# OPTIMIZING VACCINE CLINIC OPERATIONS IN LOW AND MIDDLE INCOME COUNTRIES

by

**Maryam Hasanzadeh Mofrad**

B.S. in Industrial Engineering, Sharif University of Technology, 2007

M.S. in Industrial Engineering, Sharif University of Technology, 2009

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Maryam Hasanzadeh Mofrad

It was defended on

March 25, 2016

and approved by

Lisa M. Maillart, PhD, Associate Professor

Bryan A. Norman, PhD, Associate Professor

Jayant Rajgopal, PhD, Professor

Ruben A. Proano, PhD, Associate Professor

Dissertation Director: Lisa M. Maillart, PhD, Associate Professor

**OPTIMIZING VACCINE CLINIC OPERATIONS IN LOW AND MIDDLE INCOME COUNTRIES**

Maryam Hasanzadeh Mofrad, PhD

University of Pittsburgh, 2016

This dissertation focuses on two open questions in operating vaccination clinics in low and middle income countries. First, as a result of limited "open vial life," clinicians face difficult tradeoffs between opening a multi-dose vial to satisfy a potentially small immediate demand versus retaining the vial to satisfy a potentially large future demand. Second, in low and middle income countries, governmental health organizations face tradeoffs between locating (additional) clinics or conducting outreach trips to vaccinate patients in remote locations.

To answer the first question, we formulate Markov decision process models that determine when to conserve vials as a function of the time of day, the current vial inventory, and the remaining clinic-days until the next replenishment. The objective is to minimize "open-vial waste" while administering as many vaccinations as possible. For the base model, we analytically establish that the optimal policy is of a threshold type; conduct extensive sensitivity analysis on model parameters; develop a practical heuristic policy; suggest operational approaches that do not overly inconvenience patients; define metrics for determining appropriate operating hours and sessions per inventory replenishment cycle; and study the impact of random vial yield. We then generalize the base model by considering a positive probability of return for patients who are not vaccinated on their first visit and incorporating non-stationary arrival rates. To study these enhancements, we perform extensive numerical experiments at the clinic level for a single replenishment cycle and then extrapolate to mul-

tiple clinics and the entire world on an annual basis. The results indicate potential savings on the order of millions of doses per year.

To answer the second question, given a network of population centers we develop a mixed integer linear programming model that determines clinic locations and outreach activities. The model minimizes cost over a specified period of time subject to constraints on coverage, trip distance, trip size, trip frequency and patient travel. We address demand uncertainty; perform sensitivity analysis on key model parameters; compare the performance of the optimal solution to heuristic policies; and conclude that while counterintuitive it is often suboptimal to locate clinics in the largest population centers.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## PREFACE

I am thankful to all the people who helped me throughout my PhD study. I would never have been able to finish my dissertation without the guidance of my committee members, support from friends and my family.

I would like to thank my advisor Dr. Lisa Maillart, first and foremost, for the continuous support of my research, for her encouragement, caring and patience. Her guidance and excellent feedbacks helped me in all the time of research and writing of this thesis. I also like to thank Drs. Bryan Norman and Jayant Rajgopal for their encouragement and insightful comments. I have learned so much from them and I deeply appreciate their support. My sincere thanks also go to Dr. Proano for his valuable suggestions and feedbacks.

Finally, I would like to thank my parents and my husband, Mohammadhosein, for supporting me spiritually and encouraging me with their best wishes throughout writing this thesis and my life in general.

# 1.0 DYNAMICALLY OPTIMIZING THE ADMINISTRATION OF VACCINES FROM MULTI-DOSE VIALS

## 1.1 INTRODUCTION

The prevalence of infectious diseases is a grave, global concern and millions of individuals – both children and adults – remain at risk for diseases that are preventable through immunization. According to the most recent World Health Organization (WHO) publication on the state of the world's vaccines and immunization [56], in 2007, "24 million children...did not get the complete routine immunizations scheduled for their first year of life." This problem is most acute in developing countries, but is of interest to a variety of stakeholders (e.g., the developing countries themselves, vaccine manufacturers, non-profit organizations such as WHO and UNICEF that assist these countries, as well as more wealthy countries that support vaccine dissemination in these locations).

The Expanded Program on Immunization (EPI) was launched by the WHO in 1974 [55] with the goal of vaccinating children throughout the world, and in 1999, the Global Alliance for Vaccines and Immunization (GAVI) was created specifically to extend the EPI program throughout the poorest countries in the world. The primary goal of these programs is to maximize the number of children who are vaccinated against common vaccine-preventable diseases such as measles, hepatitis, yellow fever, meningitis, etc. A key step in achieving this goal is to reduce the large amount of vaccine "wastage," i.e., doses that are manufactured and shipped, but then not administered while still viable, in developing countries. WHO [54] estimates the overall global vaccine wastage rate to be a staggering 50% which varies with vial size, vaccine type, location, etc. More specifically, a recent study conducted in Bangladesh [16] reports average wastage rates of 85.1% for BCG (tuberculosis), 71.2% for

measles, 36.6% for TT (tetanus) and 44.2% for DPT (diphtheria, pertussis and tetanus combination). Reducing wastage has been mandated by WHO [54] and identified as a key factor in maintaining the financial sustainability of immunization programs in the world's poorest countries [19] and in increasing the delivery of vaccines to patients. In the overall effort to improve vaccine coverage rates, it is critical that clinics manage vaccine supplies efficiently in the field to minimize wastage and maximize the number of immunizations administered.

The primary driver of vaccine wastage is the fact that many vaccines are manufactured in large, multi-dose vials, which are more economical to produce than single-dose vials due to their lower production, transportation and storage costs; however, they result in so-called "open-vial" waste [16]. Vaccines used in worldwide immunization programs are typically manufactured in two forms: a liquid form that can be directly administered and a freeze-dried powder that must be reconstituted with a diluent before administration. In the powder form, the primary concern prior to reconstitution is the *shelf life* of the vaccine; it determines the expiration date of the vaccine. After reconstitution, the remaining lifetime of the vaccine is called *open vial life*, which is considerably shorter than shelf life. For example, a 10-dose vial of MMR vaccine has a 48-month shelf life, but only an 8-hour open vial life [4]. When a multi-dose vial is reconstituted or "opened," but not completely used during its open vial life, the unused doses are discarded. This type of vaccine wastage is called *open vial waste* (OVW) [7].

A number of issues must be considered in designing an effective immunization program with low wastage. The majority of these issues are addressed at higher levels of decision making within governmental immunization organizations, e.g., determining the best vial size; designing the distribution chain; setting storage capacities, replenishment frequencies, and order quantities at various levels of the vaccine supply chain; and deciding on transportation modes and their capacities. Vial size and inventory considerations in particular have been examined in several previous studies. For example, [36, 24, 23] and [60] perform economical analysis to determine the appropriate vial size. More specifically, Parmar et al. in [36] estimate the potential wastage cost associated with different vial sizes in various countries. Lee et al. in [24] show that the most economical vial size is a function of mean daily demand

(i.e., vials with a greater number of doses produce less OVW as the mean daily demand increases). Lee et al. in [23] determine a threshold on the mean daily demand for the adoption of a specific vial size; they also argue that using single-dose vials (which are physically larger on a per-dose basis) can severely constrain transportation capacity in the vaccine distribution supply chain and result in increased risk of vial breakage. Lastly, Yang et al. in [60] investigate the economical impact of vial size using empirical data. Related to this work, Assi et al. in [1] study the impact of different vial sizes on the vaccine supply chain. Lastly, Dhamodharan & Proano in [7] integrate an optimization model and simulation to simultaneously determine vial size and ordering policy under the assumption of 100% coverage. Regardless of how vial size and replenishment frequency are determined, questions surrounding the downstream issue of how to best administer doses from multi-dose vials remain open.

The administration policy most commonly used in practice is to never turn away a patient as long as the required vaccine is available. Thus, vials are typically opened and vaccines administered in a "greedy" fashion with no consideration of long-term consequences, resulting in high open-vial wastage rates and frequent stockouts or "downtime", i.e., sessions at the end of the replenishment cycle on which the clinic cannot open at all. Clearly, however, a tradeoff exists between opening a vial to satisfy a potentially small immediate demand (and likely subjecting a significant number of doses to open-vial waste) versus retaining the vial to satisfy a potentially large future demand. For example, when the last dose in a vial is administered late in the day, should the clinic discontinue service for the rest of the day? Indeed, some clinics recognize this tradeoff and, as Hutchins [18] explains, incorporate the practice of "not opening a multidose vaccine vial for a small number of persons to avoid vaccine wastage..."

We seek to find the *optimal* balance between these competing factors such that the vaccination rate is maximized. More specifically, we take a Markov decision process (MDP) approach in which the decision of whether or not to remain open is made as a function of time of day, current vial inventory and the remaining sessions until the next replenishment. The resulting dynamic policies, compared to the greedy approach, optimally spread the clinic downtime throughout the replenishment period so that a significantly greater fraction of demand can be met.

There is a limited body of work on open-vial waste [7, 9, 16, 23, 24]. This majority of this literature focuses on either quantifying current open-vial waste rates [16] or developing simple cost models to quantify the economic impact of open-vial waste. In performing the latter, Drain et al. [9] assume that the open-vial waste rate is known. In contrast, the spreadsheet model in [24] does not make this assumption, although, unlike our approach, Lee et al. [24] assume fixed clinic hours and that every arriving patient is immunized; the model is used to calculate expected cost figures for various (vaccine, vial size) combinations and determine thresholds on the mean demand above which increases in vial size are economical for each type of vaccine considered. Lee et al. [23] also consider open-vial waste in their examination of how vial size affects vaccine availability and coverage, but their focus is on supply chain functions such as shipping frequency and storage capacity rather than vaccine administration. Similarly, Dhamodharan et al. [7] also focus on the question of vial size, but in an optimization context. They seek an optimal vial size, order quantity (in vials) and reorder point to minimize expected procurement, holding and open-vial wastage costs while satisfying 100% of the demand.

The vaccine administration problem considered here can also be viewed as a novel perishable inventory control problem (see surveys in [20] and [33]) involving batch "ordering" decisions for a perishable product from a finite supply to satisfy as much stochastic demand as possible over a finite horizon with zero lead time and discontinuous hours of operation. The "ordering" decisions correspond to opening a vial; the reason they are "batch" orders is because each vial contains some fixed number of doses. The "inventory level" corresponds to the number of doses in the currently open vial. The doses are perishable because they only last a matter of hours once opened. The supply is finite because the planning horizon is the time between replenishments, the lead time is zero because the doses are already on hand and the "ordering" decision is really an "opening" decision, and service is not continuous because the clinics open and close each day. To the best of our knowledge, the combination of finite supply and discontinuous operation is previously unexplored in the context of perishable inventory control. A few studies consider products that, similar to vaccines that are prone to "open-vial" waste, start to decay not once they arrive, but once they are "unpacked" [41, 48]. These studies, however, focus on the non-zero lead time case for con-

tinuously operating systems with infinite supply and therefore seek an optimal reordering policy that minimizes cost.

The remainder of this chapter is organized as follows. In Section 1.2, we present an MDP formulation of the problem and a recursive expression to evaluate the greedy policy. In Section 1.3, we analytically establish the threshold type structure of the optimal policy and present an insightful numerical example. In Section 1.4, we perform sensitivity analysis on each of the model parameters and investigate two easy-to-implement heuristic policies. In Section 1.5, we conclude and comment on future work.

## 1.2 MODEL FORMULATION

In this section, we present an MDP model that determines when to discontinue service given that each vial contains a fixed number of doses and has a limited shelf-life after opening. We assume that: $(i)$ the maximum number of clinic hours per session is equal to the shelf life of an open vial (i.e., doses cannot expire midday), $(ii)$ the daily demands are iid, $(iii)$ the minimum number of working hours per session is fixed *a priori*, $(iv)$ opting to discontinue service does not affect the demand distribution on subsequent sessions and $(v)$ the replenishment schedule is fixed. Although assumption $(i)$ may not hold for some vaccines with shorter open vial lifetimes (e.g., six hours) in clinics with longer sessions (e.g., eight hours), many clinics only vaccinate for six or fewer hours per day. Overall, this assumption allows us to cover a large proportion of clinic environments while capturing the most important problem dynamics without considering the age of the currently open vial when making decisions; we leave this generalization for future work. Assumption $(iv)$ may also be violated, i.e., service interruptions may impact demand, however, service interruptions are already commonplace in many locations due to stockouts at the end of the replenishment cycle. Therefore, this assumption is not overly restrictive and we leave its relaxation for Chapter 3.

Although from an implementation perspective this problem is viewed as one of deciding when to close the clinic for the day, we take a different, but equivalent, modeling approach. More specifically, we formulate a model to determine whether or not to open a vial *when*

*a patient arrives, no vial is currently open* and the clinic has already met the prespecified minimum number of working hours for the day. We divide the session into $\eta$ timeslots of length $\tau$ such that the maximum number of arrivals per timeslot is 1. Let $\hat{h}$ be the minimum number of working timeslots, i.e., the earliest time that service may be discontinued. The equivalence between the approaches can then be seen by noting that if it is optimal *not* to open a vial when a patient arrives after timeslot $\hat{h}$ and no vial is open, then it would have been optimal for the clinic to close for the day if no patient arrived in the previous timeslot. In Sections 1.3, we present the optimal policy in this latter form.

Let $T$ be the number of sessions between replenishments, $Q$ be the replenishment quantity, i.e., the initial inventory in vials, and $z$ be the number of doses per vial. (We assume $T > 1$ throughout because when $T = 1$ acting "greedily" is optimal.) Let $h$ be the timeslot of the arrival prompting the current decision, $h = 1, 2, \ldots, \eta$. We assume that $(i)$ vaccine demand (if any) occurs at the beginning of each timeslot, $(ii)$ the earliest the first patient can arrive is at time zero, which corresponds to the beginning of timeslot $h = 1$, $(iii)$ the latest the last patient can arrive is time $(\eta - 1)\tau$, which corresponds to the beginning of timeslot $h = \eta$ and $(iv)$ the latest the clinic may close is after it administers a vaccine at time $(\eta - 1)\tau$ to the last patient (if any).

Let the patient interarrival times (measured in timeslots) be iid geometric($p$) random variables and let $X$ denote the arrival time (slot) of the first arrival. Hence,

$$p_X(x) = \begin{cases} p(1-p)^{x-1}, & x = 1, \ldots, \eta, \\ (1-p)^{\eta}, & x = \eta + 1, \end{cases} \tag{1.1}$$

and the daily demand, $D$, is a binomial $(\eta, p)$ random variable. (Note that for $\eta$ large and $p$ small, $D$ is approximately Poisson.) Let $D_h$ be the total demand in timeslot $h$ through the end of the day given an arrival in timeslot $h$, i.e.,

$$p_{D_h}(d) = \begin{cases} \binom{\eta-h}{d-1} p^{(d-1)}(1-p)^{(\eta-h-(d-1))}, & d = 1, \ldots, \eta - h + 1; h = 1, \ldots, \eta, \\ 0, & \text{otherwise.} \end{cases} \tag{1.2}$$

Lastly, let $Y_h$ be the timeslot in which the $z^{\text{th}}$ arrival after timeslot $h$ occurs and let $Y_h = \eta+1$ denote the case in which the number of arrivals after timeslot $h$ before the end of the day is less than $z$. Note that $Y_h$ is a shifted negative binomial random variable, i.e.,

$$p_{Y_h}(y) = \begin{cases} 0, & y < h+z, \\ \binom{y-h-1}{z-1} p^z (1-p)^{y-h-z}, & y = h+z, h+z+1, \ldots, \eta, \\ 1 - \sum_{i=h+z}^{\eta} p_{Y_h}(i), & y = \eta+1. \end{cases} \tag{1.3}$$

The state of the process at the time of an arrival when there are currently no open vials is given by $(t, q, h)$ where $t$ is the number of sessions (including the current session) until the next replenishment arrives, $q$ is the number of vials on hand and $h$ is the current timeslot. In Figure 1 we depict a schematic representation of the replenishment cycle in which we also indicate the state of the process. Let $V(t, q, h)$ denote the maximum expected number of vaccines administered prior to the next replenishment starting from state $(t, q, h)$. Hence, for $t > 0$, $q > 0$, $h \leq \eta$,

$$V(t,q,h) = \begin{cases} g(t,q,h), & h \leq \hat{h}, \qquad (1.4a) \\ \max\left\{\nu(t-1,q), g(t,q,h)\right\}, & h > \hat{h}, \qquad (1.4b) \end{cases}$$

where

$$g(t,q,h) \equiv \sum_{y=h+z}^{\eta} \Big(z + V(t,q-1,y)\Big) p_{Y_h}(y) + \sum_{d=1}^{z} \Big(d + \nu(t-1,q-1)\Big) p_{D_h}(d), \tag{1.5}$$

and

$$\nu(t,q) \equiv \sum_{h=1}^{\eta+1} V(t,q,h) p_X(h). \tag{1.6}$$

Expression (1.4a) corresponds to the situation in which a new vial must be opened because a patient arrives at or before timeslot $\hat{h}$ and no vial is currently open. The first term on the right-hand-side of Equation (1.5) corresponds to the case in which a vial is opened and the rest of the day's demand (including the current arrival) exceeds $z$. The second term corresponds to the opposite case in which a vial is opened, but is not used completely before the end of the day.

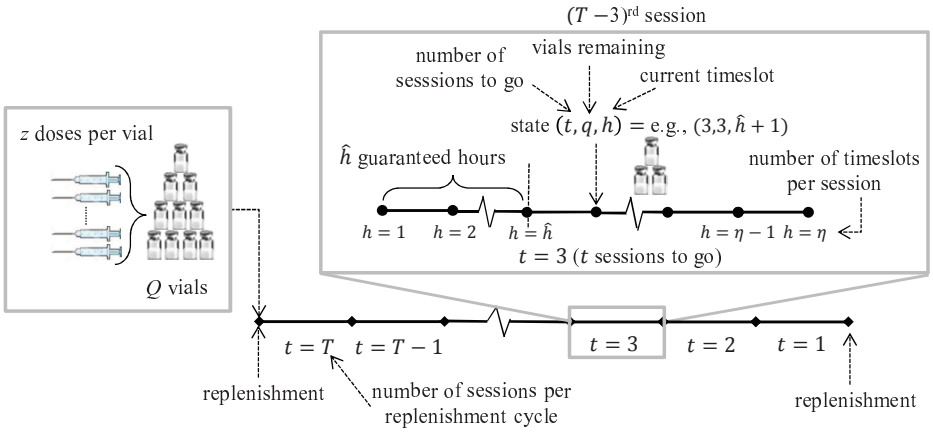



Figure 1: Replenishment cycle timeline.

Expression (1.4b) specifies the maximum expected number of vaccines administered prior to the next replenishment given that an arrival occurs in timeslot $h > \hat{h}$ and no vial is currently open. In this situation, clinicians must decide between opening a new vial (Equation (1.5)) and discontinuing service for the day (Equation (1.6)). Equation (1.6) is simply the maximum expected number of vaccinations to go at the beginning of the $t^{\text{th}}$ session until replenishment (i.e., prior to the first arrival) when there are $q$ vials on hand.

Clearly, if there is no demand on any given session, then there is no decision to be made and $V(t, q, \eta + 1) = \nu(t - 1, q)$. Similarly, if the vial inventory is exhausted, then there is no more reward to be earned, i.e., $V(\cdot, 0, \cdot) = 0$. Lastly, because the problem horizon is defined as the replenishment interval, we also have $V(0, \cdot, \cdot) = 0$.

As described in Section 1.1, current practice is typically to immunize every patient as long as inventory is on hand. Let $\tilde{V}(t, q, h)$ denote the expected number of vaccines administered prior to the next replenishment when starting from state $(t, q, h)$ under this policy. We can evaluate this "greedy policy" by removing the option not to open a vial from Equation (1.4). Hence for $t > 0$, $q > 0$, $h \leq \eta$,

$$\tilde{V}(t, q, h) = \sum_{y=h+z}^{\eta} \Big( z + \tilde{V}(t, q - 1, y) \Big) p_{Y_h}(y) + \sum_{d=1}^{z} \Big( d + \tilde{\nu}(t - 1, q - 1) \Big) p_{D_h}(d), \qquad (1.7)$$

where

$$\tilde{\nu}(t,q) \equiv \sum_{h=1}^{\eta+1} \tilde{V}(t,q,h) p_X(h). \tag{1.8}$$

The boundary cases are treated the same as in the MDP formulation, i.e., $\tilde{V}(t,q,\eta+1) = \tilde{\nu}(t-1,q)$, $\tilde{V}(\cdot,0,\cdot) = 0$ and $\tilde{V}(0,\cdot,\cdot) = 0$.

## 1.3 A NUMERICAL EXAMPLE AND POLICY STRUCTURE

In this section, we first present a numerical example for which the optimal policy exhibits intuitive structure. We then analytically establish that this structure holds in general. Subsequently, we compare the performance of the optimal policy to that of the greedy policy in terms of the percentage of demand vaccinated and the expected number of doses lost due to OVW.

Consider a problem instance for which $T = 20$ sessions, $\eta = 480$ timeslots per session and $\hat{h} = 0$ (i.e., there is no restriction on the minimum number of working hours per session). Suppose the expected daily demand is $\mu = 11$ patients/session so that $p = \frac{\mu}{\eta} = 0.0229$ (this value falls within the range of values seen in many catchment areas, e.g., Bangladesh [16]), and that $z = 10$ doses per vial. Further, suppose that the initial inventory level equals the total expected demand, i.e., $Q = (11 \times 20)/10 = 22$ vials. We implement the backward induction algorithm [38] to solve for the optimal policy using Equation (1.4) as well as to evaluate the greedy policy using Equation (1.7).

For this example, the optimal policy exhibits intuitive structure. More specifically, for each session $t$ and inventory level $q$, there exists a timeslot $h^*(t,q)$ such that for $h > h^*(t,q)$, $V(t,q,h) = \nu(t-1,q)$. That is, if when timeslot $h^*(t,q)$ occurs there is no vial open, it is optimal to discontinue service for the day. Alternatively, if when timeslot $h^*(t,q)$ occurs there is a vial open, it is optimal to discontinue service when all of its doses have been administered or at the end of the day, whichever occurs first.

Figure 2 plots $h^*(t,q)$ as a function of $q$ for $t \in \{5,10,15,20\}$. As expected, the optimal policy is more conservative when the vial inventory is low; that is, $h^*(t,q)$ decreases in $q$ for

a given $t$ so that given a lower inventory of vials, it is optimal to discontinue service earlier in the day. Starting from timeslot zero with $t$ sessions remaining and some vial inventory $q$, a sample path would move down and to the left as vaccines are administered until it hits the corresponding threshold at which point it is optimal to discontinue service immediately or when the vial currently open is emptied. That is, for all $(h, q)$ pairs below the curve $h^*(t, q)$ in Figure 2, opening a new vial is suboptimal. A sharp change in $h^*(t, q)$ occurs when the number of vials remaining equals the number of sessions remaining, which is intuitive for this example in which the daily demand and the vial size are approximately equal. For this example with $\eta = 480$, if the maximum length of a session is eight hours (e.g., 8am-4pm), then each timeslot is one minute long and corresponds to the right-hand vertical axis.

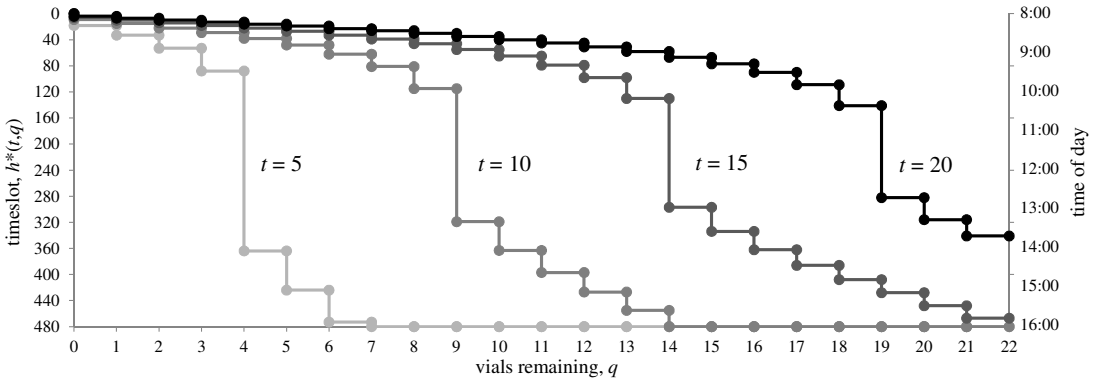



Figure 2: Optimal threshold, $h^*(t, q)$, as a function of $t$ and $q$ for the numerical example.

Proposition 1 establishes that the policy structure observed in Figure 2 holds in general. For ease of notation, let $a^*(t, q, h)$ be the optimal action in state $(t, q, h)$ in response to the arrival at time $h$, where action '0' corresponds to not opening a vial and action '1' corresponds to opening a vial.

**Proposition 1.** *For every $(t, q)$ pair, $t \in \{0, 1, \ldots, T\}$ and $q \in \{0, 1, \ldots, Q\}$, there exists an $h^*(t, q)$ such that for every $h > h^*(t, q)$, $a^*(t, q, h) = 0$.*

**Proof:** If $\hat{h} = \eta$, then the optimal policy is equivalent to the greedy policy and $h^*(t,q) = \eta$ for all $(t,q)$. Now consider $\hat{h} \leq \eta - 1$. Because $\nu(t,q)$ is independent of $h$, it is sufficient to show that $g(t,q,h)$ is monotone nonincreasing in $h$. We proceed by induction, but first note that

$$\sum_{i=1}^{z} p_{D_h}(i) + \sum_{y=h+z}^{\eta} p_{Y_h}(y) = 1. \tag{1.9}$$

Second, note that $D_h$ and $D_{h+1}$ are (shifted) binomial random variables and satisfy the conditions of Theorem (1.a) in [21] which implies that $D_{h+1} \leq_{st} D_h$, or that for all nondecreasing functions $\gamma$,

$$\mathbb{E}[\gamma(D_{h+1})] \leq \mathbb{E}[\gamma(D_h)]. \tag{1.10}$$

As a result,

$$\begin{aligned}
g(t,1,h+1) &= \sum_{y=h+z+1}^{\eta} z p_{Y_{h+1}}(y) + \sum_{d=1}^{z} d p_{D_{h+1}}(d) \\
&= \sum_{y=z+1}^{\eta-h} z p_{D_{h+1}}(y) + \sum_{d=1}^{z} d p_{D_{h+1}}(d) && (1.11) \\
&\leq \sum_{y=z+1}^{\eta-h+1} z p_{D_h}(y) + \sum_{d=1}^{z} d p_{D_h}(d) && (1.12) \\
&= \sum_{y=h+z}^{\eta} z p_{Y_h}(y) + \sum_{d=1}^{z} d p_{D_h}(d) && (1.13) \\
&= g(t,1,h),
\end{aligned}$$

where (1.11) and (1.13) follow from (1.9), and (1.12) follows from (1.10) with $\gamma$ defined as

$$\gamma(d) = \begin{cases} d, & 1 \leq d \leq z, \\ z, & z \leq d \leq z - h + 1. \end{cases} \tag{1.14}$$

Now, assume that $g(t',q',h)$ is nonincreasing in $h$ for all $t' > 0$ and $q \leq q'$. By definition,

$$
\begin{aligned}
& g(t', q'+1, h+1) \\
&= \sum_{y=h+z+1}^{\eta} \Big( z + V(t, q', y) \Big) p_{Y_{h+1}}(y) + \sum_{d=1}^{z} \Big( d + \nu(t-1, q') \Big) p_{D_{h+1}}(d) \\
&= \sum_{y=h+z+1}^{\eta} V(t, q', y) p_{Y_{h+1}}(y) + \nu(t-1, q') \sum_{d=1}^{z} p_{D_{h+1}}(d) + z \sum_{y=h+z+1}^{\eta} p_{Y_{h+1}}(y) + \sum_{d=1}^{z} d p_{D_{h+1}}(d) \\
&= \sum_{y=h+z}^{\eta-1} V(t, q', y+1) p_{Y_h}(y) + \nu(t-1, q') p_{Y_{h+1}}(\eta+1) + z \sum_{d=z+1}^{\eta-h} p_{D_{h+1}}(d) + \sum_{d=1}^{z} d p_{D_{h+1}}(d) \qquad (1.15) \\
&\leq \sum_{y=h+z}^{\eta-1} V(t, q', y) p_{Y_h}(y) + \nu(t-1, q') \Big( p_{Y_h}(\eta) + p_{Y_h}(\eta+1) \Big) + z \sum_{d=z+1}^{\eta-h+1} p_{D_h}(d) + \sum_{d=1}^{z} d p_{D_h}(d) \qquad (1.16) \\
&\leq \sum_{y=h+z}^{\eta} V(t, q', y) p_{Y_h}(y) + \nu(t-1, q') p_{Y_h}(\eta+1) + z \sum_{d=z+1}^{\eta-h+1} p_{D_h}(d) + \sum_{d=1}^{z} d p_{D_h}(d) \qquad (1.17) \\
&= \sum_{y=h+z}^{\eta} V(t, q', y) p_{Y_h}(y) + \nu(t-1, q') \sum_{d=1}^{z} p_{D_h}(d) + z \sum_{y=h+z}^{\eta} p_{Y_h}(y) + \sum_{d=1}^{z} d p_{D_h}(d) \qquad (1.18) \\
&= g(t', q'+1, h),
\end{aligned}
$$

where (1.15) follows from the fact that $p_{Y_{h+1}}(y+1) = p_{Y_h}(y)$ and Equation (1.9); (1.16) follows from the induction assumption, inequality (1.10) and Equation (1.9); (1.17) follows from the fact that, by definition, $V(t', q', \eta) \geq \nu(t'-1, q')$ for $\hat{h} \leq \eta - 1$; and (1.18) follows from Equation (1.9).

Hence, $g(t, q, h)$ is nonincreasing in $h$ for all $t$ and $q$, which implies the existence of $h^*(t, q) \geq \hat{h}$ for all $(t, q)$ pairs. □

We can exploit the structure established in Proposition 1 computationally by tailoring the backward induction algorithm to only consider policies of this form. That is, we can tailor the algorithm to consider timeslots in increasing order for each $(t, q)$ pair until the first $a^*(t, q, h) = 0$ is observed at which point we know that $h^*(t, q) = h$ and can set $a^*(t, q, h') = 0$ for all $h' > h$.

Table 1 summarizes the resulting policy performances for the same numerical example introduced at the beginning of this section. Because the greedy policy is likely to stock out

near the end of the replenishment cycle, the optimal policy outperforms the greedy policy by nearly 36 vaccinations, or 16.2% of the total expected demand. Furthermore, as shown in the last column of Table 1, the expected amount of time that the clinic is closed under the greedy policy (5.6 sessions per cycle) is more than double the expected amount of time that the clinic is closed under the optimal policy (2.4 sessions per cycle).

Table 1: Performance of the optimal and greedy policy for the numerical example.

| | expected number of vaccinations | percentage of demand vaccinated | expected open vial waste (doses) | expected number of closed timeslots (sessions) |
|---|---|---|---|---|
| greedy policy | 157.9 | 71.8% | 62.1 | 5.6 |
| optimal policy | 193.6 | 88.0% | 26.0 | 2.4 |

Table 1 also reports the expected amount of OVW, in doses, over the problem horizon. Let $W(T,Q)$ denote this quantity under the optimal policy. We calculate $W(T,Q)$ by first calculating the expected number of vials that are opened and then calculating the difference between the corresponding number of doses and the expected number of doses administered. Let $N(T,Q)$ denote the total expected number of vials opened under the optimal policy and $N(t,q,h)$ be the analogous quantity starting from state $(t,q,h)$. Hence, $N(T,Q)$ can be computed recursively using the fact that for all $t$ and $q$, analogous to Equation (1.6),

$$N(t,q) = \sum_{h=1}^{\eta+1} N(t,q,h) p_X(h), \tag{1.19}$$

where

$$N(t,q,h) = \begin{cases} \sum_{y=h+z}^{\eta} \Big(1 + N(t,q-1,y)\Big) p_{Y_h}(y) + \sum_{d=1}^{z} \Big(1 + N(t-1,q-1)\Big) p_{D_h}(d), & \\ & h \le h^*(t,q), \\ N(t-1,q), & h > h^*(t,q). \end{cases}$$

The first expression in Equation (1.20) corresponds to an arrival in timeslot $h$, $h \le h^*(t,q)$, when there is no vial currently open. In this case, one new vial is opened immediately and the expected number of additional vials opened is $N(t,q-1,y)$ if another decision is faced

today in timeslot $y$, and $N(t-1, q-1)$ otherwise. The second expression corresponds to the opposite case in which no new vial is opened and service is discontinued for the day. Therefore, the expected OVW in doses is given by

$$W(T, Q) = zN(T, Q) - \nu(T, Q). \tag{1.20}$$

Note that the sum of $W(T, Q)$ and the expected number of vaccines administered may be less than the initial inventory level of $zQ$ doses; the difference corresponds to the expected number of doses that are never reconstituted, i.e., those that are left over in unopened vials at the end of the problem horizon. For the numerical example considered in this section, this value is $220 - (193.6 + 26) = 0.4$ doses under the optimal policy. The expected OVW under the greedy policy is computed similarly.

## 1.4 NUMERICAL RESULTS

In this section, we present the results of our numerical experimentation. In Section 1.4.1, we briefly explore the impact of different values of $\hat{h}$, which dictates the minimum number of working hours per session. In Section 1.4.2, we examine the effect of using different values of $\eta$, the number of timeslots per session. In Section 1.4.3, we conduct sensitivity analyses on the model parameters $T$, $Q$ and $z$. Finally, in Section 1.4.4, we construct some intuitive, easy-to-implement heuristic policies and compare their performance to that of the optimal policy and the greedy policy. Throughout this section, we consider variants of the base case parameter values used in Section 1.3, namely $T = 20$ sessions, $Q = 22$ vials, $z = 10$ doses/vial, $\mu = 11$ patients/session.

To facilitate the discussion, we introduce several metrics. Let $\phi(T, Q)$ be the expected percentage of demand vaccinated under the optimal policy, i.e.,

$$\phi(T, Q) = \frac{\nu(T, Q)}{\mu T}. \tag{1.21}$$

Recall that $W(T,Q)$ is the expected OVW (in doses) as defined by Equation (1.20) and let $\omega(T,Q)$ be the expected percentage of doses wasted during the problem horizon under the optimal policy, i.e.,

$$\omega(T,Q) = \frac{W(T,Q)}{\nu(T,Q) + W(T,Q)}. \tag{1.22}$$

Let $\tilde{\phi}(T,Q)$, $\tilde{W}(T,Q)$ and $\tilde{\omega}(T,Q)$ represent the analogous quantities under the greedy policy.

### 1.4.1 Minimum Number of Guaranteed Timeslots Per Session, $\hat{h}$

The larger the value of $\hat{h}$, the greater the number of guaranteed clinic-hours per Session. However, constraining the problem in this way leaves less room for optimization, which is the main focus of our analysis. Therefore, in the remainder of this section, we set $\hat{h} = 0$.

First, however, to briefly investigate the effect of varying $\hat{h}$ we assume a maximum eight-hour workday divided into 480 timeslots of length 1 minute and vary $\hat{h}$ from 0 (the "unrestricted" policy) to 480 (the greedy policy) in increments of 30 timeslots for the base case outlined at the beginning of Section 1.4. Figure 3 plots the percent loss in the expected number of vaccinations for the unrestricted policy compared to the restricted policy as a function of $\hat{h}$. For example, if the clinic is required to remain open for a minimum of 360 timeslots (i.e., a minimum of 6 hours), then the total expected number of vaccinations would be approximately 4% less than if no requirement were imposed.

For this example, the percent loss in the expected number of vaccines administered is negligible when the required number of working hours per session is less than approximately 5 hours, or 62.5% of the maximum number of working hours, and then increases sharply. This behavior indicates that under the optimal policy, the clinic rarely discontinues service before its fifth hour of operation.

### 1.4.2 Number of Timeslots, $\eta$

The larger the value of $\eta$, the greater the number of timeslots in a session and hence the more justifiable the assumption of at most one arrival per timeslot. However, increasing $\eta$ also increases the problem size. To investigate the effect of varying $\eta$, we assume a maximum eight-hour workday and consider intervals of $\tau$ ranging from 30 minutes ($\eta$=16) to 15 seconds

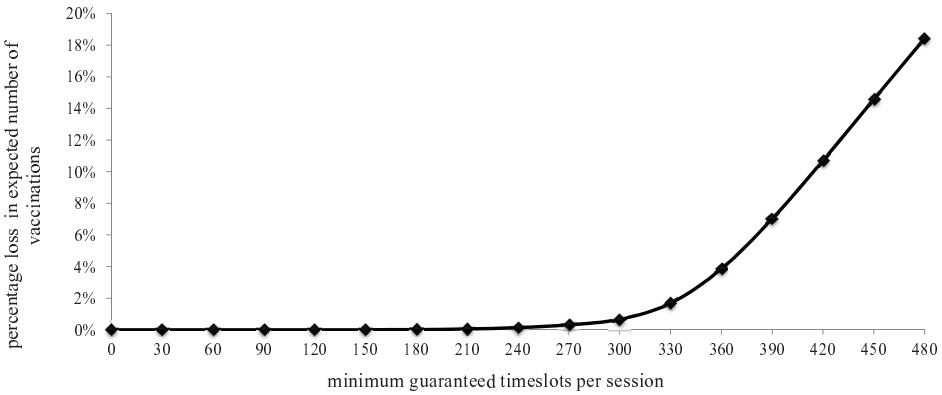



Figure 3: Percentage loss in the expected number of vaccinations between the unrestricted and restricted policy.

($\eta$=1920). We also adjust the value of $p$ corresponding to each value of $\eta$ such that the expected demand over a session is the constant $\mu = \eta p$. Specifically, we consider $(\eta, p) \in \{(16, 0.6875), (32, 0.3438), (96, 0.1146), (480, 0.0229), (960, 0.0115), (1920, 0.0055)\}$ for $T = 20$, $Q = 22$, $z = 10$ and $\mu = \eta p = 11$. Table 2 presents the values of the performance metrics for each value of $\eta$.

Table 2: Sensitivity analysis on the number of timeslots.

| | $\eta = 16$ $p = 0.6875$ | $\eta = 32$ $p = 0.3438$ | $\eta = 96$ $p = 0.1146$ | $\eta = 480$ $p = 0.0229$ | $\eta = 960$ $p = 0.0115$ | $\eta = 1920$ $p = 0.0055$ |
|---|---|---|---|---|---|---|
| $\mu T$ | 220 | 220 | 220 | 220 | 220 | 220 |
| $v(T,Q)$ | 199.8 | 196.3 | 194.3 | 193.6 | 193.5 | 193.4 |
| $\phi(T,Q)$ | 90.8% | 89.2% | 88.3% | 88.0% | 87.9% | 87.9% |
| $W(T,Q)$ | 19.9 | 23.2 | 25.2 | 26.0 | 26.1 | 26.1 |
| $\omega(T,Q)$ | 9.1% | 10.6% | 11.5% | 11.8% | 11.9% | 11.9% |

Note that as $\eta$ increases, we decrease $p$ to maintain constant daily demand, and hence the variability of the daily demand, $\eta p(1-p)$, increases in $\eta$. Therefore, as expected, the average

percentage of demand met, $\nu(T, Q)$, decreases in $\eta$ as shown in Table 2. While the results corresponding to larger values of $\eta$ are in general more "accurate," i.e., closer to continuous decision making, from Table 2 it is clear that there is little to be gained by increasing the value of $\eta$ beyond 480. Therefore, we use this value (corresponding to $\tau = 1$ minute) for the number of timeslots per session for all of our subsequent analysis. In other words, we assume at most one patient arrives per minute throughout each session.

### 1.4.3 Parameters $T$, $Q$ and $z$

One important factor that significantly affects coverage rates is the number of sessions between two successive replenishments, $T$. Examining this parameter has practical value as well, given that the policies for EPI vaccinations vary widely by country: clinics in some locations vaccinate patients every day of the week, others only once a week, and yet others only once a month. In this section, we first analyze the relationship between the number of sessions between replenishments and the expected number of vaccinations. In our comparisons, we vary $T$, but adjust $p$ such that demand remains constant, i.e., we consider $(T, p)$ $\in \{(20, 0.0229), (16, 0.0286), (12, 0.0382), (8, 0.0573), (4, 0.1146), (1, 0.4583)\}$. The first five pairs of values correspond to 5, 4, 3, 2 and 1 sessions per week, respectively; the last pair of values corresponds to one session per month.

For each of these values of $T$, Figure 4 plots the percentage of demand satisfied, $\phi(T, Q)$, and the percentage of vaccines wasted, $\omega(T, Q)$, under the optimal policy and the greedy policy using the same baseline parameter values ($Q = 22$ vials, $z = 10$ doses/vial, $\eta = 480$). As Figure 4 indicates, if there are fewer sessions between successive replenishments, then the optimal policy is able to fulfill a greater proportion of the total demand. This observation validates the observation in the field [24] and in other contexts, that in general, consolidating the same total demand into fewer discrete time periods results in less OVW and hence greater coverage rates.

Although the results in Figure 4 favor a scenario in which the clinic is open for fewer sessions per replenishment period, it does so under the assumption that the *total* demand within the replenishment period is insensitive to the number of sessions per period. An

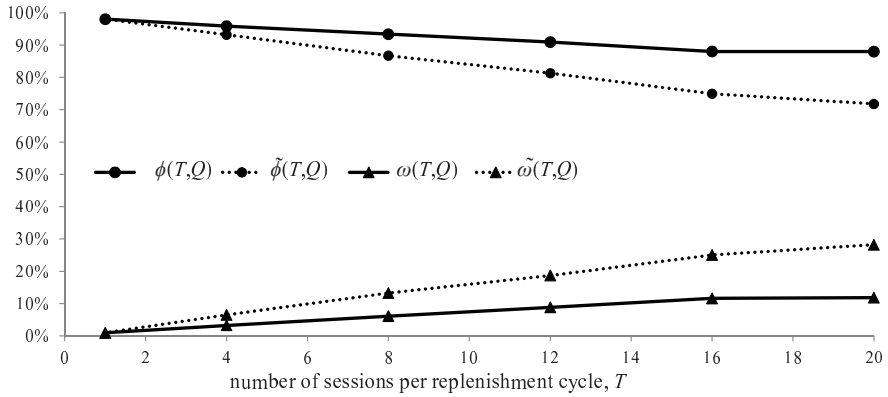



Figure 4: Percentage of demand vaccinated and percentage of doses wasted as a function of the number of sessions per replenishment cycle.

argument could be made, however, that this assumption is not necessarily true because reducing the number of sessions that the clinic is open may create an inconvenience for some patients. This phenomenon is especially likely to be true in countries with dispersed populations and poor infrastructure, and indeed, is one of the arguments offered in favor of keeping clinics open more often. If as $T$ decreases, the expected demand per session decreases, then the advantage from consolidation will be offset by a decrease in the total number of vaccinations because of decreased overall demand. Thus, the total number of vaccinations corresponding to the values plotted in Figure 4 is likely to overstate the true number of vaccinations.

For this reason, for each value of $T$ we also determine the maximum amount by which demand could decrease without resulting in poorer performance than the base case with $T = 20$ sessions. From Table 2 (in the column for $\eta = 480$), the expected *number* of vaccinations over the problem horizon is 193.6 out of an expected total demand of 220. Let $\delta_T$ be the maximum amount by which the total mean demand can decrease (in which case the daily expected demand is $\frac{220-\delta_T}{T}$) under value $T$, and yet result in no fewer than 193.6 vaccinations on average.

Figure 5 plots the values of $\frac{\delta_T}{T} \times 100\%$ as a function of $T$. As the figure indicates, switching from 5 to 4 sessions requires almost no change in demand in order to be worthwhile, whereas switching from 5 to 3 sessions would require a decrease of more than 6% in the daily demand rate for the consolidation to be unattractive. Under a just one vaccination session per month policy, a reduction in demand of up to 12% would achieve at least the same number of vaccinations as in the base case. Note that the corresponding values of $\phi(T, 22)$ are given by $(\frac{193.6}{220-\delta_T})100\%$ and these values are all greater than the values plotted in Figure 4. In summary, Figure 5 allows policy makers to weigh the pros and cons of reducing the number of sessions per period based on their estimates of any corresponding reductions in demand and the targeted expected number of vaccinations.

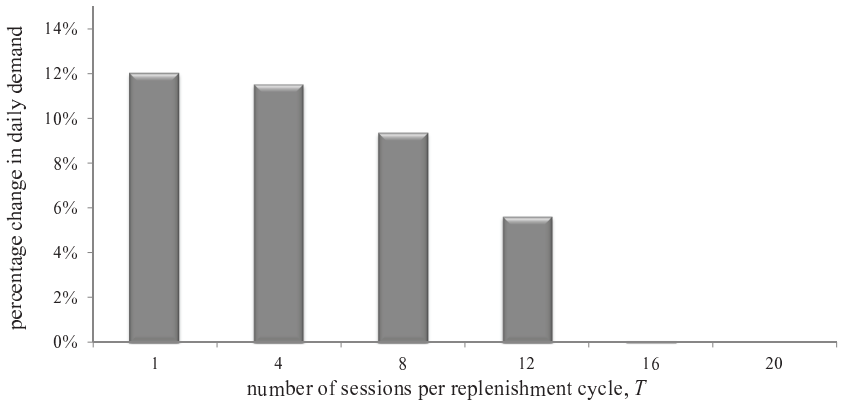



Figure 5: Sensitivity analysis on number of sessions per replenishment cycle.

In our analysis thus far, the initial inventory, $Q$, has been set equal to $\mu T$, the total expected demand over the problem horizon. Next, we relax this assumption and analyze the effect of carrying buffer stock. Let $\alpha$ be the percentage of the initial inventory that is buffer stock, i.e.,

$$\alpha = \left(\frac{Qz - \mu T}{\mu T}\right) 100\%.$$

Assuming the baseline values of $T = 20$ sessions, $\mu = 11$ patients/session, $\eta = 480$ per session and $z = 10$ doses/vial, we examine values of $Q \in \{22, 23, \ldots, 44\}$, which correspond to

$\alpha \in \{4.6\%, 9.1\%, \ldots, 100\%\}$, respectively. It is worth mentioning that the WHO recommends that clinics use a buffer of 25% when setting their order-up-to levels. Figure 6 displays the results and, as expected, indicates that a higher initial inventory results in higher coverage. Also, the greater the buffer, the better the greedy policy performs compared to the optimal; this observation is intuitive because having more vaccine on hand renders the myopic nature of the greedy policy closer to optimal.

On the other hand, as seen in Figure 6, it is interesting that the percentage of vaccines wasted increases in $\alpha$ under the optimal policy, $\omega(T, Q)$, but is constant in $\alpha$ under the greedy policy, $\tilde{\omega}(T, Q)$. This observation can be explained by the fact that the greedy policy does not depend on the number of sessions remaining until replenishment. That is, on average, the wastage rate under the greedy policy is the same every session until the initial inventory is exhausted, at which point there is no additional wastage. Under the optimal policy, however, a larger buffer results in less conservative actions, resulting in greater wastage. Overall, the optimal policy requires smaller buffers to achieve the same level of service as the greedy policy, but with less wastage.

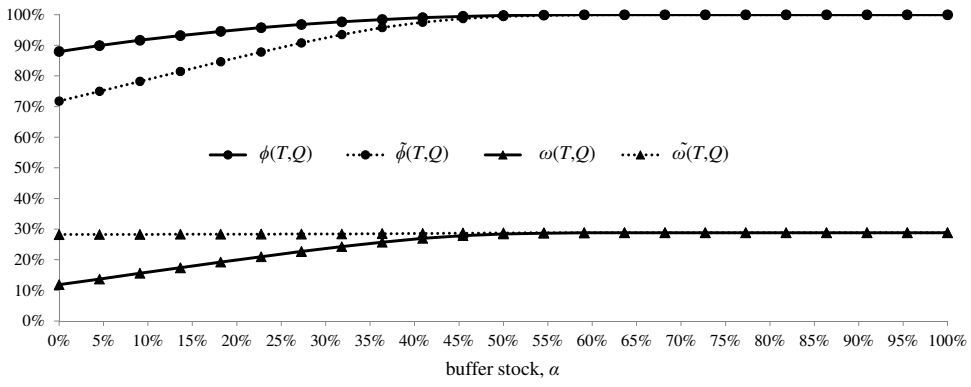



Figure 6: Percentage of demand vaccinated and percentage of doses wasted as a function of buffer stock.

Finally, we examine the effect of varying the vial size, $z$. Vaccine manufacturers often produce vials of various sizes, and the question of an "optimal" vial size remains open [23, 24]. Vials containing fewer doses result in less OVW, but in general, they also tend to be more expensive to manufacture on a per-dose basis and take up more cold storage space in the vaccine supply chain. We examine the effect of various values of $z$ between 1 and 25 doses under three scenarios. Figure 7 graphs demand fulfillment, $\phi(T, Q)$, and wastage, $\omega(T, Q)$, as a function of $z$ for: $(i)$ $T = 4$ sessions (one session per week), $\mu = 55$ patients/session, $(ii)$ $T = 12$ sessions (3 sessions per week), $\mu = 18.33$ and $(iii)$ $T = 20$ sessions (5 sessions per week), $\mu = 11$ patients/session. Ideally, each case considered would have the same constant buffer size (e.g., $\alpha = 0$), however, because the number of doses per vial is discrete, rounding is necessary to maintain an integer value of $Q$. This rounding results in the slightly bumpy nature of the curves in Figure 7.

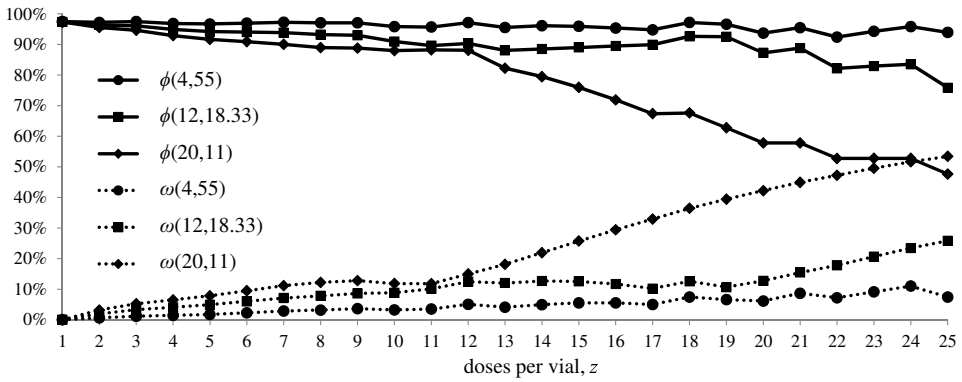



Figure 7: Percentage of demand vaccinated and percentage of doses wasted as a function of vial size.

As shown in Figure 7, smaller values of $z$ result in less wastage and a greater percentage of demand fulfilled. It is worth noting that for each $T$, there exists a threshold close to $\mu$ (daily demand) for the number of doses per vial such that values above this threshold sharply reduce coverage and increase waste. This result stems from the fact that if there are fewer

than $T$ vials on hand at the beginning of the replenishment period, then unless there are sessions with *no* demand, stocking out is almost certain. Furthermore, with fewer sessions, the benefits of smaller vial sizes are less pronounced.

### 1.4.4 Heuristic Policies

In this section, we consider two heuristic policies to investigate the degree to which the greedy approach may be improved upon without resorting to the complexities involved in implementing the optimal policy. These heuristics are based on the insights gained through the numerical experimentation, but are simpler in form in that they do not depend on the time of day.

First, consider the following heuristic: if at the beginning of the session the current stock level is not sufficient to serve the expected demand over the remaining sessions (not including the current session), then the clinic remains closed. Otherwise the clinic opens and serves all of the daily demand, as the greedy policy would. Let $\hat{V}(t,q,h)$ denote the expected number of vaccines administered prior to the next replenishment starting from state $(t,q,h)$ under this heuristic policy. Then, for $t>0$, $q>0$, $h\leq\eta$,

$$\hat{V}(t,q,h)=\sum_{y=h+z}^{\eta}\Big(z+\hat{V}(t,q-1,y)\Big)p_{Y_h}(y)+\sum_{d=1}^{z}\Big(d+\hat{V}(t-1,q-1,0)\Big)p_{D_h}(d), \quad (1.23)$$

where

$$\hat{V}(t,q,0)=\begin{cases}\hat{V}(t-1,q,0), & q\leq\frac{(t-1)\mu}{z},\\ \sum_{x=1}^{\eta}\hat{V}(t,q,x)p_X(x)+\hat{V}(t-1,q,0)p_X(\eta+1), & q>\frac{(t-1)\mu}{z},\end{cases} \quad (1.24)$$

corresponds to the total expected number of vaccines to be administered over $t$ sessions when starting the current session with $q$ vials.

Somewhat surprisingly, this heuristic performs nearly identically to the greedy policy. The reason behind this observation is that both policies result in roughly the same number of open sessions (on which they both attempt to serve all of the demand and hence both result in approximately the same expected volume of OVW) and the same number of closed sessions. The only difference is that under the greedy policy, the closed sessions all occur at

the end of the replenishment cycle, whereas under the heuristic policy, the closed sessions are scattered throughout the planning period. For this reason, we do not report results for this heuristic.

Instead, we examine a modified version of this heuristic under which the decision of whether or not to open the clinic at the beginning of each session is made using the same rule. However, under the modified heuristic, once the decision is made to open the clinic, not all of the session's demand is necessarily fulfilled. Rather, whenever a vial is emptied, the current inventory is compared to the expected demand over the remaining sessions (not including the current session) and if the inventory is less than or equal to this value, then the clinic closes for the day; otherwise, the clinic remains open. Figure 8 depicts this time-independent modified heuristic policy graphically for $t = 10$ and $t = 20$ sessions remaining (the solid lines). For the sake of comparison, Figure 8 also includes the corresponding optimal time-dependent thresholds from Figure 2.

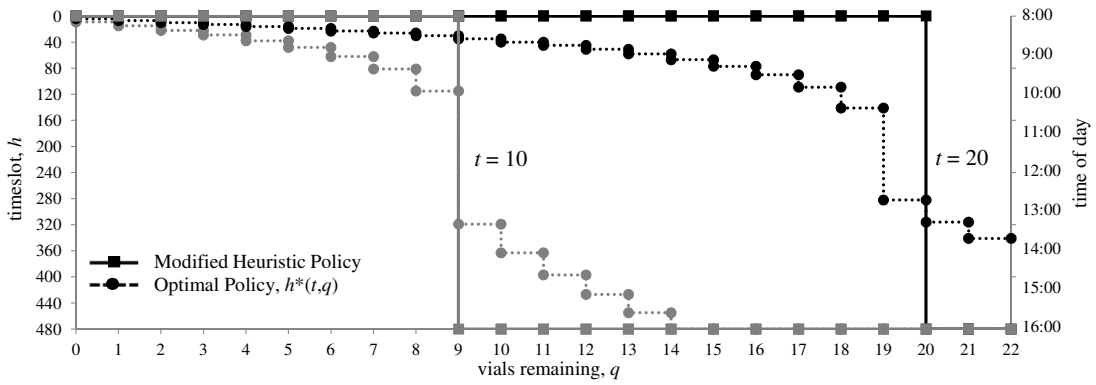



Figure 8: Optimal and modified heuristic policy threshold.

Let $\check{V}(t, q, h)$ denote the expected number of vaccines administered prior to the next replenishment starting from state $(t, q, h)$ under this modified heuristic policy. Then for

$t > 0$, $q > 0$, $h \leq \eta$,

$$\check{V}(t,q,h) = \begin{cases} \check{\nu}(t-1,q), & q \leq \frac{(t-1)\mu}{z}, \\ \sum_{y=h+z}^{\eta} \Big( z + \check{V}(t,q-1,y) \Big) p_{Y_h}(y) + \sum_{d=1}^{z} \Big( d + \check{\nu}(t-1,q-1) \Big) p_{D_h}(d), & q > \frac{(t-1)\mu}{z}, \end{cases} \tag{1.25}$$

where

$$\check{\nu}(t,q) \equiv \sum_{x=1}^{\eta+1} \check{V}(t,q,x) p_X(x) \tag{1.26}$$

is the total expected number of vaccines to be administered over $t$ sessions when starting the current session with $q$ vials. Table 3 compares the performance of the modified heuristic, the greedy policy and the optimal policy for the baseline set of parameter values presented in Section 1.3. The results show that the modified heuristic yields a considerable improvement in performance over the greedy policy and is quite comparable with the optimal policy, while being easier to implement than the optimal policy.

Table 3: Policy performance.

| | expected number of vaccinations | percentage of demand vaccinated |
|---|---|---|
| greedy policy | 157.9 | 71.8% |
| modified heuristic policy | 190.0 | 86.4% |
| optimal policy | 193.6 | 88.0% |

Lastly, in Figures 9 and 10, we present results of three policies for two different vial sizes ($z = 10$ and $z = 20$, respectively) for various values of the daily demand. We consider the baseline values of $T = 20$ sessions, $\eta = 480$ per session and $\alpha = 0\%$. If the mean daily demand, $\mu$, is significantly less than $z$, then as $\mu$ increases there is a steep increase in the percentage of demand that is satisfied; this result can be explained by the fact that it is beneficial to have a value of $\mu$ closer to $z$ than to zero, because there is less of a chance of wasting a portion of a vial that is opened. However, this effect diminishes, particularly

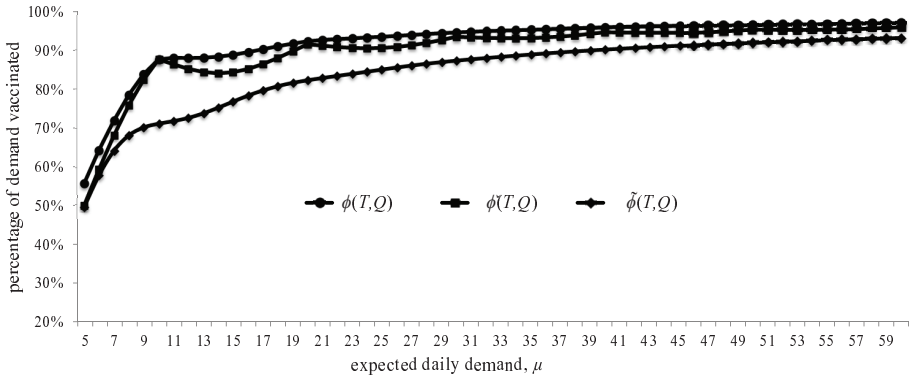



Figure 9: Sensitivity analysis of policies over daily demand for $z = 10$.

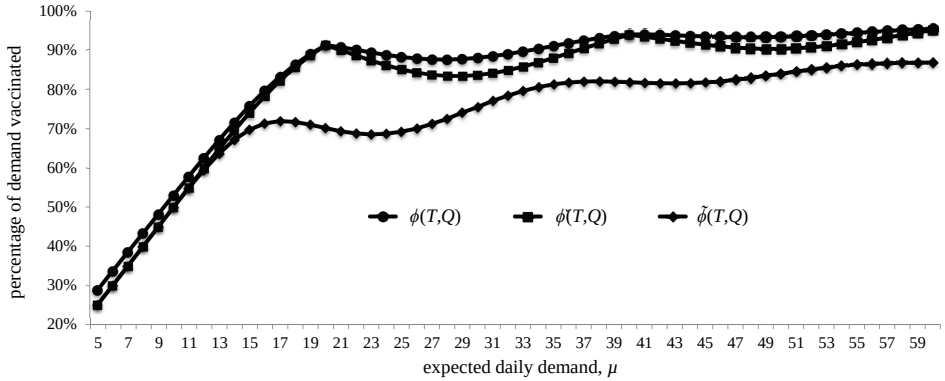



Figure 10: Sensitivity analysis of policies over daily demand for $z = 20$.

for the greedy heuristic, as $\mu$ approaches $z$ because the probability that the daily demand will slightly exceed $z$ begins to increase, which results in a large amount of waste under the greedy policy. This effect becomes less pronounced as $\mu$ assumes larger values between integer multiples of $z$. As expected, all three policies perform better when $\mu$ is an integer multiple of $z$.

## 1.5 CONCLUSION

In Chapter 1, we address the problem of how to optimally conserve multi-dose vaccines in remote locations in order to minimize OVW while administering as many vaccinations as possible. We analytically establish that the optimal policy is of a threshold type. The problem is a significant one, especially in lower and middle-income countries administering vaccines as part of the WHO-EPI program, but it is also of interest to other stakeholders in the vaccine delivery process. Policies that dynamically determine optimal clinic hours are derived using a novel MDP approach. Sensitivity analysis lends insight into the benefits of consolidating demand on fewer sessions between replenishments, adopting different buffer stock strategies and using different vial sizes. Lastly, a simple heuristic is shown to provide significant improvement over existing practice and perform competitively with the optimal policy. Although the current approach assumes that each patient demands only one type of vaccine, the results can be used in settings where patients require multiple vaccines if decisions for each type of vaccine can be made separately, i.e., the optimal policies for each type of vaccine can be implemented in parallel.

We implement several extensions to the model. For example, we relax the assumption that the number of doses per vial is a known constant, i.e., treating the vaccine yield from each vial as a random variable in Section 2.5. In practice, this variability often occurs, and for example, a 20 dose vial might not yield exactly 20 doses. Another extension which is considered in Chapter 3, is incorporating the option of delaying service and/or asking patients to return at a later time/session rather than assuming that an unserved patient is always lost. Delaying service, for example, could be captured by assuming that if a patient arrives to find the clinic closed that patient will return to the clinic the following session with some given probability.

# 2.0 CUSTOMIZING IMMUNIZATION CLINIC OPERATIONS TO MINIMIZE OPEN VIAL WASTE

## 2.1 INTRODUCTION

The existing literature on multi-dose vials is rather limited, and primarily focuses on the economic implications of single- vs. multi-dose vials. In contrast, in Chapter 1, we address another means for controlling OVW, namely that of vaccine *administration* from multi-dose vials, using a rigorous mathematical approach. In this chapter we consider the same decision making problem and evaluate operating strategies that maximize patient coverage by controlling open vial waste, whereas, the focus in Chapter 1 is on model formulation, policy structure and limited sensitivity analysis under the single objective of maximizing the mean number of vaccinations administered.

The contributions of this chapter are five-fold. First, we pair combined analysis of the MDP model with simulation to perform descriptive analysis of the distribution of session duration induced by an optimal administration policy. We examine this novel metric of policy performance because a coverage-maximizing/ waste-minimizing administration policy that induces large variability in a clinic's hours of operation may inconvenience patients and lead to undesirable long-term consequences. Second, we explore means by which a clinic can directly control patient convenience by imposing a minimum number of guaranteed hours per session or increasing the session frequency. We explore how these two means of control interact with each other as well as when the latter can counterintuitively affect coverage and wastage. Third, we propose a novel, easy to implement static heuristic policy that induces zero variability in session duration and compare its performance to that of two other heuristic policies. Fourth, we introduce the concepts of random vial-yield and vial failures to

this problem setting and assess their impact on the performance of the optimal and heuristic policies. Lastly, we use data available for three countries to perform novel costs analyses for a single vaccine over all GAVI countries, which suggest potential savings on the order of $4.6 million.

The remainder of this chapter is organized as follows. In Section 2.2, we provide a new simulation model created to evaluate additional performance metrics of interest. In Section 2.3, we conduct extensive computational analyses to generate insights on the relationships between day-to-day clinic operations and the vaccine administration policy. In Section 2.4, we present our heuristic policy analysis. We then, in Section 2.5, study the performance of the optimal policy and the heuristic policies in the presence of vial failures and random vial yield. In Section 2.6, we summarize the results in the form of some general operational recommendations based on the analysis in Sections 2.3 and 2.5 and estimate the procurement cost savings realized by switching to the optimal policy. Lastly, in Section 2.7, we discuss limitations of the work and possible future extensions.

## 2.2 SIMULATION MODEL

In this section, we present a simulation model used to assess the performance of a given vaccine administration policies generate by the MDP model developed in Section 1.2. This model simulates a clinic that administers one type of vaccine over one replenishment cycle. Similar to the MDP model, sessions are divided into $\eta = 480$ timeslots, and arrivals occur at the beginning of each timeslot according to a Bernoulli distribution with probability $\frac{\mu}{\eta}$.

A flowchart of the simulation model is shown in Figure 11. In each replication, the clinic starts at the beginning of the replenishment cycle with $Q$ vials on hand. Arrivals occur randomly in each timeslot and $h$ denotes the current timeslot. Whenever an open vial is available, clinicians vaccinate arriving patients. Whenever an arrival occurs and no vial is open, the decision to open a new vial is made according to the vaccine administration policy, $h^*(t, q)$. Hence, clinicians discontinue service at the first timeslot greater than $h^*(t, q)$ when no open vial is available. As mentioned in Section 1.2, unserved patients are lost and leave

the system with no effect on future demand. Note that the simulation of each session extends beyond the time that the clinic discontinues service so that the number of unserved patients can be tallied.

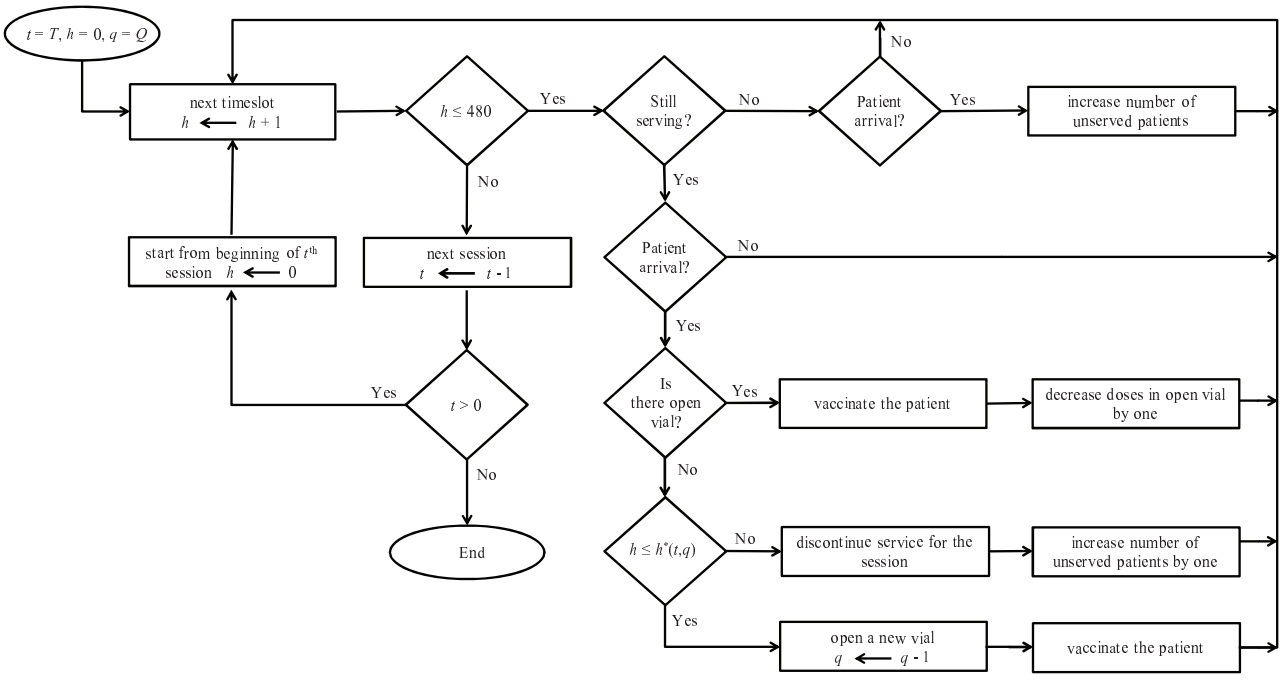



Figure 11: Simulation model flowchart.

## 2.3 IMPACT OF CLINIC OPERATIONS ON POLICY PERFORMANCE

As explained in Section 1.2, the MDP model maximizes the expected number of vaccinations under a given set of parameter values. In this section, we analyze how the parameters under the clinic's control (the number of vaccination sessions per replenishment cycle and guaranteed hours per session) interact with parameters that the clinic does not typically

control (e.g., vial size, replenishment interval and initial inventory) to determine policy performance. We assess these relationships and their inherent trade-offs as a function of coverage rate, OVW and additional new metrics (e.g., closing time distribution, prediction intervals on the number of open hours, and percentage gain/loss compared to benchmark vaccine administration policies).

More specifically, in Section 2.3.1, we analyze session duration under the simplest form of the optimal policy (i.e., one with no set number of guaranteed hours). That is, we evaluate, via simulation, how session duration changes as a function of the number of sessions per cycle and observe that in some cases, considerable variation exists from session to session within a replenishment cycle. Such variability could create operational challenges. Hence, in Section 2.3.2, with the aim of preventing discontinuation of vaccine administration "too early" during a session, we consider imposing some minimum number of working hours per session and propose a metric for determining the most desirable duration of these guaranteed hours. In Section 2.3.3, we evaluate the effect of the number of sessions per replenishment cycle on the performance of the optimal policy while operating the clinic for a specified number of guaranteed hours (i.e., using the results of Section 2.3.2). In doing so, we focus on determining the appropriate number of sessions per replenishment cycle in order to achieve a coverage target under an optimal policy with guaranteed hours. The goal is to offer as many sessions as possible to facilitate patient access to the clinic, without compromising coverage.

The main criterion used to evaluate the performance of a vaccine administration policy is the expected number of vaccinations per cycle, i.e., coverage rate which is introduced in Equation 1.21 in Section 1.4. By coverage rate, we mean the expected percentage of patients who arrive and are vaccinated.

Throughout this section, we consider values of $T \in \{4, 8, 12, 16, 20\}$, which are equivalent to 1-5 sessions per week for a one-month replenishment cycle; values of $z \in \{1, 5, 10, 20\}$ which represent the most common vial sizes; and set the initial inventory to:

$$Q = 1.25\frac{\mu T}{z}, \tag{2.1}$$

which corresponds to a 25% buffer stock, i.e., the clinic starts each replenishment cycle with 25% more doses than the expected cycle demand, $\mu T$. These ranges represent values

commonly seen in practice [24, 10]. Furthermore, consistent with Chapter 1, in the remainder of this section we assume that the number of sessions per cycle does not significantly affect the distribution of the demand between two stock replenishments. That is, in comparing different values of $T$, we assume that $\mu$ scales accordingly so that $\mu T$ remains constant. We consider two values for the expected number of arrivals per cycle, namely $\mu T = 96$ and $\mu T = 288$. These are typical values observed in practice for medium-size and large clinics, respectively, for a two-dose vaccine schedule (see Section 2.6.2). Note that to maintain consistency in buffer stock across all of the problem instances considered, both of these values always result in an integer value for $Q$, the initial vial inventory, as defined by Equation (2.1).

### 2.3.1 Session Duration

We refer to the optimal vaccine administration policy with no requirements on the minimum number of working hours per session (i.e., $\hat{h} = 0$, where $\hat{h}$ is the earliest timeslot in which the clinic may elect to discontinue service) as the "unrestricted policy." In this section, we first analyze the frequency with which it is optimal to discontinue service early under the unrestricted policy as a function of the number of sessions per cycle for various vial sizes. We then investigate the effect of the number of sessions per cycle on the full distribution of the closing time and the variability of the number of operating hours per session under the unrestricted policy.

Based on $10,000$ replications of the simulation model, the black regions in Figure 12 indicate the percentage of sessions that terminate before the end of the scheduled operating hours under the unrestricted policy. As can be seen, a larger vial size generally results in a higher frequency of early termination. This behavior can be explained by the fact that, on average, a smaller vial size results in less OVW, and therefore higher inventory levels throughout the replenishment cycle, which permits the clinic to remain open longer. (The exception to this observation is the ($z = 20$, $T = 20$) case compared to the ($z = 10$, $T = 20$) case. Note, however, that the $z = 20$, $T = 20$ case is somewhat extreme in that the initial inventory is so low compared to the number of sessions that the clinic is unable to open a vial in each of the 20 planned sessions.)

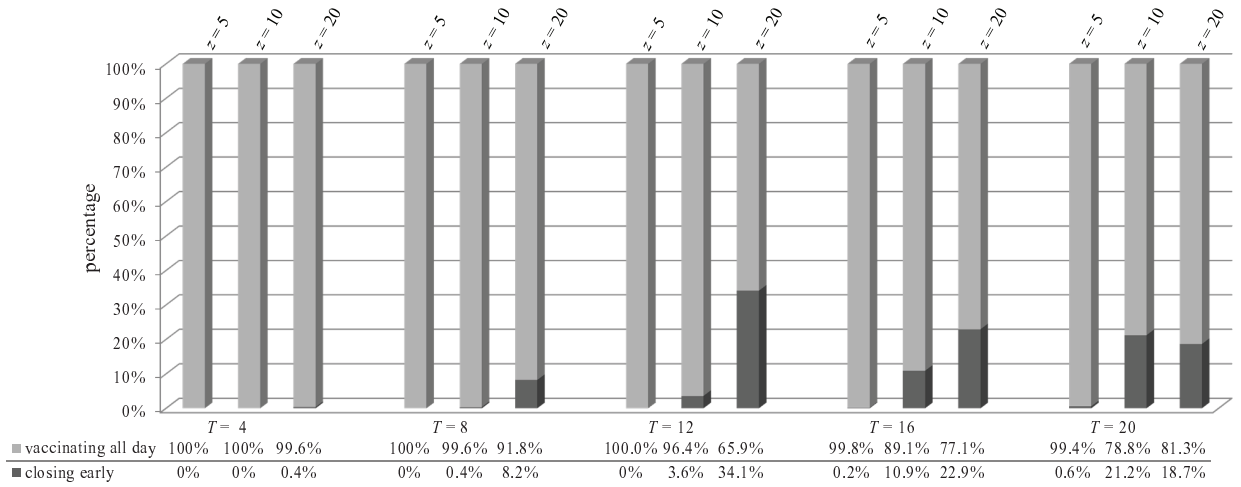
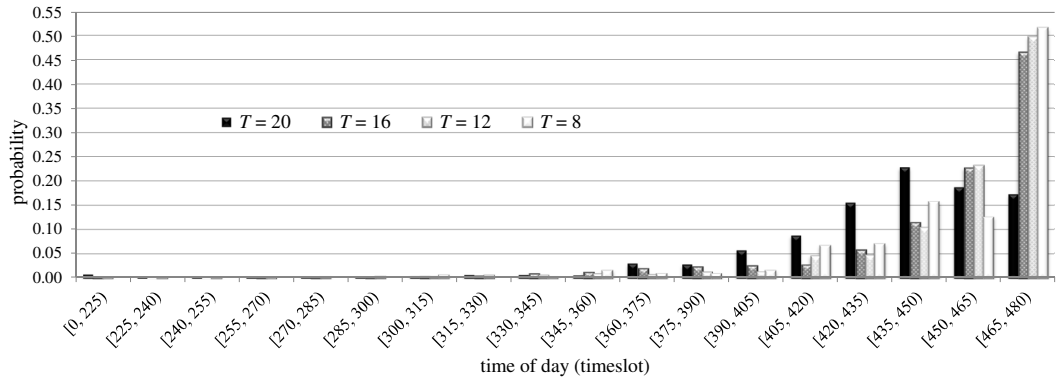


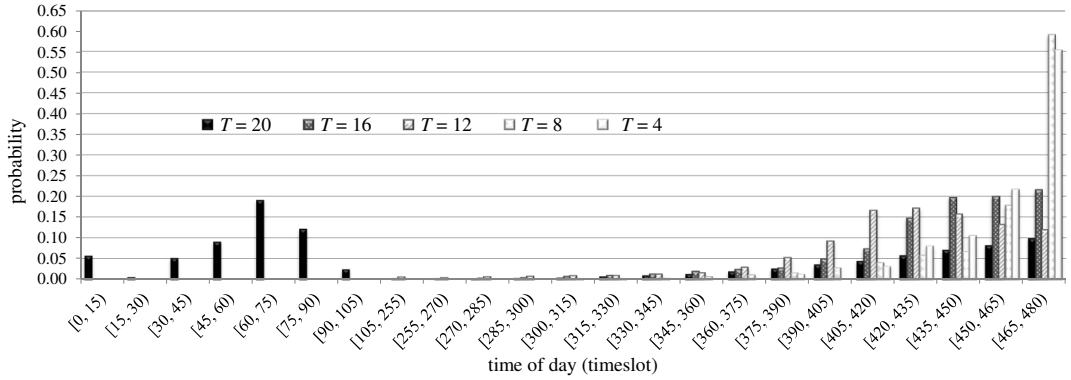

Figure 12: Percentage of sessions that terminate before the end of the scheduled operating hours ($zQ = 360$, $\mu T = 288$ and $\hat{h} = 0$).

Interestingly, however, Figure 12 suggests that the frequency of early closure is not necessarily monotone in the number of sessions per cycle. For $z = 20$, the percentage of sessions that terminate early is larger for $T = 12$ than for $T \in \{8, 16, 20\}$. This behavior is somewhat counterintuitive because increasing the number of sessions generally increases opportunities for OVW and hence entices the clinic to act more conservatively by closing early. To explain this behavior, consider the ratio $\frac{Q}{T}$, i.e., the "number of vials per session" if the vials are divided equally across the sessions. Intuitively, the vaccine administration policy is expected to result in fewer early closures when this ratio is large (plenty of inventory) and/or close to an integer value (an integral number of vials are needed per session, in expectation). For example, for $z = 20$ in Figure 12, the number of vials per session is 1.5 for $T = 12$ sessions, whereas the number of vials per session is 1.1 for $T = 16$ sessions. Although offering 4 more sessions, when $T = 16$ a smaller percentage of these sessions terminate early because each session is likely to need one (and only one) vial to satisfy demand. When $T = 12$, on the other hand, there is a good chance that each session will need a second vial, but that much of this vial will go unused if opened, resulting in the optimal policy dictating more frequent early closures.

Based on the same simulation data, Figure 13 explores the entire distribution of the closing time, as opposed to simply whether or not the closing time occurs before the scheduled session end time, under the unrestricted policy. Because a large percentage of the simulated sessions operate for the fully planned duration (see Firgure 12), Figures 13(a) and 13(b) plot the closing time distribution *conditional on* terminating the vaccination session before the end of the scheduled operating hours.




(a) $z = 10$, $Q = 36$, $\mu T = 288$, $\hat{h} = 0$.



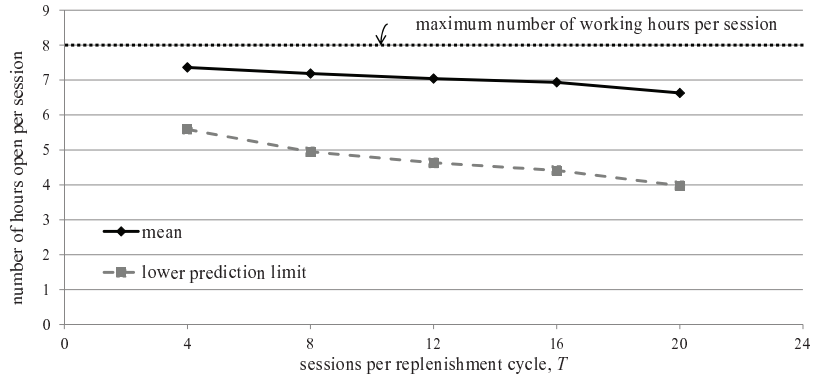

(b) $z = 20$, $Q = 18$, $\mu T = 288$, $\hat{h} = 0$.

Figure 13: Closing time distribution conditional on discontinuing service before the end of the scheduled operating hours.

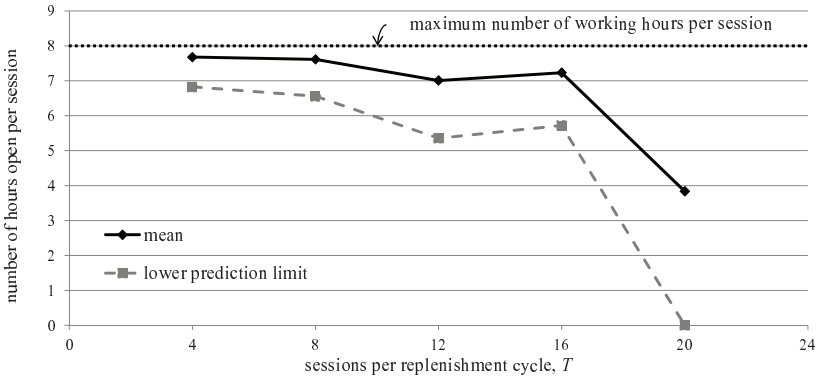
Based on Figures 13(a) and 13(b), the closing time distribution increases monotonically in the time of day for fewer sessions per cycle, whereas this behavior is not monotonic for a greater number of sessions per cycle. For example, in Figure 13(a), for $T = 20$ the frequency of closing early becomes significant after timeslot 360 (e.g., 2pm for an 8am-4pm session). After timeslot 360, the distribution increases up to timeslot 450 (e.g., 3:30 pm for an 8am-4pm session), at which point it attains its highest value and then decreases. That is, early closing times tend to occur earlier in the day (though generally less often) when the number of sessions per cycle is larger. This non-monotonic behavior occurs because at the beginning of the cycle under larger $T$, there is a higher chance that preserving vials for future use results in a larger number of vaccinations in the subsequent sessions. Therefore, the optimal policy is more conservative regarding opening the last vial at the beginning of replenishment cycle and as a result, an earlier closing time occurs more frequently. Indeed, for the extreme case ($T = 20$) in which the initial number of vials is less than the number of sessions in Figure 13(b), very early closures are not uncommon. Very early closures occur when the first arrival occurs relatively late in the day, in which case it is optimal to retain all vials for later sessions in which the first arrival is likely to occur earlier in the day resulting in a higher chance of vaccinating more patients.

Next, we consider the variability of the hours of operation under the unrestricted policy by constructing prediction intervals. Clearly, larger variability is undesirable from a patient perspective. The number of hours open is bounded above by the maximum number of hours per session. Therefore, to characterize the variability of the operating hours we construct a 99% one-sided prediction interval based on 10,000 replications of the simulation. Figure 14 reports the resulting prediction intervals for several problem scenarios. Generally speaking, the mean (respectively, the variance) of the number of operating hours decreases (respectively, increases) in the number of sessions per cycle. Therefore, for a larger number of sessions, the closing time is less predictable, which results in potentially more inconvenience for patients. In Figure 14(a), the variability is strictly increasing in the number of sessions per cycle, whereas in Figure 14(b), the variability is not monotonic. In addition to having a higher percentage of sessions that terminate before the end of the scheduled operating hours (see Figure 12), the $T = 12$ scenario in Figure 14(b) has more variability in the conditional

number of hours open than $T = 8$ and $T = 16$. That is, for the case with $\frac{Q}{T} = 1.5$, not only does the optimal policy result in more frequent early closures, but these early closure times are more erratic. (The dramatic increase in the width of the prediction interval for $T = 20$ is again due to the fact that for this scenario, the initial number of vials is less than the number of sessions, resulting in at least two sessions with zero operating hours.)




(a) $z = 5$, $Q = 24$, $\mu T = 96$, $\hat{h} = 0$.




(b) $z = 20$, $Q = 18$, $\mu T = 288$, $\hat{h} = 0$.

Figure 14: 99% prediction interval on the number of hours open per 8-hour session, conditional on discontinuing service before the end of scheduled operating hours.

Unpredictable closing times can be inconvenient, especially in a developing country in which travel logistics to reach a vaccination site can be difficult for a large portion of the

population. This unpredictability motivates the need to define a fixed minimum number of working hours per session during which clinicians vaccinate unconditionally. In the following section, we investigate the impact of imposing such guaranteed hours on the vaccine administration policy.

### 2.3.2 Guaranteed Hours per Session

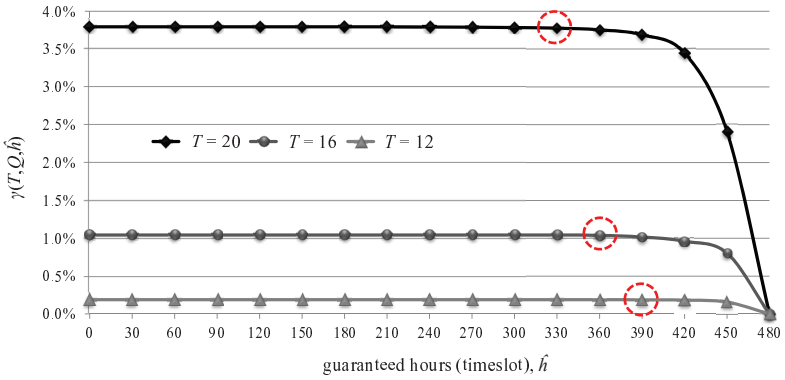
In Section 1.2, we introduce the concept of a minimum number of working hours or "guaranteed hours" per session, i.e., an earliest closing time, in order to decrease the inconvenience associated with unpredictable operating hours. The analysis of this parameter in Section 1.4.1, however, is limited in scope. Hence, in this section, we analyze the performance of the "restricted policy" (i.e., the optimal vaccine adminstration policy under a required minimum number of working hours per session) as a function of the number of guaranteed hours per session. We also propose a metric that can be used to determine an appropriate number of guaranteed hours per session based on an acceptable percentage drop in the performance of the restricted policy versus the unrestricted policy.

Let $\nu(T, Q, \hat{h})$ be the expected number of vaccinations administered between two stock replenishments over $T$ sessions, with $Q$ initial vials and $\hat{h}$ guaranteed hours per session (expressed in timeslots, e.g., $\eta = 480$ corresponds to the maximum number of working hours per session, say 8, in which case $\hat{h} = 240$ corresponds to a minimum of four hours per session). Furthermore, let $\gamma(T, Q, \hat{h})$ be the percentage gain in the expected number of vaccinations under the restricted vaccine administration policy versus the greedy policy (as described in Section 2.1) over $T$ sessions, with $Q$ initial vials and $\hat{h}$ guaranteed hours per session, i.e.,

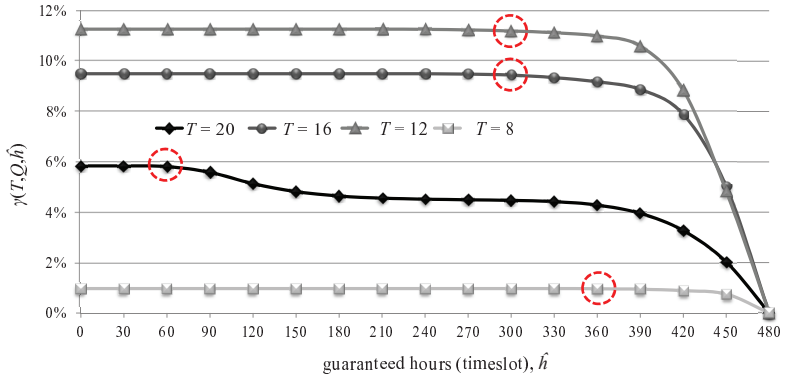
$$\gamma(T, Q, \hat{h}) = \frac{\nu(T, Q, \hat{h}) - \nu(T, Q, \eta)}{\nu(T, Q, \eta)} \times 100. \tag{2.2}$$

Figure 15 parts (a) and (b) report the percentage gain as a function of guaranteed hours per session based on the expected number of vaccinations as determined by the MDP model. As seen in these figures, a sharp drop occurs in the performance of the restricted policy for large $\hat{h}$. This drop in performance is caused by the fact that imposing a large number of guaranteed hours may require clinicians to open new vials close to the end of the day,

which are then not used completely. Prior to this point, however, the curves are relatively flat, indicating that offering patients the convenience of a reasonable number of guaranteed hours does not significantly degrade policy performance. Furthermore, as anticipated based on Figure 12, imposing guaranteed hours does not significantly affect the performance of the vaccine administration policy for smaller vial sizes because the optimal policy does not typically stop vaccinating until the end of the session as planned. When larger vial sizes are used, defining the number of guaranteed hours is more critical, as doing so can negatively impact the performance of the vaccine administration policy.




(a) $z = 10$, $Q = 36$, $\mu T = 288$.




(b) $z = 20$, $Q = 18$, $\mu T = 288$.

Figure 15: Percentage gain in expected number of vaccinations versus the greedy policy as a function of guaranteed hours.

A reasonable value of the minimum number of working hours per session would guarantee acceptable performance of the vaccine administration policy without eroding patient convenience. To explore this trade off, let

$$\hat{h}^*(T, Q, \alpha\%) = \max\left\{30\left\lfloor\frac{\hat{h}}{30}\right\rfloor : \frac{\gamma(T, Q, 0) - \gamma(T, Q, \hat{h})}{\gamma(T, Q, 0)} \leq \alpha\%, 0 \leq \hat{h} \leq \eta\right\}. \tag{2.3}$$

That is, let $\hat{h}^*(T, Q, \alpha\%)$ be the greatest number of guaranteed hours per session (rounded down to the nearest 30 minute increment to avoid unnatural closing times) that results in at most an $\alpha\%$ drop in performance compared to the unrestricted vaccine administration policy. The dashed circles in Figure 15 indicate $\hat{h}^*(T, Q, 1\%)$ for each scenario considered. For example, in Figure 15(a), $\hat{h}^*(16, 36, 1\%) = 360$, i.e., an earliest closing time of 2pm for a clinic that operates from 8am-4pm. Throughout the remainder of Chapter 2, we suppress dependence on $T$, $Q$ and $\alpha$ for notational convenience and let $\hat{h}^* \equiv \hat{h}^*(T, Q, 1\%)$.

In Figure 16, we plot $\hat{h}^*$ as a function of the number of sessions per replenishment cycle. The value of $\hat{h}^*$ generally decreases in $T$ because the arrival rate per timeslot decreases in $T$ and to prevent high OVW, it is optimal to discontinue vaccination earlier in the day. However, there are ranges of $T$ over which $\hat{h}^*$ is nonmonotone (e.g., $T$ = 5-10 for $z = 20$) because the ratio $\frac{Q}{T}$ becomes far an integer value and the optimal policy is such that it is likely to close early in the day. Therefore, smaller values of $\hat{h}$ satisfy the inequality in Equation (2.3). Finally, as vial size increases, $\hat{h}^*$ decreases because a larger vial size results in higher OVW and, consequently, requires a more flexible guaranteed hours policy.

### 2.3.3 Vaccination Session Frequency

Next, we investigate the degree to which the number of sessions per cycle can be increased, providing more days of clinic access to patients, without significantly impacting the coverage rate. Note that similar analysis can be performed to find the number of sessions per cycle that minimizes OVW; because our results indicate that these two values of $T$ are typically close if not the same, we present the results for coverage only.

To evaluate the effect of the number of sessions per cycle on the coverage rate, we define the criterion $\beta(T, Q, \hat{h})$ as the percentage loss in the expected number of vaccinations administered over $T$ sessions with $Q$ vials initially on hand and $\hat{h}$ guaranteed hours (in timeslots),

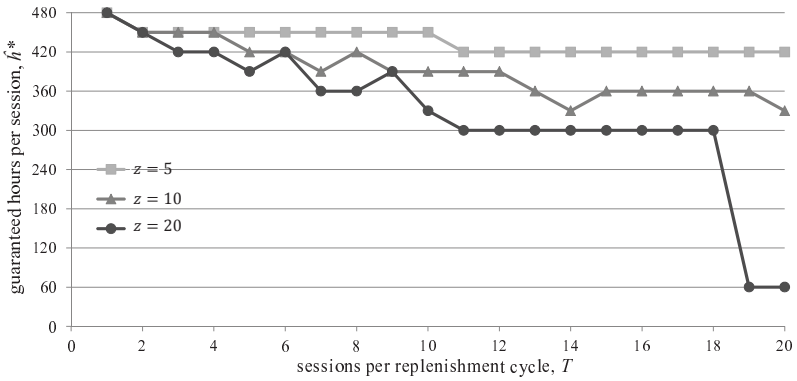



Figure 16: Value of $\hat{h}^*$ as a function of number of sessions per replenishment cycle ($zQ = 360$ and $\mu T = 288$).

as compared to that under the administration policy with the same initial inventory, but only one session per replenishment cycle (i.e., $T = 1$) and guaranteed hours of $\eta$, i.e.,

$$\beta(T, Q, \hat{h}) = \frac{\nu(1, Q, \eta) - \nu(T, Q, \hat{h})}{\nu(1, Q, \eta)} \times 100. \tag{2.4}$$

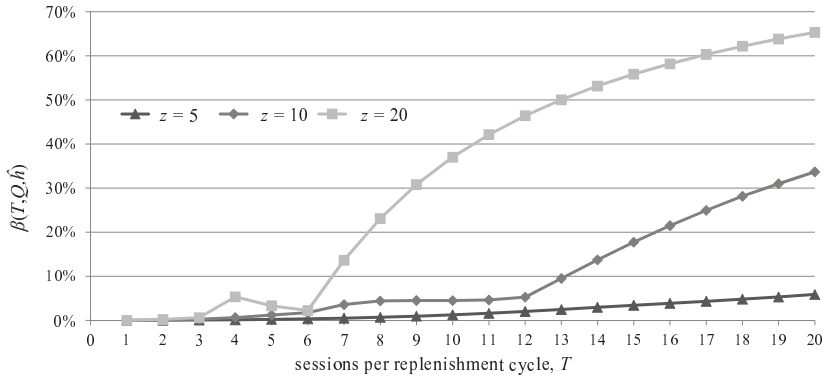
That is, $\beta(T, Q, \hat{h})$ compares the performance of the optimal restricted vaccine administration policy to the idealized policy where the clinic is open for exactly one full session between two stock replenishments. For a fixed demand per cycle, the latter policy has the highest coverage rate of all values of $T$ and the lowest OVW. Similar to Equation (2.3) for determining the number of guaranteed hours, this criterion can be used to determine the greatest number of sessions that results in an at most $\alpha\%$ loss in coverage as compared to the single-session case, denoted $\hat{T}(Q, \hat{h}, \alpha\%)$, using

$$\hat{T}(Q, \hat{h}, \alpha\%) = \max\left\{T : \beta(T, Q, \hat{h}) \leq \alpha\%, 0 \leq T \leq \bar{T}\right\}, \tag{2.5}$$

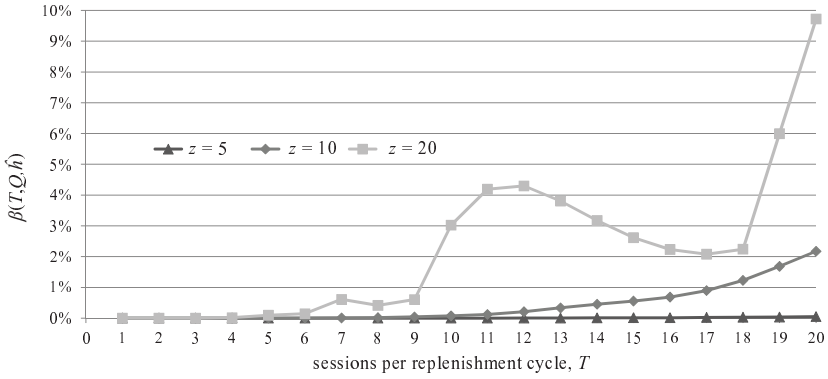
where $\bar{T}$ is the number of calendar days between two replenishments. For example, in Figure 17(b), for $z = 20$, $\hat{T}(18, 390, 1\%) = 9$ sessions per cycle.

Figure 17 plots the percentage loss as defined in Equation (2.4) as a function of the number of sessions based on the expected number of vaccinations as determined by the

MDP model. Note that the percentage loss is not necessarily monotone in the number of sessions. This complicated behavior is related to the ratio of the initial inventory over the number of sessions, as defined in Section 2.3.1. In Figure 17(b), for $z = 20$ ($Q = 18$), the percentage loss increases considerably for $9 < T \leq 12$ and then decreases up to $T = 18$, after which there is a sharp rise. In the range $9 \leq T \leq 12$ the difference between the ratio $\frac{Q}{T}$ and its nearest integer increases; in the range $12 \leq T \leq 18$, this difference decreases. Lastly, for $z = 20$ and more than 18 sessions, the performance of the vaccine administration policy drops quickly, because the number of initial vials is less than the number of sessions.




(a) $\mu T = 96$, $Qz = 120$, $\hat{h} = \hat{h}^*$.




(b) $\mu T = 288$, $Qz = 360$, $\hat{h} = \hat{h}^*$.

Figure 17: Percentage loss in the expected number of vaccinations administered $\beta(T, Q, \hat{h})$ as a function of number of sessions per cycle.

The better performing ratios of $\frac{Q}{T}$ correspond to problem instances that strike a balance between minimizing expected OVW and having additional vials to satisfy peak demands. For example, the coverage for $T = 17$ is higher than for $T = 16$ because there is less expected lost coverage for each period in which exactly one vial is opened. The benefit of $T = 16$ having two additional vials to meet peak demand days, versus the one additional vial of $T = 17$, is not enough to overcome the expected lost coverage on the other 14 days when only one vial can be used.

Next, we study the impact of the initial inventory or equivalently the buffer stock, on $\hat{T}(Q, \hat{h}^*, 1\%)$. To do so, for a fixed set of initial inventories, we first compute $\hat{h}^*$ for potential values of $T$ and then find the corresponding value of $\hat{T}(Q, \hat{h}^*, 1\%)$. In Figure 18, we plot $\hat{T}(Q, \hat{h}^*, 1\%)$ and the total number of guaranteed hours (in timeslots) as a function of the initial inventory. As $Q$ increases, $\hat{T}(Q, \hat{h}^*, 1\%)$ increases because more buffer stock is available to mitigate the increased OVW associated with a larger number of vaccination sessions. On the other hand, although increasing the number of sessions results in fewer guaranteed hours per session, the total number of guaranteed hours over the replenishment cycle generally increases in $Q$.

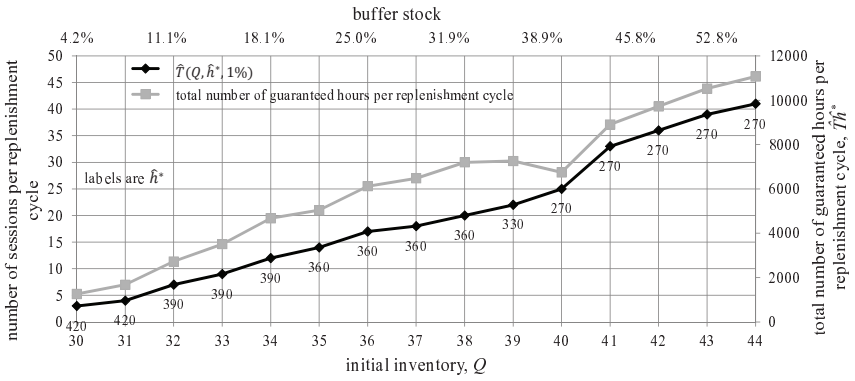



Figure 18: The value of $\hat{T}(Q, \hat{h}^*, 1\%)$ and its corresponding number of guaranteed hours per replenishment cycle as a function of initial inventory ($z = 10$, $\mu T = 288$, $\bar{T} = \infty$ and $\hat{h} = \hat{h}^*$).

Additionally, in Figure 19 we plot the "most convenient" pair of $(\hat{h}, \hat{T}(Q, \hat{h}, \alpha\%))$ values, i.e., that which results in the maximum total number of guaranteed hours per replenishment cycle, as a function of $Q$. In contrast to Figure 18 in which we fix $\hat{h}$ to $\hat{h}^*$, to generate Figure 19 we determine $\hat{T}(Q, \hat{h}, \alpha\%)$ for all values of $\hat{h}$ in the set $\{240, 300, 360, 420, 480\}$ using Equation (2.5), and then select the pair $(\hat{h}, \hat{T}(Q, \hat{h}, \alpha\%))$ that results in the maximum guaranteed hours per replenishment cycle. (Note that we restrict $\hat{T}(Q, \hat{h}, \alpha\%) \leq 20$.) The resulting pairs are the most favorable ones from the patient perspective (in terms of operation hours per cycle) which simultaneously result in at most $\alpha\%$ loss in coverage. In Figure 19, the value of $\hat{T}(Q, \hat{h}, \alpha\%)$ increases in $Q$ and $\alpha$. Increasing $Q$ or $\alpha$ provides more buffer stock and, consequently, a greater number of sessions becomes acceptable. Interestingly, as seen in Figure 19, 7 hours (i.e., 420 timeslots) is almost consistently the best choice for the number of guaranteed hours.

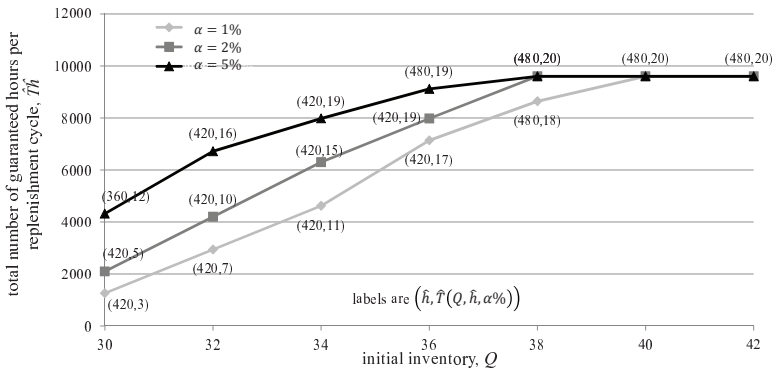


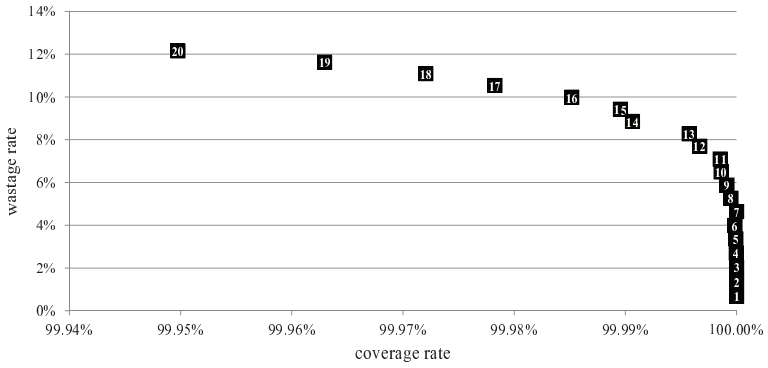

Figure 19: Maximum number of guaranteed hours per replenishment cycle and its corresponding pair of $(\hat{h}, \hat{T}(Q, \hat{h}, \alpha\%))$ as a function of $Q$ ($z = 10$, $\mu T = 288$, $1 \leq \hat{T} \leq 20$ and $\hat{h} \in \{240, 300, 360, 420, 480\}$).

Another metric by which we can evaluate the performance of a vaccine administration policy is the expected amount of OVW (in doses) per cycle, i.e., wastage rate which is introduced in Equation (1.22), Section 1.4. In other words, by wastage rate we mean the

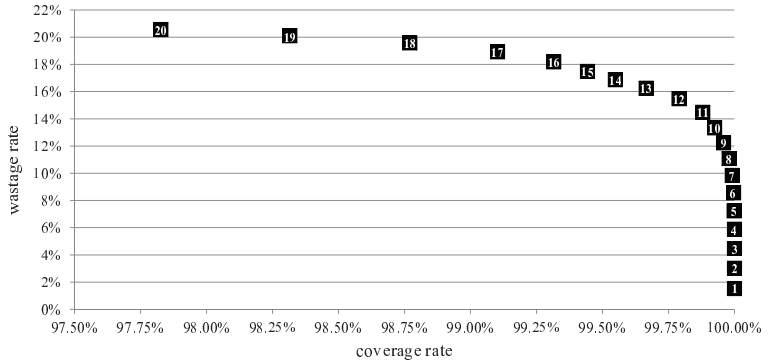
expected percentage of doses that are reconstituted and discarded. In Figure 20, we visually assess the relationship between wastage, coverage and the number of sessions. Specifically, Figure 20 shows the wastage rate versus the coverage rate as a function of the number of sessions for three vial sizes. That is, each point corresponds to a different value of $T$, as labeled. For smaller vial sizes, the relationship between wastage and coverage is monotone in $T$, i.e., the fewer the number of sessions, the lower the wastage rate and the higher the coverage rate. When the vial size is 20, however, the relationship is not monotone, i.e., fewer sessions does not necessarily result in lower wastage or higher coverage. For example, compared to $T = 8$, $T = 7$ reduces waste, but also reduces coverage. As marked by the light grey points, some values of $T$ may actually be dominated by others and should therefore not be considered for adoption. More specifically a value of 16 sessions performs so well that it dominates $T \in \{10, 11, 12, 13, 14, 15\}$. That is, compared to $T \in \{10, 11, 12, 13, 14, 15\}$, $T = 16$ offers more frequent sessions while producing less wastage and providing higher coverage.

The unexpected behavior in Figure 20(c) (highlighted in gray) can be explained through the average number of vials available per session, $\frac{Q}{T}$, introduced in Section 2.3.1. The fractional part of $\frac{Q}{T}$ represents vials which theoretically should be shared between two or more sessions although it is impossible to do in practice. As a result, if the fractional part is farther from 0 or 1 then a larger number of open doses may be discarded. In Figure 20(c) although we decrease the number of sessions from 16 to 11, lower coverage and higher OVW are seen because $\frac{Q}{T}$ moves farther from an integer value as $T$ decreases and given that the vial size is large (20 vs. 5 and 10), the number of unused doses which are discarded at the end of a session is more substantial.

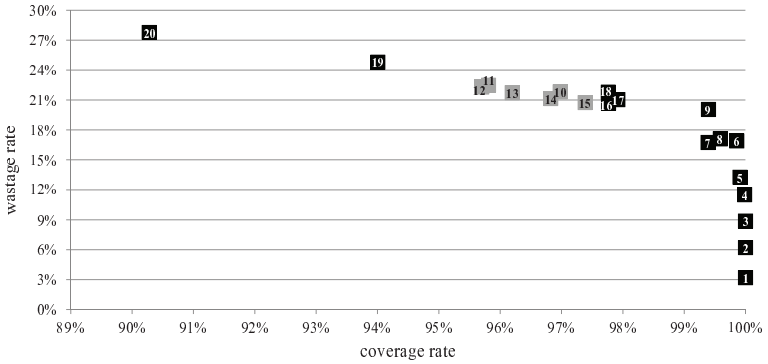
Lastly, we consider the variability of the number of vaccinations under the optimal restricted policy. Variability analysis provides insight into the unpredictability of the number of vaccinations and facilitates worst case analysis. Based on 10,000 replications of the simulation model, we compute a 99% prediction interval on the number of vaccinations administered. Figure 21 shows the prediction interval for the number of vaccinations under the optimal restricted vaccine administration policy. The variance associated with the number of vaccinations per cycle is rather large due to the high variability in the arrival process,




(a) $\mu T = 288$, $Q = 72$, $z = 5$, $\hat{h} = \hat{h}^*$.




(b) $\mu T = 288$, $Q = 36$, $z = 10$, $\hat{h} = \hat{h}^*$.

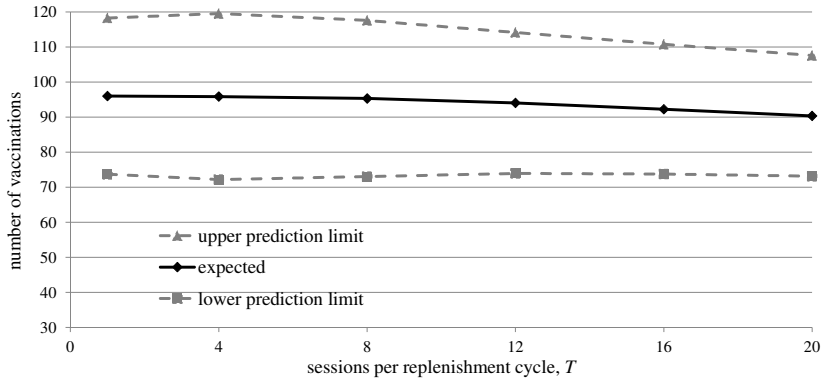



(c) $\mu T = 288$, $Q = 18$, $z = 20$, $\hat{h} = \hat{h}^*$.

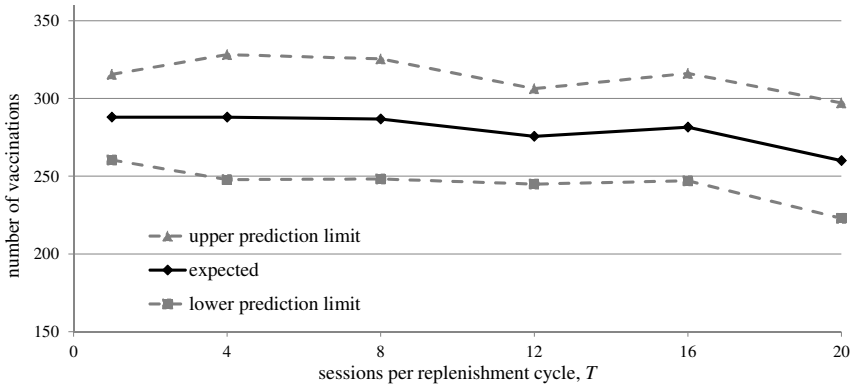
Figure 20: Open vial wastage rate vs. coverage rate as a function of $T$.

e.g., in Figure 21(b), for $T = 1$, the standard deviation of the number of arrivals is 10.7 patients. Furthermore, the variability does not appear to be very sensitive to the number of

sessions, especially for larger vial sizes. Depending on the number of sessions, the widths of the 99% prediction intervals range from (approximately) 50 to 75 vaccinations when using larger vials ($z = 20$), to 35 to 50 vaccinations when using smaller vials ($z = 5$).




(a) $z = 5$, $Q = 24$, $\mu T = 96$, $\hat{h} = \hat{h}^*$.




(b) $z = 20$, $Q = 18$, $\mu T = 288$, $\hat{h} = \hat{h}^*$.

Figure 21: Prediction interval for the number of vaccinations as a function of the number of sessions per cycle.

## 2.4 HEURISTIC POLICIES

In this section, we consider two heuristic vial administration policies. The first is the dynamic heuristic policy proposed in Section 1.4.4 under which a new vial is opened after guaranteed hours as long as the number of vials remaining is at least as large as the number of vials required to satisfy the expected demand in the remaining sessions. The second is a static heuristic policy that specifies a fixed cutoff time after which clinicians do not open new vials. More specifically, this static heuristic approach evaluates all possible cutoff times using the policy evaluation algorithm in [38] and selects the one that results in the maximum expected number of vaccinations per replenishment cycle. The resulting static policy is easier to implement than the dynamic heuristic policy proposed in Section 1.4.4 because the same cutoff time is used in each session as opposed to the optimal policy where cutoff points would be different. However, determining the cutoff point of the static heuristic policy requires considerably more computational effort.

The results suggest that the static heuristic performs very close to the optimal policy and in some cases outperforms the dynamic heuristic. To evaluate the quality of the heuristics, in Table 4 we report the percentage of the gap between the optimal and greedy policies that is covered by the heuristic policies for several problem instances, i.e.,

$$\text{percentage of gap covered} = \frac{v_h - v_g}{v^* - v_g}, \tag{2.6}$$

where $v^*$ is the expected number of vaccinations under the optimal policy (with $\hat{h} = \hat{h}^*$), $v_g$ is the expected number of vaccinations under the greedy policy and $v_h$ is the expected number of vaccinations under the heuristic policy. As seen in Table 4, a large portion of the gap is often covered by these policies. For example, for the case in which $T = 12$ and $\mu = 8$ the percentage of the gap covered by the static and dynamic heuristics is approximately 71% and 66%, respectively. In Section 2.5, we study the behavior of these policies more deeply in the presence of random vial yield.

Table 4: The percentage of the gap between the optimal and greedy policies covered by the static and dynamic heuristics ($z = 10$, $Q = 12$, $\mu T = 96$, $\hat{h} = \hat{h}^*$).

| $T$ | $\hat{h}^*$ | cutoff point (static heuristic) | percentage of gap covered | |
|---|---|---|---|---|
| | | | static heuristic | dynamic heuristic |
| 1 | 480 | 480 | 0.0% | 0.0% |
| 4 | 390 | 480 | 0.0% | 72.4% |
| 8 | 240 | 420 | 71.2% | 65.9% |
| 12 | 240 | 300 | 94.5% | 56.2% |
| 16 | 90 | 180 | 76.9% | 62.8% |
| 20 | 75 | 120 | 84.5% | 64.5% |

## 2.5 IMPACT OF VIAL-YIELD ON POLICY PERFORMANCE

The MDP model proposed in Section 1.2 generates an optimal policy under the simplifying assumption of perfect vial yield, i.e., a 5-dose vial yields exactly 5 doses. However, in practice, the number of doses per vial is a random variable due to clinician variation, syringe type and spoilage (i.e., a yield of zero) prior to reconstitution [14]. In this section, we study the impact of random vial yield on the performance of the optimal policy as well as that of the two proposed heuristic policies defined in Section 2.4. The results in this section are obtained using the simulation model detailed in Section 2.2.

Let the random variable $Z$ be the number of doses yielded by a multi-dose vial. To analyze the impact of vial yield, we construct a set of stochastically ordered distributions shown in Figure 22. To generate these probability distributions, we consider a fixed set of expected vial yields, $\mathrm{E}[Z] \in \{7, 8, 9, 10, 11\}$, and a base probability distribution function from [44]. To generate the probability distribution associated with each $\mathrm{E}[Z]$, we raise the base cumulative distribution function to the power of the positive number such that it equates to the expected value under the new cumulative distribution and $\mathrm{E}[Z]$.

In Figure 23, we present the coverage and wastage rates under the optimal policy, the two heuristic policies and the greedy policy for three different vial sizes ($z = 10, 9$, and 8)

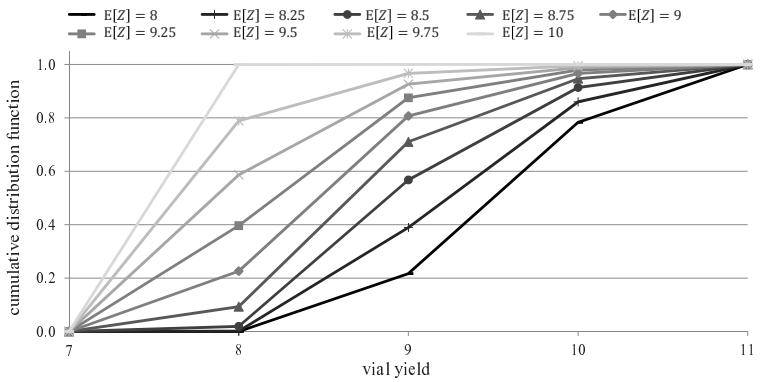
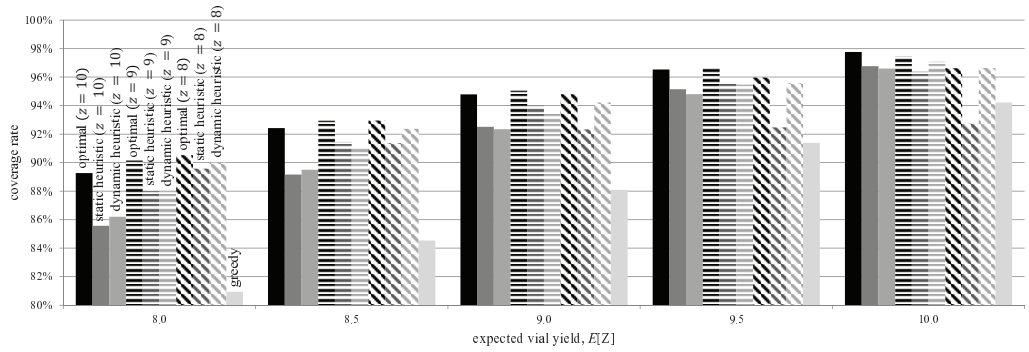



Figure 22: Cumulative distribution functions of vial yield.

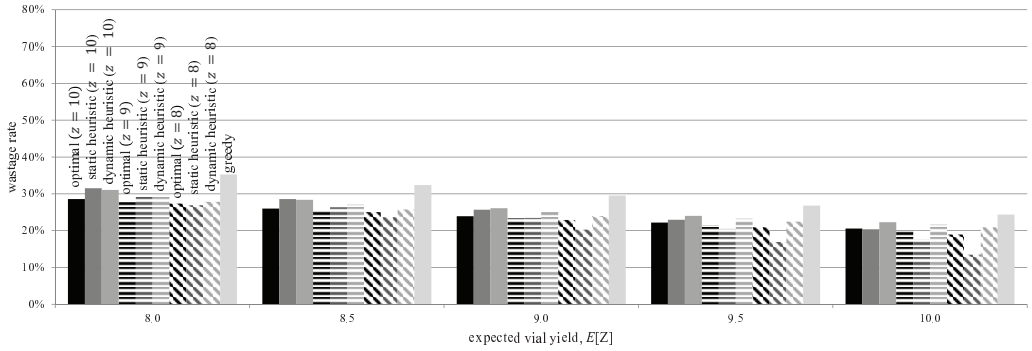
as a function of the vial yield distributions shown in Figure 22. We set $z = 10, 9$ and 8 for the optimal and two heuristic policies because the expected vial yield usually falls between 8 and 10. First, as seen in Figure 23(a), the optimal and two heuristic policies derived for a vial size closer to the expected vial size result in a larger coverage rate. For example in Figure 23(b), the coverage rate of the optimal policy with vial size 10 increases from 89.3% to 97.8% as E[$Z$] increases from 8 to 10. However, regardless of the expected vial yield, as the vial size decreases the optimal and heuristic policies become more conservative which results in lower wastage rates.

Second, the performance of the static and dynamic heuristic policies falls between that of the optimal and greedy policies for $z = 10$ and lower than the optimal policy for $z = 9$ and 8. The greedy policy does not account for vial yield so its performance drops as the expected vial yield decreases. The drop in the performance happens because fewer doses are available and the greedy policy does not adapt to this effective change in the vial size as well as the optimal policy does for a smaller expected vial size than the nominal vial size.

Next, let $p$ be the probability of vial failure due to spoilage. Figure 24 shows the coverage and wastage rate of the optimal policy, the heuristic policies and the greedy policy as a function of the probability of vial failure, $p$. First, as seen in Figure 24, as the probability of vial failure increases, the coverage (wastage) rate of all policies decreases (increases), because a smaller number of vials is available. Second, the wastage rate of the greedy and




(a) Coverage rate.

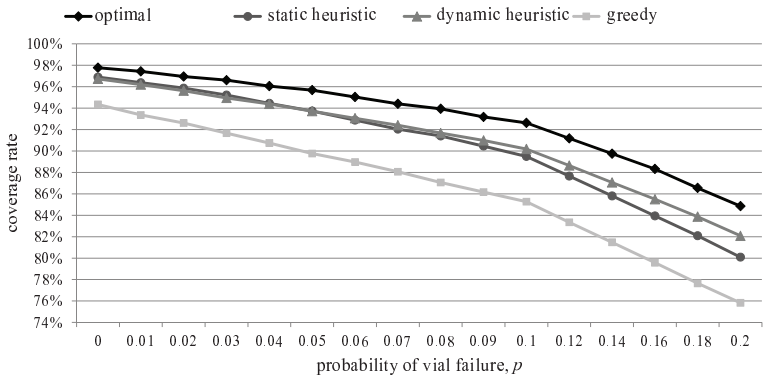



(b) Wastage rate.

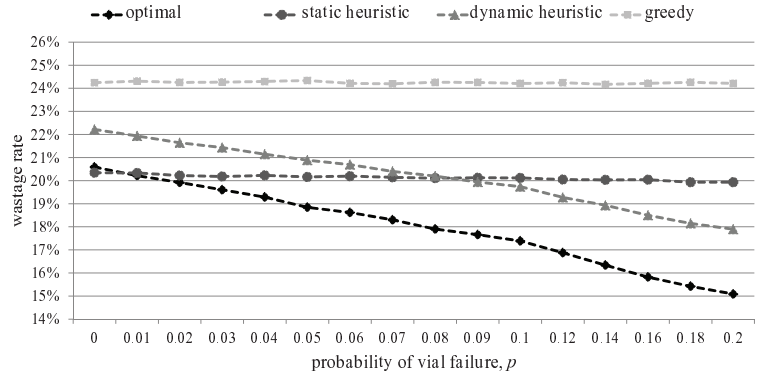
Figure 23: Coverage and wastage rate as a function of vial yield distribution in Figure 22 ($\mu T = 288$, $T = 20$, $Q = 36$, $z = 10$, $\hat{h} = \hat{h}^*$).

static heuristic policy are nearly constant as the probability of vial failure increases because these policies have a fixed cutoff time after which clinicians do not open new vials and are therefore independent of the number of vials remaining and subsequently the probability of vial failure. While the optimal and dynamic heuristic policies can better adapt to the loss of vials due to vial failure because in these policies, clinicians make the decision based on the remaining number of vials. Third, the gap between the coverage and wastage rate of the optimal and greedy policies increases in the probability of vial failure. Having a smaller number of vials available results in a more conservative optimal policy, but has no impact

on the greedy policy. Therefore, the greedy policy performance deteriorates more rapidly as the probability of failure increases.




(a) Coverage rate.




(b) Wastage rate.

Figure 24: Coverage and wastage rate as a function of the probability of vial failure ($\mu T = 288$, $T = 20$, $Q = 36$, $z = 10$, $\hat{h} = \hat{h}^*$).

## 2.6 DISCUSSION AND COST IMPLICATIONS

In this section, we begin (Section 2.6.1) with some general recommendations derived from our analysis of the optimal policy in Sections 2.3 and 2.5. In Section 2.6.2, we estimate the potential reduction in vaccine procurement costs under an optimal policy in three GAVI countries (Mozambique, Benin and Kenya) and then extrapolate these results to the set of all GAVI countries.

### 2.6.1 Operational Insights

Clearly, offering more frequent vaccination sessions and guaranteeing longer hours during these sessions is desirable from a patient perspective. However, doing so can also result in more vaccine wastage and lower overall coverage. The analysis in Section 2.3 shows that these two factors interact with each other, and that the nature of this interaction depends on the vial size and demand rate. Focusing on service to a fixed catchment population, a smaller number of sessions per cycle generally decreases the likelihood of discontinuing service before the end of each session and increases the expected number of vaccination hours per session (although not always, as illustrated by Figures 12, 14(a) and 14(b)).

In Section 2.3.1, we show that a clinic is more likely to have full length sessions when the average number of vials allocated per session ($\frac{Q}{T}$) slightly exceeds an integer value. If this fraction is less than one, then the clinic will be forced to skip at least one entire session; therefore, clinic administrators should consider the initial number of vials that they receive during each cycle ($Q$) to be an upper bound on the number of sessions per cycle. Furthermore, choosing $T$ such that $\frac{Q}{T}$ is slightly larger than a small integer helps maintain longer sessions, even when they are terminated early.

In terms of planned session hours, adopting a large number of guaranteed hours, although desirable from a service perspective, constrains the clinic and can result in poor coverage. The analysis in Section 2.3.2 shows that there is a relatively clear threshold for guaranteed vaccination hours based on the expected gain in coverage over the greedy policy, especially for larger vial sizes. However, as seen in Figures 15(a) and 15(b), the threshold value can

vary widely, ranging from $300-400$ minutes for typical parameter values during a 480-minute session. Forcing the clinic to guarantee hours beyond this threshold can lead to degradation in coverage, however, guaranteeing fewer hours does not significantly improve coverage.

As discussed in Section 2.3.3, although having a larger number of sessions can increase patient convenience, it can also decrease the coverage rate. The "best" scenario in terms of maximizing coverage would be a single session per cycle that all patients in the catchment area attend. Of course, this scenario is not practical; however, it does provide an upper bound on the expected coverage. Increasing the number of sessions per cycle can therefore be evaluated with respect to this bound in terms of the percentage loss in the expected number of vaccinations, and clinicians can increase the number of sessions as long as this loss is not unacceptably large. It should be mentioned that trying to predict the exact number of vaccinations that a clinic can provide for a given vial size and a given number of sessions per replenishment cycle is difficult. There can be considerable variability, especially when using larger vial sizes. Thus, clinicians should not be alarmed if they see some variation with respect to the number of vaccinations provided across different replenishment cycles.

Furthermore, in Section 2.3.3, we study the impact of guaranteed hours per session and number of sessions per cycle, simultaneously, on the policy performance. We choose patient convenience to determine the best pair while an acceptable deviation from the maximum coverage is achieved. Interestingly, as seen in Figure 19, in majority of cases it appears that the most desirable pair corresponds to the cases with lower number of sessions but with larger guaranteed hours per session.

Finally, in Section 2.5, we show that in the presence of random vial yield, implementing the optimal policy for a vial size close to the expected vial yield results in better coverage than merely using the nominal vial size value (Figure 23).

Though we have been able to provide *some* general guidelines for vaccine administration, our computational results also indicate that the operating strategies that maximize vaccine coverage often depend heavily on the characteristics of the clinic. Therefore, we have developed a decision support tool to permit users to conduct analysis similar to that performed in Sections 2.3 and 2.5 for their particular setting. The tool is called the Vaccine Clinic Reconstitution Optimizer (VaCRO). VaCRO develops a vaccine adminstration policy based on the

MDP model in Section 1.2. We have made VaCRO freely available online for download [51] and promoted it via The Technical Network for Strengthening Immunization Services [47], an on-line forum for global health logisticians and others interested in improving immunization programs and their effectiveness.

### 2.6.2 Cost Analysis

In this section, we estimate the reduction in procurement costs associated with switching from the greedy policy to an optimal vial allocation policy. To do so, we use available data for the annual birth rate of clinics in three GAVI countries - Mozambique, Benin and Kenya. Because of the large variation in the clinic birth rates across these countries, we divide the clinics into three sizes - small, medium, and large. Clinics with less than 365 births per year (i.e., one birth per day) are classified as small clinics, clinics with at least 365 births per year and at most 1095 births per year (i.e., 3 births per day) are classified as medium clinics and the remaining clinics are classified as large clinics. We assume a constant birth rate over the year; monthly replenishment in all clinics; and as an illustrative vaccine, we choose measles, which requires 2 doses per child. We then estimate the monthly demand using the average monthly birth rates of the clinics in the middle 80% of each size group. In Table 5, we report the estimated monthly demand for each clinic size.

Table 5: Average number of arrivals per replenishment cycle in Mozambique, Benin and Kenya for a two-dose measles vaccine schedule.

| country | total number of clinics | average number of arrivals per month | | |
|---|---|---|---|---|
| | | small clinics | medium clinics | large clinics |
| Mozambique | 114 | 31.4 | 95.5 | 284.8 |
| Benin | 658 | 43.1 | 97.9 | 260.7 |
| Kenya | 2733 | 34.3 | 94.4 | 348.9 |

To compare the performance of the greedy and optimal policy, we set the minimum coverage rate to be 95% and compute the required initial inventory under each policy using the monthly demand data presented in Table 5 and parameter values stated in Table 6. We then compute the reduction in the procurement cost per vaccination realized by switching to the optimal policy. Finally, using a weighted average based on the proportion of patients who fall into each clinic size in each country, we compute the average reduction in the procurement cost per vaccination for each country. The resulting reduction in procurement costs per vaccination in Mozambique, Benin and Kenya are approximately \$0.051/administered dose, \$0.033/administered dose and \$0.031/administered dose, respectively, which are equivalent to 13.8%, 9.1% and 8.7% of the procurement cost per vaccination under the greedy policy.

Table 6: Parameter values used in the cost analysis of the optimal and greedy policy ($\hat{h} = \hat{h}^*$).

| vaccination policy | country | buffer stock | small-size clinics | | | | | medium-size clinics | | | | | large-size clinics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mu$ | $Q$ | $z$ | $T$ | $\hat{h}$ | $\mu$ | $Q$ | $z$ | $T$ | $\hat{h}$ | $\mu$ | $Q$ | $z$ | $T$ | $\hat{h}$ |
| optimal policy | Mozambique | 25% | 7.85 | 4 | 10 | 4 | 240 | 7.96 | 12 | 10 | 12 | 240 | 14.24 | 33 | 10 | 20 | 345 |
| | Benin | | 10.78 | 6 | | | | 8.16 | 13 | | 12 | 255 | 13.03 | 31 | | | 330 |
| | Kenya | | 8.58 | 5 | | | | 7.87 | 12 | | 12 | 255 | 17.44 | 39 | | | 375 |
| greedy policy | Mozambique | 25% | 7.85 | 5 | 10 | 4 | 480 | 7.96 | 14 | 10 | 12 | 480 | 14.24 | 37 | 10 | 20 | 480 |
| | Benin | | 10.78 | 7 | | | | 8.16 | 14 | | | | 13.03 | 34 | | | |
| | Kenya | | 8.58 | 5 | | | | 7.87 | 14 | | | | 17.44 | 42 | | | |

Next, we extrapolate these results to all GAVI countries. To do so, we first estimate the average reduction in the procurement cost per vaccination when the optimal policy is adopted in all GAVI countries using a weighted average of the reductions in Mozambique, Benin and Kenya. More specifically, we use the newborn rate of each country as its corresponding weight and estimate ($\frac{51826\times\$0.051+371022\times\$0.033+1516221\times\$0.031}{51826+371022+1516221} \approx$) \$0.032/administered dose reduction in procurement costs on average in all GAVI countries. We then estimate a ($72,952,312 \times \$0.032 \times 2 \approx$) \$4.6 million reduction in procurement costs per year across all the GAVI countries given that 2 doses are required per child and the total birth cohort of all GAVI

countries is approximately 73 million [12]. It is important to note that this cost reduction is computed for measles which is a relatively inexpensive vaccine; other types of vaccines are often 10 times (or more) as expensive. Therefore, this $4.6 million savings is a conservative estimate of the reduction in procurement costs per vaccine type.

## 2.7 LIMITATIONS

In this section, we address some of the limitations of our analysis and present directions for future work. From a modeling perspective, the MDP model in Section 1.2 assumes that arrival rates during and across sessions are stationary and that patients who are not vaccinated are lost. Modeling these types of nonstationarities, patient returns and other adaptive behaviors of patients, adds considerable complexity to the MDP model and is explored in Chapter 3.

Due to data scarcity, we evaluate proposed recommendations and criteria using representative problem instances. Although the parameter values selected are representative of values observed in practice, performing empirical studies regarding patient arrival patterns would allow us to gain better understanding on how to inform the models and generate recommendations. Additionally, we focus on in-clinic vaccine administration in isolation. In reality, in-clinic operations and outreach strategies are often linked together by a shared supply of vaccines and healthcare workers.

# 3.0 MULTI-DOSE VIAL ADMINISTRATION WITH NON-STATIONARY DEMAND AND DELAYED SERVICE

## 3.1 INTRODUCTION

In this chapter, we formulate a new model by relaxing the limiting demand-related assumptions imposed in Section 1.2. More specifically, we (*i*) consider non-stationarities in the within-day arrival rate, (*ii*) consider non-stationarities in the mean daily demand throughout the replenishment cycle and (*iii*) allow unvaccinated patients to possibly return. These are important, practical generalizations because (*i*) establishing guaranteed hours in which all patients are vaccinated unconditionally, incentivizes patients to arrive earlier in the day; (*ii*) implementing a non-greedy vaccine administration policy, which uses vials more conservatively near the end of each replenishment cycle, incentivizes patients to arrive earlier in the month to increase their chances of being vaccinated; and (*iii*) discontinuing service before the end of a session may result in increased demand in the following session because some portion of the patients denied service are likely to return, especially if guaranteed vaccination upon return.

The non-stationary demand considerations in generalizations (*i*) and (*ii*) appear often in the inventory and queuing literature (e.g., [13, 25, 26, 46]). However, the focus of this section is vaccine administration policies, not ordering policies; moreover, queue formation need not be considered here because service times are negligible in comparison to the arrival rates. Furthermore, although multi-dose vials are a perishable product after reconstitution, our focus is on how to optimally consume this product rather than on replenishment decisions, which is the focus of the majority of the perishable inventory control literature [20]. The multi-period newsvendor literature (e.g., [2, 28]) deals with seasonal or perishable inventory

over multiple periods, however, here, unopened vials can be carried over from day to day and again, the focus is not on replenishment decision making.

Generalization (*iii*), i.e., explicitly modeling second attempts at service, is similar in spirit to the concept of an "orbit" in the retrial queueing literature [59]. In retrial queueing systems, customers who arrive to the system and observe that the server is busy, enter the orbit and reattempt service at a later time. Similarly, we assume that patients who arrive after vaccinations have been suspended for the day are invited to return during guaranteed hours in the following session, but may or may not actually return.

The remainder of this chapter is organized as follows. In Section 3.2, we present a new MDP formulation of the problem and explain the methods used to incorporate the non-stationarities in the mean daily demand and the arrival rate. In Section 3.3, we evaluate the impact of the generalizations on the performance of the optimal policy. In Section 3.4, using field data, we estimate the potential improvements realized by implementing the optimal policy in the sate of Bihar, India (and beyond) over a one-year period. Finally, in Section 3.5, we summarize and discuss general recommendations to improve vaccine administration policy performance.

## 3.2 MODEL FORMULATION

In this section, we first (Section 3.2.1) formulate an MDP model to maximize the expected number of vaccinations during a replenishment cycle given a limited supply of multi-dose vials. We assume that incoming patients who arrive before some prespecified point in time during the day, i.e., during "guaranteed hours," are always vaccinated by clinicians as long as vaccines are available. Patients arriving after guaranteed hours are also vaccinated as long as a vial is already open. Otherwise, the MDP model determines whether to open a new vial or discontinue service for the remainder of the session as a function of time of day, the current vial inventory and the remaining number of sessions until the next replenishment. That is, clinicians discontinue service for the rest of day when the last dose in a vial is administered and the optimal policy indicates that it is optimal to suspend service until the

beginning of the next session. Patients who do not receive service are assumed to return with some probability during guaranteed hours in the following session. Similar to Section 1.2, we assume that $(i)$ the maximum number of clinic hours per day is equal to the shelf life of an open vial (i.e., doses cannot expire midday) and $(ii)$ the replenishment schedule is fixed.

In Section 3.2.1, we formally describe the model and identify several special cases of interest. In Section 3.2.2, we explain the methods used to model non-stationarities in both the mean daily demand across the replenishment cycle and the within-day arrival rate.

### 3.2.1 MDP Model

Let $T$ be the number of sessions per replenishment cycle. We divide each session into $\eta$ timelots and let $\hat{h}$ be the timeslot after which guaranteed hours end in each session. Let the symbols $\hat{}$ and $\check{}$ correspond to the parameters associated with timeslots during and after guaranteed hours, respectively. We assume a maximum of one arrival per timeslot and that the arrival rate during each of the intervals $[0, \hat{h}]$ and $[\hat{h}+1, \eta]$ is constant. Let $\hat{p}_t$ and $\check{p}_t$, respectively, be the probability of arrival per timeslot before and after $\hat{h}$ where $t$ is the number of sessions (including the current session) remaining until the next replenishment. Consequently, the patient interarrival times (measured in timeslots) for those who arrive (for the first time) before and after $\hat{h}$ are iid geometric random variables with parameter $\hat{p}_t$ and $\check{p}_t$, respectively. Lastly, let $\rho$ be the probability that each patient who is not vaccinated on his\her first attempt accepts the invitation and returns during guaranteed hours in the following session.

As seen in Figure 25, we consider two types of non-stationarities, resulting in four versions of the MDP model. The simplest case of the MDP model, the SS-MDP model, does not consider any non-stationarities, in which case $\hat{p}_t$ and $\check{p}_t$ are equal and constant in $t$; the MDP model developed in Section 1.2 is a special case of the SS-MDP model in which it is also assumed that the probability of return for unvaccinated patients is zero, i.e., $\rho = 0$. If the arrival rate varies across sessions, but not within each session, then $\hat{p}_t$ and $\check{p}_t$ are equal, but nonconstant in $t$; this model is referred to as the SN-MDP model. If the arrival rate varies

only within each session, then $\hat{p}_t$ and $\check{p}_t$ are not equal, but constant in $t$; this model is referred to as the NS-MDP model. In the remainder of this section, we focus on most general case, namely the NN-MDP model, in which both the within-day arrival rate and daily demand are non-stationary and unvaccinated patients return with some positive probability.

| $0 \le \rho \le 1$ | | daily demand | |
|---|---|---|---|
| | | stationary | non-stationary |
| within-day arrival rate | stationary | SS-MDP model<br>$\hat{p}_t = \check{p}_t = p$ | SN-MDP model<br>$\hat{p}_t = \check{p}_t = p_t$ |
| | non-stationary | NS-MDP model<br>$\hat{p}_t = \hat{p}$<br>$\check{p}_t = \check{p}$ | NN-MDP model<br>$\hat{p}_t > \check{p}_t > 0$ |

Figure 25: Special cases of the MDP model.

For model simplicity, similar to model developed in Section 1.2, we define the decision epochs, i.e., the points in time at which decisions are made to either open a vial or suspend service for the day, to be the timeslots greater than $\hat{h}$ in which an arrival occurs and no vial is open. In implementing the resulting policy, however, the clinician would not wait for an arrival to occur before discontinuing service for the day; service would be discontinued when no vial is currently open and the policy indicates that it is not optimal to open another vial in the current session regardless of the time of the next arrival.

Let $Q$ be the initial inventory and $z$ be the vial size, i.e., number of doses per vial. The state of the process is $(t, q, h)$ where $t \in \{0, 1, \ldots, T\}$ is the number of sessions remaining (including the current session) until the next replenishment, $q \in \{0, 1, \ldots, Q\}$ is the number of vials on hand and $h \in \{\hat{h}+1, \hat{h}+2 \ldots, \eta+1\}$ is the current timeslot. Let $V(t, q, h)$ be the maximum expected number of vaccinations prior to the next replenishment starting from state $(t, q, h)$ given that an arrival occurs at timeslot $h$. Hence, for $t > 0$, $q > 0$, $\eta \ge h > \hat{h}$,

$$V(t,q,h) = \max\Big\{ g_1(t,q,h), g_2(t,q,h) \Big\} \tag{3.1}$$

determines whether to open a new vial or discontinue vaccinations for the rest of day where $g_1(t,q,h)$ corresponds to the maximum expected number of vaccinations to go until the next replenishment if the clinician opens a new vial and $g_2(t,q,h)$ corresponds to the same quantity if vaccinations are suspended for the remainder of the current session.

Next, let $\check{Y}_{k,t}$ be the number of timeslots until the $k^{\text{th}}$ arrival after $\hat{h}$, given that at least $k$ arrivals occur after $\hat{h}$, when there are $t$ sessions remaining, i.e.,

$$p_{\check{Y}_{k,t}}(y) = \begin{cases} 0, & y < k, \\ \check{p}_t B_{y-1,\check{p}_t}(k-1), & y = k, k+1, \ldots, \eta, \end{cases} \tag{3.2}$$

where $B_{n,p}(k)$ corresponds to a binomial pmf with parameters $n$ and $p$. Hence,

$$g_1(t,q,h) = \sum_{j=z}^{\eta-h} \Big( z + V(t,q-1,j+h) \Big) p_{\check{Y}_{z,t}}(j) + \sum_{i=1}^{z} \Big( i + \nu_0(t-1,q-1) \Big) B_{\eta-h,\check{p}_t}(i-1), \tag{3.3}$$

and

$$g_2(t,q,h) = \sum_{\bar{y}=1}^{\eta-h+1} \nu_{\bar{y}}(t-1,q) B_{\eta-h,\check{p}_t}(\bar{y}-1), \tag{3.4}$$

where $\nu_{\bar{y}}(t,q)$ is the expected number of vaccinations to go when there are $t$ full sessions remaining, $q$ vials on hand and $\bar{y}$ patients that went unvaccinated in the previous session due to service suspension, i.e.,

$$\begin{aligned} \nu_{\bar{y}}(t,q) \;=\; & \sum_{\hat{y}_1=0}^{\hat{h}} \Bigg( \sum_{\hat{y}_2=0}^{\bar{y}} \Bigg( \hat{y} + \sum_{i=0}^{\hat{d}} \Big( i + \nu_0(t-1,\hat{q}) \Big) B_{\eta-\hat{h},\check{p}_t}(i) \\ & + \sum_{h=\hat{d}+1}^{\eta-\hat{h}} \Big( \hat{d} + V(t,\hat{q},\hat{h}+h) \Big) p_{\check{Y}_{\hat{d}+1,t}}(h) \Bigg) B_{\bar{y},\rho}(\hat{y}_2) \Bigg) B_{\hat{h},\hat{p}_t}(\hat{y}_1), \end{aligned} \tag{3.5}$$

where

$$\hat{y} = \min(\hat{y}_1 + \hat{y}_2, zq), \tag{3.6}$$

$$\hat{q} \equiv q - \left\lceil \frac{\hat{y}}{z} \right\rceil, \tag{3.7}$$

$$\hat{d} \equiv z \left\lceil \frac{\hat{y}}{z} \right\rceil - \hat{y}. \tag{3.8}$$

In Equation (3.3) the first term on the right hand side corresponds to the case in which the rest of the current session's demand exceeds $z$ and the second term corresponds to the case in which there are fewer than $z$ arrivals over the remainder of the session. Note that we assume that $\bar{y}$ is fully observable in each session; because administering vaccinations is just one of many services provided by a clinic, clinics are typically open until the end of their working hours and able to observe patient arrivals after vaccinations have been suspended and invite these patients to return during guaranteed hours in the following session. In Equations (3.5) and (3.6), $\hat{y}_1$ and $\hat{y}_2$ are the number of arrivals during guaranteed hours from the general population (i.e., for the first time) and from the previously unvaccinated patients (i.e., for the second time), respectively; $\hat{y}$ is the number of patients vaccinated during guaranteed hours, which cannot exceed the number of doses available at the beginning of session. In Equation (3.5), $\hat{q}$ is the remaining number of unopened vials at the end of guaranteed hours and and $\hat{d}$ is the remaining number of unused doses in the most recently opened vial at the end of guaranteed hours. On the right hand side of Equation (3.5), the inner two summations compute the expected number of additional vaccinations after guaranteed hours through the end of the replenishment cycle. More specifically, the first summation corresponds to the case in which the rest of the current session's demand does not exceed $\hat{d}$, whereas the second summation is associated with the opposite case.

Obviously, if no arrival occurs after guaranteed hours when there are $t-1$ sessions to go, then $V(t,q,\eta+1)=\nu_0(t-1,q)$. Similarly, if the vial inventory is exhausted, then there is no more reward to be earned, i.e., $V(\cdot,0,\cdot)=0$. Lastly, because the problem horizon is defined as a single replenishment interval, $\nu(0,\cdot,\cdot)=0$.

### 3.2.2 Modeling Non-Stationarity

We consider two types of non-stationary demand behavior: variation in the mean daily demand over the course of the replenishment cycle and variation in the arrival rate within each session. First, consider the variation in the mean daily demand. This variability is a consequence of having a limited number of vials available during the replenishment cycle, which encourages patients to arrive in earlier sessions to increase their chance of

being vaccinated; consequently, the mean daily demand decreases as the end of the cycle approaches. We model this behavior by fixing the percent reduction in the mean daily demand from session to session. More specifically, let $\mu_0$ be the base mean daily demand value, i.e., the mean daily demand if the total cycle demand is evenly distributed across the sessions, $\mu_t$ be the mean daily demand when there are $t$ sessions to go and $\gamma$ be the coefficient of proportionality between two consecutive sessions. Hence, fixing the total expected cycle demand at $\mu_0 T$, i.e., assuming $\sum_{t=1}^{T} \mu_t = \mu_0 T$,

$$\mu_1 = \begin{cases} \frac{\mu_0 T(1-\gamma)}{1-\gamma^T}, & 0 \leq \gamma < 1, \\ \mu_0, & \gamma = 1, \end{cases} \tag{3.9}$$

$$\mu_t = \gamma \cdot \mu_{t-1}, \quad t = 2, 3, \ldots, T. \tag{3.10}$$

For example, Figure 26 plots the mean daily demand over the replenishment cycle for different values of $\gamma$ when $\mu_0 T = 220$ patients are expected to arrive during the replenishment cycle. As seen in Figure 26, even with a small reduction in the value of $\gamma$, we observe a substantial decrease in the mean daily demand from the beginning of the cycle to the end of the cycle.

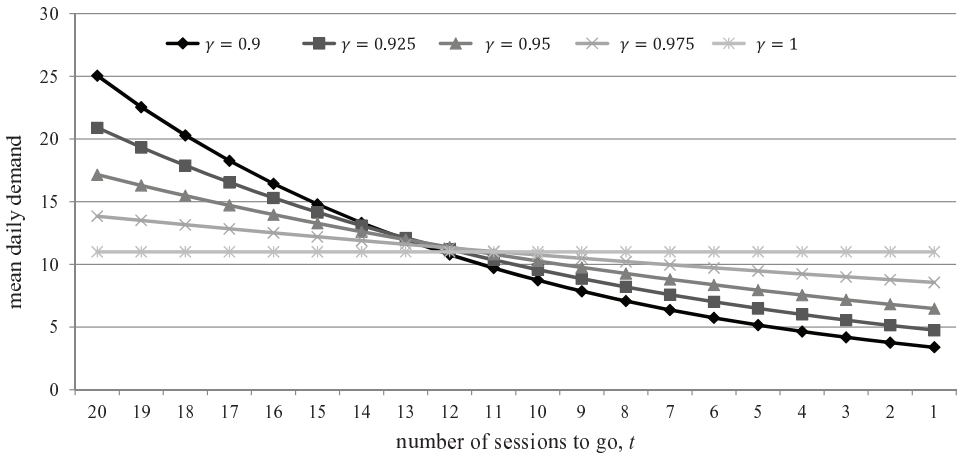



Figure 26: Mean daily demand as a function of $t$ and $\gamma$ for $\mu_0 = 11$ and $T = 20$.

Next, consider the variation in the arrival rate within each session. This variability is a consequence of establishing guaranteed hours, which encourages patients to arrive earlier in the day to guarantee their vaccinations. It is also natural to assume that as $\hat{h}$ approaches zero or $\eta$, its impact on the arrival rate decreases. That is, when the window of guaranteed hours is very narrow or very wide, a smaller portion of the arrivals shift to guaranteed hours than when the window closes midday. To capture these properties, we assume that there is a directly proportional relationship between $\hat{p}_t$ and $\check{p}_t$. Let $\beta$ be the coefficient of proportionality between $\hat{p}_t$ and $\check{p}_t$, i.e.,

$$\hat{p}_t = \beta \check{p}_t, \tag{3.11}$$

where

$$\check{p}_t = \frac{\mu_t}{\eta + \hat{h}(\beta - 1)} \qquad \text{and} \qquad 1 \leq \beta \leq \begin{cases} \frac{\eta - \hat{h}}{\mu_t - \hat{h}}, & \mu_t - \hat{h} > 0, \\ \infty, & \mu_t - \hat{h} \leq 0. \end{cases} \tag{3.12}$$

Based on Equation (3.12), when there are no guaranteed hours of operation, i.e., $\hat{h} = 0$, the arrival rate during the session is stationary and equal to $\frac{\mu_t}{\eta}$. As $\hat{h}$ increases, the arrival rate after the guaranteed hours, $\check{p}_t$, decreases. Therefore, because of the proportional relationship between $\check{p}_t$ and $\hat{p}_t$, $\hat{p}_t$ is inversely related to $\hat{h}$. For example, Figure 27 plots the percentage of patients who arrive during guaranteed hours as a function of $\hat{h}$ and $\beta$. As seen in Figure 27, as $\beta$ and $\hat{h}$ increase, the percentage of arrivals during guaranteed hours increases. As intended, the impact of $\beta$ on the proportion of arrivals during guaranteed hours is smaller for more extreme values of $\hat{h}$.

In summary, based on the values of $\gamma$ and $\beta$, we can easily specify the special cases of the MDP model established in Figure 25. Setting $\gamma \in (0, 1)$ and $\beta = 1$ allows for non-stationarity in the daily demand across the replenishment cycle only and corresponds to the SN-MDP model. If $\gamma = 1$ and $\beta \in (1, \infty)$, then only the arrival rate within each session is non-stationary, which corresponds to the NS-MDP model. The SS-MDP model corresponds to the completely stationary case in which $\gamma = \beta = 1$. Setting $\gamma \in (0, 1)$ and $\beta \in (1, \infty)$ corresponds to the most general model, NN-MDP, in which both types of non-stationarities are present.

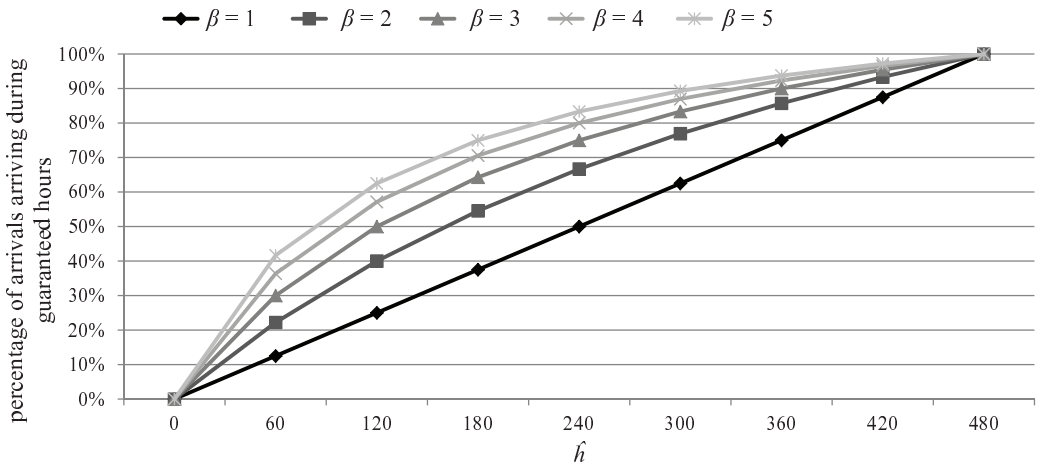



Figure 27: Proportion of patients who arrive during guaranteed hours as a function of $\hat{h}$ and $\beta$ for problem instances with $\eta = 480$ and $\gamma = 1$.

## 3.3 POLICY PERFORMANCE

In this section, we analyze how the presence of a non-stationary arrival rate within each session ($\beta > 1$), non-stationary demand across sessions ($\gamma < 1$) and the possibility that unvaccinated patients will return during guaranteed hours in the following session ($\rho > 0$) impacts the performance of the optimal vaccine administration policy. To do so, we define the following five mutually exclusive categories of patients, where the associated $\phi$ value denotes the fraction of patients that fall into that category:

**Category 1:** Patients vaccinated on their first attempt, $\phi$.

**Category 2:** Patients not vaccinated on their first attempt because the inventory is exhausted (and hence do not make a second attempt), $\phi'$.

**Category 3:** Patients not vaccinated on their first attempt because vaccinations have been suspended and who do not make a second attempt, $\phi''$.

**Category 4:** Patients who make a second attempt (during guaranteed hours), and are vaccinated on that attempt, $\hat{\phi}$.

**Category 5:** Patients who make a second attempt, but are not vaccinated on that attempt because the inventory is exhausted, $\hat{\phi}'$.

To compute the percentage of patients in each category we use the simulation model developed in Section 2.2. This model simulates a clinic that implements the optimal vaccine adminstration policy over one replenishment cycle for one type of vaccine. For each problem instance, we run 10,000 replications of the simulation and compute the percentage of patients in each category. In the remainder of this section, rather than reporting $\phi'$ and $\hat{\phi}'$ individually, we report their sum, which represents the total percentage of patients who are not vaccinated due to stock outs (although in all of the instances we simulated, $\hat{\phi}' \approx 0$).

In every problem instance considered, we set $T = 20$ days, $z = 10$ doses per vial, $\eta = 480$ timeslots and $\hat{h} = 240$ timeslots. The remaining parameters vary within the ranges reported in Table 7. The values of $\alpha$ presented in Table 7 correspond to buffer stock and are the closest achievable values under $z = 10$ to the values 10%, 15% and 25%, respectively. Additionally, for computational ease we limit the maximum number of arrivals during each session. More specifically, we define $\check{y}_1^*(h)$ to be the maximum number of arrivals after timeslot $h$ ($> \hat{h}$) and let

$$\check{y}_1^*(h) \equiv \min\left\{ y \middle| \sum_{i=y}^{\eta-h} B_{\eta-h,\check{\rho}_2}(i) < 0.01 \right\}. \tag{3.13}$$

Equation (3.13) computes the minimum number of arrivals for which the cumulative probability of arrival is larger than 0.99.

Table 7: Parameter values used in Section 3.3.

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| $T$ | 20 | $\beta$ | 1, 2, 3, 4, 5 |
| $z$ | 10 | $\gamma$ | 0.9, 0.925, 0.95, 0.975, 1 |
| $\eta$ | 480 | $\rho$ | 0, 0.25, 0.5, 0.75, 1 |
| $\hat{h}$ | 240 | $\alpha$ | 9.1%, 18.2%, 27.3% |
| $\mu_0$ | 11 | | |

### 3.3.1 Impact of Non-Stationarity

In this section, we investigate the impacts of different types of non-stationarities in the patient arrival process (i.e., both during and across sessions) on the performance of the optimal vaccine administration policy.

**3.3.1.1 NS-MDP Model** First, consider arrival rate variability within each session only, i.e., the case in which $\beta > 1$, but $\gamma = 1$, or the mean demand per session is constant over the replenishment cycle, and the MDP model given by Equation (3.1) reduces to the NS-MDP model (Figure 25). We refer to the corresponding optimal vaccine administration policy as the NS-policy. Figure 28 shows the performance of the NS-policy under different combinations of $\beta$, $\rho$ and $Q$. For a fixed probability of return, $\rho$, as the difference between the arrival rates during and after guaranteed hours increases (i.e., as $\beta$ increases), the total coverage, $\phi + \hat{\phi}$, decreases. As the difference in arrival rates increases, we can expect to open more vials during guaranteed hours. As a result, there is less of an opportunity for the NS-policy to optimize the vaccine administration process.

Furthermore, as the difference between the arrival rates increases, the number of arrivals after guaranteed hours decreases (Figure 27) and consequently, the NS-policy favors earlier closing times to ensure better use of the last vial opened during the session (Figure 29). In addition, the percentage of second attempt vaccinations also decreases. That is, even though the optimal policy favors earlier closing times when we have a larger difference between the arrival rate during and after the guaranteed hours, the percentage of patients who arrive after vaccinations have been discontinued for the rest of the day, i.e., $\phi'' + \hat{\phi} + \hat{\phi}'$, does not increase. This behavior indicates that the model is able to adequately respond to the reduction in the arrival rate after guaranteed hours. For example, in Figure 29, for $Q = 24$ and $\rho = 0.5$, as $\beta$ increases and the number of patients arriving after guaranteed hours drops, the total expected number of hours closed over the replenishment interval increases from approximately 18 hours to 34 hours while $\phi'' + \hat{\phi} + \hat{\phi}'$ *decreases* from 10.6% to 3.5%.

Under a fixed probability of return ($\rho$), the percentage of patients vaccinated on their first attempt, $\phi$, does not follow a predictable pattern with respect to within-day variations

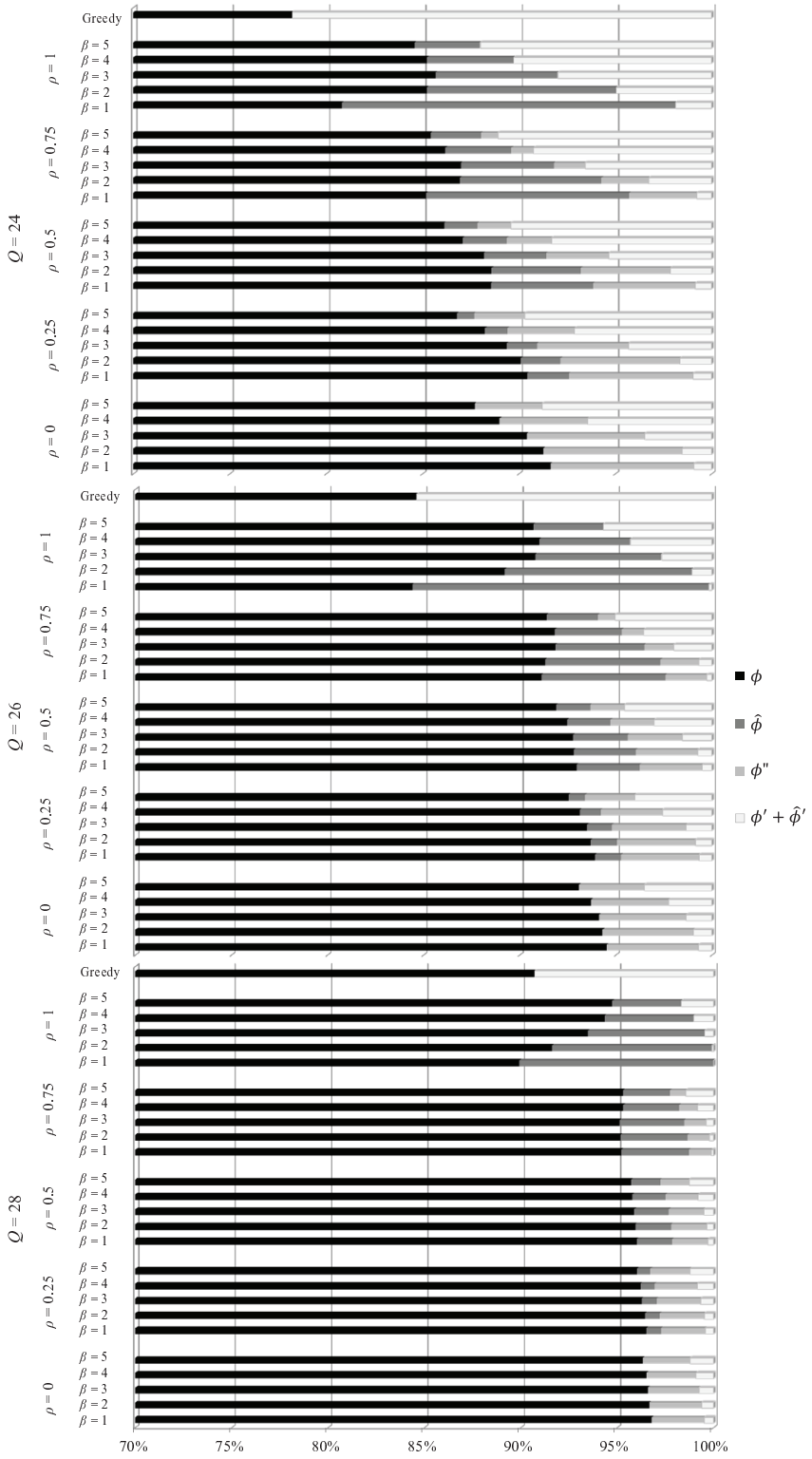



Figure 28: Proportion of patients in each category under the NS-policy as a function of $\beta$, $\rho$ and $Q$ for the problem instance defined in Table 7 with $\mu_0 = 11$ and $\gamma = 1$.

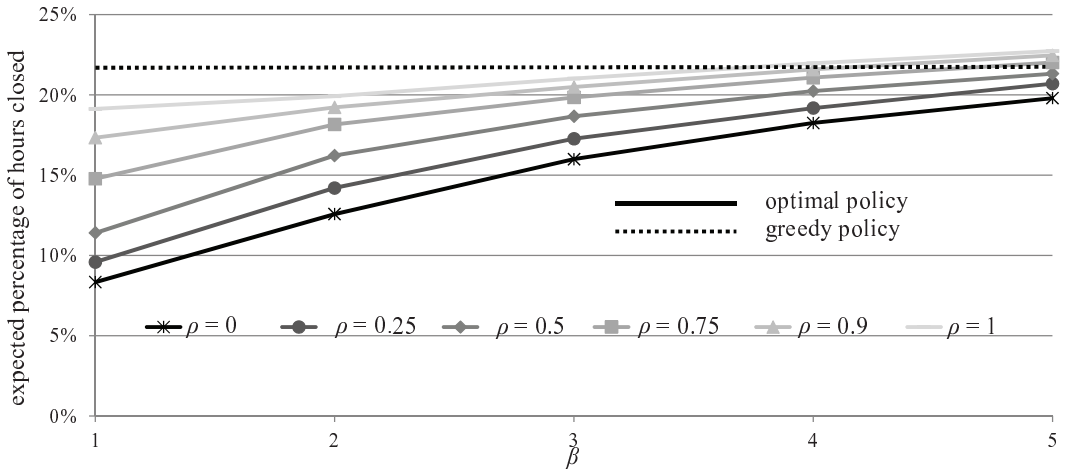



Figure 29: Expected percentage of hours closed under NS-Policy as a function of $\beta$ and $\rho$ for the problem instance defined in Table 7 with $\mu_0 = 11$, $Q = 24$ and $\gamma = 1$.

in the arrival rate. For example, in Figure 28, for $Q = 26$ and $\rho = 0.75$, as $\beta$ increases, $\phi$ first increases from 91.1% to 91.9% and then decreases to 91.4% whereas for $\rho = 0.25$ it decreases from 93.9% to 92.5%. Also evident in Figure 28 is the fact that larger total coverage rates do not always coincide with a greater percentage of vaccinations on the first attempt. For example, for $Q = 24$ and $\rho = 0.75$, the total coverage rate increases from 91.8% to 95.7% as $\beta$ decreases from 3 to 1 while $\phi$ decreases from 87.0% to 85.2%. If a clinician's main priority is vaccinations on the first attempt, increasing initial inventory is an option; based on Figure 28, increasing $Q$ results in a larger portion of patients being vaccinated on their first attempt, because the NS-policy is more willing to open vials close to the end of a session when a greater amount of buffer stock is available.

Next, consider the roles played by the probability of return and within-day arrival rate variation in determining the required buffer stock to reach a predefined coverage proportion on the first attempt. Let $\alpha_x$ be the buffer stock required to reach a minimum first attempt coverage proportion of $x$, i.e., $\alpha_x = \min\{\alpha|\phi > x\}$. Figure 30 demonstrates, for $x = 90\%$, that when the probability of return increases, $\alpha_{90}$ also increases. However, the relationship between $\alpha_{90}$ and within-day arrival rate variation is not monotonic. As the probability of

return increases, the NS-policy becomes more conservative regarding opening a new vial as the end of each session approaches and as a result, a larger proportion of patients are vaccinated on their second attempt. Therefore, when the probability of return is higher, having more buffer stock increases the number of vaccination-hours per session, which in turn leads to a larger percentage of vaccinations on the first attempt. When the probability of return is very large ($\rho$ close to one) and the within-day arrival rate variation is very small ($\beta$ close to one), larger buffer stock is required to reach $\phi = 90\%$ than under other combinations of $\rho$ and $\beta$, because under these conditions a relatively large number of patients arrive after guaranteed hours and may experience discontinued vaccination. Therefore, if determining precise values of $\rho$ and $\beta$ is not possible, to reach a specific minimum fraction vaccinated on the first attempt, $x$, a conservative buffer stock level would be the $\alpha_x$ value corresponding to the case in which all unvaccinated patients return for a second attempt and the arrival rate is stationary within each session.

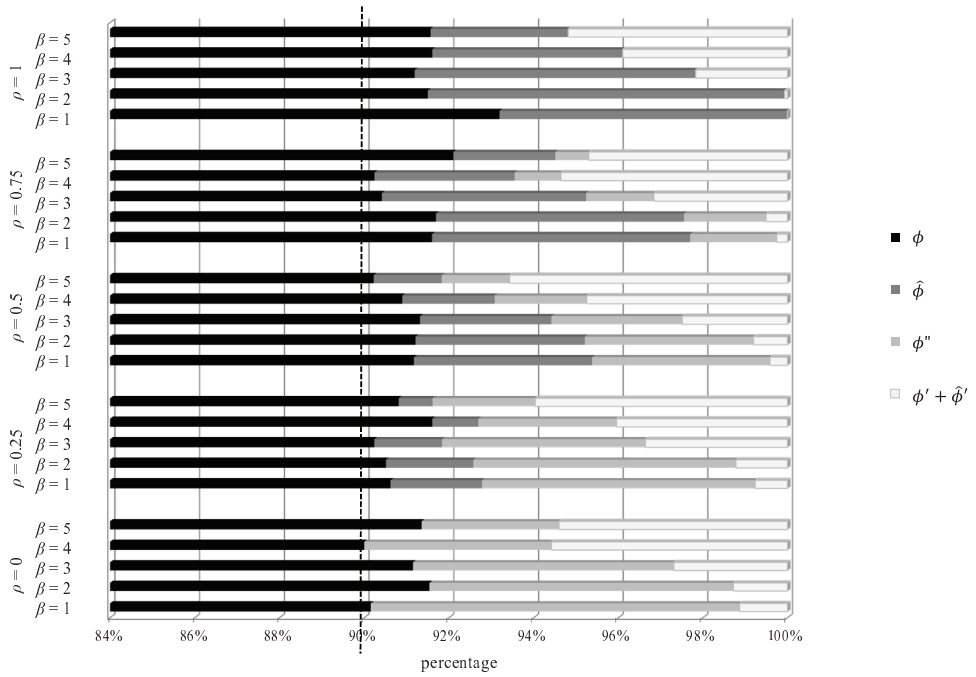



Figure 30: Proportion of patients in each category under NS-Policy as a function of $\beta$ and $\rho$ for the problem instance defined in Table 7 with $\mu_0 = 15.125$, $\gamma = 1$ and $\alpha = \alpha_{90}$.

**3.3.1.2 SN-MDP Model** Next, we explore the effect, in isolation, of non-stationary mean daily demand on the performance of the vaccine administration policy by assuming that the arrival rate within each session is constant, i.e., $\beta = 1$. Consequently, the MDP model given by Equation (3.1) reduces to the SN-MDP model (Figure 25). We refer to the corresponding optimal vaccine administration policy as the SN-policy.

Figure 31 reports the performance of the SN-policy as a function of $\gamma$ (mean daily demand variation) and $\rho$ for base demand values $\mu_0 \in \{9.165, 11, 15.125\}$. These values are chosen because to evaluate the performance of the SN-policy, it is important to consider base demands that are both smaller and larger than the vial size, $z = 10$, and because we choose base demands with equivalent buffer stock percentages so as not to confound the results. That is, maintaining a buffer stock of 9.1%, as is the case for $\mu_0 = 11$ and $Q = 24$, requires $\mu_0$ values of 9.165 and 15.125 under initial inventory levels 20 and 33, respectively.

As seen in Figure 31, for different values of the base demand, $\mu_0$, but a fixed probability of return, $\rho$, as the variation in the mean daily demand increases, i.e., as $\gamma$ decreases, different patterns emerge in the total coverage rate. For example, for $\mu_0 = 11$ and $\rho = 0.25$, as the mean daily demand variation across sessions decreases the coverage rate decreases from 92.6% to 90.3%, whereas for $\mu_0 = 15.125$ and $\rho = 0.25$, the coverage rate first increases from 94.1% to 95.0% and then decreases to 94.6%. However, for a fixed base demand, the impact of day-to-day mean daily demand variation on the coverage rate is similar across different values of $\rho$. For example, for $\mu_0 = 11$, as day-to-day mean demand variation increases, the coverage rate increases consistently across different values of $\rho$ (Figure 31). It should be noted, too, that the greedy policy performs better as demand variation across sessions increases regardless of the value of base demand (Figure 31) because under the greedy policy, clinics usually run out of vials at the end of the cycle, which coincides with the time when the mean daily demand is low.

**3.3.1.3 NN-MDP Model** Finally, we consider the most general form of the MDP model in which both types of arrival non-stationarities (within and between days) are considered simultaneously, namely the NN-MDP. We refer to the corresponding optimal vaccine administration policy as the NN-policy.

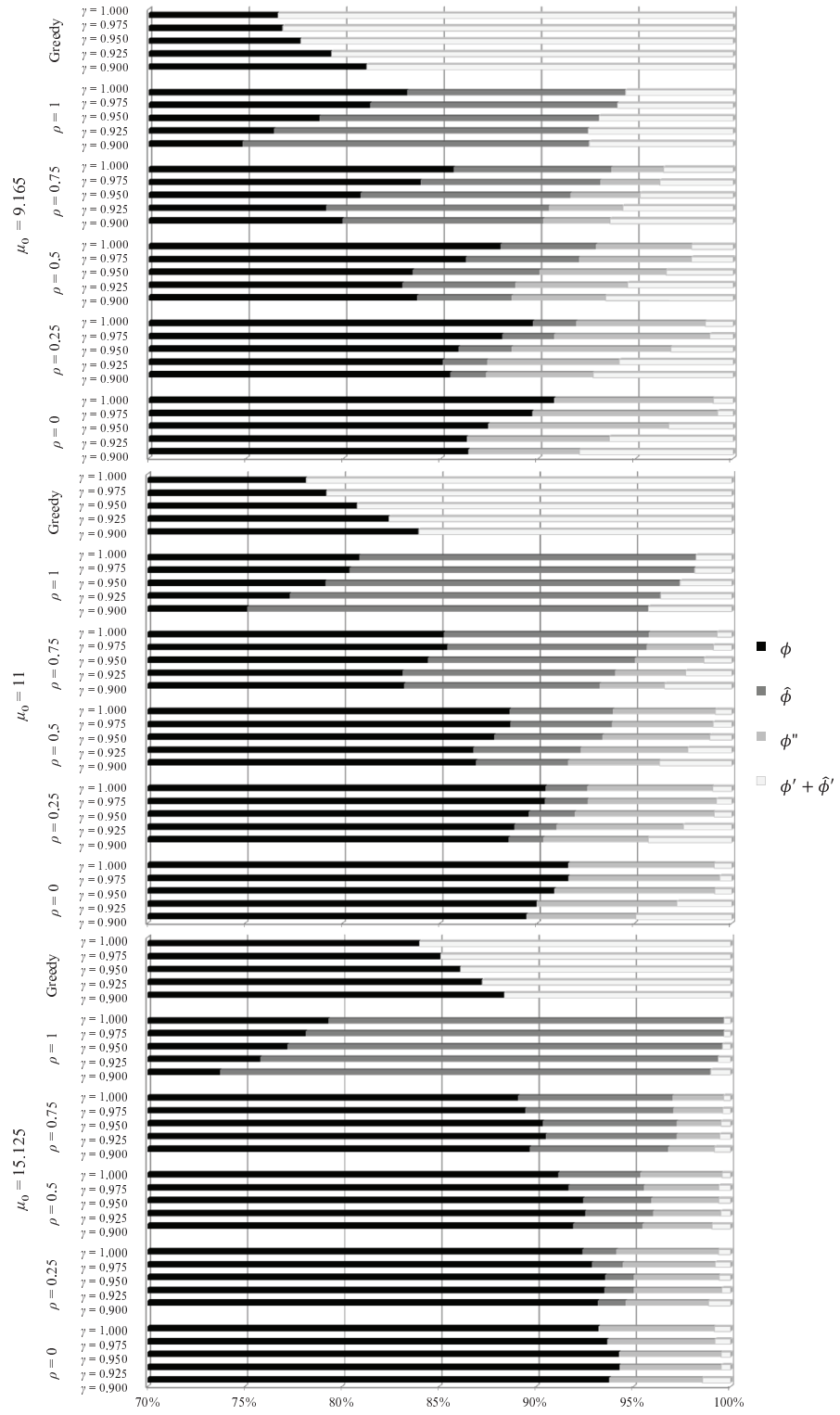



Figure 31: Proportion of patients in each category under SN-policy as a function of $\gamma$, $\rho$ and $\mu_0$ for the problem instance defined in Table 7 with $\alpha = 9.1\%$ and $\beta = 1$.

Figure 32 reports the performance of the NN-policy as a function of $\beta$, $\gamma$ and $\rho$. First, observe that the coverage rate decreases as the arrival rate variation within a session increases, regardless of the variation across sessions (consistent with Figure 28). That is, larger differences in the arrival rates during and after guaranteed hours result in poorer performance of the vaccine administration policy. Second, similar to the observation in Section 3.3.1.2, when the arrival rate is stationary within a session, variation in mean demand across sessions within the cycle has a similar impact, regardless of the probability of return, $\rho$ (e.g., for $\mu_0 = 11$, as $\gamma$ increases, the coverage rate increases). Finally, we see that the impact of day-to-day mean daily demand variation ($\gamma$) on performance when there is variation in the within-day arrival rate, i.e., $\beta > 1$, is inconsistent for different values of $\rho$. For example, for $\beta = 3$ and $0 \leq \rho \leq 0.5$, as $\gamma$ increases, the coverage rate increases, whereas for $0.75 \leq \rho \leq 1$, the coverage rate first increases and then decreases.

### 3.3.2 Impact of Delayed Service

Overall, under the optimal vaccine administration policy with a positive probability of return ($\rho > 0$), although clinicians sometimes discontinue vaccinations earlier in the day (Figure 29), they always vaccinate more patients compared with the policy in Chapter 1. That is, an increase in the probability of return increases the coverage rate because a portion of unvaccinated patients return during the following session, and thus, clinicians can vaccinate more patients by possibly discontinuing vaccinations earlier on certain days. The improvement in the performance of the vaccine administration policy is evident in Figures 28, 31 and 32. For example, in Figure 28, for $Q = 24$ and $\beta = 2$, as the probability of return increases from 0 to 1, the coverage rate increases from 91.3% to 95.1%. An increase in $\rho$ decreases the percentage of patients vaccinated on their first attempt, $\phi$. For example, in Figure 28, for $Q = 24$ and $\beta = 2$, as $\rho$ increases from 0 to 1, $\phi$ decreases from 91.3% to 85.2%. However, the increase in the percentage of patients vaccinated on their second attempt, $\hat{\phi}$, is larger than the reduction in the first attempt percentage. Lastly, for instances with a probability of return close to 1, it is often optimal to suspend vaccinations, which increases the number of hours closed (Figure 29). However, the overall coverage rate is still higher (as explained in

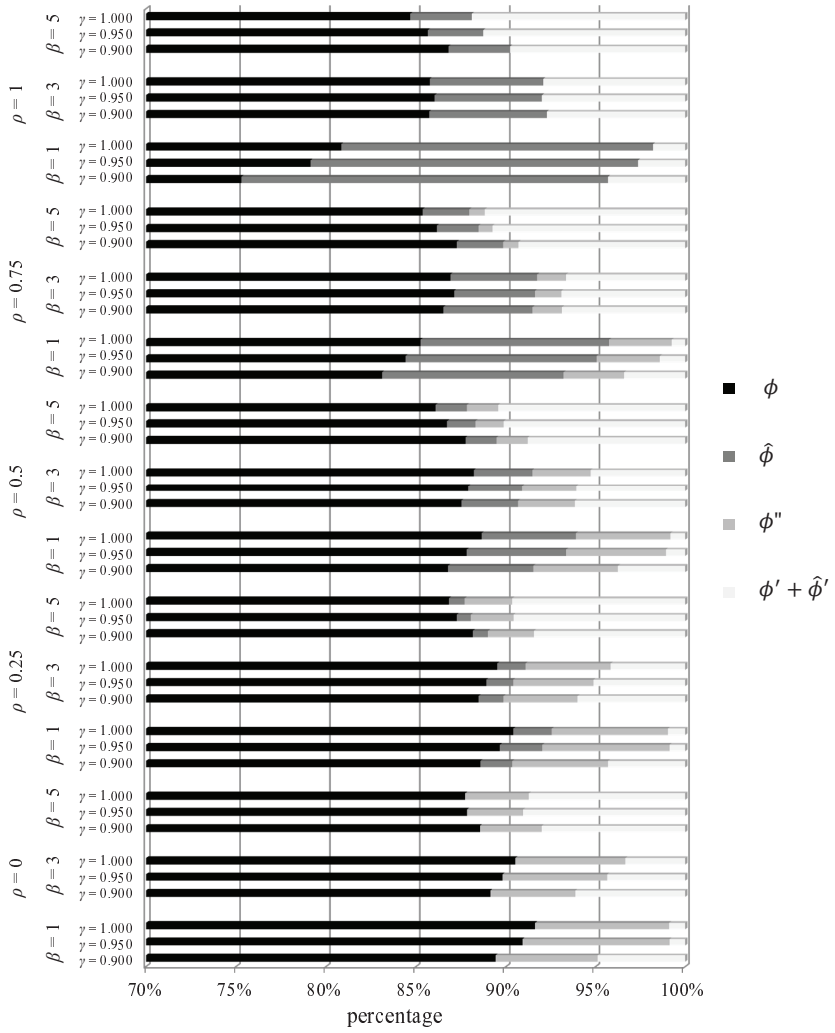



Figure 32: Proportion of patients in each category under NN-Policy as a function of $\beta$, $\rho$ and $\gamma$ for the problem instance defined in Table 7 with $\mu_0 = 11$ and $Q = 24$.

Section 3.3.1.1). In contrast to the optimal vaccine administration policy, the greedy policy is unable to take advantage of the fact that a patient turned away on one day has a positive probability of returning the next day because all patients are vaccinated as long as inventory is on hand.

In summary, the computational results presented in Section 3.3 demonstrate three main relationships. First, if there is significant variation within a day (large $\beta$), then most of

the patients arrive during the guaranteed hours and must be served, hence the optimal policy performs similarly to the greedy policy because there are fewer decision opportunities. Second, the effects of demand variation between days in the replenishment cycle ($\gamma$) are very problem specific. Changes in $\gamma$ can increase or decrease coverage. An increase in the value of $\gamma$ generally improves the performance of the greedy policy because there are fewer arrivals at the end of the month when the inventory of vials is most likely to be exhausted. Third, as $\rho$ increases, patients can be pooled (patients from late in the day on day $i$ are asked to return on day $i+1$) to reduce OVW. Therefore, although fewer patients might be vaccinated on their first attempt, the overall coverage increases.

## 3.4 EXTRAPOLATING BEYOND A SINGLE CLINIC

Whereas Section 3.3 examines the performance of the optimal policy under different sets of parameter values for a single replenishment cycle in a single clinic, in this section we explore the performance of the optimal policy on a broader scale. More specifically, we estimate the expected number of additional vaccinations that could be administered (along with the corresponding percentage reduction in OVW) across multiple clinics over the course of an entire year if the current policy is replaced by the optimal policy. We then extrapolate these results to estimate the potential benefits across all developing countries supported by the Global Alliance for Vaccines and Immunization (GAVI).

Currently, 53 countries are eligible for support from GAVI based on having a Gross National Income per capita below US $\$1,570$. As an illustrative example from this list we choose India, with a population of approximately 1.27 billion and an annual birth cohort of approximately 25.5 million [12]. We focus on the state of Bihar with an annual birth cohort of approximately 3 million. To obtain an estimate of the improvement in Bihar over a one-year period (the year 2015), we assume that the birth rate is constant throughout the year and that all clinics are identical. We also assume a monthly replenishment cycle, as is most common in practice, and that there are 20 sessions per replenishment cycle (equivalent to the clinic being open 5 days a week). As an illustrative vaccine, we choose measles. Under

the EPI schedule, each child requires two doses of measles vaccine, which is manufactured in 10-dose vials. To estimate the average number of arrivals per session we consider 160 clinics in Bihar for which we have data on the infant population (age 0-11 months) covered by each clinic; these clinics vaccinate approximately 31.8% of the Bihar infant population. In 2013, the total number of infants age 0-11 months in these 160 clinics in Bihar was approximately 858,800 and approximately 25% of them, or 214,700 infants, were vaccinated in clinics; the remaining 75% were vaccinated through outreach sessions. The infant population for 2015 is estimated to increase by 2.01% (two years of a 1% annual growth rate in India [49]) over that of 2013. Thus, the expected number of daily arrivals for measles vaccination at each clinic is $\frac{858800\times2\times0.25\times(1+0.0201)}{160\times12\times20} = 11.41$ patients per session. Unfortunately, we do not have reliable data on the day-to-day and within-day arrival patterns, nor do we have estimates of the likelihood that a patient would return the next day when denied vaccination on a given day. Therefore, we evaluate the gains realized under the optimal policy across a range of values for the corresponding parameters. In particular, the following sets of parameters values are considered: $\beta \in \{1, 1.5, 2, 2.5, 3, 5\}$, $\gamma \in \{0.9, 0.925, 0.95, 0.975, 1\}$ and $\rho \in \{0, 0.25, 0.5, 0.75, 1\}$.

Table 8 summarizes the improvements in the state of Bihar over a one-year period achieved by implementing the optimal policy as opposed to a greedy policy. We consider two different levels of buffer stock, a standard value of 27% and a reduced value of 10%. (Ideally, we would choose the WHO recommendation of 25% for the standard value, but because vials can only be ordered in discrete quantities, 27% is the closest achievable value to 25%.) For each case, we evaluate the improvement for each of the $6 \times 5 \times 5 = 150$ possible combinations of values considered for $\beta$, $\gamma$ and $\rho$. Table 8 reports the maximum (best case), minimum (worst case) and average values across all of these combinations for (*i*) the coverage rate under the optimal policy, (*ii*) the number of additional vaccinations administered and (*iii*) the reduction in open vial waste. Several observations may be drawn from the results.

First, under 27% buffer stock, the optimal policy results in over 72,000 more vaccinations annually than the greedy policy when averaged across all parameter values; this corresponds to approximately 5.3% of the total demand. With only 10% buffer stock, the improvement is even more dramatic with over 146,000 additional vaccinations on average (slightly more than

10% of the total demand). Second, in the best case scenario there are situations in which implementing the optimal policy can result in approximately 99.99% vaccination coverage, and in the worst case scenario, even with just 10% buffer stock, the optimal policy yields above 88% coverage in all cases while the greedy policy never yields more than 85% coverage.

The potential number of additional vaccinations is even more astonishing on a global level. With a 27% buffer, based on Table 8 and an annual birth cohort in Bihar of 2,997,369, 25% of which is vaccinated in clinics, the worst-case rate of additional vaccinations per year is $\frac{31{,}260}{0.25\times 2{,}997{,}369} \times 100\% = 4.2\%$. For a birth cohort in the remaining states of India and the other GAVI countries of 69,956,943 [12], even under the assumption that only 25% of the population is vaccinated through clinics and extrapolating from the worst-case increase in coverage in Bihar, the results suggests that implementing the optimal policy worldwide would result in more than $69{,}956{,}943 \times 25\% \times 4.2\% \approx$ *730 thousand* additional measles vaccinations at no additional cost. In reality, far more than 25% of patients worldwide are vaccinated through in-clinic operations; if we assume this percentage is (for example) 80%, implementing the optimal policy would result in more than approximately *2.3 million* additional measles vaccinations. Furthermore, there are many combinations of parameter values for which the increase is much more than the conservative estimate of 4.2%, in which case the corresponding increase in the number of vaccinations would be significantly higher.

Implementing the optimal policy also reduces OVW and therefore decreases the vaccination cost per child. For the Bihar data, as seen in Table 8, the percent reduction in OVW averaged across all parameter combinations is approximately 6% for the 27% buffer stock case and approximately 10% for a buffer stock of 10%. At a cost of US $3.0 per 10-dose vial [49], with buffer stocks of 27% and 10%, the cost per vaccination under the optimal policy is US $0.38 and US $0.35, respectively, whereas under the greedy policy, these costs are approximately US $0.41 and US $0.40 per vaccination. Although the measles vaccine is relatively inexpensive, this reduction in OVW can result in significant savings for other, more expensive vaccines (e.g., Hib, Haemophilus Influenzae Type B). In summary, when viewed from a world-wide perspective, the benefits of using an optimal policy could be tremendous in terms of coverage as well as cost savings.

Table 8: Summarized results of implementing the optimal policy in Bihar, India.

| buffer stock | | 10% | 27% |
|---|---|---|---|
| coverage rate under optimal policy | maximum | 98.6% | 99.99% |
| | average | 92.4% | 98.1% |
| | minimum | 88.3% | 96.4% |
| expected number of additional vaccinations per year (versus the greedy policy) | maximum | 271,154 | 122,431 |
| | average | 146,273 | 72,412 |
| | minimum | 64,074 | 31,260 |
| reduction in OVW rate | maximum | 20.0% | 12.7% |
| | average | 10.0% | 5.7% |
| | minimum | 4.3% | 2.0% |

## 3.5 CONCLUSIONS

Limited supplies of (lyophilized, multi-dose) vials in remote clinics gives rise to difficult vial-opening decision problems. In Chapter 1, we address these problems, but under three simplifying assumptions: (*i*) demand is stationary within each session, (*ii*) demand is stationary across sessions in each replenishment cycle, and (*iii*) patients who are turned away because it is suboptimal to open a new vial at the time of their arrival, are lost. In this chapter we formulate a new MDP model that relaxes these assumptions. This more general model determines how to optimally administer vaccines from multi-dose vials when nonstationary demand and patient returns are possible, thereby outperforming current practice and the policies in Chapter 1.

We perform an extensive numerical study to evaluate the impact of the probability of return and demand variation on the performance of the optimal vaccine administration policy. As expected, higher coverage is gained as the probability of return increases; hence,

motivating unvaccinated patients to return in the following session by guaranteeing vaccinating can improve overall performance. If the vast majority of patients arrive during the guaranteed hours then there is less decision making opportunity to optimize vial usage and the optimal coverage rate is reduced. Variation in the daily demand across the days of the month has a complicated effect on the optimal policy, sometimes increasing coverage and sometimes decreasing coverage, that depends on other factors such as the expected daily demand and the number of sessions between replenishments. Finally, we use data from the state of Bihar, India as a case study to estimate the increase in the expected number of vaccinations over a one-year period under the optimal policy; extrapolating these results to the other GAVI countries around the world suggests that behaving optimally has the potential to significantly increase coverage and reduce waste.

# 4.0 OPTIMAL DESIGN OF FIXED AND OUTREACH VACCINATION SERVICES

## 4.1 INTRODUCTION

The World Health Organization (WHO) operates the Expanded Program on Immunization (EPI) with the objective of maximizing the number of children who are immunized against many vaccine-preventable diseases including measles, hepatitis and polio. The WHO places special emphasis on lower and middle income countries that have low vaccination budgets, limited resources and often inadequate infrastructure. After launching the EPI, the coverage rate for core vaccines such as diphtheria, tetanus and pertussis (DTP), Bacille Calmette-Gurin (BCG), polio and measles increased significantly from less than 5% in the 1970s to more than 85% in 2014 [35]. For DTP and measles alone this has helped to prevent approximately 2 to 3 million additional deaths each year and has protected many more people from illness and disability [58].

Despite these efforts there are still a large number of children (1.5 million in 2013 [58]) who die from vaccine-preventable disease each year. Moreover, as reported by UNICEF [50] in 2014, more than 30 million children did not receive recommended vaccinations. UNICEF cites poorly managed or inaccessible health services as one of the main reasons for such a large volume of uncovered children. Indeed, in many countries (e.g., Niger, Ethiopia, Kenya), a substantial fraction of the population lives in remote locations without access to private or public transportation and hence have limited access to clinics [3]. This problem is particulary acute during the wet season and under severe weather conditions. Consequently, transporting vaccines to hard-to-access areas (i.e., outreach) can significantly improve vaccination coverage [57].

To alleviate these barriers to coverage, in 2002 the WHO, UNICEF and other partners introduced the "reaching every district" (RED) strategy [52]. One of the operational components of RED is "reaching targeted populations," which focuses on developing a vaccine delivery strategy for remote areas with poor access to health services. Trips made from clinics to these remote areas to provide vaccination are referred to as outreach trips.

When using strategies involving outreach, government health organizations face a tradeoff between establishing clinics and performing outreach. Daskin et al. [6] state poorly located facilities or the use of too many or too few facilities can result in unnecessary costs and/or degraded customer service. In the content of clinic and outreach trip trade-off, establishing more clinics is desirable from a coverage and patient perspective because it provides everyday vaccination opportunities for patients in traveling-distance locations as long as there is inventory on-hand. However, the high cost associated with facilities, salaried laborers and equipment makes it impractical to establish clinics in all locations and requiring patient to travel to a clinic negatively influences coverage (e.g., [3, 29, 34]). In contrast outreach can be more affordable than establishing a clinic and eliminates patient travel, but provides a limited number of vaccination opportunities. Moreover, outreach trips from a single clinic cannot cover all locations due to constraints on outreach trip size and distance. Therefore, we propose a mechanism to find the best combination of clinic locations and outreach trips from these clinic locations to cover an entire set of locations over a given time horizon.

Considerable research has been done on various healthcare facility location problems; see Daskin and Dean [6] for a review of single-level models in this area, Rahman et al. [39] for a review specific to developing countries and the more recent review by Rais et al. [40] for problems in providing adequate and proper healthcare service. However, to the best of our knowledge, [15, 45, 53, 27] are the only works which incorporate the fact that requiring patients to travel to a vaccination site negatively influences coverage in healthcare facility location problems. These studies focus on finding the optimal clinic locations only whereas we simultaneously consider planning outreach trips which expand a clinic's area of influence. That is, the combination of clinic location and outreach planning has not been previously studied.

The actual demand in population centers is usually unknown, which implies that demand uncertainty must be considered in healthcare facility location problem as well. A variety of approaches are proposed to deal with the demand uncertainty involved in facility location problems and these are summarized in [43]. We use a chance-constrained programming approach to model the uncertainty that exists in the demand by imposing a minimum coverage requirement at each location individually, similar to [30].

Combining clinic location and outreach planning is challenging due to the operational constraints associated with outreach trips, the most influential of which are the upper bounds on (*i*) the travel distance (trips typically have to be completed in a single-day), (*ii*) the number of outreach trips performed during a specific period of time (due to budget and manpower limitations), and (*iii*) the number of potential doses of vaccine per outreach trip (as result of vaccine carrier capacity).

The outreach trip planning problem from a given location(s) has not received much attention in the literature despite its critical role in improving coverage for hard-to-access populations. Lim et al. [27] are the only ones who develop a mathematical framework to model outreach coverage. Lim et al. in [27] develop multiple models to plan outreach trips for one or a given set of clinic locations and compare their performance. In each model, they consider different types of outreach coverage as a function of traveling distance, e.g., binary coverage, step-wise coverage (decreasing function of traveling distance) and partial coverage from multiple locations. Similar to outreach planning, the mobile healthcare facilities planning problem [8, 17] aims to improve accessibility to health care facilities by performing trips to hard-to-access areas. However, mobile facility location planning considers trips longer than a single day that include multiple destinations, whereas outreach involves single-day trips to one destination.

The problem we considered can also be reviews as a set covering problem (SCP). SCPs find the optimal set of facility locations that minimizes the cost or required number of facilities to cover a set of locations. A considerable amount of work has been done in this area and comprehensive literature reviews [31, 42] exist on the SCP. Most of the SCP work in the literature, is on the uncapacitated problem [11]; our problem, however, is a capacitated SCP [5]. Limitations on the size and number of outreach trips restrict the number of patients

and consequently the number of locations that could be covered by a clinic. Our problem is unique among existing capacitated SCP models for several reasons. First, a clinic's area of influence can be extended by performing outreach trips, which adds considerable complexity to the problem. Second, we consider capacity restrictions on outreach trips, which indirectly imposes capacity restrictions on the clinics. Finally, outreach trips have their own area of influence, which is a function of outreach size and existing demand within traveling distance of their destination.

Our problem can also be considered as a location-routing problem (LRP) because a clinic can perform outreach trips to remote locations. LRPs are facility location problems that simultaneously take into consideration routing aspects. Because of the interrelation between the facility location and the routing problem, interest in this type of problem has been growing in recent years and much work has been done in this area. Detailed literature reviews on the LRP are provided in [22, 32] and [37]. Our problem is a special case of an LRP problem because outreach trips are performed daily and restricted to just one destination in contrast to the general routing problem in which multiple destinations are visited in one trip. However, our problem is distinct from the existing literature because (*i*) demand cannot only be covered by an outreach trip but also by clinics, (*ii*) one outreach trip can simultaneously cover any subset of locations that are within traveling distance of the outreach location (it is a function of the outreach trip size), and lastly (*iii*) because of the trip size constraint, outreach trips can result in partial coverage and consequently, one location may covered through multiple outreach trips from different origins.

The remainder of this chapter is organized as follows. In Section 4.2, we first develop a mixed integer linear model which determines the appropriate clinic locations and plans the outreach trips from each location. We then discuss how to incorporate demand uncertainty into the model. In Section 4.3, we solve a small numerical example and explore its optimal solution. In Section 4.4, we perform sensitivity analysis on the primary model parameters. In Section 4.5, we propose two alternative heuristic approaches to obtain benchmarks and compare their performance with that of the proposed model. Finally, in Section 4.6, we discuss the limitations of our approach and provide some summary remarks.

## 4.2 MODEL FORMULATION

### 4.2.1 MILP Model

Consider a set of locations with known demand that must be vaccinated within a specified time horizon (we relax the known demand assumption in Section 4.2.2). We set the length of the time horizon based on dosage-schedule requirement, i.e., the frequency required to deliver vaccines to outreach locations [57], e.g., if each location is required to be visited twice a year then the time horizon is 6 months. The model formulated in this section determines the optimal set of clinic locations among these potential locations and their corresponding outreach activities. More specifically, the mixed integer linear programming (MILP) model minimizes the costs associated with operating the clinics, performing outreach trips and requiring patient travel, while simultaneously enforcing limits on outreach trip size and travel distance for both outreach trips and patients. To develop this model, we make the following assumptions:

1. There are upper bounds on the number of outreach trips per clinic and the number of potential vaccinations per outreach trip.
2. Under any given solution, the total required number of doses for the entire horizon can be made available at each clinic location at no cost.
3. The distance between all locations is fixed and known.
4. Round trip outreach trips must be performed within a single day. That is, only locations within daily travel distance of a clinic can be covered through outreach trips from that clinic.
5. There is a limit on the distance that patients are willing to travel to get vaccinated.
6. Patients within traveling distance of multiple clinics travel to the closest clinic.
7. In locations within traveling distance of at least one clinic, the entire demand is covered in the least costly way, i.e., either at the closest clinic or via outreach trips to the location. That is, patients do not travel to an outreach destination if there is a clinic within traveling distance of the location.

Accordingly, as seen in Figure 33, each location is covered in one of five ways: $(i)$ an on-site clinic (location 0), $(ii)$ patient travel to a clinic within traveling distance (location 1), $(iii)$ a direct outreach trip(s) (locations $2, 3, 4$ and 5), $(iv)$ an indirect outreach trip(s) (locations 6 and 8), i.e., patients travel to a nearby location that is an outreach trip destination site, or $(v)$ a combination of direct and indirect outreach trips (location 7).

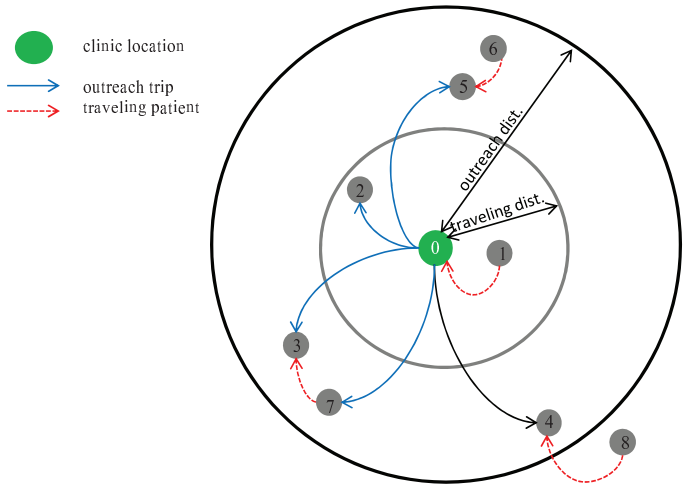



Figure 33: Different ways locations may be covered.

We don't model the timing of the outreach trips during the planning horizon. That is, given a feasible number of trips to perform, we assume that each clinic spreads the trips out in a roughly even fashion. Beyond the first time horizon, we implement the same solution unless drastic change in the population observed in locations. In this case, we may resolve considering the fixed set of clinic locations.

Let $n$ be the number of locations, $\rho_i$ be the known demand in location $i$ and $d_{ij}$ be the travel distance between location $i$ and $j$. Let $\Delta$ be the maximum patient travel distance and $S_i$ be the set of locations within $\Delta$ distance units of location $i$, i.e., $S_i = \{j | d_{ij} \leq \Delta\}$. Similarly, let $\hat{\Delta}$ be the maximum outreach travel distance and $\hat{S}_i$ be the set of locations within $\hat{\Delta}$ distance units of location $i$, i.e., $\hat{S}_i = \{j | d_{ij} \leq \hat{\Delta}\}$.

We consider two types of outreach vehicles, motorbikes and four-wheel vehicles. Four-wheel vehicles have 3-4 times the capacity of motorbikes, but also have a higher cost per unit of distance traveled. For the sake of model simplicity, we assume that only one type of vehicle is available in each clinic, however, we can easily relax this assumption. Let $\delta_M$ and $\delta_F$ be the maximum number of vaccinations per outreach trip performed by a motorbike and four-wheel vehicle, respectively.

Let $x_i$ be a zero-one decision variable that equals 1 when there is a clinic at location $i$ and 0 otherwise. Based on Assumption (7), if location $i$ is within traveling distance of at least one clinic, it is covered through either traveling-distance clinic or direct outreach trips. Let $\hat{x}_i$ and $\check{x}_i$ be zero-one decision variables which determine whether the entire population in location $i$ travels to a traveling-distance clinic or clinicians perform direct outreach trips to location $i$. That is, $\hat{x}_i$ equals 1 if the entire population in location $i$ travels to a traveling-distance clinic and 0 otherwise as opposed to $\check{x}_i$ which equals 1 if the entire population in location $i$ is covered by outreach and 0 otherwise.

The optimization model consists of 5 sets of constraints. To better describe the model, we first present each set of constraints individually and then combine them into a single optimization model. Constraints (4.1a)-(4.1g) determine how each location that is within traveling distance of at least one clinic is covered:

$$x_i + \hat{x}_i + \check{x}_i \leq 1, \qquad i = 1, \ldots, n, \tag{4.1a}$$

$$\hat{x}_i + \check{x}_i \leq \sum_{j \in S_i \backslash \{i\}} x_j, \qquad i = 1, \ldots, n, \tag{4.1b}$$

$$x_j \leq x_i + \hat{x}_i + \check{x}_i, \qquad j = 1, \ldots, n, \qquad i \in S_j \backslash \{j\}, \tag{4.1c}$$

$$\sum_{i \in S_k \backslash \{k\}} \kappa_{ik} = \hat{x}_k, \qquad k = 1, \ldots, n, \tag{4.1d}$$

$$\kappa_{ik} \leq x_i, \qquad k = 1, \ldots, n \quad i \in S_k \backslash \{k\}, \tag{4.1e}$$

$$\hat{d}_k \leq \Delta(1 - x_i) + d_{ik} x_i, \qquad k = 1, \ldots, n, \quad i \in S_k \backslash \{k\}, \tag{4.1f}$$

$$\sum_{i \in S_k \backslash \{k\}} d_{ik} \kappa_{ik} \leq \hat{d}_k, \qquad k = 1, \ldots, n. \tag{4.1g}$$

Let define $\kappa_{ik}$ to be a binary variable that equals 1 when the closest clinic within traveling distance of location $k$ is in location $i$ and 0 otherwise. If a clinic is located in location $i$ then

constraint (4.1a) forces $\hat{x}_i$ and $\check{x}_i$ to 0. Similarly, if there is no clinic within traveling distance of location $i$ then constraint (4.1b) forces $\hat{x}_i$ and $\check{x}_i$ to 0. If there is a clinic in location $j$ (i.e., $x_j = 1$) then, for each location within the traveling distance of location $j$, constraint (4.1c) guarantees coverage either through an on-site clinic (i.e., $x_i = 1$), a traveling-distance clinic (i.e., $\hat{x}_i = 1$), or direct outreach (i.e., $\check{x}_i = 1$), but there is no indirect outreach in this case.

If multiple clinics are within traveling distance of a location, we assume that the patients in that location travel to the closest clinic unless the location is covered by a direct outreach trip. Let $\hat{d}_k$ be the minimum travel distance from location $k$ to a clinic within traveling distance of location $k$, i.e.,

$$\hat{d}_k = \begin{cases} \min_{i \in S_k \backslash \{k\}} \{d_{ik} | x_i = 1\}, & S_k \backslash \{k\} \neq \varnothing, \\ 0, & S_k \backslash \{k\} = \varnothing. \end{cases}$$

We determine $\hat{d}_k$ through constraints (4.1d)-(4.1g). Specifically, constraints (4.1d) and (4.1e) select a clinic in the traveling distance of location $k$ if location $k$ is covered by a clinic within traveling-distance (i.e., $\hat{x}_k = 1$). Simultaneously, constraints (4.1f) and (4.1g) force the clinic selected by constraints (4.1d) and (4.1e) to be the closest one to location $k$ and determine the minimum traveling distance, $\hat{d}_k$.

Generally, a limited number of clinicians are available in a clinic and are required to be present for in-clinic operation. Therefore, increasing the number of outreach trips above a threshold (e.g., one outreach trip per week) requires hiring more clinicians which increases the outreach cost per trip. Hence, let $\check{\theta}$ be the number of outreach trips after which the outreach cost increases and $\hat{\theta}$ be the maximum number of outreach trips per clinic ($\check{\theta}$ and $\hat{\theta}$ can readily be extended to be location-specific parameters $\check{\theta}_i$ and $\hat{\theta}_i$). Obviously, if $\check{\theta} = \hat{\theta}$, no such threshold exists. Let decision variables $\check{y}_{ijl}$ be the number of low-cost outreach trips from location $i$ to $j$ with a vehicle of type $l$ and $\hat{y}_{ijl}$ be the analogous number of high-cost outreach trips.

Because, we consider indirect outreach as a means of coverage, we need to determine the number of patients that we plan to vaccinate from locations within traveling distance of each outreach destination. To do so, let $s_{k,k'}$ be a decision variable that indicates the number of patients from location $k'$ (that is within traveling distance of location $k$) whom we plan to

vaccinate through direct outreach trips to location $k$. We assume that patients in location $k'$ that need to travel to location $k$ for vaccination are notified in advance.

The second set of constraints, which enforce outreach operational constraints, are as follows:

$$\sum_{j\in\hat{S}_i\setminus\{i\}} \check{y}_{ijl} \leq \check{\theta}\bar{x}_{il}, \qquad i=1,2,\ldots,n, \quad l\in\{M,F\}, \tag{4.2a}$$

$$\sum_{j\in\hat{S}_i\setminus\{i\}} \hat{y}_{ijl} \leq \big(\hat{\theta}-\check{\theta}\big)\bar{x}_{il}, \qquad i=1,2,\ldots,n, \quad l\in\{M,F\}, \tag{4.2b}$$

$$\sum_{l\in\{M,F\}} \bar{x}_{il} \leq x_i, \qquad i=1,2,\ldots,n, \tag{4.2c}$$

$$\sum_{k'\in S_k\setminus\{k\}} s_{k'k} \leq (1-\check{x}_k)\rho_k, \qquad k=1,2,\ldots,n, \tag{4.2d}$$

$$\sum_{k'\in S_k\setminus\{k\}} s_{kk'} \leq (1-\hat{x}_k)\sum_{k'\in S_k}\rho_{k'}, \qquad k=1,2,\ldots,n, \tag{4.2e}$$

where $\bar{x}_{il}$ is a zero-one variable that determines the type of outreach vehicle ($l$) in location $i$. Constraints (4.2a) and (4.2b) restrict the total number of outreach trips per clinic to $\hat{\theta}$ while bounding low-cost outreach trips by $\check{\theta}$. Constraints (4.2a)-(4.2c) determine the type of outreach vehicle in location $i$. Constraint (4.2d) prevents any coverage resulting from indirect outreach trips to locations within traveling distance of at least one clinic. Finally, constraint (4.2e) prevents outreach to locations covered by a traveling-distance clinic.

Next, constraint (4.3) guarantees full coverage in all locations as follows:

$$\sum_{i\in\hat{S}_k\setminus\{k\}}\sum_{l\in\{M,F\}} \delta_l\big(\check{y}_{ikl}+\hat{y}_{ikl}\big) + \sum_{k'\in S_k\setminus\{k\}}\Big(s_{k'k}-s_{kk'}\Big) + \rho_k\Big(x_k+\hat{x}_k\Big) \geq \rho_k, \quad k=1,\ldots,n. \tag{4.3}$$

Constraint (4.3) forces the model to cover demand in location $k$ by direct outreach trips, indirect outreach trips, an on-site clinic or a traveling-distance clinic. The first term on the left hand side of constraint (4.3) corresponds to the maximum number of vaccinations that can be performed through direct outreach trips to location $k$. The second term assigns the extra outreach capacity in location $k$ to other locations and vice versa. Finally, the last term on the left hand side inactivates constraint (4.3) if either a clinic is located in location $k$ or

its patients travel to a traveling-distance clinic; that is, no outreach is required to satisfy the demand in location $k$.

To restrict patients in locations to which outreach trips are made from seeking vaccination elsewhere (when an adequate number of doses is available), we add the following constraint:

$$\sum_{k' \in S_k \setminus \{k\}} s_{kk'} \leq \max \Big\{ \sum_{i \in \hat{S}_k \setminus \{k\}} \sum_{l \in \{M,F\}} \delta_l \big( \check{y}_{ikl} + \hat{y}_{ikl} \big) - \rho_k (1 - x_k - \hat{x}_k), 0 \Big\},$$

where the first term on the right-hand side inside the maximization corresponds to the total direct outreach capacity for location $k$ and the second term is associated with the total demand in location $k$ which is not covered by an on-site or a traveling-distance clinic. This constraint is linearized using auxiliary binary variables $\beta_k, k = 1, 2 \ldots, n$ and real-valued variables $\hat{w}_k, k = 1, 2 \ldots, n$ as follows:

$$\sum_{i \in \hat{S}_k \setminus \{k\}} \sum_{l \in \{M,F\}} \delta_l \big( \check{y}_{ikl} + \hat{y}_{ikl} \big) - \rho_k (1 - x_k - \hat{x}_k) = w_k - \hat{w}_k, \qquad k = 1, \ldots, n, \tag{4.4a}$$

$$\sum_{k' \in S_k \setminus \{k\}} s_{kk'} \leq w_k, \qquad k = 1, \ldots, n, \tag{4.4b}$$

$$w_k \leq \beta_k \sum_{k'=1}^{n} \rho_{k'}, \qquad k = 1, \ldots, n, \tag{4.4c}$$

$$\hat{w}_k \leq (1 - \beta_k) \rho_k, \qquad k = 1, \ldots, n. \tag{4.4d}$$

We assume that clinic operation costs are a function of the number of patients served in the clinic. Initially, increasing the number of walk-in patients does not considerably change clinic operation costs. However, clinic operation costs increase if the number of walk-in patients grows significantly such that the clinic needs to acquire more resources. We assume that clinic operation cost is initially fixed until the number of walk-in patients hits a specific level and then increases linearly as a function of additional walk-in patients. Let $s$ be the threshold for the number of walk-in patients after which the clinic operation cost increases and $z_i$ be the number of additional walk-in patients vaccinated in the clinic in location $i$. The following constraint determines the value of $z_i$ :

$$z_i \geq \rho_i x_i + \sum_{k \in S_i \setminus \{i\}} \rho_k \kappa_{ik} - s, \qquad i = 1, \ldots, n. \tag{4.5}$$

With simple modifications in constraint (4.5), we can model any kind of piecewise linear function representing clinic operation costs.

Let $c$ be the fixed cost of clinic operation over the planning horizon, $\tilde{c}$ be the cost of clinic operation per patient when the number of walk-in patients exceeds the existing immunization resource, and $\lambda$ be the patient travel cost per unit of distance. We assume clinics are not built from scratch; rather immunization activities are added to existing clinic functions and the fixed costs correspond to extra resources required for these activities (e.g., additional equipment, staff, transport, etc.). We also assume these fixed costs are identical at all clinics which can easily be relaxed by considering different fixed cost across locations. Let $\check{c}_{ijl}$ represent the cost of an outreach trip for the first $\check{\theta}$ outreach trips and $\hat{c}_{ijl}$ represent the analogous cost after $\check{\theta}$ outreach trips. Hence, the entire model formulation can be expressed as:

$$\min z = \sum_{i=1}^{n} \left(cx_i + \tilde{c}z_i\right) + \sum_{i=1}^{n} \sum_{j \in \hat{S}_i \backslash \{i\}} \sum_{l \in \{M,F\}} \left(\check{c}_{ijl}\check{y}_{ijl} + \hat{c}_{ijl}\hat{y}_{ijl}\right) + \sum_{k=1}^{n} 2\lambda \left(\rho_k \hat{d}_k + \sum_{k' \in S_k \backslash \{k\}} d_{k'k} s_{k'k}\right)$$

s.t. Constraints (4.1a)-(4.1g)

Constraints (4.2a)-(4.2e)

Constraint (4.3)

Constraints (4.4a)-(4.4d)

Constraint (4.5)

$$x_i, \hat{x}_i, \check{x}_i, \beta_i, \gamma_i, \kappa_{ij} \in \{0,1\}, \qquad i = 1, \ldots, n, \quad j \in S_i \backslash \{i\},$$

$$\hat{d}_i, y_{ijl}, s_{ij}, z_i \in \mathbf{N}, \qquad i = 1, \ldots, n, \quad j \in \hat{S}_i \backslash \{i\}, \quad k \in S_j \backslash \{i,j\}, \quad l \in \{M,F\},$$

$$w_k, \hat{w}_k \geq 0, \qquad k = 1, \ldots, n, \quad k' \in S_k \backslash \{k\}.$$

The first and second terms in the objective function correspond to the cost of clinic operation and outreach trips, respectively. The last term is associated with the patient travel cost to either a traveling-distance clinic or outreach trip destination. In this model, we assume identical patient travel cost for all locations, but this assumption can easily be relaxed. For

example, we can set a higher value of $\lambda$ for harder to access locations; consequently, those locations would be more probable to be a clinic location or an outreach destination.

### 4.2.2 Demand Uncertainty

The model formulated in Section 4.2.1 is a deterministic optimization model. It generates an optimal solution which specifies clinic locations and outreach trips to cover a known set of demands. However, demand is not always known and its estimation involves uncertainty. Hence, the model may generate suboptimal or infeasible solutions. In this section, we proposed a method to incorporate demand uncertainty into the optimization model.

The point estimate of demand at each location, $\rho_i$, appears in many constraints in the optimization model. To capture uncertainty in these estimates we divide the constraints into two groups. The first group includes constraints (4.2d), (4.2e), (4.4c) and (4.4d) in which we set $\rho_i$ to be sufficiently large such that these four constraints are never tight. Therefore, we set $\rho_i$ to an upper bound and this does not impact solution optimality and feasibility.

The second group of constraints includes constraints (4.3), (4.4a) and (4.5), in which underestimating demand can result in low coverage and/or larger distances traveled by patients, whereas overestimating demand can increase clinic operation and outreach costs. Let the demand in location $k$ be a positive random variable $V_k$ which follows a log-normal distribution similar to [30] with parameters $\mu_k$ and $\sigma_k$, i.e., $V_k = e^{\mu_k + Z\sigma_k}$ where $Z \sim N(0,1)$. Furthermore, let $\rho_k^*(\tau)$ be such that $\Pr\left(V_k \geq \rho_k^*(\tau)\right) = \tau$. That is, $\rho_k^*(\tau)$ is the $100(1-\tau)$th percentile of the demand distribution in location $k$. We substitute $\rho^*(\tau)$ for $\rho$ in constraints (4.3), (4.4a) and (4.5) and we can control the coverage by determining the appropriate value of $\tau$. Given that the demand in location $k$, $V_k$, follows a log-normal distribution with parameters $\mu_k$ and $\sigma_k$, $\rho_k^*(\tau)$ is given by:

$$\rho_k^*(\tau) = e^{\mu_k + \Phi^{-1}(1-\tau)\sigma_k}. \tag{4.7}$$

Based on Assumption (2), in locations covered with an on-site clinic or a walking-distance clinic, the entire demand is guaranteed to be vaccinated. Based on constraint (4.3), in locations covered by direct and/or indirect outreach trips, at most $\rho_k^*(\tau)$ patients can be

vaccinated due to the constraint on the size of outreach trips. Therefore, let $\varphi_k(\tau)$ be the coverage rate (i.e., the fraction of patients who are vaccinated) in location $k$ given that we vaccinate at most $\rho_k^*(\tau)$ patients. We can compute the expected value of $\varphi_k(\tau)$ as follows

$$\begin{aligned}
\mathrm{E}[\varphi_k(\tau)] &= 1 - \mathrm{E}\Big[\frac{(V_k - \rho_k^*(\tau))^+}{V_k}\Big] && \text{(4.8a)}\\
&= 1 - \Pr\Big(V_k \geq \rho_k^*(\tau)\Big)\mathrm{E}\Big[1 - \frac{\rho_k^*(\tau)}{V_k}\Big|V_k \geq \rho_k^*(\tau)\Big] && \text{(4.8b)}\\
&= 1 - \tau\left(1 - \rho_k^*(\tau)\mathrm{E}\Big[\frac{1}{V_k}\Big|\frac{1}{V_k} \leq \rho_k^{*-1}(\tau)\Big]\right) && \text{(4.8c)}\\
&= 1 - \tau + \tau\rho_k^*(\tau)e^{-\mu_k+\sigma_k^2/2}\frac{\Phi\Big(\frac{\ln\left(\rho_k^{*-1}(\tau)\right)+\mu_k-\sigma_k^2}{\sigma_k}\Big)}{\Phi\Big(\frac{\ln\left(\rho_k^{*-1}(\tau)\right)+\mu_k}{\sigma_k}\Big)} && \text{(4.8d)}\\
&= 1 - \tau + \tau\rho_k^*(\tau)e^{-\mu_k+\sigma_k^2/2}\frac{\Phi\Big(-\Phi(1-\tau)-\sigma_k\Big)}{\Phi\big(-\Phi(1-\tau)\big)} && \text{(4.8e)}\\
&= 1 - \tau + e^{\sigma_k\left(\Phi^{-1}(1-\tau)+\sigma_k/2\right)}\left(1 - \Phi\Big(\Phi^{-1}(1-\tau)+\sigma_k\Big)\right), && \text{(4.8f)}
\end{aligned}$$

where Equation (4.8b) follows from the law of total expectation and the fact that $\mathrm{E}\big[\frac{(V_k-\rho_k^*(\tau))^+}{V_k}|V_k < \rho_k^*(\tau)\big] = 0$; Equation (4.8d) follows from the fact that, if $V \sim \ln N(\mu,\sigma^2)$, then $\frac{1}{V_k} \sim \ln N(-\mu,\sigma^2)$ and $\mathrm{E}\Big[V\Big|V \leq v\Big] = \frac{\Phi\left(\frac{\ln(v)-\mu-\sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln(v)-\mu}{\sigma}\right)}e^{\mu+\sigma^2/2}$; and Equation (4.8f) follows from $\Phi\big(-\Phi^{-1}(\alpha)\big) = 1-\alpha$ and Equation (4.7). Interestingly, the expected coverage rate in location $k$ only depends on the values of $\tau$ and $\sigma_k$.

Moreover, let $\varphi_k^\alpha(\tau)$ be the $100\alpha$th percentile of the coverage rate in location $k$ given that we vaccinate at most $\rho_k^*(\tau)$ patients. We can compute the value of $\varphi_k^\alpha(\tau)$ as follows:

$$\varphi_k^\alpha(\tau) = \frac{\rho_k^*(\tau)}{\rho_k^*(\alpha)} = e^{\sigma_k\left(\Phi^{-1}(1-\tau)-\Phi^{-1}(1-\alpha)\right)}, \tag{4.9}$$

which follows from the fact that for $0 \leq \varepsilon \leq 1$,

$$\begin{aligned}
\Pr(\varphi_k(\tau) \geq \varepsilon) &= \Pr\left(\frac{(V_k - \rho_k^*(\tau))^+}{V_k} \leq 1-\varepsilon\right) && \text{(4.10a)}\\
&= \Pr\left(\frac{V_k - \rho_k^*(\tau)}{V_k} \leq 1-\varepsilon\middle|V_k \geq \rho_k^*(\tau)\right)\Pr\Big(V_k \geq \rho_k^*(\tau)\Big) + \Pr\Big(V_k < \rho_k^*(\tau)\Big) && \text{(4.10b)}\\
&= \Pr\left(V_k \leq \frac{\rho_k^*(\tau)}{\varepsilon}\middle|V_k \geq \rho_k^*(\tau)\right)\tau + (1-\tau) && \text{(4.10c)}
\end{aligned}$$

$$= \frac{\Pr\left(\rho_k^*(\tau) \leq V_k \leq \frac{\rho_k^*(\tau)}{\varepsilon}\right)}{\Pr\left(V_k \geq \rho_k^*(\tau)\right)}\tau + \left(1-\tau\right) \tag{4.10d}$$

$$= \left(1 - \frac{\Pr\left(V_k \geq \frac{\rho_k^*(\tau)}{\varepsilon}\right)}{\tau}\right)\tau + \left(1-\tau\right) \tag{4.10e}$$

$$= \Pr\left(V_k < \frac{\rho_k^*(\tau)}{\varepsilon}\right) = \Phi\Big(\frac{\ln\left(\frac{\rho_k^*(\tau)}{\varepsilon}\right) - \mu_k}{\sigma_k}\Big). \tag{4.10f}$$

Equation (4.10b) follows from the law of total probability; and Equation (4.10f) follows from the fact that, if $V \sim \ln N(\mu, \sigma^2)$, then $\Pr(V < x) = \Phi\Big(\frac{\ln(x)-\mu}{\sigma}\Big)$. Similar to the expected value in Equation (4.8f), the percentile of the coverage rate in location $k$ only depends on the values of $\tau$, $\alpha$ and $\sigma_k$.

## 4.3 NUMERICAL EXAMPLE

In this section, we present a small numerical example and explore its corresponding optimal solution generated by the MILP model. To evaluate its performance, we develop a simulation model which simulates the vaccination demand in each location according to the underlying demand distributions discussed in Section 4.2.2. Consider a problem instance that consists of 15 locations with the coordinates, population mean, and variance reported in Table 9. All of the locations are potential clinic locations as well as outreach destinations and they are connected through a road network. We compute the travel distance between two locations connected by a road using a Euclidean distance metric and set $d_{i,j}$ to the shortest travel distance between location $i$ and $j$ considering the existing road network.

Over a 6 month time horizon, let the fixed cost of clinic operation, $c$, be \$500 for the capacity to see at most 3600 walk-in patients ($s$ = 6 months $\times$20 days/month $\times$30 patients/day) and the variable cost per extra patient, $\tilde{c}$, be \$0.5/patient. Let $\lambda$ = \$0.02/km, $\Delta$ = 10 km, $\hat{\Delta}$ = 5 km, $\check{\theta}$ = 24 outreach trips, $\hat{\theta}$ = 120 outreach trips, $\delta_M$ = 25 vaccinations/outreach trip and $\delta_F$ = 100 vaccinations/outreach trip. Let $\tau = 0.5$ and subsequently, $\rho_k = \rho_k^*(0.5)$ patients for each location $k$.

Table 9: Location specifications

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mu_k$ | 400 | 100 | 300 | 200 | 100 | 200 | 350 | 150 | 200 | 500 | 100 | 300 | 250 | 100 | 200 |
| $\sigma^2{}_k$ | 66.7 | 16.7 | 50.0 | 33.3 | 16.7 | 33.3 | 58.3 | 25.0 | 33.3 | 83.3 | 16.7 | 50.0 | 41.7 | 16.7 | 33.3 |
| Coordinates | (1,2) | (1,6) | (3,4) | (9,1) | (3,1) | (3,7) | (5,9) | (5,11) | (7,8) | (12,6) | (12,3) | (12,13) | (16,4) | (17,14) | (18,10) |

One approach to compute the outreach cost is to decompose it into a fixed and a variable cost which are functions of the number of outreach trips and the vehicle type used, respectively. Let the fixed cost of an outreach trip be $\check{c}^f$ for the first $\check{\theta}$ outreach trips and then let it increase to $\hat{c}^f$. This increase in fixed costs is due to the need for additional resources such as the cost of hiring additional staff. Let the variable cost per unit of distance traveled by a vehicle of type $l$ ($\in$ {F,M}) be $c_l^v$. Consequently, $\check{c}_{ijl} = \check{c}^f + 2c_l^v d_{ij}$ and $\hat{c}_{ijl} = \hat{c}^f + 2c_l^v d_{ij}$. For further clarification, in Figure 34, we plot the different types of outreach trip costs according to the proposed structure.

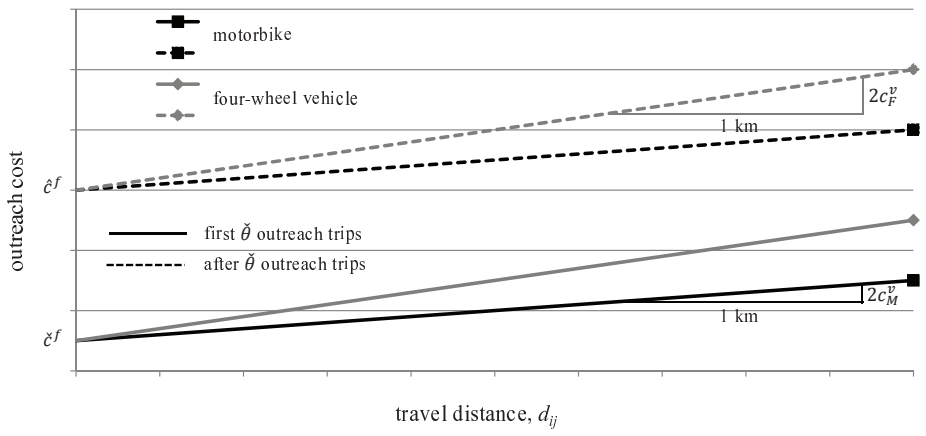



Figure 34: Outreach trip cost structure.

In the numerical example, we let the fixed cost of an outreach trip for the first 24 outreach trips be \$10 which increases to \$30 per outreach trip after 24 outreach trips. Finally, we set the variable cost of motorbikes and four-wheel vehicles, $c_M^v$ and $c_F^v$, to be \$0.15/km and \$0.5/km, respectively.

Figure 35 demonstrates the optimal clinic locations and outreach trips to cover the entire region. In Figure 35, we set the node size proportional to the population mean and use straight lines and curved lines to draw the network of roads and outreach trips, respectively. As seen in Figure 35, in the optimal solution we put clinics in locations 3 and 15 and assign four-wheel vehicles to both clinics. The clinic in location 3 performs 6 outreach trips to location 10, 4 outreach trips each to locations 1 and 7, 2 outreach trips each to locations 4 and 9, and one outreach trip to location 8. Outreach trips to location 8 partially cover the location and its remaining patients travel to location 7. Furthermore, outreach trips to location 10 simultaneously cover locations 10 and 11. The clinic in location 3 directly covers locations 2, 5 and 6 which are within traveling distance of location 3. The clinic in location 15 performs 3 outreach trips each to locations 12 and 13 and one outreach trip to location 14.

Table 10 summarizes the performance of the optimal solution considering metrics such as the expected coverage rate (i.e., the expected fraction of vaccinated patients) and the expected travel distance per walk-in patient using the simulation model. Even though we set $\tau = 0.5$ (i.e., $\rho_k^*(0.5) = e^{\mu_k}, \forall k$), the optimal solution exceeds a 95% expected coverage rate and 15.7% of all patients travel an expected distance of 5.8 km to the vaccination site.

## 4.4 SENSITIVITY ANALYSIS

In this section, we perform sensitivity analysis on key problem parameters using the problem example in Section 4.3. In Section 4.4.1, we investigate the relationship between $\tau$ and the performance of the optimal solution, and discuss determining an appropriate value for $\tau$. In Sections 4.4.2 and 4.4.3, we explore how the values of $\lambda$, i.e., the patient travel cost or

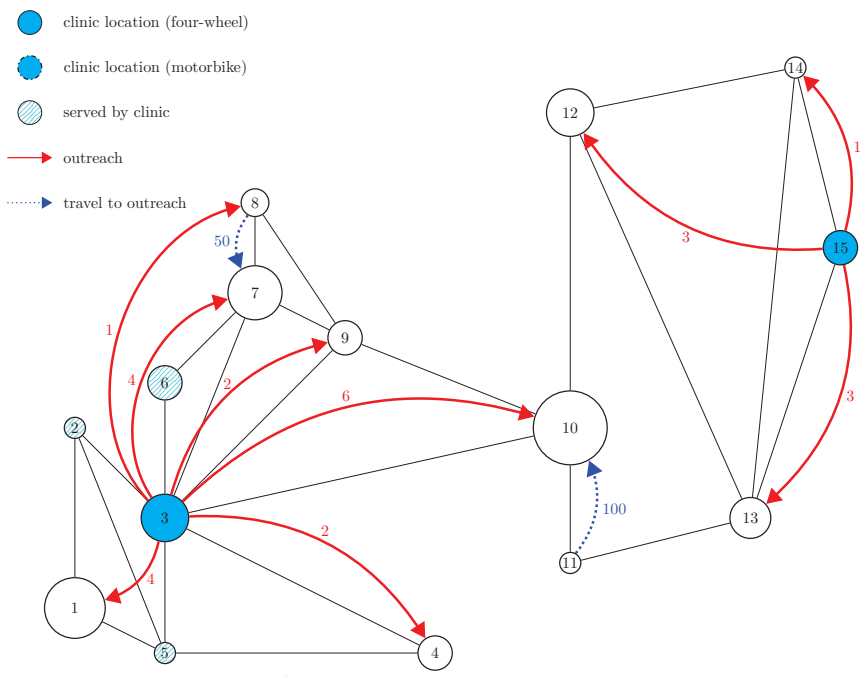



Figure 35: Optimal clinic locations and outreach trips for the numerical example ($\tau = 0.5$).

Table 10: Summary of the numerical example performance ($\tau = 0.5$).

| Expected number of vaccinations | Expected coverage rate | Expected fraction of patients who travel | Expected total travel distance | Expected travel distance per traveling patient |
|---|---|---|---|---|
| 3295 | 95.6% | 15.7% | 3135 | 5.8 |

willingness to travel, and $\delta_F$, i.e., the number of vaccinations per outreach trip, impact the cost of clinic operations, outreach and patient travel.

### 4.4.1 Probability of Underestimating Demand, $\tau$

The smaller the value of $\tau$, the more conservative the solution and hence the higher the expected coverage. However, decreasing the value of $\tau$ increases the number of clinics and outreach trips, and consequently requires more resources. In this section, we compute the

expected coverage rate and percentiles derived in Section 4.2.2 as a function of $\tau$ in order to identify an appropriate value for $\tau$.

As $\tau$ decreases, $\rho_k^*$ increases and consequently, under the optimal solution more vaccination opportunities are provided by clinics in locations which are covered by direct and/or indirect outreach trips. However, the appropriate value of $\tau$ depends on the extent of improvement in the coverage rate, which can be explored through the expected value and percentiles in Figure 36. The expected coverage rate and percentiles are equal for all locations because in the numerical example we set $\sigma_k = 0.1655$ for all $k$. Based on Figure 36, we consider values of $\tau = 0.5$ and $\tau = 0.1$ and suggest picking $\tau \in \{0.1, 0.5\}$ based on how conservative we wish the solution to be. The value of $\tau = 0.5$ corresponds to the median which results in an expected coverage rate of 94% at individual locations while its corresponding 1st and 5th percentiles are 76% and 68%, respectively. Likewise, for $\tau = 0.1$, the expected coverage rate for each location individually is 99.3%; and its 1st and 5th percentiles are 94% and 84%, respectively. Both values of $\tau$ result in reasonably high expected coverage, while the variation around the coverage rate is lower for $\tau = 0.1$, i.e., the percentiles are closer to the expected coverage rate.

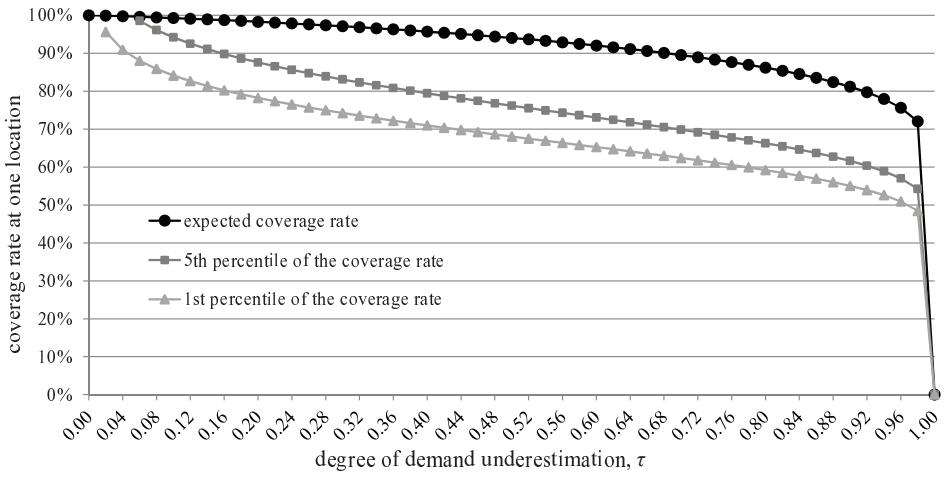



Figure 36: Expected coverage rate and 1st and 5th percentiles of the coverage rate as a function of $\tau$.

### 4.4.2 Patient Travel Cost, $\lambda$

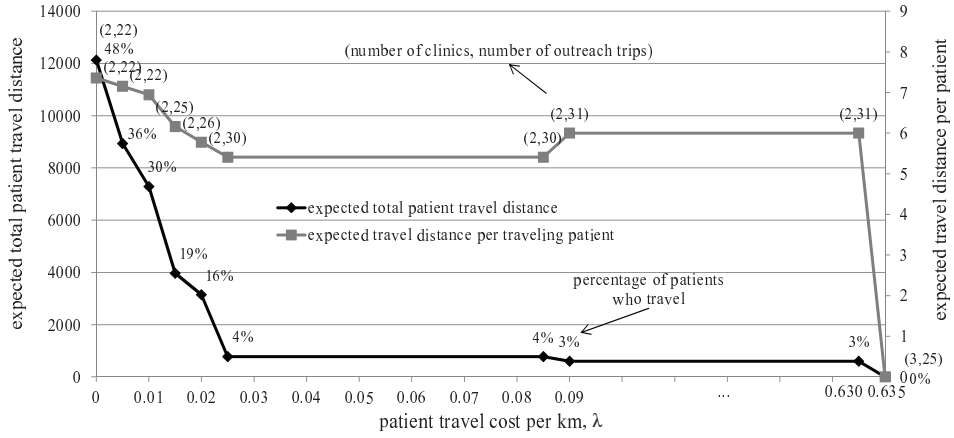
The parameter $\lambda$ prevents the model from generating a solution that imposes a major travel burden on patients. In particular, $\lambda$ specifies the amount that patients are willing to pay per unit of distance traveled to a vaccination site. We can also interpret $\lambda$ as the patients' willingness to travel. High values imply a reduced willingness to travel. In this section, we investigate the impact of $\lambda$ on the optimal solution.

Figure 37 shows the expected total travel distance and travel distance per patient (among traveling patients) as a function of $\lambda$ for $\tau = 0.5$ and 0.1. As expected, larger values of $\lambda$ result in smaller values of the expected total travel distance to avoid the high cost associated with patient travel. This decrease is the result of relocating clinics and outreach trip destinations, increasing the number of outreach trips and in some cases, increasing the number of clinics. For example, in Figure 37(b), by increasing $\lambda$ above \$0.27/km, the number of clinics increases from 2 to 3 while the number of outreach trips decreases from 41 to 35.

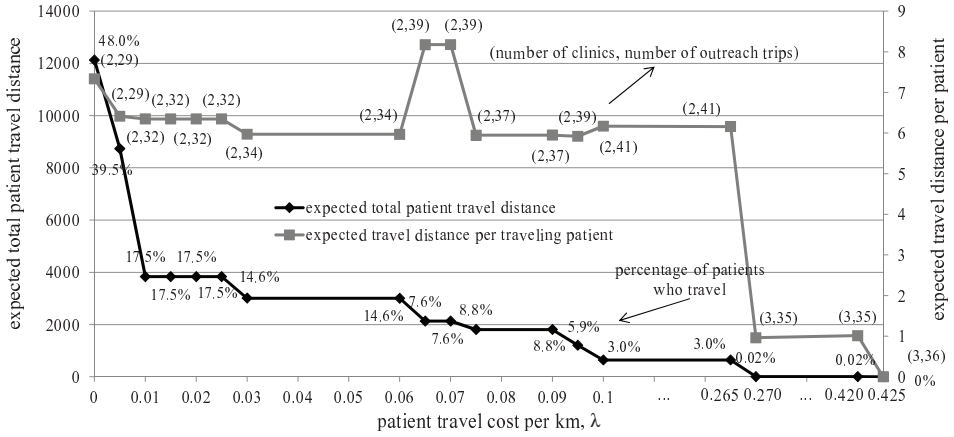
There is a threshold on the patient travel cost above which we reach an optimal solution with no traveling patients. Under this solution, vaccinations are provided at all locations either by an on-site clinic or by performing direct outreach trips. For example, in Figure 37(a), in the optimal solution for $\lambda \geq$ \$0.635/km, the entire population is covered by establishing 3 clinics at locations $3, 10$, and 12, and performing 25 direct outreach trips to the remaining locations.

In addition, the expected travel distance per patient generally decreases as $\lambda$ increases. However, there are exceptions. For example, in Figure 37(b), as $\lambda$ increases from \$0.060/km to \$0.065/km, the expected travel distance per patient increases from 6.0 km/patient to 8.2 km/patient. This increase is the result of relocating clinics and replanning outreach trips so that a smaller fraction of patients have to travel, but the travel distances are longer on average. For example, as seen in Figure 37(b), the fraction of traveling patients drops from 14.6% to 7.6% as $\lambda$ increases from \$0.06/km to \$0.065/km.

Figure 38 plots the clinic operations and outreach costs as well as the expected patient traveling cost as a function of $\lambda$. As $\lambda$ increases, the optimal solution reduces the distance traveled by patients. This reduction is achieved by establishing more clinics and performing




(a) $\tau = 0.5$.

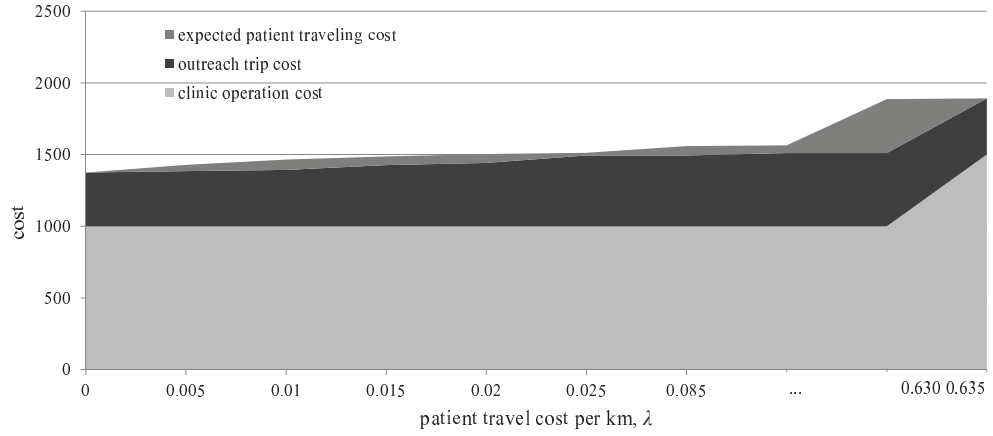



(b) $\tau = 0.1$.

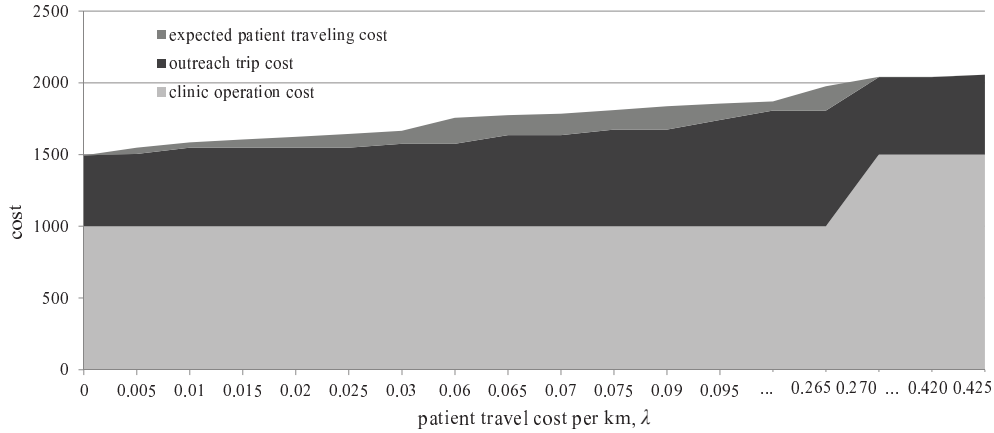
Figure 37: Expected total patient travel distance and expected travel distance per traveling patient as a function of $\lambda$.

more, and longer, outreach trips. Consequently, the total costs of clinics and outreach increase.

Furthermore, as seen in Figure 37, increasing $\lambda$ from 0 by a small amount considerably reduces the expected total patient travel distance, while Figure 38 shows that the total cost of clinics and outreach does not increase significantly. For example, in Figure 38(a), increasing $\lambda$ from 0 to 0.025 decreases the expected patients travel distance from 12,147 km to 773 km (a 94% decrease) but the cost increases from \$1,375 to \$1,512 (an increase of only 10%).




(a) $\tau = 0.5$.




(b) $\tau = 0.1$.

Figure 38: Cost of clinics, outreach and travel as a function of $\lambda$.

Similarly, in Figure 38(b), increasing $\lambda$ from 0 to 0.095 decreases the expected patients travel distance from 1,2132 km to 1,208 km (a 90% decrease) but the cost increases from \$1,494 to only \$1,855 (a 24% increase). Consequently, in cases under which decreasing the clinic operations and outreach cots is our first priority, by setting $\lambda$ to a very small positive value, we can achieve a more practical solution.

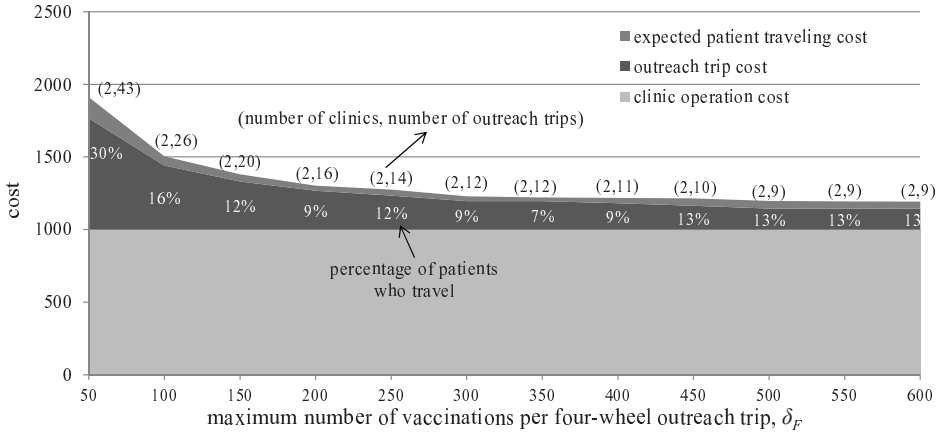
### 4.4.3 Maximum Outreach Trip Size

In this section, we perform sensitivity analysis on $\delta_F$ which provides insight on the impact of cold-box capacity. While the capacity of a motorbike is limited, four-wheel drive vehicles are capable of carrying significantly larger volumes. Determining an appropriate capacity for these vehicles, $\delta_F$, that is, the size of the cold box, is of importance. A small value of $\delta_F$ can increase the total distance traveled by patients or lead to numerous outreach trips. On the other hand, a large value of $\delta_F$ might increase the costs of equipment (i.e., cold boxes) and maintenance; it can also result in a higher risk of vaccine wastage as a consequence of cold-box failure.

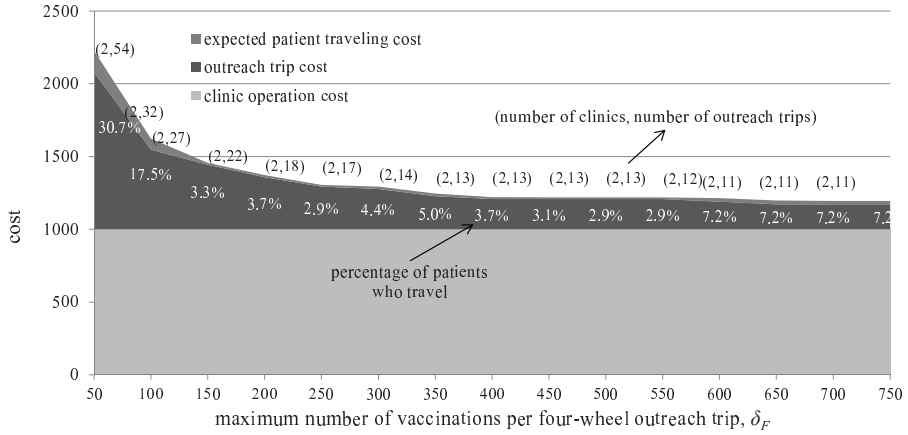
Figure 39 demonstrates the impact of $\delta_F$ on the cost of clinic operations, outreach and patient travel, and indicates the corresponding number of clinics and outreach trips. As seen in Figure 39, initially, a slight increase in the value of $\delta_F$ decreases total cost considerably and also decreases the required number of outreach trips. For example, in Figure 39(b), if we increase $\delta_F$ from 50 to 150, the required number of outreach trips decreases from 43 to 20 and the total cost decreases from \$1,911 to \$1,380. Similarly, in Figure 39(b) the required number of outreach trips decreases from 54 to 27 and the total cost decreases from \$2,226 to \$1,457. However, larger values of $\delta_F$ have less influence because at least one outreach trip is required to vaccinate each remote location. For example, in Figure 39(a), the costs and the number of outreach trips do not change significantly for values of $\delta_F$ above 350. Moreover, in this example, $\delta_F$ does not influence the optimal number of clinic and therefore, the clinic cost.

## 4.5 HEURISTIC COMPARISONS

In this section, we compare the performance of the MILP model proposed in Section 4.2.1 with two benchmark heuristic approaches, H1 and H2. An intuitive approach to locating clinics is to place them in highly populated centers to minimize patient travel and perform less outreach. Hence, in the first heuristic, H1, we propose a simple algorithm that prioritizes




(a) $\tau = 0.5$.




(b) $\tau = 0.1$.

Figure 39: Cost of clinics, outreach and travel as a function of $\delta_F$.

locating clinics in highly populated locations. To determine how much better we perform by simultaneously locating clinics and specifying outreach trips, we propose the second heuristic approach. The second approach, H2, is a combination of a linear integer programming model (which determines the minimum number of clinics required to cover all locations) and an outreach planning algorithm that first determines the clinic locations and then plans outreach trips.

In the first heuristic approach, H1, we follow the steps below:

**Step 1:** Place a clinic in the most populous uncovered location.

**Step 2:** Cover all locations within its traveling distance and remove any outreach trips from other clinics to destinations within traveling distance of the most recently added clinic.

**Step 3:** Assign outreach trips to uncovered locations starting from the closest location not within traveling distance of the most recently added clinic.

**Step 4:** Fully cover as many locations as possible through outreach trips subject to constraints on the outreach distance, size (minimum outreach size, $\delta_M$) and frequency ($\hat{\theta}$).

**Step 5:** Repeat Steps 1-4 if there is any uncovered location, otherwise, go to Step 6.

**Step 6:** Determine the outreach vehicle type at each clinic location based on the number of patients covered by outreach from each clinic.

**Step 6.a** If in a clinic the population covered through outreach is above a specific level ($\leq \hat{\theta}\delta_M$)), e.g., $\check{\theta}\delta_M$, then assign a four-wheel vehicle to that clinic location, otherwise, a motorbike.

**Step 6.b** If a four-wheel vehicle is assigned to a clinic then recalculate its corresponding number of outreach trips.

Figure 40 illustrates the heuristic solution for the numerical example in Section 4.3. For the H1 solution, clinics are located at locations 1, 10 and 15 and the remaining locations are covered by performing a total of 33 outreach trips. Motorbikes are assigned to locations 1 and 15 and a four-wheel vehicle is assigned to location 10. In the optimal solution only 2 clinics with four-wheel vehicles are located in locations 3 and 15 and the set of all locations is covered by a smaller number of outreach trips. The optimization model generates a more effective solution than the H1 approach by locating a clinic in location 3. Specifically, the clinic in location 3 covers all locations which are covered by 1 and 10 in the H1 solution except location 13 which is within the outreach distance of the clinic in location 15.

For the second approach, H2, instead of simultaneously determining clinic locations and planning outreach trips, they are done sequentially. First we solve a simple optimization problem to specify clinic locations and then determine vehicle type, outreach destinations and frequency based on these clinic locations. Let $v_{ij}$ be a zero-one variable which equals 1 if a clinic is assigned to location $i$ and covers location $j$ and 0 otherwise and $h_{ij}$ be the required number of outreach trips from location $i$ to location $j$. The following set covering

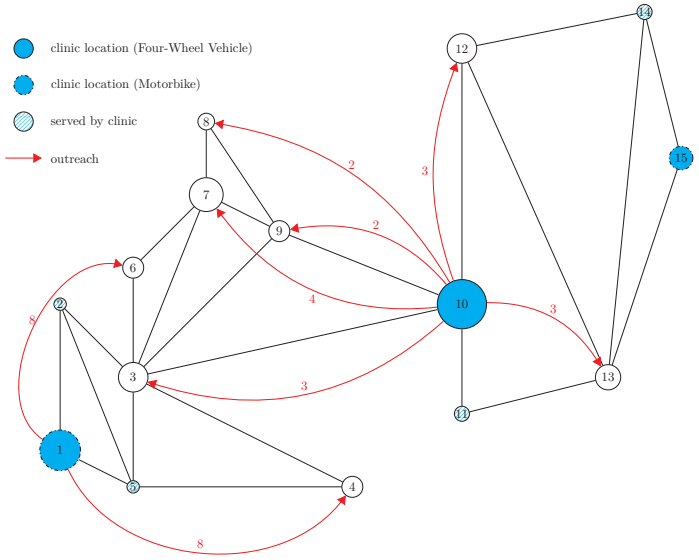



Figure 40: Clinic locations and outreach trips based on the H1 approach.

model determines the clinic locations that minimize the number of clinics needed to cover all locations through travel-distance proximity and outreach:

$$\min z = \sum_{i=1}^{n} v_{ii}$$

$$\text{s.t.} \quad v_{ii} + \sum_{j \in \hat{S}_i \backslash \{i\}} v_{ji} = 1, \qquad i = 1, 2, \ldots, n, \qquad (4.11a)$$

$$\sum_{j \in \hat{S}_i \backslash \{i\}} v_{ij} \leq n v_{ii}, \qquad i = 1, 2, \ldots, n, \qquad (4.11b)$$

$$\rho_j v_{ij} \leq \delta_F h_{ij}, \qquad i = 1, 2, \ldots, n, \quad j \in \hat{S}_i \backslash S_i, \qquad (4.11c)$$

$$\sum_{j \in \hat{S}_i \backslash S_i} h_{ij} \leq \hat{\theta}, \qquad i = 1, 2, \ldots, n, \qquad (4.11d)$$

$$v_{ij} \in 0, 1, \quad i, j = 1, \ldots, n,$$

$$h_{ij} \in \mathbf{N}, \quad i = 1, 2, \ldots, n, \quad j \in \hat{S}_i \backslash S_i.$$

Constraint (4.11a) guarantees coverage through either an on-site clinic or a clinic within outreach distance. Constraint (4.11b) requires that we cover locations within outreach distance

of location $i$ if there is a clinic in location $i$, i.e., $v_{ii} = 1$. Constraint (4.11c) determines the required number of outreach trips to location $j$ and finally constraint (4.11d) restricts the outreach capacity to be $\hat{\theta}$ in location $i$.

In the second step of H2, if a location is within traveling distance of at least one clinic, its patients travel to the closest clinic; otherwise, the location is covered via outreach by the clinic to which it is assigned by the optimization model. We then determine the outreach vehicle type. Similar to the heuristic approach, we allocate a four-wheel vehicle to a clinic if the population covered by its outreach trips exceeds a specific level, e.g., $\check{\theta}\delta_M$. Otherwise, we assign a motorbike to the clinic. Lastly, we compute the required number of outreach trips to cover each location.

Figure 41 represents the H2 solution for the numerical example in Section 4.3. In the H2 solution, 3 clinics are located at locations 3, 10 and 15 and the remaining locations are covered by performing a total of 38 outreach trips. Motorbikes are assigned to locations 3 and 15 and a four-wheel vehicle is assigned to location 10. Although, the H2 approach minimizes the number clinics, it locates one more clinic than the optimal solution (Figure 35). Higher number of clinics in the H2 solution is the result of restricting coverage to an on-site clinic, a traveling-distance clinic or direct outreach trips. Specifically, in the H2 solution, clinics in locations 3 and 15 can cover all locations except location 11 which is not in their outreach distance. Therefore, t locates three clinics. While, the optimal solution instinctively cover location 11 by indirect outreach trip.

Moreover, in Figure 42, we compare the performance of the optimal solution with the heuristc solutions for $\tau = 0.5$ and 0.1. In contrast to the heuristic approaches, the optimal solution depends on the value of $\lambda$, hence, in Figure 42, we include the optimal solutions corresponding to all possible values of $\lambda$. We find all optimal solutions by gradually increasing $\lambda$ with a step size of 0.05 from 0 up to a solution in which all patients are covered by either an on-site clinic or a direct outreach trip. Interestingly, in Figure 42(a), regardless of the value of $\lambda$, the cost of establishing clinics and performing outreach trips under the optimal solution is lower than that of the heuristic solutions. Even when $\lambda \geq 0.635$, in which case it is optimal for no patient to travel, the optimal solution results in a lower total cost. As expected, we achieve a lower combined cost for clinic and outreach trips than the heuristic

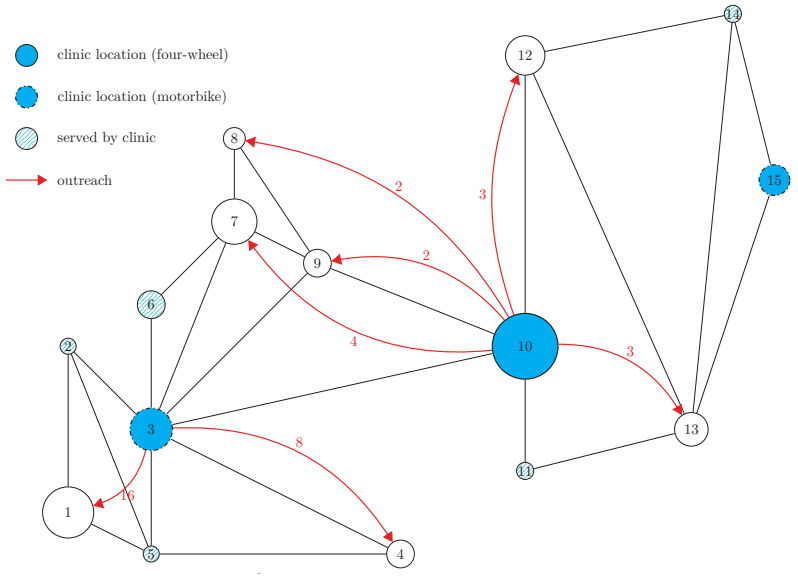


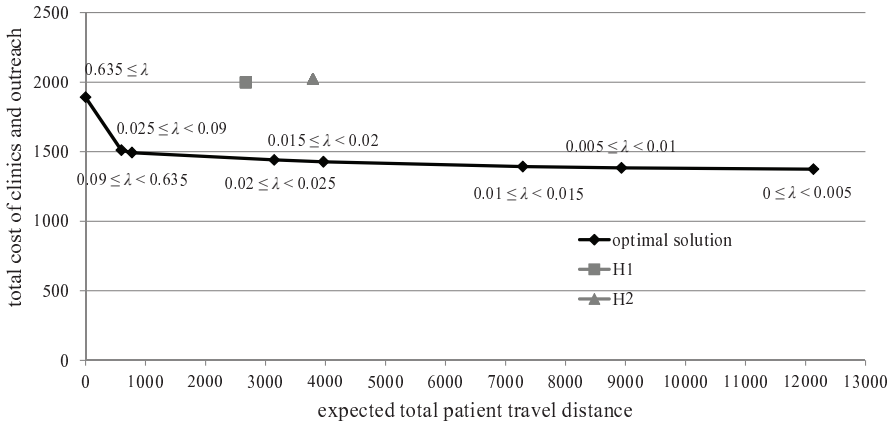

Figure 41: Clinic locations and outreach trips based on the H2 approach.

solutions while achieving almost the same expected patient travel distance. For example, in Figure 42(a), the optimal solution corresponding to $0.015 \leq \lambda \leq 0.02$ has lower cost than the two-step optimization approach with almost the same patient travel distance.

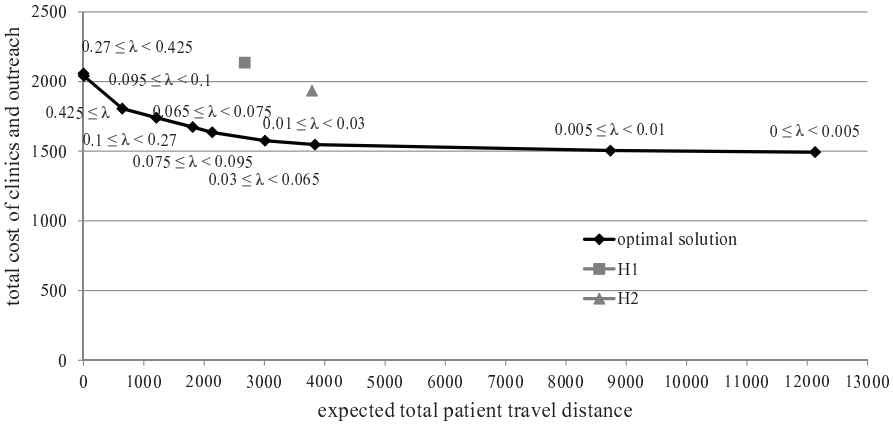
Lastly, the H2 approach can result in higher patient travel distance than the H1 approach because it minimizes the number of clinics required to cover the entire region. While, the H1 approach can result in more clinics in high population locations which can reduce the total patient travel distance. Thus, overall, the proposed methodology outperforms both heuristics and generates different solutions depending on the value placed on patient travel.

## 4.6 CONCLUSIONS AND LIMITATIONS

In this chapter, we address the problem of simultaneously determining clinic locations and outreach trips with the objective of minimizing the total costs of clinic operation, outreach and patient travel. Our model incorporates demand uncertainty and we provide insight on




(a) $\tau = 0.5$.




(b) $\tau = 0.1$.

Figure 42: Cost of clinic and outreach trips versus expected total travel distance.

how to manage the uncertainty. We investigate the impact of key parameters including patient travel cost and outreach trip size on the optimal solution. Our results suggest that modest increases in patient travel cost can substantially improve the performance of the optimal solution from the patient perspective (i.e., travel distance), however, there are marginally diminishing returns. Similarly, initial increases in outreach trip capacity decrease the total cost much more significantly than additional increases. We also found that it is not always best to locate clinics in the largest population centers. Finally, we compare the optimal solution found by our method with those found by two heuristic approaches.

There are limitations which could be modified or considered as possible future research directions. Our model currently determines optimal clinic locations and outreach trips for a blank region (i.e., a region without any vaccination clinics). However, we can simply modify the model to fix a set of pre-existing clinic locations and determine additional clinic locations (and/or possibly close or relocate existing ones) along with determining the optimal set of outreach trips. Second, in the current model we do not consider how population can change over time or seasonal migration patterns. Third, we assume that patients within traveling distance of a clinic are covered by that clinic regardless of the distance. Future research could consider coverage functions that are more explicitly dependent on patient travel distance. Fourth, we cannot draw any strong general conclusions based on our limited numerical experimentation, therefore, future work might include executing a designed experiment over a larger set of problem instances. Finally, future work could explore the solution time for large-scale problems and potential scalability issues in more detail.

# BIBLIOGRAPHY


[1] T.-M. Assi, S.T. Brown, A. Djibo, B.A. Norman, J. Rajgopal, J.S. Welling, S.-I. Chen, R.R. Bailey, S. Kone, H. Kenea, et al. Impact of changing the measles vaccine vial size on Niger's vaccine supply chain: A computational model. *BMC Public Health*, 11(1):425, 2011.

[2] A. Bensoussan, M. Çakanyildirim, and S.P. Sethi. A multiperiod newsvendor problem with partially observed demand. *Mathematics of Operations Research*, 32(2):322–344, 2007.

[3] J.I. Blanford, S. Kumar, W. Luo, and A.M. MacEachren. Its a long, long walk: Accessibility to hospitals, maternity and integrated health centers in Niger. *International Journal of Health Geographics*, 11(1):24, 2012.

[4] Center for Disease Control and Prevention. *Vaccine Management: Recommendations for Handling and Storage of Selected Biologicals*. US Department of Health and Human Services, Public Health Service, Centers for Disease Control, 2007.

[5] J. Current and J. Storbeck. Capacitated covering models. *Environment and Planning B*, 15:153–164, 1988.

[6] M.S. Daskin and L.K. Dean. Location of health care facilities. In *Operations Research and Health Care*, pages 43–76. Springer, 2004.

[7] A. Dhamodharan and R.A. Proano. Determining the optimal vaccine vial size in developing countries: A Monte Carlo simulation approach. *Health Care Management Science*, 15(3):188–196, 2012.

[8] K. Doerner, A. Focke, and W.J. Gutjahr. Multicriteria tour planning for mobile healthcare facilities in a developing country. *European Journal of Operational Research*, 179(3):1078–1096, 2007.

[9] P.K. Drain, C.M. Nelson, and J.S. Lloyd. Single-dose versus multi-dose vaccine vials for immunization programmes in developing countries. *Bulletin of the World Health Organization*, 81(10):726–731, 2003.

[10] P.K. Drain, J.S. Ralaivao, A. Rakotonandrasana, and M.A. Carnell. Introducing autodisable syringes to the national immunization programme in madagascar. *Bulletin of the World Health Organization*, 81(8):553–560, 2003.

[11] R.Z. Farahani, N. Asgari, N. Heidari, M. Hosseininia, and M. Goh. Covering problems in facility location: A review. *Computers and Industrial Engineering*, 62(1):368–407, 2012.

[12] Global Alliance for Vaccines and Immunization. Country Hub, 2014. Available from http://www.gavialliance.org/country/, information and data accessed on July 14, 2014.

[13] L.V. Green, P.J. Kolesar, and W. Whitt. Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management*, 16(1):13–39, 2007.

[14] U.K Griffiths, A.C. Santos, N. Nundy, E. Jacoby, and D. Matthias. Incremental costs of introducing jet injection technology for delivery of routine childhood vaccinations: Comparative analysis from Brazil, India, and South Africa. *Vaccine*, 29(5):969–975, 2011.

[15] W. Gu, X. Wang, and S.E. McGregor. Optimization of preventive health care facility locations. *International Journal of Health Geographics*, 9(1):17, 2010.

[16] S. Guichard, K. Hymbaugh, B. Burkholder, S. Diorditsa, C. Navarro, S. Ahmed, and M.M. Rahman. Vaccine wastage in Bangladesh. *Vaccine*, 28(3):858–863, 2010.

[17] M. J. Hodgson, G. Laporte, and F. Semet. A covering tour model for planning mobile health care facilities in Suhum district, Ghana. *Journal of Regional Science*, 38(4):621–638, 1998.

[18] S.S. Hutchins, H.A.F.M. Jansen, S.E. Robertson, P. Evans, and R.J. Kim-Farley. Studies of missed opportunities for immunization in developing and industrialized countries. *Bulletin of the World Health Organization*, 71(5):549–560, 1993.

[19] L. Kamara, J.B. Milstien, M. Patyna, P. Lydon, A. Levin, and L. Brenzel. Strategies for financial sustainability of immunization programs: A review of the strategies from 50 national immunization program financial sustainability plans. *Vaccine*, 26(51):6717–6726, 2008.

[20] I.Z. Karaesmen, A. Scheller-Wolf, and B. Deniz. Managing perishable and aging inventories: Review and future research directions. In *Planning Production and Inventories in the Extended Enterprise*, pages 393–436. Springer, 2011.

[21] A. Klenke and L. Mattner. Stochastic ordering of classical discrete distributions. *Advances in Applied Probability*, 42(2):392–410, 2010.

[22] A. Klose and A. Drexl. Facility location models for distribution system design. *European Journal of Operational Research*, 162(1):4–29, 2005.

[23] B.Y. Lee, T.M. Assi, K. Rookkapan, D.L. Connor, J. Rajgopal, V. Sornsrivichai, S.T. Brown, J.S. Welling, B.A. Norman, S.I. Chen, R.R. Bailey, A.E. Wiringa, A.R. Wateska, A. Jana, W.G. Van Panhuis, and D.S. Burke. Replacing the measles ten-dose vaccine presentation with the single-dose presentation in Thailand. *Vaccine*, 29(21):3811–3817, 2011.

[24] B.Y. Lee, B.A. Norman, T.M. Assi, S.I. Chen, R.R. Bailey, J. Rajgopal, S.T. Brown, A.E. Wiringa, and D.S. Burke. Single versus multi-dose vaccine vials: An economic computational model. *Vaccine*, 28(32):5292–5300, 2010.

[25] B. Li, H. Wang, J. Yang, M. Guo, and C. Qi. A belief-rule-based inventory control method under nonstationary and uncertain demand. *Expert Systems with Applications*, 38(12):14997–15008, 2011.

[26] S. Liao, G. Koole, C. Van Delft, and O. Jouini. Staffing a call center with uncertain non-stationary arrival rate and flexibility. *OR Spectrum*, 34(3):691–721, 2012.

[27] J. Lim, E. Claypool, B.A. Norman, and Rajgopal J. Coverage models to determine outreach vaccination center locations in low and middle income countries. *Operations Research for Health Care*, 2016.

[28] X. Lu, J. Song, and K. Zhu. Dynamic inventory planning for perishable products with censored demand data. Technical report, Working Paper, 2004.

[29] I. Müller, T. Smith, S. Mellor, L. Rare, and B. Genton. The effect of distance from home on attendance at a small rural health centre in Papua New Guinea. *International Journal of Epidemiology*, 27(5):878–884, 1998.

[30] P. Murali, F. Ordóñez, and M.M. Dessouky. Facility location under demand uncertainty: Response to a large-scale bio-terror attack. *Socio-Economic Planning Sciences*, 46(1):78–87, 2012.

[31] A.T. Murray, D. Tong, and K. Kim. Enhancing classic coverage location models. *International Regional Science Review*, 33(2):115–133, 2010.

[32] G. Nagy and S. Salhi. Location-routing: Issues, models and methods. *European Journal of Operational Research*, 177(2):649–672, 2007.

[33] S. Nahmias. Perishable inventory theory: A review. *Operations Research*, 30(4):680–708, 1982.

[34] Y.B. Okwaraji, K. Mulholland, J.R.M.A. Schellenberg, G. Andarge, M. Admassu, and K.M. Edmond. The association between travel time to health facilities and childhood vaccine coverage in rural Ethiopia. A community based cross sectional study. *BMC Public Health*, 12(1):476, 2012.

[35] World Health Organization. Global routine vaccination coverage, 2014. *Weekly Epidemiological Record*, 90(46):617–623, 2015.

[36] D. Parmar, E.M. Baruwa, P. Zuber, and S. Kone. Impact of wastage on single and multi-dose vaccine vials: Implications for introducing pneumococcal vaccines in developing countries. *Human Vaccines*, 6(3):270–278, 2010.

[37] C. Prodhon and C. Prins. A survey of recent research on location-routing problems. *European Journal of Operational Research*, 238(1):1–17, 2014.

[38] M.L. Puterman. *Markov decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[39] S. Rahman and D.K. Smith. Use of location-allocation models in health service development planning in developing nations. *European Journal of Operational Research*, 123(3):437–452, 2000.

[40] A. Rais and A. Viana. Operations research in healthcare: A survey. *International Transactions in Operational Research*, 18(1):1–31, 2011.

[41] N. Ravichandran. Stochastic analysis of a continuous review perishable inventory system with positive lead time and poisson demand. *European Journal of Operational Research*, 84(2):444–457, 1995.

[42] C.S. Revelle, H.A. Eiselt, and M.S. Daskin. A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, 184(3):817–848, 2008.

[43] L.V. Snyder. Facility location under uncertainty: A review. *IIE Transactions*, 38(7):547–564, 2006.

[44] K. Strauss, A. van Zundert, A. Frid, and V. Costigliola. Pandemic influenza preparedness: The critical role of the syringe. *Vaccine*, 24(22):4874–4882, 2006.

[45] F. Tanser. Methodology for optimising location of new primary health care facilities in rural communities: A case study in KwaZulu-Natal, South Africa. *Journal of Epidemiology and Community health*, 60(10):846–850, 2006.

[46] S.A. Tarim and B.G. Kingsman. Modelling and computing $(R^n, S^n)$ policies for inventory systems with non-stationary stochastic demand. *European Journal of Operational Research*, 174(1):581–599, 2006.

[47] TechNet-21. VaCRO: Vaccine clinic reconstitution optimizer, 2015. Available from http://technet-21.org/en/forums/vacro-vaccine-clinic-reconstitution-optimizer and data accessed on September 30, 2015.

[48] Gürler Ü. Tekin, E. and E. Berk. Age-based vs. stock level control policies for a perishable inventory system. *European Journal of Operational Research*, 134(2):309–329, 2001.

[49] UNICEF. At a glance: India, 2014. Available from http://www.unicef.org/infobycountry/india-statistics.html, information and data accessed on July 14, 2014.

[50] United Nations Childrens Fund (UNICEF). Immunization: Why are children dying, 2015. Available from http://www.unicef.org/immunization/index_why.html, information and data accessedon October 7, 2015.

[51] Vaccine clinic reconstitution optimizer. VaCRO, 2015. Available from http://www.pitt.edu/~vacro/ and data accessed on September 30, 2015.

[52] J. Vandelaer, J. Bilous, and D. Nshimirimana. Reaching every district (RED) approach: A way to improve immunization performance. *Bulletin of the World Health Organization*, 86(3):A–B, 2008.

[53] V. Verter and S.D. Lapierre. Location of preventive health care facilities. *Annals of Operations Research*, 110(1-4):123–132, 2002.

[54] WHO. *Monitoring Vaccine Wastage at Country Level: Guidelines for Programme Managers*. World Health Organization, 2005.

[55] WHO. *Immunization Service Delivery*. World Health Organization, 2015.

[56] World Bank WHO, UNICEF. *State of the World's Vaccines and Immunization*. World Health Organization, 3rd edition, 2009. Available from http://whqlibdoc.who.int/publications/2009/9789241563864_eng.pdf, information and data accessed on June 26, 2012.

[57] World Health Organization and others. In-depth evaluation of the reaching every district approach in the African region, 2007. Available from www.who.int/immunization/sage/1_AFRO_1_RED_Evaluation_Report_2007_Final.pdf, information and data accessed on July 30, 2014.

[58] World Health Organization and others. Media center: Global vaccination targets off-track warns WHO, 2010. Available from http://www.who.int/mediacentre/news/releases/2015/global-vaccination-targets/en/, information and data accessedon Oct 7, 2015.

[59] T. Yang and J.G.C. Templeton. A survey on retrial queues. *Queueing Systems*, 2(3):201–233, 1987.

[60] W. Yang, M. Parisi, B.J. Lahue, M.J. Uddin, and D. Bishai. The budget impact of controlling wastage with smaller vials: A data driven model of session sizes in Bangladesh, India (Uttar Pradesh), Mozambique, and Uganda. *Vaccine*, 32(49):6643–6648, 2014.