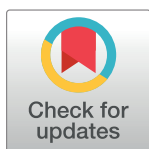RESEARCH ARTICLE

COVID-19 vaccination policies under uncertain transmission characteristics using stochastic programming

Krishna Reddy Gujjula⊕, Jiangyue Gong⊕, Brittany Segundo⊕, Lewis Ntaimo⊕⊕*

Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, College Station, Texas, United States of America

⊕ These authors contributed equally to this work.
* ntaimo@tamu.edu

Abstract

We develop a new stochastic programming methodology for determining optimal vaccination policies for a multi-community heterogeneous population. An optimal policy provides the minimum number of vaccinations required to drive post-vaccination reproduction number to below one at a desired reliability level. To generate a vaccination policy, the new method considers the uncertainty in COVID-19 related parameters such as efficacy of vaccines, age-related variation in susceptibility and infectivity to SARS-CoV-2, distribution of household composition in a community, and variation in human interactions. We report on a computational study of the new methodology on a set of neighboring U.S. counties to generate vaccination policies based on vaccine availability. The results show that to control outbreaks at least a certain percentage of the population should be vaccinated in each community based on pre-determined reliability levels. The study also reveals the vaccine sharing capability of the proposed approach among counties under limited vaccine availability. This work contributes a decision-making tool to aid public health agencies worldwide in the allocation of limited vaccines under uncertainty towards controlling epidemics through vaccinations.

Introduction

COVID-19 caused by the Severe Acute Respiratory Syndrome CoronaVirus (SARS-CoV-2) was declared a global pandemic by the World Health Organization in early 2020. The first reported outbreak occurred in Wuhan, China in December, 2019 and has spread to every region of the world [1]. To control the spread of COVID-19, governments have introduced vaccines and implemented a variety of non-pharmaceutical interventions such as social distancing, restrictions on gatherings, mask mandates, closures of businesses, religious institutions, and schools, travel restrictions and border closures, quarantining, and contact tracing [2–4]. In this paper, we consider a stochastic optimization methodology to determine optimal vaccination policies for a multi-community heterogeneous population to control the spread of the disease. The basic reproduction number, R_0, is used to measure infectious disease

community transmission and is defined as the average number of secondary infections caused by a primary case within a completely susceptible population [5, 6]. Under the impact of mitigation measures, the change in transmissibility of the disease over time is evaluated using the *effective reproduction number R_t*. It is the average number of secondary infections caused by a primary case at a given time t [7, 8]. R_t suggests that an outbreak will continue if it is greater than one and end if it has a value less than one.

In addition to the use of non-pharmaceutical mitigation measures, an effective vaccine can slow down the spread of the disease [9]. Mass vaccination campaigns reduce the susceptible population in a community and can be used to end an outbreak and mitigate any future outbreaks [10]. The reduction in a community's susceptible population decreases the number of interactions between infectious and susceptible persons, which in turn reduces R_t and will eventually drive R_t to be less than one [10]. Due to the limited availability of vaccines in certain parts of the world, developing an optimized vaccine allocation plan is critical. An optimal allocation would minimize the number of vaccinations required to ensure that $R_t \leq 1$. We believe this to be imperative under the following situations: 1) when vaccines are not widely available during the initial stages of distribution [11, 12]; 2) when governments want to reduce the implementation of non-vaccination type mitigation strategies which have a negative impact on socioeconomic activities [13, 14]; 3) when health authorities want to quickly reduce the mortality and morbidity due to COVID-19 in relatively susceptible age groups; 4) when there is vaccine hesitancy, i.e., when a portion of the population is reluctant to receive vaccines [15–17]; and 5) when the emergence of SARS-CoV-2 variants may reduce the efficacy of currently available vaccines.

Mathematical models have been developed to attempt to define an optimal vaccine allocation strategy and these can broadly be categorized as deterministic or stochastic. Deterministic models [18–21, e.g.] include dynamic and optimal control models and do not consider parameter uncertainty. These model are generally sensitive to the vaccine and epidemiological characteristics of the virus. COVID-19 disease characteristics are uncertain at best and studies report values that vary significantly [22, 23]. Thus, it is advantageous to use stochastic modeling approach [24] that account for the uncertainty. Estimating these parameter values, however, is challenging due to the complex nature of human interactions, emerging variants of the virus, as well as the cultural and demographic variability among different communities.

In this work, we build on the stochastic programming (SP) [25, 26] optimal vaccine allocation methodology proposed by [24], which extends the deterministic epidemiology model by [27]. That model was developed to find optimal vaccination policies for a community of households and we extend it to a multi-community framework that considers uncertainties in parameters related to the COVID-19 virus and its vaccines. The SP framework we take accounts for the uncertainty in the model parameters and provides solutions that hedge against unforeseen future scenarios. Unlike solutions obtained from deterministic models using point estimates for the parameters, SP solutions are in fact policies, i.e., decisions prescribed for a given state and/or level of reliability. We implement this multi-community model on a set of neighboring counties, i.e., a population center and its surrounding communities with a sparse population. Generally, the epidemic has a higher likelihood of occurring in a densely populated area due to the average person's higher number of social contacts [28]. Then the epidemic will eventually propagate to the surrounding communities as a result of inter-community social contacts [29]. The multi-community stochastic model informs a vaccine allocation policy that considers a set of communities together rather than as isolated entities.

The main contribution of this work is an SP based methodology for determining the minimum number of vaccines required to control COVID-19 outbreaks ($R_t \leq 1$) through a

vaccination campaign in a multi-community setting. At the core of this new methodology is a stochastic disease spread model for an age-based heterogeneous population that considers uncertainty in parameters related to vaccine efficacy, disease transmission characteristics, and human interactions. The model also takes into account demographic variations such as household types and age distribution of the communities to decide optimal allocation of vaccines. In addition to determining optimal vaccine policies, the new methodology addresses potential vaccine shortages and the benefits of vaccine resource sharing among neighboring communities. Another contribution of this work is a computational study based on a real setting that illustrates how the results of this can be used to guide health officials in mitigating epidemics.

The rest of this paper is organized as follows: We derive SP models for vaccine allocation for multiple communities in the next section. We then present population datasets and the uncertain parameters used in the model. Next, we report and discuss the results for two cases: unlimited and limited vaccine availability. We end the paper with concluding remarks and directions for future work.

Materials and methods

We consider a model of disease transmission in a community based on the work of Becker and Starczak [27] and Tanner et al. [24]. The former work derive a deterministic model of disease transmission which the latter extends to the stochastic setting, where disease transmission parameters are uncertain. Both models consider a single community of households. However, this work extends this approach to multiple communities in a stochastic setting with age-based heterogeneous populations. The aim of a vaccination strategy is to prevent epidemics by immunizing a sufficiently large number of members of a community to force R_t to be less than one. The proportion of individuals in each household that must be vaccinated to prevent an epidemic depends on, among other things, the distribution of household sizes and how the vaccines are allocated to households. In this work, we are interested in determining an optimal strategy for vaccinating members of the community based on household size, given a finite amount of vaccine doses. A vaccination policy or strategy provides *critical vaccination coverage* that reduces R_t to one or less. The term "vaccination coverage" refers to the proportion of individuals who are vaccinated. Our objective is to identify the minimal vaccination coverage.

The critical vaccination coverage for a given vaccination strategy is typically based on the *effective reproduction number* for infected individuals, which is the average number of secondary cases generated by an infected person. We consider the effective reproduction number for infected households in a community under vaccination, which is denoted R_{HVc} and called the *post-vaccination reproduction number*, as defined by [27]. The aim of a vaccination strategy is to keep $R_{HVc} \leq 1$ to ensure that introductions of the disease do not lead to an epidemic. Therefore, we want to compute the vaccine coverage required so that the vaccine induced herd immunity is at a sufficiently high level to prevent epidemics. We define the nomenclature we use in our mathematical model in Table 1.

Assuming a *homogeneous* population model in which we assume there are no significant age-related differences in susceptibility and infectivity of individuals within the communities, Becker and Starczak's [27] define the expression for R_{HVc} for a deterministic model of disease spread as follows: Given x_{nvc} and the proportion of n-sized households in which vaccination policy v has been implemented, R_{HVc} for a community c is expressed as

$$R_{HVc} = \sum_{n \in \mathbb{N}} \sum_{v \in \mathbb{V}} a_{nvc} x_{nvc}. \tag{1}$$

Table 1. Notation used in defining the models.

	Sets and Indices
$\mathbb{C}$	Set of communities, element $c \in \mathbb{C}$.
$\mathbb{N}$	Set of household types, element $n \in \mathbb{N}$.
$\mathbb{K}$	Set of person groups, element $k \in \mathbb{K}$.
$\mathbb{V}$	Set of vaccination policies, element $v \in \mathbb{V}$.
Ω_c	Set of outcomes (scenarios) for community $c \in \mathbb{C}$, element $\omega_c \in \Omega_c$.
	Parameters
$R_{HVc}(\tilde{\omega}_c)$	Uncertainty post-vaccination reproduction number for community $c \in \mathbb{C}$.
$\tilde{\omega}_c$	Multivariate random variable whose outcome is $\omega_c \in \Omega_c$ with probability of occurrence p_{ω_c}; describes the uncertain parameters for R_{HVc}.
$a_{nvc}(\tilde{\omega}_c)$	Uncertain R_{HVc} parameter that captures the impact of vaccination policy $v \in \mathbb{V}$ in a type n household in community $c \in \mathbb{C}$.
$m_c(\tilde{\omega}_c)$	Uncertain number of close contacts that an infective makes on average with persons from other household in the course of his/her infectious period in a community $c \in \mathbb{C}$.
H_c	Number of households in community $c \in \mathbb{C}$.
$p(n)$	Number of persons in a household of type n.
$f(v)$	Number of persons vaccinated in a household when vaccination policy $v \in \mathbb{V}$ is implemented.
h_{nc}	Proportion of type n households in community $c \in \mathbb{C}$.
μ_c	Average household size in a community, $\mu_c = \sum_{n \in \mathbb{N}} p(n) h_{nc}$.
π_{kc}	Proportion of group $k \in \mathbb{K}$ persons in community $c \in \mathbb{C}$.
$b(\tilde{\omega}_c)$	Uncertain transmission rate within a household.
$\beta_{kc}(\tilde{\omega}_c)$	Uncertain susceptibility for type $k \in \mathbb{K}$ person in community $c \in \mathbb{C}$.
$\lambda_{kc}(\tilde{\omega}_c)$	Uncertain infectivity for type $k \in \mathbb{K}$ person in community $c \in \mathbb{C}$.
$\epsilon(\tilde{\omega}_c)$	Uncertain vaccine efficacy.
V	Total number of available vaccines.
α_c	User-set model reliability level for community $c \in \mathbb{C}$.
$\bar{\alpha}_c^e$	Excess allowed on model reliability level for community $c \in \mathbb{C}$.
$\bar{\alpha}_c^s$	Deficit allowed on model reliability level for community $c \in \mathbb{C}$.
M_e, M_s	Sufficiently large numbers.
	Decision Variables
x_{nvc}	Proportion of n sized households with vaccination policy $v \in \mathbb{V}$ implemented in community $c \in \mathbb{C}$.
α_c^e	Excess amount above reliability level α_c level for community $c \in \mathbb{C}$.
α_c^s	Deficit amount below reliability level α_c level for community $c \in \mathbb{C}$.

The parameter a_{nvc} for a homogeneous population of communities is defined as given by

$$a_{nvc} = \frac{m_c h_{nc}}{\mu_c}[(1-b)(p(n)-f(v)\epsilon) + b(p(n)-f(v)\epsilon)^2 + bf(v)\epsilon(1-\epsilon)], \qquad (2)$$

where b is the transmission proportion within a household. When $b = 0$ it means that there is no transmission within a household, whereas when $b = 1$ it means that there is complete transmission in an infected household. Defining b in this manner allows us to capture the continuum of transmission rates within households. Next, we give a derivation of the model and provide an extension to the stochastic setting.

The reproduction number for infected households, R_H, is given by

$$R_H = m_c v_c, \qquad (3)$$

where v_c is the mean size of the household outbreak in community c when a randomly selected previously uninfected individual has a close contact with an infective. This model can be traced back to Bartoszyński [6]. Let $\bar{h}_{nsc}$ denote the mean size of an outbreak in household of size n with s susceptible members when the disease is introduced by one of the susceptible being infected from outside the household. Then the probability that the individual contacted is one the susceptibles of the households is s/n.

Now let π_{nc} denote the proportion of individuals who belong to a house of size n. Also, let η_{nsc} be the proportion of households of size n of whom exactly s are initially susceptible. Then we have that

$$v_c = \sum_n \pi_{nc} \sum_{s=1}^{n} \frac{s}{n} \eta_{nsc} \bar{h}_{nsc}. \tag{4}$$

Furthermore, $\pi_{nc} = nh_{nc}/\mu_{Nc}$, where μ_{Nc} is the mean household size. The reproduction number for infected households is

$$R_H = \frac{m_c}{\mu_{Nc}} \sum_n h_{nc} \sum_{s=1}^{n} s\eta_{nsc} \bar{h}_{nsc}. \tag{5}$$

When there is no immunity $s = n$ and the basic reproduction number, R_{HO}, is given as

$$R_{HO} = \frac{m_c}{\mu_{Nc}} \sum_n nh_{nc} v_{nnc}. $$

To prevent epidemics without vaccinations $R_{HO} \leq 1$. In this study, we assume that $R_{HO} > 1$ and we want to use vaccinations to bring R_{HO} below one. Each vaccination, however, provides immunity with probability ϵ, which is the vaccine efficacy. Letting x_{nvc} be the proportion of n sized households with vaccination policy v (v members vaccinated) implemented in community c, then we have that

$$\eta_{nsc} = \sum_{v=n-s}^{n} \binom{v}{n-s} \epsilon^{n-s} (1-\epsilon)^{v-(n-s)} x_{nvc}. \tag{6}$$

Now consider Eqs 5 and 6, and define

$$a_{nvc} = \frac{m_c h_{nc}}{\mu_{Nc}} \sum_{s=n-v}^{n} s \binom{v}{n-s} \epsilon^{n-s} (1-\epsilon)^{v-(n-s)} \bar{h}_{nsc}. \tag{7}$$

Then

$$R_H = \sum_n \sum_{v=0}^{n} a_{nvc} x_{nvc}. \tag{8}$$

Assuming that $\bar{h}_{nsc} = 1 - b + bs$, $b \in [0, 1]$, where $b = 0$ corresponds to no disease transmission within households and $b = 1$ corresponds to total infection, the expression for a_{nvc} can be

written as follows:

$$
\begin{aligned}
a_{nvc} &= \frac{m_c h_{nc}}{\mu_{Nc}} \sum_{s=n-v}^{n} s \binom{v}{n-s} \epsilon^{n-s} (1-\epsilon)^{v-(n-s)} (1-b+bs) \\
&= \frac{m_c h_{nc}}{\mu_{Nc}} \Big[(1-b) \sum_{s=n-v}^{n} s \binom{v}{n-s} \epsilon^{n-s} (1-\epsilon)^{v-(n-s)} + \\
&\quad b \sum_{s=n-v}^{n} s^2 \binom{v}{n-s} \epsilon^{n-s} (1-\epsilon)^{v-(n-s)} \Big] \\
&= \frac{m_c h_{nc}}{\mu_{Nc}} \Big[(1-b)(n-v\epsilon) + b(n-v\epsilon)^2 + bv\epsilon(1-\epsilon) \Big],
\end{aligned}
\tag{9}
$$

where the last expression comes from applying the binomial theorem.

In this study, we extend the expression for a_{nvc} in Becker and Starczak's model [27] to the *heterogeneous* and *stochastic setting*, i.e., where disease spread affects specific age groups differently and the disease spread and vaccine parameters are assumed to be unknown. Therefore, we model a_{nvc} as a random variable, $a_{nvc}(\tilde{\omega}_c)$, in which an outcome (scenario) ω_c of $\tilde{\omega}_c$ is given by the triple, $\omega_c := \{m_c(\omega_c), b(\omega_c), \epsilon(\omega_c)\}$, with probability of occurrence p_{ω_c}. Consequently, we have that

$$
\begin{aligned}
a_{nvc}(\tilde{\omega}_c) &= \frac{m_c(\tilde{\omega}_c) h_{nc}}{\mu_c} \big[(1-b(\tilde{\omega}_c))(p(n) - f(v)\epsilon(\tilde{\omega}_c)) + b(\tilde{\omega}_c)(p(n) - f(v)\epsilon(\tilde{\omega}_c))^2 \\
&\quad + b(\tilde{\omega}_c) f(v) \epsilon(\tilde{\omega}_c)(1 - \epsilon(\tilde{\omega}_c)) \big].
\end{aligned}
\tag{10}
$$

Therefore, R_{HVc} is also a random variable and is denoted $R_{HVc}(\tilde{\omega}_c)$ and is expressed as

$$
R_{HVc}(\tilde{\omega}_c) = \sum_{n \in \mathbb{N}} \sum_{v \in \mathbb{V}} a_{nvc}(\tilde{\omega}_c) x_{nvc}.
\tag{11}
$$

Heterogeneous population model

In the *heterogeneous population* model, we assume that there are significant age-related differences in the susceptibility and infectivity of individuals in all the communities involved. To capture these differences, we define a set of groups of people $\mathbb{K}$ in which susceptibility and infectivity are differentiated by age. We denote the susceptibility and infectivity of group k in community c, respectively, by $\beta_{kc}(\tilde{\omega}_c)$ and $\lambda_{kc}(\tilde{\omega}_c)$. We define three age groups, A, B, and C, as follows: $A = (age \leq 19)$, $B = (20 \leq age \leq 64)$, and $C = (age \geq 65)$. The number of age groups can be increased based on the real setting. For each household of type n with $p(n)$ members, we must specify the number of persons, $p_k(n)$ belonging to each of the *three* age groups A, B and C. The possible vaccination policies for a type n household are represented by $(f_A(v), f_B(v), f_C(v))$, the number of household members vaccinated in group A, B, and C, respectively. An example illustration of $p(n)$ values of 1 and 2 is shown in Table 2.

The heterogeneous model requires know the number of members in a household and to what age group each one belongs. The second column $p(n)$ in Table 2 gives the household size. For example, for $p(n) = 2$ household type $n = 4, 5, 6, 7, 8, 9$. Notice that each household type has different age compositions. For instance, for household type $n = 4$ household composition is (2, 0, 0). This means that there are two individuals in age group A and zero individuals in age groups B and C. For a household type $n = 3$, the household composition is (0, 2, 0), i.e., there are no individuals in age group A, two in age group B, and none in age group C.

Table 2. Example household types and vaccination policies under heterogeneous population for $p(n) = 1$ and $p(n) = 2$.

Household Type	Household Size	Household Composition	Total vaccination policies	Possible vaccination policies for a type n Household
n	$p(n)$	$(p_A(n), p_B(n), p_C(n))$	$(p_A(n) + 1)^* (p_B(n) + 1)^* (p_C(n) + 1)$	$(n, f_A(v), f_B(v), f_C(v))$
1	1	(1, 0, 0)	2	(1, 0, 0, 0), (1, 1, 0, 0)
2	1	(0, 1, 0)	2	(2, 0, 0, 0), (2, 0, 1, 0)
3	1	(0, 0, 1)	2	(3, 0, 0, 0), (3, 0, 0, 1)
4	2	(2, 0, 0)	3	(4, 0, 0, 0), (4, 1, 0, 0), (4, 2, 0, 0)
5	2	(0, 2, 0)	3	(5, 0, 0, 0), (5, 0, 1, 0), (5, 0, 2, 0)
6	2	(0, 0, 2)	3	(6, 0, 0, 0), (6, 0, 0, 1), (6, 0, 0, 2)
7	2	(1, 1, 0)	4	(7, 0, 0, 0), (7, 0, 1, 0), (7, 1, 0, 0), (7, 1, 1, 0)
8	2	(0, 1, 1)	4	(8, 0, 0, 0), (8, 0, 0, 1), (8, 0, 1, 0), (8, 0, 1, 1)
9	2	(1, 0, 1)	4	(9, 0, 0, 0), (9, 0, 0, 1), (9, 1, 0, 0), (9, 1, 0, 1)

Given the proportion of type n household with v vaccinated members, x_{nvc}, the post-vaccination reproduction number $R_{HVc}(\tilde{\omega}_c)$ for some community c is given by Eq (11). Under the assumption of heterogeneity, the explicit expression for $R_{HVc}(\tilde{\omega}_c)$ considers the age-stratified groups. Consequently, the uncertain parameter $a_{nvc}(\tilde{\omega}_c)$ for a heterogeneous population of communities can be defined as follows:

$$a_{nvc}(\tilde{\omega}_c) =$$
$$\frac{m_c(\tilde{\omega}_c)h_{nc}}{\mu_c}\left\{ \sum_{k\in\mathbb{K}}\beta_{kc}(\tilde{\omega}_c)\lambda_k(\tilde{\omega}_c)[(1 - b(\tilde{\omega}_c))(p_k(n) - f_k(v)\epsilon(\tilde{\omega}_c)) + \right.$$
$$b(\tilde{\omega}_c)f_k(v)\epsilon(\tilde{\omega}_c)(1 - \epsilon(\tilde{\omega}_c)) +$$
$$\left. b(\tilde{\omega}_c)\sum_{k\in\mathbb{K}}\sum_{r\in\mathbb{K}}\beta_{kc}(\tilde{\omega}_c)\lambda_{rc}(\tilde{\omega}_c)(p_k(n) - f_k(v)\epsilon(\tilde{\omega}_c))(p_r(n) - f_r(v)\epsilon(\tilde{\omega}_c)) \right\},$$

(12)

where $\sum_{k\in\mathbb{K}}\beta_{kc}(\tilde{\omega}_c)\pi_{kc} = 1$ and $\sum_{k\in\mathbb{K}}\lambda_{kc}(\tilde{\omega}_c)\pi_{kc} = 1$ for each community c, which restricts the scale of susceptibility and infectivity.

The impact on vaccine policies caused by different age compositions is captured in Eq 12. The interaction among members of the same age group is calculated using

$$\sum_{k\in\mathbb{K}}\beta_{kc}(\tilde{\omega}_c)\lambda_k(\tilde{\omega}_c)[(1 - b(\tilde{\omega}_c))(p_k(n) - f_k(v)\epsilon(\tilde{\omega}_c))b(\tilde{\omega}_c)f_k(v)\epsilon(\tilde{\omega}_c)(1 - \epsilon(\tilde{\omega}_c))],$$

whereas the interactions of different age groups is captured calculated using

$$\sum_{k\in\mathbb{K}}\sum_{r\in\mathbb{K}}\beta_{kc}(\tilde{\omega}_c)\lambda_{rc}(\tilde{\omega}_c)(p_k(n) - f_k(v)\epsilon(\tilde{\omega}_c))(p_r(n) - f_r(v)\epsilon(\tilde{\omega}_c)).$$

We should point out that under the homogeneity assumption in population, i.e, $\beta_{kc}(\tilde{\omega}_c) = 1$ and $\lambda_{kc}(\tilde{\omega}_c) = 1$ for all k, the parameter $a_{nvc}(\tilde{\omega}_c)$ reduces to:

$$a_{nvc}(\tilde{\omega}_c) = \frac{m_c(\tilde{\omega}_c)h_{nc}}{\mu_c}[(1 - b(\tilde{\omega}_c))(p(n) - f(v)\epsilon(\tilde{\omega}_c)) + b(\tilde{\omega}_c)(p(n) - f(v)\epsilon(\tilde{\omega}_c))^2$$
$$+ b(\tilde{\omega}_c)f(v)\epsilon(\tilde{\omega}_c)(1 - \epsilon(\tilde{\omega}_c))],$$

(13)

which was derived earlier in Eq (2). The scenario $\omega_c \in \Omega_c$ of $\tilde{\omega}_c$ specifies the quintuple, $\omega_c := \{m_c(\omega_c), b(\omega_c), \epsilon(\omega_c), \beta_{kc}(\omega_c), \lambda_{kc}(\omega_c)\}$, with probability p_{ω_c}.

Computing the post-vaccination reproduction number $R_{HVc}(\tilde{\omega}_c)$ requires several parameters. Due to the uncertainty in the parameters, not all vaccination coverage choices will necessarily bring R_{HVc} below one. To accommodate the fact that in reality R_{HVc} will often exceed one under certain scenarios, we use the chance (probabilistic) constraints SP [30, 31] approach to impose a chance constraint on $R_{HVc} \leq 1$ for each community. A chance constraint requires specifying a *reliability level* $\alpha \in (0.5, 1)$, which sets the minimum total probability of scenarios that must be satisfied to achieve $R_{HVc} \leq 1$. Mathematically, this is expressed as

$$\mathbb{P}\{R_{HVc} \leq 1\} \geq \alpha,$$

meaning that $R_{HVc} \leq 1$ holds at least $\alpha \times 100\%$ amount of the time, but not necessarily for every scenario. In other words, the constraint can be violated at most $(1 - \alpha) \times 100\%$ of the time, such as when a certain vaccination policy is ineffective and results in an epidemic.

We are now ready to define the minimum coverage problem using chance constraints SP. We start with a basic formulation in which we assume there is an unlimited number of vaccines available. For some specified reliability level $\alpha_c \in (0.5, 1)$, the basic optimal vaccine allocation problem for a heterogeneous population can be formally stated as follows:

$$\text{Min} \quad \sum_{c\in\mathbb{C}} \sum_{n\in\mathbb{N}} \sum_{v\in\mathbb{V}} f(v) h_{nc} x_{nvc} \tag{14a}$$

$$\mathbb{P}\left\{ \sum_{n\in\mathbb{N}} \sum_{v\in\mathbb{V}} a_{nvc}(\tilde{\omega}_c) x_{nvc} \leq 1 \right\} \geq \alpha_c, \quad \forall c \in \mathbb{C} \tag{14b}$$

$$\sum_{v\in\mathbb{V}} x_{nvc} = 1, \quad \forall n \in \mathbb{N}; \quad \forall c \in \mathbb{C} \tag{14c}$$

$$x_{nvc} \geq 0, \quad \forall n \in \mathbb{N}; \quad \forall v \in \mathbb{V}; \quad \forall c \in \mathbb{C}. \tag{14d}$$

The objective function (14a) determines the minimum vaccination coverage across communities. Constraints (14b) are comprised of the chance constraints requiring that $R_{HVc}(\tilde{\omega}_c) \leq 1$ to prevent an epidemic for each community at least α_c amount of the time. This reliability level corresponds to the decision-maker's risk in the sense that this constraint cannot be violated more than $1 - \alpha_c$ of the time; for α_c proportion of scenarios, epidemics will be prevented. This violation is inevitable if, for example, the vaccine efficacy is not sufficiently large. The outcome (scenario) ω_c that determines the value $a_{nvc}(\omega_c)$ of the random parameter $a_{nvc}(\tilde{\omega}_c)$ will depend on the distribution of $\tilde{\omega}_c$. This problem, then, is generally a nonconvex problem and difficult to solve. However, if $\tilde{\omega}_c$ is discretely distributed with a finite number of outcomes, the problem can be reformulated as a deterministic equivalent problem using mixed-integer programming (MIP) and then solved using MIP techniques. Furthermore, this basic model is separable, which allows each problem to be solved for each community separately. Constraints (14c) determine the proportion of persons to vaccinate for each household size in each community. Finally, constraints (14d) are non-negativity restrictions.

We extend heterogeneous model ([14]) to the case with limited vaccine availability V and allowing for deviations from the specified α_c level for each community as follows:

$$\text{Min} \quad \sum_{c \in \mathbb{C}} \sum_{n \in \mathbb{N}} \sum_{v \in \mathbb{V}} f(v) h_{nc} x_{nvc} - \sum_{c \in \mathbb{C}} M_e \alpha_c^e + \sum_{c \in \mathbb{C}} M_s \alpha_c^s \tag{15a}$$

$$\mathbb{P} \left\{ \sum_{n \in \mathbb{N}} \sum_{v \in \mathbb{V}} a_{nvc}(\tilde{\omega}_c) x_{nvc} \leq 1 \right\} - \alpha_c^e + \alpha_c^s \geq \alpha_c, \quad \forall c \in \mathbb{C} \tag{15b}$$

$$\sum_{v \in \mathbb{V}} x_{nvc} = 1, \quad \forall n \in \mathbb{N}; \quad \forall c \in \mathbb{C} \tag{15c}$$

$$\sum_{c \in \mathbb{C}} \sum_{n \in \mathbb{N}} \sum_{v \in \mathbb{V}} H f(v) h_{nc} x_{nvc} \leq V \tag{15d}$$

$$\alpha_c^e \leq 1 - \alpha_c, \quad \forall c \in \mathbb{C} \tag{15e}$$

$$\alpha_c^e \leq \bar{\alpha}_c^e, \quad \forall c \in \mathbb{C} \tag{15f}$$

$$\alpha_c^s \leq \bar{\alpha}_c^s, \quad \forall c \in \mathbb{C} \tag{15g}$$

$$x_{nvc}, \alpha_c^e, \alpha_c^s \geq 0, \quad \forall n \in \mathbb{N}; \quad \forall v \in \mathbb{V}; \quad \forall c \in \mathbb{C}. \tag{15h}$$

The objective function ([15a]) determines the minimum vaccination coverage while adjusting for the deviation above and below the specified reliability levels for each community. The chance constraints ([15b]) now include deviation variables α_c^e and α_c^s to adjust the reliability level for each community. The decision-maker's risk is such that the constraint $R_{HVc} = \sum_{n=1}^{N} \sum_{v \in \mathbb{V}} a_{nvc}(\tilde{\omega}_c) x_{nvc} \leq 1$ can be violated in no more than $1 - (\alpha_c + \alpha_c^e - \alpha_c^s)$ proportion of scenarios. Constraints ([15c]) remain as defined before, while constraint ([15d]) is added to ensure that the total number of vaccines allocated does not exceed the total number of vaccines available for all communities. This constraint links all the communities and is therefore a complicating constraint, which means that the problem is no longer separable. Constraints ([15e]) ensure that the deviation above α_c does not exceed the allowable amount $1 - \alpha_c$ for each community. Constraints ([15g]) and ([15f]) limit the deviations based on the user-specified bounds α_c^e and α_c^s, respectively. Finally, constraints ([15h]) are non-negativity restrictions on all the decision variables.

Model parameters

We are now ready to describe the parameters used in the two models. The communities are characterized by the distribution of household types within the community, while the stochastic parameters are represented by discrete probability distributions. For the rest of the model parameters, we created discrete distributions based on the information available for COVID-19 transmission characteristics, historical values for the effective reproduction number, and the advertised efficacy of approved vaccines.

Demographic data. We implemented the SP models using data for a multi-community setting comprising seven neighboring counties in the state of Texas, namely; Travis, Williamson, Bastrop, Caldwell, Hays, Burnet, and Blanco. In the model, communities are defined by a multivariate discrete distribution of different household types, which is defined by the size of
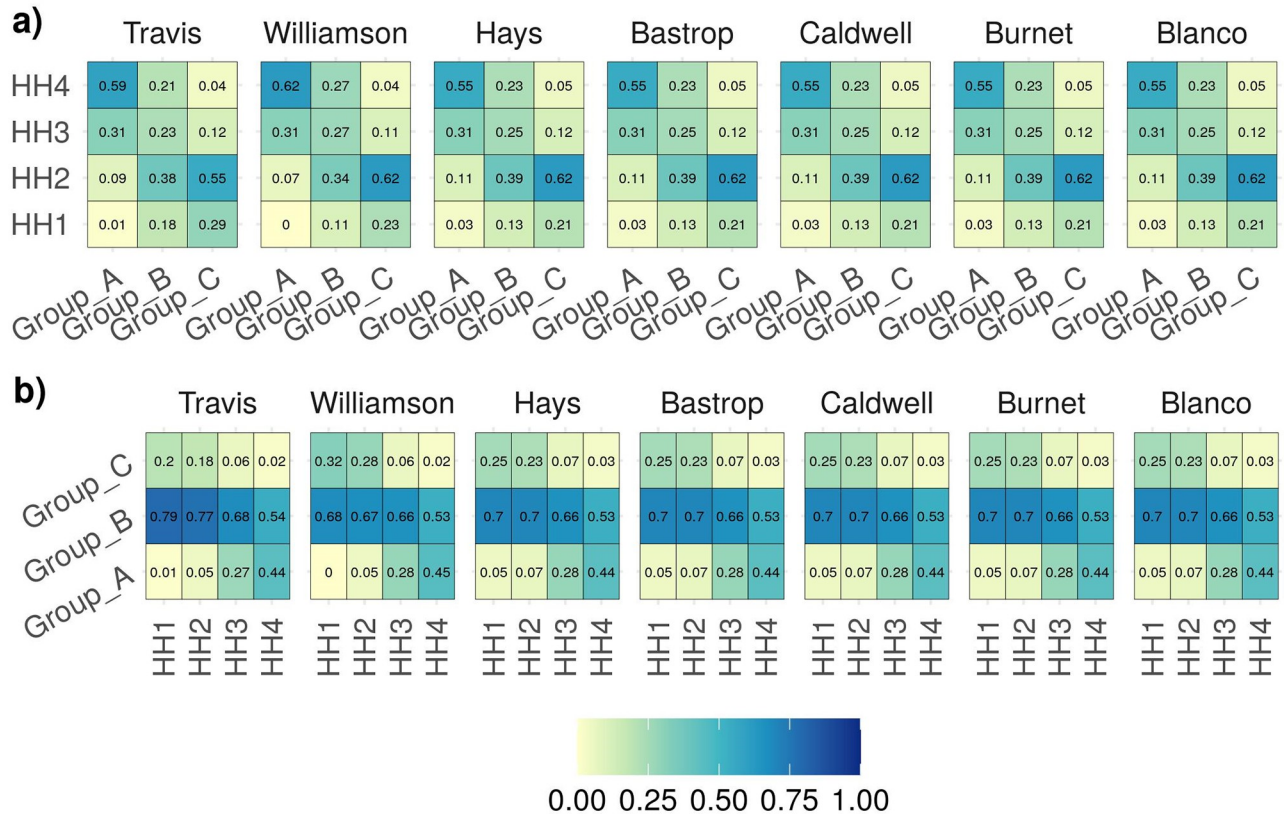
Fig 1. This figure shows the demographic distribution for each county. Fig **a)** shows the heatmap of household sizes in which a particular age group resides. We observe that the majority of the younger population, Group A, tends to belong to larger households along with members of Groups B and C, whereas higher proportion of Group B and C population occupy smaller household of size of one and two. Fig **b)** shows the heatmap of age groups residing in each household size. We observe that smaller households are comprised predominantly of members of Group B population followed by Group C population, and larger households tend to include members of Groups A and B.

https://doi.org/10.1371/journal.pone.0270524.g001

the household and the number of household members in each age group *A*, *B*, and *C*. We consider household sizes of one to seven and three age groups: *A*: *age* ≤ 19*yrs*., *B*: 20 ≤ *age* ≤ 64*yrs*., and *C*: *age* ≥ 65*yrs*. These age groups are defined on the basis of variation in the effect of COVID-19 on different age groups and can be expanded to include more refined age groups as more information on susceptibility and infectivity for different age groups becomes available. The household size distribution for each of the seven counties was downloaded from 5-year American Survey data (ACS) from (https://data.census.gov/cedsci/) for years 2014–2018 [32]. The household type distribution data was downloaded from https://usa.ipums.org/usa/ [33]. For each county, IPUMS provides a down sampled data along with the weights for each data points. Using the weights, the household type data is scaled up to represent the complete household type distribution for each county. The IPUMS database contains the household data for Travis, Williamson, and Hays counties; for the remaining counties, we assume that the household distribution is similar to that of Hays County. The demographic distribution data utilized in the model is available in S1 File. Fig 1 shows the heat map of the U.S. census demographics data for the different household sizes and age groups.

Household transmission rate $b(\tilde{\omega}_c)$**.**   This parameter captures the transmissibility of a contagion within a household in a communuty *c*, where $0 \leq b(\tilde{\omega}_c) \leq 1$. The extreme value $b(\tilde{\omega}_c) = 0$ is equivalent to no disease transmission within a household, and $b(\tilde{\omega}_c) = 1$ would

Table 3. The discrete probability distribution represents the within-household transmission rate (b) used in this study.

Within-household transmission rate $b(\tilde{\omega}_c)$	0.40	0.30	0.20	0.10
Probability	0.10	0.40	0.40	0.10

https://doi.org/10.1371/journal.pone.0270524.t003

mean that all members of the household become infected [27]. This parameter is analogous to household Secondary Attack Rate (SAR), which is defined as the probability that an infected individual will transmit the disease to a susceptible individual [34]. According to various studies, the estimated household SAR varies from 3.9% to 36.4%; when pooled the estimate is 17.1% with a confidence interval (CI) of [13.7%, 21.2%] [35]. Another review estimates that the household SAR value ranges from 4.6% to 49.56% [36]. Based on these values, we generated a discrete distribution (see Table 3) for the within-household transmission rate, $b(\tilde{\omega}_c)$.

Vaccine efficacy $\epsilon(\tilde{\omega}_c)$. Beginning in December 2020, multiple vaccines were approved for emergency use, and additional vaccines are undergoing vaccine trials and approval. Two of the vaccines approved by the FDA were Pfizer-BioNTech and Moderna, with efficacies of around 95% in clinical trials, and the others vaccines were AstraZeneca-University of Oxford, Johnson & Johnson, and Novavax, which have reported efficacies that range from 60% to 85% [37]. The vaccine efficacy also varies as a result of variability in real-world conditions, such as how the vaccine is transported, how the vaccine is administered, and the medical condition of vaccinated person. Other important factors that may affect the effectiveness of these vaccines include the emergence of new and evasive variants of SARS-CoV-2 and age of the person receiving the vaccine. To complete mass vaccination campaigns, we will need to use multiple vaccine candidates which have different reported and actual efficacies. Under such consideration, a discrete distribution representing vaccine efficacy (see Table 4) is used in the model.

Relative susceptibility $\beta(\tilde{\omega}_c)$. The heterogeneous model assumes age-related differences in susceptibility to COVID-19. *Relative susceptibility* quantifies the variation in susceptibility due to biological susceptibility and social mixing between individuals in different age groups. There is age-dependent variation in susceptibility to COVID-19; studies showed elevated susceptibility in adults over 65 years old and generally lower in the younger population [38, 39]. For each county, we generated three levels of relative susceptibility with associated probabilities (see S1 File). The data for Travis county is shown in Table 5.

Relative infectivity $\lambda(\tilde{\omega}_c)$. Variation in infectiousness between infected individuals due to differences is biological infectivity and in social mixing between individuals in different age

Table 4. The discrete probability distribution represents vaccine efficacy ($b(\tilde{\omega}_c)$) used in this study.

Vaccine efficacy $\epsilon(\tilde{\omega}_c)$	0.95	0.90	0.80	0.60
Probability	0.20	0.30	0.35	0.15

https://doi.org/10.1371/journal.pone.0270524.t004

Table 5. Discrete probability distribution of relative susceptibility $\beta(\tilde{\omega}_c)$ used in this study for Travis county.

Travis county	Age group A	Age group B	Age group C	Scale constraints
Population proportion π_k	0.26	0.64	0.10	$\Sigma_{k\in K}\,\pi_k = 1.00$
Probability	β_A	β_B	β_C	$\Sigma_{k\in K}\,\beta_k\pi_k$
0.50	0.66	1.00	1.91	1.00
0.50	0.80	1.00	1.52	1.00

https://doi.org/10.1371/journal.pone.0270524.t005

Table 6. Discrete probability distribution of relative infectivity $\lambda(\tilde{\omega}_c)$ used in this study for Travis county.

Travis county	Age group A	Age group B	Age group C	Scale constraints
Population proportion π_k	0.26	0.64	0.10	$\sum_{k \in K} \pi_k = 1.00$
Probability (Case 1)	λ_A	λ_B	λ_C	$\sum_{k \in K} \lambda_k \pi_k$
0.50	0.95	1.00	1.13	1.00
0.50	0.90	1.00	1.26	1.00
Probability (Case 2)	λ_A	λ_B	λ_C	$\sum_{k \in K} \lambda_k \pi_k$
0.50	1.10	1.00	0.74	1.00
0.50	1.15	1.00	0.61	1.00

https://doi.org/10.1371/journal.pone.0270524.t006

groups is characterized by *relative infectivity*. Goldstein et al. [39] surveyed multiple studies and suggest that there is little evidence that relative infectivity of older age groups is slightly higher than younger population. For infectivity, we assume two cases: *Case 1*. The younger population has lower infectivity compared to the older population (the older population's biological infectivity outweighs the higher social mixing of the younger population); and *Case 2*. The younger population has higher infectivity compared to the older population (the higher social mixing of the younger population outweighs the older population's biological infectivity). This nuance is important to the transmission of COVID-19, because the younger population generally has more human interactions [40] and do not develop severe symptoms as compared to older populations. A member of the younger population, then, is more likely to infect a susceptible person. For each county, we generated two levels of relative infectivity with associated probabilities (see S1 File). The data for Travis county is given in Table 6.

Outside household close contact $m(\tilde{\omega}_c)$. An important parameter in the model is $m(\tilde{\omega}_c)$, it is defined as the number of close contacts that an infective makes with persons of other households and close contact being sufficient for transmitting the disease when the contact is with a susceptible person. We should point out that $m(\tilde{\omega}_c)$ is variable due to differences in human interactions under the impact of various mitigation measures and demographics of a community. To estimate the distribution of $m(\tilde{\omega}_c)$, we collected historical time series data for effective reproduction number R_t for Travis county from University of Austin, Texas, COVID-19 dashboard [41]. Using the value of R_t and $a_{nvc}(\tilde{\omega}_c)$ in Eq 1 we estimate the value of $m(\tilde{\omega}_c)$. Note that in Eq 1, R_{HVc} is analogous to R_t, $a_{nvc}(\tilde{\omega}_c)$ is calculated for homogeneous population, under no vaccination, i.e, vaccine efficacy, $\epsilon(\tilde{\omega}_c) = 0$, and point estimate of within household transmission rate, $b(\tilde{\omega}_c) = 0.3$. Probability associated with the estimated value of m is the proportion of time period the value of R_t was observed. R_t values observed in Travis County for the region of Austin metropolitan area varied from 0.5 to 3.5 and for remaining counties from their respective historically observed *effective reproduction number*, R_t. The discrete distribution for $m(\tilde{\omega}_c)$ is available in S1 File.

Reliability level α. Typically health officials prescribe acceptable reliability levels based on the historical severity of the epidemic. For this study, we use three levels for reliability: *Low*, *Medium* and *High* (see Table 7). Note that at the highest level of reliability, Travis county has the largest reliability number of 0.990 while the surrounding counties have lower reliability

Table 7. Reliability levels for each community used in this study.

Reliability Level α	Travis	Williamson	Hays	Bastrop	Caldwell	Burnet	Blanco
High	0.990	0.980	0.970	0.970	0.955	0.955	0.955
Medium	0.985	0.975	0.965	0.965	0.950	0.950	0.950
Low	0.980	0.970	0.960	0.960	0.945	0.945	0.945

https://doi.org/10.1371/journal.pone.0270524.t007

numbers with lowest being 0.955 for Burnet and Blanco counties. These values are set assuming that the epidemic outbreak is more severe at the population centers than in the sparsely populated counties.

Results

The stochastic models were implemented in the CPLEX 12.9 Callable Library [42] using C++ and solved using a set of predetermined levels of reliability α along with discrete distributions for the model parameters vaccine efficacy, within-household transmission rate, county level household size distribution, and outside-household close contact rate. We solved several instances of the model to generate vaccination policies for the seven Texas counties (Travis, Williamson, Bastrop, Caldwell, Hays, Burnet, and Blanco) based on the all those uncertain parameters. Travis County is the most densely populated of all the seven counties, includes the city of Austin, and is surrounded by the other counties. Due to a lack of extensive studies on age-related differences in infectivity of COVID-19 at the time of this study, we investigated two cases: *Case 1*—Group A has lower relative infectivity than Group C, and *Case 2*—Group C has lower relative infectivity than Group A. We present results for two cases: *unlimited vaccine availability* and *limited vaccine availability*.

Unlimited vaccine availability

We show the results the three levels of reliability α, the two cases of relative infectivity, and the assumption of unlimited vaccine availability. The results are summarized in Table 8. For reliability level *High* and infectivity Case 1, the proportions of the population to be vaccinated to control the epidemic for Travis, Williamson, Hays, Bastrop, Caldwell, Burnet and Blanco counties are 0.94, 0.81, 0.75, 0.71, 0.65, 0.67 and 0.65, respectively. The average proportion of the population across all counties to be vaccinated is 0.87. Observe that when the reliability level is decreased, the proportion of the population to be vaccinated decreases. We observe a similar trend for infectivity Case 2. The vaccination policies prescribed under the assumption of heterogeneous model for unlimited vaccine availability, under infectivity Case 1 and the *High* reliability level are plotted in Fig 2. Those for infectivity Case 2 are plotted in Fig 3.

Limited vaccine availability

We also considered the situation in which vaccines are available in limited quantity. The results are provided for the two infectivity scenarios and vaccine shortages of 2.5%, 5.0%, 7.5%. Fig 4 shows and describes the reliability adjustment and vaccine sharing capability prescribed by the heterogeneous model under infectivity Case 1, limited vaccine availability and *High*

Table 8. The minimum number of vaccinations required to bring $R_{HVc} \leq 1$ under unlimited vaccine availability for heterogeneous populations. The proportion of the population to be vaccinated is in parentheses.

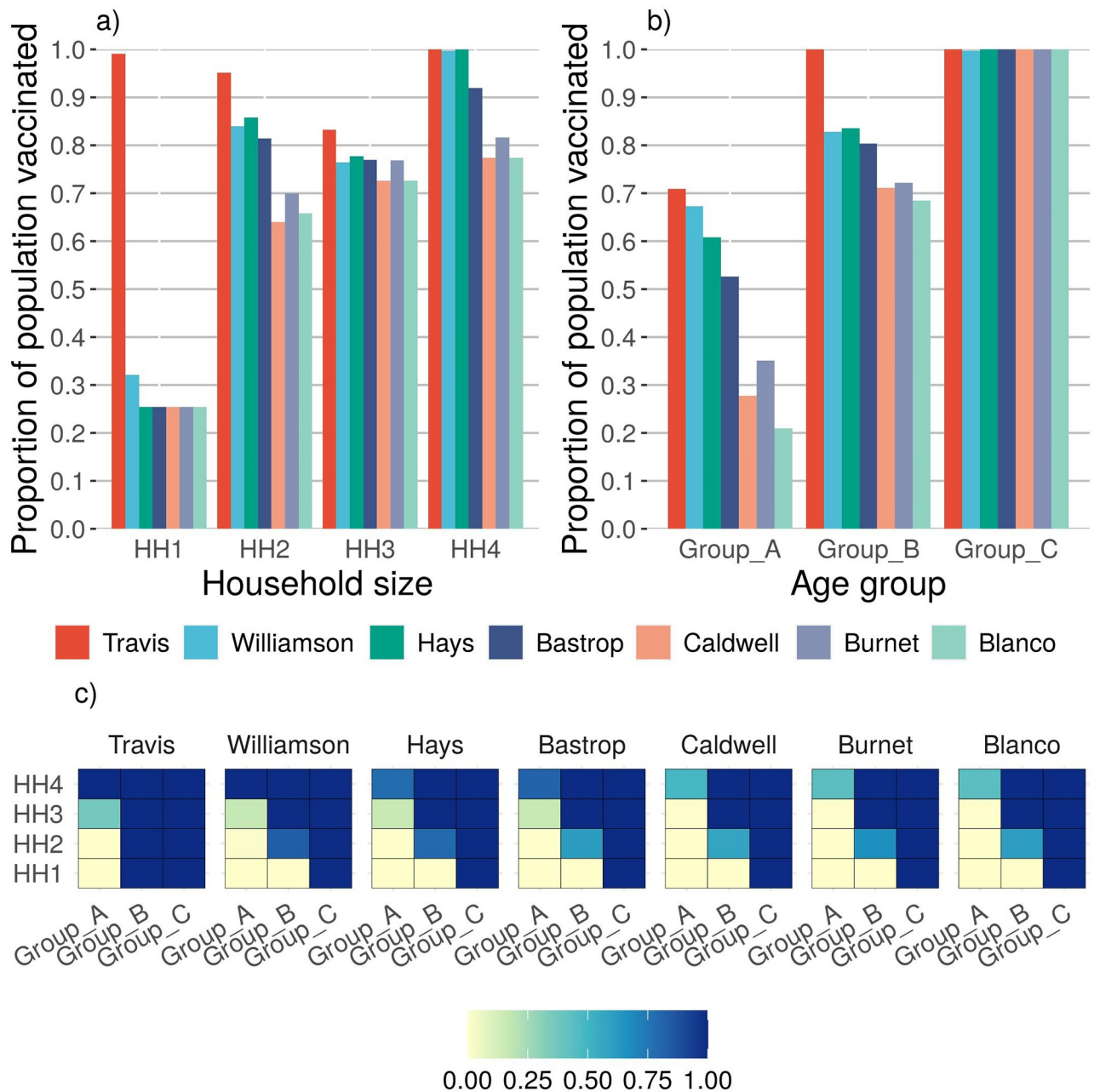| Reliability Level α							
Case 1	Travis	Williamson	Hays	Bastrop	Caldwell	Burnet	Blanco
High	831153 (0.94)	297649 (0.81)	106120 (0.75)	34837 (0.71)	16694 (0.65)	22394 (0.67)	5383 (0.65)
Medium	776772 (0.88)	283034 (0.77)	106451 (0.75)	34771 (0.71)	16573 (0.65)	22297 (0.66)	5322 (0.64)
Low	710316 (0.80)	281390 (0.77)	99689 (0.70)	32232 (0.66)	15044 (0.59)	20325 (0.60)	4895 (0.59)
Case 2							
High	795161 (0.90)	277961 (0.76)	100010 (0.71)	32603 (0.67)	15713 (0.61)	20748 (0.62)	4925 (0.59)
Medium	757667 (0.86)	267998 (0.74)	99488 (0.70)	32470 (0.67)	15474 (0.60)	20426 (0.61)	4836 (0.58)
Low	678574 (0.77)	264651 (0.72)	94743 (0.67)	31107 (0.64)	14901 (0.59)	19716 (0.59)	4771 (0.57)

Fig 2. This figure shows the vaccine policy prescribed under the assumption of heterogeneous model for unlimited vaccine availability, under infectivity Case 1 and the *High* reliability level. Fig **a)** depicts the proportion of the population to be vaccinated for each county and household size combination. The figure shows that the higher household sizes tend to be vaccinated at a higher rate. Smaller households of size one and two do have some vaccinations, albeit a lower percentage. These vaccinations are due to the fact that a majority of the population residing in smaller households are from Group B and C. Fig **b)** depicts the proportion of population to be vaccinated by county and age group. The figure illustrates the optimal policy, which is to vaccinate members of Group C population followed by members of Group B and A as a result of the higher relative susceptibility and infectivity for members of Group C. In Fig **c)** a series of heatmaps depict the proportion of population to be vaccinated by the model by household size and age group for each county. The figure indicates that communities should prioritize Group C, followed by Groups B and then A. The priority is to vaccinate Groups B and C, .i.e., populations with higher relative infectivity and susceptibility, and within each population, prioritize members residing in larger households.

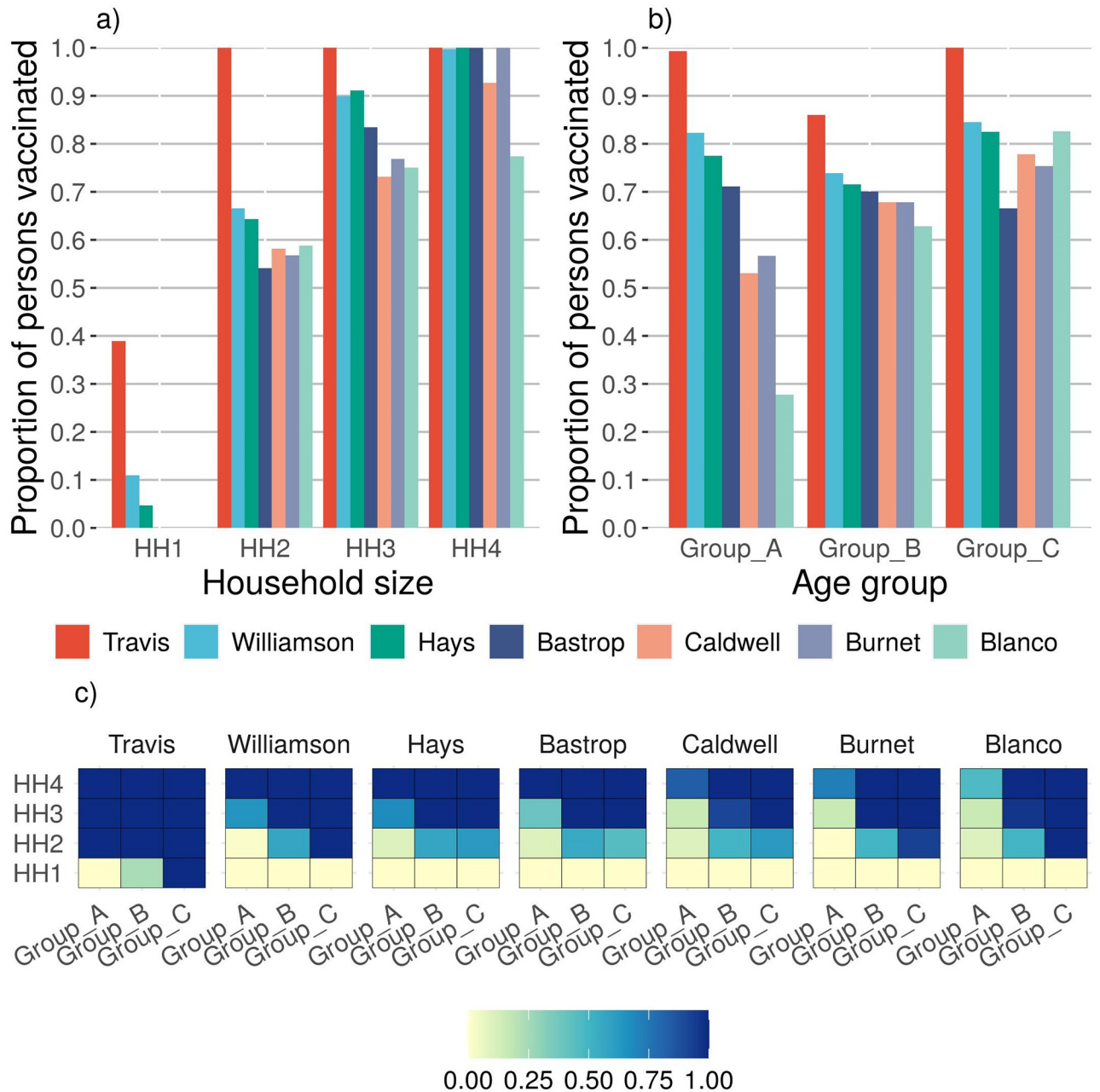https://doi.org/10.1371/journal.pone.0270524.g002

Fig 3. This figure illustrates the vaccination policy prescribed by the heterogeneous model for unlimited vaccine availability, under infectivity Case 2 and *High* reliability. Case 2 is defined such that Group A has a higher infectivity than Group C. **a)** The figure depicts the proportion of the population to be vaccinated by the model by county and household size. The figure shows that communities should vaccinate the larger households at a greater rate. This trend is due to the fact that a large number of members of Group A reside with members of Group B and C, and so a higher vaccination rate is required in this situation in order to effectively block the contagion transmission from a group of higher infectivity to a group of higher susceptibility. **b)** Fig depicts the proportion of population to be vaccinated by county and age group. The figure shows that the optimal policy recommend vaccinating the higher infectivity population, Group A, and the higher susceptibility population, Group C in Travis, Williamson and Hays counties. For Caldwell, Burnet and Blanco the priority is given to Group C followed by Group B and A respectively. The heatmaps in figure **c)** depict the proportion of the population to be vaccinated by household size and age group. The heatmaps illustrate the trend to vaccinate larger households first, and for smaller households, the preference is to vaccinate Group C followed by Group B.

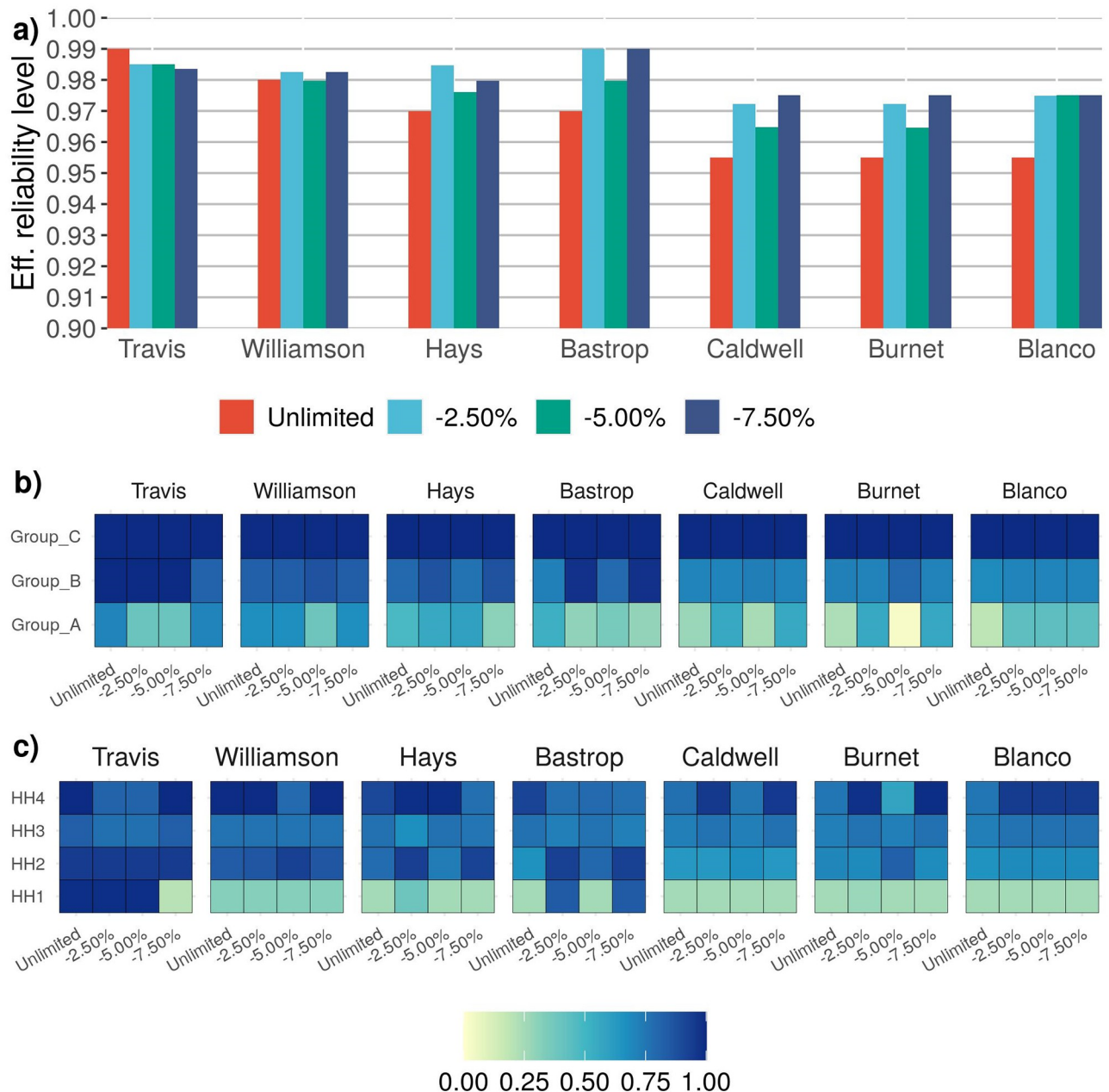https://doi.org/10.1371/journal.pone.0270524.g003

Fig 4. This figure illustrates the reliability adjustment and vaccine sharing capability prescribed by the heterogeneous model under infectivity Case 1, limited vaccine availability and _High_ reliability level. a) The bar-plot illustrates that reliability adjustment feature of the heterogeneous model under limited vaccination. For _Unlimited_ vaccine availability the model achieves the required _High_ reliability but as the vaccine availability reduces the reliability levels are adjusted to achieve an optimal vaccination policy. Generally, Travis county reliability is reduced and for other counties reliability is increased. **b)** For each county, the heatmap illustrates proportion of population to be vaccinated within an age Group for a case of vaccine availability. It shows that for Case 1 of infectivity if the reliability is lowered and vaccines are released from a county and assigned to counties where additional reliability is achieved, generally vaccines are first released from Group A followed by Group B. And the counties receiving these additional vaccines assign them to Group C first followed by Group B. **c)** For each county, the heatmap illustrates proportion of population to be vaccinated within a household size for a case of vaccine availability. Across all the counties there is no clear pattern but the plot shows the vaccination policy per household size within a county.

https://doi.org/10.1371/journal.pone.0270524.g004

reliability level. The reliability adjustment and vaccine sharing capability prescribed by the heterogeneous model under infectivity Case 2, limited vaccine availability and *High* reliability level are shown and described in Fig 5. Due to space limitations, complete results for infectivity Case 1 are given in S2 File, and for infectivity Case 2 in S3 File.

Discussion

Recall that in the heterogeneous model the household types are defined by the size of the household as well as the age groups of the members residing in the household. Therefore, the vaccination policy generated by the heterogeneous model is dependent on both the size and age composition of the members residing in the household. When we consider infectivity Case 1, the results indicate that Group C, which has higher relative susceptibility and infectivity than Group A and B, should be prioritized. The model prefers to vaccinate Group C irrespective of the size of household in which they reside. Group C is followed by Group B and then Group A in terms of the proportion of people vaccinated, where Group A has the lowest relative susceptibility and infectivity (refer to Fig 2b and 2c). With respect to household size, the model recommends to vaccinate similar proportion of inhabitants in households of size two ore more (refer to Fig 2a). This trend occurs because the majority of households of size two and larger tend to have more members of the Group B and C populations than members of Group A (refer to Fig 1b), and, in this case. Groups B and C have higher susceptibility and infectivity than Group A. In short, the model prioritizes vulnerable age groups and, within each group, prioritizes members of larger households (see Fig 2c). This vaccination policy is similar to the one implemented in the U.S. at the early stages of vaccine roll out, in which priority was given to the oldest age group. At the time of this study, the majority of what we call Group A was not eligible for vaccination, because the FDA had not approved the vaccines for members of the population 16-years-old and younger.

Under infectivity Case 2, although the required number of vaccinations is similar to that of Case 1, the model suggests a different vaccination policy. In this case, we assume that the relative infectivity of Group A is higher than that of Group C, and the relative susceptibility of Group A is lower that that of Group C. If we observe the household composition by age, the majority of the people in Group A tend to reside with members of Groups B and C population in larger households of three or more (refer to Fig 1a). In a larger household, the risk of transmission between a member of Group A and a member of Group B or C is higher. As a result, the counties should prioritize members of larger households for vaccination (refer to Fig 3a). Referring to Fig 3b), the model results show that for Travis, Williamson, and Hays counties a higher proportion of Group A and C are vaccinated when compared to Group B. In scarcely populated counties of Caldwell, Burnet and Blanco, the priority for vaccination is given to Group C followed by Group B and Group A, respectively. Fewer members of Group A reside with members of Group B or C in one- and two-person households, so the optimal vaccine policy prescribes fewer vaccinations for smaller households. Unlike Case 1, in Case 2 the solution indicates prioritizing members residing in larger household sizes, and then prioritizing vulnerable age groups—in this case Groups B and C (refer to Fig 3c). The results from this case become more relevant as younger members of the population become eligible for the vaccine.

In the case of vaccine shortages, some counties are unable to achieve the defined reliability levels, and, as a result, the results suggest lower *effective* reliability levels for Travis County, Williamson and Hays counties. However, the model increases the reliability levels for Bastrop, Blanco, Caldwell, and Burnet counties. The increased reliability suggests that additional vaccines can be transferred to these counties in order to mitigate additional outbreak scenarios. For infectivity Case 1, the vaccine policy indicates that instead of vaccinating Group A in the
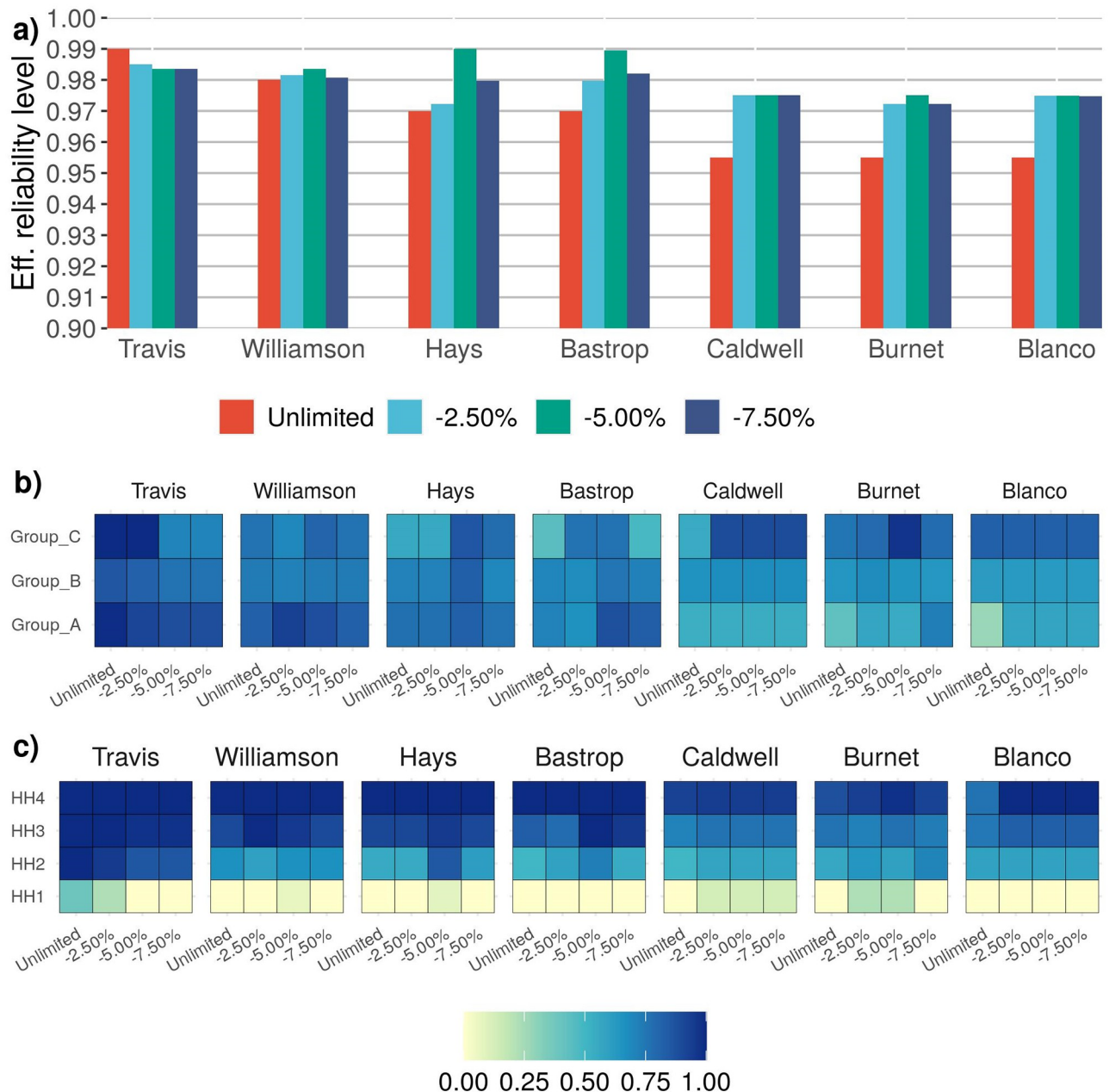
Fig 5. This figure illustrates the reliability adjustment and vaccine sharing capability prescribed by the heterogeneous model under infectivity Case 2, limited vaccine availability and *High* reliability level. a) The bar-plot illustrates that reliability adjustment feature of the heterogeneous model under limited vaccination. For *Unlimited* vaccine availability the model achieves the required *High* reliability but as the vaccine availability reduces the reliability levels are adjusted to achieve an optimal vaccination policy. Generally, Travis county reliability is reduced and for other counties reliability is increased. **b)** For each county, the heatmap illustrates proportion of population to be vaccinated within an age group for a case of vaccine availability. Across all the counties there is no clear pattern but the plot shows the vaccination policy per age group within a county. **c)** For each county, the heatmap illustrates proportion of population to be vaccinated within a household size for a case of vaccine availability. It shows that for Case 2 of infectivity if the reliability is lowered and vaccines are released from a county and assigned to counties where additional reliability is achieved, generally vaccines are first released from HH1 followed by HH2. And the counties receiving these additional vaccines assign them to higher household size of three and four.

counties where reliability is lowered, those vaccines can be distributed to members of Group C, followed by members of Group B in counties where it is possible to satisfy additional scenarios (see Fig 4). In infectivity Case 2, the solution indicates that vaccines for smaller households in the counties where reliability cannot be met should be redistributed to larger households in counties where additional reliability can be achieved (see Fig 5).

Conclusion

We introduce a multi-community household-based SP methodology for generating an optimal vaccination policy to control the outbreak of COVID-19. While generating a vaccination policy, this new methodology considers uncertainty inherent in both COVID-19 and human interactions. We develop two stochastic models that use the demographic structure of households based on census data, as well as age-related heterogeneity to COVID-19 in the sub-populations of each community. The models generate vaccination policies under unlimited and limited vaccine availability, respectively, and incorporates the idea of vaccine sharing between communities in order to control COVID-19 outbreaks. The model was implemented and tested based on seven neighboring counties in the U.S. state of Texas. Computational results show that to control the outbreak at least a certain percentage of the population in each county should be vaccinated, depending on the pre-determined reliability levels. The study also reveals the vaccine sharing capability of the proposed model among the counties under limited vaccine availability. This work contributes a new decision-making tool to aid public health agencies in the optimal allocation of vaccines under uncertainty for multiple communities to control epidemics through vaccinations. Future research along this line of work include extending the proposed methodology to include refined age classes and the vaccination status of individuals. Another direction is to include new models to estimate the outside household contact rate and incorporating the logistics of vaccine delivery.

Supporting information

S1 File. This file contains the discrete distribution for $m(\tilde{\omega}_c)$, which is the number of close contacts that an infective makes with persons of other households.
(XLSX)

S2 File. This file contains computational results for infectivity Case 1.
(XLSX)

S3 File. This file contains computational results for infectivity Case 2.
(XLSX)

Author Contributions

Conceptualization: Krishna Reddy Gujjula, Jiangyue Gong, Brittany Segundo, Lewis Ntaimo.

Data curation: Krishna Reddy Gujjula.

Formal analysis: Lewis Ntaimo.

Investigation: Jiangyue Gong, Brittany Segundo.

Methodology: Lewis Ntaimo.

Software: Jiangyue Gong, Lewis Ntaimo.

Supervision: Lewis Ntaimo.

Validation: Krishna Reddy Gujjula, Jiangyue Gong, Brittany Segundo, Lewis Ntaimo.

Visualization: Krishna Reddy Gujjula, Jiangyue Gong.

Writing – original draft: Krishna Reddy Gujjula, Jiangyue Gong.

Writing – review & editing: Brittany Segundo, Lewis Ntaimo.

References

1. The World health organization. https://covid19.who.int/table, 2020.

2. Chikina M. and Pegden W. Modeling strict age-targeted mitigation strategies for COVID-19. *PLoS ONE*, 15(7):e0236237, 2020. https://doi.org/10.1371/journal.pone.0236237 PMID: 32706809

3. Goldsztejn U., Schwartzman D., and Nehorai A. Public policy and economic dynamics of COVID-19 spread: A mathematical modeling study. *PLoS ONE* 15(12):e0244174, 2020. https://doi.org/10.1371/journal.pone.0244174 PMID: 33351835

4. Bokharaie V.S. A study on the effects of containment policies and vaccination on the spread of SARS-CoV-2. *PLoS ONE* 16(3):e0247439, 2021. https://doi.org/10.1371/journal.pone.0247439 PMID: 33661929

5. Heesterbeek J.A.P. and Dietz K. The concept of R_0 in epidemic theory. *Statistica Neerlandica*, 50 (1):89–110, 1996. https://doi.org/10.1111/j.1467-9574.1996.tb01482.x

6. Bartoszyński R.On a certain model of an epidemic. *Applied Mathematics*, 13(2):139–151, 1972.

7. Nishiura H. and Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. *Mathematical and Statistical Estimation Approaches in Epidemiology*, p. 103–121, 2009.

8. Ng T-C. and Wen T-H. Spatially adjusted time-varying reproductive numbers: Understanding the geographical expansion of urban dengue outbreaks. *Scientific Reports*, 9(1):19172, 2019. https://doi.org/10.1038/s41598-019-55574-0 PMID: 31844099

9. Luca B., Leslie R., Giulio D.L. Transmission heterogeneity and control strategies for infectious disease emergence. *PLoS ONE*, 2(8): e747, 2007. https://doi.org/10.1371/journal.pone.0000747

10. Delamater P.L., Street E.J., Leslie T.F., et al. Complexity of the basic reproduction number (R0). *Emerging Infectious Diseases*, 25(1):1–4, 2019. https://doi.org/10.3201/eid2501.171901 PMID: 30560777

11. The World Health Organization. WHO SAGE roadmap for prioritizing uses of COVID-19 vaccines in the context of limited supply. https://www.who.int/docs/default-source/immunization/sage/covid/sage-prioritization-roadmap-covid19-vaccines.pdf?Status=Temp&sfvrsn=bf227443_2, 2020.

12. The Centers for Disease Control and Prevention. COVID-19 vaccination program interim playbook for jurisdiction operations. https://www.cdc.gov/vaccines/imz-managers/downloads/COVID-19-Vaccination-Program-Interim_Playbook.pdf, 2020.

13. Habersaat K.B., Betsch C., Danchin M., et al. Ten considerations for effectively managing the COVID-19 transition. *Nature Human Behaviour*, 4(7):677–687, 2020. https://doi.org/10.1038/s41562-020-0906-x PMID: 32581299

14. McKee M., Stuckler D. If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. *Nature Medicine*, 26(5):640–642, 2020. https://doi.org/10.1038/s41591-020-0863-y PMID: 32273610

15. Lazarus J.V., Ratzan S.C., Palayew A., et al. A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine*, 27:225–228, 2021. https://doi.org/10.1038/s41591-020-1124-9 PMID: 33082575

16. Neumann-Böhme S., Varghese N.E., Sabat I., et al. Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. *The European Journal of Health Economics*, 21 (7):977–982, 2020. https://doi.org/10.1007/s10198-020-01208-6 PMID: 32591957

17. Johnson N.F., Velásquez N., Restrepo N.J., et al. The online competition between pro- and anti-vaccination views. *Nature*, 582(7811):230–233, 2020. https://doi.org/10.1038/s41586-020-2281-1 PMID: 32499650

18. Matrajt L., Eaton J., Leung T., et al. Vaccine optimization for COVID-19: Who to vaccinate first? *Science Advances*, 7(6), 2020. https://doi.org/10.1126/sciadv.abf1374 PMID: 33536223

19. Bubar K.M., Reinholt K., Kissler S.M., et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532):916–921, 2021. https://doi.org/10.1126/science.abe6959 PMID: 33479118

20. Shen Zhong-Hua, Chu Yu-Ming, Khan Muhammad Altaf, et al. Mathematical modeling and optimal control of the COVID-19 dynamics. *Results in Physics*, 31(7):105028, 2021. https://doi.org/10.1016/j.rinp.2021.105028 PMID: 34868832

21. Hoan L.V.C., Akinlar M.A., Inc M., et al. A new fractional-order compartmental disease model. *Alexandria Engineering Journal*, 59(5): 3187–319, 2020. https://doi.org/10.1016/j.aej.2020.07.040

22. Pandey P., Chu Y-M, Gómez-Aguilar J.F., et al. A novel fractional mathematical model of COVID-19 epidemic considering quarantine and latent time. *Results in Physics*, 26:104286, 2021. https://doi.org/10.1016/j.rinp.2021.104286 PMID: 34028467

23. Anderson R.M., Heesterbeek H., Klinkenberg D., et al. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395(10228):931–934, 2020. https://doi.org/10.1016/S0140-6736(20)30567-5 PMID: 32164834

24. Tanner M.W., Sattenspiel L., Ntaimo L. Finding optimal vaccination strategies under parameter uncertainty using stochastic programming. *Mathematical Biosciences*, 215(2):144–151, 2008. https://doi.org/10.1016/j.mbs.2008.07.006 PMID: 18700149

25. Birge J.R., Louveaux F.V. *Introduction to Stochastic Programming*. Springer, New York, 1997.

26. Ruszczyński A., Shapiro A. Stochastic Programming. Handbooks in Operations Research and Management Science. Elsevier, 10, 2003.

27. Becker N.G., Starczak D.N. Optimal vaccination strategies for a community of households. *Mathematical Biosciences*, 139(2):117–132, 1997. https://doi.org/10.1016/S0025-5564(96)00139-3 PMID: 9009574

28. Wang Y., Liu Y., Struthers J., et al. Spatiotemporal characteristics of the COVID-19 epidemic in the United States. Clinical Infectious Diseases, 72(4):643–651, 2021 https://doi.org/10.1093/cid/ciaa934 PMID: 32640020

29. Resilient Infrastructure and Smart Cities (RISC) Lab. https://covid19-usa-ut-austin.hub.arcgis.com/, 2020.

30. Charnes A., Cooper W.W., Symonds G.H. Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil. *Management Science*, 4:235–636, 1958. https://doi.org/10.1287/mnsc.4.3.235

31. Pre'kopa A. Contributions to the theory of stochastic programming. *Mathematical Programming*, 4:202–221, 1973. https://doi.org/10.1007/BF01584661

32. U.S. Census Bureau. Tenure by household size, 2014-2018, American community survey 5-year estimates-b25009. https://data.census.gov/cedsci/table?q=B25009&g=0400000US48.050000&tid=ACSDT5Y2018.B25009&hidePreview=false, 2018.

33. S. Ruggles, et al., IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V10.0, 2020.

34. Jing Q-L., Liu M.J., Zhang Z.B., et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *The Lancet Infectious Diseases*, 20 (10):1141–1150, 2020. https://doi.org/10.1016/S1473-3099(20)30471-0 PMID: 32562601

35. Fung H.F., Martinez L., Alarid-Escudero F., et al., Stanford-CIDE Coronavirus Simulation Model (SC-COSMO) Modeling Group. The household secondary attack rate of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): A rapid review. *Clinical Infectious Diseases*, 73(Suppl 2):S138–S145, 2020.

36. Shah K., Saxena D., Mavalankar D. Secondary attack rate of COVID-19 in household contacts: a systematic review. *QJM: An International Journal of Medicine*, 113(12):841–850, 2020. https://doi.org/10.1093/qjmed/hcaa232 PMID: 32726452

37. The Centers for Disease Control. COVID-19 and Your Health. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness.html, 2020.

38. Davies N.G., Klepac P., Liu Y., et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8):1205–1211, 2020. https://doi.org/10.1038/s41591-020-0962-9 PMID: 32546824

39. Goldstein E., Lipsitch M., Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools, and the community. *The Journal of Infectious Diseases*, 223(3):362–369. 2021. https://doi.org/10.1093/infdis/jiaa691 PMID: 33119738

40. Prem K., Cook A.R., Jit M., Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13(9):e1005697, 2017. https://doi.org/10.1371/journal.pcbi.1005697 PMID: 28898249

41. Austin Dashboard. https://covid-19.tacc.utexas.edu/dashboards/austin/, 2020.

42. *IBM ILOG CPLEX Optimization Studio CPLEX User's Manual*, Version 12 Release 9, Armonk, NY, IBM Corporation, 2019.