Finding Optimal Vaccination Strategies under Uncertainty using Stochastic Programming

Matthew W. Tanner *
Department of Industrial and Systems Engineering
Texas A&M University

Lisa Sattenspiel
Department of Anthropology
University of Missouri

Lewis Ntaimo
Department of Industrial and Systems Engineering
Texas A&M University

Abstract

We present a stochastic programming formulation for finding the optimal vaccination policy for preventing infectious disease epidemics. Stochastic programming is a popular framework for including the effects of parameter uncertainty in a mathematical optimization model. The problem is formulated with probabilistic constraints to find the minimum cost vaccination policy with the constraint that the proportion of epidemics that are prevented is greater than some value α that is set by the modeler. We also show how to formulate the problem of finding the optimal vaccination policy when vaccine supply is limited. The class of epidemic models for which this method is valid is described and we present an example

*Corresponding Author: mtanner@tamu.edu 241 Zachry, 3131 TAMU, College Station, TX 77843-3131

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 the effects of using stochastic programming to find vaccination strategies.
Keywords{parameter uncertainty, vaccination, stochastic programming}

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, 24, 35, 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 [1, 3, 21, 37], rubella [2, 14, 20, 28], pertussis [19, 22, 23, 25], 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 [24] and to guide the development of vaccination policies for specific diseases, including tuberculosis [35], influenza [29], and smallpox [17].

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 [1] with both local mixing at the household level and global mixing at the community level. They introduce the notation, R_*, 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_* 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_*, then formulate and numerically solve a linear program to find the minimum vaccination coverage under the constraint $R_* \leq 1$. This constraint ensures that the disease will tend to die out before causing an epidemic. (Becker and Starczak [9] use the notation R_{HV} rather than R_*, 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 $\{n\mu_n\}$ is convex, where μ_n is the mean size of a local outbreak within a household of size n, 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

[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.

the conditions needed for optimal vaccination coverage of individuals as opposed to entire populations. Hill and Longini [26] 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) with and without vital dynamics). They develop a method to derive optimal vaccination strategies for populations divided into m heterogenous 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. [29] 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. [15] 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 anti-virals, 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, 16, 18, 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 which means that 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. It is also important to note that, 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_* 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 effect 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 probabilistic constraints to either require a certain proportion of disease epidemics to be prevented or to minimize the probability that an epidemic will occur under resource constraints. The problem is formulated in a general framework that is valid for any epidemic model for which a closed form equation for the post-vaccination reproductive number exists. We illustrate the stochastic 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. 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 random 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 ω of the vector of random variables $\tilde{\omega}$ of the problem is known as a scenario and is defined on the sample space Ω. The decision variables of the mathematical programs are given by the vector $x \in \mathcal{R}^n$. Equation (1) defines the stochastic programming problem solution (SPP) when minimizing the

expected value of a function $z(x, \omega)$ in terms of the random variables $\tilde{\omega}$[10].

$$SPP = \min_x E_{\tilde{\omega}}\big[z(x, \omega)\big] \tag{1}$$

For this paper, we consider two possible formulations of the problem as a stochastic program. The first is minimizing the cost of vaccination under a probabilistic constraint that requires epidemics to be prevented with at least probability α. The particular value of α is set by the modeler. The second is the case where the vaccination supply is limited and the probability of an epidemic occurring is minimized. This section gives a general stochastic programming formulation that can be applied to any disease spread model that allows a closed form equation for the post-vaccination reproductive number. The useful information that can be gained by solving the problem as a stochastic program is explained, and also some background on possible solution methods for the problems is given.

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. Hence, an optimal solution can be found for each scenario as if the parameters were distributed deterministically according to that scenario. The expected value of these solutions can then be computed. The resulting solution is called the "wait-and-see" solution (WS) and is given by equation (2)[10].

$$WS = E_{\tilde{\omega}}\big[\min_x z(x, \omega)\big] \tag{2}$$

The objective value of the stochastic programming solution minus the cost of the "wait-and-see" solution is known as the value of perfect information (VPI) as it gives a measure of 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 statistic found by analyzing the stochastic model is the value of the stochastic solution. 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. Then, for each scenario,

the cost of using these decision variable values is computed. The average value over all the scenarios of these costs is known as the expected result of using the expected value solution (EEV). It is given in equation (3), where $\bar{\omega}$ is the expected value of the stochastic parameters and $\bar{x}(\bar{\omega})$ is the expected value solution[10].

$$EEV = E_{\tilde{\omega}}\big[z(\bar{x}(\bar{\omega}), \omega)\big] \tag{3}$$

Since the goal is to prevent epidemics, 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.

2.2 Probabilistic Programming

In probabilistic programming either some subset of constraints are allowed to be violated an acceptable amount of the time or else the objective function involves minimizing a probability [10]. The stochastic programming formulations of this section are applicable to any disease spread model for which R_*, the post-vaccination reproduction number, can be explicitly defined by a closed form equation. Since these formulations are general, the specific structures of the decision variables and constraint sets are all problem dependent.

In the context of finding vaccination policies, decision variables $x \in \mathcal{R}^n$ 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 ω and is given by $R_* = h(x, \omega)$. The cost of a vaccination policy is a function of x and is given by $c(x)$.

Equations (4a)-(4c) give a general formulation of the problem as a probabilistically constrained stochastic program. The objective function (4a) is to minimize the cost of the vaccination policy. The constraint (4b) is the probabilistic constraint and requires that $R_* \leq 1$ with probability greater than α. The constraint (4c) defines the feasible space of allowable vaccination policies.

$$\begin{aligned} \min \quad & c(x) && \text{(4a)} \\ \text{s.t.} \quad & \mathbb{P}\big(h(x,\omega) \leq 1\big) \geq \alpha && \text{(4b)} \\ & x \in X && \text{(4c)} \end{aligned}$$

Higher values of α 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 α with the costs of allowing too many infeasible policies.

A possible criticism of using probabilistic constraints to formulate this problem is that the only goal is to prevent epidemics. So explicitly finding a vaccination strategy that allows for an acceptable failure rate is not realistic. The problem of finding the vaccination strategy that prevents all epidemics 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 prevent the vast majority of epidemics. Also, since estimating upper bounds is just as inexact as estimating average values, there is no guarantee that the upper bound estimate will also prevent all epidemics. We feel that probabilistic constraints allow for a more natural way to plan for bad scenarios and better reflect disaster planning in the real world. If a higher proportion of epidemics need to be prevented, then the value of α can always be increased.

In the general case, the functions $h(x,\omega)$, $c(x)$, and the constraints that define the feasible set X are allowed to be nonlinear. However, solving such a nonlinear mathematical programming problem may not possible. It is possible to use nonlinear programming techniques or heuristics to find local minimums or feasible solutions that one hopes will be good enough [27], 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 linear functions. Obviously this requires a tradeoff in the form of more assumptions in the disease spread model being used.

There are several ways of solving formulation (4) when all the functions are linear. With strong assumptions on the distribution of $a_{ij}(\omega)$, the feasible space of the problem

is convex and the problem can be solved directly [13]. However since the relationship between the model parameters and the constraint matrix elements is usually nonlinear, these convexity results cannot be applied. In general, the problem is highly nonconvex and specialized algorithms must be used to find optimal strategies. Most of these algorithms make the assumption that the distributions are discrete. Ruszczyński [36] gives a formulation of the problem as a mixed-binary program that can either be solved directly by a commercial solver or with a specialized algorithm [36, 39].

Formulation (5) is a mixed-binary reformulation of problem (4) that can be solved directly. By assumption, there are a finite number of scenarios ω, each with a probability p_ω. Every scenario has a corresponding binary variable z_ω which takes the value of 0 if an epidemic is prevented in that scenario by the optimal vaccination policy. The variable takes a value of 1 if an epidemic is not prevented in that scenario. In the following formulations M is a large number. A knapsack constraint (5c) ensures that the probabilistic constraint is satisfied by forcing the sum of the probabilities of scenarios where epidemics occur to be less than $1 - \alpha$.

$$\begin{aligned}
\min \quad & c(x) && \text{(5a)}\\
\text{s.t.} \quad & h(x,\omega) - Mz_\omega \leq 1 \ \forall \omega \in \Omega && \text{(5b)}\\
& \sum_{\omega \in \Omega} p_\omega z_\omega \leq 1 - \alpha && \text{(5c)}\\
& x \in X,\ z \in \mathcal{B}^{|\Omega|} && \text{(5d)}
\end{aligned}$$

Another situation for which an optimal vaccination policy might be required is when the vaccination budget is limited. In this case, there is a limit to the proportion of epidemics that can be prevented, 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 probabilistic constraints is to optimally distribute the vaccines while minimizing the probability that an epidemic occurs.

This problem is also formulated with a binary decision variable z_ω for each scenario. The formulation is similar to formulation (5), but the objective is now to minimize the sum of the probabilities of the unsatisfied scenarios. The vaccine coverage is now a constraint forced to be less than or equal to D, where D is the budget constraint for the vaccination policy.

$$\min \quad \sum_{\omega \in \Omega} p_\omega z_\omega \tag{6a}$$

$$\text{s.t.} \quad h(x,\omega) - M z_\omega \leq 1 \ \forall \omega \in \Omega \tag{6b}$$

$$c(x) \leq D \tag{6c}$$

$$x \in X, \ z \in \mathcal{B}^{|\Omega|} \tag{6d}$$

3 Example Model

This section gives an example formulation of a stochastic program in the case where the constraints that define $c(x)$, $h(x,\omega)$, and X are all linear. In Section 3.1, the SIR epidemic model and a formulation of the optimal vaccination problem as a linear program as given in Becker and Starczak [9] are described. Note that although the linear programming formulation given here is the 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

Before formulating a linear program for finding optimal vaccination strategies, Becker and Starczak [9] show how to compute R_0 for their model and also how to compute R_*, the post-vaccination reproduction number. The assumptions of the model are 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 in order 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_{fv} 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.

********table 1**********

The full formulation of the linear program is given in equations (7a)- (7d). The objective function (7a) minimizes the vaccine coverage. The first constraint (7b) balances all the decision variables for each family type, ensuring that the proportions assigned sum to one. The second constraint (7c) requires that that reproductive number of the disease be brought below one. This constraint is a linear function of a_{fv}, which is itself a function of the parameters given by equation (8). The parameter a_{fv} is derived in [9] and the value $\sum_{f\in F}\sum_{v\in V} a_{fv}x_{fv}$ gives the post-vaccination reproduction number of the model.

$$\begin{aligned}
\min: \quad & \sum_{f\in F}\sum_{v\in V}\sum_{t\in T} v_t h_f x_{fv} && \text{(7a)}\\
\text{s.t.} \quad & \sum_{v\in V} x_{fv} = 1 \quad \forall f \in F && \text{(7b)}\\
& \sum_{f\in F}\sum_{v\in V} a_{fv}x_{fv} \leq 1 && \text{(7c)}\\
& 0 \leq x_{fv} \leq 1 \quad \forall f \in F, v \in V && \text{(7d)}
\end{aligned}$$

$$\begin{aligned}
a_{fv}(\omega) = \quad & \frac{mh_f}{\mu_F}\Big(\sum_{t\in T} u_t s_t\big[(1-b)(f_t - v_t\epsilon) + bv_t\epsilon(1-\epsilon)\big] && \text{(8)}\\
& + b\sum_{t\in T}\sum_{r\in T} u_r s_t (f_t - v_t\epsilon)(f_r - v_r\epsilon)\Big)
\end{aligned}$$

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 (7c) 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 (7a). This formulation is given by equations (9a) - (9d).

$$\begin{aligned}
\min \quad & \sum_{f\in F}\sum_{v\in V} a_{fv}x_{fv} && \text{(9a)}\\
\text{s.t.} \quad & \sum_{v\in V} x_{fv} = 1 \quad \forall f \in F && \text{(9b)}\\
& \sum_{f\in F}\sum_{v\in V}\sum_{t\in T} v_t h_f x_{fv} \leq D && \text{(9c)}\\
& 0 \leq x_{fv} \leq 1 \quad \forall f \in F, v \in V && \text{(9d)}
\end{aligned}$$

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 ϵ, 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. It is certainly possible to have more or less random parameters in the stochastic model depending on the goals of the modeler. Methods for estimating the distributions of the random parameters can be found in Becker [8].

We reformulate problem (7) 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 equations (10a) - (10d).

$$\min \quad \sum_{f\in F}\sum_{v\in V}\sum_{t\in T} v_t h_f x_{fv} \tag{10a}$$

$$\text{s.t.} \quad \sum_{v\in V} x_{fv} = 1 \quad \forall f \in F \tag{10b}$$

$$\mathbb{P}\Big(\sum_{f\in F}\sum_{v\in V} a_{fv}(\omega) x_{fv} \leq 1\Big) \geq \alpha \tag{10c}$$

$$0 \leq x_{fv} \leq 1 \quad \forall f \in F, v \in V \tag{10d}$$

where α_{fv} now becomes a function of the random parameters that we have defined

$$a_{fv}(\omega) = \frac{m(\omega)h_f}{\mu_F}\Big(\sum_{t\in T} u_t(\omega)s_t(\omega)\big[(1-b(\omega))(f_t - v_t\epsilon(\omega)) + b(\omega)v_t\epsilon(1-\epsilon)\big] + b\sum_{t\in T}\sum_{r\in T} u_r(\omega)s_t(\omega)(f_t - v_t\epsilon(\omega))(f_r - v_r\epsilon(\omega))\Big) \quad (11)$$

Formulation (10) can be reformulated as a mixed-binary program as was described in Section 2. This is the formulation that we will deal with throughout the rest of the paper (equations (12a) - (12e)). It can be solved without any modification. For the case with a limited vaccination budget, the mixed-binary formulation can be similarly set up.

$$\min \quad \sum_{f\in F}\sum_{v\in V}\sum_{t\in T} v_t h_f x_{fv} \quad (12a)$$

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

$$\sum_{f\in F}\sum_{v\in V} a_{fv}(\omega)x_{fv} - Mz_\omega \leq 1 \ \forall \omega \in \Omega \quad (12c)$$

$$\sum_{\omega\in\Omega} p_\omega z_\omega \leq 1-\alpha \quad (12d)$$

$$0 \leq x_{fv} \leq 1 \ \ \forall f \in F, v \in V, \ z_\omega \in \{0,1\} \ \forall \omega \in \Omega \quad (12e)$$

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 linear 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. The two types of parameters that we needed to define were the family group parameters that define the makeup of the population, and the parameters that control the spread of the disease. We gave the parameters distributions that seem

plausible according to the 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. [30] 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 rest of the 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 (12). 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 are weighted so that the numerical value is equal to the percentage of people that would need to be vaccinated in the optimal strategy.

Figure 1 shows how increasing the proportion of epidemics that must be prevented affects the percentage of people that need to be vaccinated in the optimal strategy. The striking detail in this plot is that the percentage of people that need to be vaccinated increases relatively linearly with the proportion of epidemics that are prevented when that percentage is between 20% and 95%. However, when more than 95% of epidemics are to be prevented, 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 how many epidemics can be prevented without extreme effort. This plot shows that the vast majority of epidemics could be stopped with relatively few doses of a vaccine, but that preventing the last few epidemics would require a huge increase in vaccine supplies. A plot such as this can also be used to set the parameter α. 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.

******figure 1********

We next solved all the test problems with $\alpha = 0.95$, 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 value solution, WS designates the wait-and-see solution, and VPI designates the value of perfect information respectively.

********table 2**********

The SSP column of Table 2 shows that 95% of epidemics can be prevented 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 to prevent 95% of epidemics, however this information alone is not particularly valuable in terms of defining a vaccination strategy to prevent a real disease. The decision variable values of this optimal solution define a specific vaccination policy that is likely to not be implementable either from a public policy standpoint or because it is unrealistically specific about which people need to be vaccinated. Nevertheless, the 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. The other main use of the optimal solution is that the decision variable values do 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 (7) was set up and solved using the expected values of the parameter distributions to find optimal decision variable x^*. Then for each scenario, we tested whether or not using the vaccination policy given by x^* would prevent an epidemic. As is shown in the EEV column of the table, in this case, an epidemic occurs an average of 55.0% 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 (7) for each of the 500 scenarios and then take the average of those values. The average optimal objective 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, epidemics could be prevented 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 a large effort may be worthwhile to better estimate parameters because parameter uncertainty has a substantial effect on policies that the program returns.

5 Conclusions

This paper introduces stochastic programming with probabilistic constraints as a method for including parameter uncertainty when finding optimal vaccination policies. We give general probabilistic formulations of the problem that can be used for any epidemic models that allow a closed form post-vaccination reproductive number. 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.

References

[1] Z. Agur, Y. L. Danon, R. M. Anderson, L. Cojocaru, and R. M. May. Measles immunization strategies for an epidemiologically heterogeneous population - the Israeli case-study. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 252(1334):81–84, 1993.

[2] R. M. Anderson and R. M. May. Vaccination against rubella and measles: quantitative investigation of different policies. *Journal of Hygiene-Cambridge*, 90: 259–325, 1983.

[3] H. R. Babad, D. J. Nokes, N. J. Gay, E. Miller, P. Morgan-Capner, and R. M. Anderson. Predicting the impact of measles vaccination in England and Wales: model validation and analysis of policy options. *Epidemiology and Infection*, 114: 319–344, 1995.

[4] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *Annals of Applied Probability*, 7:46–89, 1997.

[5] Frank Ball, Tom Britton, and Owen Lyne. Stochastic multitype epidemics in a community of households: estimation and form of optimal vaccination schemes. *Mathematical Biosciences*, 191(1):19–40, 2004.

[6] Frank G. Ball and Owen D. Lyne. Optimal vaccination policies for stochastic epidemics among a population of households. *Mathematical Biosciences*, 177-178: 333–354, 2002.

[7] Shweta Bansal, Babak Pourbohloul, and Lauren Ancel Meyers. A comparative analysis of influenza vaccination programs. *PLoS Medicine*, 3:e387, 2006.

[8] Niels Becker. Estimation of parameters relevant for vaccination strategies. *Bulletin de l'Institut International de Statistique*, 56(2):1279–1289, 1995.

[9] Niels G. Becker and Dianna N. Starczak. Optimal vaccination strategies for a community of households. *Mathematical Biosciences*, 139(2):117–132, 1997.

[10] John R. Birge and Francois Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer-Verlag, New York, 1997.

[11] S. Brogger. Systems analysis in tuberculosis control: a model. *American Review of Respiratory Diseases*, 95:419–434, 1967.

[12] Damian Clancy and Nathan Green. Optimal intervention for an epidemic model under parameter uncertainty. *Mathematical Biosciences*, 205:297–314, 2007.

[13] Darinka Dentcheva, Bogumila Lai, and Andrzej Ruszczyński. Efficient point methods for probabilistic optimization problems. *Mathematical Methods of Operations Research*, Vol. 60:331–346, 2004.

[14] Klaus Dietz. The evaluation of rubella vaccination strategies. In R. W. Hiorns and D. Cooke, editors, *The Mathematical Theory of the Dynamics of Biological Populations II*, pages 81–97. Academic Press, London, 1981.

[15] Jonathan Dushoff, Joshua B. Plotkin, Cecile Viboud, Lone Simonsen, Mark Miller, Mark Loeb, and David J. D. Earn. Vaccinating to protect a vulnerable subpopulation. *PLoS Medicine*, 4:e174, 2007.

[16] Neil M. Ferguson, Derek A. T. Cummings, Christophe Fraser, James C. Cajka, Philip C. Cooley, and Donald S. Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442:448–452, 2006.

[17] James C. Frauenthal. *When should routine vaccination be discontinued?* The UMAP Expository Monograph Series. Birkhäuser, Boston, 1981.

[18] Timothy C. Germann, Kai Kadau, Jr. Longini, Ira M., and Catherine A. Macken. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Science, USA*, 103(15):5935–5940, 2006.

[19] Herbert W. Hethcote. New vaccination strategies for pertussis. In Carlos Castillo-Chavez, Sally Blower, Pauline van den Driessche, Denise Kirschner, and Abdul-Aziz Yakubu, editors, *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction*, The IMA Volumes in Mathematics and its Applications, pages 97–118. Springer, New York, 2002.

[20] Herbert W. Hethcote. Measles and rubella in the united states. *American Journal of Epidemiology*, 117:2–13, 1983.

[21] Herbert W. Hethcote. Optimal ages of vaccination for measles. *Mathematical Biosciences*, 89:29–52, 1988.

[22] Herbert W. Hethcote. An age-structured model for pertussis transmission. *Mathematical Biosciences*, 145:89–136, 1997.

[23] Herbert W. Hethcote. Simulations of pertussis epidemiology in the United States. *Mathematical Biosciences*, 158:47–73, 1999.

[24] Herbert W. Hethcote and Paul Waltman. Optimal vaccination schedules in a deterministic epidemic model. *Mathematical Biosciences*, 18:365–381, 1973.

[25] Herbert W. Hethcote, Peter Horby, and Peter McIntyre. Using computer simulations to compare pertussis vaccination strategies in Australia. *Vaccine*, 22:2181–2191, 2004.

[26] Andrew N. Hill and Ira M. Longini, Jr. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical Biosciences*, 181:85–106, 2003.

[27] Reiner Horst, Panos M. Pardalos, and Nguyen V. Thoai. *Introduction to Global Optimization*. Kluwer Academic Publishers, Dordtrecht, The Netherlands, 2000.

[28] E. G. Knox. Strategy for rubella vaccination. *International Journal of Epidemiology*, 9:13–23, 1980.

[29] Ira M. Longini, Jr., Eugene Ackerman, and Lila R. Elveback. An optimization model for influenza A epidemics. *Mathematical Biosciences*, 38:141–157, 1978.

[30] Ira M. Longini, Jr., M. Elizabeth Halloran, Azhar Nizam, and Yang Yang. Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, 159(7):623–633, 2004.

[31] R. S. Morris, J. W. Wilesmith, M. W. Stern, R. L. Sanson, and M. A. Stevenson. Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. *Veterinary Record*, 149(5):137–144, 2001.

[32] Johannes Müller. Optimal vaccination strategies–for whom? *Mathematical Biosciences*, 139(2):133–154, 1997.

[33] Rajan Patel, Ira M. Longini, Jr, and Elizabeth M. Halloran. Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. *Journal of Theoretical Biology*, 234(2):201–212, 2005.

[34] Babak Pourbohloul, Lauren Ancel Meyers, Danuta M. Skowronski, Mel Krajden, David M. Patrick, and Robert C. Brunham. Modeling control strategies of respiratory pathogens. *Emerging Infectious Diseases*, 11(8):1249–1256, 2005.

[35] Charles S. Revelle, Walter R. Lynn, and Floyd Feldmann. Mathematical models for the economic allocation of tuberculosis control activities in developing nations. *American Review of Respiratory Diseases*, 96:893–909, 1967.

[36] Andrzej Ruszczyński. Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming*, 93(2):195–215, 2002.

[37] Boris Shulgin, Lewi Stone, and Zvia Agur. Pulse vaccination strategy in the sir epidemic model. *Bulletin of Mathematical Biology*, 60:1–26, 1998.

[38] G. Szegö. Measures of risk. *Journal of Banking and Finance*, 26:1253–1272, 2002.

[39] Sridhar R. Tayur, Rekha R. Thomas, and N.R. Natraj. An algebraic geometry algorithm for scheduling in presence of setups and correlated demands. *Mathematical Programming*, 69:369–401, 1995.

[40] H. Waaler, A. Geser, and S. Anderson. The use of mathematical models in the study of the epidemiology of tuberculosis. *American Journal of Public Health*, 52: 1002–1013, 1962.

Appendix

********table 3********** ********table 4**********

Table 1: Model Parameters

Sets	
F	set of family types
T	set of types of people
V	set of vaccine policies
Ω	the set of scenarios
Indices	
f	index for a family type in F
v	index for a vaccination policy in V
t	index for a person type in T
f_t	index for the number of people of type t in a family of type f
v_t	index for the number of people of type t vaccinated in v
ω	index for a particular scenario in Ω
Parameters	
h_f	the proportion of households in the population that are of type f
a_{nv}	computed parameter for impact of immunization decisions
μ_F	the average size of a household
Parameters to compute $a_{nv}(\omega)$	
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
ϵ	the vaccine efficacy
Decision Variables	
x_{fv}	the proportion of families of type f vaccinated under policy v

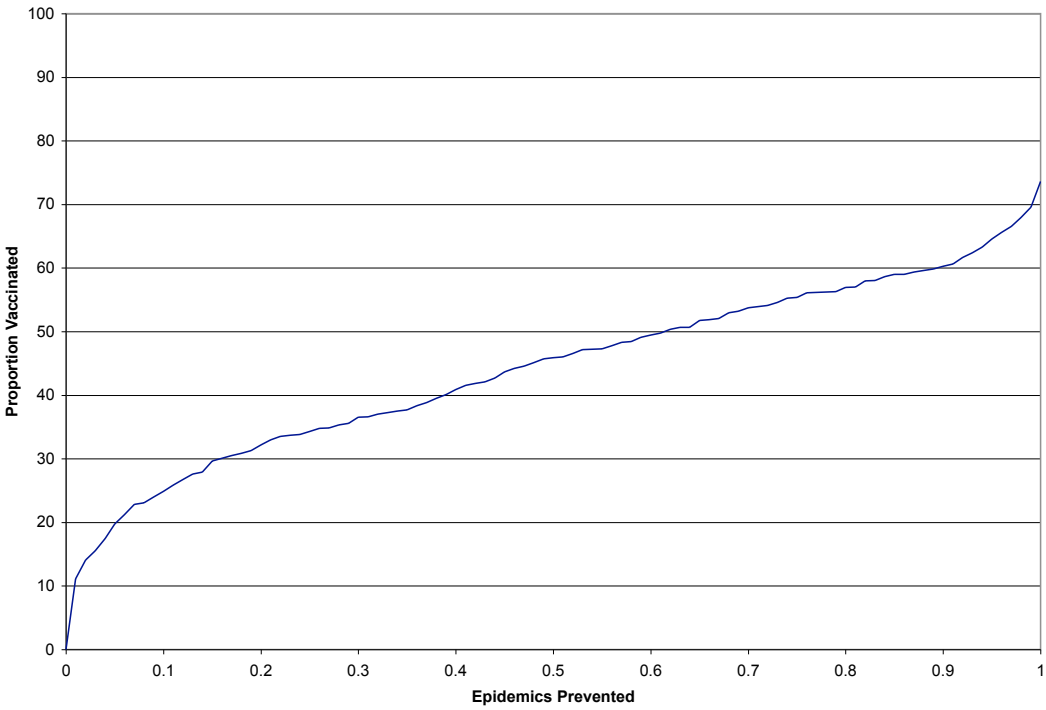

Figure 1: Plot of Vaccine Proportion vs Epidemic Prevention Rate

Test Instance	SPP	EEV	WS	VPI
vac500a	64.53	0.52	40.64	23.89
vac500b	65.49	0.57	40.38	25.11
vac500c	66.41	0.56	41.42	25.00
vac500d	66.63	0.56	40.90	25.73
vac500e	65.16	0.56	41.37	23.79
Average	65.54	0.55	40.94	24.70

Table 2: Computational Results

Table 3: List of Family Types and Frequency

Household Size	Children	Adults	Elderly	Frequency
1	0	1	0	0.05
1	0	0	1	0.05
2	0	2	0	0.10
2	0	0	2	0.05
2	1	1	0	0.08
2	0	1	1	0.02
3	1	2	0	0.10
3	0	2	1	0.05
3	0	0	3	0.05
3	1	0	2	0.05
3	0	3	0	0.05
4	2	2	0	0.03
4	3	1	0	0.03
4	0	2	2	0.03
4	0	4	0	0.03
4	0	0	4	0.03
5	3	2	0	0.03
5	2	2	1	0.03
5	0	5	0	0.02
5	0	0	5	0.02
6	4	2	0	0.01
6	0	6	0	0.01
6	0	0	6	0.01
6	3	2	1	0.01
7	2	2	2	0.01
7	5	2	0	0.01
7	0	7	0	0.01
7	0	0	7	0.01
7	4	2	1	0.01
7	3	2	2	0.01

Table 4: List of Parameters and Distributions

Parameter Name	Symbol	Distribution
vaccine efficacy	$\epsilon(\omega)$	truncated Normal(0.85, 0.32) in interval [0,1]
inter-household contact rate	$m(\omega)$	truncated Normal(1, 0.5) in interval $[0, \infty]$
intra-household spread rate	$b(\omega)$	truncated Normal(0.6, 0.32) in interval [0,1]
relative infectivity, person type t	$\mu_t(\omega)$	low value 0.7, $p = 0.5$, high value 1.3, $p = 0.5$
relative susceptibility, person type t	$\mu_t(\omega)$	low value 0.7, $p = 0.5$, high value 1.3, $p = 0.5$