ADVISER: AI-Driven Vaccination Intervention Optimiser for Increasing Vaccine Uptake in Nigeria

Vineet Nair[1], Kritika Prakash[1], Michael Wilbur[2], Aparna Taneja[1], Corinne Namblard[3], Oyindamola Adeyemo[3], Abhishek Dubey[2], Abiodun Adereni[3], Milind Tambe[1], Ayan Mukhopadhyay[2*]

[1]Google Research, India
[2]Vanderbilt University, USA
[3]HelpMum, Nigeria

{vineetn, kritikaprakash, aparnataneja, milindtambe}@google.com,
{michael.p.wilbur, abhishek.dubey, ayan.mukhopadhyay}@vanderbilt.edu,
corinne.namblard@gmail.com, {oyindamola, biodun}@helpmum.org

Abstract

More than 5 million children under five years die from largely preventable or treatable medical conditions every year, with an overwhelmingly large proportion of deaths occurring in under-developed countries with low vaccination uptake. One of the United Nations' sustainable development goals (SDG 3) aims to end preventable deaths of newborns and children under five years of age. We focus on Nigeria, where the rate of infant mortality is appalling. We collaborate with HelpMum, a large non-profit organization in Nigeria, to design and optimize the allocation of heterogeneous health interventions under uncertainty to increase vaccination uptake, the first such collaboration in Nigeria. Our framework, ADVISER: AI-Driven Vaccination Intervention Optimiser, is based on an integer linear program that seeks to maximize the cumulative probability of successful vaccination. Our optimization formulation is intractable in practice. We present a heuristic approach that enables us to solve the problem for real-world use-cases. We also present theoretical bounds for the heuristic method. Finally, we show that the proposed approach outperforms baseline methods in terms of vaccination uptake through experimental evaluation. HelpMum is currently planning a pilot program based on our approach to be deployed in the largest city of Nigeria, which would be the first deployment of an AI-driven vaccination uptake program in the country and hopefully, pave the way for other data-driven programs to improve health outcomes in Nigeria.

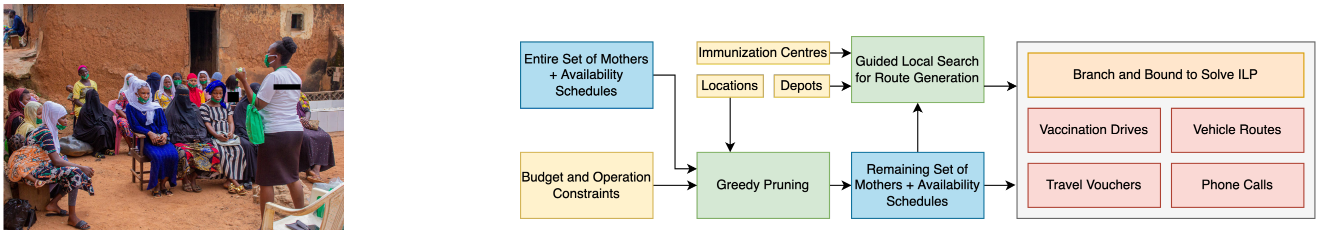
1 Introduction

The state of maternal and infant health in Nigeria is appalling. The estimated maternal mortality rate in Nigeria is about 814 per 100,000 live births; in comparison, Poland and Italy's maternal mortality rate is 2 deaths per 100,000 live births. In fact, Nigeria alone accounts for more than **10% of maternal deaths globally**, while only accounting for 2.6% of the world's population [World Health Organization, 2014]. Infant deaths in the country are also shockingly high—**Nigeria loses 2300 children under five years of age daily** [Okwuwa and Adejo, 2020]. The sustainable development goals (SDG 1 and SDG 3) aim to mobilize resources to the developing world to address inequity due to poverty and end preventable deaths of infants completely [United Nations, 2020]. However, we are far from achieving these goals.

In collaboration with HelpMum, a large non-profit organization based in Nigeria, we identify three significant challenges contributing to high mortality rates among mothers and infants. First, with an immunization rate of 13% for children between 12-23 months, Nigeria has the **lowest vaccination rate in Africa**. While vaccination is available for free in Nigeria, lack of awareness about the importance of vaccination is one of the major concerns for the low uptake of vaccination. Second, HelpMum identified that a primary driver for mothers not taking their children for vaccination is the **high transportation cost relative to their income**; we found that 46% of families analyzed as a part of this study earned less than $25 per month. Third, although several organizations, such as HelpMum, strive to design interventions for at-risk mothers and children, there is a gross **imbalance between resource availability and demand for healthcare services**.

HelpMum works closely with several local and state governments in Nigeria. Even though it uses a vaccination tracking system to remind mothers of upcoming vaccination, it has proven to be ineffective in practice. As part of this project, four new interventions were designed to increase vaccination uptake in Nigeria. HelpMum (including its advisory board) and domain experts guided the design of each intervention. The interventions (described later in the paper) are geared towards increasing the awareness of vaccination, reminding mothers about upcoming vaccinations for their children, providing accessibility to vaccination centers by operating a pick-up and drop-off service, and conducting a door-to-

*Corresponding Author



(a) (b)

Figure 1: (a) HelpMum reaches out to low-income neighborhoods to distribute clean birth kits and increase awareness of getting vaccinated (Image courtesy of HelpMum; permission was taken from each participant about sharing the picture). (b) A high-level overview of **ADVISER**: **A**I **D**riven **V**accination **I**ntervention Optimi**SER**. We formulate the allocation of heterogeneous resources as an integer linear program (ILP). We use a greedy pruning strategy to make the ILP tractable and use guided local search to generate promising vehicle routes.

door vaccination delivery program. However, matching the interventions to specific individuals presents a challenge—there exist far too many eligible recipients for the interventions as compared to the available resources. Moreover, the interventions do not necessarily guarantee successful vaccination. For example, HelpMum has observed that, at times, parents fail to bring their children for immunization despite repeated phone calls about upcoming vaccination schedules. These domain-specific challenges require that the allocation of limited resources is optimized under uncertainty of the outcomes.

We present a principled framework for optimizing the allocation of heterogeneous health interventions under uncertainty. Our approach, named **ADVISER**: **A**I **D**riven **V**accination **I**ntervention Optimi**SER**, can guide non-profit organizations and government agencies to increase vaccination uptake in resource-constrained geographies, that are crucial to achieving key goals outlines in SDG 1 and SDG 3 of the United Nations [United Nations, 2020]. Our approach is based on formulating an integer linear program (ILP) to maximize the cumulative probability of success of heterogeneous interventions under uncertainty. However, in our setting, it is infeasible to generate or solve the ILP directly. In this paper, we tackle the challenges in a principled manner through the following contributions: **1)** We present a formulation for optimizing the allocation of heterogeneous health resources under uncertainty. **2)** We present a heuristic approach to prune the decision space of the ILP by leveraging the structure of the problem. We also present a theoretical bound on the objective value attainable by the heuristic with respect to the optimal solution to the ILP. **3)** We show how guided local search can be used to generate promising vehicle routes based on the probability of specific individuals *requiring* the pickup service for vaccination. **4)** We estimate the success of interventions through historical data and community surveys. **5)** Experimental results using real-world data demonstrate that the proposed approach significantly outperforms baseline approaches. HelpMum is currently developing a pilot plan to deploy ADVISER in the largest city in West Africa. To the best of our knowledge, our solution would be the first AI-enabled program for increasing vaccination uptake in Nigeria.

2 Related Work

We present a condensed literature review here; a more extensive description of prior work is provided in the appendix (see section A.1)[1]. The general problem of resource allocation has been well studied in literature in the context of health interventions [Bishop *et al.*, 2021; Mukhopadhyay *et al.*, 2017]. Augmenting traditional vaccination campaigns by increasing awareness and optimizing vaccine supplies has shown increased uptake in South Africa [Bishop *et al.*, 2021]. Dynamically adapting the vaccine delivery under limited resources and an evolving epidemiological situation has also been explored in the context of the COVID-19 pandemic [Matrajt *et al.*, 2021]. The optimization of healthcare resources can either be done in a *single-shot* manner or by considering the sequential nature of the decision-making problem, depending on the specific use-case. Single-shot optimization problems have been used for the optimization of patient admissions in hospitals [Hulshof *et al.*, 2013], allocating home healthcare services [Aiane *et al.*, 2015], and redistribution of patients among hospitals in case of a surge in demand [Parker *et al.*, 2020]. In principle, our problem setting and formulation is most similar to the work done by Aiane *et al.* [Aiane *et al.*, 2015]; however, their setting only optimizes for travel times of resources (e.g., healthcare providers) and is not scalable (the number of decision variables and constraints in our problem is 10^6 times higher). Prior work has also explored performing sequential decision-making in the context of resource allocation in healthcare settings [Mate *et al.*, 2021; Nishtala *et al.*, 2021]. The key difference of such approaches to our setting is that we focus on a single-shot optimization problem within a fixed time window (which can be repeated to multiple time windows). The sequential nature of the problem is irrelevant in our setting—each individual in our case is only provided with a single intervention during the time period considered.

3 Problem Formulation

Problem Setting. Our problem setting involves resource allocation to M individuals (e.g., mothers) over T days. We use $[M]$ and $[T]$ as shorthand for $[1, \ldots, M]$ and $[1, \ldots, T]$ respectively. While our goal is to ensure that children get vac-

[1]https://ayanmukhopadhyay.github.io/files/adviserAppendix.pdf

cinated, mothers typically take their children for vaccination in our geographic area of interest. As a result, we say that the interventions are designed for mothers. We assume that each mother is eligible for an intervention for a fixed number of contiguous days within these T days (depending on the last date when her child was vaccinated). The binary variable a_{mt} denotes whether mother $m \in [M]$ is eligible at time $t \in [T]$; $a_{mt} = 1$ if and only if the mother is eligible on day t, and is 0 otherwise. In order to get vaccinated, mothers can either travel to designated health centers, or healthcare officials can visit a mother's house. We divide the region of interest into a grid G consisting of equally sized cells. Each mother's residence and each health center therefore map to unique cells in G. We use d_{mg} to denote the distance of mother m's residence from cell $g \in G$.

Interventions. In collaboration with HelpMum and domain experts, we design four new interventions: **1)** *Phone call*: A phone call is made to the mother reminding her about upcoming vaccination. We denote this intervention by i_c. **2)** *Travel Voucher*: A travel voucher is provided to the mother to commute to vaccination centers. We denote this intervention by i_t. **3)** *Bus Pickup*: A bus can pick up a mother (and her child) from her residence and drop them at a vaccination center. Each bus has a capacity of γ_ℓ (for ease of exposition, we assume that γ_ℓ denotes the number of mothers that a bus can accommodate with their children). We denote this intervention by i_ℓ. F denotes the set of buses. **4)** *Vaccine Drive*: A health worker goes to a designated locality and vaccinates mothers (children) living nearby who are eligible for vaccination. Naturally, there is a cap on the number of vaccinations a health worker can provide in a day. We denote this cap by γ_v, and denote this intervention by i_v.

We use I to denote the set of interventions, and for notational convenience, add no-intervention/empty-intervention, denoted as i_n, to this set. HelpMum considers i_v to be highly effective in practice, followed by i_b, i_t and i_c (in decreasing order of effectiveness). Each intervention has a cost associated with it; we use e_j to denote the cost associated with intervention $i_j \in I$. Naturally $e_n = 0$ (the cost of no-intervention is 0), and $e_\ell > e_v \gg e_t > e_c$. In particular, employing a bus pickup or the cost of conducting a vaccine drive is relatively much more expensive than giving a travel voucher or making a phone call to a single mother.

Outcomes. Let p_{mj} be the probability of mother m taking her child for vaccination given intervention $i_j \in I$.

Decision Variables. Given grid G, mothers M, and a time horizon of $[T]$ days, we optimize over the allocation of interventions I. We use g and t to denote an arbitrary cell in G and an arbitrary day in $[T]$ respectively. Let x_{tg} be a binary variable that denotes the decision to conduct a vaccine drive, i.e., x_{tg} is 1 if and only if there is a vaccine drive at cell g on day t, and 0 otherwise. We point out that a vaccination drive at a cell does not necessarily target every mother in that cell. A healthcare official can only visit a fixed number of households, and our optimization formulation must optimize which mothers to target during a drive. If possible, the healthcare worker will travel to nearby cells as well.

Let R_f denote the set of routes that a bus $f \in F$ can operate (we explain constraints specific to routes later; we first present our optimization formulation here for ease of exposition). We use a binary variable q_{tfr} to denote the routes that are chosen for operation, i.e., $q_{tfr} = 1$ if bus f operates on route $r \in R_f$ on day $t \in [T]$. Note that a specific route can only potentially target a subset of the mothers based on their locations. We use binary values s_{mtfr} to denote whether mother m can be picked up by a bus $f \in F$ operating on route $r \in R_f$ on day $t \in [T]$.

We use additional u, y, and z variables to match specific interventions to each mother. The variable $y_{mtj} = 1$ if mother m is given intervention $i_j \in \{i_n, i_c, i_t\}$ at time t. For interventions i_ℓ and i_v, we have variables u and z such that: a) $u_{mtfr} = 1$ if mother m is picked up by bus f employing route $r \in R_f$ on day $t \in [T]$, and b) $z_{mtg} = 1$ if mother m is targeted on day t by a vaccination drive conducted at cell g.

Objective Function. Formally, we seek to optimize the following objective:

$$u^*, q^*, x^*, y^*, z^* = \underset{u,x,y,z}{\arg\max} \sum_{m\in[M]} \sum_{i_j\in\{i_n,i_c,i_t\}} y_{mtj} p_{mj} \tag{1}$$

$$+ \sum_{m\in[M]} \sum_{t\in[T]} \sum_{g\in[G]} z_{mtg} p_{mv} + \sum_{m\in[M]} \sum_{t\in[T]} \sum_{f\in[F]} \sum_{r\in R_f} u_{mtfr} p_{m\ell}$$

We seek to maximize the cumulative probability of successful vaccination given a fixed overall budget b by finding the optimal allocation of interventions I among the M mothers.

Constraints. We need to enforce the following constraints given our problem setting:
1. *Eligibility Constraints*: Each mother must be eligible for the vaccine when she is being targeted for an intervention.

$$y_{mtj} \le a_{mt} \quad \forall m \in [M], t \in [T], i_j \in \{i_n, i_c, i_t\}$$
$$z_{mtg} \le a_{mt} \quad \forall m \in [M], t \in [T], g \in [G]$$
$$u_{mtfr} \le a_{mt} \quad \forall m \in [M], t \in [T], f \in [F], r \in R_f$$

2. *Vaccine Drive Constraints*: a) If a mother is being targeted for a vaccination drive at a given location and time, there must exist such a drive, b) only mothers that live within distance σ of a drive can be targeted for the drive, and c) at most γ_v mothers can be targeted by a single drive. The last two constraints denote operational limitations of conducting door-to-door vaccination drives.

$$z_{mtg} \le x_{tg} \quad \forall m \in [M], t \in [T], g \in [G]$$
$$z_{mtg}\, d_{mg} \le \sigma \quad \forall m \in [M], t \in [T], g \in [G]$$
$$\sum_{m\in[M]} z_{mtg} \le \gamma_v \quad \forall t \in [T], g \in [G]$$

3. *Route Constraints*: If a mother is being being picked by a bus on a route on a particular day, then a) the mother should be eligible to be picked on that route, b) the bus must employ that route on that day, and c) each bus can pick up at most γ_ℓ mothers, and d) in addition, given current resource limitations of our partner agency, we consider that a bus can only operate a single route on a given day.

$$u_{mtfr} \le s_{mtfr} \quad \forall m \in [M], t \in [T], f \in [F], r \in R_f$$
$$u_{mtfr} \le q_{tfr} \quad \forall m \in [M], t \in [T], f \in [F], r \in R_f$$
$$\sum_{m\in[M]} u_{mtfr} \le \gamma_\ell \quad \forall t \in [T], f \in [F], r \in R_f$$

$$\sum_{r\in R_f} q_{tfr} \leq 1 \quad \forall t \in [T], f \in [F]$$

Note that each vehicle route must obey general routing constraints, e.g., there are restrictions on the earliest pick-up times and the latest drop-off times in our setting. We assume that all routes in R_f, $\forall f \in F$ obey these constraints (for now) to simplify the discussion (discussed in detail in section 3.1).
4. *Intervention constraint*: We consider that each mother can be targeted for at most one intervention, i.e. for all $m \in [M]$,

$$\sum_{t\in T}\sum_{i_j\in\{i_c,i_t\}} y_{mtj} + \sum_{t\in T}\sum_{g\in G} z_{mtg} + \sum_{t\in T}\sum_{f\in[F]}\sum_{r\in R_f} u_{mtfr} \leq 1$$

5. *Budget Constraint*: The total cost of the interventions can not exceed the monetary budget b of the organization.

$$\sum_{m\in M}\sum_{t\in T}\sum_{i_j\in\{i_c,i_t\}} y_{mtj}\cdot e_j + \sum_{t\in T}\sum_{g\in G} x_{tg}\cdot e_v$$
$$\sum_{t\in T}\sum_{f\in[F]}\sum_{r\in R_f} q_{tfr}\cdot e_\ell \quad \leq b$$

3.1 Routing Formulation

We formulate a vehicle routing problem with time windows (VRPTW) [Toth and Vigo, 2002] to schedule vehicles to pick up mothers (and their children) and take them to a vaccination center. In practice, the mothers need to be taken to the health centers and dropped back to their resp. residences. However, we only discuss routing to the health centers to simplify the discussion. All vehicles begin operation from fixed spots (parking locations rented by HelpMum) called *depots*. Note that on day t, only a subset of mothers are eligible for vaccination, i.e., $a_{mt} = 1$. Let $\beta_e(m)$ and $\beta_l(m)$ denote the earliest and latest times on which mother m can be picked up. The times vary across the population based on occupation and other beneficiary specific constraints. Let the set of vaccination centers operating on day t day be S_t. HelpMum requires that mothers are dropped off at a vaccination center early so that there is sufficient time for them to get their children vaccinated. Let the earliest and latest drop-off times for a vaccination centre $s \in S_t$ be denoted $\beta_e(s)$ and $\beta_l(s)$ respectively. The set of pick-up locations (mothers' residences) and drop-off locations (vaccination centers) represent the nodes (N) of a graph with the road network being the edges.

A route plan is denoted by an ordered sequence of nodes $\theta = \{n_1, n_2, ...\}$, where $n_j \in N$ is an arbitrary node that the vehicle needs to visit en-route. We attach the following information with each node in a route plan. First, $\beta_{\theta,e}(n_k)$ and $\beta_{\theta,l}(n_k)$ is the earliest and latest the vehicle can arrive at location n_k, and is set to $\beta_e(m)$ and $\beta_{\theta,l}(n_k)$ respectively, when the node corresponds to a pickup location for a mother $m \in M$, or $\beta_e(s)$ and $\beta_l(s)$ respectively when the node corresponds to a vaccination site $s \in S_t$. Second, the scheduled arrival time for the vehicle servicing route plan θ to location n_k is $\delta_\theta(n_k)$. A route plan is feasible if all time window constraints are satisfied and all mothers who are picked up (up to a maximum capacity of each vehicle) are dropped off at a vaccination site. The time window constraints are satisfied for each location if $\beta_{\theta,e}(n_k) \leq \delta_\theta(n_k) \leq \beta_{\theta,l}(n_k), \forall n_k \in \theta$. For each vehicle $f \in F$, the set of feasible routes contain all routes that obey all the routing and capacity constraints.

4 Approach

We face three specific challenges in directly solving optimization problem (1). First, the ILP consists of more than 10^9 decision variables and constraints. Second, we must generate the set of feasible routes as an input to the ILP. In our problem setting, the number of routes exceed 10^{200}. Third, we must estimate the success of each intervention for each mother. We tackle these challenges in a principled manner below. Our approach (shown in Figure 1 (b)) is based on pruning the search space of the decision variables by greedily conducting the most *efficient* intervention; specifically, we greedily use the given budget to conduct the intervention that has the highest success-to-cost ratio. In this case, conducting a vaccination drive is relatively more expensive than making a phone call, but it enables HelpMum to target more mothers and guarantee more successful vaccinations. After the greedy allocation of the vaccination drives we use guided local search to generate promising vehicle routes, which is fed to the ILP as an input.

Greedy Pruning. We assume that the probabilities p_{mj} are known for each mother and intervention. We describe how such parameters can be estimated later. We use an iterative approach for pruning the size of the ILP. At each step, a set of mothers are chosen for intervention and removed from consideration. Let $[M_w]$ denote the set of mothers at the beginning of iteration w. Let H^w denote a matrix of size $G \times T$, where each entry in the matrix (denoted by H^w_{gt}) captures the utility of conducting a vaccination drive at cell $g \in [G]$ and time $t \in [T]$. At iteration w, let g^*_w on day t^*_w denote the optimal cell and day to conduct an intervention, given $[M_w]$. The matrix H^w is used to choose cell-time combinations to conduct vaccination drives in each iteration. The cell-time positions chosen in the previous iterations are updated to -1 in the matrix to remove them from consideration in future iterations, i.e., $g^*_{w'}, t^*_{w'}$ for all $w' < w$ is set to -1. We point out that conducting a vaccination drive at a cell can potentially target mothers from nearby cells as well, depending on the number of households in consideration and the manner in which door-to-door vaccine delivery is done. Let M_{gt} denote a subset of mothers in $[M_w]$ who live within some exogenously specified distance σ of cell g and are eligible for vaccination at time t. H^w is then computed such that $H^w_{gt} = -1$ if $g = g^*_{w'} \wedge t = t^*_{w'}\ \forall w' < w$ and $U^w(g,t)$ otherwise, where $U^w(g,t) = \max_{S\subseteq M_{gt}, |S|\leq\gamma_v} \sum_{m\in S}(p_{mv} - p_{mn})$ denotes the utility of conducting a vaccine drive at cell g at time t on iteration w by targeting the subset of mothers who provide the most *gain* over no interventions. Note that the utility depends on w, because mothers who are targeted for intervention in an iteration are removed from consideration on the next iteration. As an example, conducting a vaccination drive at the same location on two successive days won't be ideal, if every mother eligible for vaccination is already benefited by the service on the first day. Then, $g^*_w, t^*_w = \arg\max_{g,t} H^w_{g,t}$. In our formulation, $p_{mv} = 1$. We denote the mothers targeted as part of the conducting a drive at g^*_w, t^*_w as $S^w_{g^*t^*} = \arg\max_{S\subseteq M_{g^*t^*}, |S|\leq\gamma_v} \sum_{m\in S}(p_{mv} - p_{mn})$.

We drop references to w to simplify the discussion. In the proposed heuristic, we decide to conduct a vaccine drive at g^*, t^* if $e_t \cdot |S_{g^*t^*}| \geq e_v$, i.e., if the cost of conducting a vaccine drive is at most the cost of giving travel vouchers to the mothers being benefited by the drive. Note that conducting a vaccine drive is a better intervention than providing travel vouchers even when they cost the same (as children are guaranteed to be vaccinated through the former strategy). However, our pruning strategy is *lazy*; we commit to conducting a drive only if the aforementioned condition is satisfied and leave other decisions for the pruned ILP. The mothers who are mapped to vaccine drives are removed from consideration at the next iteration and the budget is updated accordingly. We stop pruning after we are left with some exogenously specified budget parameter (b'). The complete algorithm is presented in the appendix (see section A.6).

Performance Bounds. Let k denote the number of vaccination drives determined in the greedy pruning phase, and M_{VH} denote the set of mothers targeted by these k vaccination drives. We bound the loss incurred through greedy pruning relative to the optimal solution of the ILP. We assume that the optimal solution has at least k vaccination drives of size at least e_v/e_t; we verify this empirically in multiple parameter settings. We arbitrarily choose k vaccine drives from such an optimal solution, and denote by M_{VI} the set of mothers targeted by these k vaccine drives. Also, let mother $m \in M$ be given intervention i^*_m by such an optimal solution. We begin by proving the following proposition, which is an outcome of our greedy choice made at every iteration during pruning.

Proposition 1. $\sum_{m\in M_{VH}}(p_{mv} - p_{mn}) \geq \sum_{m\in M_{VI}}(p_{mv} - p_{mn})$ *(proof in Appendix A.2).*

Let $M_{VH} \setminus M_{VI}$ (resp. $M_{VI} \setminus M_{VH}$) denote the set of mothers in M_{VH} (resp. M_{VI}) but not in M_{VI} (resp. M_{VH}). We use Proposition 1 to prove the following theorem.

Theorem 1. *Let O_H be the objective value of the solution derived from our heuristic procedure and O^* be the objective value of the optimal ILP solution. Then $O_H \geq O^* - (\sum_{m\in M_{VH}\setminus M_{VI}}(p_{mi^*_m} - p_{mn}))$ (proof in Appendix A.3).*

The lower bound on the objective value of the heuristic approach in Theorem 1 depends on the interventions provided in the optimal solution to mothers in M_{VH} but not in M_{VI}.

Route Generation. In principle, we could generate all the feasible routes given the routing constraints, which can then be provided as an input to the optimization problem. However, route generation is intractable in our setting. The total number of routes in our setting exceeds 10^{200}. As a result, we focus on generating a smaller subset of *promising* routes. Recall that our overall goal is to maximize the cumulative probability of successful vaccination; as a result, it is imperative that given the limited number of vehicles, we pickup mothers that need the ride the most. We capture this idea to define the utility of a route plan. Let $p(\theta, n_k) = p_{m\ell} - p_{mn}$ denote the utility of an arbitrary pick-up node $n_k \in N$ in the routing graph (recall that each pickup node corresponds to a unique mother). The quantity $p_{m\ell} - p_{mn}$ captures the importance of providing a bus pickup to mother m over giving no intervention. Given the utility function, we use guided local search [Kilby *et al.*, 1999] to generate a subset of routes that maximize the utility (see section A.9 in the appendix).

Parameter Estimation. Note that optimization problem (1) requires estimates of the probability of success of each intervention for each mother. However, estimating the probabilities presents a challenge—the interventions — conducting vaccine drives, operating vehicle routes, and providing travel vouchers, are designed as part of this research; as a result, we lack exact historical data about the interventions. We only have data about phone calls that HelpMum made to all mothers. We compute the probability of success of untested interventions (i.e., vaccine drives, bus pickups, and travel vouchers) through a community survey that HelpMum performed. We estimate the probability of successful vaccination through phone calls and the effect of no interventions by learning a regression model on historical data. For brevity, we describe the estimation procedure in the appendix (see section A.5).

5 Experiments

Data. We collect anonymous information from HelpMum for 500 mothers registered as part of a vaccination tracking system operated by HelpMum. Each data point consists of several features such as the income level of the family, whether the mother received a reminder about the upcoming vaccination appointment, whether she took her child for vaccination, and the age of her child, among others. HelpMum obtained consent from each beneficiary for anonymous data sharing. We present a detailed description of each feature in the appendix (see A.4). We also collect the geographic locations of all 32 vaccination centers in our area of interest. A map with the vaccination centers and the rented parking locations is presented in the appendix (see section A.4).

HelpMum plans to deploy the ADVISER framework to all mothers in their vaccination tracking system (about 40000 mothers). As a result, we use the available data to generate two synthetic datasets (D1 and D2), consisting of 40000 mothers each. We generate features of mothers in D1 and D2 by sampling each feature independently and uniformly at random from the original data of 500 mothers. For the mothers in $D1$, we compute the probabilities of success for each mother given an intervention as follows: for each $m \in [M]$, we choose p_{mn} uniformly at random from $(0, 1)$; followed by p_{mc} uniformly at random from $(p_{mn}, 1)$; p_{mt} uniformly at random from $(p_{mc}, 1)$; and $p_{m\ell}$ uniformly at random from $(p_{mt}, 1)$. We set $p_{mv} = 1$ based on community feedback (see section A.4). D1 essentially captures the domain knowledge that we have about the interventions, specifically, for a mother m, $p_{mn} \leq p_{mc} \leq p_{mt} \leq p_{m\ell} \leq p_{mv} = 1$. For D2, we estimate the probability of vaccination given no intervention (and vaccination given phone calls) by training a logistic regression model on the original data (details in Appendix A.4). The probabilities for remaining interventions, p_{mt}, $p_{m\ell}$, and p_{mv} are chosen in a similar manner as in D1.

Baseline Algorithms. While prior work does not consist of approaches that optimize the allocation of heterogeneous health resources under uncertainty, we consider the following baselines: **1) Real-world Baseline (RWB)** For the first baseline, we asked HelpMum to allocate the interventions solely based on domain expertise. HelpMum identified key areas of importance to conduct vaccination drives in each lo-

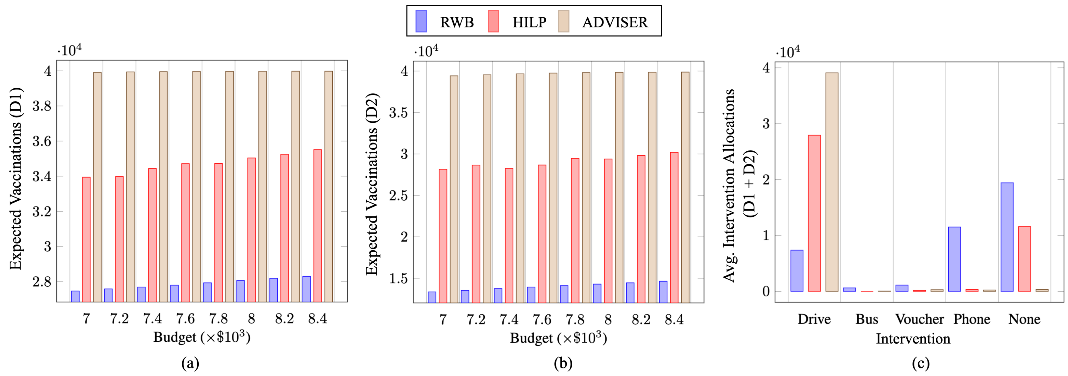


Figure 2: (a) Expected number of vaccinations in D1. (b) Expected number of vaccinations in D2. (c) The distribution of the interventions through the different algorithms. We observe that ADVISER outperforms the other two baselines. Also, both ADVISER and HILP choose vaccination drives as the dominant intervention; RWB's poor performance can be explained by having fixed allocations for each intervention.

cal government (similar to administrative jurisdictions), operate fixed routes to each vaccination center, distribute travel vouchers to mothers living more than 10 kms from a vaccination center, and call mothers by prioritizing the age of the child (the younger the child, the higher the priority). **2) Hierarchical ILP (HILP)** Motivated by the use of hierarchical planning to create tractable approaches for resource allocation [Zhang *et al.*, 2016; Pettet *et al.*, 2021], we design a baseline that solves optimization problem (1) in a hierarchical manner. First, we leverage the geographic density of the beneficiaries to identify clusters (using k-means [MacQueen and others, 1967]). The overall budget is distributed across clusters in proportion to the number of mothers in each cluster. A separate ILP is then solved directly for each cluster. We describe the exact parameters of both the approaches and results from the clustering algorithm in the appendix (see section A.7).

Experiment Setup. In consultation with HelpMum, we set the costs as follows: $e_c = \$0.1, e_t = \$1.1$, $e_v = \$15$, and $e_\ell = \$20$. We optimize the allocation of resources for $T = 30$ days and $\gamma_v = 100$. We vary the overall budget b between \$7000 to \$8400, and use $b - \$1000$ as a threshold for the greedy pruning procedure. Our implementation is available online[2]. All experiments were run on a Linux machine with 64GB RAM and an 8-core AMD processor. We implement ADVISER using the Python programming language and solve the ILP using Google OR Tools with SCIP solver.

Results. We show the expected number of successful vaccinations in D1 and D2 in Figure 2 (a) and Figure 2 (b) respectively. We observe that the expected vaccine uptake achieved via ADVISER is more than 39970 for all the budgets considered in the experiment, whereas the average vaccine intake achieved by the baseline algorithm is at most 26000 (we performed the simulation on 40000 mothers). The average number of mothers who received intervention through ADVISER is 39672, in comparison to 28426 and 20588 through HILP and RWB respectively. We also observe in Figure 2 (c) the distribution of the interventions; as expected, both the ILP-based approaches capitalize on solutions with more vaccination drives. However, we point out the importance of the other interventions as well. In practice, the number of vaccination drives is bound by the number of available healthcare workers for the service, whose regular job is to work at healthcare centers. We observed that when the number of vaccination drives is restricted to 400 per month, the average number of mothers who are targeted for pickups more than triples as compared to Figure 2 (c) (result presented in appendix A.8). Moreover, HelpMum seeks to utilize ADVISER to improve antenatal care for pregnant mothers as well, which will lower the realization of γ_v (number of children who can be vaccinated during a drive). Additional empirical results, such as the distribution of each intervention for each budget, outputs of the clustering approach, and running time of the approaches are provided in Appendix A.8.

[2]See https://github.com/ayanmukhopadhyay/ADVISER_IJCAI

Deployment. HelpMum is currently planning a pilot program in the largest city of Nigeria. We show an initial version of the tool based on the ADVISER framework in Figure 3.

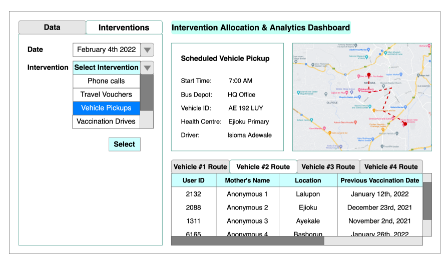


Figure 3: An initial version of the tool that HelpMum will use for deployment. The tool is under construction.

6 Conclusion

In collaboration with HelpMum, a non-profit organization in Nigeria, we present ADVISER: AI Driven Vaccination Intervention OptimiSER. Our framework can accelerate our progress towards goals in SDG 1 and SDG 3 by increasing access to healthcare services and vaccination, and by reducing maternal and infant mortality in resource-constrained settings.

Ethical Statement

Our goal in this project is to improve vaccination uptake in Nigeria. The specific interventions were designed by HelpMum in collaboration with domain experts and its advisory board. HelpMum is currently planning to pilot the ADVISER framework in collaboration with the local governments. The implications of the solution quality of any AI-driven framework for public intervention need to be studied carefully. While a large part of the interventions suggested by ADVISER are targeted towards low-income individuals (by construction), the objective function of our framework can be modified to add a score function that measures other parameters, e.g., fairness of an allocation. Also, it is possible to add arbitrary constraints on such parameters. However, the very nature and form of such constraints needs to be determined in collaboration with local stakeholders and governments.

Acknowledgements

We would like to acknowledge funding from the Google AI for Social Good Grant, Google AI for Social Good "Impact Scholars" program, National Science Foundation grant CNS-1952011, HelpMum, and Vanderbilt University. We would like to thank the staff at HelpMum for helping conduct community surveys and perform data collection.

References

[Aiane *et al.*, 2015] Daouia Aiane, Adnen El-Amraoui, and Khaled Mesghouni. A new optimization approach for a home health care problem. In *Conference on Industrial Engineering and Systems Management*, 2015.

[Bishop *et al.*, 2021] Kate Bishop, Meredith McMorrow, Susan Meiring, Sibongile Walaza, Liza Rossi, et al. An evaluation of an influenza vaccination campaign targeting pregnant women in 27 clinics in two provinces of south africa, 2015–2018. *BMC Health Services Research*, 21(1):1–8, 2021.

[Hulshof *et al.*, 2013] Peter JH Hulshof, Richard J Boucherie, Erwin W Hans, and Johann L Hurink. Tactical resource allocation and elective patient admission planning in care processes. *Health Care Management Science*, 16(2):152–166, 2013.

[Kilby *et al.*, 1999] Philip Kilby, Patrick Prosser, and Paul Shaw. Guided local search for the vehicle routing problem with time windows. In *Meta-heuristics*, pages 473–486. Springer, 1999.

[MacQueen and others, 1967] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967.

[Mate *et al.*, 2021] Aditya Mate, Lovish Madaan, Aparna Taneja, et al. Field study in deploying restless multi-armed bandits: Assisting non-profits in improving maternal and child health. *arXiv preprint arXiv:2109.08075*, 2021.

[Matrajt *et al.*, 2021] Laura Matrajt, Julia Eaton, Tiffany Leung, and Elizabeth R Brown. Vaccine optimization for covid-19: Who to vaccinate first? *Science Advances*, 7(6):eabf1374, 2021.

[Mukhopadhyay *et al.*, 2017] Ayan Mukhopadhyay, Yevgeniy Vorobeychik, Abhishek Dubey, and Gautam Biswas. Prioritized allocation of emergency responders based on a continuous-time incident prediction model. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 168–177, 2017.

[Nishtala *et al.*, 2021] Siddharth Nishtala, Lovish Madaan, Aditya Mate, Harshavardhan Kamarthi, Anirudh Grama, et al. Selective intervention planning using restless multi-armed bandits to improve maternal and child health outcomes. *arXiv preprint arXiv:2103.09052*, 2021.

[Okwuwa and Adejo, 2020] Charles Onuora Okwuwa and Simon M Adejo. Infant mortality, access to primary health care and prospects for socio-economic development in bwari area council of niger state, nigeria. *Journal of International Women's Studies*, 21(1):289–308, 2020.

[Parker *et al.*, 2020] Felix Parker, Hamilton Sawczuk, Fardin Ganjkhanloo, et al. Optimal resource and demand redistribution for healthcare systems under stress from covid-19. *arXiv preprint arXiv:2011.03528*, 2020.

[Pettet *et al.*, 2021] Geoffrey Pettet, Ayan Mukhopadhyay, Mykel J Kochenderfer, et al. Hierarchical planning for resource allocation in emergency response systems. In *International Conference on Cyber-Physical Systems*, pages 155–166, 2021.

[Toth and Vigo, 2002] Paolo Toth and Daniele Vigo. *The Vehicle Routing Problem*. SIAM, 2002.

[United Nations, 2020] United Nations. Sustainable development goals. https://www.un.org/sustainabledevelopment/health/, 2020. Accessed: 2022-05-10.

[World Health Organization, 2014] World Health Organization. Facility-based maternal death review in nigeria. https://www.who.int/maternal_child_adolescent/epidemiology/maternal-death-surveillance/case-studies/nigeria-study/en/, 2014. Accessed: 2022-05-10.

[Zhang *et al.*, 2016] Chao Zhang, Victor Bucarey, Ayan Mukhopadhyay, et al. Using abstractions to solve opportunistic crime security games at scale. In *Conference on Autonomous Agents and Multiagent Systems*, pages 196–204, 2016.