

Robust Epidemiological Prediction and Optimization

Chenyi Fu

Institute of Systems Engineering, College of Management and Economics, Tianjin University, China
chenyi_fu_201509@163.com

Melvyn Sim

Department of Analytics & Operations (DAO), NUS Business School, National University of Singapore
melvynsim@gmail.com

Minglong Zhou

Institute of Operations Research and Analytics, National University of Singapore
oramz@nus.edu.sg

The COVID-19 pandemic has brought many countries to their knees, and the urgency to return to normalcy has never been greater. Epidemiological models, such as the SEIR compartmental model, are indispensable tools for, among other things, predicting how pandemic may spread over time and how vaccinations and different public health interventions could affect the outcome. However, deterministic epidemiological models do not reflect the stochastic nature of the actual infected populations for which the true distribution can never be determined precisely. When embedded in an optimization model, the impact of ambiguous risk can influence the desired outcomes of the mitigating strategy. To address these issues, we first propose a robust epidemiological model, which provides prediction intervals that is specified by the Aumann and Serano (2008) *riskiness index*. With suitable approximations, the robust epidemiological optimization model that minimizes the riskiness index can be formulated as a mixed integer linear optimization problem. We illustrate how we can apply the robust epidemiological optimization model for strategic vaccine allocation by minimizing the model's riskiness indices for all the constraints on limiting infections across all time periods, and within a given budget for vaccinations. We conduct a simulation study using parameters estimated from open-source datasets on the COVID-19 pandemic. Simulation results illustrate that our robust vaccine allocation model yields solutions that outperform the benchmark models in controlling the spread of infections.

Key words: COVID-19, epidemiological model, riskiness index, robust optimization, vaccine allocation

History: June 18, 2021.

1. Introduction

The COVID-19 pandemic has resulted in an unprecedented volume of infections and mass mortality within a short span of time. Countries urge to mitigate the propagation of this global pandemic and they have employed various strict non-pharmacological intervention policies (NPIs) to control the impacts of COVID-19 and save numerous lives. However, it has been challenging to control

the propagation of this pandemic while maintaining an active economy. Global manufacturing production, aviation industry, labor market, and education face the deepest crisis in 2020 (ESCAP et al. 2020).

The pandemic has once again raised our attention to how we can better control pandemic propagation through mass vaccinations and public health interventions such as, *inter alia*, isolation and treatment of infected patients, contact tracing and dedicate quarantine facilities for potentially infected population. As we have already witnessed in some countries, inadequate public health interventions can lead to severe shortages of critical healthcare supplies and result in dire escalation of morbidity and mortality. However, building a prescriptive analytics model for this purpose is challenging due to risk and ambiguity, where risk is associated with the random nature of the infection outcomes, while ambiguity is associated with the unobservable stochastic model for generating the random infections. Indeed, epidemiological models have been studied since the early 20th century (Ross 1916, Kermack and McKendrick 1927) and they are indispensable tools for predicting infection trends during a pandemic. The SEIR compartmental model captures the transition of the population among four different population compartments labeled by: *Susceptible*, *Exposed*, *Infectious*, and *Removed* (*i.e.*, either recovered or deceased). It is built on a system of ordinary differential equations to predict the average number of individuals in each of the four population compartments at each time epoch. As a deterministic model, the SEIR model does not account for infection risks and its predictions on population averages are subject to uncertainty due to, among other things, the model having unobservable parameters that could only be estimated from data. When embedded in decision problems that could affect how the pandemic might propagate, using a predictive model that is oblivious to risk and uncertainty may underestimate the efforts and costs needed to achieve the desired outcomes for controlling the pandemic (a phenomenon known as the *Optimizer's Curse* by Smith and Winkler 2006).

Several variations of the SEIR compartmental model have been proposed to forecast the propagation trajectory of COVID-19. For instance, Li et al. (2020) modify the SEIR model and consider compartments for the symptomatic and asymptomatic infected individuals. Based on the data collected from Wuhan, China, they use their model to infer various epidemiological parameters. Chang et al. (2021) suggest that integrating the mobility network derived from mobile phone data into the SEIR model can help better predict the pandemic spread. Bertsimas et al. (2021a) design a data-driven framework to forecast the evolution of COVID-19. They modify the SEIR model and consider 13 compartments rather than four. These epidemiological models focus on incorporating more realistic states in the context of COVID-19; however, most of them are deterministic models and they rely on accurate input parameters, which are difficult to come by, especially when we are

dealing with new strains of virus that we are not familiar with. Stochastic SEIR models with random epidemiological parameters (*i.e.*, transmission rate, latency rate and recovery rate) assumed to follow given probability distributions have been proposed (Zhang and Wang 2014, Liu et al. 2017), though these studies focus on the stochastically asymptotic stability of the solution and the sufficient condition for disease-free equilibrium. Related to our model, Lekone and Finkenstädt (2006) consider a stochastic SVEIR model, utilizing binomial distributions to describe the compartmental state variables and their transitions. While stochastic epidemiological models can be more realistic, most of them are not analytically tractable.

The COVID-19 pandemic would likely become an endemic, hence, vaccination is possibly the best option available to facilitate safe reopening of the economy. However, to date, vaccines are still extremely limited resources. How vaccines should be properly distributed is thus an important problem. It is common for countries to prioritize vaccinations for people under high transmission and exposure risks, such as healthcare workers. However, the distribution of vaccines to the general population is subject to multiple constraints and its impact is subject to uncertainty. For instance, different population groups can have different exposure probabilities due to intra-group and inter-group interactions; the number of exposures and infections are stochastic; and vaccines can have different effectiveness to different population groups. Other realistic components such as fairness and social welfare should also be taken into considerations. Therefore, it is critical to optimize the vaccine distribution policy and utilize these limited resources effectively and properly.

From the operations perspective, vaccine treatment capacity management for pandemic control has well been studied in the literature. Sun et al. (2014) present a static model to allocate different types of patients into different hospitals with various equipment resources (*e.g.*, ICU beds and ventilators) for a planning horizon in which the length of stay is dependent on the types of patients. Some literature further integrate epidemiological model into the treatment capacity management (Liu and Zhang 2016, Büyüktaktakın et al. 2018). Long et al. (2018) focus on optimizing when and where to allocate bed capacity across geographic regions during the outbreak's early phases of Ebola. Three solution approaches, namely, greedy policy, myopic linear problem and approximate dynamic programming approach are presented to solve the non-convex model. A large proportion of the literature focus on vaccine supply chain (Duijzer et al. 2018b), while others focus on minimizing the number of vaccines needed to reasonably control the pandemics (Hill and Longini Jr 2003, Tanner and Ntamo 2010, Yarmand et al. 2014, Enayati and Özalpın 2020). Duijzer et al. (2018a) present a vaccine distribution model embedded with the SIR model to maximize the number of individuals averting infection under a given budget of vaccines; they demonstrate that allocating to all populations equally would reduce the effect of herd immunity. Sun et al. (2009) present a decentralized drug allocation game among several countries, which captures the uncertainties in

the number of initial infected individuals, the spread of pandemics, and the effect of drugs. Bertsimas et al. (2021b) integrate an epidemiological prediction model and various fairness measures into a novel data-driven model for a vaccination facility location model. Their epidemiological optimization model modifies the classical SEIR model and leads to a non-convex optimization problem, which can be solved by a coordinate descent algorithm. Mak et al. (2021) design several dynamic stocking policies for two-dose COVID-19 vaccine administration to determine the amount of vaccines needed to be reserved for second doses. Existing models in optimizing vaccination and healthcare resource allocation are mostly cost-driven, *i.e.*, they minimize the estimated total costs associated with each infection, which can be difficult to determine precisely. Alternatively, we can also control the total number of infections at each time point by setting intervention targets subject to available budget, which has an effect on relieving the pressure on the healthcare system. Such model aims to manage healthcare resource capacities by controlling the overall exposure rate in the society.

To account for ambiguous risks in the disease propagation, we propose a robust epidemiological model that provides prediction intervals specified by the Aumann and Serrano (2008) *riskiness index*. We focus on a SVEIR compartmental model, which is an extended SEIR model with an additional compartment label for vaccinated population. With suitable approximations, the robust epidemiological optimization model that minimizes the riskiness index can be formulated as a mixed integer linear optimization problem. This robust epidemiological model is closely related to the robustness optimization framework (Brown and Sim 2009, Zhou et al. 2021, Long et al. 2021). We illustrate how we can apply the robust epidemiological optimization model for strategic vaccine allocation by lexicographically minimizing the model’s riskiness indices for all the constraints on limiting infections across all time periods, and within a given budget for vaccinations. The optimal solutions of this robustness optimization model can be reliably obtained using state-of-the-art solvers such as CPLEX and Gurobi. Our model aims to satisfy a prescribed capacity of the healthcare system under uncertainty. Specifically, we restrict the number of infections at any time point to be below a prescribed capacity as much as possible under uncertainty. Such an objective can effectively spread out the peak of infection to alleviate the pressure on the healthcare system. The optimization model can incorporate realistic considerations in the pandemic progression and is formulated as a mixed-integer linear optimization model, where the solutions can be obtained by solving a sequence of mixed-integer linear optimization problems. In the simulation study, we elucidate that our proposed model is practically solvable and provides new insights on vaccine distribution. The average infection level of our robust optimization model outperforms benchmarks in most of the tests and achieves the lowest number of cumulative infections with high confidence.

With parameters estimated from real data, the elderly is prioritized for vaccination, which is consistent with the policy implemented in the New York City. When we vary the contact rates of different age groups, the vaccination distribution strategy also changes in a reasonable manner.

Notation. We typically use boldface lowercase letters for vectors (*e.g.*, θ), and calligraphic letters for sets (*e.g.*, \mathcal{X}). We use $[N]$ to denote the running index $\{1, 2, 3, \dots, N\}$ for N a known integer, and we use $[0, N]$ to denote $\{0\} \cup [N]$. We adopt the convention that $\inf \emptyset = +\infty$, where \emptyset is the empty set. A random variable \tilde{v} is denoted with a tilde sign such as $\tilde{v} \sim \mathbb{P}, \mathbb{P} \in \mathcal{P}_0$, where \mathcal{P}_0 to represent the set of all possible distributions. We use $\mathbb{E}_{\mathbb{P}}[\cdot]$ to signify the corresponding expectation.

2. Epidemiological prediction models

In this section, we demonstrate how we can extend a deterministic epidemiological compartmental model to address risk and uncertainty. We will focus on a SVEIR model, which is an extension of the celebrated SEIR model and is governed by a system of nonlinear ordinary differential equations across the following five population compartments labeled by: *Susceptible*, *Vaccinated*, *Exposed*, *Infectious*, and *Removed*. The population is also segmented into J different groups, and the interactions within and among the groups can lead to spread of infections.

Deterministic SVEIR model

In its discrete time form, the deterministic SVEIR forecasts the average populations segmented by compartments and groups over the time horizon of $t \in [T]$ based on their initial populations at time $t = 1$. As a simplified model, we assume that recovered individuals become immune to the disease, at least for a substantial amount of time greater than T . Hence, the recovered individuals do not re-enter the population of susceptible individuals, and we take the recovered population and deaths collectively as the removed population. For the purpose of modeling the impact of vaccination, we also include the decision variables $x_{j,t} \geq 0$, $j \in [J], t \in [T]$ that denote the number of scheduled vaccinations for the j th group at period t . The vaccinated population would be less likely to be exposed in subsequent periods. To account for different vaccinations and their efficacy with doses, we have to introduce new population compartments, which we can easily incorporate in an extended epidemiological compartmental model (see Section 4).

For a given vaccination allocation decision $x_{j,t} \in \{0, 1, \dots, \bar{S}_{j,t}\}$, $j \in [J]$, $t \in [T]$, the populations of different compartments in groups $j \in [J]$ at time period $t \in [T - 1]$ are as follows:

$$\bar{S}_{j,t+1} := (\bar{S}_{j,t} - x_{j,t}) (1 - \hat{q}_{j,t}) \quad (1a)$$

$$\bar{V}_{j,t+1} := \bar{V}_{j,t} (1 - \hat{\omega}_j \hat{q}_{j,t}) + x_{j,t} \quad (1b)$$

$$\bar{E}_{j,t+1} := (1 - \hat{\alpha}_j) \bar{E}_{j,t} + (\bar{S}_{j,t} - x_{j,t}) \hat{q}_{j,t} + \bar{V}_{j,t} \hat{\omega}_j \hat{q}_{j,t} \quad (1c)$$

$$\bar{I}_{j,t+1} := \hat{\alpha}_j \bar{E}_{j,t} + (1 - \hat{\gamma}_j) \bar{I}_{j,t} \quad (1d)$$

$$\bar{R}_{j,t+1} := \bar{R}_{j,t} + \hat{\gamma}_j \bar{I}_{j,t}. \quad (1e)$$

Equations (1a) and (1b) of the SVEIR model characterize how the average susceptible and vaccinated populations would evolve over time, respectively. The estimated contraction rate for the j th group at time t for the susceptible population is denoted by $\hat{q}_{j,t}$. Because a vaccinated person is less likely to be infected than a susceptible person, the corresponding estimated contraction rate for the vaccinated population would be $\hat{\omega}_j \hat{q}_{j,t}$, where $\hat{\omega}_j \in [0, 1]$ relates to the inefficacy of the vaccination for the j th group. Equation (1c) describes the average exposed population, where $\hat{\alpha}_j$ denotes the incubation rate for a person in the j th group. Equation (1d) governs the dynamics of the average infectious population, where $\hat{\gamma}_j$ represents the recovery rate for the patients in the j th group. Finally, Equation (1e) models the average removed population.

The main complexity of SVEIR model is due to the fact that the estimated contraction rate of an individual in any population group is a function of his interaction with his own group as well as infectious individuals in other groups as follows,

$$\hat{q}_{j,t} := \eta \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \right).$$

Specifically, the parameter $\hat{\zeta}_{k,j}$ relates to the estimated interaction-transmission rate between groups k and j , which can be expressed as $\hat{\zeta}_{k,j} = \frac{\hat{\beta}_k \hat{w}_{k,j}}{N_k}$, where $\hat{\beta}_k$ is the transmission rate of group k with total population N_j , and the interaction rate with the j th group is represented by $\hat{w}_{k,j}$. Several studies suggest using track records derived from mobile data to estimate the interaction rate (Birge et al. 2020, Chang et al. 2021). Hence, $\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t}$ represents the estimated interaction-transmission rate within the j th group and among other groups. In practice, the contraction rate may not necessarily be a linear function of the interaction-transmission rate, which can be due to, among other things, the complex dynamics involving public and governmental interventions. We cater for this flexibility by introducing the non-decreasing function η , which can be determined empirically to provide a more flexible mapping to the estimated contraction rate, $\hat{q}_{j,t}$ as a function of the interaction-transmission rate for the j th group. As we will reveal, for the optimization problem, we will also confine η to step functions to facilitate modeling as a mixed-integer linear optimization problem.

2.1. Stochastic SVEIR model

As a deterministic model, the SEIR model does not account for infection risks and prediction uncertainty. The next step is to extend to more realistic stochastic epidemiological models. We first introduce a stochastic SVEIR model that captures the populations as random variables instead

of deterministic averages. We use $\tilde{S}_{j,t}$, $\tilde{V}_{j,t}$, $\tilde{E}_{j,t}$, $\tilde{I}_{j,t}$ and $\tilde{R}_{j,t}$, $j \in [J], t \in [T]$ to respectively characterize the random populations of the susceptible, vaccinated, exposed, infectious, and removed compartments, which are stochastic processes that depend on the vaccination decision, \mathbf{x} . Based on binomial distributions (see Lekone and Finkenstädt 2006), the following stochastic SVEIR model (S-SVEIR) is a natural extension that captures the random populations of different compartments in groups $j \in [J]$ at time period $t \in [T - 1]$,

$$\tilde{S}_{j,t+1} := \text{Bin} \left(\tilde{S}_{j,t} - \min \left\{ x_{j,t}, \tilde{S}_{j,t} \right\}, 1 - \tilde{q}_{j,t} \right) \quad (2a)$$

$$\tilde{V}_{j,t+1} := \min \left\{ x_{j,t}, \tilde{S}_{j,t} \right\} + \text{Bin} \left(\tilde{V}_{j,t}, 1 - \hat{\omega}_j \tilde{q}_{j,t} \right) \quad (2b)$$

$$\tilde{E}_{j,t+1} := \text{Bin} \left(\tilde{E}_{j,t}, 1 - \hat{\alpha}_j \right) + \text{Bin} \left(\tilde{S}_{j,t} - \min \left\{ x_{j,t}, \tilde{S}_{j,t} \right\}, \tilde{q}_{j,t} \right) + \text{Bin} \left(\tilde{V}_{j,t}, \hat{\omega}_j \tilde{q}_{j,t} \right) \quad (2c)$$

$$\tilde{I}_{j,t+1} := \text{Bin} \left(\tilde{E}_{j,t}, \hat{\alpha}_j \right) + \text{Bin} \left(\tilde{I}_{j,t}, 1 - \hat{\gamma}_j \right) \quad (2d)$$

$$\tilde{R}_{j,t+1} := \tilde{R}_{j,t} + \text{Bin} \left(\tilde{I}_{j,t}, \hat{\gamma}_j \right), \quad (2e)$$

where

$$\tilde{q}_{j,t} := \eta \left(\sum_{k \in [J]} \frac{\hat{\beta}_k \hat{w}_{k,j}}{N_k} \tilde{I}_{k,t} \right),$$

and $\text{Bin}(\tilde{v}, p)$ describes a composite random variable that is drawn from a binomial distribution with success trial probability p over \tilde{v} independent Bernoulli trials. The stochastic population dynamics (2a)—(2e) describes the complex joint distributions of the populations in the various compartments in different groups over the different time periods. Unfortunately, the S-SVEIR model is complex to analyze; even to evaluate the average population would require numerical simulation, which inevitably introduces additional simulation uncertainty to the model. It is also not a particularly useful model for prescriptive analytics, especially when the feasible set of decisions is exponential in size or even infinite.

2.2. Robust SVEIR model

To motivate the robust prediction model, we first consider how we would evaluate the riskiness of a constraint under distributional ambiguity imposed on a random number of infected population denoted by $\tilde{I} \sim \hat{\mathbb{P}}$ to be held below the intervention capacity H . In evaluating riskiness under ambiguity, we would typically examine how likely the infected population would exceed the intervention capacity under the actual distribution, \mathbb{P}^* , which is not observed but could be proximal to $\hat{\mathbb{P}}$. Although we may use the feasibility probability, $\hat{\mathbb{P}} \left[\tilde{I} \leq H \right]$, the distribution $\hat{\mathbb{P}}$ generated by a stochastic SVEIR model would likely deviate from the true distribution, \mathbb{P}^* .

We propose using the riskiness index of Aumann and Serrano (2008) to evaluate the riskiness of a constraint under distributional ambiguity. Let \mathcal{L} be the space of all bounded discrete random

variables. For a given random variable $\tilde{v} \in \mathcal{L}$ with discrete probability distribution, $\hat{\mathbb{P}}, \tilde{v} \sim \hat{\mathbb{P}}$ and support

$$\mathcal{V} = \left\{ v \in \mathbb{R} \mid \hat{\mathbb{P}}[\tilde{v} = v] > 0 \right\},$$

the Aumann and Serrano (2008) riskiness index is a functional, $\rho: \mathcal{L} \mapsto [0, \infty]$ defined as follows

$$\rho[\tilde{v}] := \min \{ \theta \mid \mu_\theta[\tilde{v}] \leq 0, \theta \geq 0 \}$$

where $\mu_\theta: \mathcal{L} \mapsto \mathbb{R}$, $\theta \in [0, \infty]$ is the certainty equivalent under the exponential disutility as follows

$$\mu_\theta[\tilde{v}] := \begin{cases} \theta \log \mathbb{E}_{\hat{\mathbb{P}}}[\exp(\tilde{v}/\theta)] & \text{if } \theta \in (0, \infty) \\ \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] & \text{if } \theta = \infty \\ \max\{v \mid v \in \mathcal{V}\} & \text{if } \theta = 0. \end{cases}$$

THEOREM 1. *Let $\tilde{v} \in \mathcal{L}$ be a discrete random variable with support \mathcal{V} over the distribution $\hat{\mathbb{P}}$ and $\theta = \rho[\tilde{v}]$. Suppose the true distribution \mathbb{P}^* is absolutely continuous with respect to $\hat{\mathbb{P}}$ then*

1.

$$\mathbb{E}_{\mathbb{P}^*}[\tilde{v}] \leq \mu_\theta[\tilde{v}] + \theta \phi_{KL}(\mathbb{P}^* \parallel \hat{\mathbb{P}})$$

where $\phi_{KL}(\mathbb{Q} \parallel \hat{\mathbb{P}})$ is the KL-divergence as follows

$$\phi_{KL}(\mathbb{Q} \parallel \hat{\mathbb{P}}) = \sum_{v \in \mathcal{V}} \left(\mathbb{Q}[\tilde{v} = v] \log \left(\frac{\mathbb{Q}[\tilde{v} = v]}{\hat{\mathbb{P}}[\tilde{v} = v]} \right) \right).$$

2. For all $a \geq 0$

$$\mathbb{P}^* \left[\tilde{v} > \theta \left(a + \Delta \left(\mathbb{P}^* \parallel \hat{\mathbb{P}} \right) \right) \right] \leq \exp(-a),$$

where

$$\Delta \left(\mathbb{P}^* \parallel \hat{\mathbb{P}} \right) = \max_{v \in \mathcal{V}} \log \left(\frac{\mathbb{P}^*[\tilde{v} = v]}{\hat{\mathbb{P}}[\tilde{v} = v]} \right).$$

Hence, from Theorem 1, when estimating the average of a random variable, \tilde{v} , for a given riskiness index θ , $\mu_\theta[\tilde{v}]$ provides an upper estimate of $\mathbb{E}_{\mathbb{P}^*}[\tilde{v}]$ and the violation of the upper estimate diminishes with the product of the riskiness index and how far the true distribution, \mathbb{P}^* may deviate from the estimated distribution, $\hat{\mathbb{P}}$. Moreover, the riskiness index can also provide a probability guarantee in the situation where the true distribution is unobservable, but is proximal to the estimate distribution. Specifically, $\rho[\tilde{I} - H]$ is associated with the probability of violations that diminishes exponentially in the magnitude of the violation as follows

$$\mathbb{P}^* \left[\tilde{I} > H + a\rho[\tilde{I} - H] \right] \leq \exp \left(- \left(a - \Delta \left(\mathbb{P}^* \parallel \hat{\mathbb{P}} \right) \right) \right),$$

for all $a > \Delta \left(\mathbb{P}^* \parallel \hat{\mathbb{P}} \right)$, so that larger violations would become more unlikely. The strength of the bound also depends on how close the estimated distribution is with respect to the true distribution. If \mathbb{P}^* and $\hat{\mathbb{P}}$ are identical distributions, then $\Delta \left(\mathbb{P}^* \parallel \hat{\mathbb{P}} \right) = 0$ and we have for all $a > 0$,

$$\mathbb{P}^* \left[\tilde{I} > H + a\rho[\tilde{I} - H] \right] \leq \exp(-a).$$

A lower riskiness index is associated with a lower risk of constraint violation even under an ambiguous distribution. If this index is zero, there will be no risk of violating the constraint at all.

Observe that for an uncertain constraint $\tilde{I} \leq H$, with riskiness index θ , we have $\mu_\theta \left[\tilde{I} \right] = H$. Likewise for an uncertain constraint $\tilde{I} \geq L$ or $-\tilde{I} \leq -L$ with the same riskiness index, we would have $-\mu_\theta \left[-\tilde{I} \right] = L$. Therefore, the robust prediction interval of \tilde{I} for a given riskiness index θ is define as $[L, H]$ or $\left[-\mu_\theta \left[-\tilde{I} \right], \mu_\theta \left[\tilde{I} \right] \right]$.

To apply the riskiness index in the robust epidemiological predication model, we need to adopt a tractable stochastic SVEIR model that would enable us to evaluate the certainty equivalent under the exponential disutility. Observe that if \tilde{v}_1 and \tilde{v}_2 are independently distributed, then

$$\mu_\theta [\tilde{v}_1 + \tilde{v}_2] = \mu_\theta [\tilde{v}_1] + \mu_\theta [\tilde{v}_2].$$

The independence property is a key idea for proposing an analytically tractable stochastic SVEIR model (T-SVEIR) that stems from the deterministic SVEIR model. The random populations of different compartments over the time period $t \in [T - 1]$ are described as follows:

$$\tilde{S}_{j,t+1} := \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, 1 - \hat{q}_{j,t} \right) \quad (3a)$$

$$\tilde{V}_{j,t+1} := \text{Bin} \left(\tilde{V}_{j,t}, 1 - \hat{\omega}_j \hat{q}_{j,t} \right) + x_{j,t} \quad (3b)$$

$$\tilde{E}_{j,t+1} := \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, \hat{q}_{j,t} \right) + \text{Bin} \left(\tilde{E}_{j,t}, 1 - \hat{\alpha}_j \right) + \text{Bin} \left(\tilde{V}_{j,t}, \hat{\omega}_j \hat{q}_{j,t} \right) \quad (3c)$$

$$\tilde{I}_{j,t+1} := \text{Bin} \left(\tilde{E}_{j,t}, \hat{\alpha}_j \right) + \text{Bin} \left(\tilde{I}_{j,t}, 1 - \hat{\gamma}_j \right) \quad (3d)$$

$$\tilde{R}_{j,t+1} := \tilde{R}_{i,t} + \text{Bin} \left(\tilde{I}_{i,t}, \hat{\gamma}_i \right). \quad (3e)$$

Note that with vaccine allocation decision $x_{j,t} \in \{0, 1, \dots, \bar{S}_{j,t}\}$, it is possible for the composite binomial distribution to evaluate a negative number of trials. In this regard, which is important for analytical tractability, the equivalence in distributions is defined on their moment generation functions, which would still be analytically defined. Nevertheless, we remark that vaccines are highly limited, and the scale of daily vaccine allocation is much lower than the scale of the susceptible population; hence, one would not encounter this issue in practice.

The distribution of the tractable stochastic SVEIR model is denoted by $\hat{\mathbb{P}}$, and observe that

$$\bar{S}_{j,t} = \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{S}_{j,t} \right], \bar{V}_{j,t} = \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{V}_{j,t} \right], \bar{E}_{j,t} = \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{E}_{j,t} \right], \bar{I}_{j,t} = \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{I}_{j,t} \right], \bar{R}_{j,t} = \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{R}_{j,t} \right].$$

Hence, the tractable stochastic SVEIR model is a natural extension to the deterministic SVEIR model that incorporates risks in its model. The robust SVEIR (R-SVEIR) model predicts the robust interval of populations of different compartments in the T-SVEIR model using the riskiness index, *i.e.*, the R-SVEIR prediction intervals for a given riskiness index θ are

$$\begin{aligned} S_{j,t} &\in \left[-\mu_\theta \left[-\tilde{S}_{j,t} \right], \mu_\theta \left[\tilde{S}_{j,t} \right] \right], \\ V_{j,t} &\in \left[-\mu_\theta \left[-\tilde{V}_{j,t} \right], \mu_\theta \left[\tilde{V}_{j,t} \right] \right], \\ E_{j,t} &\in \left[-\mu_\theta \left[-\tilde{E}_{j,t} \right], \mu_\theta \left[\tilde{E}_{j,t} \right] \right], \\ I_{j,t} &\in \left[-\mu_\theta \left[-\tilde{I}_{j,t} \right], \mu_\theta \left[\tilde{I}_{j,t} \right] \right], \\ R_{j,t} &\in \left[-\mu_\theta \left[-\tilde{R}_{j,t} \right], \mu_\theta \left[\tilde{R}_{j,t} \right] \right]. \end{aligned}$$

Now, we show how we can recursively evaluate these certainty equivalents of the random populations, scaled by a constant, that arises in the different population segments.

PROPOSITION 1. *Ignoring the integrity of the number of trials parameter of the composite Binomial generating function, the certainty equivalent of the random populations in different compartments can be calculated recursively via the following system of equations:*

For $v \in \mathbb{R}$, $j \in [J]$, $t \in [T - 1]$,

$$\begin{aligned} \mu_\theta \left[v\tilde{S}_{j,t+1} \right] &= \mu_\theta \left[\tilde{S}_{j,t}\kappa_\theta(1 - \hat{q}_{j,t}, v) \right] - x_{j,t}\kappa_\theta(1 - \hat{q}_{j,t}, v) \\ \mu_\theta \left[v\tilde{V}_{j,t+1} \right] &= \mu_\theta \left[\tilde{V}_{j,t}\kappa_\theta(1 - \hat{\omega}_j\hat{q}_{j,t}, v) \right] + vx_{j,t} \\ \mu_\theta \left[v\tilde{E}_{j,t+1} \right] &= \mu_\theta \left[\tilde{S}_{j,t}\kappa_\theta(\hat{q}_{j,t}, v) \right] - x_{j,t}\kappa_\theta(\hat{q}_{j,t}, v) + \mu_\theta \left[\tilde{V}_{j,t}\kappa_\theta(\hat{\omega}_j\hat{q}_{j,t}, v) \right] + \mu_\theta \left[\tilde{E}_{j,t}\kappa_\theta(1 - \hat{\alpha}_j, v) \right] \\ \mu_\theta \left[v\tilde{I}_{j,t+1} \right] &= \mu_\theta \left[\tilde{E}_{j,t}\kappa_\theta(\hat{\alpha}_j, v) \right] + \mu_\theta \left[\tilde{I}_{j,t}\kappa_\theta(1 - \hat{\gamma}_j, v) \right] \\ \mu_\theta \left[v\tilde{R}_{j,t+1} \right] &= \mu_\theta \left[v\tilde{R}_{j,t} \right] + \mu_\theta \left[\tilde{I}_{j,t}\kappa_\theta(\hat{\gamma}_j, v) \right], \end{aligned} \quad (4)$$

where the function, $\kappa_\theta(q, v)$ is defined as follows

$$\kappa_\theta(q, v) := \theta \log(1 - q + q \exp(v/\theta)).$$

By Proposition 1, the R-SVEIR prediction intervals can be calculated via recursion, illustrating the practicality of the R-SVEIR model. In addition, both predictions coincide when $\theta \rightarrow +\infty$ because μ_θ becomes the expectation measure, illustrating that the R-SVEIR model is closely related to the deterministic SVEIR model.

2.3. Illustrations

Now, we provide two illustrations of the predictions of the deterministic SVEIR model and the R-SVEIR model. In Figure 1, we first plot the predictions of the size of infectious population under the SVEIR model over 400 days, labeled as SVEIR (β), indicating that the SVEIR model assumes a transmission rate of β . To illustrate the impact of β to the predictions, we also plot two sets of

SVEIR predictions under transmission rates of 0.99β and 1.01β , respectively. As we can see, even a 1% change in β could lead to a non-negligible impact to the predictions. In practice, the true transmission rate is most likely to deviate from the estimated value β , and the mean predictions would lie in a large range. In Figure 1, we also include the predictions of R-SVEIR with $\theta = 2$ and $\theta = 5$, both assuming the transmission rate is β . Note that a R-SVEIR model gives a robust prediction interval, *e.g.*, $[-\mu_\theta [-\tilde{I}], \mu_\theta [\tilde{I}]]$, it naturally hedges against the impact of inaccuracy in estimating the transmission rate β . Specifically, if we look at the predictions under the R-SVEIR model ($\theta = 5, \beta$), the robust prediction interval could very well describe the interval of possible mean predictions of the SVEIR model if the transmission rate was to deviate from β by 1%. The robust prediction interval of the R-SVEIR model ($\theta = 2, \beta$) accommodates great ambiguity that might arise.

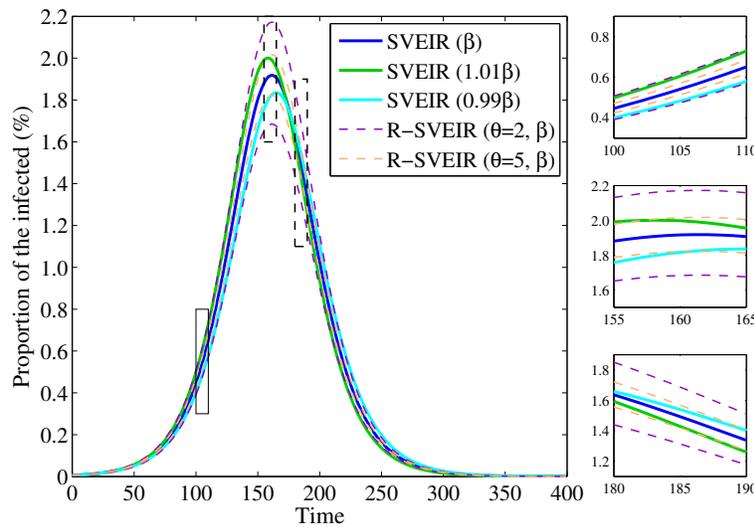


Figure 1 Deterministic SVEIR model and robust SVEIR model.

In Figure 2, we illustrate the distribution of the number of active infections at time 160, corresponding to the peak number of active infections in Figure 1. This histogram is generated by simulation under the S-SVEIR model with random transmission rate following uniform distribution $\text{Uniform}(0.99\beta, 1.01\beta)$. Note that the number of active infections generated by S-SVEIR model can vary from its deterministic SVEIR prediction (blue vertical line) by more than 6%, indicating that there is non-negligible risks in the number of infections. As the riskiness index θ varies, the robust prediction interval of the R-SVEIR model would capture the corresponding variation of the number of infections.

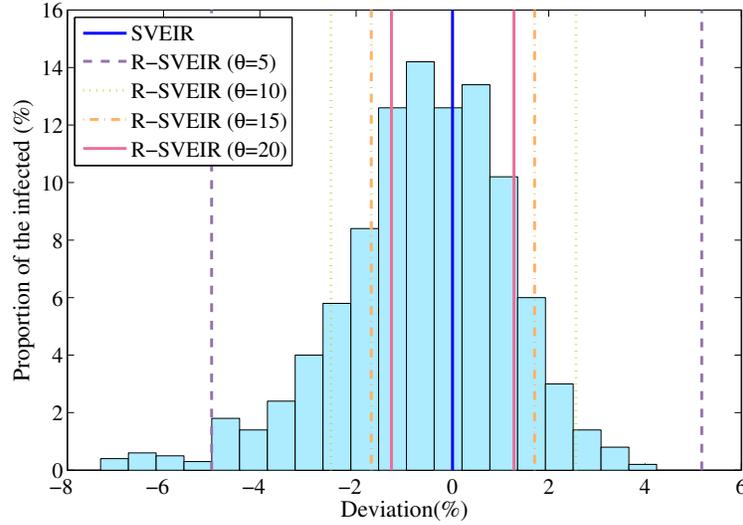


Figure 2 Histogram of the deviations of the number of active infections from its mean at time 160.

3. Robust Epidemiological Optimization

We incorporate the R-SVEIR model within a prescriptive analytics model by considering a vaccine allocation problem. By embedding the epidemiological prediction model in the optimization problem, we can anticipate how the vaccination allocation decision \mathbf{x} would affect the disease propagation over time. First, we consider the baseline model, which incorporates the deterministic SVEIR model in a vaccine allocation model that minimizes the total costs of vaccinations, while ensuring that the total infections at time period t do not exceed the healthcare capacity limits H_t . There is also a restriction on the number of vaccinations that can be carried out to at most B_t at time period t , representing the capacity of vaccine resources. We present this optimization model as follows:

$$\begin{aligned}
Z_0 = \min & \sum_{j \in [J]} \sum_{t \in [T]} x_{j,t} \\
\text{s.t.} & \sum_{j \in [J]} x_{j,t} \leq B_t \quad \forall t \in [T] \\
& \sum_{j \in [J]} \bar{I}_{j,t} \leq H_t \quad \forall t \in [T] \\
& \hat{q}_{j,t} = \eta \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \right) \quad \forall j \in [J], t \in [T-1] \\
(1a) \text{--}(1d) & \quad \forall j \in [J], t \in [T-1] \\
x_{j,t} & \leq \bar{S}_{j,t} \quad \forall j \in [J], t \in [T] \\
\bar{S}_{j,t}, \bar{V}_{j,t}, \bar{E}_{j,t}, \bar{I}_{j,t}, x_{j,t} & \geq 0 \quad \forall j \in [J], t \in [T].
\end{aligned} \tag{5}$$

The tractability of this model depends on the form of the contraction rate function η ; if it is a constant, then Problem (5) is a linear optimization problem. For practicality, we consider η being a step function, which allows the problem to be formulated as a mixed-integer linear optimization problem that we can solve to optimality using commercial solvers such as CPLEX and Gurobi.

THEOREM 2. *By restricting η to a step function*

$$\eta^\dagger = \begin{cases} 0 & \text{if } v \leq \bar{v}_1 \\ \underline{v}_l & \text{if } v \in (\bar{v}_l, \bar{v}_{l+1}], \text{ for some } l \in [L-1] \\ \underline{v}_L & \text{if } v > \bar{v}_L \end{cases}$$

where the parameters of the step function satisfy $0 = \underline{v}_0 < \underline{v}_1 < \dots < \underline{v}_L \leq 1$ and $0 = \bar{v}_0 \leq \bar{v}_1 < \dots < \bar{v}_L \leq \bar{v}_{L+1} = 1$, we can formulate the constraints

$$\begin{aligned} \hat{q}_{j,t} &= \eta^\dagger \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \right) & \forall j \in [J], t \in [T-1] \\ (1a) \text{---}(1d) & & \forall j \in [J], t \in [T-1] \end{aligned}$$

as the following set of mixed-integer linear constraints:

$$\begin{aligned} \bar{S}_{j,t+1} &= \bar{S}_{j,t} - x_{j,t} - \bar{W}_{j,t}, & \forall j \in [J], t \in [T-1] \\ \bar{W}_{j,t} &\leq (\bar{S}_{j,t} - x_{j,t}) \underline{v}_l + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\ \bar{W}_{j,t} &\geq (\bar{S}_{j,t} - x_{j,t}) \underline{v}_l - M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\ \bar{V}_{j,t+1} &= x_{j,t} + \bar{V}_{j,t} - \bar{Y}_{j,t} & \forall j \in [J], t \in [T-1] \\ \bar{Y}_{j,t} &\leq \hat{\omega}_j \bar{V}_{j,t} \underline{v}_l + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\ \bar{Y}_{j,t} &\geq \hat{\omega}_j \bar{V}_{j,t} \underline{v}_l - M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\ \bar{E}_{j,t+1} &= (1 - \hat{\alpha}_j) \bar{E}_{j,t} + (\bar{W}_{j,t} + \bar{Y}_{j,t}) & \forall j \in [J], t \in [T-1] \\ \bar{I}_{j,t+1} &= \hat{\alpha}_j \bar{E}_{j,t} + (1 - \hat{\gamma}_j) \bar{I}_{j,t} & \forall j \in [J], t \in [T-1] \\ \sum_{l \in [0, L]} y_{j,t}^l &= 1 & \forall j \in [J], t \in [T] \\ \bar{v}_l - M(1 - y_{j,t}^l) &\leq \sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \leq \bar{v}_{l+1} + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T], l \in [0, L] \\ \bar{W}_{j,t}, \bar{Y}_{j,t} &\geq 0 & \forall j \in [J], t \in [T] \\ y_{j,t}^l &\in \{0, 1\} & \forall j \in [J], t \in [T], l \in [0, L]. \end{aligned} \tag{6}$$

Hence, by restricting η to a step function as in Theorem 2, we can formulate Problem (5) as the following mixed-integer linear optimization problem.

$$\begin{aligned}
Z_0 = \min & \sum_{j \in [J]} \sum_{t \in [T]} x_{j,t} \\
\text{s.t.} & \sum_{j \in [J]} x_{j,t} \leq B_t & \forall t \in [T] \\
& \sum_{j \in [J]} \bar{I}_{j,t} \leq H_t & \forall t \in [T] \\
& x_{j,t} \leq \bar{S}_{j,t} & \forall j \in [J], t \in [T] \\
& \text{Constraint set (6)} \\
& \bar{S}_{j,t}, \bar{V}_{j,t}, \bar{E}_{j,t}, \bar{I}_{j,t}, x_{j,t}, \bar{W}_{j,t}, \bar{Y}_{j,t} \geq 0 \quad \forall j \in [J], t \in [T].
\end{aligned} \tag{7}$$

The deterministic optimization model is oblivious to risk and uncertainty; hence, it can underestimate the efforts and costs needed to control the actual number of infections below the allotted healthcare capacities. The R-SVEIR model can also be embedded in an optimization model that is comparable in size as the deterministic counterpart. As an extension to the deterministic SVEIR model of Problem (5), for a given acceptable cost target $\tau \geq Z_0$, we propose the following the *robustness optimization* model that lexicographically minimizes (see, *e.g.*, Isermann 1982, Qi 2017) the riskiness indices associated with keeping the infection constraints below desired healthcare capacities as follows:

$$\begin{aligned}
& \text{lexmin } \boldsymbol{\theta} \\
\text{s.t.} & \sum_{j \in [J]} \sum_{t \in [T]} x_{j,t} \leq \tau \\
& \sum_{j \in [J]} x_{j,t} \leq B_t & \forall t \in [T] \\
& \mu_{\theta_t} \left[\sum_{j \in [J]} \tilde{I}_{j,t} \right] \leq H_t & \forall t \in [T] \\
& \text{(1a)—(1d)} & \forall j \in [J], t \in [T-1] \\
& \hat{q}_{j,t} = \eta \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \right) & \forall j \in [J], t \in [T-1] \\
& x_{j,t} \leq \bar{S}_{j,t} & \forall j \in [J], t \in [T-1] \\
& \bar{S}_{j,t}, \bar{V}_{j,t}, \bar{E}_{j,t}, \bar{I}_{j,t}, x_{j,t}, \theta_t \geq 0 \quad \forall j \in [J], t \in [T].
\end{aligned} \tag{8}$$

Note that in the deterministic SVEIR model, we only need to track $5JT$ state variables (predictions). However, in the R-SVEIR model, we are required to track more variables. For instance, even for the susceptible population, we need to track $\mu_{\theta} \left[v \tilde{S}_{j,t} \right]$ for $j \in [J], t \in [T]$ and for a large number (finite) of possible constant multiplier values, $v \in \mathbb{R}$.

3.1. Normal approximation

In the following, we propose a normal approximation method to reduce the complexity of the robustness optimization model to the same level as the deterministic epidemiological optimization model. Specifically, we consider the normal approximation of the certainty equivalent, $\nu_\theta : \mathcal{L} \mapsto \mathbb{R}$ as follows

$$\nu_\theta [\tilde{v}] = \mathbb{E}_{\hat{\mathbb{P}}} [\tilde{v}] + \frac{1}{2\theta} \text{Var}_{\hat{\mathbb{P}}} [\tilde{v}],$$

so that if \tilde{v} is normally distributed, then $\nu_\theta [\tilde{v}] = \mu_\theta [\tilde{v}]$.

Observe that the random populations across different groups are stochastically independent under the joint distribution $\hat{\mathbb{P}}$ so that

$$\nu_\theta \left[\sum_{j \in [J]} \tilde{I}_{j,t} \right] = \sum_{j \in [J]} \nu_\theta [\tilde{I}_{j,t}] = \sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta} \text{Var}_{\hat{\mathbb{P}}} [\tilde{I}_{j,t}] \right).$$

Under the normal approximation, we consider the following robustness optimization model

$$\begin{aligned} & \text{lexmin } \boldsymbol{\theta} \\ & \text{s.t. } \sum_{j \in [J]} \sum_{t \in [T]} x_{j,t} \leq \tau \\ & \sum_{j \in [J]} x_{j,t} \leq B_t \quad \forall t \in [T] \\ & \sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta_t} \text{Var}_{\hat{\mathbb{P}}} [\tilde{I}_{j,t}] \right) \leq H_t \quad \forall t \in [T] \\ & (1a) \text{---}(1d) \quad \forall j \in [J], t \in [T-1] \\ & \hat{q}_{j,t} = \eta \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \right) \quad \forall j \in [J], t \in [T-1] \\ & x_{j,t} \leq \bar{S}_{j,t} \quad \forall j \in [J], t \in [T-1] \\ & \bar{S}_{j,t}, \bar{V}_{j,t}, \bar{E}_{j,t}, \bar{I}_{j,t}, x_{j,t}, \theta_t \geq 0 \quad \forall j \in [J], t \in [T], \end{aligned} \tag{9}$$

where difference lies in using normal approximation for evaluating the certainty equivalent of the infection populations. It suffices to evaluate the variance terms. To ease the notation, let $S_{j,t}^\dagger$, $V_{j,t}^\dagger$, $E_{j,t}^\dagger$, $I_{j,t}^\dagger$, $R_{j,t}^\dagger$ represent the variance terms, $\text{Var}_{\hat{\mathbb{P}}} [\tilde{S}_{j,t}]$, $\text{Var}_{\hat{\mathbb{P}}} [\tilde{V}_{j,t}]$, $\text{Var}_{\hat{\mathbb{P}}} [\tilde{E}_{j,t}]$, $\text{Var}_{\hat{\mathbb{P}}} [\tilde{I}_{j,t}]$, $\text{Var}_{\hat{\mathbb{P}}} [\tilde{R}_{j,t}]$, respectively. In the following proposition, we show how we can determine the variance terms recursively.

PROPOSITION 2. *The variance terms can be calculated recursively as follows:*

For $j \in [J], t \in [2, T]$,

$$\begin{aligned}
S_{j,t}^\dagger &= (\bar{S}_{j,t-1} - x_{j,t-1})\hat{q}_{j,t-1}(1 - \hat{q}_{j,t-1}) + (1 - \hat{q}_{j,t-1})^2 S_{j,t-1}^\dagger \\
V_{j,t}^\dagger &= \bar{V}_{j,t-1}\hat{\omega}_j\hat{q}_{j,t-1}(1 - \hat{\omega}_j\hat{q}_{j,t-1}) + (1 - \hat{\omega}_j\hat{q}_{j,t-1})^2 V_{j,t-1}^\dagger \\
E_{j,t}^\dagger &= (\bar{S}_{j,t-1} - x_{j,t-1})\hat{q}_{j,t-1}(1 - \hat{q}_{j,t-1}) + \hat{q}_{j,t-1}^2 S_{j,t-1}^\dagger \\
&\quad + \bar{V}_{j,t-1}\hat{\omega}_j\hat{q}_{j,t-1}(1 - \hat{\omega}_j\hat{q}_{j,t-1}) + (\hat{\omega}_j\hat{q}_{j,t-1})^2 V_{j,t-1}^\dagger \\
&\quad + \bar{E}_{j,t-1}(1 - \hat{\alpha}_j)\hat{\alpha}_j + (1 - \hat{\alpha}_j)^2 E_{j,t-1}^\dagger \\
I_{j,t}^\dagger &= \bar{I}_{j,t-1}(1 - \hat{\gamma}_j)\hat{\gamma}_j + (1 - \hat{\gamma}_j)^2 I_{j,t-1}^\dagger + \bar{E}_{j,t-1}\hat{\alpha}_j(1 - \hat{\alpha}_j) + \hat{\alpha}_j^2 E_{j,t-1}^\dagger \\
R_{j,t}^\dagger &= R_{j,t-1}^\dagger + \bar{I}_{j,t-1}(1 - \hat{\gamma}_j)\hat{\gamma}_j + \hat{\gamma}_j^2 I_{j,t-1}^\dagger.
\end{aligned} \tag{10}$$

By Proposition 2, the recursive equations for the variance terms have the same complexity as that of the recursion of the deterministic SVEIR model. More importantly, as we will reveal later, having the normal approximation would help the robustness optimization model maintain the same computational complexity as the deterministic optimization model.

The system of equations in Proposition 2 is analogous to the state dynamics of the deterministic SVEIR model. Hence, they can be readily incorporated into an optimization model. In the following theorem, we confirm that we can reformulate the constraint

$$\sum_{i \in [J]} \nu_\theta [\tilde{I}_{j,t}] = \sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta} I_{j,t}^\dagger \right) \leq H_t$$

as a set of mixed-integer linear constraints when η is given by a step function.

THEOREM 3. *By restricting η to a step function*

$$\eta^\dagger = \begin{cases} 0 & \text{if } v \leq \bar{v}_1 \\ \underline{v}_l & \text{if } v \in (\bar{v}_l, \bar{v}_{l+1}], \text{ for some } l \in [L-1] \\ \underline{v}_L & \text{if } v > \bar{v}_L \end{cases}$$

where the parameters of the step function satisfy $0 = \underline{v}_0 < \underline{v}_1 < \dots < \underline{v}_L \leq 1$ and $0 = \bar{v}_0 \leq \bar{v}_1 < \dots < \bar{v}_L < \bar{v}_{L+1} \leq 1$, we can model the constraint $\sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta_t} \text{Var}_{\hat{\mathbb{P}}} [\tilde{I}_{j,t}] \right) \leq H_t$ with

$$\sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta_t} I_{j,t}^\dagger \right) \leq H_t \quad \forall t \in [T]$$

and the following set of mixed-integer linear constraints:

$$\begin{aligned}
I_{j,t+1}^\dagger &= \bar{I}_{j,t}(1 - \hat{\gamma}_j)\hat{\gamma}_j + (1 - \hat{\gamma}_j)^2 I_{j,t}^\dagger + \bar{E}_{j,t}\hat{\alpha}_j(1 - \hat{\alpha}_j) + \hat{\alpha}_j^2 E_{j,t}^\dagger & \forall j \in [J], t \in [T-1] \\
S_{j,t+1}^\dagger &\leq (\bar{S}_{j,t} - x_{j,t})\underline{v}_l(1 - \underline{v}_l) + (1 - \underline{v}_l)^2 S_{j,t}^\dagger + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\
S_{j,t+1}^\dagger &\geq (\bar{S}_{j,t} - x_{j,t})\underline{v}_l(1 - \underline{v}_l) + (1 - \underline{v}_l)^2 S_{j,t}^\dagger - M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\
V_{j,t+1}^\dagger &\leq \bar{V}_{j,t}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (1 - \hat{\omega}_j\underline{v}_l)^2 V_{j,t}^\dagger + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\
V_{j,t+1}^\dagger &\geq \bar{V}_{j,t}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (1 - \hat{\omega}_j\underline{v}_l)^2 V_{j,t}^\dagger - M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\
E_{j,t+1}^\dagger &\leq \bar{E}_{j,t}(1 - \hat{\alpha}_j)\hat{\alpha}_j + (1 - \hat{\alpha}_j)^2 E_{j,t}^\dagger & \\
&\quad + (\bar{S}_{j,t} - x_{j,t})\underline{v}_l(1 - \underline{v}_l) + \underline{v}_l^2 S_{j,t}^\dagger & \\
&\quad + \bar{V}_{j,t}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (\hat{\omega}_j\underline{v}_l)^2 V_{j,t}^\dagger + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L] \\
E_{j,t+1}^\dagger &\geq \bar{E}_{j,t}(1 - \hat{\alpha}_j)\hat{\alpha}_j + (1 - \hat{\alpha}_j)^2 E_{j,t}^\dagger & \\
&\quad + (\bar{S}_{j,t} - x_{j,t})\underline{v}_l(1 - \underline{v}_l) + \underline{v}_l^2 S_{j,t}^\dagger & \\
&\quad + \bar{V}_{j,t}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (\hat{\omega}_j\underline{v}_l)^2 V_{j,t}^\dagger - M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], l \in [0, L].
\end{aligned} \tag{11}$$

We remark that the computational complexity required to evaluate this constraint is the same as that to evaluate the corresponding constraint under the deterministic optimization model. By Theorem 3, the robustness optimization model embedded with R-SVEIR has the same complexity as the deterministic optimization model embedded with deterministic SVEIR model.

THEOREM 4. *The robustness optimization model can be reformulated as a lexicographical minimization problem with only mixed-integer linear constraints. In addition, the numbers of variables and constraints of the robustness optimization problem are of the same order as those of the deterministic optimization model (7). Specifically, the robustness optimization model can be written as the following mixed-integer linear optimization problem:*

lexmin θ

$$\begin{aligned}
\text{s.t. } & \sum_{j \in [J]} \sum_{t \in [T]} x_{j,t} \leq \tau \\
& \sum_{j \in [J]} x_{j,t} \leq B_t & \forall t \in [T] \\
& \sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta_t} I_{j,t}^\dagger \right) \leq H_t & \forall t \in [T]
\end{aligned}$$

Constraint set (6)

Constraint set (11)

$$x_{j,t} \leq \bar{S}_{j,t}, \quad \forall j \in [J], t \in [T]$$

$$\bar{S}_{j,t}, \bar{V}_{j,t}, \bar{E}_{j,t}, \bar{I}_{j,t}, x_{j,t}, \bar{W}_{j,t}, \bar{Y}_{j,t}, S_{j,t}^\dagger, V_{j,t}^\dagger, E_{j,t}^\dagger, I_{j,t}^\dagger \geq 0 \quad \forall j \in [J], t \in [T].$$

For tractability, we have restricted η to a step function. We remark that restricting this contraction rate to a step function is a mild and safe approximation if the true contraction rate is a linear function. Figure 3 displays the histogram of the number of cumulative infections at the end of the planning horizon, where the simulation is conducted based on both the T-SVEIR model with linear and step contraction functions. The interval in the step function between any two levels is $\bar{v}^l - \bar{v}^{l-1} = 1 \times 10^{-5}$. Note that the gap between \bar{v}_1 and \bar{v}_L is small in practice, and the number of pieces, L , is manageable. As we can see, the two distribution well overlap, and the T-SVEIR model with step contraction function has a slightly higher mean.

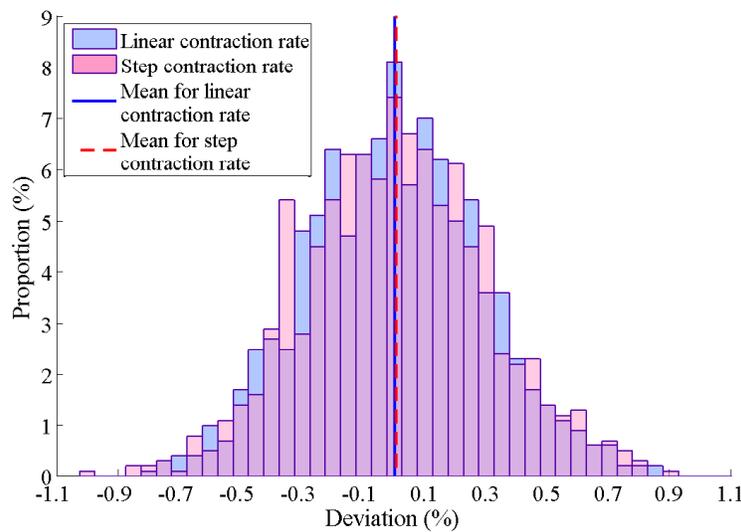


Figure 3 Linear contraction rate vs. step contraction rate

We also remark that the normal approximation is mild. It is well accepted that a binomial random variable, $\text{Bin}(N, p)$, can be closely approximated by a normal random variable with the same mean and variance when $Np \geq 5$ and $N(1 - p) \geq 5$. This is easily true in the context of pandemics, *e.g.*, the number of daily infections is often high. We should expect that the predictions under normal approximation are close to the exact predictions. In addition, the normal approximation is a conservative approximation. As an illustration, we plot the the percentage gap between the magnitude of $\nu_\theta[\tilde{I}_t]$ and $\mu_\theta[\tilde{I}_t]$ over the planning horizon in Figure 4. As we see, $\nu_\theta[\tilde{I}_t]$ overestimates $\mu_\theta[\tilde{I}_t]$, indicating it is a conservative, or safe approximation while implemented in the optimization model. The gap between $\mu_\theta[\tilde{I}_t]$ and $\nu_\theta[\tilde{I}_t]$ becomes larger as the value of θ becomes smaller. Although there is a gap between $\mu_\theta[\tilde{I}_t]$ and $\nu_\theta[\tilde{I}_t]$, it would not lead to over-conservativeness in our robustness optimization model. Note that the parameters θ_t , for all $t \in [T]$ appear in the objective function and both μ_θ and ν_θ are monotone in θ . Because θ_t only appears in the robustness constraint of

time t , the conservativeness due to the gap between $\mu_\theta[\tilde{I}_t]$ and $\nu_\theta[\tilde{I}_t]$ would not impact the solution much because the model would be able to scale down θ_t to “compensate” the gap.

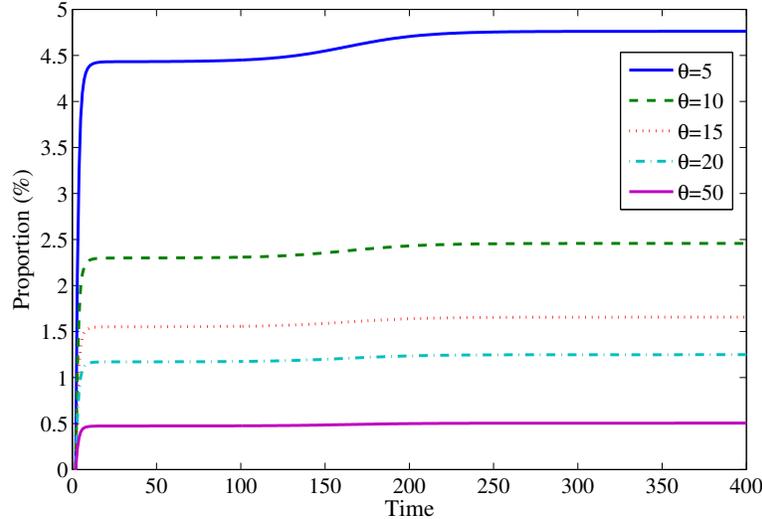


Figure 4 Percentage increase of $\nu_\theta[\tilde{I}_t]$ relative to $\mu_\theta[\tilde{I}_t]$.

3.2. Solution algorithm: Bisection and lexicographical minimization

Our robustness optimization model minimizes the lexicographical order of the vector θ . The solution algorithm relies on a bisection search subroutine.

Bisection search

1. Initialize with two given input sets, \mathcal{T}_0 and \mathcal{T}_1 , and values θ_t^* for $t \in \mathcal{T}_0$.
2. Set $\bar{\theta} = M$ for a large number, M , and set $\underline{\theta} = 0$.
3. If $\bar{\theta} - \underline{\theta} \leq \epsilon$, then we record $\theta^* = (\underline{\theta} + \bar{\theta})/2$ and find the index set $\mathcal{T}_b \subseteq \mathcal{T}_1$, such that the robustness constraint is binding for any time index $\hat{t} \in \mathcal{T}_b$, and we terminate the algorithm. Otherwise, set $\theta = (\underline{\theta} + \bar{\theta})/2$ and continue with the next step.
4. Solve the following mixed-integer linear feasibility problem.

$$\begin{aligned}
 \min \quad & \sum_{j \in [J]} \sum_{t \in [T]} x_{j,t} \\
 \text{s.t.} \quad & \sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta_t^*} I_{j,t}^\dagger \right) \leq H_t \quad \forall t \in \mathcal{T}_0 \\
 & \sum_{j \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta} I_{j,t}^\dagger \right) \leq H_t \quad \forall t \in \mathcal{T}_1
 \end{aligned} \tag{12}$$

The rest of the constraints,

where we omit the explicit expression of the rest of the constraints (see Theorem 4) for brevity.

5. If the objective value of the Problem (12) is less than τ , set $\bar{\theta} = \theta$. Otherwise, set $\underline{\theta} = \theta$. Repeat from step 2.

The lexicographical minimization procedure iteratively incurs the above bisection search algorithm. We summarize this procedure below.

Lexicographical minimization

1. Initialize with $\mathcal{T}_0 = \emptyset$ and $\mathcal{T}_1 = [T]$.
2. Solve the bisection search with input \mathcal{T}_0 , \mathcal{T}_1 , and values θ_t^* for $t \in \mathcal{T}_0$ if $\mathcal{T}_0 \neq \emptyset$. Get the output θ^* and \mathcal{T}_b . Let $\theta_{\hat{t}}^* = \theta^*$ for all $\hat{t} \in \mathcal{T}_b$. Then, include all elements of \mathcal{T}_b in \mathcal{T}_0 and exclude them from \mathcal{T}_1 .
3. If $\mathcal{T}_1 = \emptyset$, then we terminate the algorithm. Otherwise, repeat from step 2.

Note that each bisection search only takes $\log_2(M/\epsilon)$ number of iterations to terminate. In practice, this is a small number. For instance, in the simulation study, we set $M = 1000$, and $\epsilon = 0.01$; each of the bisection search terminates within ten iterations. By definition, the lexicographical minimization procedure terminates in at most $T \log_2(M/\epsilon)$ iterations, which scales linearly in T .

4. Extensions

In this section, we briefly discuss several extensions of the robust epidemiological prediction and optimization model.

Incorporating additional population compartments. For the ease of exposition, we only considered five population compartments in our main model. Here, we remark that our model can be tractably extended to incorporate more realistic population compartments such as population groups that are vaccinated with different types of vaccines, hospitalized, isolated, asymptomatic, infected with mutated viruses, and *etc.* One could refer to existing literature on examples of realistic population compartments (Bertsimas et al. 2021b).

As a concrete example, we can consider P different types of vaccines by incorporating state variables $\tilde{V}_{j,t}^{(1)}, \dots, \tilde{V}_{j,t}^{(P)}$ in the T-SVEIR model. Similarly, we can model multiple variants of the virus by incorporating state variables $\tilde{E}_{j,t}^{(1)}, \dots, \tilde{E}_{j,t}^{(R)}$ in the T-SVEIR model. To do so, for any $t \in [T-1], j \in [J]$, we modify Equations (3a)—(3d) to the following set of equations

$$\begin{aligned}
\tilde{S}_{j,t+1} &:= \text{Bin} \left(\tilde{S}_{j,t} - \sum_{p \in [P]} x_{j,t,p}, 1 - \sum_{p \in [R]} \hat{q}_{j,t,p} \right) \\
\tilde{V}_{j,t+1}^{(p)} &:= \text{Bin} \left(\tilde{V}_{j,t}^{(p)}, 1 - \sum_{r \in [R]} \hat{\omega}_{j,p,r} \hat{q}_{j,t,r} \right) + x_{j,t,p} \quad \forall p \in [P] \\
\tilde{E}_{j,t+1}^{(r)} &:= \text{Bin} \left(\tilde{S}_{j,t} - \sum_{p \in [P]} x_{j,t,p}, \hat{q}_{j,t,r} \right) + \text{Bin} \left(\tilde{E}_{j,t}^{(r)}, 1 - \hat{\alpha}_{j,r} \right) + \sum_{p \in [P]} \text{Bin} \left(\tilde{V}_{j,t}^{(p)}, \hat{\omega}_{j,p,r} \hat{q}_{j,t,r} \right) \quad \forall r \in [R] \\
\tilde{I}_{j,t+1}^{(r)} &:= \text{Bin} \left(\tilde{E}_{j,t}^{(r)}, \hat{\alpha}_{j,r} \right) + \text{Bin} \left(\tilde{I}_{j,t}^{(r)}, 1 - \hat{\gamma}_{j,r} \right) \quad \forall r \in [R],
\end{aligned}$$

where $\hat{\omega}_{j,p,r}$ denote the inefficacy of the type- p vaccine to the type- r virus, $\hat{q}_{j,t,r}$ denotes the contraction rate of type- r virus, $\hat{\alpha}_{j,r}$ is the incubation rate of type- r virus, and $\hat{\gamma}_{j,r}$ represents the recovery rate of type- r virus.

The general guideline is that as long as we can describe the system of dynamics under a T-SVEIR model in a similar form as (3a)—(3e), we can derive a counterpart of Proposition 1 and 2. Hence, we can get the robust prediction intervals of the different population compartments and derive a tractable reformulation of the robustness vaccine allocation optimization model. The rest of the results also follow similarly.

Government and healthcare interventions. In practice, once the result of the PCR swab test is positive, the tested individual has to be quarantined either at home or at some designated facilities so that they cannot contact with the general population. In this case, we can incorporate some decisions in $\hat{\mathbf{q}}$, *e.g.*,

$$\hat{q}_{j,t} := \eta \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} (\bar{I}_{k,t} - y_{k,t}) \right),$$

for some quarantine decisions $y_{j,t}$, $j \in [J]$, $t \in [T]$. In other words, only non-quarantined infectious individuals could potentially infect the susceptible population. Note that this does not affect the system of dynamics in (3a)—(3e), except the contraction rate $\hat{q}_{j,t}$; hence, the reformulation of the robustness optimization model follows similarly as illustrated in the main text.

Similar to Birge et al. (2020), we can also incorporate decisions on the implementation of NPIs to control the pandemics, that is, we can let

$$\hat{\zeta}_{k,j} = \sum_{r \in [R]} \frac{\beta_k w_{k,j,r} a_{k,t,r}}{N_k},$$

where R denotes the number of potential NPI policies to implement. The binary NPI policy implementation decision $a_{k,t,r}$ equals 1 if NPI policy r is adopted in group k at time t and 0 otherwise. Accordingly, $w_{k,j,r}$ denotes the interaction rate under NPI policy r . For instance, if NPI policy r represents a complete lockdown, then $w_{k,j,r}$ would almost be zero. In general, each NPI policy r can represent a bundle/set of intervention rules determined beforehand; hence, we can restrict $\sum_{r \in [R]} a_{j,t,r} = 1$ for $j \in [J]$, $t \in [T]$ without loss of generality. Based on Theorems 2 and 3, we can incorporate such NPI implementation decisions tractably as before.

Fairness. Vaccine distribution decision may be subject to fairness considerations among population groups. Some examples of fairness considerations in vaccine allocation are considered in Bertsimas et al. (2021b). We can incorporate similar fairness constraints as they are deterministic

mixed-integer linear constraints. For instance, the following fairness constraint ensures that the number of vaccines distributed to group j does not deviate too much from that to group k , *i.e.*,

$$\underline{F}_t \leq \frac{x_{j,t}}{N_j} - \frac{x_{k,t}}{N_k} \leq \bar{F}_t \quad \forall t \in [T], \forall j, k \in [J], j \neq k,$$

for some fairness parameters \underline{F}_t and \bar{F}_t . Because this is a deterministic linear constraint, it does not affect the established tractability results of our model.

5. Numerical Study

In this section, we test the proposed robustness vaccine distribution model using open-source datasets for COVID-19 pandemics in New York City.

5.1. Simulation setting

We consider a planning horizon of 16 days from Jan 31, 2021 (*i.e.*, $T = 16$) and plan for the vaccine allocation in the first two weeks. We define four population groups based on age. The age groups are split into 0-14, 15-34, 35-64, and 65 years old and above. The population sizes, N_j , for $j \in [J]$, are reported by US Census Bureau (2018).

Parameters. Age groups have different latency rates, transmission rates, and infection rates. We assume that $\hat{\alpha}_j = \alpha \Delta \alpha_j$, $\hat{\beta}_j = \beta \Delta \beta_j$ and $\hat{\gamma}_j = \gamma \Delta \gamma_j$, where α , β , and γ are the reference rates, and $\Delta \alpha_j$, $\Delta \beta_j$, and $\Delta \gamma_j$ are the deviations of the respective rates of group j . Li et al. (2020) present an inference model with mobility characteristics and use observations of reported infections in China in the early state of the outbreak. Their estimates of reference latency period (α^{-1}) and reference infectious period (γ^{-1}) during 24 January to 8 February are 3.42 days (95% credible interval (CI): 3.30- 3.65) and 3.31 days (95% CI: 2.96-3.88), respectively. We will use these values in our simulation.

The reference transmission rate, β , depends on the effectiveness of different non-pharmacological interventions and can vary among different population groups. We estimate the transmission rate by $\beta = R_t/\gamma$, where R_t is the effective reproduction rate at time t and is estimated in Abbott et al. (2020). Related to the interaction rate, we define $\hat{w}_{j,j} = 0.05$ and $\hat{w}_{j,k} = 0.4\hat{w}_{j,j}$ for $j \neq k$, indicating that the intra-group interaction rate is larger than the inter-group interaction rate. For the effectiveness of vaccines, we define $\hat{\omega}_j = 0.2$, implying that vaccination can reduce the infection rate by 80%. We summarize the values of all exogenous parameters in Tables 1 and 2.

For simplicity, the capacity of vaccines B_t (in thousands) in each period belongs to $\{20, 40\}$. In practice, the healthcare capacity would depend on the budget of vaccines. Hence, to ensure the healthcare capacity limit is not loose, we first solve the deterministic SVEIR prediction model with uniform vaccine distribution (*i.e.*, $x_{j,t} = B_t/J$) and obtain $\bar{I}_{j,t}$. Then, we set $H_t(B_t) = \sum_{j \in [J]} \bar{I}_{j,t}$,

where we use the notation $H_t(B_t)$ to emphasize that this value depends on B_t . In our simulation, we present the results under $H_t(20)$. Subsequently, we solve the deterministic model (5) to obtain Z_0 and let the cost target parameter τ in robustness model (9) equal to ϕZ_0 for some normalized target $\phi \geq 1$. Under this setup, both the deterministic and robustness optimization models are feasible.

Parameters	Values
Effective reproduction rate (R_t)	0.9
Reference latency rate (α , days ⁻¹)	1/3.42
Reference infectious rate (γ , days ⁻¹)	1/3.31
Reference transmission rate (β , days ⁻¹)	R_t/γ
Contact rate ($\hat{w}_{j,j}$)	0.05
Effect of vaccine ($\hat{\omega}_j$)	0.2

Table 1 Value of epidemiological parameters and vaccine budget

Parameters	0-14 years	15-34 years	35-64 years	65 years and above
Deviation of latency rate ($\Delta\alpha$)	1.0	1.0	1.0	1.0
Deviation of infectious rate ($\Delta\gamma$)	1.2	1.1	1.0	0.8
Deviation of transmission rate ($\Delta\beta$)	1.0	1.4	1.2	0.6

Table 2 Value of parameters $\Delta\alpha$, $\Delta\gamma$ and $\Delta\beta$

NYC Health (2020) has reported the daily confirmed cases for each county in New York City. However, there is a reporting delay between the daily confirmed cases and the daily new infection cases. Li et al. (2020) assume that this reporting delay follows a gamma distribution $\Gamma(2.34, 2.59)$. It implies that the mean reporting delay is about one day. For simplicity, we assume that the reporting delay is one day, and $\bar{E}_{j,1}$ can be estimated by the number of daily new infection cases at time two divided by $\hat{\alpha}_j$. We also estimate the infections at the beginning of the horizon, $\bar{I}_{j,1}$ by deterministic SVEIR model $\bar{I}_{j,t} = (1 - \hat{\gamma}_j)\bar{I}_{j,t-1} + \hat{\alpha}_j\bar{E}_{j,t-1} \approx \sum_{\tau \in [T]} (1 - \hat{\gamma}_j)^{r-1} \hat{\alpha}_j \hat{E}_{j,t-r}$. Finally, we assume that $\bar{S}_{j,1}$ equals to the total population size in region j , N_j , less the sum of (i) the cumulative number of infected cases during Feb. 29, 2020 to Jan 30, 2021, and (ii) the number of current exposed cases.

We use a step function to model the contraction rate, as in Theorem 2 and 3. We set the interval length (*i.e.*, $\bar{v}_{l+1} - \bar{v}_l$) equals 5×10^{-6} and $\underline{v}_l = (\bar{v}_l + \bar{v}_{l+1})/2$. The gap between \bar{v}_1 and \bar{v}_L is small in practice, and the same applies in this simulation; hence, The number of infection levels L is manageable. This number depends on transmission rate and cost budget of vaccines and is set accordingly.

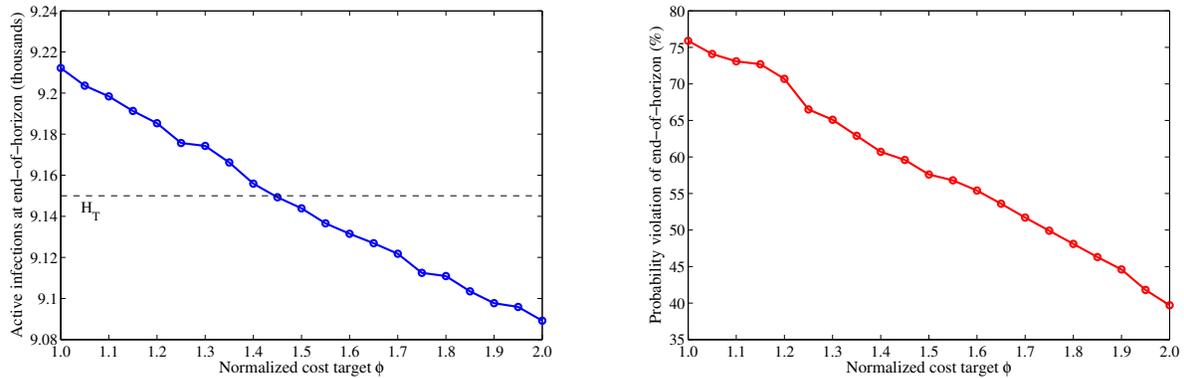
Simulation procedure. For all performance analysis, we generate 1,000 sample paths by simulating the state transitions according to the S-SVEIR model, which reflect the true distribution of the various state variables. The parameter values in S-SVEIR model is the same as those listed in Table 1 and Table 2. We refer to this as the fixed parameter setting.

In reality, the transmission rate is highly dependent on effects of NPIs and population behaviours. For example, government may reopen economic activities after there is a certain number of vaccinated individuals so that the transmission rate may increase. Hence, we also test our model under misspecification of transmission rates.

5.2. Results

We first illustrate the impact of total vaccination cost target to the robustness of the model. We illustrate how increasing total cost target could help reduce the average number of end-of-horizon active infections and the probability of violating the healthcare capacity. Figure 5 plots these two metrics under the instance with $B_t = 40$. The x-axis indicates the value of the normalized cost target ϕ , *i.e.*, the cost target τ is given by ϕZ_0 . When $\phi = 1$, the robustness optimization model would use at most the same amount of vaccines as the deterministic model. As we can see in Figure 5, the average number of end-of-horizon active infections exceeds the healthcare capacity H_t when $\phi = 1$, and the probability of violating the healthcare capacity is remarkably high. This indicates that the deterministic optimization model would significantly underestimate the risks in disease propagation, and governments and healthcare planners are highly likely to suffer from the cost of violating healthcare capacity due to vastly insufficient vaccination. This indicates the need of robustness by increasing total cost target of vaccine allocation. As we increase ϕ , both the average number of active infections and probability of exceeding healthcare capacity decrease. Hence, the tradeoff for the planner to address is between the total cost target and the robustness to satisfy the healthcare capacity.

At the first glance, we reduce the end-of-horizon active infections by less than 100 when we increase ϕ from 1.0 to 1.5, indicating that a 50% increase in vaccination cost only leads to a relatively small number of reduction in end-of-horizon active infections. However, we remark that vaccination has a long-term effect that is not fully reflected within the planning horizon of $T = 16$. Even a small difference within our 16-day planning horizon could lead to a significant change in the long run. To illustrate this, we further simulate disease propagation for 100 days under a no-vaccine policy starting from the end of our 16-day planning horizon, within which we implementing the vaccination policies of the deterministic model and robustness optimization with different normalized cost targets. Specifically, there are $T + 100 = 116$ days in the simulation, and the last 100 days are used for investigating how the differences of the 16-day vaccine allocation policies would magnify under no additional intervention, *e.g.*, under a no-vaccine policy.



(a) Average end-of-horizon active infections (b) Probability of exceeding healthcare capacity
Figure 5 Out-of-sample metrics of the deterministic and robustness optimization models (with different normalized cost target ϕ) under $B_t = 40$.

Figure 6 plots that the average number and 95% quantile of active infections from time T to $T + 100$ under a no-vaccine policy, under different initial states due to different vaccine allocation policies in the planning horizon $t \in [T]$. As an additional benchmark, we also plot the trend under a no-vaccine policy over the entire duration of $T + 100$ days. As we can see, the differences of daily active infections among different models would become more significant in the long run, even when there is no additional intervention. Hence, an increase in vaccination cost target in a planning horizon of only 16 days can help reduce the active infections significantly in the long run. Similarly, we also look at the average number and 95% quantile of the number of cumulative infections at the end of planning horizon (time T) and $T + 100$ in Figure 7. The result is also strong. Consider $\phi = 1.05$ (represented by the green bar), *i.e.*, we only increase the total vaccination cost by 5% compared to the deterministic model. At the end of planning horizon, T , we reduce the average number of cumulative infections by about 30, while the reduction of average cumulative infections increases to nearly 1,500 in the long run at no additional cost and additional intervention.

By above results and observations, we remark that our model can lead to a significantly better long-term performance at the cost of incurring a slightly higher vaccination cost over a short planning horizon. Note that the simulation of the additional 100 days is conducted under a no-vaccine policy to illustrate how the performance improvement can magnify over time. In practice, one could apply our model in a rolling horizon fashion with a sliding window, which would lead to larger improvements.

Now, we investigate the model robustness to satisfying the healthcare capacity in Figure 8. In Figure 8(a), we plot the probability that the number of active infections exceeds the healthcare capacity. As we can see, bearing a small increase in the total vaccination cost can lead to a lower probability of violating healthcare capacity in all time periods. In Figure 8(b), we focus on the end

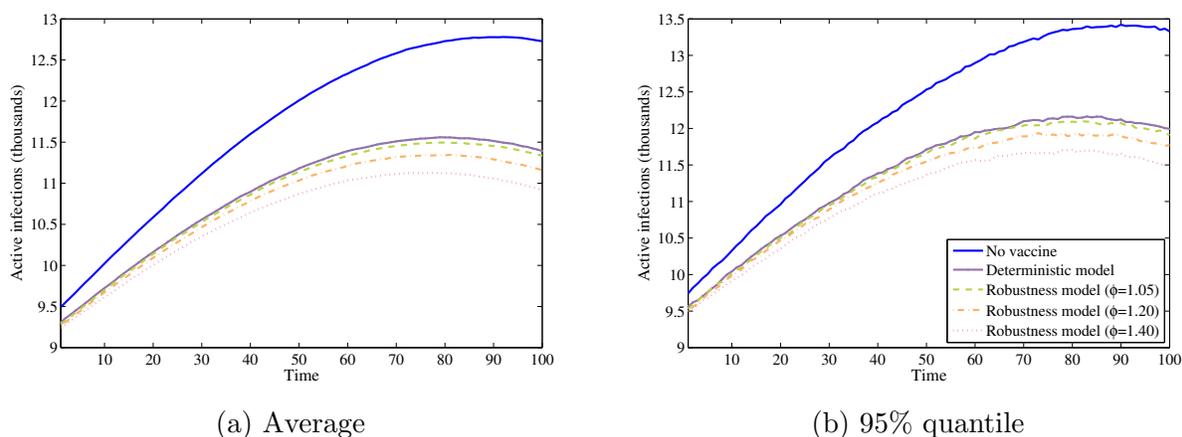


Figure 6 Long-term impact of the deterministic and robustness optimization models from time T to $T + 100$ under $B_t = 40$.

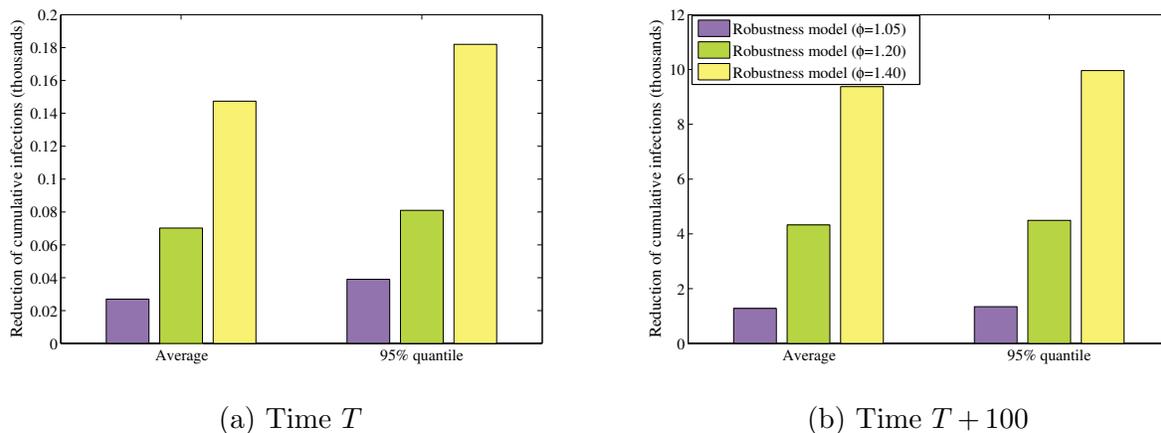


Figure 7 Reduction of the number of cumulative infections with respect to the deterministic model at time T and $T + 100$ under $B_t = 40$.

of planning horizon, time T . Specifically, we plot the probability that the number of active infections at time T exceeds the healthcare capacity by the range of values on the x-axis. For the same magnitude of violation threshold, the robustness optimization models achieve lower probabilities of violation. From the magnitude of the violation probability, we see that the deterministic model would severely underestimate the risks in disease propagation and use only an insufficient amount of vaccines, which eventually leads to high risks of violating the healthcare capacity. This indicates the inefficacy of minimizing the total vaccination cost in the deterministic model. We believe it is more practical to use the robustness optimization model, which specifies an acceptable cost target and finds the most robust solution to managing the healthcare capacity.

Impact of misspecification. In above analysis, we assumed that there is no misspecification of model parameters, *e.g.*, the true transmission rate β that generates the data is the same as

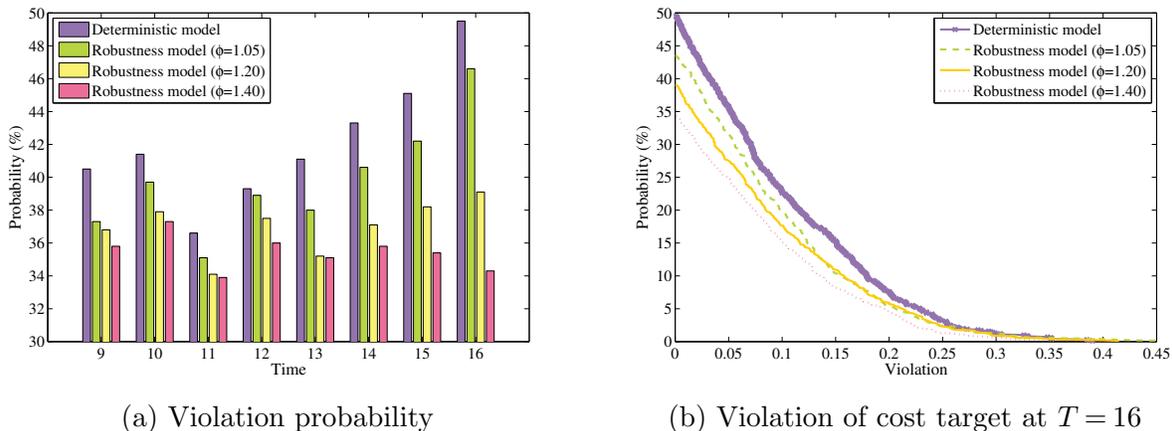


Figure 8 Model robustness to satisfying the healthcare capacity under $B_t = 20$.

the one used in the model. In practice, the transmission rate cannot be estimated accurately with limited and potentially non-stationary data. Here, we investigate the model performance under misspecification of transmission rate. Specifically, the models assume a nominal value of transmission rate, $\hat{\beta}$; however, the out-of-sample test data is generated with the true transmission rate, $\bar{\beta}$. The true transmission rate is set as $\bar{\beta} = \delta\hat{\beta}$, where δ varies from 0.7 to 1.3. When $\delta = 1$, the model has no misspecification of transmission rates.

In Figure 9, we evaluate the reduction of end-of-horizon active infections at time T with respect to the no-vaccine policy. In Figure 9(a), we focus on the reduction of the average number of end-of-horizon active infections. We see that robustness optimization models always lead to a higher reduction and thus fewer average active infections. Figure 9(b) plots the reduction of the 95% quantile of the number of end-of-horizon active infections. The observations are consistent.

Similar to our discussion of Figure 5, we remark that the seemingly small improvement over the deterministic model would magnify over time even without any additional cost and intervention, due to the inherent nature that the effect of vaccination is a long-term one. In addition, under adversarial transmission rates, *e.g.*, $\delta \in \{1.1, 1.2, 1.3\}$, the improvement would magnify more aggressively over time. As an illustration, we provide a counterpart of Figure 7. In Figure 10, we show how the seemingly small reduction of the cumulative infections over the deterministic model at time T would magnify in an additional period of 100 days. Specifically, when $\delta = 1.3$, the long-term effect of implementing our robustness optimization in only 16 days can lead to a reduction of nearly 12,000 cumulative infections in 100 days at no additional cost and intervention.

Next, we investigate the case where the true transmission rate, $\bar{\beta}$, is drawn from a uniform distribution, $\text{Uniform}(\underline{\delta}\hat{\beta}, \bar{\delta}\hat{\beta})$. Specifically, when generating the sample paths, we first sample a transmission rate from this uniform distribution. We test the model under two broad scenarios. In the first scenario, the nominal transmission rate $\hat{\beta}$ used in the model is unbiased, *i.e.*, $(\underline{\delta} + \bar{\delta})/2 = 1$.

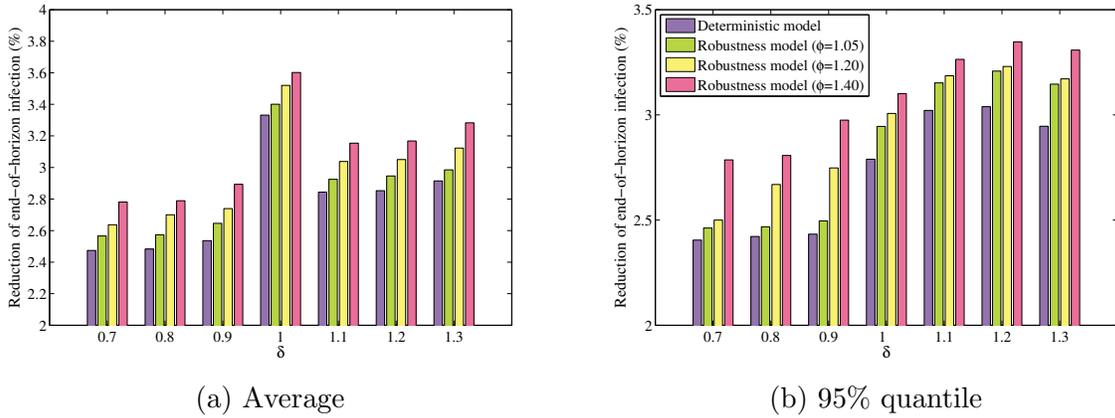


Figure 9 Effect of misspecification of transmission rate under $B_t = 20$ – Reduction of end-of-horizon infection with respect to the no-vaccine policy under non-stochastic transmission rate.

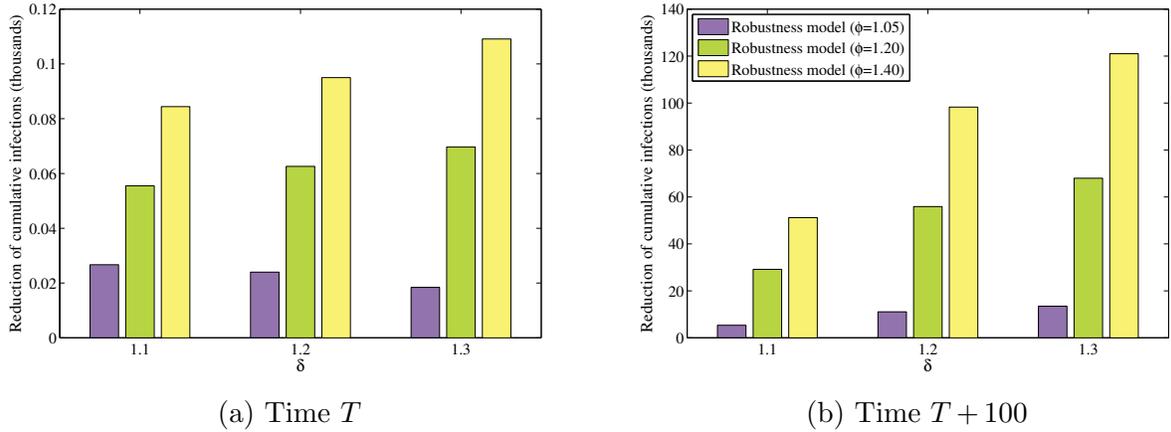


Figure 10 Reduction of the average number of cumulative infections with respect to the deterministic model at time T and $T + 100$ under adversarial transmission rates.

In the second scenario, the nominal transmission rate $\hat{\beta}$ is biased (consistently underestimate the true transmission rate), *i.e.*, $\bar{\delta} > \underline{\delta} = 1$. The performance summary of six instances is plotted in Figure 11. As we can see, the robustness optimization model with only 5% increase in the vaccination cost would help to better reduce the number of end-of-horizon active infections compared to the deterministic optimization model in all instances, indicating that our model improvement is still consistent under model misspecification.

Illustration of vaccine allocation policy. Finally, we provide an illustration on the impact of interaction rates to the vaccine allocation policy. The goal is to verify that the optimal solution of our model is reasonable. Besides the nominal interaction rate listed in Table 1, we also consider three particular settings: (i) High intra-interaction (we increase the value of $\hat{w}_{j,j}$ by 50%, for all $j \in [J]$), (ii) high inter-interaction (we increase the value of $\hat{w}_{j,k}$ by 50%, for all $j, k \in [J], j \neq k$),

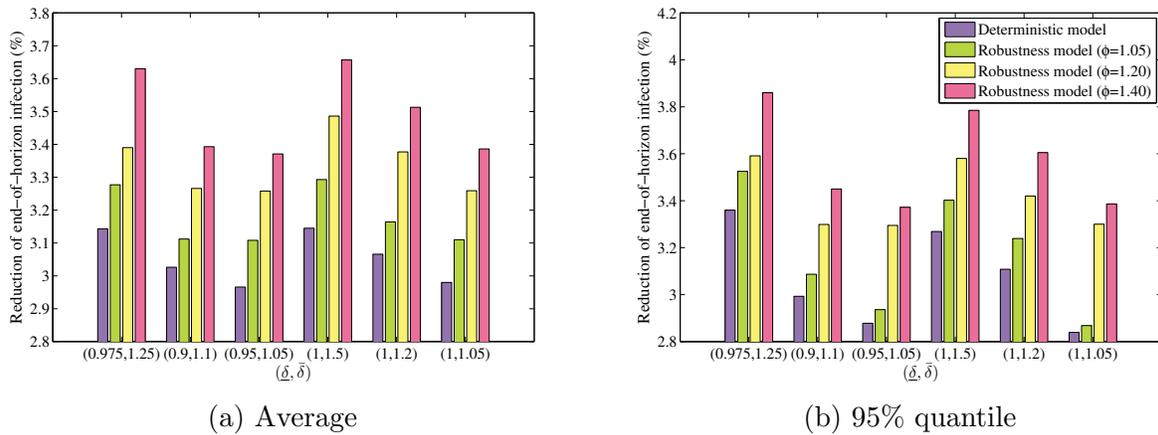


Figure 11 Effect of misspecification of random transmission rate under $B_t = 20$ – Reduction of end-of-horizon infection with respect to the no-vaccine policy under stochastic transmission rate.

and (iii) low interaction rate for the elderly (we reduce the interaction rate of people aged over 65, $\hat{w}_{1,j}$, by 50%, for all $j \in [J]$).

The optimal vaccine allocation policies are summarized in Figure 12. We remark that the corresponding changes in the vaccine allocation policies are reasonable. In Figure 12(b), we allocate more vaccines to the group aged 15-34. When intra-interaction rate increases, the contraction rate would increase more significantly in this group because it has the highest transmission rate. Hence, it is reasonable to prioritize vaccination in this group. In Figure 12(c), we allocate more vaccines to the elderly. This is because the high inter-interaction rate would expose groups with low transmission rates (*e.g.*, the elderly) to more dangers. Finally, in Figure 12(d), we reduce the vaccination to the elderly significantly because they now have a low level of interaction and are thus considerably safe from the disease.

References

- Abbott, Sam, Joel Hellewell, Robin N Thompson, Katharine Sherratt, Hamish P Gibbs, Nikos I Bosse, James D Munday, Sophie Meakin, Emma L Doughty, June Young Chun, et al. 2020. Estimating the time-varying reproduction number of sars-cov-2 using national and subnational case counts. *Wellcome Open Research* **5**(112) 112.
- Aumann, Robert J, Roberto. Serrano. 2008. An economic index of riskiness. *Journal of Political Economy* **116**(5).
- Bertsimas, Dimitris, Leonard Boussiou, Ryan Cory-Wright, Arthur Delarue, Vassilis Digalakis, Alexandre Jacquillat, Driss Lahlou Kitane, Galit Lukin, Michael Li, Luca Mingardi, et al. 2021a. From predictions to prescriptions: A data-driven response to covid-19. *Health care management science* 1–20.
- Bertsimas, Dimitris, Vassilis Digalakis Jr, Alexander Jacquillat, Michael Lingzhi Li, Alessandro Previero. 2021b. Where to locate covid-19 mass vaccination facilities? *arXiv preprint arXiv:2102.07309* .

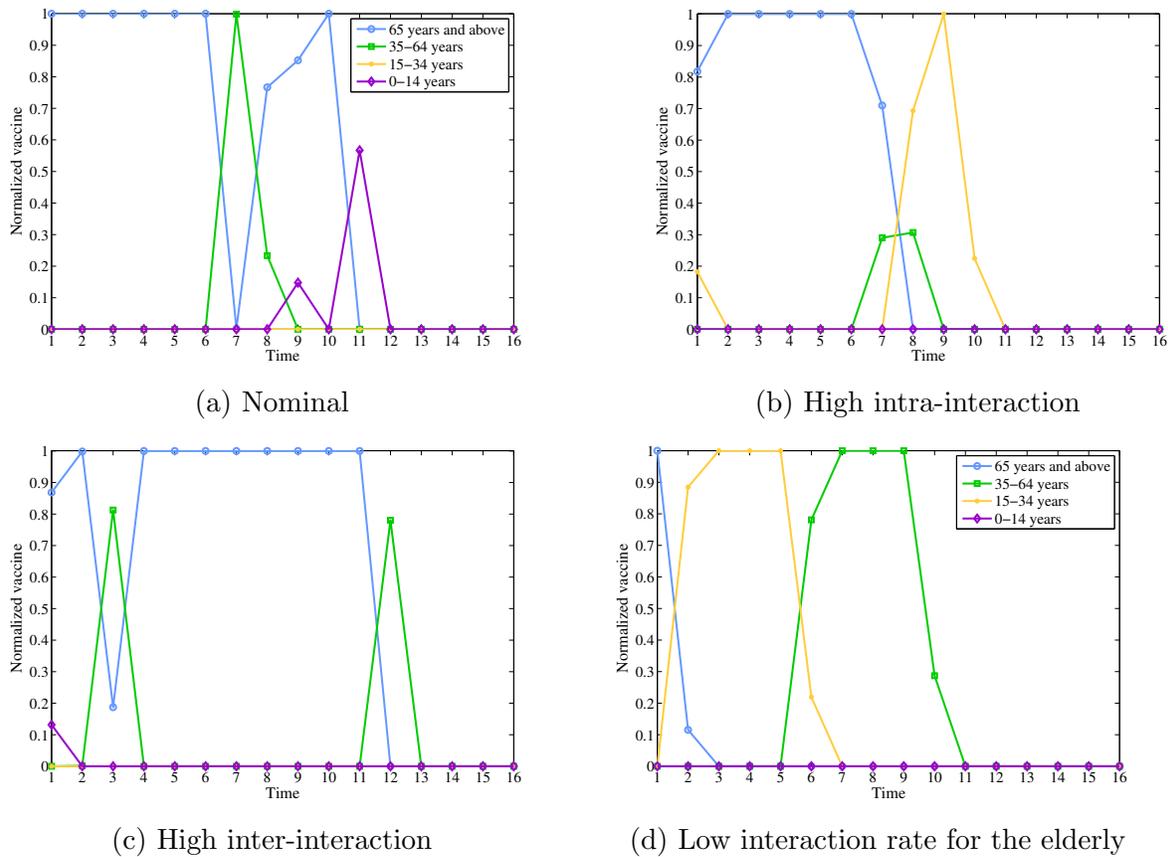


Figure 12 Vaccine distribution policies of the Robustness SVEIR model under $B_t = 20$ and $\phi = 1.02$

Birge, John R, Ozan Candogan, Yiding Feng. 2020. Controlling epidemic spread: reducing economic losses with targeted closures. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-57).

Brown, David B, Melvyn Sim. 2009. Satisficing measures for analysis of risky positions. *Management Science* **55**(1) 71–84.

Büyüktaktakın, İ Esra, Emmanuel des Bordes, Eyyüb Y Kıbış. 2018. A new epidemics–logistics model: Insights into controlling the ebola virus disease in west africa. *European Journal of Operational Research* **265**(3) 1046–1063.

Chang, Serina, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, Jure Leskovec. 2021. Mobility network models of covid-19 explain inequities and inform reopening. *Nature* **589**(7840) 82–87.

Duijzer, Lotty E, Willem L van Jaarsveld, Jacco Wallinga, Rommert Dekker. 2018a. Dose-optimal vaccine allocation over multiple populations. *Production and operations management* **27**(1) 143–159.

Duijzer, Lotty Evertje, Willem van Jaarsveld, Rommert Dekker. 2018b. Literature review: The vaccine supply chain. *European Journal of Operational Research* **268**(1) 174–192.

-
- Enayati, Shakiba, Osman Y Özaltın. 2020. Optimal influenza vaccine distribution with equity. *European Journal of Operational Research* **283**(2) 714–725.
- ESCAP, UN, Universal Postal Union UPU, World Health Organization (WHO and others). 2020. How covid-19 is changing the world: A statistical perspective .
- Föllmer, Hans, Alexander Schied. 2002. Convex measures of risk and trading constraints. *Finance and stochastics* **6**(4) 429–447.
- Hill, Andrew N, Ira M Longini Jr. 2003. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical biosciences* **181**(1) 85–106.
- Isermann, H. 1982. Linear lexicographic optimization. *Operations-Research-Spektrum* **4**(4) 223–228.
- Kermack, William Ogilvy, Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115**(772) 700–721.
- Lekone, Pheny E, Bärbel F Finkenstädt. 2006. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics* **62**(4) 1170–1177.
- Li, Ruiyun, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, Jeffrey Shaman. 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science* **368**(6490) 489–493.
- Liu, Ming, Ding Zhang. 2016. A dynamic logistics model for medical resources allocation in an epidemic control with demand forecast updating. *Journal of the Operational Research Society* **67**(6) 841–852.
- Liu, Qun, Daqing Jiang, Ningzhong Shi, Tasawar Hayat, Bashir Ahmad. 2017. Stationary distribution and extinction of a stochastic seir epidemic model with standard incidence. *Physica A: Statistical Mechanics and its Applications* **476** 58–69.
- Long, Daniel Zhuoyu, Melvyn Sim, Minglong Zhou. 2021. The dao of robustness: Achieving robustness in prescriptive analytics URL <https://ssrn.com/abstract=3478930>.
- Long, Elisa F, Eike Nohdurft, Stefan Spinler. 2018. Spatial resource allocation for emerging epidemics: A comparison of greedy, myopic, and dynamic policies. *Manufacturing & Service Operations Management* **20**(2) 181–198.
- Mak, Ho-Yin, Tinglong Dai, Christopher S Tang. 2021. Managing two-dose covid-19 vaccine rollouts with limited supply URL <https://ssrn.com/abstract=3790836>.
- NYC Health. 2020. Covid-19: Data. <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>.
- Qi, Jin. 2017. Mitigating delays and unfairness in appointment systems. *Management Science* **63**(2) 566–583.
- Ross, Ronald. 1916. An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character* **92**(638) 204–230.

-
- Smith, James E, Robert L Winkler. 2006. The optimizer’s curse: skepticism and postdecision surprise in decision analysis. *Management Science* **52**(3) 311–322.
- Sun, Li, Gail W DePuy, Gerald W Evans. 2014. Multi-objective optimization models for patient allocation during a pandemic influenza outbreak. *Computers & Operations Research* **51** 350–359.
- Sun, Peng, Liu Yang, Francis De Véricourt. 2009. Selfish drug allocation for containing an international influenza pandemic at the onset. *Operations Research* **57**(6) 1320–1332.
- Tanner, Matthew W, Lewis Ntaimo. 2010. Iis branch-and-cut for joint chance-constrained stochastic programs and application to optimal vaccine allocation. *European Journal of Operational Research* **207**(1) 290–296.
- US Census Bureau. 2018. The american community survey. <https://www.census.gov/programs-surveys/acs/>.
- Yarmand, Hamed, Julie S Ivy, Brian Denton, Alun L Lloyd. 2014. Optimal two-phase vaccine allocation to geographically different regions under uncertainty. *European Journal of Operational Research* **233**(1) 208–219.
- Zhang, Xianghua, Ke Wang. 2014. Stochastic seir model with jumps. *Applied Mathematics and Computation* **239** 133–143.
- Zhou, Minglong, Gar Goei Loke, Chaithanya Bandi, Zi Qiang Glen Liao, Wilson Wang. 2021. Intraday scheduling with patient re-entries and variability in behaviours. *Manufacturing & Service Operations Management* .

A. Proof of Results

Proof of Theorem 1. The first result has been observed in Long et al. (2021) using the fact (see, e.g. Föllmer and Schied 2002) that the certainty equivalent under the exponential disutility can be expressed as

$$\mu_\theta [\tilde{v}] = \sup_{\mathbb{Q} \in \hat{\mathcal{P}}} \{ \mathbb{E}_{\mathbb{Q}} [\tilde{v}] - \theta \phi_{KL}(\mathbb{Q} || \hat{\mathbb{P}}) \}.$$

where $\hat{\mathcal{P}}$ is the set of all probability distributions that are absolutely continuous with respect to $\hat{\mathbb{P}}$. Observe that with $\theta = \mu_\theta [\tilde{v}] > 0$, we have

$$\begin{aligned} & \mu_\theta [\tilde{v}] \leq 0 \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \hat{\mathbb{P}}) & \forall \mathbb{Q} \in \hat{\mathcal{P}} \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \mathbb{P}^*) + \theta (\phi_{KL}(\mathbb{Q} || \hat{\mathbb{P}}) - \phi_{KL}(\mathbb{Q} || \mathbb{P}^*)) & \forall \mathbb{Q} \in \hat{\mathcal{P}} \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \mathbb{P}^*) + \theta \sup_{\mathbb{Q}^\dagger \in \hat{\mathcal{P}}} \{ (\phi_{KL}(\mathbb{Q}^\dagger || \hat{\mathbb{P}}) - \phi_{KL}(\mathbb{Q}^\dagger || \mathbb{P}^*)) \} & \forall \mathbb{Q} \in \hat{\mathcal{P}} \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \mathbb{P}^*) + \theta \sup_{\mathbb{Q}^\dagger \in \mathcal{P}^*} \{ (\phi_{KL}(\mathbb{Q}^\dagger || \hat{\mathbb{P}}) - \phi_{KL}(\mathbb{Q}^\dagger || \mathbb{P}^*)) \} & \forall \mathbb{Q} \in \hat{\mathcal{P}} \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \mathbb{P}^*) + \theta \sup_{\mathbb{Q}^\dagger \in \mathcal{P}^*} \left\{ \sum_{v \in \mathcal{V}: \mathbb{P}^*[\tilde{v}=v] > 0} \left(\mathbb{Q}[\tilde{v}=v] \log \left(\frac{\mathbb{P}^*[\tilde{v}=v]}{\hat{\mathbb{P}}[\tilde{v}=v]} \right) \right) \right\} & \forall \mathbb{Q} \in \hat{\mathcal{P}} \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \mathbb{P}^*) + \theta \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) & \forall \mathbb{Q} \in \hat{\mathcal{P}} \\ \implies & \mathbb{E}_{\mathbb{Q}} [\tilde{v}] \leq \theta \phi_{KL}(\mathbb{Q} || \mathbb{P}^*) + \theta \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) & \forall \mathbb{Q} \in \mathcal{P}^* \\ \implies & \theta \log \mathbb{E}_{\mathbb{P}^*} [\exp(\tilde{v}/\theta)] \leq \theta \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) \\ \implies & \mathbb{E}_{\mathbb{P}^*} \left[\exp \left(\left(\tilde{v} - \theta \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) \right) / \theta \right) \right] \leq 1, \end{aligned}$$

where \mathcal{P}^* is the set of all probability distributions that are absolutely continuous with respect to \mathbb{P}^* .

Therefore, from Chernoff bound, we have for all $a > 0$

$$\begin{aligned} \mathbb{P}^* \left[\tilde{v} > \theta \left(a + \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) \right) \right] &= \mathbb{P}^* \left[\exp \left(\left(\tilde{v} - \theta \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) \right) / \theta \right) > \exp(a) \right] \\ &\leq \mathbb{E}_{\mathbb{P}^*} \left[\exp \left(\left(\tilde{v} - \theta \Delta \left(\mathbb{P}^* || \hat{\mathbb{P}} \right) \right) / \theta \right) \right] \exp(-a) \\ &\leq \exp(-a). \end{aligned}$$

□

Proof of Theorem 2 By step function, the value of contraction rate, $\eta(v_{j,t})$ will belong to one interval $(\bar{v}_l, \bar{v}_{l+1}]$. We define binary variable $y_{j,t}^{l+1}$ equaling 1 if $\eta(v_{j,t})$ lies in level l in group j at time t (i.e., $\eta(v_{j,t}) \in (\bar{v}_l, \bar{v}_{l+1}]$). Hence, we have

$$\begin{aligned} \sum_{l \in [L]} y_{j,t}^l &= 1 & \forall j \in [J], t \in [T] \\ \eta(v_{j,t}) &\leq \bar{v}_1 + M(1 - y_{j,t}^1) & \forall j \in [J], t \in [T] \end{aligned}$$

$$\begin{aligned} \bar{v}_{l-1} - M(1 - y_{j,t}^l) &\leq \eta(v_{j,t}) \leq \bar{v}_l + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T], 2 \leq l \leq L \\ \eta(v_{j,t}) &\geq \bar{v}_L + M(1 - y_{j,t}^{L+1}) & \forall j \in [J], t \in [T]. \end{aligned}$$

Hence, in Problem (5), $\hat{q}_{j,t}$ can be restricted as the following functions in terms of $y_{j,t}^l$:

$$\hat{q}_{j,t} = \sum_{l \in [L]} \underline{v}_l y_{j,t}^l \quad \forall j \in [J], t \in [T-1]$$

Although the function of $\hat{p}_{j,t}$ is non-linear, it can be easily linearized because \mathbf{y} is binary variable. Here, we first try to linearize the generality function $z_{j,t} \hat{p}_{j,t}$ (e.g., $z_{j,t} = \bar{S}_{j,t} - \bar{U}_{j,t}$ in dynamics constraint involving $\bar{S}_{j,t}$, and $z_{j,t} = \hat{\omega}_j \bar{V}_{j,t}$ in dynamics constraint involving $\bar{V}_{j,t}$). We also define a new variable $a_{j,t}$, for all $j \in [J]$, $t \in [T]$ then we have

$$\begin{aligned} -M(1 - y_{j,t}^1) &\leq \bar{a}_{j,t} \leq M(1 - y_{j,t}^1) & \forall j \in [J], t \in [T-1] \\ z_{j,t} \underline{v}_{l-1} - M(1 - y_{j,t}^l) &\leq \bar{a}_{j,t} \leq z_{j,t} \underline{v}_{l-1} + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], 2 \leq l \leq L+1. \end{aligned}$$

Hence, Problem (5) can be reformulated as a mixed-integer linear optimization problem. For example, dynamics constraint involving $\bar{S}_{j,t}$ can be rewritten as the following set of constraints:

$$\begin{aligned} \bar{S}_{j,t+1} &= \bar{S}_{j,t} - \bar{U}_{j,t} - \bar{W}_{j,t}, & \forall j \in [J], t \in [T-1] \\ -M(1 - y_{j,t}^1) &\leq \bar{W}_{j,t} \leq M(1 - y_{j,t}^1) & \forall j \in [J], t \in [T-1] \\ (\bar{S}_{j,t} - \bar{U}_{j,t}) \underline{v}_{l-1} - M(1 - y_{j,t}^l) &\leq \bar{W}_{j,t} \leq (\bar{S}_{j,t} - \bar{U}_{j,t}) \underline{v}_{l-1} + M(1 - y_{j,t}^l) & \forall j \in [J], t \in [T-1], 2 \leq l \leq L+1. \end{aligned}$$

□

Proof of Proposition 1. Observe that

$$\begin{aligned} \mu_\theta [v \tilde{S}_{j,t+1}] &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \tilde{S}_{j,t+1} \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, 1 - \hat{q}_{j,t} \right) \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, 1 - \hat{q}_{j,t} \right) \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, 1 - \hat{q}_{j,t} \right) \right) \mid \tilde{S}_{j,t} \right] \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, 1 - \hat{q}_{j,t} \right) \right) \mid \tilde{S}_{j,t} \right] \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left((\tilde{S}_{j,t} - x_{j,t}) \kappa_\theta (1 - \hat{q}_{j,t}, v) / \theta \right) \right] \\ &= \mu_\theta \left[\tilde{S}_{j,t} \kappa_\theta (1 - \hat{q}_{j,t}, v) - x_{j,t} \kappa_\theta (1 - \hat{q}_{j,t}, v) \right]. \end{aligned}$$

In addition, we have

$$\begin{aligned} \mu_\theta [v \tilde{S}_{j,2}] &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(S_{j,1} - x_{j,1}, 1 - \hat{q}_{j,1} \right) \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left((S_{j,1} - x_{j,1}) \kappa_\theta (1 - \hat{q}_{j,1}, v) / \theta \right) \right] \\ &= \mu_\theta [S_{j,1} \kappa_\theta (1 - \hat{q}_{j,1}, v) - x_{j,1} \kappa_\theta (1 - \hat{q}_{j,1}, v)] \\ &= S_{j,1} \kappa_\theta (1 - \hat{q}_{j,1}, v) - x_{j,1} \kappa_\theta (1 - \hat{q}_{j,1}, v), \end{aligned}$$

where the last equality is due to the fact that $S_{j,1}$ is a known parameter. Hence, $\mu_\theta [v\tilde{S}_{j,t+1}]$ can be evaluated recursively.

We now focus on the vaccinated population, $\mu_\theta [v\tilde{V}_{j,t+1}]$, which, because of the assumption of independence, can be evaluated as follows:

$$\begin{aligned}\mu_\theta [v\tilde{V}_{j,t+1}] &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, 1 - \hat{\omega}_j \hat{q}_{j,t} \right) + \frac{v}{\theta} x_{j,t} \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, 1 - \hat{\omega}_j \hat{q}_{j,t} \right) \right) \right] + vx_{j,t}.\end{aligned}$$

By the same reasoning as above, the first term can be evaluated recursively as

$$\begin{aligned}& \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, 1 - \hat{\omega}_j \hat{q}_{j,t} \right) \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, 1 - \hat{\omega}_j \hat{q}_{j,t} \right) \right) \mid \tilde{V}_{j,t} \right] \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\tilde{V}_{j,t} \kappa_\theta (1 - \hat{\omega}_j \hat{q}_{j,t}, v) / \theta \right) \right] \\ &= \mu_\theta \left[\tilde{V}_{j,t} \kappa_\theta (1 - \hat{\omega}_j \hat{q}_{j,t}, v) \right]\end{aligned}$$

Hence, we have

$$\mu_\theta [v\tilde{V}_{j,t+1}] = \mu_\theta \left[\tilde{V}_{j,t} \kappa_\theta (1 - \hat{\omega}_j \hat{q}_{j,t}, v) \right] + vx_{j,t}.$$

Because $\mu_\theta (v\tilde{V}_{j,1}) = \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} V_{j,1} \right) \right] = vV_{j,1}$ is a constant parameter, the above can be evaluated recursively.

For the exposed individuals, $\mu_\theta [v\tilde{E}_{j,t+1}]$, we have

$$\begin{aligned}\mu_\theta [v\tilde{E}_{j,t+1}] &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, \hat{q}_{j,t} \right) \right) \right] + \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, \hat{\omega}_j \hat{q}_{j,t} \right) \right) \right] \\ &\quad + \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, 1 - \hat{\alpha}_j \right) \right) \right].\end{aligned}$$

By the same reasoning as above, the first term can be evaluated recursively as

$$\begin{aligned}& \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, \hat{q}_{j,t} \right) \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{S}_{j,t} - x_{j,t}, \hat{q}_{j,t} \right) \right) \mid \tilde{S}_{j,t} \right] \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left((\tilde{S}_{j,t} - x_{j,t}) \kappa_\theta (\hat{q}_{j,t}, v) / \theta \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\tilde{S}_{j,t} \kappa_\theta (\hat{q}_{j,t}, v) / \theta \right) \right] - x_{j,t} \kappa_\theta (\hat{q}_{j,t}, v) \\ &= \mu_\theta \left[\tilde{S}_{j,t} \kappa_\theta (\hat{q}_{j,t}, v) \right] - x_{j,t} \kappa_\theta (\hat{q}_{j,t}, v).\end{aligned}$$

The second term can be evaluated as:

$$\begin{aligned}& \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, \hat{\omega}_j \hat{q}_{j,t} \right) \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, \hat{\omega}_j \hat{q}_{j,t} \right) \right) \mid \tilde{V}_{j,t} \right] \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{V}_{j,t}, \hat{\omega}_j \hat{q}_{j,t} \right) \right) \mid \tilde{V}_{j,t} \right] \right) \right] \\ &= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\tilde{V}_{j,t} \kappa_\theta (\hat{\omega}_j \hat{q}_{j,t}, v) / \theta \right) \right] \\ &= \mu_\theta \left[\tilde{V}_{j,t} \kappa_\theta (\hat{\omega}_j \hat{q}_{j,t}, v) \right].\end{aligned}$$

The third term can be evaluated as:

$$\begin{aligned}
& \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, 1 - \hat{\alpha}_j \right) \right) \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, 1 - \hat{\alpha}_j \right) \right) \mid \tilde{E}_{j,t} \right] \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, 1 - \hat{\alpha}_j \right) \right) \mid \tilde{E}_{j,t} \right] \right) \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\tilde{E}_{j,t} \kappa_{\theta} (1 - \hat{\alpha}_j, v) / \theta \right) \right] \\
&= \mu_{\theta} \left[\tilde{E}_{j,t} \kappa_{\theta} (1 - \hat{\alpha}_j, v) \right].
\end{aligned}$$

Hence, we have

$$\mu_{\theta} \left[v \tilde{E}_{j,t+1} \right] = \mu_{\theta} \left[\tilde{S}_{j,t} \kappa_{\theta} (\hat{q}_{j,t}, v) \right] - x_{j,t} \kappa_{\theta} (\hat{q}_{j,t}, v) + \mu_{\theta} \left[\tilde{V}_{j,t} \kappa_{\theta} (\hat{\omega}_j \hat{q}_{j,t}, v) \right] + \mu_{\theta} \left[\tilde{E}_{j,t} \kappa_{\theta} (1 - \hat{\alpha}_j, v) \right].$$

Because $\mu_{\theta} \left[v \tilde{E}_{j,1} \right] = \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} E_{j,1} \right) \right] = v E_{j,1}$ is a constant parameter, the above can be evaluated recursively.

Now, we look at the infectious individuals, $\mu_{\theta} \left[v \tilde{I}_{j,t+1} \right]$:

$$\mu_{\theta} \left[v \tilde{I}_{j,t+1} \right] = \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, \hat{\alpha}_j \right) \right) \right] + \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{I}_{j,t}, 1 - \hat{\gamma}_j \right) \right) \right].$$

By the same reasoning as above, the first term can be evaluated as:

$$\begin{aligned}
& \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, \hat{\alpha}_j \right) \right) \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{E}_{j,t}, \hat{\alpha}_j \right) \right) \mid \tilde{E}_{j,t} \right] \right) \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\tilde{E}_{j,t} \kappa_{\theta} (\hat{\alpha}_j, v) / \theta \right) \right] \\
&= \mu_{\theta} \left[\tilde{E}_{j,t} \kappa_{\theta} (\hat{\alpha}_j, v) \right].
\end{aligned}$$

The second term can be evaluated as:

$$\begin{aligned}
& \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{I}_{j,t}, 1 - \hat{\gamma}_j \right) \right) \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{I}_{j,t}, 1 - \hat{\gamma}_j \right) \right) \mid \tilde{I}_{j,t} \right] \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{I}_{j,t}, 1 - \hat{\gamma}_j \right) \right) \mid \tilde{I}_{j,t} \right] \right) \right] \\
&= \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\tilde{I}_{j,t} \kappa_{\theta} (1 - \hat{\gamma}_j, v) / \theta \right) \right] \\
&= \mu_{\theta} \left[\tilde{I}_{j,t} \kappa_{\theta} (1 - \hat{\gamma}_j, v) \right].
\end{aligned}$$

Hence, we have

$$\mu_{\theta} \left[v \tilde{I}_{j,t+1} \right] = \mu_{\theta} \left[\tilde{E}_{j,t} \kappa_{\theta} (\hat{\alpha}_j, v) \right] + \mu_{\theta} \left[\tilde{I}_{j,t} \kappa_{\theta} (1 - \hat{\gamma}_j, v) \right].$$

Because $\mu_{\theta} \left[v \tilde{I}_{j,1} \right] = \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} I_{j,1} \right) \right] = v I_{j,1}$ is a constant parameter, the above can be evaluated recursively.

Finally, we look at the removed individuals, $\mu_{\theta} \left[v \tilde{R}_{j,t+1} \right]$:

$$\mu_{\theta} \left[v \tilde{R}_{j,t+1} \right] = \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \tilde{R}_{j,t} \right) \right] + \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} \text{Bin} \left(\tilde{I}_{j,t}, \hat{\gamma}_j \right) \right) \right].$$

By the same reasoning as above, the above can be written as:

$$\mu_\theta \left[v\tilde{R}_{j,t+1} \right] = \mu_\theta \left[v\tilde{R}_{j,t} \right] + \mu_\theta \left[\tilde{I}_{j,t} \kappa_\theta(\hat{\gamma}_j, v) \right].$$

Because $\mu_\theta \left[v\tilde{R}_{j,1} \right] = \theta \log \mathbb{E}_{\hat{\mathbb{P}}} \left[\exp \left(\frac{v}{\theta} R_{j,1} \right) \right] = vR_{j,1}$ is a constant parameter, the above can be evaluated recursively. \square

Proof of Proposition 2. We start with $R_{j,t}^\dagger$. By the independence assumption, we have

$$R_{j,t}^\dagger := \text{Var}_{\hat{\mathbb{P}}}[\tilde{R}_{j,t}] = \text{Var}_{\hat{\mathbb{P}}}[\tilde{R}_{j,t-1}] + \text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{I}_{j,t-1}, \hat{\gamma}_j)]$$

Now, by the law of total variance, we have

$$\begin{aligned} \text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{I}_{j,t-1}, \hat{\gamma}_j)] &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{I}_{j,t-1} \right] (1 - \hat{\gamma}_j) \hat{\gamma}_j + \hat{\gamma}_j^2 \text{Var}_{\hat{\mathbb{P}}}[\tilde{I}_{j,t-1}] \\ &= \bar{I}_{j,t-1} (1 - \hat{\gamma}_j) \hat{\gamma}_j + \hat{\gamma}_j^2 \text{Var}_{\hat{\mathbb{P}}}[\tilde{I}_{j,t-1}] \\ &= \bar{I}_{j,t-1} (1 - \hat{\gamma}_j) \hat{\gamma}_j + \hat{\gamma}_j^2 I_{j,t-1}^\dagger. \end{aligned}$$

Hence, we have

$$R_{j,t}^\dagger = R_{j,t-1}^\dagger + \bar{I}_{j,t-1} (1 - \hat{\gamma}_j) \hat{\gamma}_j + \hat{\gamma}_j^2 I_{j,t-1}^\dagger.$$

Then, we rewrite $I_{j,t}^\dagger$ as follows:

$$I_{j,t}^\dagger := \text{Var}_{\hat{\mathbb{P}}}[\tilde{I}_{j,t}] = \text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{I}_{j,t-1}, 1 - \hat{\gamma}_j)] + \text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{E}_{j,t-1}, \hat{\alpha}_j)].$$

By the law of total variance, we have

$$\begin{aligned} \text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{I}_{j,t-1}, 1 - \hat{\gamma}_j)] &= \mathbb{E}_{\hat{\mathbb{P}}} \left[\tilde{I}_{j,t-1} \right] (1 - \hat{\gamma}_j) \hat{\gamma}_j + (1 - \hat{\gamma}_j)^2 \text{Var}_{\hat{\mathbb{P}}}[\tilde{I}_{j,t-1}] \\ &= \bar{I}_{j,t-1} (1 - \hat{\gamma}_j) \hat{\gamma}_j + (1 - \hat{\gamma}_j)^2 \text{Var}_{\hat{\mathbb{P}}}[\tilde{I}_{j,t-1}] \\ &= \bar{I}_{j,t-1} (1 - \hat{\gamma}_j) \hat{\gamma}_j + (1 - \hat{\gamma}_j)^2 I_{j,t-1}^\dagger. \end{aligned}$$

Now, we focus on $\text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{E}_{j,t-1}, \hat{\alpha}_j)]$. By the law of total variance, we have

$$\text{Var}_{\hat{\mathbb{P}}}[\text{Bin}(\tilde{E}_{j,t-1}, \hat{\alpha}_j)] = \bar{E}_{j,t-1} \hat{\alpha}_j (1 - \hat{\alpha}_j) + \hat{\alpha}_j^2 \text{Var}_{\hat{\mathbb{P}}}[\tilde{E}_{j,t-1}] = \bar{E}_{j,t-1} \hat{\alpha}_j (1 - \hat{\alpha}_j) + \hat{\alpha}_j^2 E_{j,t-1}^\dagger.$$

Hence, the variance term $I_{j,t}^\dagger$ can be written as:

$$I_{j,t}^\dagger = \bar{I}_{j,t-1} (1 - \hat{\gamma}_j) \hat{\gamma}_j + (1 - \hat{\gamma}_j)^2 I_{j,t-1}^\dagger + \bar{E}_{j,t-1} \hat{\alpha}_j (1 - \hat{\alpha}_j) + \hat{\alpha}_j^2 E_{j,t-1}^\dagger.$$

Now, we derive the recursive equation for $E_{j,t}^\dagger$. By independence,

$$\begin{aligned} &E_{j,t}^\dagger \\ &= \text{Var}_{\hat{\mathbb{P}}} \left[\text{Bin} \left(\tilde{S}_{j,t-1} - x_{j,t-1}, \hat{q}_{j,t-1} \right) \right] + \text{Var}_{\hat{\mathbb{P}}} \left[\text{Bin} \left(\tilde{V}_{j,t-1}, \hat{\omega}_j \hat{q}_{j,t-1} \right) \right] + \text{Var}_{\hat{\mathbb{P}}} \left[\text{Bin} \left(\tilde{E}_{j,t-1}, 1 - \hat{\alpha}_j \right) \right] \\ &= (\bar{S}_{j,t-1} - x_{j,t-1}) \hat{q}_{j,t-1} (1 - \hat{q}_{j,t-1}) + \hat{q}_{j,t-1}^2 S_{j,t-1}^\dagger \\ &\quad + \bar{V}_{j,t-1} \hat{\omega}_j \hat{q}_{j,t-1} (1 - \hat{\omega}_j \hat{q}_{j,t-1}) + (\hat{\omega}_j \hat{q}_{j,t-1})^2 V_{j,t-1}^\dagger \\ &\quad + \bar{E}_{j,t-1} (1 - \hat{\alpha}_j) \hat{\alpha}_j + (1 - \hat{\alpha}_j)^2 E_{j,t-1}^\dagger. \end{aligned}$$

It remains to derive the recursive equations for $S_{j,t}^\dagger$ and $V_{j,t}^\dagger$ as follows:

$$\begin{aligned} S_{j,t}^\dagger &:= \text{Var}_{\hat{\mathbb{P}}} \left[\tilde{S}_{j,t} \right] = \text{Var}_{\hat{\mathbb{P}}} \left[\text{Bin} \left(\tilde{S}_{j,t-1} - x_{j,t-1}, 1 - \hat{q}_{j,t-1} \right) \right] \\ &= (\bar{S}_{j,t-1} - x_{j,t-1}) \hat{q}_{j,t-1} (1 - \hat{q}_{j,t-1}) + (1 - \hat{q}_{j,t-1})^2 S_{j,t-1}^\dagger \\ V_{j,t}^\dagger &:= \text{Var}_{\hat{\mathbb{P}}} \left[\tilde{V}_{j,t} \right] = \text{Var}_{\hat{\mathbb{P}}} \left[\text{Bin} \left(\tilde{V}_{j,t-1}, 1 - \hat{\omega}_j \hat{q}_{j,t-1} \right) \right] \\ &= \bar{V}_{j,t-1} \hat{\omega}_j \hat{q}_{j,t-1} (1 - \hat{\omega}_j \hat{q}_{j,t-1}) + (1 - \hat{\omega}_j \hat{q}_{j,t-1})^2 V_{j,t-1}^\dagger \end{aligned}$$

Given that $\tilde{S}_{j,1}$, $\tilde{V}_{j,1}$, $\tilde{E}_{j,1}$, and $\tilde{I}_{j,1}$ are known, constant parameters, we can calculate $S_{j,t}^\dagger$, $V_{j,t}^\dagger$, $E_{j,t}^\dagger$, $I_{j,t}^\dagger$, and $R_{j,t}^\dagger$ via the above system of recursive equations. \square

Proof of Theorem 3. The constraint $\sum_{i \in [J]} \nu_\theta \left[\tilde{I}_{j,t} \right] \leq H_t$ can be written as

$$\sum_{i \in [J]} \left(\bar{I}_{j,t} + \frac{1}{2\theta} I_{j,t}^\dagger \right) \leq H_t,$$

where $I_{j,t}^\dagger$ follows the recursive dynamics as described in (10). We need to show that the system dynamics in (10) can be described by mixed-integer linear constraints. Hence, we only need to focus on the recursions of $S_{j,t}^\dagger$, $V_{j,t}^\dagger$, $E_{j,t}^\dagger$.

Note that

$$\hat{q}_{j,t} := \eta \left(\sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \right).$$

Without loss of generality, let $\underline{v}_0 := 0$, $\bar{v}_0 := 0$ and $\bar{v}_{L+1} := 1$. By the definition of the step function η , for any $j \in [J]$ and $t \in [T]$, we can write

$$\begin{aligned} \hat{q}_{j,t} &= \sum_{l \in [0,L]} \underline{v}_l y_{j,t}^l \\ \sum_{l \in [0,L]} y_{j,t}^l &= 1 \\ \bar{v}_l - M(1 - y_{j,t}^l) &\leq \sum_{k \in [J]} \hat{\zeta}_{k,j} \bar{I}_{k,t} \leq \bar{v}_{l+1} + M(1 - y_{j,t}^l) \quad \forall l \in [0, L] \\ y_{j,t}^l &\in \{0, 1\} \quad \forall l \in [0, L] \end{aligned}$$

In other words, we can represent $\hat{q}_{j,t}$ with $\sum_{l \in [0,L]} \underline{v}_l y_{j,t}^l$ where \mathbf{y} is subject to some additional mixed-integer linear constraints. With such a construction, for any functional f , note that

$$f(\hat{q}_{j,t}) = \sum_{l \in [0,L]} f(\underline{v}_l) y_{j,t}^l.$$

Then, we can equivalently evaluate $S_{j,t}^\dagger$ as:

$$\begin{aligned} S_{j,t}^\dagger &= (\bar{S}_{j,t-1} - x_{j,t-1}) \hat{q}_{j,t-1} (1 - \hat{q}_{j,t-1}) + (1 - \hat{q}_{j,t-1})^2 S_{j,t-1}^\dagger \\ &= \sum_{l \in [0,L]} (\bar{S}_{j,t-1} - x_{j,t-1}) \underline{v}_l (1 - \underline{v}_l) y_{j,t-1}^l + \sum_{l \in [0,L]} (1 - \underline{v}_l)^2 y_{j,t-1}^l S_{j,t-1}^\dagger \\ &= \sum_{l \in [0,L]} \left((\bar{S}_{j,t-1} - x_{j,t-1}) \underline{v}_l (1 - \underline{v}_l) + (1 - \underline{v}_l)^2 S_{j,t-1}^\dagger \right) y_{j,t-1}^l. \end{aligned}$$

We can equivalently reformulate the above relation using the big-M method as follows:

$$\begin{aligned} S_{j,t}^\dagger &\leq (\bar{S}_{j,t-1} - x_{j,t-1})\underline{v}_l(1 - \underline{v}_l) + (1 - \underline{v}_l)^2 S_{j,t-1}^\dagger + M(1 - y_{j,t-1}^l) \quad \forall l \in [0, L] \\ S_{j,t}^\dagger &\geq (\bar{S}_{j,t-1} - x_{j,t-1})\underline{v}_l(1 - \underline{v}_l) + (1 - \underline{v}_l)^2 S_{j,t-1}^\dagger - M(1 - y_{j,t-1}^l) \quad \forall l \in [0, L]. \end{aligned}$$

Similarly, we can reformulate the dynamics of $V_{j,t}^\dagger$ as:

$$\begin{aligned} V_{j,t}^\dagger &\leq \bar{V}_{j,t-1}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (1 - \hat{\omega}_j\underline{v}_l)^2 V_{j,t-1}^\dagger + M(1 - y_{j,t-1}^l) \quad \forall l \in [0, L] \\ V_{j,t}^\dagger &\geq \bar{V}_{j,t-1}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (1 - \hat{\omega}_j\underline{v}_l)^2 V_{j,t-1}^\dagger - M(1 - y_{j,t-1}^l) \quad \forall l \in [0, L], \end{aligned}$$

and the dynamics of $E_{j,t}^\dagger$ as:

$$\begin{aligned} E_{j,t}^\dagger &\leq \bar{E}_{j,t-1}(1 - \hat{\alpha}_j)\hat{\alpha}_j + (1 - \hat{\alpha}_j)^2 E_{j,t-1}^\dagger \\ &\quad + (\bar{S}_{j,t-1} - x_{j,t-1})\underline{v}_l(1 - \underline{v}_l) + \underline{v}_l^2 S_{j,t-1}^\dagger \\ &\quad + \bar{V}_{j,t-1}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (\hat{\omega}_j\underline{v}_l)^2 V_{j,t-1}^\dagger + M(1 - y_{j,t-1}^l) \quad \forall l \in [0, L] \\ E_{j,t}^\dagger &\geq \bar{E}_{j,t-1}(1 - \hat{\alpha}_j)\hat{\alpha}_j + (1 - \hat{\alpha}_j)^2 E_{j,t-1}^\dagger \\ &\quad + (\bar{S}_{j,t-1} - x_{j,t-1})\underline{v}_l(1 - \underline{v}_l) + \underline{v}_l^2 S_{j,t-1}^\dagger \\ &\quad + \bar{V}_{j,t-1}\hat{\omega}_j\underline{v}_l(1 - \hat{\omega}_j\underline{v}_l) + (\hat{\omega}_j\underline{v}_l)^2 V_{j,t-1}^\dagger - M(1 - y_{j,t-1}^l) \quad \forall l \in [0, L]. \end{aligned}$$

□

Proof of Theorem 4. This follows from Theorem 3 and the proof of Theorem 2

□