Optimal distribution strategy of a prevention tool in a heterogeneous population

Nada Ben Amor

Supervised by Dr. Eugenio Valdano and Dr.Michele Re Fiorentin

June 2023

Politecnico di Torino

université PARIS-SACLAY

Acknowledgements

I would like to express my gratitude to everyone who has supported me throughout my master's program and the completion of this thesis.

First and foremost, I would like to thank my supervisors, Eugenio and Michele, for their guidance, and continuous support throughout the duration of my internship.

I also want to thank my family for their love, encouragement, and understanding throughout this journey. They have always trusted me and supported me in my choices and projects.

Lastly, I would like to thank my friends and Andrea for their love and their unwavering faith in me.

Abstract

Response to infectious disease epidemics requires an optimal distribution strategy for tools of primary prevention(vaccines or prophylaxis for instance) with the aim of protecting those susceptible from acquiring the infection. We propose various distribution scenarios based on different epidemic phases and the goal one wants to reach : outbreak prevention (before the disease actually becomes an epidemic), containment (when the outbreak already occurred but stochastic fluctuations may still lead to the extinction of the disease), mitigation (once the endemic state is well established, one needs to adapt the strategy in order to eliminate the disease). Through the introduction of both a mean field and a stochastic view of an epidemiological model of a disease spread coupled with a prevention strategy, this research provided some insights into an adaptive approach. By analyzing various epidemiological indicators, it was possible to devise a strategy that takes into account the specific phase of the epidemic. The results indicate that a risk-based strategy, which prioritizes individuals with a higher number of contacts, does not always yield optimal results.

Contents

1 Introduction

Response to infectious disease epidemics requires an optimal distribution strategy for tools of primary prevention like vaccines or prophylaxis, with the aim of protecting those susceptible from acquiring the infection. However, achieving full coverage of those preventive tools remains a challenge due to various factors, including limited access to healthcare facilities, financial constraints, inadequate insurance coverage, and insufficient awareness campaigns. These limitations make it difficult for preventive interventions to be widely available and used.

One such prevention tool is Pre-Exposure Prophylaxis (PrEP) for the human immunodeficiency virus (HIV), a medication developed to significantly decrease the risk of acquiring HIV, making it an important tool in the fight against HIV epidemics. However, maximizing the benefits of PrEP poses considerable challenges: optimal utilization requires consistent daily adherence and regular follow-up. Moreover, stigmatization associated with HIV, limited healthcare infrastructure and the high cost of PrEP medication, create significant obstacles in reaching the target population and ensuring accessibility.

In this study, our objective is to develop an optimal strategy for distributing prevention tools, with a specific focus on PrEP for HIV. We aim to address the challenges associated with PrEP and propose various distribution scenarios based on different epidemic phases and the goal one wants to reach: outbreak prevention (before the disease actually becomes an epidemic), containment (when the outbreak already occurred but stochastic fluctuations may still lead to the extinction of the disease), mitigation (once the endemic state is well established, one needs to adapt the strategy to eliminate the disease). We will consider both the epidemiological context and broader societal factors that can influence the effectiveness of prevention campaigns. For example, the number of sexual partners or the level of adherence to the medication may impact the effectiveness of PrEP.

Through a comprehensive examination of existing literature and the development of a suitable model, our study aims to offer valuable insights that can guide policy decisions and enhance the distribution of PrEP and similar prevention tools. In particular, we demonstrate that risk-based strategies, which prioritize individuals at the highest risk of exposure, may not always be the optimal approach for reducing infection risk or the duration of the infection. Despite significant efforts to supply high-risk communities such as men who have sex with men or sex workers, previous research has revealed that the effectiveness of this strategy varies depending on the epidemiological context and the efficacy of the treatment [1]. Moreover, it can be demonstrated that, based on the epidemic phase and the desired goals, the relevant variables to be optimized need to be adapted, subsequently necessitating adjustments to the distribution strategy.

In the following, a network based-model will be used to describe a population of susceptible and infected individuals. Since this work aims to develop a distribution strategy for medicines having a limited coverage as for the PrEP,

one will use a susceptible-infected-susceptible (SIS) model coupled to an immunization strategy. The SIS epidemiological model captures the transmission dynamics of disease that do not present a long-term immunity and for which recovery may occur as for influenza or HIV. The population will be described as a network with individuals (nodes) separated into several classes depending on the number of contacts (degree) that they present. After defining the model properly, one needs to establish a distribution strategy based on the goal that one wants to reach.

The SIS model will be combined with a PrEP rollout strategy and will be applied to a heterogeneous population (individuals may have a different number of contacts, the efficacy of the prevention may vary from an individual to the other) using a formalism that enables us to address a stochastic process that exhibits nontrivial dependencies on network and prevention characteristics.

This manuscript is organized as follows: Section 2 will introduce the mean-field (MF) formalism for the SIS model integrated with an immunization strategy. Section 3 will provide the necessary tools for describing a process accounting for stochastic dynamics. In Section 4, we will present the key findings derived from the stochastic description, along with a suggested approach for constructing a prevention strategy based on the quantity that requires minimization.

2 Mean Field view

2.1 The classic SIS equation

In the classic SIS compartmental model, the population is divided into infected and susceptible individuals, and individuals may transition from one compartment to the other with a transmission (λ) or recovery (μ) rate.

$$\mathrm{I} \xrightarrow{\mu} \mathrm{S}$$

$$\mathrm{S} \xrightarrow{\lambda} \mathrm{I}$$

For a homogeneous network where all nodes (individuals) have the same degree (number of contacts), the system can be fully described by a single variable x = fraction of infected individuals. The MF dynamic is described by

$$\dot{x} = -\mu x + \lambda(1-x)x. \tag{1}$$

The epidemic threshold corresponds to the point in which we may or may not have an epidemic outbreak and it is characterized by the basic reproduction number, an epidemiological metric representing the average number of new infections that one infected individual is expected to generate in a population of entirely susceptible individuals. In the classic SIS model, the basic reproduction number R_0 is $\frac{\lambda}{\mu}$. By setting the left hand side of equation (1) to 0, one can identify two fixed points : $x^* = 0$ or $1 - \frac{\mu}{\lambda}$ whose stability depends on the value of $R_0 = \frac{\lambda}{\mu}$. As this number crosses the critical value 1, the system undergoes

a transcritical bifurcation, which is illustrated in figure 1a. When $R_0 < 1$, the only physical fixed point is $x^* = 0$ and the disease dies out: the system is below the epidemic threshold. The regime where $R_0 > 1$ is more interesting. In that case, the system is said to be above the epidemic threshold and the free-from disease state $x^* = 0$ is unstable while the endemic state $x^* = 1 - \frac{\mu}{\lambda}$ is stable.

This model can be easily generalized to the case where a fraction of the population received immunization against the disease. The effect of prevention will be, at the MF level, to reduce the basic reproduction number. This aspect will be explained with more details in the following section.

2.2 The SIS model coupled to an immunization strategy in a heterogeneous population

The SIS model described by equation (1) can be extended by coupling it to an immunization strategy. In that context, one should to further divide the population according to whether each person received prevention or no. The variables x and y will be respectively the conditional probabilities of being infected given that an individual did not receive prevention (x) or the conditional probabilities of being infected given that an individual received prevention (y). Indeed, to allow for an epidemic breakthrough, we assume that the efficacy of the prevention is lower than 100%, therefore, an individual that received prevention may still get infected.

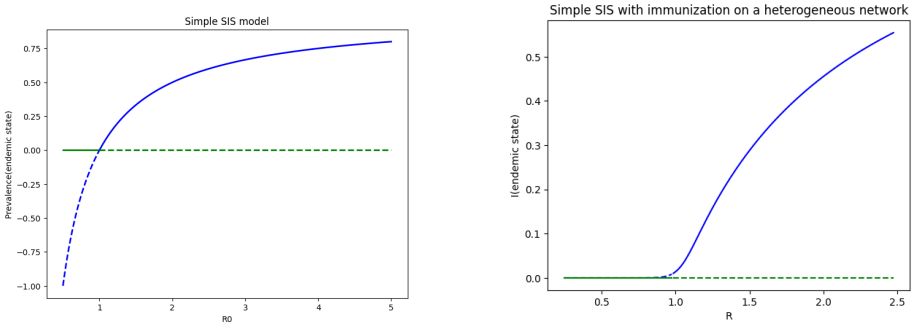
The heterogeneity in the number of contacts will be taken into account by dividing the population into n degree classes denoted by indices $i = 1, ..., n$. Each class i will then be described by $n_\epsilon + 1$ variables : y_{ia} = probability that an individual is infected given that he belongs to degree class i and received prevention with efficacy $\epsilon_a, (a = 1, ..., n_\epsilon)$ and x_i is the probability of being infected given that an individual belongs to class i and did not receive any prevention. In total, the population will be divided into $n(n_\epsilon + 1)$ classes.

Let us keep in mind that the variables x and y correspond to conditional probabilities. This choice was done to make a generalization of the common SIS equation but it will be important to introduce joint probabilities when comparing MF equations with other results as presented in sections 3.1. Moreover, we will apply the annealed network approximation (this is equivalent to a MF approximation for a heterogeneous network). For simplicity, we assume that there is no assortativity (i.e. no degree-degree correlation). The resulting dynamics is described by

$$\dot{x}_i = -\mu x_i + \frac{\lambda}{\langle k_i \rangle} k_i (1 - x_i) \xi, \tag{2}$$

$$\dot{y}_{ia} = -\mu y_{ia} + \frac{\lambda}{\langle k_i \rangle} (1 - \epsilon_a) k_i (1 - y_{ia}) \xi, \tag{3}$$

$$\xi = \sum_j p_j k_j \left[(1 - g_j) x_j + g_j \sum_b h_b y_{jb} \right] \tag{4}$$



(a) (b)

Figure 1: The bifurcation diagrams were obtained by solving the deterministic dynamic. The value of the prevalence is computed on the stable fixed point (solid line) and on the unstable fixed point (dashed ine). Figure1a shows the two well known fixed point for the simple SIS model (no immunization, homogeneous network). The unstable fixed point in the region with $R_0 < 1$ doesn't exist physically but the line was kept for better visualisation of the bifurcation. Figure1b shows the value of the prevalence in the endemic fixed point and in the free from disease state for an SIS model coupled to an immunization strategy on a heterogeneous network, with two degree classes.

In this set of equations, p_i is the probability of finding an individual belonging to degree class i (number of contacts) while g_i is the fraction of available medication allocated to degree class i. Finally, h_a is the probability that the prevention works with efficacy ϵ_a. Limited coverage c implies that the quantity $c = \sum_i p_i g_i$ is less than 1.

2.2.1 Epidemic threshold

The previous set of equations can once again lead to two different regimes depending on whether the system is above or below the epidemic threshold. Our goal is to describe the dynamic when the system is above the epidemic threshold and we should therefore look for the corresponding conditions on the epidemiological parameters. The conditions on the parameters can be obtained by performing a linear stability analysis around the free-from-disease state. In the case of two populations with no heterogeneity in the efficacy, equations (2) and (3) can be linearized around small x and y, which leads us to the following expression

$$\begin{pmatrix} \dot{x}_1 \\ \dot{y}_1 \\ \dot{x}_2 \\ \dot{y}_2 \end{pmatrix} = J \begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{pmatrix}, \tag{5}$$

with J being the Jacobian matrix (of size 4×4 in the case of two degree classes with homogeneous efficacy). If all eigenvalues of J, computed in 0, are negative, then the free-from-disease state is stable, and we are below the epidemic threshold. On the contrary, it is enough to find one positive eigenvalue to show that we are above the epidemic threshold. It is possible to diagonalize J analytically for no prevention ($g_1 = g_2 = 0$), and we find that its spectrum is: $[\lambda\frac{\langle k^2\rangle}{\langle k\rangle} - \mu, -\mu, -\mu, -\mu]$. This leads to the following condition on the parameters to be above the epidemic threshold : $\lambda > \mu\langle k\rangle/\langle k^2\rangle$. More generally for $g_1, g_2 \neq 0$, the spectrum of J is given by $[-\epsilon\langle gk^2\rangle\frac{\lambda}{\langle k\rangle} + \langle k^2\rangle\frac{\lambda}{\langle k\rangle} - \mu, -\mu, -\mu, -\mu]$. The epidemic threshold is thus given by

$$\lambda_c = \frac{\mu\langle k\rangle}{\langle k^2\rangle - \epsilon\langle gk^2\rangle}. \tag{6}$$

One can see in Figure 1b the new bifurcation diagram corresponding to the classic SIS model coupled to an immunization strategy.

In the most general case where both the network and the efficacy of the prevention are heterogeneous, one can show [2] that the epidemic threshold is given by

$$\lambda_c = \frac{\mu\langle k\rangle}{\langle k^2\rangle - \langle \epsilon gk^2\rangle}, \tag{7}$$

and the basic reproduction number of this model can be identified as $R = \frac{\lambda}{\mu}\frac{\langle k^2\rangle - \langle \epsilon gk^2\rangle}{\langle k\rangle}$ [2].

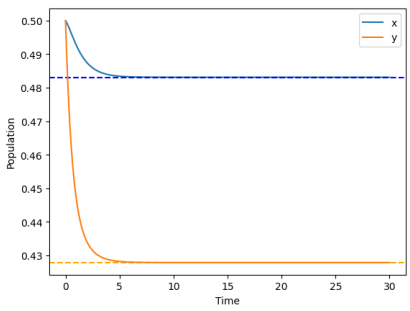
2.2.2 Endemic state

In the general case of heterogeneous network and efficacy, the endemic state is obtained by setting the left-hand-side of equations (2) and (3) to 0 leading to the following expressions:

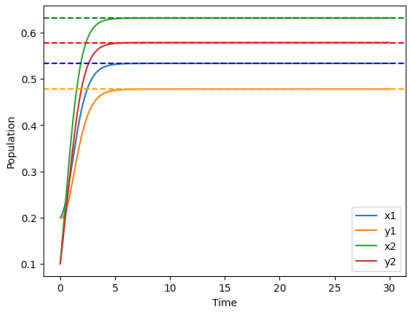
$$x_i = \frac{\lambda\frac{k_i}{\langle k\rangle}\xi}{1 + \lambda\frac{k_i}{\langle k\rangle}\xi}, \tag{8}$$

$$y_{ia} = \frac{x_i(1-\epsilon_a)}{1 - \epsilon_a x_i}. \tag{9}$$

Equations (8) and (9) are not analytically solvable in general, due to the term ξ that couples all classes. However, for homogeneous degree and efficacy, solving this system of equations reduces to solving a one-dimensional equation for x and the endemic state can be expressed in terms of the parameters of the problem. Moreover, the value of the endemic state can be obtained by solving numerically equations (2) and (3) using finite differences methods. For two degree classes, the solution converges to the endemic state within a few iterations (less than 10). This numerical approach is robust as long as one chooses reasonable initial conditions (x_i and y_i are conditional probabilities, so one should take initial



(a) $n = 1, k = 2, g = 0.2, \lambda/\mu = 0.99, \epsilon = 0.2$



(b) $n = 2, k_1 = 2, k_2 = 3, \lambda/\mu = 0.99, \epsilon = 0.2, g_1 = 0.2, g_2 = 0.8$

Figure 2: The analytical prediction of the endemic state is shown in dashed lines (solved with mathematica) and the numerical estimation is in solid lines. Figure 2a: The endemic state is calculated for a single degree class, x corresponds to infected individuals that did not receive any prevention while y corresponds to infected individuals that received immunization. Figure 2b : The endemic state is calculated for two degree classes denoted by index i= 1, 2. In both cases, the analytical prediction (in dashed lines) is in perfect agreement with the numerical estimation (solid lines).

points $\in$ [0, 1]). Finally, one can see in Figure 2 that the numerical solution agrees perfectly with the theoretical prediction (solution of the endemic state obtained with mathematica for one and two degree classes).

3 Stochastic dynamics and WKB methods

In the previous section, we considered a set of equations based on a mean field view : the equations were purely deterministic and did not take the intrinsic stochasticity of the model into account. To assess the impact of including heterogeneous efficacy on the effectiveness of treatment, it is now necessary to consider stochastic differential equations.

In this context, the endemic state appears to be a metastable state and the free-from disease state corresponds to an absorbing state : there is a very small but finite probability that the disease will disappear and no individual will no longer be in the infected state.In addition to the prevalence of the disease, one can look at the average time spent in the metastable endemic state. Therefore, building an optimal distribution strategy relies also on defining properly the goal of the immunization campaign in order to choose the relevant quantity that needs to be optimized.

In order to treat this complex stochastic process, one can use the Wentzel-Kramers-Brillouin (WKB) formalism [3]. Initially introduced in quantum mechanics as a semi classical limit, this set of methods can be applied in the context

of stochastic processes in the limit of a very large system in which stochasticity can be treated as a perturbation around the MF solution. As we shall see later, the endemic state is nothing but the stochastic counterpart of the deterministic fixed point derived from the MF equations (2) and (3). Going beyond the deterministic description enables us to compute [4], [5] the optimal path towards the absorbing state and other quantities that derive from it.

In the following, the WKB formalism will be presented with more details and we will compare two methods allowing for the evaluation of the average time spent in the endemic state, denoted as τ in the following and also referred to as the mean time to extinction. First, we will consider a numerical approach based on the master equation associated to the SIS model coupled to an immunization strategy. Then, we will use the results presented in [6] which provides the system size dependence of the mean time to extinction.

Section 3.1 introduces the master equation describing the stochastic process underlying our model. In section 3.2, the WKB Ansatz will be presented and adapted to the SIS model coupled to an immunization strategy with or without heterogeneity in the network and/or in the efficacy of the immunization. The latter can be described in a very general manner by following the notations used in [5], [6], [7]. Section 3.3 will be dedicated to the derivation of Hamilton's equations and their fixed points. A vector approach will then be presented in section 3.4, providing a numerical estimation of the mean time to extinction. Finally, in section 3.5, we will present an analytical approach to the heterogeneous SIS model based on the results of Assaf et al. [2] as well as Clancy et al. [6] [8].

3.1 Master equation

Let us consider a population of N individuals assigned to n degree classes, with homogeneous efficacy. Then, $\forall j = 1, ..., n$, x_j = P(infected | class j, no prevention), y_j = P(infected | class j, prevention) and p_j = P(class j) while g_j = P(prevention | class j). Therefore, P(infected, class j, no prevention) $= p_j(1 - g_j)x_j$ while P(infected, class j, prevention) $= p_j g_j y_j$. The variable $I_j^x = Np_j(1 - g_j)x_j$ corresponds to the number of infected individuals in class j that did not receive prevention and $I_j^y = Np_j g_j y_j$ is the number of individuals in class j that received prevention.

At each time step, an individual of class j gets infected at a rate w^+ and recovers at a rate w^-. It is important to understand that there are several ways of defining the SIS model. In the context of network-based models (which are the ones used in our case), connections between individuals are random variables and can be described by an adjacency matrix. The rates w^+ and w^- can then be defined in several ways (see [9]), depending on whether connected individuals are in perpetual contact or no. If we follow what was done in reference [9], we can assume that connected individuals are in perpetual contact. The explicit

expression of the rates is given in equations (10)-(13)

$$w^+(I_j^x) = \frac{\lambda}{N}(Np_j(1-g_j) - I_j^x)\beta_j^x \sum_{i=1}^{n} \alpha_i \left(I_i^x + I_i^y\right), \tag{10}$$

$$w^+(I_j^y) = \frac{\lambda}{N}(Np_j g_j - I_j^y)\beta_j^y \sum_{i=1}^{n} \alpha_i \left(I_i^x + I_i^y\right), \tag{11}$$

$$w^-(I_j^x) = \mu I_j^x, \tag{12}$$

$$w^-(I_j^y) = \mu I_j^y. \tag{13}$$

In these equations, λ and μ are the overall infection and recovery rates. In order to be consistent with the notation in [5], we introduce the infectiousness $\alpha_i \equiv k_i$ as well as the susceptibility $\beta(x_i) \equiv \frac{k_i}{\langle k \rangle}$, $\beta(y_i) \equiv \frac{k_i}{\langle k \rangle}(1-\epsilon)$.

These quantities will be defined in details in the following section : it will allow us to use the results presented in [6] and extend the methodology to heterogeneous efficacy (implemented in β). The heterogeneity of the network (the difference in degree) is already implemented and impacts both α ans β.

It is clear from those equations that the term $\sum_{i=1}^{n} \alpha_i \left(I_i^x + I_i^y\right)$ corresponds to the variable ξ that was introduced in the MF equations (2)-(4) but it is now rescaled by a factor N. Note that this choice for the infection and recovery rates was validated in several ways: the basic reproduction number found in section 2.2.1 is exactly the same as the one presented in ref [6] using this new parametrization. Additionally, we will see later that the expression of the endemic state obtained by solving the Hamilton equations that derive from the master equation defined with those rates coincides with the generic expression presented in [6] as well as the endemic state derived from the deterministic dynamic. The master equation for the probability distribution of a state $\mathbf{I} = (I_1^x, I_1^y, ..., I_n^x, I_n^y)$ can then be derived and one gets :

$$\begin{aligned} \frac{\partial P(\boldsymbol{I},t)}{\partial t} = \sum_{j=1}^{n} \sum_{q=x,y} & w^+(\boldsymbol{I} - \mathbf{1_j})P(\boldsymbol{I} - \mathbf{1_j}) - w^+(\boldsymbol{I})P(\boldsymbol{I}), \\ & + w^-(\boldsymbol{I} + \mathbf{1_j})P(\boldsymbol{I} + \mathbf{1_j}) - w^-(\boldsymbol{I})P(\boldsymbol{I}). \end{aligned} \tag{14}$$

Here, $1_j$is a vector whose components are all 0 except for the j^{th} component that is equal to 1.

3.2 WKB Ansatz or large deviation principle

Let us consider a population whose state is described by the vector $\mathbf{q}$ and that is divided into n degree classes. An SIS model coupled to an immunization strategy on a heterogeneous network can be described following a common notation found in the literature [5], [10]. The state variable q_i is the joint probability of being infected, belonging to class i and being in a given prevention state: if $q = \tilde{x}$, the individual did not receive prevention and $q = \tilde{y}$ if the individual received

prevention. Each class being characterized by some infectiousness α_i^q, some susceptibility β_i^q and by the fraction of individuals f_i^q belonging to class i.

Then, the probability that an individual of class i is infected by an individual of class j is given by : $\lambda\alpha_j^q(f_i^q - q_i)\beta_i^q$. In order to express those new quantities in terms of the parameters of the model, one can use the following definitions: $\alpha_i^q \equiv k_i$, $\beta_i^x \equiv \frac{k_i}{\langle k\rangle}$, $\beta_i^y \equiv \frac{k_i(1-\epsilon)}{\langle k\rangle}$, $f_i^x = p_i(1-g_i)$, $f_i^y = p_i g_i$.

For simplicity, the definitions were given in the case of a heterogeneous network with homogeneous efficacy but they can be easily generalized to heterogeneous efficacy, simply by adding a layer of classification in the model. The population that received prevention in class i can be further divided into efficacy classes,leading to a description with state variables $\tilde{y}_i^a$. Only the susceptibility β_i^y and the fraction of immunized individuals in class i $f(y_i)$ will be modified to $\beta(\tilde{y}_i^a) = \frac{k_i(1-\epsilon_a)}{\langle k\rangle}$, $f(\tilde{y}_i^a) = p_i g_i h_a$ where h_a is the probability that the immunization works with an efficacy ϵ_a.

To solve this stochastic process, one can use the WKB Ansatz (15) and perform a first order approximation in $\frac{1}{N}$ of the argument $S(q)$ in the exponent (16) [3].

$$P(\boldsymbol{q}) = e^{NS(\boldsymbol{q})}, \tag{15}$$

$$S(\boldsymbol{q} + d\boldsymbol{q}) \simeq S_0(\boldsymbol{q}) + \frac{1}{N}\nabla S_0 \Delta\boldsymbol{q}. \tag{16}$$

This is equivalent to assuming that the probability distribution follows a **large deviation principle**. Indeed, introducing weak noise allows for transitions between the metastable state and the absorbing one. These transitions correspond to rare events involving large fluctuations [11]. Finally, the infection and recovery rates in equation (14) are expanded in terms of $\frac{1}{N}$ but only the 0 order term is kept. This gives us an equation for $S_0(\mathbf{q})$ (17), which can be solved in some cases. S_0 will be referred to as the action.

$$\partial_t S_0(\boldsymbol{q}, t) = \sum_{j=1}^{n} \sum_{q=x,y} \sum_{\Delta q = \pm\frac{1}{N}} w(\boldsymbol{q} \longrightarrow \boldsymbol{q} + \mathbf{1_j}\delta q_j) \left[e^{\Delta q \partial_q S_0(\boldsymbol{q},t)} - 1\right]. \tag{17}$$

The analogy with a mechanical system can be made by defining the momentum $\boldsymbol{\theta} = \boldsymbol{\nabla} S$ which is the conjugate variable of the state variable $\boldsymbol{q}$. One can find the conditions under which one is allowed to use this analogy in the article by Onasager and Mashlup ([12]). It appears that in our case, the state variable $\boldsymbol{q}$ describing the stochastic process verifies all the conditions necessary to define a conjugate variable $\boldsymbol{\theta}$ and some further considerations on the action S_0 can be made as we will explain later. One can re-write equation (17) as a Hamilton-Jacobi equation since the right hand-side can be identified as the Hamiltonian of the system : $\partial_t S_0(\boldsymbol{q}, \boldsymbol{\theta}, t) = -H(\boldsymbol{q}, \boldsymbol{\theta}, t)$.

From this analogy, one can derive Hamilton's equations for the "position" $\boldsymbol{q}$ and the "momenta" $\boldsymbol{q}$. The variable $\boldsymbol{\theta}$ carries some information on the stochastic nature of the system. Let us take a diffusion process as described in section C.2 of [9]. The Langevin equation associated to this process is given by $\dot{x} =$

$F(x) + \sqrt{D}\xi(t)$ (where F comes from the deterministic force driving the motion of the particle while D is the diffusion constant and ξ is the noise term) and the equation of motion of the particle is $\dot{x} = \nabla_p H = F(x) + p$ (here F is comes again from an external force and p is the momentum of the particle). By comparing those two equations, the momentum can be identified to the noise term in the Langevin equation Onsager and Mashlup talk about a thermodynamic force that quantifies the tendency of the system to reach equilibrium. These considerations lead to a better understanding of the existing connection between the stochastic and the deterministic descriptions of the problem. Any fixed point of the MF description has a counter-part that is a fixed point of Hamilton's equations with a momentum $\boldsymbol{\theta} = \mathbf{0}$.

The definition of the Hamiltonian enables us to derive the usual equations of motion $\dot{q} = \frac{\partial H}{\partial \theta}$ and $\dot{\theta} = -\frac{\partial H}{\partial q}$, underlying the evolution of the state variable $\boldsymbol{q}$ and the associated momentum $\boldsymbol{\theta}$. The fixed points of these equations can then be obtained by setting the l.h.s to 0. Among them, one of the fixed point will correspond to the endemic state (with $\boldsymbol{q^*} \neq \mathbf{0}$, $\boldsymbol{\theta^*} = \mathbf{0}$) while the free-from-disease state will correspond to another fixed point (identified by $\boldsymbol{q^*} = \mathbf{0}$, $\boldsymbol{\theta^*} \neq \mathbf{0}$). It can be shown [2], that the free-from-disease state has a momentum $\theta_i^*(q) = ln(w^+/w^-)$.

Solving the Hamilton equations between two fixed extremes (endemic and absorbing state) provides the optimal path minimizing $\Delta S \sim S_0$ and therefore corresponding to the most likely path to extinction. This can be done with kinetic Monte-Carlo simulations on heterogeneous networks [5] or using an Iterative action minimization algorithm [4].

3.3 Hamilton's equations

In the following, we will apply the general theory to the specific problem of the SIS model coupled to a prevention strategy on a homogeneous network (single degree) with homogeneous efficacy of the prevention. We will determine the value of the endemic state, to be compared with the one obtained in a mean field fashion.

Let us consider again a population divided into several degree classes denoted by indices $i, j = 1, .., n$. In order to be consistent with the notation used earlier, we define $\tilde{x_i}$ = P(infected, class i , no prevention) and $\tilde{y_i}$ = P(infected, class i, prevention) while x_i = P(infected | class i , no prevention) and y_i = P(infected | class i, prevention). Again, we can use a more general variable $q_i = \tilde{x}$ or $\tilde{y}$. The infection and recovery rates can then be expressed in terms of the state variable q_i as follows:

$$w^+(q_i) = \lambda \left(f_i^q - q_i\right) \beta_i^x \sum_{j=1}^{n} N \alpha_j q_j, \tag{18}$$

$$w^-(q_j) = \mu N q_j. \tag{19}$$

The Hamiltonian can then be expressed in terms of the infection and recovery rates as in equation (17), setting $\Delta_{q_i} = \frac{1}{N}$. According to [9], any fixed point

of the mean-field dynamics has a corresponding fixed point in the Hamiltonian dynamics with the same $\boldsymbol{q}$ and $\boldsymbol{\theta} = 0$. This correspondence allows us to find easily the endemic state starting from Hamilton's equations. Indeed, one can set $\theta_i^q = 0$ and look for the fixed points of Hamilton's equations:

$$\tilde{x_i} = \frac{\lambda p_i (1 - g_i)\beta_i^x \xi}{\mu + \lambda \beta_i^x \xi}, \tag{20}$$

$$\tilde{y_i} = \frac{\lambda p_i g_i \beta_i^y \xi}{\mu + \lambda \beta_i^y \xi}, \tag{21}$$

$$\xi = \sum_{j=1}^{n} \alpha_j^x (x_j + y_j). \tag{22}$$

Remembering that $\tilde{x_i} = x_i f_i^x$ and $\tilde{y_i} = y_i f_i^y$, we can express the equations of the endemic state in terms of the conditional probabilities x_i and y_i. These equations are exactly the same as the ones obtained with the MF formalism (see equation (8) and (9)). Note that the free-from-disease state corresponds to the fixed point such that $\boldsymbol{x}, \boldsymbol{y} = \mathbf{0}$ and $\boldsymbol{\theta} \neq \mathbf{0}$. Finally, the most likely path to extinction, also called the "WKB" path will be the one connecting the endemic and the absorbing (free-from-disease) states. It can be obtained by noticing that$H(\boldsymbol{q}, \boldsymbol{\theta} = \mathbf{0}) = 0$ and since the Hamiltonian is constant along a trajectory, it will remain 0. This allows us to compute $\boldsymbol{q}(t)$ along the path as presented in [9] and can be used to determine the evolution of the prevalence along the trajectory to the extinction. This might be used to build an optimal strategy in which we choose to target the degree class that minimizes the prevalence as it evolves in time but it will not be presented here.

3.4 Eigenvalue problem

Equation (14) can be written in a vectorial form such that

$$\dot{P} = PW \tag{23}$$

where W is the transition rate matrix such that the element $W_{ij} = \text{rate(i}\rightarrow\text{j)}$. The stationary distribution P^* is obtained by setting the left hand side of equation (23) to 0. P^* appears to be an eigenvector of the matrix W, with eigenvalue 0. In fact, this eigenvector corresponds to the absorbing state in which the system will end up at some point.

However, since the endemic state is a long-lived meta-stable state, it can be assumed ([13],[3]) that the solution of the master equation is of the form :

$$P(\boldsymbol{I} \neq \mathbf{0}, t) \simeq \pi_I e^{-t/\tau}, \tag{24}$$

$$P(\mathbf{0}, t) \simeq 1 - e^{-t/\tau}. \tag{25}$$

Here, π_I is the quasi-stationary distribution (QSD) describing the process when the system is close to the endemic state while τ is the decay time from the endemic to the absorbing state (the mean time to extinction we are looking

for). Note that π is normalized to one if we sum over all states different from the absorbing state.

The spectrum of W contains information on the dynamic of the system. If we order the eigenvalues of W in the following way : $0 > Re(\lambda_1) \geq Re(\lambda_2) \geq Re(\lambda_3) \geq ...$ and the associated eigenvectors $(l_0, l_1, l_2, l_2, ...)$. For the SIS model, the first eigenvalue is always 0 and corresponds to the long term behaviour in which the system goes to the absorbing state, so l_0 is the stationary distribution. Since l_0 corresponds to a probability distribution, it should be renormalized such that all entries sum up to 1. All other eigenvalues are real and negative.

If we consider the reduced matrix W_0 (that is W from which we remove the first line and column), the first eigenvector corresponds to the quasi-stationary distribution and the associated eigenvalue is related to τ by the following : $\tau \propto \frac{1}{\lambda_1}$. If the quasi-stationary state was a real stable point, $\lambda_1 = 0$ and $\tau = \infty$. It is important to note that this approach assumes that the time it takes to reach the endemic state is much shorter than the mean time to extinction. This assumption holds if $|\lambda_2 - \lambda_1| >> |\lambda_1|$, which corresponds to the fact that the time spent in the transient before reaching the endemic state is negligible with respect to the mean extinction time [13]. Since most entries of the matrix W (or W_0) are 0, one can use efficient algorithms to compute its first eigenvalues [14]. However, the size of W increases as $O(N^n)$ which makes the computational time grow exponentially with the number of classes n.

We will now compare the results of the analytical approaches (see next section) with our numerical estimation only in the case of a single degree class divided into immunized and non immunized individuals, with homogeneous efficacy. This corresponds to having $p_j = 1$, $\alpha_j = k$.

The existence of an analytical approach is all the more useful as we increase the number of classes since the solution of the vectorial problem becomes more challenging. Indeed, this method presents clearly some limitations since the eigenvalues of the transition matrix becomes smaller and smaller as the system size increases (of order 10^{-15} for N = 500), which makes the algorithm highly sensitive to numerical errors, in addition to the computational time that increases exponentially with he number of classes. This is an important limitation when we are looking for a methodology to treat an increasing number of layers of the population in the model.

Note that an important assumption is verified also for low values of N: in order to approximate τ by $\frac{-1}{\lambda_1}$, it is indeed necessary [3] to assume that $|\lambda_1| << |\lambda_2 - \lambda_1|$. After numerical evaluation of the first eigenvalues, one can check that this assumption is well verified (in figure 5b).

To conclude this section, one can make a few observations on the QSD. Figure 3 shows the prevalence distribution in the quasi-stationary state. This gives us another way of comparing the analytical results with our numerical approach : one can see that the QSD is indeed maximal and centered around the endemic state prevalence, which was obtained by the method presented in the next section. It is also interesting to see how the QSD varies as we change some parameters. Increasing the system size N will decrease the skeweness of the distribution (Figure 4a). As the efficacy of the immunization increases (while

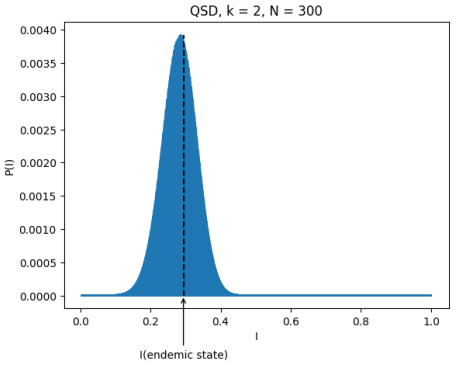


Figure 3: The eigenvector corresponding to the first eigenvalue is shown for k = 3, $\epsilon = 0.5$, g = 0.2 and N = 300. The dashed line in black corresponds to the value of the endemic state prevalence, computed analytically. The maximum of the QSD is localised around this value.

keeping the basic reproduction number constant) the QSD becomes slightly narrower and the probability of finding the system in the endemic state is higher. Finally, the number of contacts k will mainly impact the value of the endemic state prevalence: as k increases, the QSD is shifted towards the left and becomes narrower.

3.5 Analytical expression for the mean time to extinction

In order to have an analytical expression of the mean time to extinction, one can use the results presented in [2]. In this article, the extinction of long-lived stochastic population, caused by intrinsic noise is explored. The method is quite generic but the results diverge depending on the stability of the absorbing state. In our case, we will follow the method and results presented for a repelling absorbing state since the system is studied above the epidemic threshold (see section 2.2.1)

3.5.1 Exact expression

The main idea of the article is to give an analytical expression for the mean time to extinction by looking at both the quasi-stationary distribution (that describes well the system when the population reached or is around the endemic state) and then by linearizing the rates around the absorbing state, a recursive relation will give us an estimate of the solution close to the absorbing state. The connection between the two regimes is then made in order to get a solution valid in both regimes.

The first step consists in determining the quasi-stationary distribution whose expression is given in equation (24). π will be given by the normalized eigenvector associated to the first eigenvalue λ_1 (smallest in absolute value). In [2],

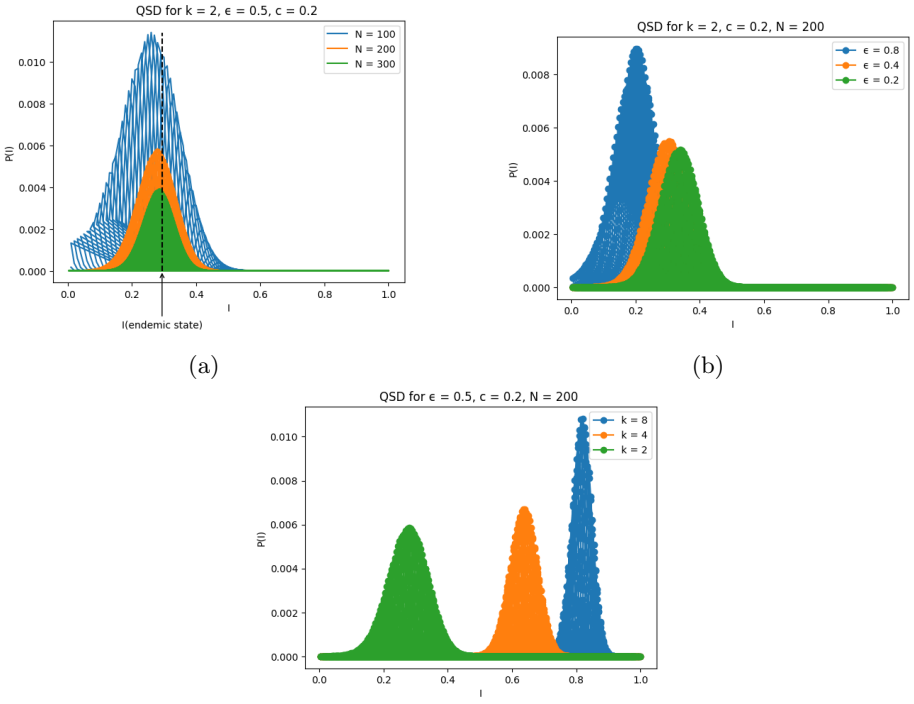


Figure 4: Dependence of QSD of the prevention on parameters characterizing the system. Figure 4a : The system size is modified while all other parameters are fixed. The main effect of increasing N is to reduce the value of the endemic state probability as well as the variance of the distribution.Figure 4b : Modifying the efficacy while keeping the basic reproduction number constant will not have a big impact on the shape of the distribution, except for the value of the endemic state probability. Figure 4c : Increasing the number of contacts shifts and deforms the distribution such that the value of the endemic state prevalence is higher and the distribution sharper.

the authors derive an expression for the quasi-stationary distribution, that is to be compared with the one obtained by the numerical solution of the eigenvalue problem. It is important to note that the following results hold only in the case where the rates can be written in the form : w(q) $= Nw_0(q) + u(q) + O(\frac{1}{N})$ where $w_0(q)$ and $u(q) = 0(1)$ when $q = 0(1)$.

In our case, one can see from equations (18) and (19) that this assumption is valid and that $u(q)$ is actually 0. Moreover, the infection and recovery rates scale like $O(N)$, as expected. Note that there are two types of trajectories with 0-energy: fast and slow modes (activation and relaxation). However, only the fast mode will matter in the case of a repelling absorbing state and it will be enough to match the recursive solution with the fast mode. Then, a recursive solution is given by solving the eigenvalue problem and by linearizing the infection and recovery rates for q_j small enough. Finally, for a single step process, the two solutions can be merged in order to compute the mean time to extinction. Note that the two solutions are compatible in the intermediate regime in which $1 << I_j^q << N$ or $\frac{1}{N} << \frac{1}{q_j} << \frac{1}{\sqrt{N}}$.

We will now generalize the results presented in [2] to higher dimensions. Let us consider a population divided into $n = n_k(n_\epsilon + 1)$. The variable $\boldsymbol{q}$ has $2n$ components where n is the number of degree classes. As always, the probability of belonging to degree class i and not having received prevention will be given by $f_i^x = p_i(1 - g_i)$ and the probability of being in class i and having received prevention will be given by $f(y_i) = p_i g_i$. For simplicity, we will assume homogeneous efficacy but the generalization to heterogeneous efficacy can be done by introducing another classification inside each degree class.

$$\tau = \sqrt{\frac{(2\pi)^n det\Sigma^{-1}}{N^{2n}}} e^{N\sum_{i=1}^{n}\sum_{q=x,y} f(q_i)\int_0^{q_i^*} ln\left(\frac{w^+(qi)}{w^-(qi)}\right)dq_i} e^{S_1(\boldsymbol{q^*})-ln(\boldsymbol{q^*})} \tag{26}$$

The evaluation of the quantity $\sum_{i=1}^{n}\sum_{q=x,y} f(q_i)\int_0^{q_i^*} ln\left(\frac{w^+(qi)}{w^-(qi)}\right)dq_i$ is not straightforward in a multidimensional problem because of the interdependence of the state variables characterizing each class of the population. Indeed, the rate $w^+(q_i)$ not only depends on q_i but on all the q_j, j = 1,..n, q=x,y (see equation (18)). Therefore, it is necessary to know the dependence of q_j on q_i along all the trajectory from 0 to q^* in order to calculate this integral. One could solve those trajectories numerically and introduce the result in the expression for the mean time to extinction which would lead to a semi-analytical estimation of the time spent in the endemic state (see section 4 for some insight on the numerical tools allowing for the evaluation of the trajectories).

3.5.2 Dependence in the system size

One can also proceed further in the analytical approach by using the results presented in an article by Clancy [6]. Depending on the type of heterogeneity that is introduced in the system, one can get an exact expression of the action variation ΔS along the path to extinction. This quantity will indeed be very

useful to get the dependence of the mean time to extinction τ on the system size N. The previous calculations have shown that τ grows exponentially with the system size and equation (26) can be written in the following form : $\tau = \frac{c}{N^{2n}} e^{N\Delta S}$ where c is a prefactor independent of the system size. Therefore, for N $\to \infty$, $\frac{ln\tau}{N} = \frac{lnc}{N} - 2n\frac{lnN}{N} + \Delta S$ can be accurately approximated by ΔS.

In [6], the author derives the Hamiltonian of the system as well as the endemic state which matches the one we obtained earlier (20)-(22). Then, a first theorem is demonstrated and gives the asymptotic limit of $\frac{ln\tau}{N}$ for N going to infinity, in the case where either the susceptibility (β) or the infectiousness (α) is heterogeneous. In the case where both are heterogeneous, an approximation of the action is given. The interested reader can find the full derivation in the original article, but here we will simply present the main elements that were used in order to derive the expression of the action associated to an SIS model coupled to an immunization strategy.

Since the action variation must be computed along a path of 0-energy, if one can find a quantity V such that $H(\boldsymbol{q}, \frac{\partial V}{\partial \boldsymbol{\theta}}) = 0$, then the calculation of ΔS is straightforward. Following a similar reasoning as the one presented in the article, one can derive an expression of the action in the general case where both the susceptibility and the infectiousness are heterogeneous:

$$\Delta S = -\sum_{i=1}^{n} \sum_{q=x,y} \int_0^{q_i^*} dq_i \frac{dV}{dq_i} = V(\mathbf{0}) - V(\boldsymbol{q}^*), \tag{27}$$

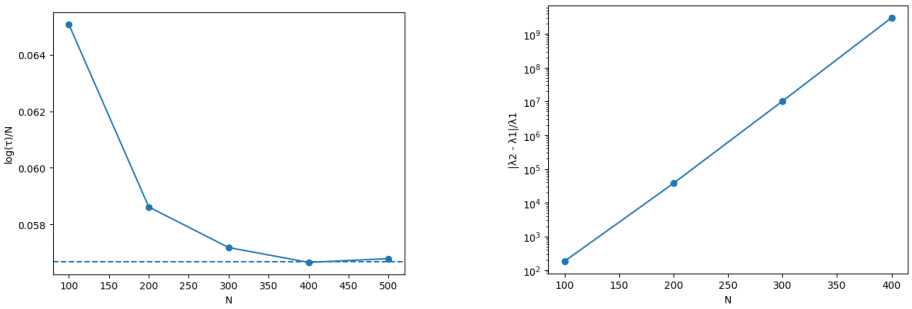
$$\frac{dV}{dq_i} = ln\left(\frac{q_i}{\frac{\lambda}{\mu}\beta(q_i)(f(q_i)-q_i)\sum_j \sum_q \alpha_j q_j}\right). \tag{28}$$

One can first consider the case of heterogeneous susceptibility with homogeneous infectiousness. According to our definition, the degree (number of contacts) must be the same for everyone but we can divide the population depending on their immunization state and even introduce heterogeneity in the efficacy. Here index $a = 0, 1, ...n_\epsilon$ will correspond to an efficacy class. We will denote by q_a the probability of being infected and in efficacy class a. The class $a = 0$ will correspond to individuals without immunization ($\epsilon_0 = 0$). For homogeneous efficacy, we would have only two classes : $q_0 = x$ and $q_1 = y$ with efficacy $\epsilon \neq 0$. The probability of not being immunized ($a = 0$) is, as always, $(1 - g)$ while the probability of being immunized with a drug that has efficacy ϵ_a is gh_a (for $a > 0$) which leads to defining $f_0 = (1 - g)$ and $f_{a>0} = gh_a$.

The settings presented above correspond to having homogeneous infectiousness ($\alpha_a = k_a = k, \forall a$) and heterogeneity in the susceptibility ($\beta_a = (1 - \epsilon_a)$). The analytical expression of the action is then given by:

$$\Delta S = V(\mathbf{0}) - V(\boldsymbol{q}^*) = \sum_a f_a ln\,(1 + \beta_a d) - \frac{\mu d}{\lambda \alpha}, \tag{29}$$

$$1 = \frac{\lambda}{\mu} \sum_a \frac{\alpha \beta_a f_a}{1 + \beta_a d}. \tag{30}$$



(a) (b)

Figure 5: Numerical solution of the eigenvalue problem. The calculations were performed for k = 2, $\epsilon = 0.5$ and g = 0.2. Figure 5a : The eigenvalue λ_1 was obtained with a function from the scipy library in python (straight line). The result was tested by comparing it with other algorithms (shift and invert algorithm and power inverse algorithm with LU decomposition [14]). The difference between the numerical value of $\frac{ln(\tau)}{N}$ (in solid line) and ΔS is decreasing with the system size. Figure 5b : The ratio $|\lambda_2 - \lambda_1|/|\lambda_1|$ is shown here and it is growing exponentially with the system size N.

Note that d is the unique positive solution of equation (30) when $\alpha_i > 1$. This expression is advantageous because it allows to get immediately ΔS, simply using the expression of the endemic state derived earlier in equations (20)-(22), without even having to determine the corresponding momentum, also in the case of heterogeneous efficacy.

By extending the approach presented in ref [6], one can get a precise expression for the action in the case of many degree classes coupled to an immunization strategy (both infectiousness and susceptibility are heterogeneous).

$$\Delta S = \sum_i \sum_a \left[f_a^i ln\left(1 + \beta_a^i d\right) - q_a^i \right], \tag{31}$$

$$1 = \frac{\lambda}{\mu} \sum_i \sum_a \frac{\alpha_i \beta_a^i f_a^i}{1 + \beta_a^i d}. \tag{32}$$

Again, one can refer to Figure 5a in order to compare the numerical estimation of the mean time to extinction and the analytical expression for a single degree class (k=2) with an immunization strategy (efficacy $\epsilon = 0.5$ and coverage $c = 0.2$).

4 Designing an optimal strategy

The reproduction ratio is an important predictor of the emergence and evolution of infectious disease outbreak. Its value determines whether the disease can

trigger an outbreak, as it indicates whether each infected individual infects more than one person.

However, when the disease is endemic, public health often focuses on minimizing another significant epidemiological measure: the prevalence of the disease within the population. This measure represents the fraction of infected individuals and indicates the extent of the disease presence within the population.

In the previous section, a new formalism was introduced to effectively address the stochastic process underlying the dynamics of our problem. This approach enables us to examine a new quantity: the mean time to extinction. This measure corresponds to the duration spent in the endemic state before transitioning to a disease-free state due to stochastic fluctuations. The mean time to extinction represents an alternative pathway for developing an optimal distribution strategy. On the one side, it prompts the question: What is the relevant measure to consider when the coverage of the prevention tool is limited? On the other side, it leads to the realization that the decision to optimize the basic reproduction number or the prevalence, or to focus on the time to extinction, depends on the specific circumstances.

Understanding the definitions of these three measures helps us identify when each measure becomes significant. To minimize the risk of outbreak, the reproduction number should be minimized. Once an outbreak has taken place, efforts can be focused on containing the infection and reducing the mean time to extinction. This measure becomes relevant when the number of infected individuals is relatively small. It is important to note that the duration spent in the metastable endemic state increases exponentially with population size. Thus, this measure becomes particularly interesting when an outbreak has occurred within a specific population which has loose ties with other populations, ie. is spatially contained. In such a scenario, if a timescale separation between the mean time to extinction τ and the timescale spreading to other populations exists, then one can consider the sub-population in which the outbreak subsides as the effective population and the spread to larger regions can be prevented by minimizing the mean time to extinction. Finally, when the endemic state is well established in a large population, the most suitable measure to consider is the prevalence of the disease. At this stage, the objective becomes the complete elimination of the disease.

4.1 Prevention

Let us first look at how we can minimize the risk of having an outbreak. As mentioned in section 2, the basic reproduction number for an SIS model coupled to a prevention strategy in a population with heterogeneous number of contacts and efficacy of prevention can be expressed as:

$$R_0 = \frac{\lambda}{\mu} \frac{\langle k^2 \rangle - \langle \epsilon g k^2 \rangle}{\langle k \rangle}. \tag{33}$$

In the previous expression, $\langle\rangle$ refer to an average over both degree distributions and efficacy distributions. Now, in order to determine which population

should be given priority to receive prevention, one should determine the prevention distribution $\boldsymbol{g}$ that will minimize the basic reproduction number. However, since $\boldsymbol{g}$ is a distribution, R_0 is a functional and this would correspond to a variational problem.

Even though an analytical solution can be obtained for R_0, the same cannot be applied to the prevalence or the mean time to extinction, due to their explicit dependence on the endemic state that doesn't have an explicit analytical solution in the general case. Therefore, our goal will be a little bit different since we will focus on finding the value k^* that maximises the response function of each quantity (here the basis reproduction number) defined as follows:

$$r(k_m) = -\frac{\partial R_0}{\partial(N p_m g_m)}. \tag{34}$$

The definition above corresponds to looking at the variation in the basic reproduction number as we make a small variation in the amount of prevention given to degree class k_m Inserting equation (33) into equation (34) leads to the following expression :

$$r(k_m) = \frac{\langle\epsilon\rangle k_m^2}{N\langle k\rangle}. \tag{35}$$

It is clear that r(k_m) is maximal for $k_m \longrightarrow \infty$ and one can immediately conclude that in the prevention phase, when the goal is to prevent an outbreak of the disease, the best strategy is the well-established risk-based distribution strategy: in order to minimize the risk of outbreak, one should give prevention in priority to individuals with the highest degree (i.e. with the largest number of contacts).

4.2 Containment and elimination

In a scenario where the disease is well established only for a fraction of the population, one can expect that the disease might die out thanks to stochastic fluctuations. In that context, we can try to determine what class to prioritize to minimize the time spent in the metastable state. In the previous section, we have applied existing results from the literature in order to get an analytical expression for the exponential rate of the mean time to extinction. Once again, we define a response function that informs us on the variation in the exponential rate of the mean time to extinction as we perform a slight change in the fraction of prevention given to class m:

$$t(k_m) = -\frac{\partial \Delta S}{\partial(p_m g_m)}. \tag{36}$$

When the outbreak already occurred, one will focus on eliminating the disease by minimizing the response function associated to the prevalence (i.e. the fraction of infected individuals).

$$f(k_m) = -\frac{\partial I}{\partial(p_m g_m)}. \tag{37}$$

The final expression of the response function can be obtained by deriving the action defined in (31). Note that the following expression is for homogeneous efficacy, but a similar result can be obtained in a heterogeneous case.

$$t(k_m) = ln\left(\frac{1+\beta_m^x d}{1+\beta_m^y d}\right) - \frac{1}{p_m}\sum_i \left(\frac{f_i^x \beta_i^x}{1+\beta_i^x d} + \frac{f_i^y \beta_i^y}{1+\beta_i^y d}\right)\frac{d\xi}{dg_m} - f(k_m). \quad (38)$$

The prevalence being given by $\sum_i p_i\left((1-g_i)\tilde{x_i} + g_i\tilde{y_i}\right)$, the response function related to this quantity is given by:

$$f(k_m) = x_m - y_m - \frac{1}{p_m}\sum_i p_i\left((1-g_i)\frac{dx_i}{dg_m} + g_i\frac{dy_i}{dg_m}\right). \quad (39)$$

It is clear from this that t and f will not have the same behaviour with respect to the degree k and one can try to find and compare the degree class k that will maximize those functions.

4.2.1 Early-stage distribution strategy for prevention

Let us consider a naive population that did not receive any prevention yet. This would be valid in the beginning of a prevention campaign, and would be equivalent to taking a value of coverage $c = 0$. One can use the previous relationship between f and k as well as the results presented in [1] in order to understand better the different strategies that one should adopt in the containment and elimination phase. At 0 coverage, one has $\forall i = 1, ..., n\ g_i = 0$ and

$$f(k_m)|_{g=0} = (x_m - y_m)|_{g=0}\left(1 + \frac{\Psi\hat{\lambda}k_m}{1-\Phi}\right), \quad (40)$$

$$\begin{aligned} t(k_m)|_{g=0} = ln\left(\frac{1+\hat{\lambda}kz}{1+(1-\epsilon)\hat{\lambda}kz}\right) \\ - (x_m - y_m)|_{g=0}\left(1 + \frac{\hat{\lambda}k_m}{1-\Phi}(\Psi - \chi)\right). \end{aligned} \quad (41)$$

Note that we have used the quantity $\hat{\lambda} = \frac{\lambda}{\mu\langle k\rangle}$ that is simply a rescaling of the overall infection rate. One can give a physical interpretation to the quantity $z = \xi|_{g=0} = \langle xk\rangle = \sum_j p_j x_j k_j$. Indeed, x_k being the probability of being infected given that we belong to class k (again, we're working in the limit of no prevention), $p_k x_k$ will be the joint probability of being infected and in class k and $\sum_j p_j x_j k_j = z$ will be the average degree of those infected. The quantities $\Psi = \langle\frac{k}{(1+\hat{\lambda}kz)^2}\rangle$, $\Phi = \langle\frac{\hat{\lambda}k^2}{(1+\hat{\lambda}kz)^2}\rangle$ and $\chi = \langle\frac{k}{1+\hat{\lambda}kz}\rangle$ are averages over the degree distribution.

The expression of f in equation (40) is a result taken from the article [1]. The authors also showed that for ϵ below a certain threshold ϵ_c, $f(k_m)|_{g=0}$ is

maximal for some $k* < \infty$. Above this threshold, the maximum of f is at ∞ and the risk-based strategy holds.

Equation (41) can be decomposed as the sum of two terms that will determine whether we can find a similar behaviour for t or no. The first term is a positive quantity that increases with k and will therefore tend to increase the value of k^*, going in the direction of the risk-based strategy. The second term is a positive contribution since $\Psi - \chi$ is negative, so the resulting term is very similar to $f(k)$. To proceed, one needs to determine the derivative of t with respect to k :

$$t'(k) \sim 2\frac{\chi - \Psi}{1 - \Phi} + \frac{(2-\epsilon)z + 2\hat{\lambda}z^2k(1-\epsilon)}{1 + (2-\epsilon)\hat{\lambda}zk + \hat{\lambda}^2k^2z^2(1-\epsilon)}\left(1 - \frac{\hat{\lambda}k(\chi - \Psi)}{1 - \Phi}\right). \quad (42)$$

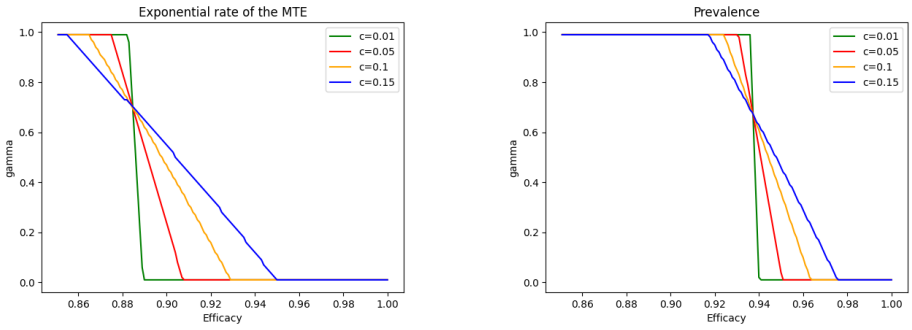
When $k \longrightarrow \infty$, $t'(k)$ tends to $2\epsilon\frac{\chi-\Psi}{1-\Phi}$. Moreover, the value of k^* such that $f'(k^*) = 0$ is not physical (negative). This indicates that the response function for the mean time to extinction is maximised by the largest value of k, going in the direction of risk-based strategies.

Those results have important implications : depending on whether one wants to prevent, control or mitigate an infectious disease, the distribution strategy of the prevention tool is not the same and one must adjust its strategy as the disease spreads among the population. When the prevention tool is distributed in a population that did not receive any prevention yet (i.e. $g_i = 0 \forall i$), the best strategy to control the spread of the disease and contain it is to focus on individuals with the highest number of contacts. Instead, when the outbreak already occurred, alternative strategies (targeting lower degree nodes) should be considered.

4.2.2 Distribution strategy for a finite coverage of the population

Even though further analytical analysis is challenging, especially if one wants to go beyond the 0 coverage limit, one can try to understand whether the mean time to extinction and the prevalence behave in a similar way. We know from a previous work [1] and from the previous observations that for a prevention tool with efficacy lower than 100%, the risk based strategy is not always the optimal choice if we want to maximize f or t. Using a numerical solution of the endemic state for a population with two degree classes, we can show that, at fixed basic reproduction number, both the prevalence and the exponential rate associated to the mean time to extinction will not be minimized by a risk based strategy when the efficacy is finite (see figure 6). Moreover, this result shows that the two quantities behave differently: if we define ϵ_c as the minimal value of efficacy needed for a risk based strategy to work, we can see that this critical value is lower for the mean time to extinction than for the prevalence. Moreover, the interval of values for which a mixed strategy is better than targeting a specific class is larger for the mean time to extinction.

Even though an analytical analysis of the response functions $t(k)$ and $f(k)$ is necessary to provide a complete strategy, one can at least conclude from



(a) (b)

Figure 6: The optimal value of gamma (i.e. fraction of prevention given to the lowest degree class) is shown as a function of the efficacy when R_0 is kept at a constant value > 1. Figure 6a shows $\gamma^*(\epsilon)$ for the mean time to extinction and for different values of coverage c. Figure 6b shows $\gamma^*(\epsilon)$ for the prevalence.

these numerical results that one should consider different strategies during the containment and the elimination phases. For the first, in which fluctuations might lead to a disease extinction, a mixed/ non-selective strategy might be the optimal choice for a wide range of efficacy values. For the latter, a non selective strategy might also work, but the range of efficacy values for which it is the case is smaller. As the coverage increases, the risk-based strategy becomes more relevant but the range of efficacy value for which a non-selective strategy is more efficient is also larger.

Again, these results have to be taken carefully : even though this behaviour was checked by varying the number of contacts in degree class 1 and 2 as well as the overall transmission rate λ, its validity is not ensured for a larger number of degree classes.

Finally, addressing the problem of heterogeneous efficacy, one can derive similar expression for $f(k)$ and $t(k)$, simply by introducing the average over the efficacy values. One can first look at the case where the efficacy of the prevention tool is all-or-nothing: an individual may be totally resistance to the medicine, or the latter could be totally inefficient if not taken in a regular way. In that situation, the heterogeneity in the efficacy (that can have a value of ϵ with probability h or 0 with probability $1-h$) simply reduces to a re-scaling of the coverage : $c \longrightarrow hc < c$. The coverage being closer to 0, the results presented at 0-coverage become more relevant. In the case of a more complicated efficacy distribution, again, an analytical advancement is more difficult to get.

5 Conclusion

In conclusion, this master thesis aimed to address the challenge of finding the optimal strategy to distribute a tool of primary prevention of infectious disease

(vaccine or prophylaxis for instance) with limited coverage in a heterogeneous population. The study focused on the context of infectious diseases, such as HIV, where risk-based strategies have been traditionally employed. These strategies aim to optimize the distribution of limited prevention tools by targeting individuals or communities that are at higher risk of creating super-spreading events. However, this is accompanied by significant drawbacks since it requires gathering a lot of information on the individuals in order to determine whether they represent a target for the prevention or no. Moreover, those strategies do not give significant results in improving the effectiveness of the campaign of distribution for the prevention tool.

Through the introduction of both a mean field and a stochastic view of an epidemiological model of a disease spread coupled with a prevention strategy, this research provided some insights into an adaptive approach. By analyzing various epidemiological indicators, it was possible to devise a strategy that takes into account the specific phase of the epidemic. The results indicate that a risk-based strategy, which prioritizes individuals with a higher number of contacts, does not always yield optimal results.

However, it is important to acknowledge the limitations of this work. We have seen that depending on the epidemic phase, the goal distribution strategy of the preventive tool should be adjusted: during the prevention phase where the outbreak did not occur yet and the fraction of infected individuals is very small, one should try to minimize the basic reproduction number by prioritizing individuals at highest risk of creating an outbreak (i.e. individuals with highest number of contacts). When the outbreak already occurred but is spatially contained, one should consider the mean time to extinction while the relevant quantity to look at when the outbreak occurred in the whole population is the prevalence of the disease (the probability of getting infected). However, the optimal distribution of the preventive tool for minimizing those epidemiological indicators could not be derived for finite coverage. Therefore, further analytical advancements are crucial in order to develop a systematic distribution strategy. Nonetheless, this research demonstrates that a risk-based strategy is not the most effective, particularly at the initial stages of distribution when the coverage is at zero.

Additionally, the inclusion of heterogeneous efficacy, characterized by all-or-nothing effectiveness, among those who received the prevention tool shows interesting results. This approach effectively lowers the value of the coverage, suggesting the importance of considering the varying levels of protection within the population.

In summary, this study contributes to the ongoing efforts in the field of epidemiology by highlighting the challenges of distributing a preventive medicine or vaccine with limited coverage. It emphasizes the need for further analytical advancements and suggests alternative strategies to improve the effectiveness of distribution efforts. Ultimately, these findings have implications for public health decision-making in combating transmittable diseases.

References

[1] B Steinegger, I Iacopini, A S Teixeira, A Bracci, P Casanova-Ferrer, A Antonioni, and E Valdano. Non-selective distribution of infectious disease prevention may outperform risk-based targeting. *Nature Communications*, 13(1):3028, 2022.

[2] Michael Assaf and Baruch Meerson. Extinction of metastable stochastic populations. *Phys. Rev. E*, 81:021116, Feb 2010.

[3] Peter Ashcroft. *The Statistical Physics of Fixation and Equilibration in Individual-Based Models*. Springer Theses. Springer Cham, 2016.

[4] Brandon S. Lindley and Ira B. Schwartz. An iterative action minimizing method for computing optimal paths in stochastic dynamical systems. *Physica D: Nonlinear Phenomena*, 255:22–30, 2013.

[5] Elad Korngut, Jason Hindes, and Michael Assaf. Susceptible-infected-susceptible model of disease extinction on heterogeneous directed population networks. *Phys. Rev. E*, 106:064303, Dec 2022.

[6] Clancy D. Persistence time of sis infections in heterogeneous populations and networks. *J Math Biol.*, 77, Sep 2018.

[7] Assaf M Hindes J. Degree dispersion increases the rate of rare events in population networks. *Phys Rev Lett.*, 123:068301, Aug 2016.

[8] Christopher J. Clancy, Damian Pearce. The effect of population heterogeneities upon spread of infection. *Journal of Mathematical Biology*, 67, Oct 2013.

[9] Javier Aguilar. *Generation of stochastic trajectories : Application to complex systems*. PhD thesis, Universitat delle illes Balears, 2023.

[10] Jason Hindes and Ira B. Schwartz. Epidemic extinction and control in heterogeneous networks. *Phys. Rev. Lett.*, 117:028302, Jul 2016.

[11] Paul C. Bressloff. *The WKB method, path integrals, and large deviations*, pages 601–680. Springer International Publishing, Cham, 2021.

[12] L. Onsager and S. Machlup. Fluctuations and irreversible processes. *Phys. Rev.*, 91:1505–1512, Sep 1953.

[13] Ross JV Keeling MJ. On methods for studying stochastic disease dynamics. *J R Soc Interface.*, Feb 2008.

[14] Gene H. Golub and Charles F. van Loan. *Matrix Computations*. JHU Press, 2013.