomas Cherian. Burden of

disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *The Lancet*, 374(9693):893–902, 2009.

[4] Sonja Lehtinen, François Blanquart, Nicholas J. Croucher, Paul Turner, Marc Lipsitch, and Christophe Fraser. Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5):1075–1080, 2017.

[5] Andrew D. Letten, Alex R. Hall, and Jonathan M. Levine. Using ecological coexistence theory to understand antibiotic resistance and microbial competition. *Nature Ecology and Evolution*, 5(April):431–441, 2021.

[6] Christophe Fraser, T. Déirdre Hollingsworth, Ruth Chapman, Frank De Wolf, and William P. Hanage. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 104(44):17441–17446, 2007.

[7] Sergey V. Melnikov, David L. Stevens, Xian Fu, Hui Si Kwok, Jin Tao Zhang, Yue Shen, Jeffery Sabina, Kevin Lee, Harry Lee, and Dieter Söll. Exploiting evolutionary trade-offs for posttreatment management of drug-resistant populations. *Proceedings of the National Academy of Sciences of the United States of America*, 117(30):17924–17931, 2020.

[8] Christophe Fraser, Katrina Lythgoe, Gabriel E. Leventhal, George Shirreff, T. Déirdre Hollingsworth, Samuel Alizon, and Sebastian Bonhoeffer. Virulence and pathogenesis of HIV-1 infection: An evolutionary perspective. *Science*, 343(6177), 2014.

[9] Centers for Disease Control and Prevention. About Pneumococcal Vaccines. https://www.cdc.gov/vaccines/vpd/pneumo/hcp/about-vaccine.html.

[10] Yonas I. Tekle, Kaare M. Nielsen, Jingzhou Liu, Melinda M. Pettigrew, Lauren A. Meyers, Alison P. Galvani, and Jeffrey P. Townsend. Controlling Antimicrobial Resistance through Targeted, Vaccine-Induced Replacement of Strains. *PLoS ONE*, 7(12):1–9, 2012.

[11] Caroline Colijn, Ted Cohen, Christophe Fraser, William Hanage, Edward Goldstein, Noga Givon-Lavi, Ron Dagan, and Marc Lipsitch. What is the mechanism for persistent coexistence of drug-susceptible and drug-resistant strains of Streptococcus pneumoniae? *Journal of the Royal Society Interface*, 7(47):905–919, 2010.

[12] Sam Mehr and Nicholas Wood. Streptococcus pneumoniae - a review of carriage, infection, serotype replacement and vaccination. *Paediatric Respiratory Reviews*, 13(4):258–264, 2012.

[13] A. Løvlie, D. F. Vestrheim, I. S. Aaberge, and A. Steens. Changes in pneumococcal carriage prevalence and factors associated with carriage in Norwegian children, four years after introduction of PCV13. *BMC Infectious Diseases*, 20(1):1–9, 2020.

[14] Nicholas J. Croucher, Jonathan A. Finkelstein, Stephen I. Pelton, Julian Parkhill, Stephen D. Bentley, Marc Lipsitch, and William P. Hanage. Population genomic datasets describing the post-vaccine evolutionary epidemiology of Streptococcus pneumoniae. *Scientific Data*, 2:1–9, 2015.

A Derivation of SIS Equilibria

Since we neglect births and deaths as seen above we get a constant population size N: $N = S + I_r + I_s$. This means the system becomes a 2x2 system of equations:

$$\frac{dI_s}{dt} = \beta(N - I_s - I_r)I_s - (\gamma + \tau)I_s$$
$$\frac{dI_r}{dt} = \frac{\beta}{c_\beta}(N - I_s - I_r)I_r - c_\gamma \gamma I_r$$

Factoring out I_r and I_s we get:

$$\frac{dI_s}{dt} = I_s\left(\beta(N - I_s - I_r) - (\gamma + \tau)\right) \stackrel{!}{=} 0$$
$$\frac{dI_r}{dt} = I_r\left(\frac{\beta}{c_\beta}(N - I_s - I_r) - c_\gamma \gamma\right) \stackrel{!}{=} 0$$

Here we get either a disease free solution ($I_s = I_r = 0$) or three endemic solutions

- disease free: $I_s = I_r = 0$
- endemic I: $I_s = I_r \neq 0$

$$N - I_s - I_r = \frac{\gamma + \tau}{\beta}$$
$$N - I_s - I_r = \frac{c_\gamma c_\beta \gamma}{\beta}$$

It follows that $S = \frac{\gamma+\tau}{\beta} = \frac{c_\gamma c_\beta \gamma}{\beta}$

If we take the inverse of S it follows that $\frac{\beta}{\gamma+\tau} = \frac{\beta}{c_\gamma c_\beta \gamma}$. This is the definition of R_0 meaning that we get the relationship $R_{0_s} = R_{0_r}$ at equilibrium.

- endemic II: $I_s = 0, I_r \neq 0$

$$\frac{dI_s}{dt} = I_s\left(\beta(N - I_s - I_r) - (\gamma + \tau)\right) \stackrel{!}{=} 0$$
$$\frac{dI_r}{dt} = I_r\left(\frac{\beta}{c_\beta}(N - I_s - I_r) - c_\gamma \gamma\right) \stackrel{!}{=} 0$$

Since $I_s = 0$ we get:

$$I_r\left(\frac{\beta}{c_\beta}(N - I_r) - c_\gamma \gamma\right) = 0$$

since $I_r \neq 0$ the brackets need to zero:

$$\frac{\beta}{c_\beta}(N - I_r) - c_\gamma \gamma = 0$$
$$\frac{\beta}{c_\beta}N - \frac{\beta}{c_\beta}I_r - c_\gamma \gamma = 0$$
$$\frac{\beta}{c_\beta}I_r = \frac{\beta}{c_\beta}N - c_\gamma \gamma$$
$$I_r = N - \frac{c_\beta c_\gamma \gamma}{\beta}$$

- endemic III: $I_s \neq 0, I_r = 0$

$$\frac{dI_S}{dt} = I_s\left(\beta(N - I_s - I_r) - (\gamma + \tau)\right) \overset{!}{=} 0$$
$$\frac{dI_R}{dt} = I_r\left(\frac{\beta}{c_\beta}(N - I_s - I_r) - c_\gamma\gamma\right) \overset{!}{=} 0$$

Since $I_r = 0$ we get:

$$I_s\left(\beta(N - I_s) - (\gamma + \tau)\right) = 0$$

since $I_s \neq 0$ the brackets need to zero:

$$\begin{aligned}\beta(N - I_s) - (\gamma + \tau) &= 0\\ \beta N - \beta I_s - \gamma + \tau &= 0\\ \beta I_s &= \beta N - \gamma + \tau\\ I_s &= N - \frac{\gamma + \tau}{\beta}\end{aligned}$$

B Derivation of SIS Stability

Stability analysis will allow us to decide which equilibrium will be attained according to the parameters. In this analysis we perform a linearisation of the system around the equilibrium point. First, we will define the Jacobian Matrix

$$\begin{aligned}J &= \begin{pmatrix}\frac{\partial f_1}{\partial I_s} & \frac{\partial f_1}{\partial I_r}\\ \frac{\partial f_2}{\partial I_s} & \frac{\partial f_2}{\partial I_r}\end{pmatrix}\\ &= \begin{pmatrix}\beta(N - 2I_s - I_r) - (\gamma + \tau) & -\beta I_s\\ -\frac{\beta}{c_\beta} I_r & \frac{\beta}{c_\beta}(N - I_s - 2I_r) - c_\gamma\gamma\end{pmatrix}\end{aligned}$$

the stability of the equilibrium point can be determined by the sign of the eigenvalues of the Jacobian. In order to solve this, we will have to define the characteristic polynomial:

$$\det(J - \lambda I) = 0$$

In order to solve this we will plug in the equilibrium solutions and then solving the resulting characteristic polynomial.

We will analyse it first generally and then plug in the corresponding values at the equilibria.

$$
\begin{aligned}
\det(J-\lambda I) &= 0 \\
&= \begin{pmatrix} \beta(N-2I_s-I_r)-(\gamma+\tau)-\lambda & -\beta I_s \\ -\frac{\beta}{c_\beta} I_r & \frac{\beta}{c_\beta}(N-I_s-2I_r)-c_\gamma\gamma-\lambda \end{pmatrix} \\
&= [\beta(N-2I_s-I_r)-(\gamma+\tau)-\lambda]\cdot\left[\frac{\beta}{c_\beta}(N-I_s-2I_r)-c_\gamma\gamma-\lambda\right]-[-\beta I_s]\cdot\left[-\frac{\beta}{c_\beta}I_r\right] \\
&= \lambda^2+\lambda\left(c_\gamma\gamma-\frac{\beta}{c_\beta}(N-I_s-2I_r)+(\gamma+\tau)-\beta(N-2I_s-I_r)\right) \\
&+\frac{\beta^2}{c_\beta}(N-2I_s-I_r)(N-I_s-2I_r)-\beta c_\gamma\gamma(N-2I_s-I_r)+c_\gamma\gamma(\gamma+\tau) \\
&-(\gamma+\tau)\frac{\beta}{c_\beta}(N-I_s-2I_r)-\frac{\beta^2}{c_\beta}I_rI_s
\end{aligned}
$$

this can then be solved using midnights formula:

$$
\lambda_{1,2}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}
$$

The parameters a, b, c for the different equilibrium points are:

- disease free $(I_s = I_r = 0)$:

$$
\begin{aligned}
a &= 1 \\
b &= \frac{\beta}{c_\beta}N+c_\gamma\gamma+\gamma+\tau-\beta N \\
c &= \frac{\beta^2}{c_\beta}N^2-\beta c_\gamma\gamma N-\frac{\beta}{c_\beta}\gamma N+c_\gamma(\gamma+\tau)-\frac{\beta}{c_\beta}N(\tau+\gamma)
\end{aligned}
$$

- endemic I $(I_s \neq 0, I_r \neq 0)$:

$$
\begin{aligned}
a &= 1 \\
b &= c_\gamma\gamma-\frac{\beta}{c_\beta}(N-I_s-2I_r)+(\gamma+\tau)-\beta(N-2I_s-I_r) \\
c &= \frac{\beta^2}{c_\beta}(N-2I_s-I_r)(N-I_s-2I_r)-\beta c_\gamma\gamma(N-2I_s-I_r) \\
&+c_\gamma\gamma(\gamma+\tau)-(\gamma+\tau)\frac{\beta}{c_\beta}(N-I_s-2I_r)-\frac{\beta^2}{c_\beta}I_rI_s
\end{aligned}
$$

- endemic II $(I_s = 0, I_r \neq 0)$:

$$
\begin{aligned}
a &= 1 \\
b &= c_\gamma\gamma-\frac{\beta}{c_\beta}(N-2I_r)+(\gamma+\tau)-\beta(N-I_r) \\
c &= \frac{\beta^2}{c_\beta}(N-I_r)(N-2I_r)-\beta c_\gamma\gamma(N-I_r) \\
&+c_\gamma\gamma(\gamma+\tau)-(\gamma+\tau)\frac{\beta}{c_\beta}(N-2I_r)-\frac{\beta^2}{c_\beta}
\end{aligned}
$$

- endemic III ($I_s \neq 0, I_r = 0$):

$$
\begin{aligned}
a &= 1 \\
b &= c_\gamma \gamma - \frac{\beta}{c_\beta}(N - I_s) + (\gamma + \tau) - \beta(N - 2I_s) \\
c &= \frac{\beta^2}{c_\beta}(N - 2I_s)(N - I_s) - \beta c_\gamma \gamma (N - 2I_s) \\
&\quad + c_\gamma \gamma(\gamma + \tau) - (\gamma + \tau)\frac{\beta}{c_\beta}(N - I_s)
\end{aligned}
$$

C Initial Growth analysis

We can analyse the initial growth of the system as follows

$$
\begin{aligned}
\frac{dS}{dt} &= (\gamma + \tau)I_s - \beta S I_s + c_\gamma \gamma I_r - \frac{\beta}{c_\beta} S I_r \\
\frac{dI_s}{dt} &= \beta S I_s - (\gamma + \tau) I_s \\
\frac{dI_r}{dt} &= \frac{\beta}{c_\beta} S I_r - c_\gamma \gamma I_r
\end{aligned}
$$

We plug in the initial conditions of the disease free state
if I_s should grow faster than I_r we get the following

$$
\beta S_{df} - (\gamma + \tau) > \frac{\beta}{c_\beta} S_{df} - c_\gamma \gamma
$$

as $S_{df} = C$ (the initial condition of S) in the disease-free state we get

$$
\begin{aligned}
\beta S_{df} - \frac{\beta}{c_\beta} S_{df} - \gamma - \tau + c_\gamma \gamma > 0 \\
\beta S_{df} \left(1 - \frac{1}{c_\beta}\right) + \gamma(c_\gamma - 1) - \tau > 0
\end{aligned}
$$

This means we only suspect a faster growth of I_s in the beginning (dominance) if the above holds true. This is the case e.g. when $c_\gamma = c_\beta \geq 1$. If only one of them is greater than 1, then the inequality above has to hold.

D Sensitivity Analysis of Parameters

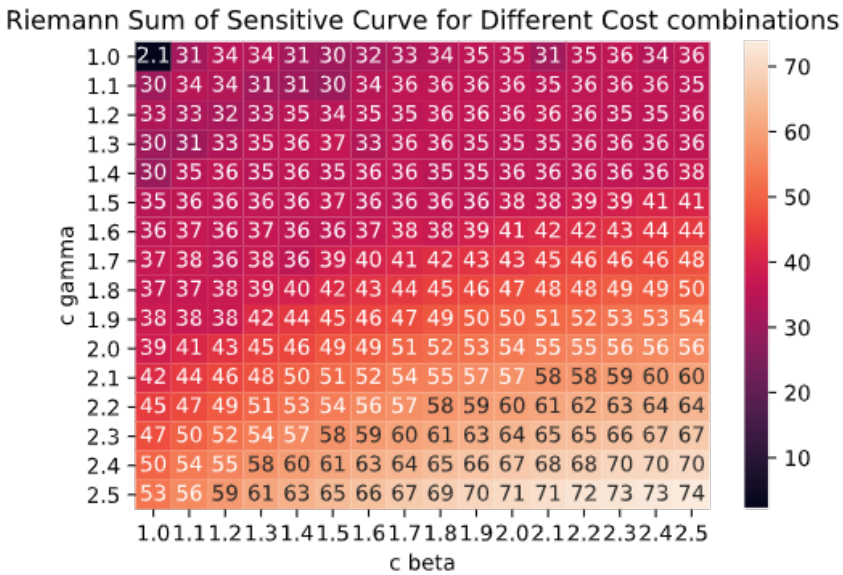

Figure 6: Sensitivity analysis for different cost values in the SIS model. For different combinations of c_γ and c_β the SIS model was simulated numerically. For each of these simulation the riemann sum was calculated for the sensitive curve. This corresponds to the numerical integral of the sensitive curves for each of the c_γ and c_β values. The heatmap indicates these integrals.