Epidemic control using stochastic and deterministic transmission models: performance comparison with and without parameter uncertainties

Julien Flaig* Nicolas Houy†

November 12, 2022

Abstract

Background: The spread of infectious diseases can be modeled using deterministic models assuming a continuous population or stochastic models assuming a discrete population. A stochastic model can be approximated by its deterministic counterpart under some conditions, however deterministic models are unable to captures effects pertaining to the discrete nature of populations, for instance epidemic extinction. We look into the choice of a model – deterministic or stochastic – from the perspective of decision making. We are interested in the influence of parameter uncertainties and of the quality of the estimates used to inform decisions.

Method: We consider an emerging disease in a closed population whose spread can be modeled by a stochastic SIR model or its deterministic version. Our objective is to minimize the cumulative number of symptomatic infected-days over the course of the epidemic by picking a vaccination policy out of three available options. We consider four decision making scenarios: based on the stochastic model or the deterministic model, and informed or under parameter uncertainty. We also consider different sample sizes covering parameter draws, stochastic model runs, or both depending on the scenario. We estimate the average performance of decision making in each scenario and for each sample size.

Results: The model used for decision making has an influence on the picked policies. The best achievable performance is obtained with the stochastic model, knowing parameter values, and for a large sample size. For small sample sizes, the deterministic model can outperform the stochastic model due to stochastic effects, both in the informed and the uninformed cases. Starting with the deterministic model under uncertainty, resolving uncertainties brings more benefit than switching to the stochastic model in our example.

*Corresponding author. Epidemiology and Modelling of Infectious Diseases (EPIMOD), F-69002 Lyon, France. Email: jflaig@epimod.fr

†University of Lyon, Lyon, F-69007, France; CNRS, GATE Lyon Saint-Etienne, F-69130, France. Email: houy@gate.cnrs.fr.

Conclusion: This article illustrates the interplay between the choice of a type of model, parameter uncertainties, and sample sizes. It points to issues to be carefully considered when attempting to optimize a stochastic model.

Keywords: epidemic control, emerging disease, stochastic models, deterministic models, stochastic effects, value of information, disease X.

1 Introduction

Infectious diseases spread in populations made up of discrete units, and events pertaining to this spread (whether individual encounters, recoveries, deaths, births or immigration) are essentially random. Stochastic transmission models – in which the population is assumed to be discrete – take these features into account. But stochastic models do not lend themselves easily to mathematical analysis. As for numerical approaches, they can be resource intensive as they may require many Monte Carlo simulations of epidemic trajectories. Early modeling efforts were then focused on deterministic transmission models in which the population is modeled as a continuum whose dynamics are represented by differential equations. A deterministic model approximates its corresponding stochastic model "for large populations". Loosely speaking, as the population increases, the proportion of individuals, say, getting infected at a given time converges to the probability that any particular individual gets infected at that time by the law of large numbers.

Importantly, absolute numbers of individuals do not matter in a deterministic model: populations can be normalized or rescaled without changing the qualitative behavior of the system. This is of course a limitation in cases where absolute numbers of individuals do matter. The most obvious and widely studied of such situations are those in which extinction is possible, whether upon introduction of a new pathogen in a population, or in the endgame of an eradication attempt [1–3]. For instance, in a deterministic framework, the basic reproduction number R_0 tells whether an emerging infectious disease will be able to spread in a population ($R_0 > 1$) or not ($R_0 < 1$) irrespective of the initial number of infectious individuals, which may assume any positive real value. In a stochastic model, by contrast, an emerging infectious disease may still become extinct before causing an outbreak if $R_0 > 1$ with a probability depending on the (discrete) number of initially infectious individuals [4]. Thus, there is a rich literature using stochastic models to estimate the probability of an outbreak or of eradication in different scenarios [3, 5–7]. Beside the possibility of extinction, absolute numbers of individuals also matter where control means come in discrete units (think of hospital beds), although such scenarios have been less studied [8].

As for comparisons of deterministic and stochastic approaches, many previous studies compared the respective ability of deterministic and stochastic models to account for observed patterns for different parameter values and population sizes [2, 9–13]. These problems were chiefly put in terms of population biology – typically: under what conditions does a deterministic approximation give an accurate picture of a stochastic system? The objective of the present article is to shift perspective from population biology (or dynamics)

to the perspective of decision making. Deterministic models are more convenient to use, but what is the cost of making decisions based on a deterministic model where a stochastic model would be more appropriate from a "biological" perspective? This issue is closely related to the problem of decision making under uncertainty. Indeed, the probability of stochastic effects such as extinction depends on the value of parameters that are most often uncertain. So, what is comparatively more costly: parameter uncertainties or the inappropriate use of a deterministic model? In addition, the quality of policy performance estimates – and hence of decision making – depends on the number of Monte Carlo samples used to produce these estimates. When using a stochastic model, Monte Carlo samples include samples of stochastic epidemic trajectories. In the case of decision making under parameter uncertainty (whether based on stochastic or a deterministic model), samples include epidemic trajectories for different random draws of parameter values. What, then, is the influence of sample size on the respective performance of stochastic and deterministic models, with or without parameter uncertainties?

We will answer these questions using two generic SIR transmission models: a stochastic model and its deterministic counterpart. In our scenario, a new pathogen is introduced in a susceptible population and the objective is to control its spread through single or double dose vaccination given a number of available vaccine doses. We consider four possible scenarios for decision making: using a stochastic model or its deterministic version, and deciding under parameter uncertainty or knowing parameter values. Notice that in these scenarios, the level of available information about the disease is a matter of context, while the use of a deterministic or stochastic model is a modeling choice (although it may depend on the context, e.g. on computational or organizational constraints). Also, we do not intend to describe the control of a specific disease or to make specific policy recommendations, but to give a qualitative and intuitive understanding of how the choice of a deterministic or stochastic model may influence decision making, and of the interplay between this choice and the level of available information about model parameters.

Yuan et al. [14] compared the use deterministic and stochastic models for decision making in a similar way as we do: they determined optimal vaccination policies based on a stochastic and a deterministic model, and assessed the performance of both policies when applied to the stochastic system. They showed for what parameter values (assumed to be known to the decision maker) the difference in performance was more or less significant. We look at the problem from a different angle by considering the perspective of a decision maker endowed with a certain level of information about model parameters, and deciding whether to use a deterministic or a stochastic model and with how many Monte Carlo samples. Another more technical difference is that Yuan et al. [14] numerically integrate

the master equation of the stochastic model using the method proposed by Jenkinson and Goutsias [15] while we use a version of the Gillespie algorithm [16] to generate samples and estimate means.

We emphasize that the present article is not concerned with optimization methods or optimal control of stochastic models. For this we refer to [14, 17–20]. We restrict ourselves to a limited number of available policies (three), so that their respective performances can easily be estimated exhaustively. Solving the optimization problem then simply consists in picking the best policy based on these estimates. This does not make our results any less relevant: they apply to decision making in any situation where stochastic effects may be critical, whatever the optimization methods used to make decisions.

Finally, it is worth mentioning the distinction between demographic and environmental stochasticity [12]. We only consider demographic stochasticity, that is the randomness of events involving discrete individuals. Environmental stochasticity refers to the fact that parameter values may be noisy, even assuming a continuous population. The results presented in this article do not so much revolve around the noise involved by stochasticity, but rather around effects pertaining to the discrete nature of populations (in our case, the possibility of extinction). Randomness is necessary for our analyses to be relevant, however it is not sufficient: we also need a discrete population. See Grandits et al. [21] for an example of decision making based on a model assuming a continuous population with random variations.

The remainder of the article is organized as follow. Our method is presented in Section 2: the disease transmission models and parameter uncertainties (Sections 2.1 and 2.2 respectively), how we assume decisions to be made in the four considered scenarios (Section 2.3), and the validation method used to assess the performance of decision making in each case (Section 2.4). The results are presented in Section 3. Section 3.1 provides a general understanding of the behavior of our stochastic transmission model and of its deterministic counterpart. Section 3.2 shows the performance of decision making based on our stochastic model and its deterministic version under parameter uncertainty. Section 3.3 shows the same but assuming that parameter values are known to the decision maker. In Section 3.4, we compare decision making in all four scenarios: based either on the stochastic or deterministic model, and either under parameter uncertainty or knowing parameter values. Section 4 concludes.

Parameter	Description
R_0	Reproduction number of unvaccinated individuals.
γ	Rate of recovery.
ρ_1	Relative infectiousness of individuals having received 1 dose of vaccine.
ρ_2	Relative infectiousness of individuals having received 2 doses of vaccine.
σ_1	Relative susceptibility of individuals having received 1 dose of vaccine.
σ_2	Relative susceptibility of individuals having received 2 doses of vaccine.
p_0	Probability of symptomatic infection of unvaccinated individuals.
p_1	Probability of symptomatic infection of individuals having received 1 dose of vaccine.
p_2	Probability of symptomatic infection of individuals having received 2 doses of vaccine.

Table 1: Transmission model parameters.

2 Method

2.1 Disease transmission models and intervention policies

We consider a total population of $n = 100$ individuals, which corresponds to a hospital ward or a nursing home. The population is assumed to be closed. Individuals can be susceptible, infected and infectious, or recovered. Individuals may receive one, two, or no dose of vaccine. We assume that there is no waning of vaccination protection. We assume homogeneous mixing. All individuals get infected with contact rate β and they recover at rate γ – the basic reproduction number is $R_0 = \beta/\gamma$. However, individuals who received $i \in \{1, 2\}$ vaccine doses are less susceptible by a factor σ_i and less infectious by a factor ρ_i compared to unvaccinated individuals. Individuals who received $i \in \{0, 1, 2\}$ doses of vaccine have have a probability p_i of developing symptoms. Model parameters are summarized in Table 1.

At time 0, one person is infected and the others are susceptible. We assume that this initial case was detected, and our objective is to minimize the cumulative number of symptomatic infection-days over the course of the disease. At time 0, 50 vaccine doses are available and can be administered either as double doses to 25 individuals, or as single doses (to 50 individuals). We denote these policies "0-50" and "50-0" respectively. We also leave the possibility not to implement vaccination (policy "0-0"), even though this policy

should be dominated. The set of available policy options is then $A = \{0\text{-}0, 0\text{-}50, 50\text{-}0\}$.

A policy is picked based either on a deterministic compartmental model M_0, or on its stochastic counterpart M_1. The compartmental models are described in more details in Appendix A. In the following, with a slight abuse of notation, we will use the same notations for deterministic and random variables.

2.2 Parameter uncertainties

We assume that the structure of the model is known as well as the number of initially infected individuals. Under parameter uncertainty, other parameters are uncertain but prior distributions are available to the decision maker. The parameter prior distributions are given in Appendix B and Figure 1 shows 512 draws from these distributions. For the sake of simplicity, we restrict $p_0 = 1$. In the base case, we also fix $\gamma = 0.1$ so that all uncertainty regarding the spread of disease without intervention is summarized by uncertainty about a single parameter (R_0 or β). All other uncertainties concern the respective levels of protection offered by single and double dose vaccination. Naturally, we impose that $p_1 > p_2$, $\rho_1 > \rho_2$ and $\sigma_1 > \sigma_2$.

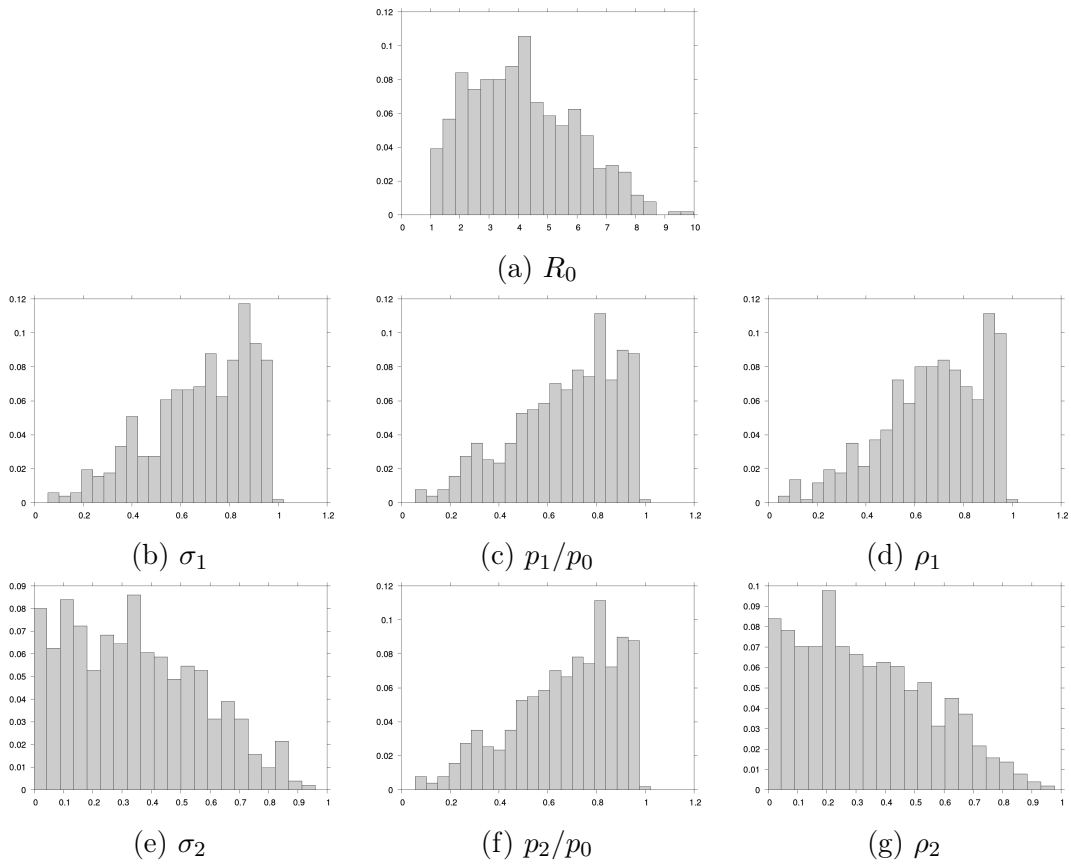
2.3 Decision making

Given parameter values ξ, policy $a \in A$ can be evaluated based on deterministic model M_0 or based on stochastic model M_1. Let $C_i(a, \xi)$ the cumulative number of symptomatic infected-days over the course of the epidemic estimated with model M_i. $C_0(a, \xi)$ is obtained directly by solving deterministic model M_0. In the stochastic framework, $C_1(a, \xi) = E\left[C(a, \xi)\right]$, where C is a random variable – the number of symptomatic infected-days obtained with a single run of M_1. C_1 is estimated as the average of realizations of C over a number of runs of M_1. This number will be varied in our analyses.

Under uncertainty about parameters ξ, policy $a_i^0 = \arg\min_{a \in A} E_\xi\left[C_i(a, \xi)\right]$ is then picked based on model M_i: a policy is picked to minimize the expected number of symptomatic infected-days over the prior distribution of parameters. If the true value of ξ is known, policy $a_i^* = \arg\min_{a \in A} C_i(a, \xi)$ is picked based on model M_i.

2.4 Validation of chosen policies

We assume that stochastic model M_1 is an accurate representation of reality. Thus, given parameter values ξ, the outcome of chosen policy a once implemented is the result $C(a, \xi)$ of a single stochastic epidemic trajectory of M_1. We evaluate the performance *a priori* of policy a by estimating $E_\xi\left[C_1(a, \xi)\right] = E_\xi\left[E\left[C(a, \xi)\right]\right]$, the expected outcome



(a) R_0

(b) σ_1 (c) p_1/p_0 (d) ρ_1

(e) σ_2 (f) p_2/p_0 (g) ρ_2

Figure 1: 512 draws from the parameter prior distributions.

over possible parameter values and epidemic trajectories. This performance can also be expressed in terms of averted symptomatic infected-days compared to the no vaccination policy: $E_\xi\left[C_1(\text{0-0},\xi) - C_1(a,\xi)\right]$. Estimating the performance of decisions made under different scenarios – parameter uncertainties, type of model, Monte Carlo sample size – allows to compare the cost of uncertainties to the cost of using the "wrong" model or a too small sample size.

3 Results

3.1 Stochastic effects

Figure 2 illustrates the spread of the disease for each of the three available policy options, as estimated with deterministic model M_0 and stochastic model M_1, for each of the 512 parameter draws shown in Figure 1. For each parameter draw ξ (y-axis), a green dot shows the cumulative number $C_0(a,\xi)$ of symptomatic infected-days under

Policy a	$E_\xi\left[C_0(a,\xi)\right]$ **estimate**	$E_\xi\left[C_1(a,\xi)\right]$ **estimate**
0-0	941.371 (95% CI: 941.169 – 941.574)	681.479 (95% CI: 665.501 – 697.457)
0-50	726.308 (95% CI: 726.022 – 726.593)	439.498 (95% CI: 424.672 – 454.324)
50-0	726.012 (95% CI: 725.646 – 726.378)	452.722 (95% CI: 435.771 – 469.673)

Table 2: Estimate expected outcome of the three available policy option based on M_0 and M_1. Averages over $512 \times 2{,}048 = 2^{20}$ simulations. In the stocastic case: 512 parameter draws, and 2,048 runs of M_1 per parameter draw.

policy $a \in A$ estimated with deterministic model M_0. Parameter draws are ordered by increasing $C_0(a,\xi)$. The heat map shows the distribution of the outcomes $C(a,\xi)$ of 2,048 runs of stochastic model M_1 for each parameter draw ξ and policy a. White dots show the average $C_1(a,\xi)$ of the 2,048 outcomes.

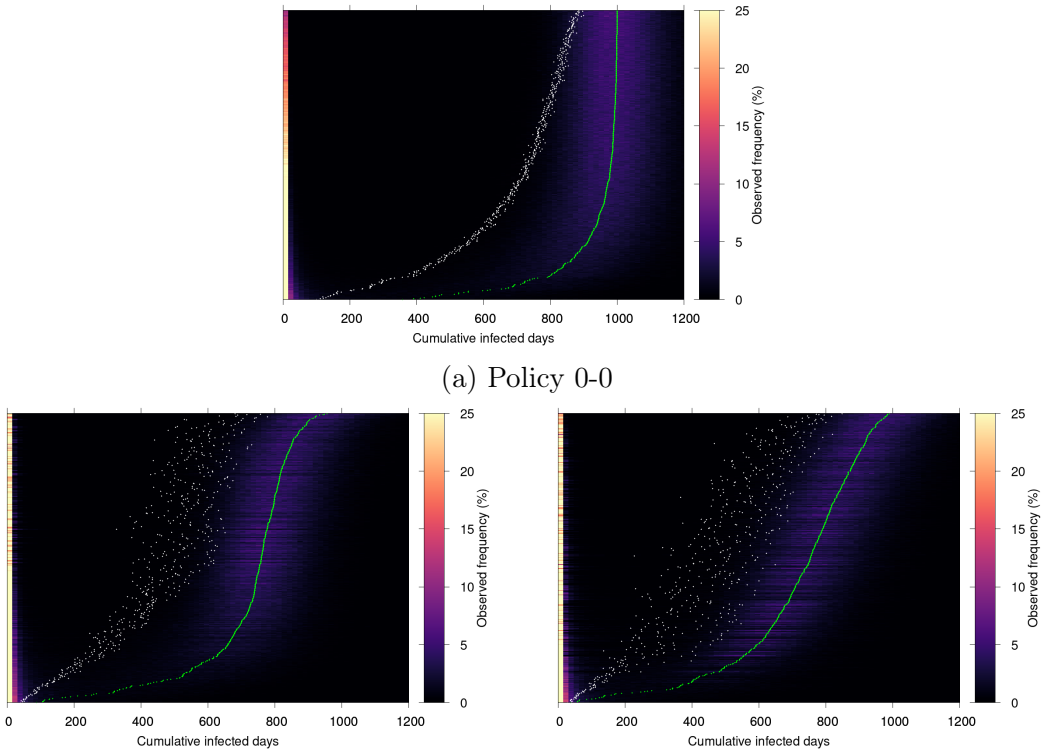
The distribution of $C(a,\xi)$ is bimodal: in a portion of stochastic epidemic trajectories the disease goes extinct before spreading to a large part of the population (clear vertical bands on the left of the heat maps), while in others it causes a major outbreak (purple areas). This is a well-known feature of stochastic disease transmission models [3, 5, 7, 22]. Conditionally on the epidemic taking off, the deterministic model M_0 gives a pretty good approximation ($C_0(a,\xi)$, green) of the of the average stochastic outcome (for each vaccination policy, the green dots are aligned with the purple areas of the heat maps). However the average stochastic outcome ($C_1(a,\xi)$, white) is dragged down by early extinction in some trajectories. By construction, these stochastic effects are not properly taken into account by the deterministic model M_0.

3.2 Decision making under parameter uncertainty

Whether stochastic effects are overlooked or properly taken into account influences decision making. This is shown in Table 2 in the case of decision making under parameter uncertainty. Recall that under parameter uncertainty, policy $a_i^0 = \arg\min_{a\in A} E_\xi\left[C_i(a,\xi)\right]$ is picked based on model M_i. If a decision is made based on deterministic model M_0, policy 50-0 is picked. If a decision is made based on stochastic model M_1, policy 0-50 is picked. Once implemented, the former yields a total of about 453 symptomatic infected-days on average, and the latter about 439 symptomatic infected-days on average.[1]

This difference is small, but it still invalidates the naive reasoning according to which the deterministic model "assumes the worst" by ignoring the possibility of early epidemic

[1]Notice that confidence intervals are displayed for the sake of completeness. However, they do not inform on statistical significance of the policy *differences* per parameter values. Relevant confidence intervals are given in the remainder of the text. We provide estimates of expected outcome differences based on M_0 in Table App-3 in Appendix, although this will not bear on our resoning.



(a) Policy 0-0

(b) Policy 0-50

(c) Policy 50-0

Figure 2: Observed frequency (over 2,048 stochastic simulations, color) of cumulative infected-days (x-axis) for each of the 512 data draws from the base case prior parameter distributions (y-axis). Green dots: outcomes of the deterministic simulations. White dots: averages of the stochastic simulations.

extinction (in the stochastic sense: number of infectious individuals equal to exactly zero), hence leading to more conservative and thus effective policies. The edge of M_1 over M_0 is better understood from Figure 3, which shows the average (over parameter draws ξ) estimate cumulative density function of $C(a, \xi)$. The two inflection points correspond to the two modes of the $C(a, \xi)$ distributions shown in Figure 2. From the zoom in the right panel, we can see that policy 0-50 is better than policy 50-0 at avoiding the epidemic to take off. This can only be shown by modeling stochastic effects properly using stochastic model M_1. Similarly, model M_1 shows that policy 0-50 is better at avoiding the largest outbreaks: the proportion of outbreaks above 900 symptomatic infected-days is lower under policy 0-50 than under policy 50-0. Deterministic model M_0, by contrast, considers a single possible trajectory per set ξ of parameter values, and predicts that on average over possible parameter values policies 0-50 and 50-0 will yield about 718 and 713 symptomatic infected-days respectively (Table 2). A vaccination policy may i) avoid the outbreak of the epidemic,

or ii) limit the number of cases given an outbreak. Using a deterministic model for decision making means dealing with only the second effect.

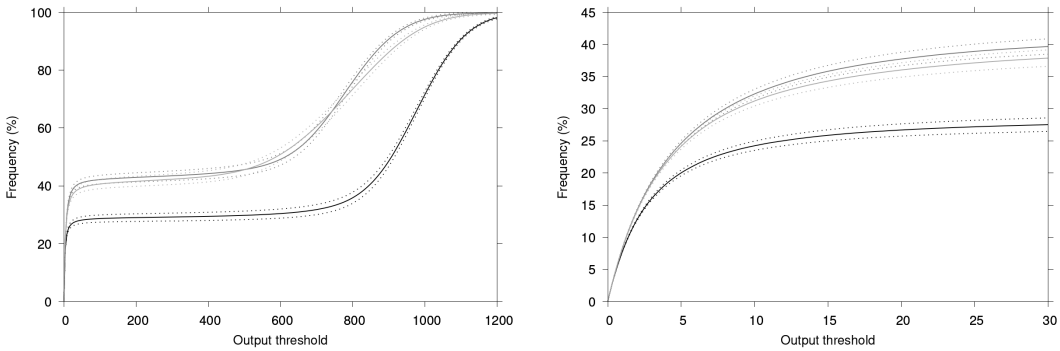
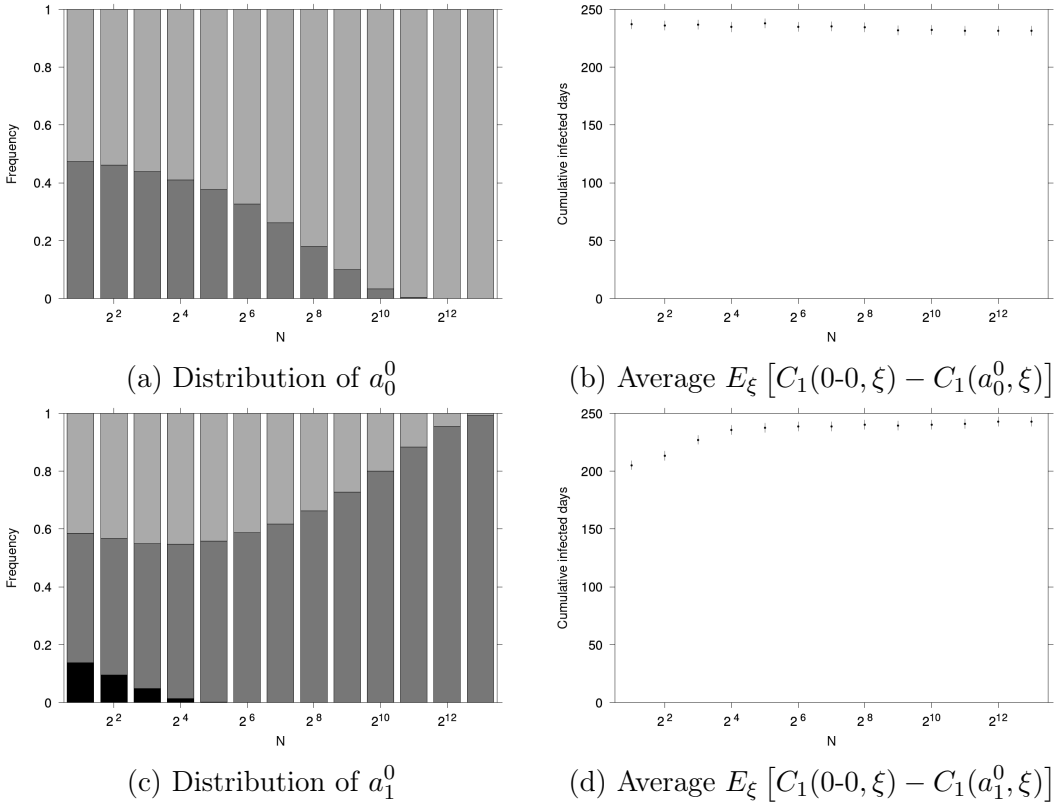


Figure 3: Average (over 512 data draws) frequency (over 2,048 stochastic simulations) of simulations yielding less cumulative symptomatic infected-days than the x-axis threshold value. Right: zoom for small values. Black: policy 0-0 (no vaccination), dark grey: policy 0-50, light grey: policy 50-0. Dotted lines: 95% CI.

So far, we assumed that policies were picked based on 512 parameter draws and, in the stochastic cases, 2,048 runs of stochastic model M_1 for each parameter draw. Obviously, the number of parameter draws and of runs of M_1 influences the quality of decision making. To illustrate this, we simulate decision making based on models M_0 and M_1 under parameter uncertainty and for varying numbers N of model simulations used to estimate $E_\xi\left[C_0(a,\xi)\right]$ and $E_\xi\left[C_1(a,\xi)\right]$ for each policy $a \in A$. When making a decision based on deterministic model M_0, N is the number of parameter draws (the deterministic model is solved once for each draw). When making a decision based on stochastic model M_1, N is the number of parameter draws and trajectory simulations – we bootstrap our $512 \times 2{,}048$ parameter draws and stochastic runs of M_1. Each simulation of decision making is repeated 65,536 times for validation. We look at how often each policy is picked out of the 65,536 decision making simulations, and at the average performance of picked policies. Our bootstrapping procedures are detailed in Appendix D.

The results are shown in Figure 4. When a policy is chosen based on deterministic model M_0 (Figure 4a), larger N means a better sampling of prior parameter distributions allowing for less volatile decision making. As N increases, decision a_0^0 converges towards policy 50-0, which is $\arg\min_{a\in A} E_\xi\left[C_0(a,\xi)\right]$ with a slight advantage over policy 0-50, as suggested by Table 2. By contrast, for $N = 2$, policies 0-50 and 50-0 are picked with almost equal probability, depending on the drawn parameter values. Since policies 0-50 and 50-0 have a close performance once implemented (see $E_\xi\left[C_1(a,\xi)\right]$ in Table 2), the average performance of decisions made based on M_0 does not much depend on N (Figure 4b).

Similarly, Figure 4c shows that decision a_1^0 made based on stochastic model M_1 under parameter uncertainty becomes less volatile as N increases and converges to 0-50, that is $\arg\min_{a \in A} E_\xi \left[C_1(a, \xi)\right]$ (Table 2). However, for small values of N, policy 0-0 (no vaccination) is sometimes picked due to stochastic effects (earlier extinction of the epidemic without vaccination than under policy 0-50 or 50-0 due to chance). Since policy 0-0 is actually suboptimal, the average performance of decisions made based on M_1 is reduced for small values of N (Figure 4d).



(a) Distribution of a_0^0

(b) Average $E_\xi \left[C_1(0\text{-}0, \xi) - C_1(a_0^0, \xi)\right]$

(c) Distribution of a_1^0

(d) Average $E_\xi \left[C_1(0\text{-}0, \xi) - C_1(a_1^0, \xi)\right]$

Figure 4: Decisions made under parameter uncertainty. Top panels: decisions based on M_0. Bottom panels: decisions based on M_1. Left panels: frequency of the policy picked, as a function of N. Black: no vaccination, dark grey: 0-50 policy, light grey: 50-0 policy. Right panels: average decision performance in averted symptomatic infected-days (vertical bars: 95 % CI). Policy distributions and average performances are estimated based on 65,536 decision making experiments.

3.3 Informed decision making

Let us now look into decision making when the true parameter values ξ^* are known. Recall that in this scenario, policy $a_i^* = \arg\min_{a \in A} C_i(a, \xi^*)$ is picked based on model M_i.

Which policy performs better depends on parameter values ξ^* as shown in Figure 5. Policies 0-50 and 50-0 are optimal for about 50 % of parameter values each. Policy 0-0 (no vaccination) is optimal for no parameter values (Figure App-2 in Appendix).

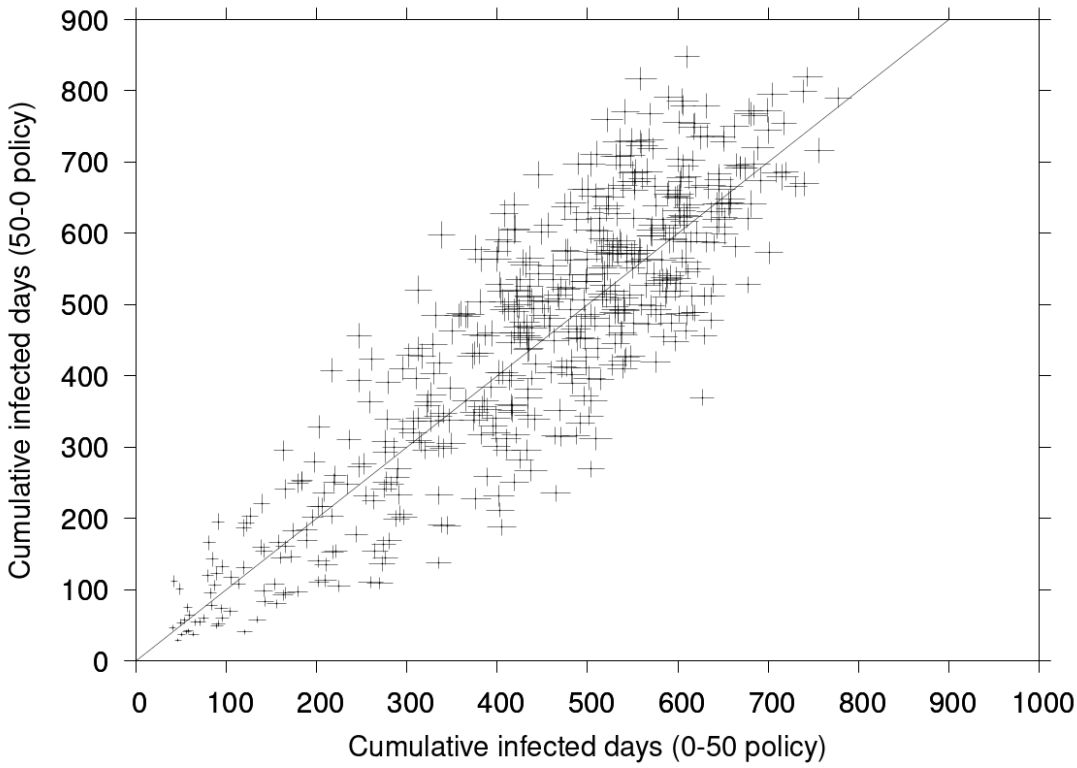
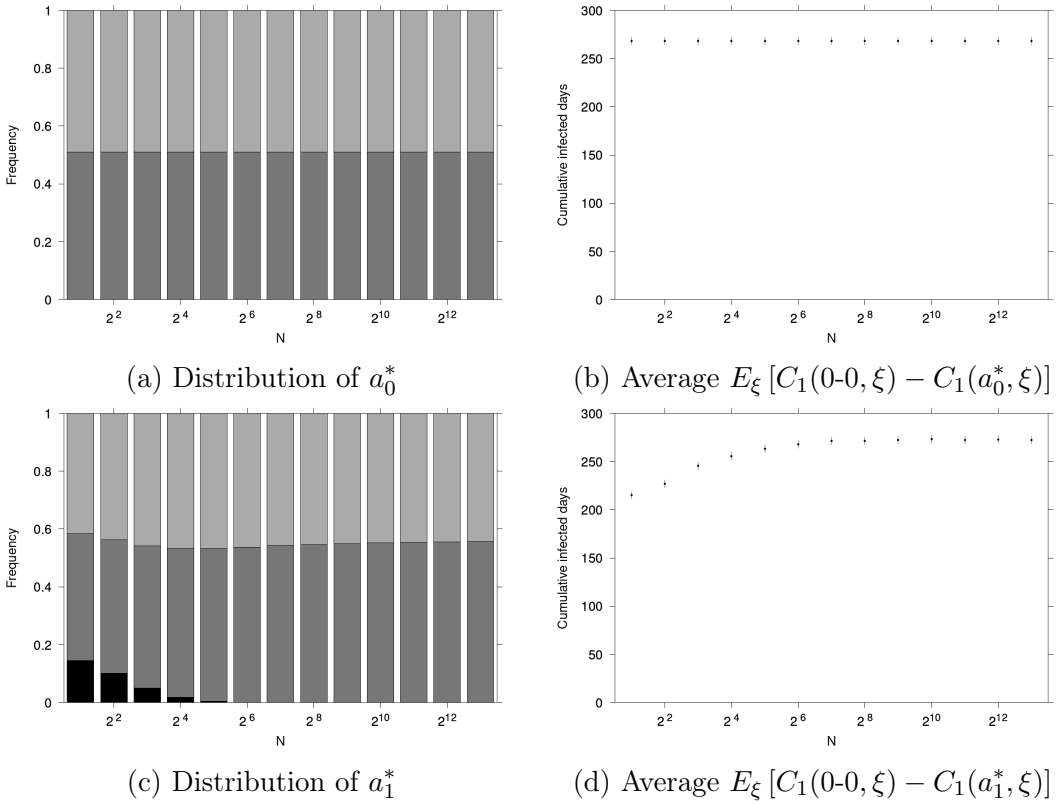


Figure 5: Cumulative symptomatic infected-days $C_1(\text{0-50}, \xi^*)$ (x-axis) and $C_1(\text{50-0}, \xi^*)$ (y-axis) for 512 data draws ξ^* from the base case parameter prior distributions. Crosses: 95% CI across 2,048 stochastic simulations of M_1.

Figure 6 shows the influence of the number of model simulations N on decision making when the true parameter values ξ^* are known. When decisions are made based on deterministic model M_0, decisions do not depend on the number of parameter runs (Figures 6a and 6b). Policies 0-50 and 50-0 are picked for about 50 % of parameter draws each (Figure 6a). When decisions are made based on stochastic model M_1, we can observe the same stochastic effects as under uncertainty (Figure 4c): for small numbers N of model runs, suboptimal policy 0-0 is picked with non-zero probability (Figure 6c), which reduces the average performance of decision making (Figure 6d). As N increases, decisions converge towards 0-50 and 50-0 in almost equal proportion, consistently with Figure 5.



(a) Distribution of a_0^* (b) Average $E_\xi\left[C_1(0\text{-}0,\xi) - C_1(a_0^*,\xi)\right]$

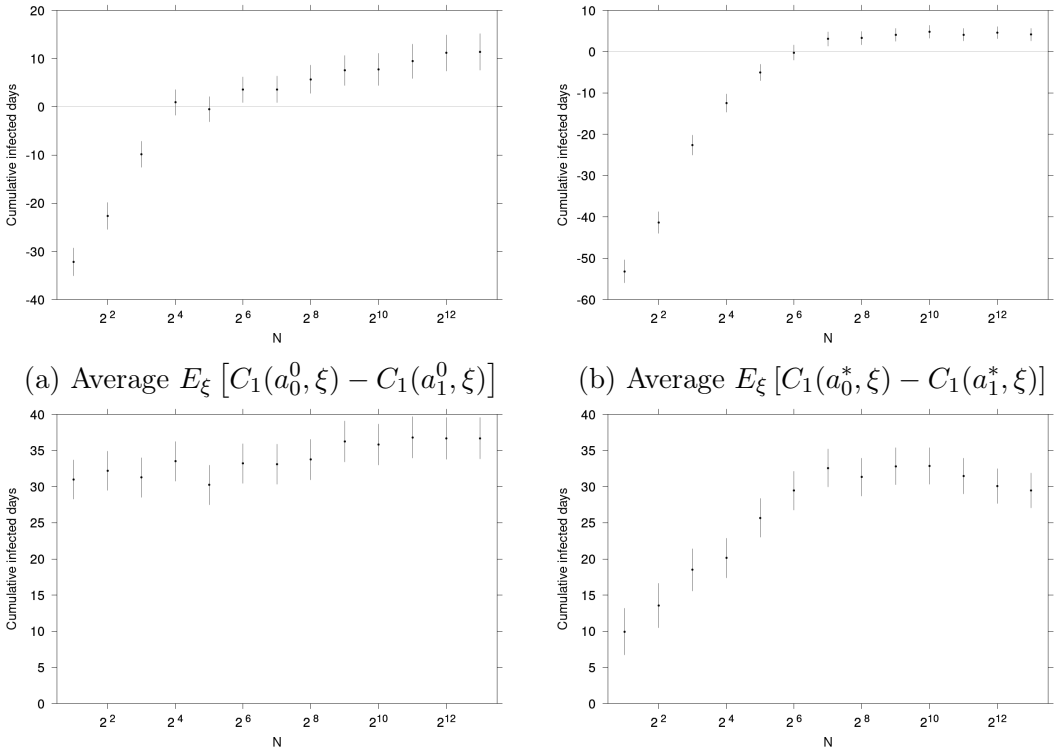
(c) Distribution of a_1^* (d) Average $E_\xi\left[C_1(0\text{-}0,\xi) - C_1(a_1^*,\xi)\right]$

Figure 6: Decisions made knowing true parameter values. Top panels: decisions based on M_0. Bottom panels: decisions based on M_1. Left panels: frequency of the policy picked, as a function of N. Black: no vaccination, dark grey: 0-50 policy, light grey: 50-0 policy. Right panels: average decision performance in averted symptomatic infected-days (vertical bars: 95 % CI). Policy distributions and average performances are estimated based on 65,536 decision making experiments.

3.4 Cost of parameter uncertainty v. of using the wrong model

We showed that using a stochastic model where appropriate rather than its deterministic counterpart does influence the performance of decision making. We also showed that the number N of model runs can influence decision making as well, in particular when using a stochastic model. These effects hold under parameter uncertainty and when parameter values are known. Hence the question: how does the cost of using a deterministic model compares to that of deciding under uncertainty? And how does the number of model runs come into play in this respect?

Figure 7a shows the difference in average performance between decisions a_1^0 and a_0^0 made under uncertainty based on stochastic model M_1 and on deterministic model M_0 respectively. It corresponds to the difference between Figures 4b and 4d. For small values of



(a) Average $E_\xi\left[C_1(a_0^0,\xi) - C_1(a_1^0,\xi)\right]$

(b) Average $E_\xi\left[C_1(a_0^*,\xi) - C_1(a_1^*,\xi)\right]$

(c) Average $E_\xi\left[C_1(a_0^0,\xi) - C_1(a_0^*,\xi)\right]$

(d) Average $E_\xi\left[C_1(a_1^0,\xi) - C_1(a_1^*,\xi)\right]$

Figure 7: Differences in average decision making performance depending on the model used (M_0 or M_1) as functions of N and on the level of parameter information. Vertical bars: 95% CI.

N, deterministic model M_0 outperforms stochastic model M_1 on average due to stochastic effects. As N increases, the difference converges to about 12 less cumulative symptomatic infected-days on average when decision are made with M_1 rather than M_0. In the context of decision making under uncertainty, N can be understood as a computational budget. Given this budget, it is possible i) to better sample the parameter prior distribution and use deterministic model M_0 (one run per parameter draw), or ii) to use stochastic model M_1 which is a more appropriate representation of reality but requires several runs for each parameter draw. This interpretation is of course limited since the computational resources actually needed for a single model run may vary greatly depending on the numerical methods or, in the case of the stochastic model, on the number of random events until simulation end (in our case, epidemic extinction). Still, we show with an example that an appropriate stochastic model can be outperformed by its deterministic version if the number of model runs is too small. We leave accurate estimations of required computational resources outside the scope of the present article since this is largely problem specific. We also do not

consider the technical issue of balancing parameter draws and model runs per draw when making decisions under uncertainty using a stochastic model. Notice finally that in our example, the threshold value of N under which M_0 outperforms M_1 is rather small (about 16 or 32 runs) considering the simplicity of our transmission model and optimization problem. This is clearly problem dependent. In our robustness checks (Appendix F), we easily found instances in which stochastic effects still had an influence on the average performance of a_1^0 for $N = 2^{11} = 2{,}048$, and up to $N = 2^{13} = 8{,}192$ model runs. Besides, the maximum manageable number of model runs depends not only on the problem but also on available computing resources.

Figure 7b shows the difference in average performance between decisions a_1^* and a_0^* made knowing true parameter values based on stochastic model M_1 and on deterministic model M_0 respectively. It corresponds to the difference between Figures 6b and 6d. We can observe the same influence of stochastic effects for small values of N as in the scenario of decision making under uncertainty. Notice that given N, the benefit of using stochastic model M_1 rather than deterministic model M_0 is smaller in the informed decision making scenario than under the uninformed scenario (Figure 7a).

Figure 7c shows the additional benefit of resolving parameter uncertainties prior to making a decision, given that deterministic model M_0 is used for decision making (comparison of a_0^0 and a_0^*). It corresponds to the difference between Figures 4b and 6b. In our example, policies 0-50 and 50-0 are picked with about 50 % chance for $N = 2$ both under parameter uncertainty and knowing parameter values (compare Figures 4a and 6a). Under parameter uncertainty, this is due to the poor sampling of the parameter prior distribution used to estimate $E_\xi\left[C_0(a,\xi)\right]$ for each policy a. For $N = 2$, decision making is essentially equivalent to picking a policy at random (among the non-dominated policies). In the informed case, $a_0^* = \arg\min_{a\in A} C_0(a,\xi^*)$ is policy 0-50 for about 50 % of the possible parameter values ξ^*, and policy 50-0 for the remaining parameter values. Policy a_0^* performs better than a_0^0 on average for $N = 2$ even though the distribution between policies 0-50 and 50-0 in each case are almost identical. The reason is that although deterministic model M_0 is unable to capture some features of the epidemic, it still carries some information and allows to make better decisions knowing parameter values than those made under uncertainty.

The quantity $E_\xi\left[C_1(a_1^0,\xi) - C_1(a_1^*,\xi)\right] = \min_{a\in A} E_\xi\left[C_1(a,\xi)\right] - E_\xi\left[\min_{a\in A} C_1(a,\xi)\right]$ shown in Figure 7d is the expected value of perfect information (EVPI), a popular value-of-information metric which quantifies the additional benefit of entirely resolving uncertainties prior to making a decision [23]. It corresponds to the difference between Figures 4d and 6d. EVPI is positive or null by definition. It is subject to stochastic effects for small values of N.

We already showed that decisions made using the appropriate model, stochastic model M_1, and in particular decisions a_1^* made based on M_1 *and* knowing parameter values, can be outperformed due to stochastic effects for small values of N. Actually for $N = 2$, the best performance is achieved by policy a_0^*. The performance of a_0^0 is decreased by using the "right" model (Figure 7a) but it can be improved by resolving uncertainties (Figure 7c). For large values of N however, the best performance is achieved by a_1^*. But still, starting from a_0^0, switching to model M_1 and resolving uncertainties do not make the same contribution to improving performance: for $N = 2^{13}$, switching to M_1 allows to avert about 12 symptomatic infected-days (Figure 7a) against about 37 symptomatic infected-days averted by resolving uncertainties (Figure 7c).[2]

Results shown in Figure 7 are based on the baseline parameter prior distributions introduced in Section 2.2. As a robustness check we performed these analyses for randomly drawn prior distributions (the distribution parameters are drawn at random) and obtained qualitatively similar results, see Appendix F.

4 Conclusion

The respective properties of deterministic and stochastic infectious disease transmission models have been largely studied and are now introduced in most epidemiological modeling textbooks. By assuming discrete populations, stochastic models allow to account for effects that deterministic models cannot explain, typically (but not only) where epidemic extinction is involved. Yet the implications for decision making of the use of a deterministic or stochastic model have been somewhat overlooked in the literature. Some authors, it is true, compared the performance of vaccination policies picked based on a stochastic model with those picked based on its deterministic counterpart. However they did so mostly in order to provide a counterfactual to their stochastic optimization methods. They took the perspective of omniscient researchers investigating "from above" the relative performance of policies based on the stochastic and deterministic models in different areas of the parameter space. By contrast, we focused on the perspective of a decision maker. In particular we considered scenarios in which model parameters were uncertain, as is most often the case in real life.

Thus, we considered a total of four decision making scenarios: based on a stochastic model or its deterministic counterpart, and under uncertainty or knowing true parameter values. We looked into these scenarios using two minimal working examples of transmission models, a deterministic and a stochastic SIR models of an emerging disease, and a simple

[2]For this value of N, all differences have converged to their large sample size values.

optimization problem consisting in picking a vaccination policies out of three alternative options (including a dominated no vaccination policy). The objective was to minimize the cumulative number of symptomatic infected-days over the course of the epidemic. In a validation step, we assessed the performance of the chosen policies by implementing them in the stochastic framework, which we assumed to be an accurate model of reality.

In our example, one of the non dominated policies was better at "nipping the epidemic in the bud" while the other was better at reducing the number of symptomatic infected-days in a large scale outbreak. Each policy was optimal for about 50 % of the possible parameter values. Under parameter uncertainty, a policy is picked that minimizes the expected number of cumulative symptomatic infected-days. We showed that in our case, different policies minimize the expected burden estimated by the deterministic model and the expected burden estimated by the stochastic model. This is because whether under parameter uncertainty or knowing parameter values, the deterministic model does not properly take the possibility of early epidemic extinction into account.

The performance of decision making largely depends on the quality of the expected performance estimates used to pick a policy. We then performed our analyses for different Monte Carlo sample sizes: this covers the size of parameter samples in the uniformed case, and the number of stochastic model runs per set of parameter values where the stochastic model is used. We showed that due to stochastic effects, the deterministic model can outperform the stochastic model for small sample sizes, both under parameter uncertainty or assuming informed decision making.

Our study comes with the usual caveats and provisions. Our aim was to provide an understanding of the interplay between the choice of a stochastic or deterministic model, parameter uncertainties, and Monte Carlo sample sizes from the perspective of decision making. Specific results presented here obviously depend on the transmission model structure or the parameter prior distributions. For instance our results suggest that resolving uncertainties brings more benefit than using the "right" (stochastic) model. We do not expect this to hold in all real life cases. We estimated the value of resolving uncertainties by assuming that true parameter values can actually be known. The value of information literature provides alternative approaches based on intermediate levels of information that may be more relevant in practice. As for our discussion of Monte Carlo sample sizes, it would ultimately need to consider specific numerical methods and available computational resources. We also left some technical issues such as finding the optimal balance between stochastic model runs and parameter sampling outside the scope of the study.

However, general results will remain true: the quality of decision making decreases with the sample size, stochastic models can be outperformed due to stochastic effects for

small sample sizes, and this is true both under uncertainty and knowing parameter values. More importantly, the reasoning presented in the article will be relevant in any situation where stochastic effects are critical. The extra burden of sampling stochastic epidemic trajectories will have to be balanced against available resources and the cost of sampling the parameter space – this could be critical when solving a difficult optimization problem e.g. with heuristics requiring some fine tuning. Value-of-information metrics provide a useful benchmark when measuring the benefit of increasing the sample size or using a different model for decision making. In this, our study points to issues to be carefully considered when embarking on optimization of stochastic transmission models.

Declarations of interest

None.

References

[1] M. S. Bartlett. The relevance of stochastic models for large-scale epidemiological phenomena. Applied Statistics, 13(1):2, 1964. doi: 10.2307/2985217.

[2] I. Nåsell. On the time to extinction in recurrent epidemics. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(2):309–330, apr 1999. doi: 10.1111/1467-9868.00178.

[3] Tom Britton. Stochastic epidemic models: A survey. Mathematical Biosciences, 225 (1):24–35, may 2010. doi: 10.1016/j.mbs.2010.01.006.

[4] Fred Brauer and Carlos Castillo-Chavez. Epidemic models. In Mathematical Models in Population Biologyand Epidemiology, pages 345–409. Springer New York, nov 2011. doi: 10.1007/978-1-4614-1686-9_9.

[5] Frank Ball. Epidemic thresholds, jul 2005.

[6] Nicolas Houy. The probability of undetected wild poliovirus circulation: Can we do better? Journal of Theoretical Biology, 382:272–278, oct 2015. doi: 10.1016/j.jtbi.2015.06.046.

[7] Francesca A. Lovell-Read, Sebastian Funk, Uri Obolski, Christl A. Donnelly, and Robin N. Thompson. Interventions targeting non-symptomatic cases can be important to prevent local outbreaks: SARS-CoV-2 as a case study. Journal of The Royal Society Interface, 18(178):20201014, may 2021. doi: 10.1098/rsif.2020.1014.

[8] R. N. Thompson, C. A. Gilligan, and N. J. Cunniffe. Will an outbreak exceed available resources for control? estimating the risk from invading pathogens using practical definitions of a severe epidemic. Journal of The Royal Society Interface, 17(172): 20200690, nov 2020. doi: 10.1098/rsif.2020.0690.

[9] Ingemar Nåsell. Measles outbreaks are not chaotic. In Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory, pages 85–114. Springer New York, 2002. doi: 10.1007/978-1-4613-0065-6_6.

[10] Ingemar Nåsell. Stochastic models of some endemic infections. Mathematical Biosciences, 179(1):1–19, jul 2002. doi: 10.1016/s0025-5564(02)00098-6.

[11] L.J.S. Allen and P. van den Driessche. Relations between deterministic and stochastic thresholds for disease extinction in continuous- and discrete-time infectious disease

models. Mathematical Biosciences, 243(1):99–108, may 2013. doi: 10.1016/j.mbs.2013.02.006.

[12] Clara Champagne and Bernard Cazelles. Comparison of stochastic and deterministic frameworks in dengue modelling. Mathematical Biosciences, 310:1–12, apr 2019. doi: 10.1016/j.mbs.2019.01.010.

[13] Milliward Maliyoni. Probability of disease extinction or outbreak in a stochastic epidemic model for west nile virus dynamics in birds. Acta Biotheoretica, 69(2):91–116, sep 2020. doi: 10.1007/s10441-020-09391-y.

[14] Edwin C. Yuan, David L. Alderson, Sean Stromberg, and Jean M. Carlson. Optimal vaccination in a stochastic epidemic model of two non-interacting populations. PLOS ONE, 10(2):e0115826, feb 2015. doi: 10.1371/journal.pone.0115826.

[15] Garrett Jenkinson and John Goutsias. Numerical integration of the master equation in some models of stochastic epidemiology. PLoS ONE, 7(5):e36160, may 2012. doi: 10.1371/journal.pone.0036160.

[16] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4): 403–434, dec 1976. doi: 10.1016/0021-9991(76)90041-3.

[17] Chantal Nguyen and Jean M. Carlson. Optimizing real-time vaccine allocation in a stochastic SIR model. PLOS ONE, 11(4):e0152950, apr 2016. doi: 10.1371/journal.pone.0152950.

[18] Nicolas Houy and Julien Flaig. Hospital-wide surveillance-based antimicrobial treatments: A monte-carlo look-ahead method. Computer Methods and Programs in Biomedicine, 204:106050, jun 2021. doi: 10.1016/j.cmpb.2021.106050.

[19] Nicolas Houy and Julien Flaig. Optimal dynamic empirical therapy in a health care facility: A monte-carlo look-ahead method. Computer Methods and Programs in Biomedicine, 198:105767, jan 2021. doi: 10.1016/j.cmpb.2020.105767.

[20] Lawrence Thul and Warren Powell. Stochastic optimization for vaccine and testing kit allocation for the COVID-19 pandemic. European Journal of Operational Research, 304(1):325–338, jan 2023. doi: 10.1016/j.ejor.2021.11.007.

[21] Peter Grandits, Raimund M. Kovacevic, and Vladimir M. Veliov. Optimal control and the value of information for a stochastic epidemiological SIS-model. Journal of

Mathematical Analysis and Applications, 476(2):665–695, aug 2019. doi: 10.1016/j.jmaa.2019.04.005.

[22] Luis F. Gordillo, Stephen A. Marion, Anders Martin-Löf, and Priscilla E. Greenwood. Bimodal epidemic size distributions for near-critical SIR with vaccination. Bulletin of Mathematical Biology, 70(2):589–602, nov 2007. doi: 10.1007/s11538-007-9269-y.

[23] Christopher H. Jackson, Gianluca Baio, Anna Heath, Mark Strong, Nicky J. Welton, and Edward C.F. Wilson. Value of information analysis in models to inform health policy. Annual Review of Statistics and Its Application, 9(1):95–118, mar 2022. doi: 10.1146/annurev-statistics-040120-010730.

Appendix

A Compartmental models

Let S_i and I_i denote the number of susceptible and infectious individuals who received $i \in \{0, 1, 2\}$ doses of vaccine, and R the number of recovered individuals. Individuals get infected with contact rate β and they recover at rate γ. The basic reproduction number is $R_0 = \beta/\gamma$. Individuals who received $i \in \{1, 2\}$ vaccine doses are less susceptible by a factor σ_i and less infectious by a factor ρ_i compared to unvaccinated individuals. We assume that vaccination is instantaneous and that there is no waning of vaccine protection. In the deterministic model M_0, the epidemic dynamics is described by Equations (1)–(5). The corresponding stochastic model M_1 is summarized by events shown in Table App-1. Figure App-1 is a sketch of the transmission models.

$$\frac{dS_0}{dt} = -\lambda S_0 \tag{1}$$

$$\frac{dI_0}{dt} = \lambda S_0 - \gamma I_0 \tag{2}$$

$$\frac{dS_i}{dt} = -\sigma_i \lambda S_i \tag{3}$$

$$\frac{dI_i}{dt} = \sigma_i \lambda S_i - \gamma I_i \tag{4}$$

$$\frac{dR}{dt} = \gamma \sum_{j=0}^{2} I_j \tag{5}$$

where

$$\lambda = \beta(I_0 + \rho_1 I_1 + \rho_2 I_2)/n. \tag{6}$$

Event	**Population changes**	**Rate**
Infection of unvaccinated	$S_0 \leftarrow S_0 - 1,\ I_0 \leftarrow I_0 + 1$	λS_0
Infection of vaccinated	$S_i \leftarrow S_i - 1,\ I_i \leftarrow I_i + 1$	$\sigma_i \lambda S_i$
Recovery	$I_i \leftarrow I_i - 1,\ R_i \leftarrow R_i + 1$	γI_i

Table App-1: Events included in stochastic model M_1.

Let p_i the probability of symptomatic infection for infected individuals who received i doses of vaccine. The incidence rates of symptomatic and asymptomatic cases among unvaccinated individuals are $\mathcal{I}_0^{\mathrm{s}} = p_0 \lambda S_0$ and $\mathcal{I}_0^{\mathrm{as}} = (1 - p_0)\lambda S_0$ respectively. Among

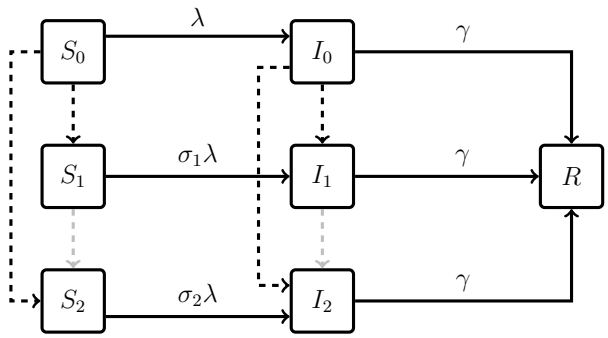


Figure App-1: Sketch of the transmission models. Dashed: possible vaccination. Grey dashed: vaccination not considered in the present article.

individuals who received $i \in \{1, 2\}$ doses of vaccine, incidence rates are $\mathcal{I}_i^{\mathrm{s}} = p_i \sigma_i \lambda S_i$ and $\mathcal{I}_i^{\mathrm{as}} = (1 - p_i)\sigma_i \lambda S_i$.

B Parameter prior distributions

For convenience, we use truncated normal distributions as parameter prior distributions. Table App-2 shows the mean and standard deviation of the underlying normal distributions, and the supports of the prior distributions. Once a set of parameter is drawn, we check that $p_1 > p_2$, $\rho_1 > \rho_2$ and $\sigma_1 > \sigma_2$. Hence, the reader should be aware that parameter values are not independent.

Parameter	**Mean**	**Std. dev.**	**Support**
R_0	4.0	4.0	$[1.2, \infty[$
γ	0.1	0.0	–
σ_1	0.8	0.3	$[0.02, 1.0]$
ρ_1	0.8	0.3	$[0.02, 1.0]$
p_1/p_0	0.8	0.3	$[0.02, 1.0]$
σ_2	0.5	0.3	$[0.02, 1.0]$
ρ_2	0.5	0.3	$[0.02, 1.0]$
p_2/p_0	0.5	0.3	$[0.02, 1.0]$

Table App-2: Base case parameter prior distributions.

C Stochastic effects: additional table

Metric	Estimate
$E_\xi\left[C_0(\text{0-0},\xi) - C_0(\text{0-50},\xi)\right]$	215.064 (95% CI: 214.947 – 215.18)
$E_\xi\left[C_0(\text{0-0},\xi) - C_0(\text{50-0},\xi)\right]$	215.359 (95% CI: 215.122 – 215.597)
$E_\xi\left[C_0(\text{0-50},\xi) - C_0(\text{50-0},\xi)\right]$	0.295686 (95% CI: 0.100199 – 0.491173)

Table App-3: Estimate differences in expected outcome between the three available policy options based on M_0. Averages over 2^{20} parameter draws (one run of M_0 per parameter draw).

D Bootstrapping procedures

We bootstrap our simulations (512 parameter draws and 2,048 stochastic simulations per available policy for each data draw) in order to obtain 65,536 total simulations of decision making.

For decision making based on deterministic model M_0 under parameter uncertainty, the bootstrapping procedure is as follow:

- 65,536 times do:
  - draw parameter values ξ^*,
  - N times do:
    - draw parameter values ξ,
    - for each policy $a \in A$, compute $C_0(a, \xi)$,
  - for each policy $a \in A$, estimate $E_\xi \left[C_0(a, \xi)\right]$ based on the N parameter draws ξ,
  - pick $a_0^0 = \arg\min_{a \in A} E_\xi \left[C_0(a, \xi)\right]$ based on these estimates,
  - validation: compute $C(a_0^0, \xi^*)$ (one run of stochastic model M_1),
- estimate $E_\xi \left[C_1(a_0^0, \xi)\right] = E_\xi \left[E \left[C(a_0^0, \xi)\right]\right]$ based on the 65,536 simulations of decision making.

For decision making based on stochastic model M_1 under parameter uncertainty, the bootstrapping procedure is as follow:

- 65,536 times do:
  - draw parameter values ξ^*,
  - N times do:
    - draw parameter values ξ,
    - for each policy $a \in A$, compute $C(a, \xi)$ (one run of M_1 for each policy),
  - for each policy $a \in A$, estimate $E_\xi \left[C_1(a, \xi)\right] = E_\xi \left[E \left[C(a, \xi)\right]\right]$ based on the N runs of M_1 and parameter draws ξ,
  - pick $a_1^0 = \arg\min_{a \in A} E_\xi \left[C_1(a, \xi)\right]$ based on these estimates,
  - validation: compute $C(a_1^0, \xi^*)$ (one run of stochastic model M_1),
- estimate $E_\xi \left[C_1(a_1^0, \xi)\right] = E_\xi \left[E \left[C(a_1^0, \xi)\right]\right]$ based on the 65,536 simulations of decision making.

For decision making based on deterministic model M_0 knowing parameter values, the bootstrapping procedure is as follow:

- 65,536 times do:
  - draw parameter values ξ^*,
  - for each policy $a \in A$, compute $C_0(a, \xi^*)$,
  - pick $a_0^* = \arg\min_{a \in A} C_0(a, \xi^*)$,
  - validation: compute $C(a_0^*, \xi^*)$ (one run of stochastic model M_1),
- estimate $E_\xi \left[C_1(a_0^*, \xi)\right] = E_\xi \left[E \left[C(a_0^*, \xi)\right]\right]$ based on the 65,536 simulations of decision making.

For decision making based on stochastic model M_1 knowing parameter values, the bootstrapping procedure is as follow:

- 65,536 times do:
  - draw parameter values ξ^*,
  - N times do:
    - for each policy $a \in A$, compute $C(a, \xi^*)$ (one run of M_1 for each policy),
  - for each policy $a \in A$, estimate $C_1(a, \xi^*) = E \left[C(a, \xi^*)\right]$ based on the N runs of M_1,
  - pick $a_1^* = \arg\min_{a \in A} C_1(a, \xi^*)$ based on these estimates,
  - validation: compute $C(a_1^*, \xi^*)$ (one run of stochastic model M_1),
- estimate $E_\xi \left[C_1(a_1^*, \xi)\right] = E_\xi \left[E \left[C(a_1^*, \xi)\right]\right]$ based on the 65,536 simulations of decision making.

E Informed decision making: additional figures

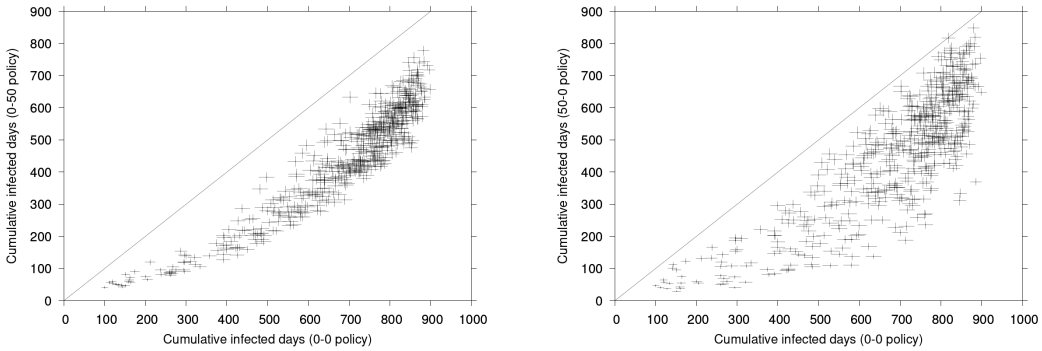


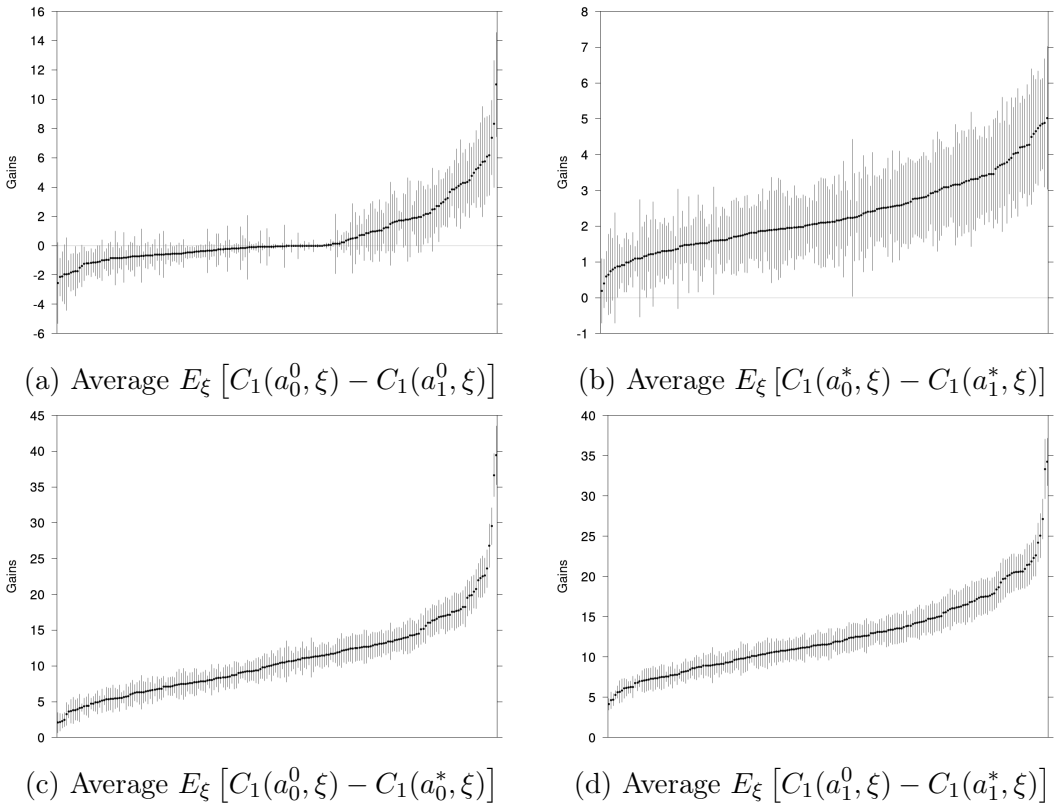
Figure App-2: Cumulative symptomatic infected-days $C_1(0\text{-}0, \xi^*)$ (x-axis) and $C_1(0\text{-}50, \xi^*)$ (y-axis, left panel) or $C_1(50\text{-}0, \xi^*)$ (y-axis, right panel) for 512 data draws ξ^* from the base case parameter prior distributions. Cross heights and widths represent 95% CI across 2,048 stochastic simulations of M_1.

F Robustness checks

In Figure App-3, we show results similar to those presented in Figure 7 for different priors drawn by chance as displayed in Table App-4.

Parameter	**Mean**	**Std. dev.**
R_0	$U(2, 10)$	$U(1, 5)$
γ	$U(0.1, 0.5)$	$U(0, 0.2)$
$\sigma_1 = \rho_1 = p_1/p_0$	$U(0.5, 0.9)$	$U(0.1, 0.4)$
$\sigma_2 = \rho_2 = p_2/p_0$	$U(0.1, \mu_{\sigma_1})$	$U(0.1, 0.4)$

Table App-4: Distributions of the parameters of parameter prior distributions. $U(a, b)$: uniform distribution with support $[a, b]$. μ_{σ_1}: mean parameter (mean of the underlying normal distribution) of the prior distribution of σ_1.



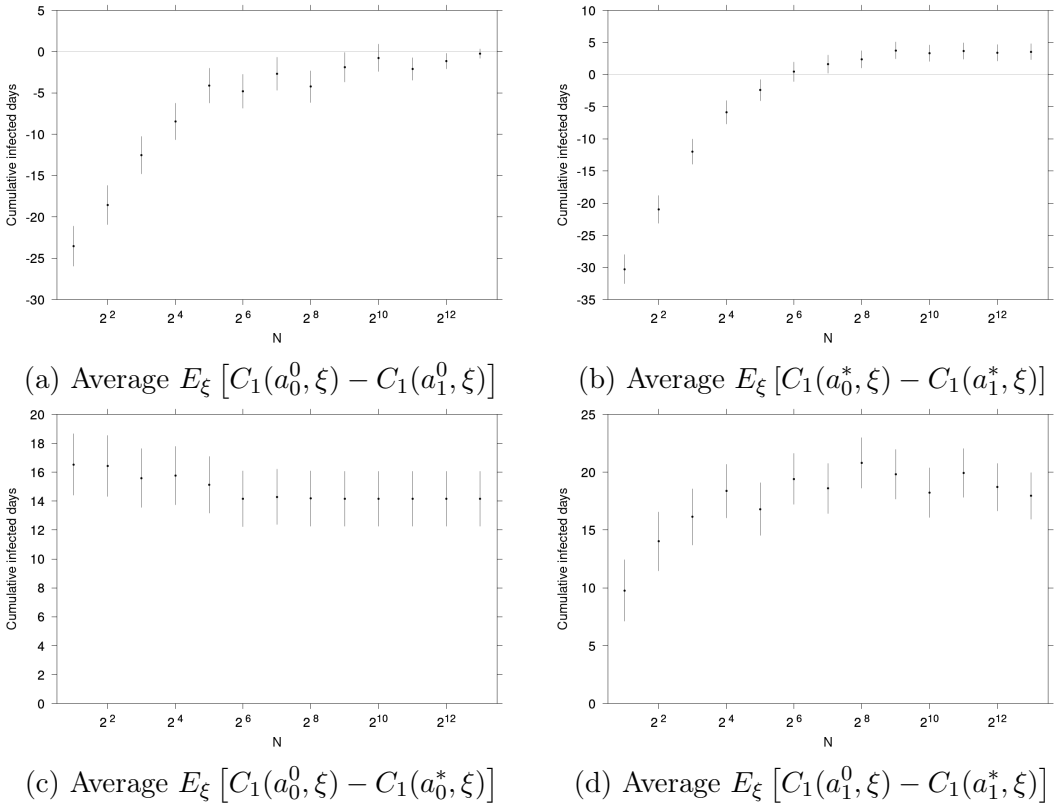
(a) Average $E_\xi \left[C_1(a_0^0, \xi) - C_1(a_1^0, \xi)\right]$ (b) Average $E_\xi \left[C_1(a_0^*, \xi) - C_1(a_1^*, \xi)\right]$

(c) Average $E_\xi \left[C_1(a_0^0, \xi) - C_1(a_0^*, \xi)\right]$ (d) Average $E_\xi \left[C_1(a_1^0, \xi) - C_1(a_1^*, \xi)\right]$

Figure App-3: Differences in cumulative symptomatic infected-days between different cases for prior distributions drawn by chance as described in Table App-4. Results are ordered by increasing value. Vertical bars: 95% CI. $N = 2,048$.

Remark: the fact that many values of the average performance difference $E_\xi \left[C_1(a_0^0, \xi) - C_1(a_1^0, \xi)\right]$ between M_0 and M_1 under uncertainty (Figure App-4a) are negative for $N = 2,048 = 2^{11}$

seems to be due to N being too small. In figure App-4, we display the same as Figure 7 for the set of values with the second lowest average value in Figure App-4a (the smallest is not statistically significant). Figure App-5 shows the corresponding decision distributions. We see from Figure App-5c that stochastic effects still play a role for $N = 2^1 3 = 8{,}192$ as the distribution of a_1^0 has not fully converged.



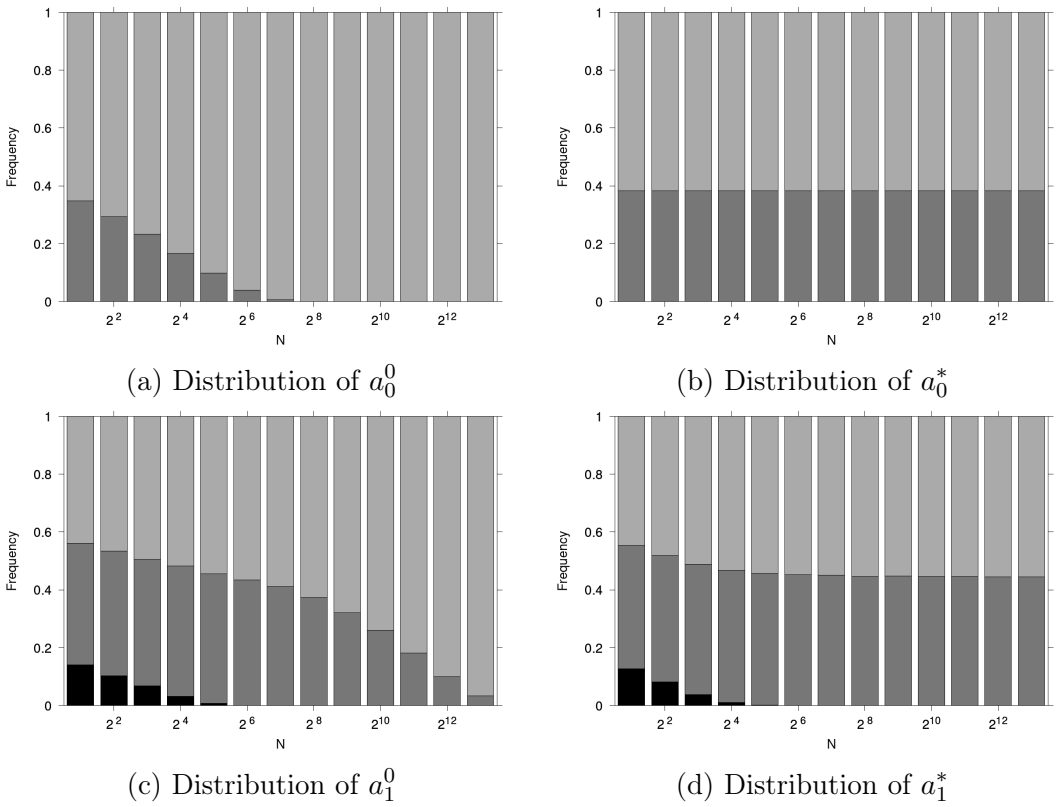
(a) Average $E_\xi\left[C_1(a_0^0,\xi) - C_1(a_1^0,\xi)\right]$

(b) Average $E_\xi\left[C_1(a_0^*,\xi) - C_1(a_1^*,\xi)\right]$

(c) Average $E_\xi\left[C_1(a_0^0,\xi) - C_1(a_0^*,\xi)\right]$

(d) Average $E_\xi\left[C_1(a_1^0,\xi) - C_1(a_1^*,\xi)\right]$

Figure App-4: Difference between numbers of cumulative infected days avoided in the different cases and as functions of N. Vertical bars: 95% CI.



(a) Distribution of a_0^0

(b) Distribution of a_0^*

(c) Distribution of a_1^0

(d) Distribution of a_1^*

Figure App-5: Frequency of the policy implemented as a function of N. Black: no vaccination, dark grey: 0-50 policy, light grey: 50-0 policy.