




# Developing childhood vaccine administration and inventory replenishment policies that minimize open vial wastage

Zahra Azadi<sup>1</sup> · Harsha Gangammanavar<sup>2</sup> · Sandra Eksioglu<sup>3</sup> 

Published online: 19 February 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

In the last century, many infectious diseases have been completely eradicated or significantly reduced because of childhood vaccinations. Ample evidence suggests that low vaccination coverage in developing countries is caused by vaccine stockout and high rates of vaccine wastage. Wastage occurs when a vaccine vial is physically damaged or exposed to extreme temperatures, or when doses from an open vial are discarded after their safe-use time expires. The latter is referred to as open vial wastage (OVW). Clinics can use single-dose vials to reduce OVW; however, such an approach is more expensive than using multi-dose vials. The focus of this research is to develop new policies that support vaccine administration and inventory replenishment. These policies are expected to reduce OVW, reduce the cost of vaccinations, and improve vaccination coverage levels in developing countries. This paper proposes a two-stage stochastic programming model that identifies an optimal combination of differently sized vaccine vials and the corresponding decisions that clinics make about opening vials in the face of uncertain patient arrivals. This work develops a case study with data gathered from Bangladesh. Experimental results indicate that using a combination of vials of different sizes reduces OVW, as opposed to the current practice of using single-sized multi-dose vials. Experimental results also point to simple and economic vaccine administration policies that health care administrators can use to minimize OVW. The model is solved using an extension of the stochastic Benders decomposition algorithm, the L-shaped method (LS). This algorithm uses Gomory mixed integer and mixed-integer rounding cuts to address the problem's non-convexity. Computational results reveal that the solution approach presented here outperforms the standard LS method.

**Keywords** Vaccine inventory replenishment · Stochastic programming · L-shaped · Perishable product · Vaccine open vial wastage

---

✉ Sandra Eksioglu  
seksiog@clemson.edu

<sup>1</sup> University of Miami Business School, Coral Gables, FL, USA

<sup>2</sup> Southern Methodist University, Dallas, TX, USA

<sup>3</sup> Clemson University, Clemson, SC, USA

## 1 Introduction

The spread of infectious diseases has significantly receded over the last century, and these widespread immunities can largely be attributed to childhood vaccinations. Childhood vaccinations have helped eradicate diseases, like smallpox, and have severely restricted diseases, such as polio, measles, and tetanus (TT). Immunization programs, led by the World Health Organization (WHO) and United Nations Children's Emergency Fund (UNICEF), have been instrumental in delivering global eradication efforts. However, achieving immunization targets in developing countries has been particularly challenging. As recently as 2014, 60% of the 18.7 million children worldwide who did not receive the diphtheria-tetanus-pertussis (DTP) lived in ten developing countries (WHO 2014a). Vaccine stockout is one of the primary reasons for the low vaccination coverage in these places. In 2014, 26% of WHO-member countries experienced a national level vaccine shortage for at least one month (Subaiya 2015). Stockout occurs, in part, because of vaccine wastage. WHO estimates that more than half of the total vaccination supply distributed around the world is wasted (WHO 2005). Wastage occurs when a vaccine vial is physically damaged or exposed to extreme temperatures, or when doses from an open vial are discarded after their safe-use time expires. The latter is referred to as open vial wastage (OVW). The Global Alliance for Vaccines and Immunizations (GAVI) requests that countries take measures to bring down their vaccine wastage rates (WHO 2005), and lowering OVW rates can be one effective approach.

In many developing countries, immunizations are administered in health care centers or at outreach sessions. For example, in Bangladesh, 94% of immunizations are accomplished at outreach sessions (Guichard et al. 2010). These sessions are organized through programs, such as Expanded Programme on Immunization (EPI), and delivered by trained nurses. Health care centers are often located in remote areas that lack proper transportation and basic amenities, including refrigeration. Childhood vaccines, including Bacille Calmette-Guérin (BCG), measles, DTP, and TT, are typically produced in multi-dose vials. Outreach sessions conducted by EPI follow a strict open vial policy that requires opened vials of multi-dose vaccines be discarded at the end of the day or 6 h, whichever comes first (WHO 2014b). Discarded doses contribute to OVW. While single-dose vials have zero OVW, they are more expensive than multi-dose vials because of additional costs for packaging, holding, and transportation. A 2010 study in Bangladesh estimated OVW for BCG at 85%, measles at 71%, DTP at 44.2%, and TT at 36% (Guichard et al. 2010).

Organizations managing immunization programs in developing countries do make vaccine replenishment decisions for a single vial size (Dhamodharan and Proano 2012; Guichard et al. 2010). Vaccines are then distributed to clinics based on a fixed replenishment schedule (GAVI 2011). Intuition suggests that allowing clinics to use a combination of vials of different sizes should reduce OVW. Better inventory management should result when clinics decide their own replenishment schedules based on observed patient arrivals and when clinics have easy-to-implement vaccine administration policies.

In this study, “inventory replenishment policy” refers to the combination of multi-dose vaccine vials, e.g., single-, two-, and ten-dose vials, that a clinic should order to reduce OVW during outreach sessions and maintain low vaccination costs. “Vaccine administration policy” refers to the conditions under which a nurse should open a new vial of a particular size during an outreach session. Such a policy accounts for session duration, the number of patients present, and/or the time of day when a multi-dose vial of a particular size should be opened. In the absence of such policies, deciding whether to open or not to open a multi-dose vial vaccine is challenging because of the uncertain number of patient arrivals. The study

presented here develops a two-stage stochastic programming model to identify the optimal combination of multi-dose vaccine vials to order, the timing of orders, and the corresponding vial opening decisions in the face of uncertain patient arrivals. The goal is to minimize OVW and achieve lower vaccination costs for developing countries. The models proposed here are tested via an extensive numerical analysis using data from Bangladesh. Numerical results inspired the development of a number of vaccine administration policies which, as demonstrated in this paper, can reduce OVW.

## 1.1 Literature review

In developing countries, EPI vaccines are purchased by governmental and non-profit agencies to satisfy annual vaccine requirements. In these countries, EPI delivers vaccines to a central store from which they are shipped to regional stores, then to district stores, and, finally, to local clinics. The number of downstream tiers, or supply chain stages, that handle vaccines vary by country. Some studies in the literature focus on national-level decisions made about how many vaccines should be purchased to ensure high vaccine coverage levels (Chen et al. 2014a). Other studies focus on the structure of the corresponding supply chain to minimize immunization costs (Assi et al. 2011, 2013; Brown et al. 2014). A few studies focus on the decisions made about clinic-level vaccine administration that can minimize OVW (Mofrad et al. 2014, 2016; Dhamodharan and Proano 2012). Similarly, this work examines clinic-level inventory replenishment decisions and vaccine administration policies with the goal of minimizing OVW and inventory replenishment costs.

The literature on vaccine supply chain management has increased in recent years. Research focused on OVW management started as vaccine vial wastage increased in outreach immunization sessions delivered by programs such as EPI in developing countries. The work conducted by Drain et al. (2003) was the first to discuss the general differences between single- and ten-dose vaccine vials in terms of cost, distribution, coverage, and safety. A survey of vaccine wastage in Bangladesh, provided in Guichard et al. (2010), details the usage and wastage rates of differently sized vaccine types in randomly selected outreach and non-outreach sessions. The aims of both studies were to estimate vaccine wastage, but neither considered the associated supply chain costs. However, these costs were considered in the analysis conducted by Parmar et al. (2010). The scope of these studies was restricted to analyzing existing vaccine administration practices, and, hence, they do not prescribe any plan for efficient inventory management. However, they recognize that “there is an urgent need for more rigorous and systematic wastage monitoring” (Parmar et al. 2010).

The following studies use simulation-based approaches to assess the impacts of different-sized multi-dose vials on OVW. Lee et al. (2011) propose a simulation model for Thailand’s Trang Province that evaluates the cost of replacing ten-dose vial vaccines with single-doses. The results of this study show that, given a fixed patient-arrival rate, the cost associated with medical waste disposal is greater than the cost savings resulting from OVW reduction. This finding, follows their earlier simulation work that models patient arrivals using a Poisson distribution (Lee et al. 2010). The authors conclude that the suitable number of doses per vial may vary by region and patient-arrival rate. However, the Poisson arrival assumption may not necessarily capture the arrival process in outreach immunization sessions due to the specific challenges of organizing these sessions in particular communities. A similar simulation study was carried out by Yang et al. (2014) to analyze the impact of different patient-arrival rates on OVW for five- and ten-dose vaccine vials. In this case, the authors use real-world data to estimate parameters for simulation. These studies examine the current practice of using a single multi-dose vial, and they recommend using a particular vial size based on patient

arrival rates. These recommendations are based on simulated cost estimates, but once again, these works do not prescribe any vaccine inventory management schemes.

To develop a vaccine management plan, optimization models are necessary. Such approaches are limited to include deterministic mixed-integer programs (MIP) and finite state probabilistic models. In Dhamodharan and Proano (2012), the total vaccination cost is minimized via optimal ordering decisions which are identified by solving a deterministic MIP integrated within a Monte Carlo simulation setup. The authors use the simulation model to generate patient arrivals and compute OVW under a fixed vaccine administration policy and for a given multi-dose vaccine vial. The estimated OVW is then used in the optimization model to determine the optimal ordering decisions. Mofrad et al. (2014) propose a Markov Decision Process (MDP) model to identify a vial-opening policy that minimizes expected OVW cost and maximizes vaccine coverage over a finite horizon. The authors use a backward induction algorithm to solve their probabilistic model. This work is further extended in Mofrad et al. (2016), who integrates the MDP model with simulation to analyze how session duration impacts the optimal policy obtained in their previous work. In both studies, the authors determine the clinic closing time or the time when patients should be rejected. However, they do not consider loss of opportunity due to not serving some patients. Moreover, they assume that an infinite amount of vaccines can be replenished, which is not always the case at outreach sessions. These studies address the inventory management of a single multi-dose vaccine vial, and patient arrival is characterized either probabilistically or through simulation. The optimization models focus either on ordering decisions or vial-opening decisions. However, these decisions are interrelated, and, therefore, a model which considers both and accounts for uncertainty in patient arrivals should be very effective in reducing OVW costs and increasing immunization rates.

Vaccines are perishable products that expire after their safe-use time. Thus, the study presented here is related to the literature on inventory replenishment models for perishable products with fixed shelf-life. Works by Nahmias (1982), Raafat (1991), Bakker et al. (2012), and Atan and Rousseau (2016) model these problems by extending the classic economic lot sizing (ELS) problem formulation. ELS and its variations are often formulated using extensions of the minimum cost network flow model (Palak 2013). While some ELS models for perishable products with fixed shelf-life consider demand to be deterministic, others have considered stochastic demand. Bakker et al. (2012) surveys this work, and Van Zyl (1963) and Nahmias and Pierskalla (1973) find optimal ordering policies, when demand follows continuous and differentiable distributions, by incorporating shortage and ordering costs. Nahmias and Pierskalla (1973) develop a cost function that also accounts for wastage costs due to expiry. An extended  $m$ -period dynamic programming model is presented both in Fries (1975) and in Nahmias (1975). The ELS model for perishable products with fixed shelf-life is solved with well-known ordering policies, such as  $(s, S)$  and  $(Q, r)$ . A Markov renewal process is used to obtain an ordering policy in Liu and Lian (1999), who use a closed-form cost function for an  $(s, S)$  policy with back-orders. Weiss (1980) shows the optimality of  $(s, S)$  policy when the demand process is compound Poisson. In these studies, the inventory replenishment model is restricted to following a specific policy that is optimal only under certain demand distributions and cost functions. For example, in Weiss (1980), the proposed inventory ordering policy is optimal only if the demand is represented with a compound Poisson process.

Using exact methods to derive optimal policies for stochastic ELS models and their extensions is computationally complex which has resulted in the development of many heuristic policies. For example, Nahmias and Wang (1979) develop a myopic-based inventory policy using bounds on wastage costs. Following this approximation method, Chen et al. (2014b) develop two heuristic policies for a joint inventory control and pricing model, but these poli-

cies perform well only when demand is non-stationary. Simulation-based methods have also been employed to obtain approximate policies; see Van Donselaar and Broekmeulen (2012). A combination of dynamic programming and simulation is used to reduce the state space of the stochastic ELS model in Haijema et al. (2009). However, optimization models that provide exact solutions to the ELS problem and its extensions have been limited to deterministic models; see Ekşioğlu et al. (2006). Stochastic demand is handled through optimal policies only when highly simplifying assumptions have been made, or through sub-optimal heuristic policies under slightly general settings (Liu and Lian 1999; Nahmias and Wang 1979; Weiss 1980; Van Donselaar and Broekmeulen 2012; Palanivel et al. 2018).

## 1.2 Our contributions

When stochastic patient arrivals cannot be described using standard distributions, probabilistic models are not effective. In this case, using optimization methods is preferred over obtaining approximate policies through simulation. Moreover, when vaccines are distributed via a combination of multiple-dose vials, inventory replenishment and vaccine administration decisions, estimating inventory states, and their interdependencies become complicated. Such realistic requirements can be captured using stochastic programming (SP) models. Furthermore, SP provides effective tools to design algorithms that can handle these comprehensive models. In this regard, the contributions of this paper are:

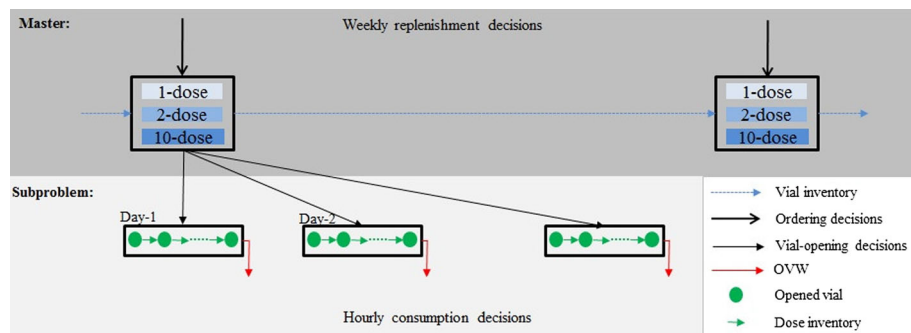
- *Vaccine administration and inventory replenishment model* A two-stage SP model is developed to address vaccine vial replenishment by capturing (a) the order frequency for respective quantities of different-sized vials, (b) the opening schedule for these vials, and (c) the administration of available doses to patients. Experimental results indicate that using a combination of different-sized vials, compared to using vials of a single size, results in monetary savings, better service, and OVW reductions at outreach sessions.
- *Vaccine administration policies* Using the solutions obtained from the vaccine administration and inventory replenishment model developed, simple and economic vaccine administration policies for outreach sessions are proposed. These policies depend on the population size and birth rate in the regions served. This study compares their effectiveness by replicating a simulation process.
- *Scalable algorithm* A new solution approach is developed for two-stage stochastic integer programs (2-SIP) with continuous recourse. This algorithm is motivated by the well-known stochastic Benders method and uses Gomory mixed integer (GMI) and mixed-integer rounding (MIR) cuts to address the non-convexity of the first-stage problem. Compared to the standard method approach, the approach presented here significantly reduces the computational requirement. Such computational enhancements allow our algorithm to be applied to real-world problems.

In the following sections, the SP model is developed and tested. The model is presented in Sect. 2, and the solution approach is discussed in Sect. 3. Data and computational results are presented in Sect. 4, and ideas regarding extensions of the current model and the solution approach are presented in Sect. 5.

## 2 Problem formulation

### 2.1 Problem description and assumptions

This section details a model that aids the design and management of immunization programs conducted through outreach sessions in developing countries. The structure of the vaccine



**Fig. 1** Inventory dynamics in the vaccine administration and inventory replenishment model

distribution system differs by country. Typically, clinics replenish inventories via shipments from a regional distribution center. Clinics run two types of immunization sessions: drop-in and outreach. In drop-in sessions, individuals visit the clinic to get vaccinated. In outreach sessions, health workers travel to provide vaccinations in rural communities during temporary gatherings (Dhamodharan and Proano 2012). For example, drop-in sessions account for 6%, and outreach sessions account for 94% (Guichard et al. 2010). This study focuses on short outreach sessions, usually lasting a few weeks, that target infant populations in rural areas. Assume that patient arrivals in outreach sessions are random. Due to the stochastic nature of patient arrivals and the open vial policy, identifying replenishment schedules and administration policies is critically important.

The model presented below is based on the following assumptions: (i) vaccines are distributed in multi-dose vials of different sizes. (ii) Vaccine purchase cost is proportional to its size because of additional packing costs. (iii) Clinics maintain differently sized inventories of multi-dose vials and inventories of doses from opened vials. (iv) OVW and storage cost the clinic.

Transportation lead time for vaccine replenishment is constant since deliveries are received from the same distribution center. Thus, this problem can be easily transformed to have zero lead time: (a) the order for a replenishment that is to arrive in time  $t$  is submitted in time  $t - t'$ , where  $t'$  is the lead time; and (b) the unit procurement cost is increased by the unit inventory cost during transit time. Thus, based on these assumptions, the remainder of this paper assumes negligible lead times.

Figure 1 provides a schematic representation of the modeling approach used for this problem. The darker shaded area represents inventory replenishment decisions for multi-dose vials during the planning horizon. These decisions are made weekly. Each box represents the type of vials in the inventory over a particular week. Boxes in successive time periods are connected to represent the inventory flow. The lighter shaded area represents vaccine administration decisions. These decisions are made daily. Each box shows when a vial is opened and how the doses are used during a given day. Boxes in successive time periods are not connected since unused doses are discarded and contribute to OVW.

## 2.2 Vaccine administration and inventory replenishment model

For the model formulation presented here, note the use of lower case variables for vector and matrix entries and that bold fonted variables represent the entire vector or matrix.

This model considers two major decisions: (a) inventory replenishment or ordering decisions; and (b) vial administration or opening decisions. Ordering decisions help to identify the optimal replenishment schedule, order mix and vial sizes. These decisions are made weekly. On the other hand, opening decisions are a function of the vaccine administration policy which depends on patient arrivals, and thus, these decisions are made much more regularly: on an hourly basis. Here,  $\mathcal{T} = 1, \dots, T$  denotes the ordering decision epochs, and each ordering period consists of  $N$  opening decision epochs. Therefore,  $\mathcal{N} = 0, \dots, NT$  denotes all the opening decision epochs over the planning horizon.

First, modeling of the inventory replenishment decisions is described. Let  $\mathcal{V}$  represent the set of multi-dose vial sizes, e.g., one-dose, five-dose and ten-dose vials, available for order. Once the vials are opened, they must be used within their safe-use time, and  $\tau$  denotes this limit. At  $t \in \mathcal{T}$ , ordering decisions  $z_t$  are made at fixed cost  $f_t$ . If an order is placed ( $z_t = 1$ ), then the replenishment quantity for each multi-dose vial size can be determined. These decisions are represented by  $r_{vt}$ , and the corresponding variable purchase cost is  $c_{vt}$ ,  $\forall v \in \mathcal{V}$ . At each opening decision epoch  $n \in \mathcal{N}$ , vial-opening decisions are made. Let  $u_{vn}$  represent the number of vials of size  $v$  opened. Replenishment quantity decisions  $r_{vt}$  and vial-opening decisions  $u_{vn}$  together determine the state of vial inventory, represented here as  $s_{vn}$ ,  $\forall v \in \mathcal{V}$ . Let  $d_{vt}$  represent the corresponding unit inventory holding cost. The evolution of inventory for each vial size  $v \in \mathcal{V}$  is captured by the following flow-balance equations:

$$s_{vNt} = s_{v(Nt-1)} + r_{vt} - u_{vNt} \quad \forall t \in \mathcal{T}, \quad (1a)$$

$$s_{vn} = s_{vn-1} - u_{vn} \quad \forall n \in \mathcal{N} \setminus \{N, 2N, \dots, TN\}, \quad (1b)$$

The initial inventory  $s_{v0}$  is assumed to be known. Since the inventory of vials is replenished weekly, Eq. (1a) represents the corresponding inventory balance constraints. However, since dose administration decisions are made during vaccination sessions and those decisions impact the inventory of vials, Eq. (1b) captures inventory balance during vaccine administration time periods. Since open vial policy does not apply to unopened vials, loss of inventories is not considered in these constraints:

$$\sum_{v \in \mathcal{V}} r_{vt} \leq M_t z_t \quad \forall t \in \mathcal{T}. \quad (2)$$

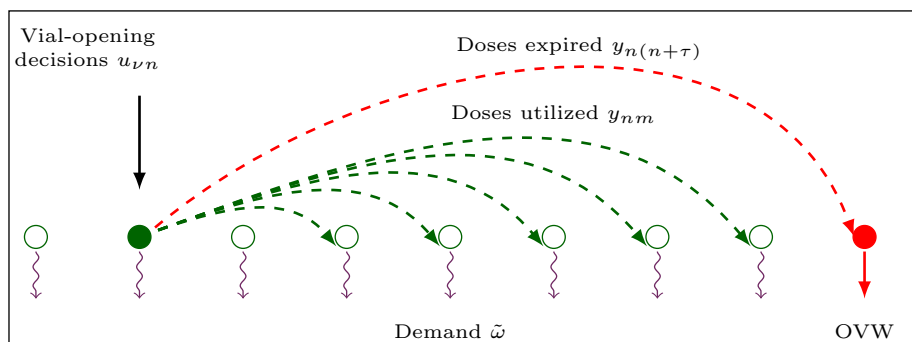
Let  $M_t$  be an upper bound (UB) on the quantity of vials ordered in a time period. Constraints (2) ensure that the inventory of vials is replenished only if an order is placed in period  $t$ . Note that decisions about vial replenishment and inventory ( $\mathbf{z}$ ,  $\mathbf{r}$ ,  $\mathbf{u}$ , and  $\mathbf{s}$ ) are made before the number of patient arrivals is realized.

Next, modeling of vial administration decisions is described. The vials opened in period  $n$  can vaccinate children in period  $m < n + \tau$  or before the end of a session, whichever occurs first. Let  $y_{nm}$  denote the number of doses obtained from vials opened in period  $n$  and used in period  $m$ . Therefore,  $y_{n(n+\tau)}$  represents the number of doses that expire after their safe-use time expires and contributes to OVW. The relationship between the number of vials opened and the number of doses available is captured by the following equation:

$$\sum_{m=n}^{n+\tau-1} y_{nm} + y_{n(n+\tau)} = \sum_{v \in \mathcal{V}} q_v u_{vn} \quad \forall n \in \mathcal{N}, \quad (3)$$

where  $q_v$  is the number of doses in a multi-dose vial of size  $v$ . The right-hand side of Eq. (3) represents the total number of doses obtained by opening vials, while the left-hand side is the sum of utilized and expired doses in period  $n$ .





**Fig. 2** Dose utilization for  $n = 2$  with  $\tau = 7$ . \*Filled circles indicate opened vials, and unfilled circles indicate unopened vials

The total number of patients arriving in time period  $n$  is denoted by  $\tilde{\omega}_n$ . If the total number of patients arriving during period  $t$  exceeds the number of doses available, then these patients will not be served. This model assumes that patients do not wait for vaccinations. This assumption is relaxed in Sect. 4.4 when the wait-to-open policy is described. A penalty is associated with each unserved patient, denoted by  $\ell_n$ . Based on these assumptions, the dose inventory must satisfy the following flow-balance equations:

$$\sum_{m=n-\tau+1}^n y_{nm} + \ell_n = \omega_n \quad \forall n \in \mathcal{N}. \quad (4)$$

Here,  $\omega_n$  is a realization of the random variable  $\tilde{\omega}_n$ . Equation (4) indicates that, of the total number of patients arriving during a session, the number of patients vaccinated depends on the total number of doses available, denoted by the first term in the left-hand side of the equation. If the number of open doses is smaller than the number of patients present during period  $t$ , then the unserved patients leave the session. Figure 2 provides a schematic representation of how doses obtained from vials opened in period  $n = 2$  are utilized over the next  $\tau = 6$  periods. OVW is represented by the arc in red.

Finally, the modeling of the objective function is described. The objective function includes the fixed and variable purchasing costs, ( $f_t$ ) and ( $c_v$ ) respectively; inventory holding costs ( $d_v$ ); and the expected values of wastage ( $g$ ) and penalty ( $p$ ) costs. This objective function is given by:

$$\sum_{t \in \mathcal{T}} (f_t z_t + \sum_{v \in \mathcal{V}} c_v r_{vt}) + \sum_{v \in \mathcal{V}} \sum_{n \in \mathcal{N}} d_v s_{vn} + \mathbb{E} \left\{ \sum_{n \in \mathcal{N}} (g y_{n(n+\tau)} + p \ell_n) \right\}. \quad (5)$$

The expectation function is calculated over the distribution of  $\tilde{\omega}$ , and its argument captures the penalties associated with unserved patients and OVW disposal. Note that the total disposal cost depends on patient arrivals and is bounded below by zero.

The vaccine vial replenishment problem can be written as a 2-SIP to capture the impact of the stochastic nature of patient arrivals on vaccine inventory replenishment and vaccine administration decisions. The complete 2-SIP is presented in “Appendix A”. To simplify the exposition of the algorithm, the remainder of the paper uses the following succinct representation of the model formulation: A single decision vector  $\mathbf{x} \in \mathcal{X}$  collectively denotes the ordering decision  $z_t$ , the replenishment quantity  $r_{vt}$ , vial-opening decisions  $u_{vn}$ , and the state variables  $s_{vn}$ , i.e.,  $\mathbf{x} = (\mathbf{z}, \mathbf{r}, \mathbf{u}, \mathbf{s})^\top$ . Note that decision vector  $\mathbf{x}$  consists of binary ( $z_t$ ),



as well as pure-integer decisions.<sup>1</sup> Since decision  $\mathbf{x}$  is made prior to realizations of patient arrivals, which is non-anticipative in nature (Birge and Francois 2011),  $\mathbf{x}$  represents the first-stage decision. On the other hand, vial-opening decisions  $y_{nm}$  depend on patient arrivals, and therefore, these decisions are made in an adaptive manner. All the adaptive decisions are collated in a single decision vector  $\boldsymbol{\gamma} = (\mathbf{y}, \boldsymbol{\ell})^\top$ . The replenishment model can then be written as:

$$\begin{aligned} \min \quad & F(\mathbf{x}) + \mathbb{E}(H(\mathbf{x}, \tilde{\omega})) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X} \end{aligned} \quad (6)$$

where the recourse function  $H(\mathbf{x}, \omega)$  uses the first-stage decisions  $\mathbf{x}$  and realization  $\omega$  of random variable  $\tilde{\omega}$  as input. The resulting model is:

$$\begin{aligned} H(\mathbf{x}, \omega) = \min \quad & G(\boldsymbol{\gamma}) \\ \text{s.t.} \quad & \mathbf{W}\boldsymbol{\gamma} = \rho(\omega) - \mathbf{T}\mathbf{x}. \end{aligned} \quad (7)$$

In the equations above, functions  $F(\cdot)$  and  $G(\cdot)$  are linear functions;  $\mathbf{W}$  is the recourse matrix; and  $\mathbf{T}$  is the technology matrix. In SP literature, formulation (6) is referred to as the first-stage problem, and formulation (7) is the second-stage problem.

Equations (3) and (4) represent the constraints set of formulation (7). The first-stage decision variable  $\mathbf{x}$  impacts the right-hand side of Eq. (3).  $\mathbf{T}\mathbf{x}$  in formulation (7) represents this relationship. The uncertainty affects the right-hand side of Eq. (4).  $\rho(\omega)$  in formulation (7) denotes this relationship. Consequently, recourse matrix  $\mathbf{W}$  represents the left-hand side of Eq. (3), and Eq. (4) is independent of randomness. Such SP models are said to have a fixed recourse (Birge and Francois 2011). Finally, the first- and second-stage problems have discrete decision variables. However, the structure of the second-stage constraint set satisfies the following proposition:

**Proposition 1** *The linear programming in second-stage problem (7) has at least one integer optimal solution.*

**Proof** Second-stage problem (7) is an uncapacitated minimum cost network flow model. Therefore, the recourse matrix satisfies the total unimodularity property (Ahuja et al. 1988). In other words, since the recourse matrix is integer, every basic feasible solution to the linear relaxation of second-stage problem will be an integer for any integer vector  $(\sum_{v \in \mathcal{V}} q_v u_{vn}, \omega, \forall n \in \mathcal{N})$ , which appears on the right-hand side of second-stage problem (7). Since the number of patients and number of doses available are integers, this condition is satisfied.  $\square$

As a consequence of Proposition 1, the integer constraints are relaxed in the second-stage problem. Thus, this problem is solved as a linear program for a given input of  $(\mathbf{x}, \omega)$ . Linearizing second-stage problem (7) is necessary for the implementation of the LS algorithm which relies on the duality of the second-stage problem. From this point the 2-SIP notation refers to the integer first-stage problem (6) and the linear relaxation of second-stage problem (7). The solution algorithm is presented in the next section.

<sup>1</sup> Notation  $\mathbf{v}$  denotes a single column vector obtained from elements  $(v_1, v_2, \dots, v_i)^\top$  for one-dimensional vectors, column-wise concatenation for two-dimensional matrices  $(v_{11}, \dots, v_{i1}, v_{12}, \dots, v_{i2}, \dots, v_{1j}, \dots, v_{ij})^\top$ , and so on.

### 3 Solution approach: an extended L-shaped method

2-SIPs have been used to model many applications in the fields of financial planning, capacity expansion, manufacturing, and resource allocation, among others (Edirisinghe and Patterson 2007; You and Grossmann 2013; Patriksson et al. 2015). Over the past several decades, different algorithms have been proposed to tackle these problems. To achieve computational tractability, many algorithms represent uncertainty through a finite number of realizations or scenarios. If  $\Omega = \{\omega_1, \omega_2, \dots, \omega_S\}$  represents this finite set of scenarios with respective probabilities  $p_i$ ,  $i = 1, \dots, S$ , then the expectation function in first-stage problem (6) can be written as:

$$\mathbb{E}\{H(\mathbf{x}, \tilde{\omega})\} = \sum_{i=1}^S p_i H(\mathbf{x}, \omega_i). \quad (8)$$

Discretizing  $\omega$  allows modeling of the corresponding deterministic equivalent formulation (DEF) of the proposed 2-SIP. In general, even if  $S$  is relatively small, the size of the DEF can increase quickly. Since the decision variables of the first-stage problem are integers, the DEF is a network flow problem with fixed charge costs, known as an NP-hard problem (Hochbaum and Segev 1989). Therefore, solving the DEF is computationally challenging.

Decomposition-based methods, including the stochastic Benders decomposition (Van Slyke and Wets 1969), Dantzig–Wolfe decomposition (Dantzig and Wolfe 1960), and progressive hedging (Rockafellar and Wets 1991), have effectively addressed this computational challenge. These methods iteratively build approximations for the expected recourse function. Dantzig–Wolfe decomposition is an inner linearization method that iteratively solves the dual of the first-stage problem. However, this method is not applicable to the class of problems with discrete first-stage variables, which is the case here. Progressive hedging is a primal-dual method where, in each iteration, a penalty is associated with a deviation from a feasible solution. However, using this method requires selecting an appropriate proximal parameter that is both instance dependent and hard to discern (Watson and Woodruff 2011). Therefore, the solution approach proposed here is based on stochastic Benders decomposition, which is also known as the LS method.

In order to build the DEF of a 2-SIP, which has a random variable with a large number of realizations, or a continuous underlying distribution as is the case in our application, a sampling-based approach is recommended (Birge and Francois 2011). In this case, one can generate  $S$  realizations of random vector  $\tilde{\omega}$  using Monte Carlo simulation. Then, set  $\Omega$  will be comprised of these simulated vectors, and each will have the same probability  $p_i = 1/S$ . Given this set of realizations, the sample average approximation (SAA) problem of the DEF can be written as:

$$\min_{\mathbf{x} \in \mathcal{X}} \hat{F}_S(\mathbf{x}) := F(\mathbf{x}) + \frac{1}{S} \sum_{i=1}^S H(\mathbf{x}, \omega_i). \quad (9)$$

The second term in problem (9) is an unbiased estimator of the expectation function in formulation (8). The SAA problem is a 2-SIP with discrete distribution and therefore can be solved using the LS method with a branch-and-cut procedure to recover the integrality of the first-stage problem. Since the SAA problem is constructed using random realization obtained by sampling with Monte-Carlo or other techniques, verifying the quality of the solutions obtained from the SAA is imperative. Thus, this study uses estimates of lower bounds (LBs) and UBs computed across multiple replications, as suggested by Mak et al. (1999a).

The LS method is an iterative method used to solve SP problems with continuous recourse. It is well known that, in these problems, the expected recourse function is piecewise linear and convex (Birge and Francois 2011). Therefore, the dual solutions of the second-stage problem are used to generate an LB and provide an outer approximation of this expected recourse function in each iteration of the LS method. When second-stage problem variables are discretely valued, then additional steps are necessary to achieve convexity; see Sen (2011) for details. Due to the total unimodularity of recourse matrix  $\mathbf{W}$ , employing such procedures is not necessary. In order to solve problems with binary constraints in the first stage, Laporte and Louveaux proposed the integer LS method in 1993 (Laporte and Louveaux 1993). Later, improved optimality cuts were introduced by Hjorring and Holt (1999) to strengthen the method. This method was further improved by using local branching techniques developed by Rei et al. (2009). These methods work for problems with only first-stage binary variables and hence do not apply to the problem presented here.

In order to speed the computational time of solving large-scale mixed-integer two-stage SP models with continuous recourse, Bodur et al. (2016) add valid integer cuts to the LS method. The cuts are added to the LP relaxation of the first-stage problem. Their computational results show an improvement in the LS algorithm's performance. The solution algorithm developed here is an extension of the LS method to solve the SAA formulation where the first-stage problem is a MIP. This work uses GMI and MIR cuts to strengthen Benders' cuts. In this regard, our approach resembles the work of Bodur et al. (2016) who add cuts to the DEF of their problem.

### 3.1 Extended LS method

When the classic LS method of Van Slyke and Wets (1969) is applied to 2-SIPs with continuous recourse, optimization is achieved by solving the first-stage problem as an MIP in every iteration. These MIP solvers often tend to be computationally expensive, prohibiting the extensive numerical experimentations necessary to conducting reliable statistical analyses of the results obtained by solving an SAA-based formulation of the problem. Herein lies the motivation for developing the algorithm presented here.

Our algorithm operates in one of two modes: (a) *optimization* mode or (b) *integer-feasibility* mode. In the optimization mode, the goal is to develop acceptable LB approximations of the first-stage objective function. In this regard, the procedure adopted is akin to the classic LS method applied to a two-stage stochastic linear program (2-SLP). The goal of the integer-feasibility mode is to determine an integer feasible solution.

The algorithm is initialized by setting the first-stage feasible region as  $\bar{\mathcal{X}}^0 = \mathcal{X}$ , the set of original constraints. Over the course of the algorithm, lower-bounding affine functions are computed in the optimization mode and integer cuts are added in the integer-feasibility mode. In iteration  $k$ , the first-stage feasible region  $\bar{\mathcal{X}}^{k-1}$  is characterized by the set of original constraints  $\mathcal{X}$ , a set of lower bounding cuts  $\mathcal{L}_{opt}^{k-1}$ , and a certain set of integer cuts  $\mathcal{L}_{int}^{k-1}$ . The lower bounding cuts are expressed using an auxiliary variable  $\eta$ , as in the case of the classic LS method. While  $\mathbf{x}$  is an integer decision variable,  $\eta$  is a free continuous variable. Therefore, the feasible region for iterations  $\bar{\mathcal{X}}^{k-1}$ ,  $k > 1$  is described in an extended space which includes both  $\mathbf{x}$  and  $\eta$ , a mixed-integer set. The second-stage problem in formulation (7) satisfies the relatively complete recourse assumption; that is, the second-stage problem is feasible for all  $\omega_i \in \Omega$  and  $\mathbf{x} \in \bar{\mathcal{X}}^k$ , which mitigates the need to generate feasibility cuts. Feasibility is achieved because not every patient is assumed to be vaccinated, and waste is allowed

and measured. However, the relatively complete recourse assumption can be relaxed, and a mechanism to generate feasibility cuts can be incorporated in the optimization mode. These feasibility cuts will be generated in a manner similar to the classic LS method, and therefore, they are not discussed here. First, the steps involved in the optimization mode procedure are described, followed by the approach used in the integer-feasibility mode.

### 3.1.1 Optimization mode

In this mode, iteration  $k$  starts by solving the following problem that identifies a candidate solution:

$$\min \{F(\mathbf{x}) + \eta \mid (\mathbf{x}, \eta) \in \bar{\mathcal{X}}_{lp}^{k-1}\} \quad (M_{lp}^k)$$

where  $\bar{\mathcal{X}}_{lp}^{k-1}$  is a linear relaxation of the first-stage feasible region  $\bar{\mathcal{X}}^{k-1}$ . The optimal solution of this problem is given by  $(\mathbf{x}^k, \eta^k)$ , with objective function value  $v_S^k$ . Using the solution  $\mathbf{x}^k$  and a realization  $\omega_i \in \Omega$ , the second-stage problem  $H(\mathbf{x}^k, \omega_i)$  is solved to obtain the optimal dual solution,  $\pi^k$ . This procedure is enumerated for every  $\omega_i \in \Omega$ . Using these dual solutions, a lower bounding affine function cut is computed by:

$$l_{opt}^k(\mathbf{x}, \eta) := \frac{1}{S} \sum_{i=1}^S (\pi^k)^\top [\rho(\omega_i) - \mathbf{T}\mathbf{x}] - \eta \leq 0. \quad (10)$$

This affine function is used to obtain the updated set  $\mathcal{L}_{opt}^k = \mathcal{L}_{opt}^{k-1} \cup \{l_{opt}^k(\mathbf{x}, \eta)\}$  which provides an LB approximation to the objective function.

The optimal objective function value  $v_S^k$  provides an LB to  $(M_{lp}^k)$ . A UB to  $(M_{lp}^k)$  is  $\hat{F}_S(\mathbf{x}^k) = l_{opt}^k(\mathbf{x}^k, \eta^k)$ . Let  $\Delta^k$  denote the relative gap between the current UB and LB defined as  $\Delta^k(\mathbf{x}^k) = (\hat{F}_S(\mathbf{x}^k) - \hat{v}^k) / \hat{F}_S(\mathbf{x}^k)$ . If  $\Delta^k(\mathbf{x}^k)$  is less than a predefined error  $\epsilon$ , then switch to the integer-feasibility mode. Otherwise, if  $\Delta^k(\mathbf{x}^k) \geq \epsilon$ , then set  $\mathcal{L}_{int}^k = \mathcal{L}_{int}^{k-1}$ ; i.e., no new integer cuts are added in iteration  $k$ . An updated first-stage feasible region  $\bar{\mathcal{X}}^k$  results, and one iteration in the optimization mode is complete.

**Remark** While the algorithm operates in the optimization mode, no new integer cuts are added. Therefore, when the relative gap  $\Delta^k(\mathbf{x}^k) < \epsilon$ , the algorithm has solved a 2-SLP with a first-stage problem characterized by a feasible region  $\{\mathbf{x}^k \mid \mathbf{x}^k \in \bar{\mathcal{X}}^k \cap \{l(\mathbf{x}^k, \eta^k) \leq 0\}, l \in \mathcal{L}_{int}^\kappa\}$ , where  $\kappa$  is the last iteration when the algorithm operates in integer-feasibility mode. This feasible region is defined by the original set of constraints, and the integer cuts added up to iteration  $\kappa$ . Therefore, the algorithm presented here converges to the optimal solution of this problem in a finite number of iterations if  $\omega$  has a finite support (Birge and Francois 2011).

### 3.1.2 Integer-feasibility mode

Let  $\kappa$  denote the iteration when switching to integer-feasibility mode. At this point, the current solution is assigned as an incumbent solution; i.e.,  $\hat{\mathbf{x}}^\kappa = \mathbf{x}^\kappa$  with the corresponding objective function value  $\hat{v}^\kappa = v^\kappa$ . If the incumbent solution satisfies the integer requirement, then the integer feasible solution has an acceptable relative gap. Therefore, the algorithm is terminated. Otherwise, an MIP is solved:

$$\min \{F(\mathbf{x}) + \eta \mid (\mathbf{x}, \eta) \in \bar{\mathcal{X}}_{lp}^{k-1} \cap l_{opt}^k(\mathbf{x}, \eta) \leq 0\}. \quad (M_{mip}^\kappa)$$

Note that the constraint set in the above problem does includes the latest optimality cut added in the optimization mode. Let  $\hat{\mathbf{x}}_S^K$  be the integer optimal solution with the objective function value  $\hat{v}_S^K$ . Since the solution is updated, the gap estimate  $\Delta^K(\hat{\mathbf{x}}^K)$  is recomputed with  $\hat{F}_S(\hat{\mathbf{x}}^K)$  as the UB. This update might result in  $\Delta^K$  becoming greater than error  $\epsilon$ . In this case, the lower bounding approximation for  $\hat{\mathbf{x}}^K$  is not acceptable, and hence, the optimization mode is necessary to improve the approximation.

Before returning to the optimization mode, GMI and MIR cuts are constructed at  $(\mathbf{x}^K, \eta^K)$  and added to the feasible region (Higle and Suvrajeet 1999). As a result, the feasible region no longer contains the non-integer solution  $\mathbf{x}^K$ . Note that integer cuts are generated based on the original constraint set, as well as the affine lower bounding functions added so far. Particularly, GMI cuts are generated for fractional basic solutions to  $(\bar{M}_{lp}^K)$ . Therefore, multiple GMI cuts are generated up to iteration  $\kappa$  using fractional basic solutions. Let  $\mathcal{L}_{gmi}^\kappa$  denote the set of all GMI cuts generated in iteration  $\kappa$ . The MIR cuts are generated only for constraints with mixed-integer variables. In the SAA model, the variables used in the constraint set in the first-stage problem (1, 2) are integers. Therefore, MIR cuts are only used for the lower bounding affine functions in  $\mathcal{L}_{opt}^\kappa$ , which contain the continuous variable  $\eta$ . Moreover, the fact that recourse function  $H(\cdot)$  is lower bounded by zero almost surely plays a critical role in deriving the MIR inequalities. The set of MIR inequalities added in iteration  $\kappa$  is denoted by  $\mathcal{L}_{mir}^\kappa$ . Together, the updated set of integer cuts becomes:

$$\mathcal{L}_{int}^\kappa = \mathcal{L}_{int}^{\kappa-1} \cup \mathcal{L}_{gmi}^\kappa \cup \mathcal{L}_{mir}^\kappa. \quad (11)$$

This formulation completes the steps involved in the proposed algorithm's integer-feasibility mode, which results in an updated first-stage feasible region  $\bar{\mathcal{X}}^K$ . The algorithm returns to a new iteration in the optimization mode.

### 3.2 Statistical evaluation

Note that  $(\hat{\mathbf{x}}_S, \hat{v}_S)$  is the optimal solution-value pair to the SAA problem in formulation (9) when the sample size is  $S$ . However, the sample size required to ensure that  $\hat{\mathbf{x}}$  is an optimal solution to the true problem (6) is problem specific. Moreover, since the samples used to setup the SAA are chosen randomly through the Monte Carlo sampling method, the corresponding optimal objective function value is also a random quantity. Therefore, concluding that the solution obtained from a single SAA is an optimal solution to first-stage problem (6) may be erroneous. In order to overcome such premature conclusions on optimality, the experiments should be replicated with different sets of scenarios. The idea of using multiple replications for estimating the LBs and UBs, which are then used to obtain estimates on an optimality gap, was first proposed by Mak et al. (1999b). This idea has been incorporated successfully in the SP literature to provide statistical performance guarantees for 2-SLPs (Linderot et al. 2006). Here, a similar experimental setting is used for the SAA approach to develop a 2-SIP with continuous recourse.

For a given selection of  $S$ , let  $\Omega_S^r$  denote the set of scenarios generated, and let  $(\hat{\mathbf{x}}_S^r, \hat{v}_S^r)$  be the optimal solution pair obtained using the extended LS algorithm in replication  $r = 1, \dots, R$ . In this regard, an estimator of the optimal objective function value of the true problem (6) is given by  $\bar{v}_S^R = \frac{1}{R} \sum_{r=1}^R \hat{v}_S^r$ . Note that  $\hat{v}_S^r$  is an LB, and therefore  $\bar{v}_S^R$  is a biased estimator of the objective function value.

In each replication, the quality of solution  $\hat{\mathbf{x}}_S^r$  is evaluated by simulating the second-stage problem using scenarios different from the ones in  $\Omega_S^r$ . This is an out-of-sample evaluation process. The evaluation continues until a sample of size  $S'$  is found, such that the correspond-

**Algorithm 1** Pseudocode for one replication of the extended L-shaped algorithm.

---

```

1: Initialize  $\epsilon, k \leftarrow 0, \mathcal{L}_{int}^0 = \emptyset, \mathcal{L}_{opt}^0 = \emptyset$ 
2: Optimality mode
3:  $k \leftarrow k + 1$ 
4: Solve  $(M_{lp}^k)$  to obtain  $(\mathbf{x}^k, \boldsymbol{\eta}^k)$  with the objective function value of  $v^k$ 
5: for all  $i \in \{1, \dots, S\}$  do
6:   Solve the second-stage problem  $H(\mathbf{x}^k, \omega_i)$  to obtain  $\boldsymbol{\pi}^k$ 
7: end for
8: Compute  $\ell_{opt}^k(\mathbf{x}^k, \boldsymbol{\eta}^k)$  using formulation (10) and add to  $\mathcal{L}_{opt}^k$ 
9:  $\hat{F}_S(\mathbf{x}^k) = \ell_{opt}^k(\mathbf{x}^k, \boldsymbol{\eta}^k)$ 
10: if  $\Delta^k(\mathbf{x}^k) = (\hat{F}_S^k(\mathbf{x}^k) - \hat{v}^k) / \hat{F}^k(\mathbf{x}^k) < \epsilon$  then
11:   Go to integer-feasibility mode
12: else
13:    $\mathcal{L}_{int}^k = \mathcal{L}_{int}^{k-1}$ 
14: end if
15: Integer-feasibility mode
16: Set  $k = \kappa, \hat{\mathbf{x}}^\kappa = \mathbf{x}^\kappa, \hat{v}^\kappa = v^\kappa$ 
17: if  $\hat{\mathbf{x}}^\kappa \in \mathbb{Z}^+$  then
18:   STOP
19: else
20:   Solve  $(M_{mip}^\kappa)$  to obtain  $\hat{\mathbf{x}}_S^\kappa$  with the objective function value  $\hat{v}_S^\kappa$ 
21: end if
22: if  $\Delta^\kappa(\mathbf{x}^\kappa) = (\hat{F}_S^\kappa(\mathbf{x}^\kappa) - \hat{v}^\kappa) / \hat{F}^\kappa(\mathbf{x}^\kappa) < \epsilon$  then
23:   STOP
24: else
25:   Add GMI and MIR cuts,  $\mathcal{L}_{int}^k = \mathcal{L}_{int}^{k-1} \cup \mathcal{L}_{gmi}^k \cup \mathcal{L}_{mir}^k$ 
26:   Go to optimization mode
27: end if

```

---

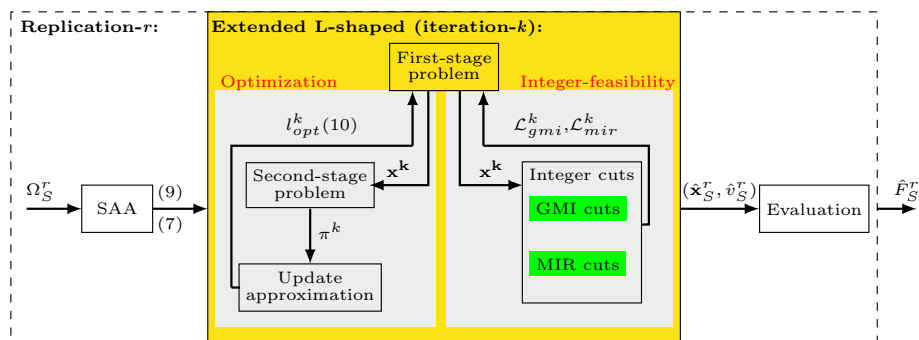
ing  $\hat{F}_{S'}^r$  is contained in the  $(1 - \alpha)$  confidence interval created using  $\hat{F}_s^r$  for  $s = 1, \dots, S'$ . Consider the estimator  $\Delta_r = \hat{F}_S^r(\hat{\mathbf{x}}_S^r) - \hat{v}_S^r$  of the optimality gap. The mean and variance of this estimator are given by:

$$\bar{\Delta}^R = \frac{1}{R} \sum_{r=1}^R \Delta^r, \quad \sigma_{\Delta}^2 = \frac{1}{R(R-1)} \sum_{r=1}^R (\Delta^r - \bar{\Delta}^R)^2. \quad (12)$$

Notice that  $\bar{\Delta}^R$  overestimates the optimality gap and therefore can be viewed as a pessimistic gap. Our computational study reports this gap and the distribution of  $\hat{F}_{S'}^r$  over the replications.

Next, the distinguishing features of the algorithmic framework presented here are compared to prior works.

1. If the LS method is implemented to solve the SAA, then the MIP first-stage problem presented in this study would be solved in every iteration of the algorithm. Instead, the algorithm proposed here solves a linear first-stage problem during the optimization mode, which significantly reduces its running time. The algorithm seeks an integer solution only during the integer-feasibility mode. The computational results in Table 4 illustrate that the proposed algorithm leads to lower running times.
2. The effectiveness of using integer cuts within the LS method has previously been explored in Bodur et al. (2016). The authors add GMI cuts in every iteration of the LS method. In addition to GMI cuts, the algorithm proposed here also exploits the impact of adding MIR cuts. These cuts are only added during the integer-feasibility mode, which reduces the computational time of the algorithm. The results in Table 4 indicate that using both



**Fig. 3** Schematic representation of the extended L-shaped algorithm

GMI and MIR cuts, compared to using GMI cuts only, results in finding solutions of higher quality.

- For the application considered in this paper, limited data is available to estimate the underlying distribution of the random variable involved. When this is the case, an evaluation process of the SP algorithms is required to validate and verify the quality of the solutions obtained (Mak et al. 1999b). While evaluation procedures have been developed for a 2-SLP (Linderoth et al. 2006), to the best of our knowledge, the work presented here reflects the first adoption of these procedures for 2-SIPs.

Figure 3 shows the schematic representation of the algorithmic framework developed in this study.

## 4 Computational experiments: a case study of Bangladesh

Instances of the vaccine administration and inventory replenishment problem were created with real-world data collected from different regions in Bangladesh. These instances were created to highlight the impact of data, deterministic as well as stochastic, on model decisions and to draw insights about the problem as it actually exists in the real world. The optimal solution of the SAA is used as a benchmark for the simple, easy-to-implement inventory replenishment and vaccine administration policies proposed here.

This section is organized as follows: First, the input data is presented; second, the experimental setup is explained; and third, the results of solving the SAA are analyzed. The inventory replenishment and vaccine administration policies derived from solving the SAA are referred to as the “base policy” to distinguish them from some easy-to-implement policies proposed at the end of this section.

### 4.1 Data input model

Two different data sources were used in this case study to generate problem instances for outreach sessions conducted by EPI in Bangladesh. Two regions, based on varying demographic data, were selected: urban Chittagong and rural Rajshahi. Deterministic data, including purchase and inventory holding costs, were compiled using the 2012 WHO vaccine volume calculator (WHO 2012). Stochastic data were compiled from Bangladesh's 2011 Demographic and Health Survey (DHS), a population-based survey executed by the National Institute for



**Table 1** Cost and weight of different vaccine vial sizes of Pentavalent

Vaccine size	Purchase cost per dose (\$)			Vaccine weight per dose (g)		
	Lower limit	Mean	Upper limit	Lower limit	Mean	Upper limit
1 dose	3.24	3.60	3.96	1.542	1.713	1.885
2 dose	3.25	3.50	3.75	1.723	1.914	2.106
10 dose	2.00	1.80	2.20	3.169	3.522	3.874

Population Research and Training (NIPORT 2011) that includes individual household-level data. The details of deterministic and stochastic parameters are described next.

#### 4.1.1 Deterministic parameters

EPI outreach sessions in rural areas are scheduled on a weekly basis (GAVI 2011). Thus, this study terms the planning horizon to include two weeks with five working days per week and two to eight working hours per day. Based on the proposed model, inventory replenishment decisions are made at the beginning of each week, and vaccine administration decisions are made hourly. Experiments for the Pentavalent vaccine, which is distributed in one-, two-, and ten-dose vials, are conducted, and this setup can apply to other pediatric vaccines as well. The Pentavalent vaccines has 6 h safe-use time before expiration (i.e.  $\tau = 6$ ). The 2012 WHO vaccine volume calculator (WHO 2012) provides data about vial purchasing costs and corresponding weights, which are summarized in Table 1. Inventory holding cost is estimated based on the weight of the Pentavalent vaccine. Sensitivity analyses, with respect to purchase and inventory holding costs, were conducted. The lower limits, means, and upper limits of the values are presented in Table 1. In all the other experiments conducted, the mean values of purchase and inventory holding costs are used. Order setup cost is assumed to be constant over the planning horizon and is set at \$10. This value is estimated based on interviews with health care workers who work in developing countries. The penalty cost associated with unserved patients is estimated to be \$100 and is based on a sensitivity analysis conducted as part of this study. A higher penalty value resulted in scalability issues, and a smaller value resulted in a large number of unserved patients. The selected penalty avoided scalability problems and provided realistic numbers of unserved patients. The UB of order size  $M_t$  is set equal to 2000 for all  $t$ . This value is large enough to capture vaccination shortages due to OVW only, not capacity. Finally, the penalty cost for OVW equals the vaccine's purchase cost.

#### 4.1.2 Stochastic parameters

The only stochastic element in the model presented here is patient arrival. In order to determine how many arrivals an outreach program might expect, this study uses the DHS data set that includes both demographic information and interviews with families from each region of Bangladesh. The interviews provide information about the vaccination history of children younger than 5. Each child's vaccination date is used to count the number of daily observations, or patient arrivals, over a 1-year time interval. Since the outreach sessions are held only on weekdays, only weekday data is considered. The data is used to estimate the distribution of the total number of patients arriving at an outreach session in one day. Goodness-of-fit

**Table 2** Data used to predict Pentavalent's daily demand distribution in different regions of Bangladesh

Region	Number of clinics	Infant population	Daily fitted distribution	Hourly estimated distribution
Barisal	41	8,147,000	NB(0.65, 0.43)	NB(5, 0.43)
Chittagong	141	288,079,000	NB(0.53, 0.81)	NB(286, 0.81)
Dhaka	171	46,729,000	NB(2.08, 0.16)	NB(2, 0.16)
Khulna	70	15,563,000	NB(0.95, 0.33)	NB(4, 0.33)
Rajshahi	182	18,329,000	NB(0.84, 0.31)	NB(2, 0.31)
Rangpur	43	15,665,000	NB(5, 0.93)	NB(327, 0.93)
Sylhet	62	9,807,000	NB(0.78, 0.38)	NB(5, 0.38)

tests were conducted to find the best distribution for this data, and summaries of results are listed in Table 2. Here NB refers to Negative Binomial distribution.

Since no information is available about the timing of vaccinations during the day, NB distribution is applied to estimate the total number of patients arriving within each hour. Furthermore, this work assumes that patients' hourly arrivals follow a uniform distribution. Such an assumption is not restrictive since the model does not penalize for patient waiting times. The last column of Table 2 shows the estimated arrival rates.

**Scenario selection** Selecting an appropriate number of scenarios for our SAA formulation is critical since generating a larger number of scenarios increases the running time of the algorithm, and generating a smaller number of scenarios may not appropriately represent uncertainty. Two different sets of analyses are used to assess the difference between the optimal solutions,  $\hat{\mathbf{x}}^S$  and  $\hat{\mathbf{x}}^{S'}$ , obtained in  $S$  and  $S'$  scenarios.

The first set of analyses captures the impact of the optimal solution on the recourse cost by fixing the optimal solution and simulating the recourse function over the two sets of scenarios. A paired  $t$  test on the function values is performed, and the corresponding  $p$  values for each region are presented in Table 3. A  $p$  value greater than 0.05 implies that the two solutions  $\hat{\mathbf{x}}^S$  and  $\hat{\mathbf{x}}^{S'}$  have a statistically indistinguishable impact on the recourse cost. The second set of analyses is based on the Euclidean norm of the two solutions. Table 3 shows the results. A high value indicates that the solutions are significantly different. As Table 3 shows,  $\hat{\mathbf{x}}^{1000}$  and  $\hat{\mathbf{x}}^{2000}$  have acceptable  $p$  values and relatively small Euclidean distances. For all regions considered in this study, increasing the number of scenarios from 1000 to 2000 is not necessary, and therefore, the number of scenarios used is  $S = 1000$ .

## 4.2 Experimental setup

The solution algorithm presented here was implemented in C programming language on a 64-bit Intel Core i5 processor @ 1.9 GHz with 8 GB RAM. All first-stage and second-stage problems were solved using CPLEX callable subroutines.

The experiments began by finding an optimal or near optimal first-stage solution (error gap  $< 0.01$ ) using the extended LS method. Next, a posterior analysis was conducted to evaluate the quality of the first-stage solutions by fixing the first-stage decisions and simulating second-stage problem (7). The scenarios used for optimization and evaluation were generated using the same distribution. Evaluation terminated when the standard deviation of the recourse function's value came within a tolerable limit. This process was replicated using independent

**Table 3** Paired  $t$  test between different number of scenarios

S; S'	$p$ value (95% CI)							$\ \hat{x}^S - \hat{x}^{S'}\  / \ \hat{x}^S\ $						
	Rajshahi	Barisal	Khuluna	Sylhet	Dhaka	Rangpur	Chittagong	Rajshahi	Barisal	Khuluna	Sylhet	Dhaka	Rangpur	Chittagong
100; 200	0.99	0.00	0.00	0.00	0.14	0.00	0.44	0.30	0.07	0.04	0.06	0.02	0.02	0.05
200; 500	0.99	0.84	0.29	0.80	0.03	0.08	0.01	0.33	0.05	0.01	0.05	0.03	0.02	0.05
500; 1000	0.99	0.67	0.89	0.35	0.48	0.00	0.96	0.22	0.05	0.03	0.05	0.02	0.02	0.02
1000; 2000	0.99	0.74	0.66	0.99	0.58	0.39	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00

samples. The experiments included  $R = 30$  replications. The results are presented in terms of the empirical distribution over replications, and heuristic policies are treated similarly.

### 4.3 Base policy

First, the performance of the solution algorithm and the model's adequacy must be verified. The inventory replenishment and vaccine administration policies identified by solving the extended LS method outlined in Sect. 3.1 are called the “base policy.” Results are provided for the base policy, and Table 4 compares the results obtained with those found with the classic LS method.

#### 4.3.1 Performance of solution algorithms

This set of experiments is implemented across the Chittagong dataset with vaccines of a single vial size, such as the ten-dose vials of Pentavalent. The goal is to evaluate the impact of MIR and GMI cuts in the performance of the extended LS method. The quality of solutions obtained under the following conditions are compared with the classic LS method: (a) use only MIR cuts (LSM), (b) use only GMI cuts (LSG), and (c) use both MIR and GMI cuts (LSMG). Table 4 summarizes the solution quality of each algorithm in terms of the CPU time, total number of iterations, LB, UB, the total number of LPs solved, and the total number of MILPs solved over 30 replications. Note that the values of LB and UB for all algorithms are similar, and the LS method uses the smallest number of iterations. However, in every iteration of the LS method, an MILP is solved, and therefore, the CPU time is significantly high. In algorithms that require integer cuts, MILPs are solved only when the error gap is below a threshold  $\epsilon = 0.1\%$  and  $\mathbf{x} \notin \mathbb{Z}^+$ . Therefore, the first-stage problem is solved as an LP in most iterations, particularly in the first iterations of the algorithm, which result in not only a higher number of iterations, but also in a shorter computational time.

Experimental results indicate that using MIR cuts takes less computational time than using GMI cuts. The greatest computational time savings, by a factor 6.2, was observed when both types of integer cuts were incorporated. Figure 4 illustrates the error gaps over algorithm iterations for one of the replications.

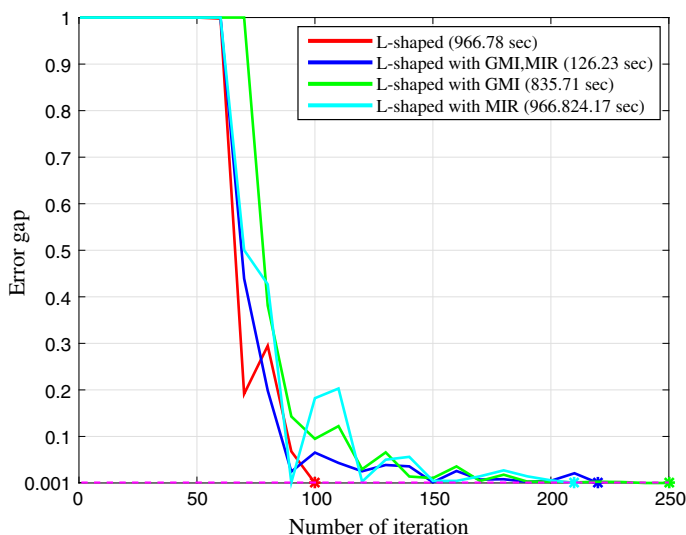
This analysis establishes that the LSMG algorithm provides solutions similar to those obtained from the LS method in significantly less CPU time. Such computational enhancements are necessary to satisfy the need to replicate SP models, which also allow the algorithm presented here to be applicable to real-world problem instances, including the one considered in our study.

#### 4.3.2 Analysis of parameters

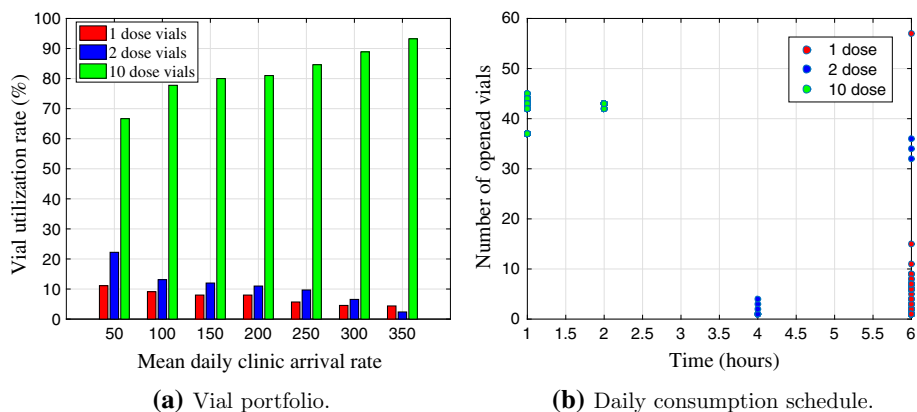
Solutions to the vaccine administration and inventory replenishment model are sensitive to the problem parameters, such as patient-arrival rate, purchase cost, and session duration. Analysis of these parameters helps health care policy makers to evaluate the effectiveness of outreach sessions and design appropriate policies. In the first set of experiments, the impact of different patient arrivals on vial portfolio selection is evaluated. The second experiment is developed to study the effects of varying purchase costs on OVW rates and the number of unserved patients. Finally, the impacts of session duration on OVW and the number of unserved patients are analyzed.

**Table 4** Comparison of solution quality

Algorithm	LB (\$)	UB (\$)	CPU time (s)	No. of iter.	No. of LP	No. of MILP
LS	21,571.7 ± 12.1	22,047.9 ± 33.8	859.7 ± 245.9	233.8 ± 11.2	0	232.7 ± 11.2
LSM	21,574.9 ± 11.8	22,048.9 ± 30.7	350.4 ± 86.1	836.3 ± 83.5	756.4 ± 79.3	78.9 ± 11.8
LSG	21,571.9 ± 12.1	22,047.1 ± 33.6	412.2 ± 160.1	837.9 ± 93.8	756.4 ± 88.5	80.5 ± 12.7
LSMG	21,571.7 ± 12.1	22,047.9 ± 33.9	137.2 ± 47.3	431.1 ± 26.3	335.6 ± 16.6	94.5 ± 12.4

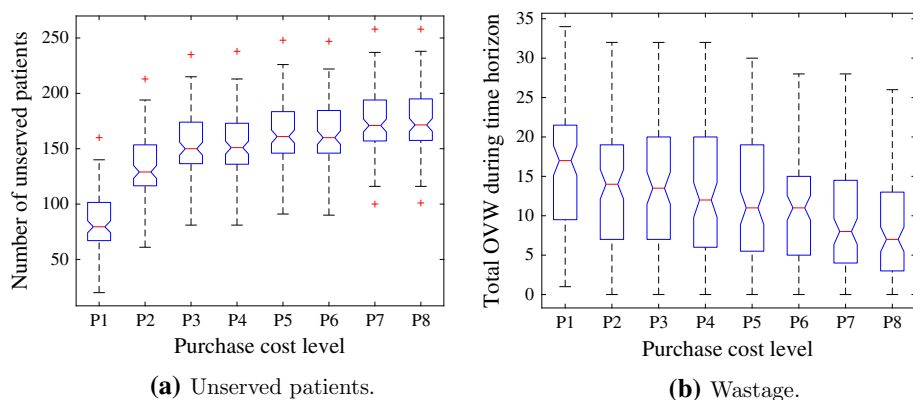


**Fig. 4** Base policy: analysis of solution algorithm



**Fig. 5** Base policy: analyzing the impact of mean daily patient-arrival rates on vial portfolios and the relationship between the daily consumption schedule and the number of opened vials

1. *Patient arrival* This experiment was conducted using regions with different patient-arrival rates. For example, the patient-arrival rate in Rajshahi is lowest since this is the least populated region with available data. The patient-arrival rate in Chittagong is highest, and it has the highest population. Figure 5a depicts the relationship among utilization rates of multi-dose vaccines of different sizes and patient-arrival rates. Going from Rajshahi to Chittagong, the arrival rate increases, as does the number of ten-dose vials ordered. This increase is attributed to the low cost per dose of multi-dose vials. For high patient-arrival rates, one-dose vials are ordered to supplement the multi-dose vials, particularly, for end-of-day arrivals. Based on the results presented in Fig. 5a, procuring larger size vials for moderately and highly populated regions is economically efficient. Likewise, the usage of one- and two-dose vials is higher in sessions with smaller patient-arrival rates, such as outreach centers in remote locations. Figure 5b shows the number of multi-dose



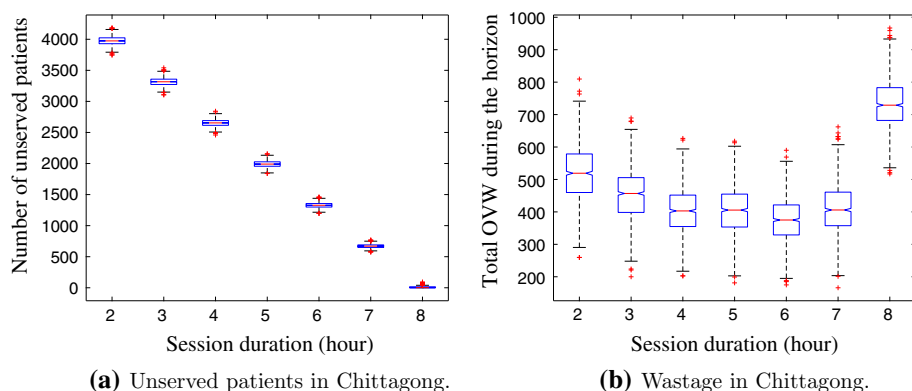
**Fig. 6** Effect of varying purchase costs on dose utilization

vials opened at different times over the duration of a session. The results indicate that multi-dose vials should be opened in the early hours of a session and one-dose vials for the final hours.

2. *Purchase cost* To conduct these experiments, eight problem instances, P1-P8, were developed. Each instance has the same patient-arrival rate as in Rajashahi, but each instance has a different purchase cost. Different values of purchase costs were used within the lower and upper limits presented in Table 1. Instance P1 has the lowest purchase costs, and costs increase from P1 to P8. Figure 6a, b present the number of unserved patients and OVW observed during the evaluation step of one of the replications. As purchase cost increases, vial-opening decisions become conservative to reduce OVW, which, in turn, increases the number of unserved patients. This analysis provides a tool to aid policy makers identifying the necessary subsidies for vaccine purchase costs and achieving a certain level of immunization in the region. For example, when purchase costs equal the lower limit in instance P1 ( $c_1 = 1.6$ ,  $c_2 = 1.5$ ,  $c_{10} = 1$ ), the total number of unserved patients during the two-week planning horizon and in all the regional outreach sessions is fewer than 100 for 99% of scenarios tested, as indicated by the top whisker. This number increased to 175 for P3-P4 and went up to 190 for P8.
3. *Session duration* The choice of session duration affects how many patients attend vaccination sessions. Since vaccines expire after their safe-use time, the choice of session duration also impacts OVW. To evaluate these impacts, problem instances for different regions were developed by examining session durations from 2 to 8 h. When session duration is less than 8 h, patients who cannot attend the session during its operational hours are lost or unserved. As a result, a shorter session duration increases the number of unserved patients; see Fig. 7a. The results indicate that OVW reaches its minimum value when the session duration equals the vaccine's safe-use time ( $\tau = 6$  h). When the session duration is short, unexpired doses also contribute to OVW. While increasing the session duration decreases the number of unserved patients, the need to open new vials to serve patients arriving in later hours also impacts OVW. Note that increasing session duration also increases operational costs, which are not considered in the model presented here.

Based on this work's analysis of the base policy, the usage of multi-dose vials with complementary one-dose vials for use in last operating hours, when  $t > \tau$ , is recommended for highly populated and well connected regions, like Chittagong. Demographic data about





**Fig. 7** Impacts of session duration in Chittagong on OVW rates and the number of unserved patients

Chittagong indicates that the majority of the population is employed in jobs that do not provide flexible working hours, and therefore, holding longer outreach sessions increases the number of vaccinated patients. Unlike Chittagong, Rajshahi is mostly rural, and the majority of its population works in the agricultural sector (Rahedul and Hassan 2011). This provides mothers with greater flexibility to attend outreach sessions. Thus, having a six-hour session duration in this location is appropriate to achieve high vaccination rates and low OVW.

The model presented here and the subsequent analyses target developing countries, and, hence, purchase cost is a critical factor in inventory replenishment and vaccine administration decisions. Therefore, health care policy makers should negotiate the necessary subsidies to achieve the target vaccination coverage levels in a cost-efficient manner. The setup in this paper provides not only a statistically optimal policy, but also a systematic tool to conduct analyses that support decisions related to vaccination in developing countries.

#### 4.4 Heuristic policies

Experimental results presented in Sect. 4.3 indicate that solving the proposed inventory replenishment and vaccine administration model to develop a base policy requires extensive use of engineering tools that are commercially expensive and are rarely accessible to health care administrators in developing countries. In order to address these concerns, this study develops and evaluates four simple, easy-to-use policies based on the base policy. These policies are: (a) first multi-dose, last single-dose (FMLS) policy, (b) shorter duration session (SDS) policy, (c) single vial size policy, and (d) wait to open (WO) policy. These policies impact both vaccine order replenishment and administration. To implement these policies, the first- and second-stage problem (6) and (7) developed above are modified and described below:

- (a) *FMLS* The principle of this policy follows the observation made in Fig. 5b regarding the use of one-dose and multi-dose vials: Health care practitioners should open multi-dose vials in the early hours of the session and one-dose vials as the session comes to an end. To implement this approach, the session length is divided into three equal intervals, such that, for  $t \in \mathcal{T}$  (a) ten-dose vials are opened during the first 2 h of a session ( $\mathcal{N}_1 = \{Nt - 5, Nt - 4\}$ ); (b) two-dose vials are opened during the next 2 h ( $\mathcal{N}_2 = \{Nt - 3, Nt - 2\}$ ); and (c) one-dose vials are opened in the last 2 h ( $\mathcal{N}_3 = \{Nt - 1, Nt\}$ ).

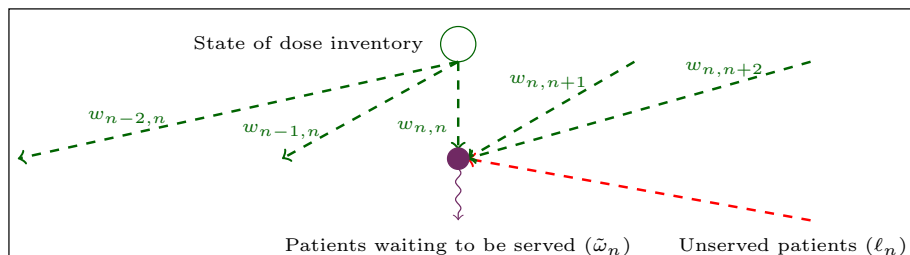
Mathematically, this approach corresponds to setting  $u_{vn} = 0$  for  $n \in \mathcal{N} \setminus \mathcal{N}_i$  when  $q_v = i$  in first-stage problem (6). For example, for time periods  $n \in \mathcal{N}_3$ , when only ten-dose vials are opened,  $u_{vn} = 0$  if  $q_v = 1$  or 2.

- (b) *SDS* Results in Fig. 7 indicate that sessions lasting 6 h have the minimum OVW, but longer sessions have the lowest number of unserved patients. The SDS policy achieves low OVW by trimming the duration of the immunization session from 8 to 6 h. The model presented here assumes that only a certain percentage of patients, who would be arriving during the last 2 h of a session, are attending the immunization session earlier in the day. The remaining patients are not served.
- (c) *Single vial size* Current vaccination practice relies on the use of single multi-dose vials (Dhamodharan and Proano 2012; Guichard et al. 2010). This policy can highlight the advantage of using a combination of multi-dose vials. The single multi-dose vial policy is implemented using one-, two, or ten-dose vials and adjusting set  $\mathcal{V}$  in the first- and second-stage problems.
- (d) *WO* Vaccine administration is commonly delayed until a sufficient number of patients have arrived at a session (Bosu et al. 1997). Such a practice reduces OVW, but if the waiting time is too long, patients may leave without being vaccinated, which adversely effects immunization coverage. In order to study the effects of this policy, an extension of our model is presented. This extension captures the impacts of patient waiting times on vaccine vial utilization decisions and considers the trade-offs among OVW, unserved patients, and waiting times.

Including patient waiting times affects only the second-stage problem in formulation (7). To capture this difference, a new variable must be introduced. Let  $w_{jn}$  be the number of patients who arrived in period  $j$  and were served in period  $n$ . Thus,  $j \leq n < j + L$ , where  $L$  is the maximum waiting time in hours. Patients not served within this time window leave the session. A non-linear function  $v(j - n)$  for the total patient waiting time is introduced, and a unit cost  $v$  is used to estimate the total cost of waiting and not serving patients. Figure 8 illustrates how vaccines are administered to patients arriving in time period  $n$ . Mathematically, flow-balance equation (4) changes to include the following constraints:

$$\sum_{m=n-\tau+1}^n y_{mn} = \sum_{j=n-L+1}^n w_{jn} \quad \forall n \in \mathcal{N}, \quad (13a)$$

$$\sum_{j=n}^{n+L-1} w_{nj} + \ell_n = \tilde{\omega} \quad \forall n \in \mathcal{N}. \quad (13b)$$



**Fig. 8** Dose utilization and patients waiting for  $n = 2$  with  $L = 2$

The left-hand side of constraints (13a) represents the inventory of doses immediately available, and the right-hand side represents the total number of patients that arrived within the last  $L$  periods. This inventory contains doses from vials opened during the last  $\tau$  periods. For example, when  $L = 2$ , the doses available in period  $n$  are used to vaccinate patients who arrived in periods  $(n - 2)$ ,  $(n - 1)$ , and  $n$ , as shown in Fig. 8. Patients who arrive in the current time period,  $\tilde{\omega}$ , can be served within the next  $L$  hours. Patients waiting longer than  $L$  hours leave the session, they are represented by the terms on the left-hand side of Eq. (13b) and represented by the incoming arcs in Fig. 8. Now, recourse function  $H(\mathbf{x}, \omega)$  can be rewritten as follows:

$$\begin{aligned}
 H(\mathbf{x}, \omega) = \min \quad & \sum_{n \in \mathcal{N}} (gy_{n(n+\tau)} + p\ell_n + \sum_{j=n+1}^{n+L} v(j-n)w_{nj}) \\
 \text{s.t. } & (3), (13) \\
 & y_{nm}, \ell_n, w_{nj} \in \mathbb{Z}^+ \quad \forall n \in \mathcal{N}, m \in \{n, \dots, n + \tau + 1\}, \\
 & j \in \{n, \dots, n + L\}.
 \end{aligned} \tag{14}$$

In these experiments, the maximum waiting time is set to  $L = 2$  hours, and the penalty for waiting to  $v = 0.24\$/\text{h}$ , the average hourly wage in Bangladesh.

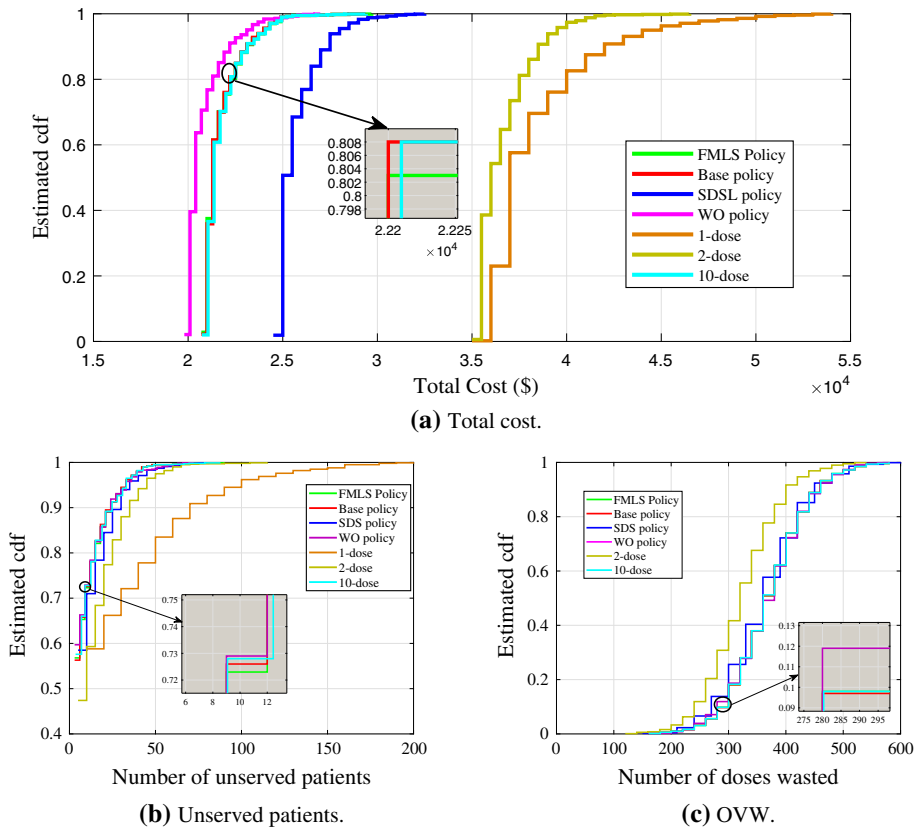
To implement this policy, we solve the 2-SIP with (6) as the first-stage problem and (14), instead of (7), as the second-stage problem. The experimental setup described in Sect. 4.2 is adopted to experiment with this policy.

#### 4.4.1 Policy analysis and comparison

The performance of the policies presented is compared based on the total costs, OVW, and number of unserved patients. Figure 9 illustrates the cumulative distribution function (CMF) of the average total cost, number of unserved patients, and OVW over 30 replications.

1. *FMLS* Results in Fig. 9a indicate that the probability of achieving a certain cost target with FMLS policy is only slightly lower than the base policy. For example, the probability that the total cost is less than \$22,200 is 0.81 for base policy and 0.8 for FMLS policy. Moreover, the results in Fig. 9b, c indicate that the number of unserved patients and OVW are comparable to the base policy, which can be attributed to the similarity in vial-opening decisions under base policy and the vial-opening windows set under FMLS policy; see Fig. 5b. However, FMLS policy mandates using one-dose vials even when multiple patients arrive in the final time window, although using multi-dose vials is more economical. In summary, FMLS policy is easy to implement and fairly economical.
2. *SDS* SDS policy increases the effective arrival rate in the six-hour time window by allowing a fraction of patients, originally scheduled to attend in the last seventh and eighth hours of the initial eight-hour session, to attend a session during the first 6 h. This increased arrival rate increases the number of vaccines used within the first 6 h, but it has only a marginal impact on OVW and the fraction of patients left unserved. The increase in the number of vaccines used and the number of unserved patients results in a higher cost, which is demonstrated in Fig. 9a.

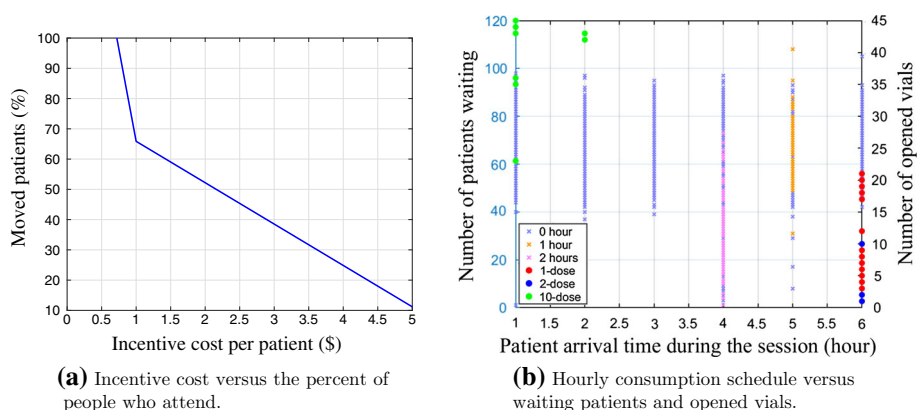
In this experimental setup with SDS policy, assume that an incentive of \$1 was provided to patients willing to be rescheduled earlier in the day. Assume that 65% of patients are willing to be rescheduled. This percentage could be higher if the incentive pay is bigger, which would result in an increase of vaccination coverage. The incentive pay of



**Fig. 9** Comparison of heuristic policies

\$1 is expected to cover the cost of a snack and drink. This approach is motivated by an experience in Nicaragua, where, in 1985, the attendance in immunization sessions increased from 77 to 94% because food was served to patients waiting for vaccinations (Loevinsohn and Loevinsohn 1987). Using such an incentive makes economical sense when the total amount paid is smaller than the penalty associated with unserved patients. Experiments were repeated to estimate the penalty from unserved patients as a function of the percentage of patients willing to move their appointments to earlier in the day. Figure 10a shows the relationship between the fraction of patients who rescheduled and the corresponding unit incentive costs, which are calculated based on the penalty costs. For example, an incentive pay of \$4 is economical if at least 20% of the patients are incentivized to reschedule.

3. *Single vial size* If a region must use a single multi-dose vial, then the choice of vial size should depend on the patient-arrival rate. In regions with higher patient-arrival rates, ten-dose vials are cost-efficient. However, in regions with lower patient-arrival rates, one- and two-dose vials are more economical than using ten-dose vials, since the latter results in higher unused doses, which contribute to OVW. Figure 9 summarizes these observations for a region with a high patient-arrival rate, like Chittagong. The results in Fig. 9b show that, with a probability of 0.95, the total number of unserved patients, for the duration of the planning horizon, is as high as 45 when two-dose vials are utilized. This number is



**Fig. 10** SDS and WO analysis results

smaller when ten-dose vials are used instead. The use of vials of small size reduces OVW; it is in fact 0 for the one-dose policy. Since the penalty for letting a patient go unserved is higher than the penalty for OVW, the total cost for the ten-dose policy is lower than the two-dose policy. In fact, the probability that the total cost is smaller than \$22,000 is 0.7 for the ten-dose policy and 0.81 for the optimal policy. These results indicate that a policy that relies only on the use of ten-dose vials is near optimal for this region.

4. **WO** This policy, by requesting that patients wait until a multi-dose vial is opened, contributes to reducing OVW and associated costs. WO policy reduces the number of unserved patients significantly, although a few patients do leave if the waiting times are too long; see Fig. 9b. This policy makes the utilization of larger-sized vials economical, as seen in Fig. 9c. For these reasons, the total cost of WO policy is smallest; see Fig. 9a. Moreover, Fig. 10b shows that WO policy, in an effort to reduce the total waiting time, recommends health care practitioners use ten-dose vials for patients arriving in the early hours of the session, and smaller-sized vials near the end.

In order to compare the proposed policies with the base policy presented in this paper, a two-sample  $t$  test is conducted on the results obtained over 30 replications. These results are summarized in Table 5. The second column of the table presents the estimated differences in total costs between the base and heuristic policies. In this table, the negative numbers indicate that the base policy has a lower total cost. The table also presents the 95% confidence interval (CI) of the difference in objective function values and the corresponding  $p$  values. A  $p$  value greater than 0.05 indicates that the null hypothesis, which states that the objective function values are statistically indistinguishable from one another, cannot be rejected at a 95% confidence level. These results indicate that WO is the only policy that has a total cost lower than the base policy. Moreover, the  $p$  values indicate that the total cost of FMLS policy is statistically indistinguishable from the total cost of the base policy. This is further corroborated by the fact that 95% CI for the estimated difference includes zero. However, the null hypothesis is rejected for other policies, meaning that they are significantly different from the base policy.

Our analysis suggests that the WO policy should be implemented during outreach sessions because it achieves the highest coverage levels with the lowest corresponding costs. However, implementation of this policy must take into consideration the population's economic and demographic characteristics, which impact patients' willingness to wait. The success of the

**Table 5** Two-sample  $t$  tests for differences between the base and heuristic policies in Chittagong

Two-sample $t$ test	Estimated difference (%)	Minimum difference (%)	Maximum difference (%)	95% CI for difference	$p$ value
Base, FMLS	− 0.2	− 0.4	− 0.1	[− 26.4, 0]	0.19
Base, SDS	− 18.2	− 19.1	− 17.6	[− 3,946, − 3,892]	0.000
Base, WO	0.2	0.1	0.3	[26.9, 58.9]	0.00
Base, 1 dose	− 73.5	− 74.1	− 73.2	[− 15,861, − 15,819]	0.00
Base, 2 dose	− 68.9	− 69.5	− 68.7	[− 14,876, − 14,834]	0.00
Base, 10 dose	− 0.1	− 0.5	0.0	[− 46.9, − 6.0]	0.01

FMLS policy applied depends on selecting the appropriate time periods to open differently sized vials. The ease of implementing FMLS policy can be enhanced by combining the FMLS and WO policies. Vaccine coverage rates are greatest under the SDS policy, which thrives with appropriate incentive programs. The additional costs incurred from these incentives are justified when the tangible benefits of disease eradication are considered and realized. Finally, in regions where vaccines are distributed in a single multi-vial size, choosing the right size is important. Factors to weigh include population size, birth rate, and the number of clinics that organize outreach sessions in the region.

## 5 Conclusion and future research

This paper presents a 2-SIP model for inventory replenishment and administration of childhood vaccines in targeted immunization outreach sessions. To the best of our knowledge, this paper presents the first stochastic optimization model developed that captures the relationships that exist among these decisions. The model presented here minimizes replenishment and OVW costs. Unlike related works in the literature and the current practice that relies on the use of a single multi-dose vial, this paper models the performance of an inventory replenishment policy that allows a combination of multi-dose vials for vaccination. An extensive numerical analysis is conducted to evaluate the performance of the proposed policy and to compare to other simple-to-implement vaccine administration policies.

In order to solve the proposed 2-SIP, the LS method is extended by incorporating GMI and MRI cuts in the first-stage problem. An extensive numerical study shows that the proposed algorithm is scalable; it outperforms the LS method by providing high quality solutions in a much shorter CPU time.

After developing a case study using real-world data from Bangladesh, a sensitivity analysis is conducted to evaluate the system's behavior. Three main observations are summarized as follows:

1. Population size impacts decisions about the combination of multi-dose vials to use in a region. The use of multi-dose with complementary single-dose vials is recommended in highly populated regions.
2. Vaccine purchasing costs impact the achievable immunization levels within a given budget limit. Thus, the models presented here aid policy makers in negotiating the subsidies necessary to achieve the targeted vaccination coverage levels.

3. Session length impacts replenishment costs, the number of patients served, and OVW. Session lengths that are the same length as the vaccine's safe-use time do minimize the total cost. Short sessions in sparsely populated regions also minimize costs and OVW.

These observations motivated the design of vaccine administration policies that are simple and economical. Numerical results demonstrate that the WO policy has the lowest total costs and, therefore, is highly recommended. For highly populated and well-connected regions, FMLS policy works well since it provides high vaccine coverage levels for lower costs. Moreover, in regions that can only use single multi-dose vials, the decision about the size of a vial to use should be based on population size, birth rate, and the number of clinics in the region (see Table 5).

This research can be extended in the following ways. First, since no clear guidelines determine how many scenarios to use, investigating applications of sequential sampling algorithms, such as the two-stage stochastic decomposition (SD) method (Higle and Sen 1991), is necessary. The SD method was originally designed for 2-SLPs and does not require a priori selection of scenarios. Since the model presented here is an MILP, this method can be extended to accommodate discrete decision variables in the first stage. Second, the proposed model identifies inventory replenishment decisions for a single outreach session, so the model can be extended to consider multiple sessions organized by the same clinic, as well as multiple clinics within a region. These clinics will coordinate their own decisions about inventory and operating hours to minimize costs and OVW. Third, the proposed model can be extended to aid replenishment decisions in clinics that handle different types of vaccines with different safe-use times, such as liquids, with twenty eight days, or lyophilized with 6 h.

## Appendix A: Two-stage stochastic programming model

See Table 6.

$$\min \sum_{t \in \mathcal{T}} \left( f_t z_t + \sum_{v \in \mathcal{V}} c_v r_{vt} \right) + \sum_{v \in \mathcal{V}} \sum_{n \in \mathcal{N}} d_v s_{vn} + \mathbb{E}\{H(x, \omega)\}, \quad (15a)$$

s.t.

$$s_{vNt} = s_{v(Nt-1)} + r_{vt} - u_{vNt} \quad \forall t \in \mathcal{T}, \quad (15b)$$

$$s_{vn} = s_{vn-1} - u_{vn} \quad \forall n \in \mathcal{N} \setminus \{N, 2N, \dots, TN\}, \quad (15c)$$

$$\sum_{v \in \mathcal{V}} r_{vt} \leq M_t z_t \quad \forall t \in \mathcal{T}, \quad (15d)$$

$$z_t \in \{0, 1\}; s_{vn}, u_{vn}, r_{vt} \in \mathbb{Z}^+ \quad \forall t \in \mathcal{T}, v \in \mathcal{V}, n \in \mathcal{N}, \quad (15e)$$

where,

$$H(x, \omega) = \min \sum_{n \in \mathcal{N}} (g y_{n(n+\tau)} + p \ell_n) \quad (16a)$$

s.t.

$$\sum_{m=n}^{n+\tau-1} y_{nm} + y_{n(n+\tau)} = \sum_{v \in \mathcal{V}} q_v u_{vn} \quad \forall n \in \mathcal{N}, \quad (16b)$$



**Table 6** Decision variables and parameters

<i>Set of indices</i>	
Set	Description
$\mathcal{T}$	Set of ordering decision epochs
$\mathcal{N}$	Set of consumption decision epochs
$\mathcal{V}$	Set of vials of different sizes
<i>Parameters</i>	
Parameter	Description
$N$	Number of consumption decision epochs within an ordering period
$T$	Number of ordering decision epochs
$f_t$	Ordering fixed cost at time period $t$
$c_{vt}$	Variable purchase cost of vial size $v$ at time period $t$
$d_{vt}$	Unit inventory holding cost of vial size $v$ at time period $t$
$q_v$	Number of doses in vial $v$
$\tilde{\omega}$	Random patient arrival
$M_t$	Limit on the number of vials ordered in time period $t$
$\tau$	Safe open vial use time in periods
$g$	Unit wastage cost
$p$	Unit penalty cost of an unserved patient
<i>Decision variables</i>	
Variable	Description
$z_t$	A binary decision variable that takes 1 if an order is placed at time period $t$ and takes 0 otherwise
$r_{vt}$	Replenishment quantity for vial size $v$ at time period $t$
$u_{vn}$	Number of vials of size $v$ to open at time period $n$
$s_{vn}$	Number of vials of size $v$ in the inventory at time period $n$
$y_{nm}$	Number of doses obtained from vials opened in period $n$ , and used in period $m$

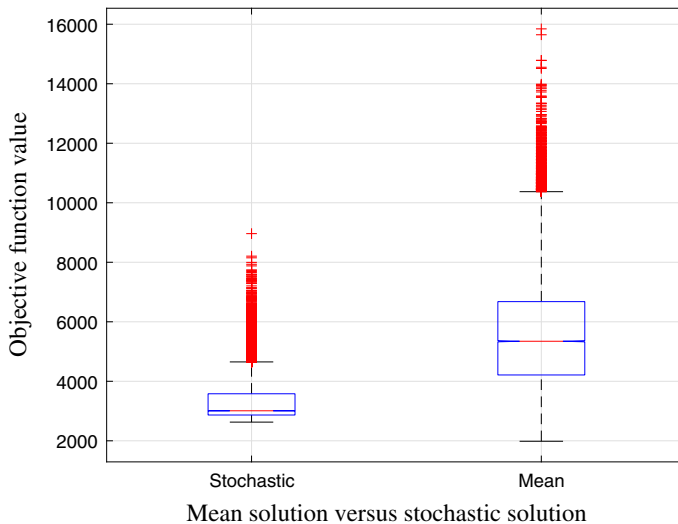
$$\sum_{m=n-\tau+1}^n y_{mn} + \ell_n = \omega \quad \forall n \in \mathcal{N}, \quad (16c)$$

$$y_{mn}, \ell_n \in \mathbb{Z}^+ \quad \forall m, n \in \mathcal{N}. \quad (16d)$$

## Appendix B: Performance evaluation of stochastic solutions

This section presents the performance evaluation of stochastic solutions obtained from solving the 2-SIP with first-stage problem (6) and second-stage problem (7). To achieve this goal, metrics, such as expected value of perfect information (EVPI) and the value of stochastic solution (VSS), are computed. EVPI and VSS are obtained for a problem instance created with data from the Barisal region. The number of scenarios generated for these experiments is  $S = 1000$ .

EVPI measures the price one is willing to pay to gain access to perfect information, and EVPI uses the difference between the objective function value of the wait-and-see and here-



**Fig. 11** The value of the stochastic solution

and-now problems. To calculate the objective function value of the wait-and-see problem, the SAA problem in formulation (9) is solved separately for each single scenario. Next, the corresponding expected value is calculated over the scenarios generated. The here-and-now problem is indeed the SAA problem in formulation (9). Then,  $EVPI = \$3300 - \$2870 = \$428$ . This means, if the number of patients arriving in a session is known, the total costs would only be \$2870. Thus, the cost of not knowing the future is  $EVPI = \$428$ .

VSS measures the impact of random patient arrivals has on the performance of the system (Birge and Francois 2011). VSS is the difference between the objective function value of SAA problem in formulation (9) and the deterministic mean value problem. To compare the objective function values of SAA and deterministic mean value problem, both problems are initially solved for a given set of scenarios. Next, the values of the first-stage solutions are fixed, and second-stage problem (7) is simulated over a different set of scenarios. The corresponding objective function values are reported in Fig. 11. Fewer variations in the objective function values obtained from the stochastic solution are observed. Also, the average objective function value of the stochastic solution is lower than the solutions to the mean value problem. These results indicate that solving the proposed stochastic model formulation, rather than the corresponding mean value problem, is valuable.

## References

- Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1988). *Network flows*. Tech. rep., DTIC Document.
- Assi, T.-M., Brown, S. T., Djibo, A., Norman, B. A., Rajgopal, J., Welling, J. S., et al. (2011). Impact of changing the measles vaccine vial size on Niger's vaccine supply chain: A computational model. *BMC Public Health*, 11(1), 425.
- Assi, T.-M., Brown, S. T., Kone, S., Norman, B. A., Djibo, A., Connor, D. L., et al. (2013). Removing the regional level from the Niger vaccine supply chain. *Vaccine*, 31(26), 2828–2834.
- Atan, Z., & Rousseau, M. (2016). Inventory optimization for perishables subject to supply disruptions. *Optimization Letters*, 10(1), 89–108.
- Bakker, M., Riezebos, J., & Teunter, R. H. (2012). Review of inventory systems with deterioration since 2001. *European Journal of Operational Research*, 221(2), 275–284.
- Birge, J. R., & Francois, L. (2011). *Introduction to stochastic programming*. Berlin: Springer.

- Bodur, M., Dash, S., Günlük, O., & Luedtke, J. (2016). Strengthened benders cuts for stochastic integer programs with continuous recourse. *INFORMS Journal on Computing*, 29(1), 77–91.
- Bosu, W. K., Ahelegbe, D., Edum-Fotwe, E., Bainsan, K. A., & Turkson, P. K. (1997). Factors influencing attendance to immunization sessions for children in a rural district of Ghana. *Acta Tropica*, 68(3), 259–267.
- Brown, S. T., Schreiber, B., Cakouros, B. E., Wateska, A. R., Dicko, H. M., Connor, D. L., et al. (2014). The benefits of redesigning Benin's vaccine supply chain. *Vaccine*, 32(32), 4097–4103.
- Chen, S.-I., Norman, B. A., Rajgopal, J., Assi, T. M., Lee, B. Y., & Brown, S. T. (2014a). A planning model for the WHO-EPI vaccine distribution network in developing countries. *IEEE Transactions*, 46(8), 853–865.
- Chen, X., Pang, Z., & Pan, L. (2014b). Coordinating inventory control and pricing strategies for perishable products. *Operations Research*, 62(2), 284–300.
- Dantzig, G. B., & Wolfe, P. (1960). Decomposition principle for linear programs. *Operations Research*, 8(1), 101–111.
- Dhamodharan, A., & Proano, R. A. (2012). Determining the optimal vaccine vial size in developing countries: A Monte Carlo simulation approach. *Healthcare Management Science*, 15(3), 188–196.
- Drain, P. K., Nelson, C. M., & Lloyd, J. S. (2003). Single-dose versus multi-dose vaccine vials for immunization programmes in developing countries. *Bulletin of the World Health Organization*, 81(10), 726–731.
- Edirisinghe, N. C. P., & Patterson, E. I. (2007). Multi-period stochastic portfolio optimization: Block-separable decomposition. *Annals of Operations Research*, 152(1), 367–394.
- Eksiöğlu, S. D., & Jin, M. (2006). Cross-facility production and transportation planning problem with perishable inventory. In *Computational Science and Its Applications-ICCSA 2006* (pp. 708–717). Springer.
- Fries, B. E. (1975). Optimal ordering policy for a perishable commodity with fixed lifetime. *Operations Research*, 23(1), 46–61.
- GAVI. (2011). *Comprehensive multi-year plan of the national immunization program of Bangladesh 2011–2016*. Tech. rep., GAVI.
- Guichard, S., Hymnbaugh, K., Burkholder, B., Diorditsa, S., Navarro, C., Ahmed, S., et al. (2010). Vaccine wastage in Bangladesh. *Vaccine*, 28(3), 858–863.
- Haijema, R., van Dijk, N., van der Wal, J., & Sibinga, C. S. (2009). Blood platelet production with breaks: Optimization by SDP and simulation. *International Journal of Production Economics*, 121(2), 464–473.
- Higle, J. L., & Sen, S. (1991). Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16(3), 650–669.
- Higle, J. L., & Suvrajeet, S. (1999). *Stochastic decomposition: A statistical method for large scale stochastic linear programming*. Hoboken: Wiley.
- Hjorring, C., & Holt, J. (1999). New optimality cuts for a single-vehicle stochastic routing problem. *Annals of Operations Research*, 86, 569–584.
- Hochbaum, D. S., & Segev, A. (1989). Analysis of a flow problem with fixed charges. *Networks*, 19(3), 291–312.
- Laporte, G., & Louveaux, F. V. (1993). The integer l-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3), 133–142.
- Lee, B. Y., Assi, T.-M., Rookkapan, K., Connor, D. L., Rajgopal, J., Sornsrivichai, V., et al. (2011). Replacing the measles ten-dose vaccine presentation with the single-dose presentation in Thailand. *Vaccine*, 29(21), 3811–3817.
- Lee, B. Y., Norman, B. A., Assi, T.-M., Chen, S.-I., Bailey, R. R., Rajgopal, J., et al. (2010). Single versus multi-dose vaccine vials: An economic computational model. *Vaccine*, 28(32), 5292–5300.
- Linderöth, J., Shapiro, A., & Wright, S. (2006). The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142(1), 215–241. <https://doi.org/10.1007/s10479-006-6169-8>.
- Liu, L., & Lian, Z. (1999). (s, s) continuous review models for products with fixed lifetimes. *Operations Research*, 47(1), 150–158.
- Loevinsohn, B. E. N. J. A. M. I. N. P., & Loevinsohn, M. I. C. H. A. E. L. E. (1987). Well child clinics and mass vaccination campaigns: An evaluation of strategies for improving the coverage of primary health care in a developing country. *American Journal of Public Health*, 77(11), 1407–1411.
- Mak, W.-K., Morton, D. P., & Wood, R. K. (1999a). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1–2), 47–56.
- Mak, W. K., Morton, D. P., & Wood, K. (1999b). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1), 47–56. [https://doi.org/10.1016/S0167-6377\(98\)00054-6](https://doi.org/10.1016/S0167-6377(98)00054-6).
- Mofrad, M. H., Garcia, G.-G. P., Maillart, L. M., Norman, B. A., & Rajgopal, J. (2016). Customizing immunization clinic operations to minimize open vial waste. *Socio-Economic Planning Sciences*, 54, 1–17.

- Mofrad, M. H., Maillart, L. M., Norman, B. A., & Rajgopal, J. (2014). Dynamically optimizing the administration of vaccines from multi-dose vials. *IIE Transactions*, 46(7), 623–635.
- Nahmias, S. (1975). Optimal ordering policies for perishable inventory—II. *Operations Research*, 23(4), 735–749.
- Nahmias, S. (1982). Perishable inventory theory: A review. *Operations Research*, 30(4), 680–708.
- Nahmias, S., & Pierskalla, W. P. (1973). Optimal ordering policies for a product that perishes in two periods subject to stochastic demand. *Naval Research Logistics Quarterly*, 20(2), 207–229.
- Nahmias, S., & Wang, S. S. (1979). A heuristic lot size reorder point model for decaying inventories. *Management Science*, 25(1), 90–97.
- NIPORT. (2011). *Bangladesh demographic and health survey 2011*. Tech. rep., NIPORT National Institute of Population Research and Training, Associates for Community and Population Research (ACPR), and ICF International.
- Palak, G. (2013). *Optimization models for cost efficient and environmentally friendly supply chain management*. Mississippi State University.
- Palanivel, M., Priyan, S., & Mala, P. (2018). Two-warehouse system for non-instantaneous deterioration products with promotional effort and inflation over a finite time horizon. *Journal of Industrial Engineering International*, 14(3), 603–612.
- Parmar, D., Baruwā, E. M., Zuber, P., & Kone, S. (2010). Impact of wastage on single and multi-dose vaccine vials: Implications for introducing pneumococcal vaccines in developing countries. *Human Vaccines*, 6(3), 270–278.
- Patriksson, M., Strömberg, A.-B., & Wojciechowski, A. (2015). The stochastic opportunistic replacement problem, part II: A two-stage solution approach. *Annals of Operations Research*, 224(1), 51–75.
- Raafat, F. (1991). Survey of literature on continuously deteriorating inventory models. *Journal of the Operational Research Society*, 42, 27–37.
- Rahedul, I. M., & Hassan, M. Z. (2011). Land use changing pattern and challenges for agricultural land: A study on Rajshahi district. *Journal of Life and Earth Science*, 6, 69–74.
- Rei, W., Cordeau, J.-F., Gendreau, M., & Soriano, P. (2009). Accelerating benders decomposition by local branching. *INFORMS Journal on Computing*, 21(2), 333–345.
- Rockafellar, R. T., & Wets, R. J.-B. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1), 119–147.
- Sen, S. (2011). Stochastic mixed-integer programming algorithms: Beyond benders' decomposition. *Wiley Encyclopedia of Operations Research and Management Science*.
- Subaiya, S., Lydon P., Gacic-Dobo, M., Eggers, R., Conklin, L., & Dumolard, L. (2015). *Global routine vaccination coverage, 2014*. Tech. rep., Centers for Disease Control and Prevention.
- Van Donselaar, K. H., & Broekmeulen, R. A. C. M. (2012). Approximations for the relative outdating of perishable products by combining stochastic modeling, simulation and regression modeling. *International Journal of Production Economics*, 140(2), 660–669.
- Van Slyke, R. M., & Wets, R. (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4), 638–663.
- Van Zyl, G. J. J. (1963). *Inventory control for perishable commodities*. Ph.D. thesis, University of North Carolina at Chapel Hill.
- Watson, J.-P., & Woodruff, D. L. (2011). Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science*, 8(4), 355–370.
- Weiss, H. J. (1980). Optimal ordering policies for continuous review perishable inventory models. *Operations Research*, 28(2), 365–374.
- WHO. (2005). *Monitoring vaccine wastage at country level: Guidelines for programme managers*. Geneva: WHO World Health Organization
- WHO. (2012). *Vaccine volume calculator*; 2009. Tech. rep., Geneva: WHO World Health Organization. [http://www.who.int/immunization\\_delivery/systems\\_policy/logistics/en/index4.html](http://www.who.int/immunization_delivery/systems_policy/logistics/en/index4.html).
- WHO. (2014a). *Global immunization data*. Tech. rep., WHO world Health Organization and others.
- WHO, World Health Organization. (2014b). *Who policy statement: Multi-dose vial policy (MDVP)-revision 2014*. Geneva, 2014. Tech. rep., WHO/IVB/14.07.
- Yang, W., Parisi, M., Lahue, B. J., Uddin, M. J., & Bishai, D. (2014). 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.
- You, F., & Grossmann, I. E. (2013). Multicut benders decomposition algorithm for process supply chain planning under uncertainty. *Annals of Operations Research*, 210(1), 191–211.