Drone-Enabled Early Infant Diagnosis of HIV Supply Chains

Alborz Hassanzadeh
HEC Montréal, Canada, al.zadeh@hec.ca

Laurent Alfandari
ESSEC Business School, France, alfandari@essec.edu

Over 85% of the burden of infant human immunodeficiency virus (HIV) is carried by African countries, where only one-third of the population resides within 2km of functioning routes. This leads to a significantly long turnaround time (TAT), which in turn results in high infant mortality rates. Uncrewed aerial vehicles (drones) hold great potential to benefit this region that is grappling with the global epidemic of HIV. We study the potential benefits of using drones for transporting samples crucial to the early infant diagnosis (EID) of HIV. We consider a resource-limited environment where samples must be transferred between healthcare facilities and laboratories located in dispersed areas and efficient sample transportation is indispensable. We first develop a mixed-integer non-linear program to optimize the assignment of clinics to labs, transportation mode selection, and sizing of the fleet of drones. Despite high non-linearity in the objective function and constraints, we derive a linear model that is computationally tractable. Using country-level data from Mozambique, an overview of our results demonstrates that drones can significantly reduce the turnaround time. The potentials of a drone network in reducing the TAT is threefold: (i) Drones significantly reduce the sample transportation time; (ii) By eliminating the waiting time for sample transportation opportunities, drones remarkably decrease clinic delays and simultaneously increase the frequency of dispatches; (iii) Increased dispatch frequency creates a continuous flow of samples between clinics and labs which decreases the variation in the random time between creation of two sample batches. This in turn decreases the total lab delay. Our analysis demonstrates that designing an effective network does not undermine fairness as much as prioritizing equitable resource allocation reduces effectiveness. We also find that optimizing for a fair distribution of resources (using a CVaR-type objective function) tends to concentrate drones in villages rather than major cities. Conversely, optimizing the effectiveness of the network generates a structure whereby drones are allocated primarily to urban areas. Finally, we find that drones have a greater impact on a small number of infants in villages, while in major cities, they have a smaller impact on a large number of infants. Therefore the structure of the network should be tailored to the policy-maker’s needs.

Key words: HIV, drones, global health operations, social impact operations, queuing

1. Introduction
Since its identification in 1983, the human immunodeficiency virus (HIV) has become a major global health challenge. In 2020, a child was infected with HIV every two minutes (UNICEF 2021). UNICEF reports that, in 2018, sub-Saharan Africa is home to nearly 90% of all children and
adolescents living with HIV (UNICEF 2023). Without treatment, over 35% of the infected infants in African countries die before their first birthday, and more than half perish before they turn two years old (Newell et al. 2004). The human immunodeficiency virus attacks and destroys the CD4 cells of the immune system. CD4 cells play a significant role in protecting the body from infection. If left untreated, HIV can lead to acquired immunodeficiency syndrome (AIDS). There is currently no effective cure, and once contracted, HIV stays in the body for life; however, with proper medical care, the virus can be controlled through antiretroviral therapy (ART), which uses a combination of medications to treat the infection.

As a result, Early Infant Diagnosis (EID) of HIV programs are designed to inform caregivers of the status of HIV-exposed infants and link those with confirmed infection to care and treatment. EID of HIV offers substantial benefits to infants and families who live with and without the virus (Ciaranello et al. 2011). One of the first steps for treatment is confirmation of infection; in adults and adolescents, chronic HIV infection can be accurately diagnosed by antibody tests. In infants born to HIV-infected mothers, however, maternal antibodies persist in the blood for up to 18 months (Chantry et al. 1995, Dabis et al. 1993), which renders antibody tests inaccurate. Due to the inaccuracy of such tests, the blood sample of infants must be tested through an enzymatic process called polymerase chain reaction (PCR) (Parpia et al. 2010). However, the technology of PCR machines is expensive and complex. As a result, EID networks are often highly centralized, with very few laboratories equipped for PCR testing.

This entire process creates a supply chain that begins with the blood sample of an HIV-exposed infant being drawn at a clinic and sent to a lab for PCR testing, thus creating a “cascade” of EID and pediatric HIV care as Ciaranello et al. (2011) put it. One of the most important measures of this cascade of care is the turnaround time (TAT) which is the delay between obtaining the sample and the time that results are available to be announced to caregivers. Minimizing the TAT is a widely sought-after objective since not only does it increase the likelihood of result collection by caregivers, it also enables treatment initiation, which can prevent infant mortality and reduce the negative health consequences of active virus’ presence (Becquet et al. 2012, Chatterjee et al. 2011).

However, in sub-Saharan African countries, the TAT is significantly prolonged due to a lack of reliable and functioning roads. It is reported that only one-third of Africans reside within 2km of a functional road. In Mozambique, only about 25% of the rural population lives within 2km of functioning routes (World Bank 2011). Only 2% of the roads are paved in South Sudan, a country larger than France; moreover, conflict, insecurity, and poor infrastructure leave most of the country’s unpaved roads impassable (Ranganathan and Briceño-Garmendia 2011). Environmental factors and high maintenance costs further exacerbate this situation. Rainy seasons render some streets in poor conditions and even wash some routes out entirely (Khazan 2016). Take the example
of the Likoma Districts in Malawi, which is accessible primarily by boat. During rainy and flooding seasons, this district is unreachable by conventional transportation modes.

There has been several propositions and analysis on how to tackle the problem of long TATs, such as analyzing the impact of point-of-care (POC) devices, the use of antigen tests, and improving resource allocation. However, very little attention has been given to how new and autonomous technologies, such as drones, can be used to improve the EID of HIV supply chains. Using drones and proper modeling of this resource-limited supply chain enables us to tackle this problem and create new opportunities, some of which we explore. We first develop a mixed-integer nonlinear program to optimize the use of drones in such a supply chain. The problem involves the assignment of clinics to laboratories, transportation mode selection, and sizing issues about the fleet of drones to carry the samples from clinics to labs.

In the optimization problem, the objective function includes a random variable (TAT), a function of other random variables, themselves nonlinear functions of several auxiliary decision variables. These variables are nonlinear functions of clinic-lab-mode assignment and drone fleet size decisions. Despite all these challenges, we derive a linear formulation of the problem that is computationally tractable. We use queuing theory and model clinic operations as an $\text{M}/\text{D}^\infty / \ell$ queue and lab operations as a $\sum_{i \in I} \sum_{k \in K} \text{GI}^{\psi_{k \imath}^{(1)}} / C_{j}^{b_{j}} / s_{j}$ queueing system. We then use state-of-the-art approximations to measure the turnaround time and the total waiting time at labs and clinics. We also conduct a sensitivity analysis on the drone budget and the dispense of this budget in the network. Finally, we analyze how optimizing effectiveness versus equity alters the network’s structure and drone allocation decisions.

Using country-level data from the EID of HIV network in Mozambique, an overview of our findings shows that the impact of drones on reducing the TAT is threefold: A modestly sized fleet of drones significantly decreases the average transportation time of clinics that are newly equipped with drones by a factor of 85, since drones fly in a straight line, do not require functioning roads, and are resistant to variable weather. We also find that drones eliminate the waiting time for transportation opportunity arrivals at healthcare facilities and increase dispatch frequencies, thereby reducing the average clinic delay by up to a factor of 13. The final contribution of drones in reducing the TAT comes through creating a continuous flow of samples between clinics and labs, thus significantly improving the average lab delay.

We also find that when minimizing the average TAT in the absence of drones, the optimal reassignment of clinics to labs improves the network’s effectiveness by only 0.7%, compared to the status quo that is assignment based on administrative regions. However, deploying a fleet of drones reduces the average TAT by up to 27.2%. An equitable distribution of resources is often a significant challenge in humanitarian problems since, in most practical settings, it collides with an effective
allocation. To this end, we also optimize for the tail of the TAT distribution (Conditional Value At Risk, or CVaR). An overview of our results for CVaR minimization shows that the optimal re-assignment of clinics to labs, in the absence of drones, decreases the conditional value at risk by a mere 0.3%. However, an optimal distribution of a fleet of drones leads to a decrease in the average TAT by up to 18.4% and reduces the CVaR by up to 18%. This suggests that we can prioritize equitable distribution of resources without compromising the network’s overall effectiveness.

Finally, we find that aiming to design an effective network produces a drone allocation policy that predominantly focuses on urban areas and major cities. Conversely, an objective function that focuses on equitable distribution of resources by optimizing the tail of the TAT distribution yields a more dispersed allocation of drones. This in turn leads to a network that prioritizes clinics located in rural areas and villages, thus promoting a more equitable allocation of resources. Note our methodology can also cater to sample transportation for other centrally tested diseases. With slight modifications, our model can accommodate the transportation of other health products such as medications, testing kits, and hygiene packages.

In the remainder of the paper, we first provide an overview of the related literature in §2. We then model a drone-enabled EID of the HIV supply chain in §3. Section 4 presents an application of our model and results using data from Mozambique and section 5 concludes the paper. Proofs and supplementary materials are available in the online appendix.

2. Background and Related Literature

Our work contributes to the growing literature on global health operations and HIV supply chains, allocation and assignment problems with congestion, and the emerging operations management/research interest in the applications of drones.

2.1. Global Health Operations and HIV Supply Chains

The operations community has built a substantial research repository on healthcare delivery systems. Researchers broadened the scope of this stream to encompass inquiries into healthcare delivery in a resource-limited setting. This avenue of research is frequently referred to as Global Health Operations (GHO) (Kraiselburd and Yadav 2013). There is a rich literature on this type of research: for example, see De Boeck et al. (2022) on vaccine supply chains, Natarajan and Swaminathan (2017) on managing and procuring health products and Gibson et al. (2023) and references therein on sample transportation.

The practical setting in GHO research often concerns nonprofit organizations and the focus of the study is often related to underserved communities. Jónasson et al. (2022) thus categorize this stream of research under the umbrella of Social Impact Operations. Within that field of research, there exists an area of inquiry that analyzes HIV supply chains. For example, Lasry et al. (2007)
analyzed the allocation of HIV prevention funds, McCoy and Eric Johnson (2014) studied clinic capacity management and de Vries et al. (2020) focused on the location of clinics in Africa to improve access to HIV-related services.

Within the domain of HIV supply chain management, there is a distinct area of investigation that places significant emphasis on the early diagnosis of HIV in infants. Deo and Corbett (2010) studied the impact of uncertainty in the supply of HIV treatment on the optimal balance between continuous treatment for current patients and initiation of treatment for new HIV+ patients. Deo and Sohoni (2015) considered an EID of HIV network with a single laboratory and several clinics and found that a proper allocation of POC devices dominates other measures such as increasing the lab capacity. Jónasson et al. (2017) studied a network with several laboratories and several clinics and provided operational recommendations on the proper assignment of clinics to labs and their impact on the number of patients initiating treatment.

While an efficient allocation of POC devices and assignment of clinics to laboratories can enhance the overall performance of the network, these endeavors may prove insufficient in exceptionally challenging circumstances as mentioned in §1. In this work, using the case of Mozambique, we explicitly demonstrate that drones have the potential to transform the EID of HIV and overcome such challenges.

2.2. Facility Location with Congestion
Our work also contributes to the vast literature on facility location problems with congestion. For location problems with mobile servers, see, for example, the seminal works of ReVelle and Hogan (1988, 1989), whereby the authors assumed the probability of two servers being busy at the same facility is independent, thereby incorporating congestion. Berman et al. (1985) considered an $M/G/1$ in each facility and to consider multiple servers, Batta and Berman (1989) analyzed the problem by employing an $M/G/s$ queueing system at each facility. While these articles analyzed problems with mobile servers, they modeled congestion implicitly; that is, the number of servers is known and fixed a priori. In some notable works such as Marianov and Serra (2002), Baron et al. (2008), congestion is captured explicitly. However, the server locations are fixed. Despite a rich body of literature, we cannot adapt the existing work to properly include essential aspects of the EID of HIV with drones in its entirety.

In contrast to previous work, we model clinic operations as a queue with mobile servers and explicitly determine the optimal number of servers that are drones. Lab operations are modeled as a queueing system whereby the servers are immobile. Moreover, the EID network consists of two echelons, each modeled as a queueing system: one with mobile servers (clinics) and the other with static servers (labs). The number of servers in the labs (PCR testing machines) is fixed a priori; however, the number of servers (drones) per clinic will be determined.
Finally, the intricate interplay among these echelons elevates model complexities. The reason hinges on samples dispatched from clinics being batched and stochastic. A confluence of such batches creates the input flow of labs. In certain aspects, our work is closely related to Deo and Sohoni (2015), Jónasson et al. (2017) who studied the EID of HIV network in Mozambique. For example, regarding modeling lab delay using queuing theory, we also benefit from approximations of Sakasegawa (1977), Hanschke (2006). However, for model tractability Deo and Sohoni (2015), Jónasson et al. (2017) invoked a set of assumptions in clinics and labs. Since certain constraints are intrinsic to drone technologies, our model inherently diverges from theirs as the same underlying assumptions cannot be imposed in a drone-enabled EID of HIV network.

2.3. Applications of Drones in Operations Management/Research

Finally, our work also contributes to the growing literature on the applications of drones from the operations management/research point of view. Once developed for a single purpose in the military, Uncrewed Aerial Vehicles (UAV) or drones have received much attention due to their potential applications in surveillance (Evers et al. 2014), agriculture (Zhang and Kovacs 2012), and commercial cargo delivery (Reed et al. 2022).

The benefits of using drones in last-mile delivery are evident: drones have low per-mile cost, they can operate without a human pilot, and finally, they have a relatively high speed regardless of road traffic or availability of proper infrastructure. The majority of research on drone applications in the OM/OR community has emphasized the traveling salesman problem with flying sidekicks, where a drone is often coordinated with a truck. See (Agatz et al. 2018, Baloch and Gzara 2020, Carlsson and Song 2018, Dayarian et al. 2020), for example.

Drones have also been academically studied in humanitarian and relief problems (Bravo et al. 2019, Glock and Meyer 2020, Adsanver et al. 2021). However, there is a handful of studies on the applications of drones in healthcare delivery, with the majority of the research focusing on the delivery of medical products and supplies to out-of-reach and rural areas (Hern 2014, Kim et al. 2017, Knoblauch et al. 2019, Mateen et al. 2020, Enayati et al. 2023). There also exists a few reports from field studies by Dirks (2017) and Phillips et al. (2016) on the use of drones in healthcare systems. Recently, Boutilier and Chan (2022) studied the benefits of a drone network for defibrillator dispatching in response to out-of-hospital cardiac arrests. They developed a location-queueing model to minimize the drone fleet size. Gao et al. (2023) embedded a fleet of drones in an ambulance network to reduce the response time to overdose incidences.

Amukele et al. (2017) found that blood samples are stable in prolonged drone flights. There also exist a few promising reports on the feasibility of blood sample transportation via drones in sub-Saharan African countries (UNICEF 2016, 2017, 2019). Despite the compelling body of evidence pointing to the potential advantages of drones in sample transformation, a unified concrete framework for operational decision-making is still missing. This gap further motivates our research.
3. Problem Statement

We study an early infant diagnosis network with $I$ clinics indexed by $i \in \mathcal{I} = \{1, \ldots, I\}$ in predetermined locations. The process begins with caregivers bringing infants exposed to HIV to a clinic for testing at a rate of $\lambda_i$. For model tractability, we assume that this infant arrival rate is independent of clinics’ method of HIV detection and of the delay in receiving results.

When an infant is brought to a clinic for HIV screening, a nurse collects a blood sample either by a heel stick or finger prick and dries it on a filter paper to be sent and tested at a laboratory (Smit et al. 2014). These dried blood samples are not infectious and can remain stable if exposed to heat (Ciaranello et al. 2011). At the same time of testing, an appointment is given to the infant’s caregivers to follow up for results collection, further treatment, and advice. The infant’s mother is also recommended to initiate a protective treatment (Creek et al. 2007).

A technician then accumulates samples until a transportation opportunity becomes available to carry the samples to a lab. As a result, there will be a preprocessing delay at clinics which we denote by $\gamma_i$. We leave a detailed discussion and calculation of this delay in section 3.2. Once a vehicle is available, samples are sent to one of the labs indexed with $j \in \mathcal{J} = \{1, \ldots, J\}$.

Once results are ready, they are communicated via short message service to clinics. When clinics receive the results, there will be a short post-processing delay at the clinic before they are communicated to parents or caregivers. We denote this delay by $W_i$ which is approximately one day, given discussion and consultation with clinic collaborators. Together, these delays constitute the total clinic delay (CD). We remark that this delay is a function of several random variables and decision vectors and differ further details for the next section.

In a typical EID of HIV in sub-Saharan African countries, samples are carried to labs by various means. For example, clinics can use informal opportunities; local staff traveling to an area close to a lab and hence deposit samples at that lab which can be done with a bike, car, or even on foot or a combination of these which we consider a “mode” and index by $k \in \mathcal{K} = \{1, \ldots, K\}$.

Naturally, the travel time $T_{ij}^k$ differs depending on which mode of transport is assigned to each clinic. Due to the lack of proper infrastructure in less developed countries, some transportation modes are highly time-consuming and impose a long delay. Drones, due to their structural abilities, have significantly lower travel time. For example, there are small drones that can carry 2 to 3kg and travel between 30 to 75km while large drones can carry 8kg and travel up to 900km. We provide further details on drone specifications in section 4. We denote the set of drone types by $\mathcal{D} \subset \mathcal{K}$, the coverage range by $R^k$ and capacity with $c^k$ for $k \in \mathcal{D}$.

Due to various requirements for drone corridors and related technologies, we assume each clinic-lab link can be equipped with only one type of drone. There usually exists an investment budget...
and, consequently, a limit on the number of drones available, which we denote by $n^k, k \in \mathcal{D}$. Moreover, in centralized networks, laboratories are usually located in populous urban centers, where they can leverage enhanced rapid technical assistance and have ample space. At the same time, each lab serves over 50 clinics. For these reasons, we assume that drone bases are in the laboratories.

Let $x^k_{ij} \in \{0, 1\}$ be the decision variable indicating the joint assignment of clinic $i$ to lab $j$ with transportation mode $k$ and $\mathbf{x}$ the clinic-lab-mode assignment vector. We index the number of drones of type $k \in \mathcal{D}$ assigned to arc $(i, j)$ by

$$\ell \in \mathcal{L}^k = \{1, \ldots, L^k\},$$

where $L^k \leq n^k$ is a tailored upper bound on $\ell$. We then define the binary variable

$$z^k_{ij\ell} \in \{0, 1\} \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{D}, \ell \in \mathcal{L}^k,$$

where $z^k_{ij\ell} = 1$ if and only if clinic $i$ is assigned to lab $j$ with $\ell$ drones of type $k$, and $z^k_{ij\ell} = 0$ otherwise. Variable $z^k_{ij\ell}$ serves as a proxy in modeling the number of drones assigned to an arc. This choice of modeling proves to significantly reduce model complexity and avoids non-linearity in constraints, as explained later in §3.2.

Let $D_{ij}$ be the Euclidean distance between clinic $i$ and lab $j$. Therefore, clinic $i$ can be assigned to lab $j$ and served with drone type $k \in \mathcal{D}$, if and only if $D_{ij} \leq R^k$, that is, the length of arc $(i, j)$ is less than or equal the coverage range of the assigned drone type. For model tractability, we assume that once a drone type $k \in \mathcal{D}$ is assigned to arc $(i, j)$, that drone only transports samples on arc $(i, j)$ and cannot be used to transport samples on arc $(i', j), i' \neq i$.

Once samples are deposited at a lab, staff accumulate samples in batches to be processed on a machine. This batching is due to the costly technology of PCR testing machines and their scarcity in less developed countries. Therefore, a batch formation delay is observed at the lab. Furthermore, there will be a congestion delay: the waiting time until a PCR testing machine becomes available. Finally, after samples are tested, staff must cross-check and record them on a computer, which leads to further post-processing delays at the lab, which we denote by $W_j$. Similarly to before, the lab technicians we consulted suggested that we consider this post-processing delay as approximately one day. Combined, these delays constitute the total laboratory delay (LD).

For a given $(i, j, k)$ assignment, for reasons discussed previously, a crucial output of clinic $i$ is its turnaround time which is the time interval that begins with an infant’s blood sample being drawn at a clinic and ends when the test result is ready to be delivered to the caregiver (Creek et al. 2007). Formulated mathematically,

$$TAT_i(\mathbf{x}, \mathbf{z}) = CD_i(\mathbf{x}, \mathbf{z}) + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} (T^k_{ij} + LD_j(\mathbf{x}, \mathbf{z})) x^k_{ij} \quad \forall i \in \mathcal{I},$$

(2)
where $T_{kj}^k$ is the transportation time from clinic $i$ to lab $j$ with mode $k$, $CD_i(x,z)$ is the total clinic delay, and $LD_j(x,z)$ is the lab cycle time at lab $j$. We include $(x,z)$ in the notation to emphasize that the TAT is a function of allocation vectors. We provide a detailed explanation of the computation of these delays in the section but for the moment, we remark that both lab and clinic delays are functions of several random variables and decision vector $(x,z)$.

Since the collection of results and treatment initiation depends heavily on the average TAT at each clinic, we dedicate the following section to calculating and estimating components of the average TAT. Moreover, since there can be various methods to estimate the expected turn-around time $E[TAT]$, we denote our estimations of the mean values of stochastic parameters of the EID of the HIV supply chain network by $\tilde{E}[\cdot]$. We provide a schematic representation of the cascade of events that we explained previously in Figure 1.

### 3.1. Clinic operations and delays

The modeling and calculation of clinic delays are among the most complicated ones in the EID of the HIV cascade due to the high number of activities and resources involved. Drone operations further complicate these calculations. In this section, we analyze and model such operations in detail.

In addition to the post-processing administrative delay $W_i$, the total clinic delay has another component: the delay before a randomly chosen sample is sent to a lab for further analysis. If a clinic is to be served with a transportation mode other than a drone ($k \in \mathcal{K} \setminus \mathcal{D}$), similar to Jónasson et al. (2017), we model this delay as the average inter-arrival time of transportation opportunities at clinic $i$ which follows an exponential distribution with parameter $1/\theta_i^k$. When such an opportunity presents itself, all samples collected until that point in time are loaded in the vehicle and dispatched to a lab. This implies that the capacity of the vehicle is large enough to contain all samples, regardless of the size of the sample batch. This is a reasonable assumption since

**Figure 1**  Sample processing timeline in the lab network
dried blood sample filter papers are lightweight and require little space. If some mode \( k \in \mathcal{K} \setminus \mathcal{D} \) is assigned to clinic \( i \), the average total delay is

\[
\mathbb{E}[CD_i] = \frac{1}{\theta_i} + W_i. \tag{3}
\]

Although it is reasonable to assume that the capacity of transportation modes such as a car or a bike is large enough to contain all samples, this assumption might be violated for drones due to their technical characteristics and limited capacity. Moreover, repeated dispatches using drones before they have full capacity can also engage a significant portion of the labor force, a resource already scarce in HIV supply chains in sub-Saharan African countries. Therefore, for labor and cost economization, we assume that a drone will dispatch samples only when its capacity is full. This assumption is not too restrictive as it naturally favors assigning drones to higher volumes of samples, which is an efficient use of costly resources. As a result, drones transport blood sample batches of size \( c_k \) for all \( k \in \mathcal{D} \).

We model the process of samples at clinic \( i \)–batched and loaded on drones and dispatched to a lab–as a queueing system where sample arrival follows a Poisson process with parameter \( \lambda_i \). The number of servers equals the number of drones \( \ell \) assigned to arc \((i, j)\). Although it takes \( T_{ij}^k \) time units for a drone of type \( k \) to carry a sample batch from clinic \( i \) to assigned lab \( j \), we note that the “processing time” of a server in this queueing system is \( 2T_{ij}^k \) since a drone requires \( T_{ij}^k \) time units to return to clinic \( i \) to pick up the next batch, if \( c_k \) samples have arrived.

Our clinic collaborators informed us that loading samples on a drone takes approximately one minute. Therefore, we assume the unloading time to be the same. Also, since travel time is the largest component of the total service time, it is a realistic assumption that service time is fixed.

These are the characteristics of an \( M/D/c_k/\ell \) queueing system, for which the average queue time is equivalent to the average waiting time of a sample before dispatch. As a result, we model and calculate the total average clinic delay in the following proposition.

**Proposition 1.** Suppose clinic \( i \) is assigned to lab \( j \) and is served with \( \ell \) drones of type \( k \in \mathcal{D} \) \((s_{ij\ell}^k = 1)\). Using the approximation of Wu (2014), the average clinic delay comprising the average waiting time of a randomly chosen sample until its dispatch, plus the clinic post-processing delay, is

\[
\mathbb{E}[CD_i] = \frac{c_k - 1}{2\lambda_i} + \frac{T_{ij}^k}{c_k} \left( \frac{\delta_{ij\ell} \sqrt{2(\ell+1)-1}}{\ell (1 - \delta_{ij\ell})} \right) + W_i, \tag{4}
\]

where \( \delta_{ij\ell} = 2T_{ij}^k \lambda_i / \ell c_k \) is the drone utilization rate.
Proof. We present the proof in Appendix B. □

The first term on the right-hand side of formulation (4) is the average batching delay. The second term is the congestion delay. We remark that the average delay before a sample is dispatched depends on the clinic-lab assignment, the type of transportation mode, and the number of drones if arc \((i,j)\) is set to be served by drones. From formulations (3) and (4), we can determine the total average clinic delay as

\[
\mathbb{E}[CD_i(x,z)] = \sum_{j \in J} \left( \sum_{k \in K \setminus D} \frac{x_{ij}^k}{\theta_i^k} + \sum_{k \in D} \sum_{\ell \in L_k} \left( \frac{c^k - 1}{2\lambda_i} + \frac{T_{ij}^k}{\ell c^k} \left( \delta_{ij} - \frac{\sqrt{2(\ell+1)-1}}{\ell} \right) \right) z_{ij}^k \right) + W_i \quad \forall i \in I.
\]

Notice from the formulation above that modeling the number of drones \(\ell\) indirectly as an index in variable \(z_{ij}^k\) results in this linear formulation. Modeling this decision directly as a non-negative integer decision variable would lead to an infeasible formulation (division by zero) since some arcs are not assigned with drones. Such alternative modeling would also result in the generation of highly non-linear constraints.

We now analyze a critical parameter \(\omega_{ij}\), that is, the mean dispatch frequency of sample batches sent from clinic \(i\) to lab \(j\) using transportation mode \(k\). As we observe in the next section, this variable directly affects lab delays and, therefore, the turnaround time.

For those \((i,j)\) pairs assigned with a mode other than a drone \((k \in K \setminus D)\), we denote by \(I_{ik}^d\) the random time between two dispatches. We also define \(I_i^s\) and \(I_i^t\) to be the random inter-arrival times of samples and transport opportunities, respectively. Therefore,

\[
\mathbb{E}[I_{ik}^d] = \mathbb{E}[I_i^s] + \mathbb{E}[I_i^t] = \frac{1}{\lambda_i} + \frac{1}{\theta_i^k} = \frac{\lambda_i + \theta_i^k}{\lambda_i \theta_i^k} \quad \text{for } k \in K \setminus D.
\]

For such \((i,j)\) pairs, the mean dispatch frequency is

\[
\omega_{ij}^k = \frac{1}{\mathbb{E}[I_{ik}^d]} = \frac{\lambda_i \theta_i^k}{\lambda_i + \theta_i^k} \quad \text{for } k \in K \setminus D.
\]

(5)

For \((i,j)\) pairs that are assigned with drones, the mean dispatch frequency depends on the number of drones \(\ell\) assigned to that arc. In a steady state where a sufficient number of drones type \(k \in D\) can be assigned to arc \((i,j)\), the time between dispatches is the time required for \(c^k\) samples to arrive at clinics \(i\), that is \(c^k/\lambda_i\). Since reaching a steady state is not likely due to the resource-limited nature of the problem in real life, each drone would make an average of \(1/2T_{ij}^k\) trips per unit of time. Therefore, with \(\ell\) drones of type \(k \in D\), there would be an average of \(\ell/2T_{ij}^k\) dispatches per unit time (a steady state would require \(\ell \geq \lceil 2T_{ij}^k \lambda_i / c^k \rceil \)). As a result, depending on the number of drones type \(k \in D\) assigned to arc \((i,j)\), we compute the mean dispatch frequencies as

\[
\omega_{ij}^k = \begin{cases} 
\frac{\ell}{2T_{ij}^k} & 1 \leq \ell < \left\lceil \frac{2T_{ij}^k \lambda_i}{c^k} \right\rceil \\
\frac{\lambda_i}{c^k} & \ell \geq \left\lceil \frac{2T_{ij}^k \lambda_i}{c^k} \right\rceil 
\end{cases} \quad \text{for } k \in D.
\]

(6)
Using (5) and (6), we formulate the mean clinic dispatch frequency as

$$\omega_{ij}(x, z) = \sum_{k \in K \setminus D} \omega_{ij}^k x_{ij}^k + \sum_{k \in D} \sum_{\ell \in L} \omega_{ij}^k z_{ij \ell}^k \quad \forall i \in I, j \in J.$$  

(7)

### 3.2. Laboratory Operations and Delays

In a typical network, results arrive at labs in batches from different clinics. Let us define $\Psi_i$ as a random variable denoting the sample batch size dispatched from clinic $i$.

If a clinic is assigned to be served by any transportation mode but drones ($k \in K \setminus D$), at any time, the probability of a transportation opportunity being available before the following sample arrives is $\frac{\theta_i^k}{\lambda_i + \theta_i^k}$. Therefore, the random batch size of clinic $i$, $\Psi_i^k$, follows a geometric distribution where “failure” is the number of samples arriving before the next transportation opportunity is available, hence

$$\mathbb{P}(\Psi_i^k = \psi) = \frac{\theta_i^k}{\lambda_i + \theta_i^k} \left(1 - \frac{\theta_i^k}{\lambda_i + \theta_i^k}\right)^{\psi}.$$  

As a result, the average shipment size from clinic $i$ is the mean of a geometric distribution $\text{geom}(\frac{\theta_i^k}{\lambda_i + \theta_i^k})$ with parameters

$$\mathbb{E}[\Psi_i^k] = \frac{1 - \frac{\theta_i^k}{\lambda_i + \theta_i^k}}{\frac{\theta_i^k}{\lambda_i + \theta_i^k}} = \frac{\lambda_i}{\theta_i^k}, \quad \text{and} \quad \mathbb{V}[\Psi_i^k] = \frac{\lambda_i (\lambda_i + \theta_i^k)}{(\theta_i^k)^2} \quad \text{for} \ k \in K \setminus D.$$  

(8)

If a clinic is assigned with drones ($k \in D$), since the drone has to be at full capacity before being dispatched, the sample batch size for that clinic is the capacity of that drone, in which case

$$\mathbb{E}[\Psi_i^k] = c^k, \quad \text{and} \quad \mathbb{V}[\Psi_i^k] = 0 \quad \text{for} \ k \in D.$$  

(9)

After data entry and sample preparation at lab $j$, samples are processed in batches with size $b_j$ to reduce various costs. Each lab has $s_j$ testing machines available, each of which has a random processing time $S_j$. After sample processing, results are analyzed, prepared, and sent back to clinics through different methods such as fax, or SMS, depending on resource availability. This transmission time is a negligible component of the TAT.

Each lab $j$ can then be viewed as a queueing system $\sum_{i \in I} \sum_{k \in K} GI[\Psi_i^{k(i)}]/G_j^{b_j}/s_j$, where $G_j^{b_j}$ captures lab-specific service time distribution and $k(i)$ is the transportation mode assigned to clinic $i$. In Appendix B, we provide detailed explanations on how to compute the average queue time of such a system. In the following proposition, we summarize the result and compute the total lab delay.
Proposition 2. By Little’s law, approximations of Sakasegawa (1977) and Hanschke (2006), the average delay at laboratory \( j \) for a randomly selected sample is

\[
\dot{E}[LD_j(x, z)] = \frac{(b_j - 1) E[S_j]}{2b_j s_j \rho_j} + \frac{(\nu_j(x, z) + SCV[S_j]) E[S_j] \rho_j^{-1/2(s_j+1)}}{2s_j (1 - \rho_j)} + E[S_j] + W_j \quad \forall j \in J,
\]

where

\[
\rho_j = \sum_{i \in I} \sum_{k \in K} \frac{E[S_j] \lambda_i x_{ij}}{b_j s_j},
\]

is the lab utilization rate, and variable

\[
\nu_j(x, z) = SCV[B_j(x, z)] \approx \frac{1}{b_j} \frac{\sum_{i \in I} \sum_{k \in K \setminus D} E[(\Psi_i)^2] \omega_{ij}^k x_{ij}^k + \sum_{i \in I} \sum_{k \in D} \sum_{\ell \in L} E[(\Psi_i^k)^2] \omega_{ij\ell}^k z_{ij\ell}^k}{\sum_{i \in I} \sum_{k \in K \setminus D} E[\Psi_i] \omega_{ij}^k x_{ij}^k + \sum_{i \in I} \sum_{k \in D} \sum_{\ell \in L} E[\Psi_i^k] \omega_{ij\ell}^k z_{ij\ell}^k},
\]

is the squared coefficient of variation of the random times between the creation of successive fully formed batches; the mean dispatch frequencies \( \omega_{ij}^k \) and \( \omega_{ij\ell}^k \) are calculated in (5) and (6), and \( E[\Psi_i] \) and \( E[(\Psi_i^k)^2] \) are computed using (8) and (9).

Proof. We present the proof in Appendix B. \( \square \)

In formulation (10), the first term on the right-hand side is the average batching delay. The second term is the average congestion delay, the third is the average sample testing time, and the last term is the lab post-processing delay.

### 3.3. Measure of Equity

Guaranteeing an equitable distribution of resources is often a crucial issue in humanitarian contexts, and achieving fairness often poses a challenge to achieving effectiveness. To design an equitable network, we optimize the tail of the average TAT distribution using the conditional value at risk. The upper CVaR for a specified probability level \( \beta \in (0, 1) \) is defined as

\[
\min_{\alpha} \left\{ \alpha + \frac{1}{I (1 - \beta)} \sum_{i \in I} E_{\lambda, \theta}[TAT_i(x, z, \lambda, \theta) - \alpha]^+ \right\},
\]

where \( [Y]^+ = \max(Y, 0) \) (see Rockafellar et al. (2000)).

In section 4, we also employ the Gini index \( G \in [0, 1] \), a measure that has gained widespread acceptance as a standard for evaluating equity in social welfare (Marsh and Schilling 1994). We briefly review the Gini index in Appendix B. For the moment, we remark that while CVaR represents the expected value of the TATs that exceed the value at risk (VaR) at the same probability level \( \beta \), the Gini index computes the degree of dispersion in the distribution of weighted average TATs. A higher Gini index indicates greater inequality or concentration of small weighted average TATs.
TATs among a few clinics. A lower Gini index, on the other hand, indicates a more even distribution of weighted average TATs. We calculate the value of the Gini index using the following proposition.

**Proposition 3.** For the problem of EID of HIV supply chain with drones, the Gini coefficient can algebraically be computed as

\[
G = \frac{1}{2\lambda^2} \sum_{i=1}^{I} \sum_{i'=1}^{I} \lambda_i \lambda_{i'} \left| \tilde{E}[TAT_i] - \tilde{E}[TAT_{i'}] \right|
\]

where

\[
\lambda = \frac{1}{I} \sum_{i=1}^{I} \lambda_i, \quad \text{and} \quad \tilde{t} = \frac{\sum_{i=1}^{I} \lambda_i \tilde{E}[TAT_i]}{\sum_{i=1}^{I} \lambda_i}.
\]

**Proof.** We present the proof in Appendix B. \(\square\)

We also explore networks that prioritize finding the most efficient allocation. To this end, we define effectiveness as the extent to which a given allocation reduces as much as possible the weighted average TAT, which is computed as

\[
TAT = \frac{\sum_{i=1}^{I} \lambda_i \tilde{E}[TAT_i]}{\sum_{i=1}^{I} \lambda_i}.
\]

### 3.4. The Optimization Model

Using previous discussions and formulations, we develop a nonlinear optimization program (NLOP) to optimize drone use in early infant diagnosis of the HIV supply chain with a CVaR objective function, as provided below. A detailed list of notations is provided in Table 1 in Appendix A.

**NLOP:**

\[
\begin{align*}
\min & \quad \alpha + \frac{1}{I(1-\beta)} \sum_{i \in \mathcal{I}} y_i \\
\text{subject to} & \quad y_i \geq \tilde{E}[TAT_i] - \alpha \quad \forall i \in \mathcal{I}, \\
& \quad \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} x_{ij}^k = 1 \quad \forall i \in \mathcal{I}, \\
& \quad \sum_{\ell \in \mathcal{L}^k} z_{ij\ell}^k = x_{ij}^k \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{D}, \\
& \quad \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{\ell \in \mathcal{L}^k} \ell z_{ij\ell}^k \leq n^k \quad \forall k \in \mathcal{D}, \\
& \quad \tilde{E}[TAT_i] = \tilde{E}[CD_i] + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \left( T_{ij}^k + \tilde{E}[LD_j] \right) x_{ij}^k \quad \forall i \in \mathcal{I}, \\
& \quad \tilde{E}[CD_i] = \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K} \setminus \mathcal{D}} \frac{\theta_{ik}}{\theta_i^k} + \sum_{k \in \mathcal{D}} \sum_{\ell \in \mathcal{L}^k} \left( \frac{\xi_{ij\ell}^k - 1}{2\lambda_i} + \frac{T_{ij}^k}{\ell} \left( \frac{\delta_{ij\ell}^k \sqrt{2(\ell+1)-1}}{\ell(1-\delta_{ij\ell}^k)} \right) \right) z_{ij\ell}^k + W_i \quad \forall i \in \mathcal{I},
\end{align*}
\]
\[
\mathbb{E}[LD_j] = \frac{(b_j - 1) \mathbb{E}[S_j]}{2b_j s_j \rho_j} + \left(\nu_j + SCV[S_j] \mathbb{E}[S_j] \rho_j \right)^{\frac{\sqrt{2(s_j+1)}}{2 \sqrt{(1 - \rho_j)}}} + \mathbb{E}[S_j] + W_j \quad \forall j \in J,
\]

\[
\rho_j = \min \left( \frac{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \mathbb{E}[S_j] \lambda_i x_{ij}^k}{b_j s_j} \right) \quad \forall j \in J,
\]

\[
\nu_j = \frac{1}{b_j} \cdot \frac{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K} \setminus D} \mathbb{E}[\Psi_i^k] \omega_{ij}^k x_{ij}^k + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{D}} \sum_{\ell \in \mathcal{L}^k} \mathbb{E}[\Psi_i^k] \omega_{ij}^k z_{ij}^k}{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{D} \setminus D} \mathbb{E}[\Psi_i^k] \omega_{ij}^k x_{ij}^k + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{D}} \sum_{\ell \in \mathcal{L}^k} \mathbb{E}[\Psi_i^k] \omega_{ij}^k z_{ij}^k} \quad \forall j \in J,
\]

\[
\alpha \geq 0,
\]

\[
y_i \geq 0 \quad \forall i \in \mathcal{I},
\]

\[
x_{ij}^k \in \{0,1\} \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K},
\]

\[
z_{ij}^k \in \{0,1\} \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{D}, \ell \in \mathcal{L}^k,
\]

\[
\nu_j, \rho_j \geq 0 \quad \forall j \in \mathcal{J},
\]

where in constraint (17), \(\delta_{ij}^k = \min \left( 2T_{ij}^k \lambda_i / k^\ell, 1 - \epsilon \right)\).

The CVaR objective function (11) minimizes the average of the tail of the TAT distribution given a probability level \(\beta\). Constraints (13) guarantee that if clinic \(i\) is in the lab network, it is assigned to a transportation mode and a laboratory. Constraints (14) compute the number of drones type \(k\) assigned to arc \((i,j)\). Constraints (15) assure that the total number of drones of type \(k \in \mathcal{D}\) deployed in the network does not exceed the total number available. Equalities (16) calculate the turnaround time of clinic \(i\) based on assignment decisions \(x\) and \(y\). Constraints (17) calculate the average waiting time of a randomly selected sample at clinic \(i\) before its departure to a lab plus the post-processing delay after results are announced to the clinic.

Constraints (18) calculate the total average delay time of laboratory \(j\). Constraints (19) calculate the lab utilization rates that result from assignment decisions and restrict this measure to be less than or equal to 1. Here, for system stability, we set the capacity reserve \(\epsilon\) to be a small number. Constraints (20) compute the squared coefficient of variation of random times between the creation of two consecutive sample batches at lab \(j\). Finally, constraints (21)-(25) give the nature of the decision variables. Note that in the optimization problem above, the objective function includes expected values of TAT random variables which comprise, in particular, the lab delay expected values in (18). These LD expected values not only depend on transport mode assignment variables \(x\) and \(z\) but also on auxiliary variables \(\rho_j\) and \(\nu_j\) in a highly non-linear way.

Despite all these complexity challenges, we derive a linear and feasible reformulation of problem (11)-(25) that is computationally tractable. We use various linearization techniques, the detail of which we provide in Appendix C. This leads to a tractable Mixed-Integer Linear Program MILP that we present in Appendix D and use in our case study.
4. Case Study: Mozambique

In this section, we demonstrate an application of our work using data from Mozambique. The EID of HIV network in Mozambique serves as a practical illustration of an application of our model and methodology for several reasons; HIV and its latest stage of infection, AIDS, is a devastating public health problem in most sub-Saharan African countries, and Mozambique is not an exception. UNICEF (2018) identified Mozambique as the second country with the highest estimated population of infants and adolescents living with HIV after South Africa. Moreover, with the extensive research carried out on this country by numerous scholars such as Mugambi et al. (2013), Deo et al. (2015), Deo and Sohoni (2015), Jónasson et al. (2017), accessing data is more attainable.

We first provide an overview of the EID of the HIV supply chain in Mozambique, followed by details on the data set on this network. We then provide information on the specifications and parameters of drones that can be used in this network. Finally, we present a detailed analysis and discussion of the numerical results from applying our model and findings.

4.1. Set-up of EID of HIV in Mozambique

The EID of HIV program in Mozambique was started in 2006 by the Ministry of Health and as of 2010, it consists of approximately 400 clinics spread over 11 regions with four laboratories capable of performing PCR tests. These labs are in the cities of Maputo, Nampula, Beira, and Quelimane (National AIDS Council 2010).

Our data set pertains to over 200 clinics across 11 regions. In general, clinics transport samples to district and regional headquarters by car, and from there, samples are flown to the city of the assigned lab. As of 2015, the assignment of these clinics to labs is based on the boundaries of administrative regions as shown in Figure 2: samples collected from clinics in the Maputo region are tested in the lab in Maputo. All samples collected in clinics in regions of Nampula, Nassa, and Cabo Delgado are sent to the lab in the city of Nampula. The laboratory in the city of Beira tests samples from its region (Sofala), Inhambane, and Manica. The Qelimane lab in the region of Zambezia is used to process samples from Zambezia and Tete regions. We call this clinic-lab pre-assignment the BaseCase and use this case as a benchmark to assess the impact of drones in the EID network of Mozambique.

In Figure 2 we observe that all the clinics are assigned either to the closest or the second closest lab, except for those in the region of Gaza. The closest lab to Gaza is the one in the capital region of Maputo. However, this lab falls short in terms of capacity. At the same time, there is an excess capacity in the lab in Nampula, and there are frequent, reliable flights between Maputo and Nampula – two of the most populated cities in Mozambique. Therefore, samples collected in Gaza are driven to Maputo and then flown to Nampula for testing.
We obtained the coordinates of clinics and labs and the estimated daily average number of samples that arrive at each clinic. The data set also includes the frequency with which a transportation opportunity becomes available to take samples from a clinic and the estimated transportation time for each clinic-lab pair. The summary statistics of clinics is provided in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily arrival rate of infants to clinic</td>
<td>1.7</td>
<td>0.8</td>
<td>1.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Inter-arrival time of transportation opportunities (days)</td>
<td>11.8</td>
<td>10.2</td>
<td>1.0</td>
<td>87.5</td>
</tr>
</tbody>
</table>

The lab in Maputo has two automatic PCR test machines, while the labs in Nampula, Quelimane, and Beira have one, each with a batch size of 92. All laboratories are located in highly populated urban areas. Such areas have better access to technical and engineering support. In addition, each laboratory serves over 50 clinics. Therefore, we assume that each laboratory is a drone base.

4.2. Drone Specifications

A wide range of technologies is used in designing, engineering, and manufacturing of drones, leading to a wide range of drone capabilities. The drone specifications used in our model are based on recent technological advancements, academic publications, data from several companies that offer drone delivery services, and discussions with drone engineers and experts (Dove Air 2022, Scott and Scott 2017, Mims 2022).

We use specifications of fuel-based fixed-wing drones. These drones have a maximum forward velocity of approximately 75 km/h and can cover a range of up to 900 km. These drones can carry up to 8 kg of payload. The time required for a drone to reach maximum speed is approximately 10 seconds. The drawback of most fixed-wing drones is that they cannot launch vertically and need more space for take-off and landing. However, our lab and clinic collaborators informed us that this technical constraint is not a significant concern since there is ample space at clinics and lab sites.

Since the maximum daily arrival rate of infants at clinics is six in our data set, we assume the capacity of the drone to be six. That is, once a batch of five samples is available, the drone is dispatched to deliver samples to the designated lab for testing. Since most companies provide drone delivery as a service, information on acquisition expenses and capital outlay for drone hubs are rarely accessible. As the drone capacity accommodates the low arrival rates $\lambda_i$, it is unnecessary to employ more than one drone for a single arc $(i,j)$, i.e., $L^k$ is set to one. In the next section, we provide a detailed analysis of the impact of drone fleet size on public health outcomes.
4.3. Results and Managerial Insights

In this section, we present the numerical results from applying our model and methodology to improve an existing EID of HIV network by deploying a fleet of drones across all regions in Mozambique. All optimization problems were coded in Julia 1.9.2 and solved using Gurobi 10.0.2 solver. We set the maximum processing time limit to 10 hours and used a Virtual Machine with an Intel® Xeon® E5-2670, 2.60-GHz processor and 24 GB of RAM.

4.3.1. Drone Network Size. We first assess the impact of the drone fleet size on various performance metrics as we list in Table 2. To this end, we first analyze the BaseCase in which clinics are assigned to labs based on administrative borders of regions, in the absence of drones. We compute several performance metrics for such an assignment policy: the average turnaround time, TAT, and the conditional value at risk, CVaR, of the TAT distribution. We also employ...
an alternative measure of equity to assess the fairness of the distribution of resources given an assignment, the Gini index, $G$.

To analyze the impact of the drone fleet size, we consider various budget levels on the number of available drones $n$. Specifically, $n \in \{0, 5, 10, 20, 30, 40\}$, that is, we equip between zero to 20% of the clinics with drones. We solve the linearized NLOP provided in Appendix D for each drone budget level, which minimizes CVaR. We denote this model by $\text{min-CVaR}$. Alternatively, we solve the optimization problem with an objective function that focuses on the effectiveness of the network, that is, to minimize the average turnaround time. Henceforth, we denote this model by $\text{min-TAT}$.

The second column in Table 2 shows the processing time for each problem instance. Except for two instances of the problem, all other instances were solved within 15 minutes. The third column shows the average turnaround time when optimizing for equity and effectiveness. For example, in the absence of drones, it takes on average 90 days for results to be announced to caregivers, under all three models. Equipping only 5% of the clinics with drones remarkably reduces this delay by between 8 and 10 days.

We also observe that as the fleet size expands, with each additional five drones, the TAT decreases by between 3 and 6 days under the $\text{min-TAT}$ model; A fleet of drones with the size of 40 reduces the average TAT by a remarkable 24.5 days, compared to the BaseCase. While the pursuit of equity reduces dispersion in the tail of the TAT distribution, it naturally increases TAT thus diminishing the effectiveness of the network. The fourth column of Table 2, which demonstrates the improvement in the $\text{TAT}$, better highlights this phenomenon. In the fifth column, we compute the “loss” in the $\text{TAT}$ when pursuing equitable distribution of drones, compared to optimizing the effectiveness of the network. As this column shows, $\text{TAT}$ experiences a marginal increment consistently below the threshold of 12.

The sixth column in the table computes the CVaR. For example, when optimizing for an equitable assignment of resources, the expected value of “losses” beyond the value-at-risk threshold is approximately 118 days, in the absence of drones. Similarly to before, we remark that there is a significant decrease of between 6 and 21 days in CVaR by adding a fleet of drones to a network operating without drones. Each additional five drones reduces CVaR by up to 6 days, under the $\text{min-CVaR}$ model. We note that by fixating solely on the effectiveness of the network, the equitable distribution of resources among clinics is likely to be compromised, as we observe in column 7 of the table. The eight column of Table 2 better demonstrates such compromise of the measure of equity.

A noteworthy observation from Table 2 pertains to the extent of trade-offs required when optimizing one network measure at the expense of another. When maximizing the effectiveness of the electronic copy available at: https://ssrn.com/abstract=4800799
Hassanzadeh and Alfandari: Infant HIV Supply Chain with Drones

The drone-enabled EID of the HIV network, the equity measure deteriorates by as much as 7.1%. Conversely, when optimizing the drone network for equity, its effectiveness (TAT) decreases by up to 12%. The numerical results in columns 5 and 8 have an important implication: an effective allocation of resources results in a significantly lower decrement in the equity of the network compared to the reduction in effectiveness caused by prioritizing the fairness of the network. In other words, the pursuit of effectiveness does not undermine fairness as much as the pursuit of fairness reduces effectiveness.

We also employ the Gini index to measure equity. According to the Gini coefficient from the BaseCase, the assignment is done fairly as the index is close to zero. However, in this assignment, about 5% of the clinics, mostly in rural areas and villages, exhibit a significantly high TAT (over three months). This underscores the nuanced challenge of employing the Gini index and the need for a measure that guarantees improvement in the turnaround time for the most devastated and vulnerable populations and clinics. For these reasons, CVaR by definition serves as a more appropriate metric for capturing and fostering equity.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Main outcomes of the optimization models min-TAT and min-CVaR compared to BaseCase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>BaseCase</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>0</td>
<td>2.6</td>
</tr>
<tr>
<td>5</td>
<td>6.3</td>
</tr>
<tr>
<td>10</td>
<td>6.7</td>
</tr>
<tr>
<td>20</td>
<td>10.5</td>
</tr>
<tr>
<td>30</td>
<td>23.2</td>
</tr>
<tr>
<td>40</td>
<td>59.7</td>
</tr>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>0</td>
<td>2.1</td>
</tr>
<tr>
<td>5</td>
<td>50.8</td>
</tr>
<tr>
<td>10</td>
<td>169.3</td>
</tr>
<tr>
<td>20</td>
<td>738.5</td>
</tr>
<tr>
<td>30</td>
<td>11340.5</td>
</tr>
<tr>
<td>40</td>
<td>36000.0</td>
</tr>
</tbody>
</table>

The plots in Figure 3 display a summary of the impact of drones: under both models (min-TAT, min-CVaR), the effectiveness and equity of the EID of HIV network improve as the drone fleet size increases. As is seen in Table 3, the average sample transportation time for clinics that use drones varies between 1.2 and 3.1 hours while other transportation modes impose an average transportation delay of between 11.21 and 11.44 days. This 85-fold improvement is owed to drones’ core functionality as they fly in a straight line and have a much higher speed compared to other transportation modes.
Hassanzadeh and Alfandari: Infant HIV Supply Chain with Drones

Figure 3  Impact of drone network size on TAT (left) and CVaR (right)

Table 3  Average transportation times of clinics equipped with a drone vs. those without drones

<table>
<thead>
<tr>
<th>n</th>
<th>( \text{Average T w/ drones (days)} )</th>
<th>( \text{Average T w/o drones (days)} )</th>
<th>( \text{Average T w/ drones (days)} )</th>
<th>( \text{Average T w/o drones (days)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>–</td>
<td>11.21</td>
<td>–</td>
<td>11.21</td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>11.27</td>
<td>0.13</td>
<td>11.23</td>
</tr>
<tr>
<td>10</td>
<td>0.07</td>
<td>11.28</td>
<td>0.11</td>
<td>11.24</td>
</tr>
<tr>
<td>20</td>
<td>0.10</td>
<td>11.34</td>
<td>0.12</td>
<td>11.23</td>
</tr>
<tr>
<td>30</td>
<td>0.10</td>
<td>11.38</td>
<td>0.13</td>
<td>11.29</td>
</tr>
<tr>
<td>40</td>
<td>0.09</td>
<td>11.44</td>
<td>0.12</td>
<td>11.36</td>
</tr>
</tbody>
</table>

Drones also help in reducing the turnaround times by decreasing clinic delays (CD) as we observe in the plot on the left in Figure 4. The average CD for clinics equipped with a drone is between six to seven times less than the average CD for those without one. In the absence of drones, clinics would have to wait a long time for a transportation opportunity to become available (see Table 1). Assigning even one drone to a clinic eliminates this waiting time and increases the dispatch frequency between the clinic and the assigned lab (see the right plot in Figure 4). Before we proceed to analyze the significant impact of this increased dispatch frequency, we note that although the average clinic delay consistently decreases with the fleet size in both optimization models, the impact of drones in enhancing the dispatch frequency diminishes under the \( \min \text{TAT} \) model. On the other hand, when optimizing the tail of the TAT distribution, the average dispatch frequency increases with the fleet size and reaches its maximum with 20 drones. The intuition behind this result is that under the \( \min \text{-CVaR} \) model, we optimize for the tail of the TAT distribution which comprises the \( 1 - \beta = 10\% \) of the 200 clinics with the highest TAT.
In our computational experiments, we observed that all the labs work at full capacity. However, even when working at a near-complete capacity, deploying more drones consistently reduces lab delays, thus further contributing to the reduction of TATs (see Figure 5). This decrease is owed to the role of drones in reducing the variation of the random times between the creation of successive fully formed batches, $\nu$ (Figure 6 displays this result). As we observe from formulation (10), the lab delay is decreasing in $\nu$. This result is intuitive: deploying more drones for sample transportation increases the frequency of dispatches $\omega$, thus creating a more continuous flow of samples arriving at labs, which in turn reduces the variation in the random time between the creation of two batches.

We finally note that the delay and variation in time between two batches does not improve for the lab in Nampula as significantly as it does for the other three laboratories. This is because clinics in the region of Gaza cannot be assigned to any other labs. This hard constraint limits the models' flexibility.

4.3.2. Equity, Effectiveness, and the Network Structure. We now analyze how the choice of modeling alters network structure, i.e., drone assignment. Given the existing trade-offs, modeling choices can be tailored to policy maker’s needs. We use results from instances with 20 and 40 drones under the min-CVaR and min-TAT models and employ logistic regression modeling. The models examine the relationship between drone network structure and sample arrival rate, $(\lambda)$, mean inter-arrival time of transportation opportunities $(\bar{\theta})$, and transportation time using the non-drone mode, which we denote by $T^K_D$. We provide the results of the logistic regression in Tables 2 and 3 in Appendix E. The results indicate a highly significant relationship between the
mean inter-arrival time of transportation opportunities ($\bar{\theta}$) at clinics and the assignment of drones to clinics. The positive coefficient of ($\bar{\theta}$) is associated with a higher probability of assigning a drone to a clinic when the waiting time for the arrival of a transportation mode increases. Table 2 for the min-CVaR model shows that the sample arrival rate does not have a significant relationship with the assignment of drones to clinics. Given that the min-CVaR model optimizes the tail of the TAT distribution rather than considering the sample arrival rate, this observation appears intuitive.

The association between sample arrival rate and drone assignment to clinics is somewhat modest when optimizing the tail of the TAT distribution. On the other hand, when optimizing the
effectiveness of the network, we observe a significant association between the two variables. This observation hinges on the fact that, unlike CVaR, the sample arrival rate is embedded in the measure of effectiveness. At the same time, urban areas typically exhibit lower mother-to-infant transmission rates, owing to better access to health facilities and prevention of mother-to-child transmission (PMTCT) programs (Dabis et al. 1999, Wiktor et al. 1999). Nevertheless, the higher population in urban areas than in rural areas often results in higher sample arrival rates. Hence, optimizing the measure of effectiveness tends to focus on assigning drones to major cities, while the min-CVaR model generates networks with a more geographically dispersed distribution of drone assignments.

This result is consistent with what Figures 7–9 display. We observe in Figure 7 that clinics in major cities (larger circles) are often assigned with drones (filled circles) under the min-TAT model. A few exceptions exist where clinics in urban areas are not equipped with drones (large empty circles). This observation stems from the fact that clinics in major cities have a high rate of transportation availability (low $\bar{\theta}$) to carry samples from clinics to the labs. In major cities, the transportation time is also low (low $T^{K\setminus D}$) compared to rural areas due to better road and infrastructure conditions. See Figures 8 and 9 for an illustration. As seen in Figures 7–9, optimizing the average TAT allocates drones to clinics in urban areas while optimizing the 90th percentile of TATs favors clinics in rural areas for drone assignment, thus ensuring an equitable assignment. Take the region Tete as an example. Most of the clinics in this region are located in villages with high waiting times for the arrival of transportation opportunities and relatively low sample arrival rates. Under the min-CVaR model, four clinics out of eight in this region are equipped with drones while optimizing the effectiveness of the network results in a network whereby only two clinics with the highest sample arrival rate are allocated with drones.

5. Discussion and Conclusion

In this work, we explored the use of unmanned aerial vehicles or drones to improve the early infant diagnosis of HIV supply chains and demonstrated the great potential of a drone fleet in reducing the TATs. We develop a mixed-integer nonlinear program to optimize the use of drones for sample transportation from clinics to laboratories. Our findings show that deploying a modestly sized fleet of drones can significantly reduce the turnaround time. In early infant diagnosis of HIV, reducing the TAT is greatly valued as it not only boosts the chances of caregivers collecting results but also facilitates prompt treatment initiation, potentially preventing infant mortality and minimizing the adverse health effects of an active virus. (Becquet et al. 2012, Chatterjee et al. 2011). The drone network’s impact on reducing the TAT stems from three key factors.

First, drones have significantly reduced transportation times due to their inherent speed advantage over other modes of transport. From the Mozambique data set, we observed that the average
Note. The colors correspond to the clinic-lab assignment. Diamonds with darker colors represent laboratories, and clinics are marked with circles with lighter colors. The size of a circle is proportionate to the value of $\lambda$. Filled circles denote drone assignment.

Figure 7 (Color Online) Interplay of $\lambda$ and drone network structure under min-CVaR model (left) and min-TAT model (right)

transportation time of those clinics that are not equipped with drones is approximately 11 days under both optimization models. Clinics equipped with a drone experience an average sample transportation time ranging between 1.2 and 3.12 hours, signifying an 85-fold reduction in transportation delays. The second domain where drones have a substantial impact is reducing clinic delays. Clinics in Mozambique would have to wait between 1 and 87 days until a transportation opportunity is available to carry the sample batch. Drones eliminate this delay and increase the dispatch frequency between clinics and laboratories. This leads to a decrease of between six and sevenfold in the average CD, which in turn leads to a decrease in the variation in the random time between the creation of two fully-formed batches. This results in the third and final area where drones make an impact, that is, reducing the lab delay.

As mentioned in the previous section, clinics in villages and rural areas typically have a higher TAT compared to those in major cities. We find that by minimizing the average TAT, drone
Note. The colors correspond to the clinic-lab assignment. Diamonds with darker colors represent laboratories, and clinics are marked with circles with lighter colors. The size of a circle is proportionate to the value of $\bar{\theta}$. Filled circles denote drone assignment.

Figure 8 (Color Online) Interplay of $\bar{\theta}$ and drone network structure under min-CVaR model (left) and min-TAT model (right)

Allocation tends to concentrate on urban areas and major cities. On the other hand, minimizing the conditional value at risk associated with the TAT distribution prioritizes clinics in rural areas and villages, resulting in a more equitable allocation of resources in terms of the diversity of populations. These results illustrate that drones have a great capability to improve testing delays in villages. However, clinics in major cities have a higher sample arrival rate, which in turn means that drones would have a higher utilization in urban areas. This means that in villages, drones have a greater impact on a small number of infants, while in urban areas and major cities, they have a smaller impact on a large number of infants.

Beyond reducing the TAT, drones have other potential benefits in addressing availability and accessibility challenges through various means. For example, sample transportation can be done regardless of the location of clinics. This is particularly useful for out-of-reach areas as explained in the introduction. With the rapid progress of technological advancement, certain models are
Note. The colors correspond to the clinic-lab assignment. Diamonds with darker colors represent laboratories, and clinics are marked with circles with lighter colors. The size of a circle is proportional to the value of $T^{K\setminus D}$. Filled circles denote drone assignment.

Figure 9 (Color Online) Interplay of $T^{K\setminus D}$ and drone network structure under min-CVaR model (left) and min-$\overline{TAT}$ model (right)

designed to be entirely resilient to varying weather conditions. The above-the-ground movement of drones eliminates the need for functioning roads, which might also be a major challenge in most developing countries. Finally, drones provide the opportunity for public excitement which might further incentivize follow-up and treatment initiation.

Despite all the potential benefits, before realizing and deploying a drone network, various operational, regulatory, technical, and educational challenges necessitate resolution. As these concerns fall beyond the scope of this paper, we do not delve into them here. However, for a comprehensive understanding of these challenges, we refer the reader to the research of Boutilier and Chan (2022) and Eksioglu et al. (2023) for detailed explanations. These challenges create interesting new research opportunities. The potential benefits of drones are not limited to sample transportation EID of HIV networks; this work is then a useful first step to analyzing and exploring drone-enabled

Electronic copy available at: https://ssrn.com/abstract=4800799
sample transportation. Another interesting direction would be the incorporation of routing decisions in sample transportation. Since blood samples are collected on special paper sheets, they do not require much space and weigh little. A drone can thus be used to collect samples from several clinics in a tour rather than performing round trips. This reduces the number of drones required, thus reducing the necessary network budget. However, using tours for sample collection may increase the TAT due to the additional collection delays. These frictions generate several interesting questions to explore.

References


Gibson E, Deo S, Jónasson JO, Kachule M, Palamountain K (2023) Redesigning sample transportation in malawi through improved data sharing and daily route optimization. Manufacturing & Service Operations Management.


Sakasegawa H (1977) An approximation formula $l_q \approx \alpha \cdot \rho^b/(1 - \rho)$. *Annals of the Institute of Statistical Mathematics* 29(1):67–75.


