Finding optimal vaccination strategies under parameter uncertainty using stochastic programming

Matthew W. Tanner\textsuperscript{a,}\textsuperscript{*}, Lisa Sattenspiel\textsuperscript{b}, Lewis Ntaimo\textsuperscript{a}

\textsuperscript{a} Department of Industrial and Systems Engineering, Texas A&M University, 241 Zachry, 3131 TAMU, College Station, TX 77843-3131, USA

\textsuperscript{b} Department of Anthropology, University of Missouri, 105 Swallow Hall, Columbia, MO 65211-1440, USA

\textsuperscript{*} Corresponding author. Tel.: +1 573 268 4019.
E-mail address: mtanner@tamu.edu (M.W. Tanner).

\textbf{Abstract}

We present a stochastic programming framework for finding the optimal vaccination policy for controlling infectious disease epidemics under parameter uncertainty. Stochastic programming is a popular framework for including the effects of parameter uncertainty in a mathematical optimization model. The problem is initially formulated to find the minimum cost vaccination policy under a chance-constraint. The chance-constraint requires that the probability that $R_e$ $\leq$ $1$ be greater than some parameter $\epsilon$, where $R_e$ is the post-vaccination reproduction number. We also show how to formulate the problem in two additional cases: (a) finding the optimal vaccination policy when vaccine supply is limited and (b) a cost–benefit scenario. The class of epidemic models for which this method can be used is described and we present an example formulation for which the resulting problem is a mixed-integer program. A short numerical example based on plausible parameter values and distributions is given to illustrate how including parameter uncertainty improves the robustness of the optimal strategy at the cost of higher coverage of the population. Results derived from a stochastic programming analysis can also help to guide decisions about how much effort and resources to focus on collecting data needed to provide better estimates of key parameters.

© 2008 Elsevier Inc. All rights reserved.

\textbf{1. Introduction}

Vaccination is one of the primary strategies used by public health authorities to control human infectious diseases. Mathematical models have long played a major role in identifying and evaluating strategies to allocate resources in order to guarantee maximum effectiveness of vaccination in controlling infectious disease outbreaks. Three primary modeling approaches have been used in this effort – deterministic analytical models, stochastic analytical models, and computer simulations. The determination of optimal vaccination strategies may be sensitive to changes in model parameter values, however, so there is a need for new methods that can take parameter uncertainty into account in order to find more robust vaccination policies. We present here a description of one such method, stochastic programming, and illustrate how this method can improve our ability to find optimal vaccination strategies.

The goal of most deterministic and stochastic epidemiological models addressing vaccination strategies is to derive appropriate strategies analytically. Deterministic models focused on identifying reasonable vaccination strategies for the control of infectious diseases date back to at least the 1960s [early papers include, for example, [11,23,36,40]]. In general, deterministic vaccination models fall into two major groups. The majority of these models are used to evaluate predetermined vaccination strategies to see which of the proposed strategies may be most effective. Analysis of most of these models generally involves exploration of the steady state behavior of the model system and determination of an epidemic threshold. The effectiveness of different proposed vaccination strategies in reducing the susceptible population below the epidemic threshold for the minimum cost is then evaluated. In some of the recent more complex models, computer simulation is used to assess the effectiveness of different strategies. Models of this type have been developed for a number of infectious diseases, including tuberculosis [11,40], measles [13,20,37], rubella [213,19,27], pertussis [18,21,22,24], and respiratory illnesses [34].

The second group of deterministic vaccination models do not start with predetermined strategies; rather, they center on the use of optimization methods in combination with deterministic epidemic models to identify the optimal vaccination strategy. Optimization methods have been used both in a theoretical framework [23] and to guide the development of vaccination policies for specific diseases, including tuberculosis [36], influenza [28], and smallpox [16].
A number of stochastic models have also been developed to determine optimal vaccination strategies. For example, Ball et al. [4] develop an SIR epidemic model with both local mixing at the household level and global mixing at the community level. They introduce the notation, $R_s$, to represent the threshold parameter for a community of households. They analyze the case of a perfect vaccine and show that under this condition, a strategy that allocates vaccines to those households with the largest number of unvaccinated individuals is best for reducing $R_s$ to a level that will control an epidemic. Becker and Starczak [9] study vaccination policies in a stochastic SIR model divided into a community of households. They derive a closed form equation for the post-vaccination reproductive number, $R_v$, then formulate and numerically solve a linear program to find the minimum vaccination coverage under the constraint $R_v < 1$. This constraint ensures that the disease will tend to die out. (Becker and Starczak [9] use the notation $R_{IH}$ rather than $R_v$, but the concepts are equivalent.) Drawing upon the earlier work of Ball et al. [4], Ball and Lyne [6] consider the case of an all-or-nothing vaccine where a person is either totally immune following vaccination or the vaccine does not work at all. They show that if the sequence $(n_0, \mu_0, \ldots, \mu_l)$ is convex, where $\mu_i$ is the mean size of a local outbreak within a household of size $n_i$, then the optimal solution to the linear programming problem formulation of Becker and Starczak [9] can be characterized explicitly. Ball et al. [5] use the model described by Ball and Lyne [6] to address the question of optimal allocation of vaccines. They show that an explicit characterization of the optimal vaccination strategy is only possible in certain special cases, such as proportionate mixing. Müller [32] uses an SIRS epidemic model to derive optimal vaccination strategies in an age structured population and compares the conditions needed for optimal vaccination coverage of individuals as opposed to entire populations. Hill and Longini [25] use a general framework that could apply to several epidemic situations (e.g., diseases with permanent immunity (SIR models), incorporation of latent periods (SEIR models), or no immunity (SIS models)) and with and without vital dynamics. They develop a method to derive optimal vaccination strategies for populations divided into $m$ heterogeneous subgroups and fully examine the use of the model in populations with two subgroups and proportionate mixing.

Very few of these analytical models include discussion of the effect of parameter uncertainty on the vaccination policies identified and/or evaluated but this uncertainty can have major consequences. For example, Longini et al. [28] show that the optimal allocation of vaccines derived from their influenza model is highly sensitive to both the epidemiological characteristics of the virus and to the choice of the objective function used in the optimization process. Similar conclusions about the sensitivity of model outcomes to epidemiological and structural uncertainty are reached by Bansal et al. [7], who use a contact network model to compare morbidity-based strategies that target high prevalence populations and mortality-based strategies that target high risk populations, Dushoff et al. [14] who use a very simple model to explore the consequences of different vaccine allocation strategies, and Clancy and Green [12] who use a Bayesian-decision theoretic approach and a general stochastic SIR model with a homogenous population under parameter uncertainty.

Computer simulation models within a fully stochastic framework have also been used to assess the effectiveness of various potential strategies to control infectious disease spread. Most of these papers focus on pure control strategies, such as antivirals, vaccines, quarantine, and travel restrictions, that are implemented over the entire population. The effect of these strategies used individually and in different combinations are analyzed through simulation (see, for example, [15,17,31]). As an example of a simulation model focused specifically on the identification of an optimal vaccination strategy, Patel et al. [33] use a genetic algorithm within the framework of a simulation of pandemic influenza. Their algorithm is a heuristic; in other words, it is designed to find feasible solutions to the problem but there is no guarantee for how close those solutions are to the true optimal solution. It is important to note that at the present time heuristic approaches are all that are available for this class of problems. Also, due to the large amount of computer time per simulation run, none of the simulation papers discussed here consider the effects of parameter uncertainty.

Both analytical models and simulation models have weaknesses that must be considered in light of the goals of a modeling project. A major criticism of analytical deterministic and stochastic vaccination models that allow closed form representations of $R_v$ is that many assumptions are needed to have this property. These assumptions generally result in a model that is only a rough approximation of the actual spread of a disease through a population. Despite this weakness, analytical models can still be useful because they can give a clearer picture of the crucial parameters in a model [5]. For the task of identifying appropriate vaccination strategies, analytical models provide a good way to find mixed strategies that can provide insight into the groups that need to be particularly targeted by health authorities. Simulation models, which generally incorporate more realistic assumptions about population structure and disease transmission processes, are usually limited to pure or simple strategies because of the time required to run simulations given their complexity and the necessity of running them repeatedly because of their inherent randomness. Another important use of optimal strategies derived from analytical disease models is as a benchmark for strategies found via a heuristic on simulation models. The cost and effectiveness of the heuristic strategies can be checked against the optimal strategies of the analytical models to provide information on the quality of the heuristic strategies.

The complexity of human interactions means that parameter estimation for epidemiological models is notoriously difficult. Thus, vaccination policies found for any kind of model should be considered very carefully, especially if the uncertainty of the parameters is not taken into account. Policies derived from models with deterministic parameters may not be robust in the sense that even an optimal strategy might be highly suboptimal or even infeasible if parameters are changed slightly. Stochastic programming is a popular method for incorporating uncertainty in mathematical optimization problems by finding optimal decisions given that some parameter values are not deterministically known [10].

Using stochastic programming to include parameter uncertainty when finding optimal vaccination strategies can give several clear benefits. The stochastic programming framework allows for more robust vaccination strategies that are not as reliant on point estimates of parameter values. Stochastic programming can also help identify parameters to which optimal decisions are particularly sensitive, and so can provide guidance for allocation of resources for estimating parameters of the model.

The formulations presented in this paper include chance constraints to require $R_v < 1$ with at least a minimum probability, a random objective to minimize the probability that an epidemic will occur under resource constraints, and a cost-benefit formulation making the required probability for $R_v < 1$ a decision variable. The problem is formulated in a general framework that is valid for a wide class of epidemic models. We illustrate the stochastic

---

1 Epidemiological models are often formulated as a series of compartments corresponding to different disease states, e.g., susceptible, exposed, infectious, recovered, etc. The models are then referred to by the series of capital letters that corresponds to the compartments within the basic model structure. For example, an SIR model considers individuals to be either susceptible (S), infectious (I), or recovered (R) and to progress through the stages in that order; an SIS model would consist of the stages susceptible – infectious – susceptible and would represent a disease for which there was no immunity.
programming formulation using the heterogeneous household model of Becker and Starczak [9] and we provide a numerical example to show why including parameter uncertainty is important when devising an optimal vaccination strategy.

The rest of this paper is organized as follows: Section 2 gives a short introduction to stochastic programming and presents a general problem framework for finding optimal vaccination strategies under parameter uncertainty including a discussion of applying this technique to a variety of disease spread models. Section 3 describes the model of Becker and Starczak [9] and some of the basic results and extensions. The section continues with an example reformulation of their linear program as a stochastic program with probabilistic constraints. Section 4 gives a numerical illustration of this technique focusing on the value of information and the effects of not including uncertainty. Section 5 finishes with some conclusions and future work.

2. Stochastic programming

2.1. Background

Stochastic programming is an extension of mathematical programming in which the assumption that all data are known is relaxed; instead, a subset of the parameter values of the problem are characterized by probability distributions [10]. The goal of a mathematical programming problem is to identify an optimal solution, where optimality is defined in terms of a cost function to be minimized or maximized. The most popular measure of optimality in stochastic programming is in terms of the expected value of the objective function, but other risk functions can also be used [38]. Except for a few cases, solving stochastic programs with continuously distributed parameters is extremely difficult, so in most cases parameters are given discrete distributions or the continuous distributions are discretized through sampling. A realization \( \omega \) of the vector of random variables \( \omega \) of the problem is known as a scenario and is defined on the sample space \( \Omega \). The decision variables of the mathematical programs are given by the vector \( x \in \mathbb{R}^n \). The types of information that can be gained by optimizing a general, random function \( z(x, \omega) \) in the stochastic programming framework can be described as follows:

**Definition 2.1.** The stochastic programming solution (SPP) is the minimum expected value of the function \( z(x, \omega) \) in terms of the random variables \( \omega \). In other words,

\[
SPP = \min_x E_\omega [z(x, \omega)].
\]

Besides the optimal objective value and optimal decision variables, other useful statistics can be computed to show the effect of parameter uncertainty on the policies suggested by the model. Assuming that the random parameters have discrete distributions, the problem has a finite number of scenarios. The 'wait-and-see' solution assumes that the decision maker can see the realization of the random variables before any decisions need to be made.

**Definition 2.2.** The 'wait-and-see' solution (WS) given in Eq. (2) is the expected value of the solutions found by assuming that decisions are made after the random parameters are realized.

\[
WS = E_\omega [\min_x z(x, \omega)].
\]

**Definition 2.3.** The value of perfect information (VPI) is given by the equation

\[
VPI = SPP - WS.
\]

VPI measures how much improvement can be gained if the true values of the parameters are known. This information can be used to decide how much effort should be expended trying to improve estimates of the parameters.

Another important statistic found by analyzing the stochastic model is the value of including uncertainty in the formulation. To compute this value, a deterministic linear program is set up using the expected values of the random parameters as deterministic parameters. This deterministic linear program is solved to find optimal decision variable values.

**Definition 2.4.** The expected result of using the expected value solution (EEV) is the expected objective value of the solution found with the mean point estimates of the random parameters. EEV is given by

\[
EEV = E_\omega [z(x(\bar{\omega}), \omega)].
\]

Since the goal is to lower the probability of a disease spreading widely, the benefit of solving these problems as a stochastic program is in the added robustness of the optimal value. In this case, the EEV shows the probability of failure for the vaccination policy found by using point estimates of the parameter values and so gives an estimate of the gain in solution robustness due to including parameter uncertainty in the problem formulation.

2.2. Stochastic programming formulations

For this paper, we consider three possible formulations of the problem as a stochastic program. We consider a disease to be controlled if \( R < 1 \) and not controlled otherwise. Setting a reliability parameter \( a \), the first formulation minimizes the cost of vaccination under a chance constraint that requires \( R < 1 \) with at least probability \( a \). The second is the case where the vaccination supply is limited and the probability that the vaccination is insufficient to control the disease spread is minimized. The final formulation that we consider is a cost/benefit analysis with \( a \) as a variable instead of a parameter. These formulations are general and can be applied to a number of disease spread models; hence, the specific structures of the decision variables and constraint sets are all problem dependent.

In the context of finding vaccination policies, decision variables \( x \in \mathbb{R}^n \) define possible vaccination policies implied by the model. These decision variables are constrained by an arbitrary set \( X \), which defines allowable vaccination policies. The post-vaccination reproductive number is a function both of the vaccination policy \( x \) and the realization of the random variables \( \omega \) and is given by \( R = h(x, \omega) \). The cost of a vaccination policy is a function of \( x \) and is given by \( c(x) \).

Eqs. (5a)–(5c) give a general formulation of the problem as a chance constrained stochastic program. The objective function (5a) is to minimize the cost of the vaccination policy. The constraint (5b) is the probabilistic constraint and requires that \( R < 1 \) with probability greater than or equal to the reliability parameter, \( a \). The constraint (5c) defines the feasible space of allowable vaccination policies.

\[
\min \quad c(x) \tag{5a}
\]
\[
s.t. \quad P(h(x, \omega) \leq a) \geq a \tag{5b}
\]
\[
x \in X. \tag{5c}
\]

Higher values of \( a \) mean higher costs for the optimal prevention strategy since the disease must be prevented for a larger number of scenarios. The parameter is often chosen through computational experimentation, trading off the much higher cost of policies under extreme values of \( a \) with the costs of allowing too many infeasible policies.
A possible criticism of using chance constraints to formulate this problem is that the goal of policy makers is to prevent a major epidemic from ever occurring. Explicitly finding a vaccination strategy that allows for an acceptable failure rate clearly goes against this ideal. However, the problem of finding the vaccination strategy that requires \( R_c < 1 \) for all scenarios consists of defining the worst possible parameter values for the disease and solving the deterministic program for those values. The weakness of this approach is that strategies that are feasible for the worst values are often much more expensive than a strategy that is feasible for the vast majority of cases. Therefore in terms of vaccination strategies, the strategy to prevent every possible epidemic might just mean that everyone has to be vaccinated. This is not as useful as knowing how many fewer vaccine doses are needed to control the spread of a disease with high probability. Also, since estimating upper bounds is just as inexact as estimating average values, there is no guarantee that the upper bound estimate will actually mean that \( R_c < 1 \) in all cases. We feel that chance constraints allow for a more natural way to plan for bad scenarios and better reflect disaster planning in the real world. If failure rate of an optimal vaccination strategy found with formulation (5) is too low, then the value of \( x \) can always be increased.

Another situation for which an optimal vaccination policy might be required is when the vaccination budget is limited. In this case, it is not always possible to vaccinate in a way that makes \( R_c < 1 \), but it is still necessary to distribute the vaccines effectively. A measure of the effectiveness of a vaccine distribution with limited supplies that is analogous to the chance constraints is to optimally distribute the vaccines while minimizing the probability that \( R_c > 1 \). Adding a budget parameter \( B \), Eqs. (6a)-(6c) give a problem formulation to do this.

\[
\begin{align*}
\min & \quad \mathbb{P}(h(x, \omega) \geq 1) \\
\text{s.t.} & \quad c(x) \leq B \\
& \quad x \in X.
\end{align*}
\]

A third possible problem formulation is to explicitly consider the costs and benefits of lowering the probability that the reproductive number is less than one. Now, instead of a problem parameter, the reliability \( \alpha \) is a decision variable of the problem. The costs of a less reliable vaccination strategy are modeled with the cost function \( p(\alpha) \) leading to the following mathematical program.

\[
\begin{align*}
\min & \quad c(x) + p(\alpha) \\
\text{s.t.} & \quad \mathbb{P}(h(x, \omega) \leq 1) - \alpha \geq 0 \\
& \quad x \in X, \quad 0 \leq \alpha \leq 1.
\end{align*}
\]

2.3. Application to various disease spread models

This section will give a short introduction to some of the types of disease spread models for which stochastic optimization can be used to find optimal vaccination strategies. We discuss two main types of models, those for which an explicit function for \( h(x, \omega) \) can be derived and those for which \( h(x, \omega) \) can only be calculated implicitly through simulation. We will include a few remarks on solution methods for these different types of problems. The results here refer to problem (5), but they can easily be extended for the other two formulations.

Under general random parameter distributions, even finding a feasible solution to problem (5) may be impossible since computing the probability in constraint (5b) may be too computationally intensive [35]. In order to avoid this, it is necessary to assume that the distributions are discrete. This assumption is not too limiting as a discrete distribution can always be created from a continuous one by sampling.

For problems with discretely distributed random data, Morgan et al. [30] gives a formulation of the problem as a mixed-binary program that is more amenable to solution. There are a finite number of scenarios \( \omega \), each with a probability \( p_\omega \). Every scenario has a corresponding binary variable \( z_\omega \in \{0,1\} \) which takes the value of 0 if the disease is controlled in that scenario by the optimal vaccination policy. The variable takes a value of 1 if the disease is not controlled in that scenario. In the following formulations \( M \) is a sufficiently large number to guarantee that constraint (8b) is satisfied under scenario \( \omega \) whenever \( z_\omega = 1 \). A knapsack constraint (8c) ensures that the probabilistic constraint is satisfied by forcing the sum of the probabilities of scenarios where epidemics occur to be less than \( 1 - \alpha \).

\[
\begin{align*}
\min & \quad c(x) \\
\text{s.t.} & \quad h(x, \omega) - Mz_\omega \leq 1 \quad \forall \omega \in \Omega \\
& \quad \sum_{\omega \in \Omega} p_\omega z_\omega \leq 1 - \alpha \\
& \quad x \in X, \quad z \in \{0,1\}^\Omega.
\end{align*}
\]

In the general case, the functions \( h(x, \omega) \), \( c(x) \), and the constraints that define the feasible set \( X \) are non-linear, non-convex functions in terms of the decision variables \( x \). However, solving such a non-linear mathematical programming problem may not possible. It is possible to use non-linear programming techniques or heuristics to find local minima or feasible solutions that one hopes will be good enough [26], but these do not give any guarantee of solution quality. The formulation is more computationally tractable in the case where \( X \), \( c(x) \), and \( h(x, \omega) \) are given by convex functions in terms of \( x \). Such problems can then be solved using commercial optimization programs. Our example formulation gives an instance where \( X \), \( c(x) \), and \( h(x, \omega) \) are given by linear functions.

The other class of epidemic models for which the stochastic programming framework can possibly be used is that for which \( R_c \) can only be calculated by means of simulation. In this case, a problem of type (5) only exists implicitly and it is necessary to use heuristic methods in the simulation optimization framework in order to search for a feasible solution [39]. Simulation optimization methods are usually not particularly good at finding optimal solutions, but even a feasible solution to problem (5) can be valuable for defining a robust vaccination policy.

Remark 2.1. Many of the diseases for which vaccination is an option vary seasonally in their transmission, a feature that is usually modeled by incorporating a sinusoidal or similarly varying transmission parameter. Most vaccine optimization models do not take seasonality into account, however. One simple way to apply the stochastic programming framework to find an optimal vaccination policy in such a situation is to estimate the parameter distributions for the high season of the disease when the transmission parameter is maximal. In this case, \( \mathbb{P}(R_c \leq 1) \geq \alpha \) for the worst season, which means that the reliability requirement is guaranteed to be satisfied for the entire year. Of course using a strategy that guarantees reliability over an entire year based on transmission rates during the worst part of the year can be expensive, especially if resources are limited. In this case it might be more expedient to use a cost/benefit formulation that is tied to the cyclicity of the transmission parameter.
same as their formulation, we have changed the notation for clarity. In Section 3.2, our extension of this linear program to a stochastic program is explained.

3.1. Linear programming formulation

As a first step in formulating their linear program for finding optimal vaccination strategies, Becker and Starczak [9] compute $R_0$ and $R_*$ (the post-vaccination reproduction number) for their model. They assume in the model that the disease spreads quickly within individual households and spreads more slowly between them through close contacts between infected and susceptible members of different households. To ensure that the problem constraints are linear, they also assume proportionate mixing between households. This allows them to find a closed form equation for the post-vaccination reproduction number.

To formulate the program, it is necessary to define groups within the population that have different susceptibilities and infectivities. It is also necessary to define the different types of families that make up the overall population. The decision variables $x_{pf}$ of the program represent the proportion of households of type $f$ that are vaccinated under policy $v$. The rest of the model parameters and their descriptions are given in Table 1.

The full formulation of the linear program is given in equations (9a)–(9d). The objective function (9a) minimizes the vaccine coverage. The first constraint (9b) balances all the decision variables for each family type, ensuring that the proportions assigned sum to one. The second constraint (9c) requires that that reproductive number of the disease be brought below one. This constraint is a linear function of $a_{pf}$, which is itself a function of the parameters given by Eq. (10). The parameter $a_{pf}$ is derived in [9] and the value $\sum_{f,F} a_{pf} x_{pf}$ gives the post-vaccination reproduction number of the model.

$$\min \sum_{f,F} \sum_{v} v_i h_j x_{pf}$$

s.t. $\sum_{f} x_{pf} = 1 \ \forall f \in F$ (9b)

$\sum_{f,F} a_{pf} x_{pf} \leq 1$ (9c)

$0 \leq x_{pf} \leq 1 \ \forall f \in F, \ v \in V.$ (9d)

Table 1

<table>
<thead>
<tr>
<th>Sets</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>Set of family types</td>
</tr>
<tr>
<td>$T$</td>
<td>Set of types of people</td>
</tr>
<tr>
<td>$V$</td>
<td>Set of vaccine policies</td>
</tr>
<tr>
<td>$D$</td>
<td>The set of scenarios</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indices</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>Index for a family type in $F$</td>
</tr>
<tr>
<td>$v$</td>
<td>Index for a vaccination policy in $V$</td>
</tr>
<tr>
<td>$t$</td>
<td>Index for a type in $T$</td>
</tr>
<tr>
<td>$f_t$</td>
<td>Index for the number of people of type $t$ in a family of type $f$</td>
</tr>
<tr>
<td>$v_t$</td>
<td>Index for the number of people of type $t$ vaccinated in $v$</td>
</tr>
<tr>
<td>$\delta_t$</td>
<td>Index for a particular scenario in $\Omega$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_j$</td>
<td>The proportion of households in the population that are of type $f$</td>
</tr>
<tr>
<td>$\alpha_{pf}$</td>
<td>Computed parameter for impact of immunization decisions</td>
</tr>
<tr>
<td>$\mu_v$</td>
<td>The average size of a household</td>
</tr>
</tbody>
</table>

Parameters to compute $a_{pf}(\omega)$

| $m$                         | The average contact rate of infected people      |
| $u_t$                       | The relative infectivity of people of type $t$   |
| $s_t$                       | The relative susceptibility of people of type $t$ |
| $b$                         | The transmission proportion within a household   |
| $\epsilon$                 | The vaccine efficacy                             |

Decision variables $x_{pf}$

The proportion of families of type $f$ vaccinated under policy $v$

$$a_{pf}(\omega) = \frac{m h_j}{\mu_v} \left( \sum_{t,T} u_t s_t [(1-b)(f_t - v_t \epsilon) + b v_t \epsilon (1-\epsilon)] \right) + b \sum_{t,T} u_t s_t (f_t - v_t \epsilon).$$ (10)

In the case where vaccination supplies are limited and planners wish to minimize $R_*$, the formulation can be modified in the following manner. The lefthand side of constraint (9c) is moved to the objective and is minimized, and a constraint limiting the total number of vaccine doses that can be allotted to $D$ is created from the objective function (9a). This formulation is given by Eqs. (11a)–(11d).

$$\min \sum_{f,F} \sum_{v} v_i h_j x_{pf}$$ (11a)

s.t. $\sum_{v} x_{pf} = 1 \ \forall f \in F$ (11b)

$\sum_{f,F} \sum_{v} v_i h_j x_{pf} \leq D$ (11c)

$0 \leq x_{pf} \leq 1 \ \forall f \in F, \ v \in V.$ (11d)

Becker and Starczak [9] and Ball et al. [5] show that this linear program does not allow an easy characterization of the optimal strategy, meaning that the optimal strategy may not be easy to implement. However, they claim that it is still useful from a policy standpoint because it gives insight into groups that should be targeted in any vaccination plan. Constraints can be added to the model if more implementable plans are desired. For example, to limit policies to those where either an entire household is vaccinated or no members are, all decision variables corresponding to partially vaccinating a household are set to zero.

3.2. Stochastic programming formulation

We extend the linear programming model of Becker and Starczak [9] to the stochastic setting by considering the following parameters as random variables: the vaccine efficacy $\epsilon$, the average contact rate of infected people $m$, the relative infectivities and susceptibilities of people of different types $u_t$ and $s_t$, and the transmission proportion within a household $b$. The rest of the parameters of the model can be estimated more easily than these from census data and similar sources so they are assumed to be deterministic in our model. Depending on the goals of the modeler, a different number of random parameters could be included in the stochastic model. Methods for estimating the distributions of the random parameters can be found in Becker [8].

We reformulate problem (9) as a stochastic program with probabilistic constraints considering the previously mentioned parameters as random variables. This formulation is a special case of the general structure that was defined in Section 2. The stochastic formulation is given by Eqs. (12a)–(12d).

$$\min \sum_{f,F} \sum_{v} v_i h_j x_{pf}$$ (12a)

s.t. $\sum_{v} x_{pf} = 1 \ \forall f \in F$ (12b)

$P \left( \sum_{f,F} a_{pf}(\omega) x_{pf} \leq 1 \right) \geq \alpha$ (12c)

$0 \leq x_{pf} \leq 1 \ \forall f \in F, \ v \in V.$ (12d)

where $a_{pf}$ now becomes a function of the random parameters that we have defined.
Eqs. (14a)–(14e), which can be solved without any modification.

Formulation (12) can be reformulated as a mixed-binary program as was described in Section 2. The remainder of this paper will be concerned with this mixed-binary formulation (given in Eqs. (14a)–(14e), which can be solved without any modification.

A mixed-binary formulation can be similarly derived for the case with a limited vaccination budget.

\[ a_{f}(\omega) = \frac{m(\omega)h_{f}}{\mu_{f}} \left( \sum_{c_{i}\in C} u_{c}(\omega)s_{c}(\omega)[1 - b(\omega)](f_{t} - v_{t}\epsilon(\omega)) \right. \]
\[ + b(\omega)v_{t}\epsilon(1 - c) + b \sum_{c_{i}\in C} u_{c}(\omega)s_{c}(\omega) \]
\[ \times (f_{t} - v_{t}\epsilon(\omega))(f_{t} - v_{t}\epsilon(\omega)) \right). \]

(13)

Formulation (12) can be reformulated as a mixed-binary program as was described in Section 2. The remainder of this paper will be concerned with this mixed-binary formulation (given in Eqs. (14a)–(14e), which can be solved without any modification.

A mixed-binary formulation can be similarly derived for the case with a limited vaccination budget.

\[
\begin{align*}
\min & \sum_{f\in F} \sum_{v\in V} \sum_{t\in T} v_{t}h_{f}x_{fv} \\
\text{s.t.} & \sum_{v\in V} x_{fv} = 1 & & \forall f \in F \tag{14a} \\
& \sum_{f\in F} \sum_{v\in V} a_{f}(\omega)x_{fv} - Mz_{\omega} \leq 1 & & \forall \omega \in \Omega \tag{14b} \\
& \sum_{\omega \in \Omega} \pi_{\omega}z_{\omega} \leq 1 - \alpha \tag{14b} \\
& 0 \leq x_{fv} \leq 1 & & \forall f \in F, v \in V, z_{\omega} \in \{0, 1\} & & \forall \omega \in \Omega. \tag{14e}
\end{align*}
\]

4. Numerical example

We generated a small set of random test instances in order to illustrate the effect of random parameter values on the optimal policies found by mathematical programming. The goal of this section is to use example instances of the problem to show why stochastic programming with probabilistic constraints is useful for finding optimal vaccination policies. In particular, this section will show the importance of using random distributions for the model parameters rather than point estimates in terms of the cost and effectiveness of the vaccination policy given in the solution.

To set up our random test instances, we generated values and distributions for the model parameters, which included family group parameters that define the makeup of the population and parameters that control the spread of the disease. For this example we chose parameter distributions that seemed plausible based on information in the epidemiological literature; to properly estimate them is beyond the scope of this project.

The family group parameters included three different types of people: children, adults, and the elderly. We defined 30 possible family groupings comprised of different numbers of these types, the details of which can be found in Appendix Table 3. The disease parameters were more difficult to estimate. According to Longini et al. [29] a plausible value of \( R_0 \) for influenza is estimated to be around 2.0. We defined our parameters so that the distribution of the \( R_0 \) values would be mostly in the interval [1.5, 2.5]. For the efficacy of the vaccine we assumed a truncated normal distribution with a mean of 0.85 and a standard deviation of 0.1. See Appendix Table 4 for the assumed distributions of the remaining parameters.

Our test instances were created by independently sampling the parameters from their defined distributions. We formulated the problem with probabilistic constraints as given in formulation (14). We limited the number of scenarios of the instance to 500 so that the instance could be solved quickly by a commercial solver. We found solutions to the problems by using the mixed-integer programming solver CPLEX 9.0. The objective values of the formulations were weighted so that the numerical value was equal to the percentage of people who need to be vaccinated in the optimal strategy.

Fig. 1 shows how increasing \( \alpha \) affects the percentage of people who need to be vaccinated in the optimal strategy. The striking detail in this plot is that the percentage of people who need to be vaccinated increases relatively linearly when the probability that \( R_0 \leq 1 \) is between 20% and 95%. However, when \( \alpha > 0.95 \), the required number of vaccine doses increases at a much faster rate. The plot gives evidence for why stochastic programming with probabilistic constraints is a good framework for finding optimal vaccination policies. Since resources for the prevention of disease are limited, it is important to be able to identify a level for which the probability that \( R_0 > 1 \) is low, but the vaccination coverage is not too extreme. This plot shows that this constraint can be satisfied with high probability using relatively few doses of a vaccine, but that increasing the probability that \( R_0 \leq 1 \) beyond that requires a huge increase in vaccine supplies. A plot such as this can be used to set the parameter \( \alpha \). Since the number of doses needed to prevent extra epidemics starts to increase quickly above \( \alpha = 0.95 \), this value is a reasonable choice for that parameter.

Fig. 1. Plot of vaccine proportion vs epidemic prevention rate.
We next solved all the test problems with $x = 0.95$, computed the expected value of the expected solution, the wait-and-see solution, and the value of perfect information. The results of these computations are given in Table 2. In this table SPP designates the stochastic problem solution, EEV designates the expected value of the expected solution, WS designates the wait-and-see solution, and VPI designates the value of perfect information, respectively.

The SPP column of Table 2 shows that $R_c \leq 1$ with probability 95% can be achieved by vaccinating on average 65.54% of the population. This number gives the absolute minimum percentage of the population that would have to be vaccinated in order to guarantee that the chance constraint is satisfied, however this information alone is not particularly valuable in terms of defining a vaccination strategy to prevent a real disease. The optimal decision variable values give the exact proportion of each family type $f \in F$ that need to be vaccinated according vaccination policy $v \in V$. This means that the optimal vaccination policy found by solving formulation (5) is not likely to be implementable from a public policy standpoint because it is unrealistically specific about which people need to be vaccinated. Nevertheless, the optimal objective value is an absolute lower bound on the cost of any effective strategy so it does give us a starting point for budgeting for a real vaccination program. Also, the optimal decision variable values can be used to identify particular groups of people that are crucial to vaccinate and hence should be specifically targeted in the actual vaccination plan.

We computed the expected value of the expected value solution (EEV) for the test instance. To compute the expected value decision, formulation (9) was set up and solved using the expected values of the parameter distributions to find optimal decision variable $x'$. Then for each scenario $\omega$, we tested whether using the vaccination policy given by $x'$ satisfies the constraint $h(x, \omega) \leq 1$. As is shown in the EEV column of the table, in this case, an $R_c > 1$ occurs on average 95.0% of the time. This is clearly unacceptable for a vaccination policy. This result indicates that the effectiveness of a vaccination policy is highly susceptible to how the parameters are distributed. Using the traditional expected values of parameters gives a solution that is not robust enough to be useful.

We also computed the value of perfect information $VPI = WS - SPP$ for our example. To compute this value it was necessary to compute the cost of the optimal objective value using formulation (9) for each of the 500 scenarios and then take the average of those values. The average optimal percentage of the population to vaccinate when the stochastic data are known is 40.94%. Hence the value of perfect information is 24.70%. This means that if parameters were able to be estimated perfectly, $R_c \leq 1$ with an average of over a third fewer vaccine doses. From a policy standpoint, this helps decide how much effort and resources should be spent finding more exact estimates of the model parameters.

The purpose of setting up and solving a small example like this is to show that traditional linear programming may be insufficient for finding good vaccination strategies even under the assumptions for which it can be used. In particular, the expected value solution shows that using point estimates of the epidemic parameters is not robust enough to be used to plan for actual epidemics. On the other hand, the stochastic programming solution gives a vaccine allocation program that does not require the vast majority of people to be vaccinated, but is robust enough to prevent most epidemics from occurring. Also, the value of perfect information shows that because parameter uncertainty has a substantial effect on policies that the program returns it may be worthwhile to expend significant effort to better estimate parameters.

5. Conclusions

This paper introduces stochastic programming with chance constraints as a framework for including parameter uncertainty when finding optimal vaccination policies. We give general stochastic programming formulations of the problem that can be used for a wide class of epidemic models. As an example, the linear programming formulation of Becker and Starczak [9] is extended to this stochastic programming framework. We then use a numerical example to show the large effect that including parameter uncertainty has on the optimal vaccination strategies. We believe that
since accurate parameter estimation can be extremely difficult for epidemic models, ignoring parameter uncertainty is not a good assumption to make when creating a vaccination policy. Extensions of this work include creating realistic estimates of the different parameter values and testing the robustness of the optimal vaccination schemes through simulation, as well as extending the stochastic programming framework to other epidemic models.

Appendix A

See Tables 3 and 4.

References