Development and Application of an Immunization Network Design Optimization Model for UNICEF

by

Henrique Ribeiro Carretti

Bachelor in Naval Engineering, Escola Politécnica, Universidade de São Paulo, 2016

and

Yuto Hashimoto

Master of International Development Policy, Sanford School of Public Policy, Duke University, 2019

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author: ____

Henrique Ribeiro Carretti
Department of Supply Chain Management
May 8, 2020

Signature of Author: ____

Yuto Hashimoto
Department of Supply Chain Management
May 8, 2020

Certified by: ____

Dr. Jarrod Goentzel
Principal Research Scientist, Center for Transportation and Logistics
Capstone Advisor

Certified by: ____

Dr. Francisco J Jauffred
Research Scientist, Center for Transportation and Logistics
Capstone Co-Advisor

Certified by: ____

Tim Russell
Research Engineer, Center for Transportation and Logistics
Capstone Co-Advisor

Accepted by:____

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Development and Application of an Immunization Network Design Optimization Model for UNICEF

by

Henrique Ribeiro Carretti

and

Yuto Hashimoto

Submitted to the Program in Supply Chain Management
on May 8, 2020 in Partial Fulfilment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Over the past five years, immunization coverage in Sub-Saharan Africa has stagnated at 72%. Immunization supply chains are expensive and complicated, and creating a model that helps optimize these supply chains is of great importance not only for people's health but also for the efficient use of limited budget in developing nations. Most existing supply chain optimization models aim at minimizing costs or maximizing profit, which do not always fit in the context of humanitarian logistics. Also, they regard the demand as exogenous, but the proximity to health centers may affect people's demand. Therefore, the objective of this project is to build an optimization model for vaccine network design that aims to maximize the access to immunization and incorporates the endogenous demand function. We take a two-step approach to build the model, and each step has a process of model formulation and validation. The first step is to formulate the toy model, a simplified version of the model using an example dataset, to understand the basic behavior of the model. With the toy model validated, we formulate the final model, which incorporates more complexities based on the real dataset. Following that, a case study of The Gambia was conducted to validate the effectiveness of our model and to provide useful insights in a real-world context regarding the applicability of our solution procedure. The results of the case study show the ability of the model to increase access to immunization. Through the opening of new outreach sites and the optimization of outreach allocation and scheduling, it would be possible to increase the immunization access from 91% to 97.1%. Furthermore, our analysis contributes by showing that the better determination of the shape of a demand coverage function is a promising area of future research.

Capstone Advisor: Dr. Jarrod Goentzel
Title: Principal Research Scientist, Center for Transportation and Logistics
Capstone Co-Advisor: Dr. Francisco J Jauffred
Title: Research Scientist, Center for Transportation and Logistics
Capstone Co-Advisor: Tim Russell
Title: Research Engineer, Center for Transportation and Logistics

ACKNOWLEDGMENTS

We would like to thank our capstone advisor, Dr. Jarrod Goentzel, and co-advisors, Dr. Francisco J Jauffred and Tim Russell, for their valuable guidance and great support. We would also like to express appreciation to our capstone partner organization, the United Nations Children's Fund (UNICEF), specially Ryan McWhorter and Olamide Folorunso, who provided us with valuable feedback, supporting data, and great insights during the whole research process. Also, from UNICEF, we express our gratitude to Buya Jallow, Richard Duncan, and Rocco Panciera, who dedicated time to our interviews and provided us with important hands-on expertise in immunization practices. Moreover, we are grateful to Pamela Siska and Toby Gooley for their advice on writing. Finally, we would like to thank our mentors, family members, and friends for their support and dedication throughout the completion of our master’s program.

Table of Contents

List of figures

List of tables

1 Introduction

A rise in immunization coverage has resulted in reduced children mortality especially in developing countries. Even though remarkable progress has been achieved in the reduction in neonatal mortality, according to an annual report issued by the United Nations Children's Fund (UNICEF), 6.3 million children lost their lives in 2017, most of it from preventable causes (UNICEF, 2019).

The context of this project is an alarming situation involving vaccination in African countries. Over the past five years, immunization coverage in sub-Saharan Africa has stagnated at 72%. These factors have prompted global immunization experts attending the 2019 Regional Immunization Technical Advisory Group (RITAG) meeting to call for strengthened routine immunization (WHO Regional Office for Africa, 2019).

UNICEF is a non-profit organization internationally recognized for its humanitarian work. One of the goals that UNICEF set is that every child survives and thrives (Goal Area 1), which involves thematic work in health, nutrition, HIV and AIDS, and early childhood development. This institution plays a vital part in the progress of immunization. In 2018, the organization provided three doses of the Pentavalent (five-in-one) vaccine for an estimated 65.5 million children. UNICEF recognizes that primary health care – integrated programs and interventions across the life cycle, delivered within strong community health systems – is the most sustainable path to achieving Goal Area 1 results and the third Sustained Development Goal: healthy lives and well-being for all (UNICEF, 2019). Through the collaboration of UNICEF's know-how in immunization and MIT's research expertise, this project tackles the issue of increasing the level of access to vaccination in less developed countries through the design of more efficient vaccination networks.

The objective of this project is to build an optimization model for vaccine network design through incorporating innovative features that enhance the maximization of immunization coverage.

1.1 Current immunization supply chains optimization model limitations

Immunization supply chains are expensive and complicated, and creating a model that helps optimize these supply chains is of great importance not only for people's health but also for the efficient use of limited budget in developing nations. Different types of network optimization can be found in the literature, such as demand allocation, vaccination location, production capacities, and batch sizes (Lemmens, Decouttere, Vandaele, & Bernuzzi, 2016). However, this project will focus on the most common and strategic decisions of the network in developing countries: the location of outreach immunization sites and fixed-health-center capacity.

Generally, supply chain optimization and facility location models ignore the impact of supply chain network design on demand and profit. Instead, they focus on cost. However, proximity to supply facilities may influence price and service sufficiently to affect customer behavior. In partnership with a chemical industry firm, MIT's Center for Transportation and Logistics (CTL) previously conducted research on the impact of supply chain network design on demand. The key research development was the creation of a flexible market share function that is endogenous to the mathematical model so that an optimized solution directly considers how the design of a supply network shapes demand. The results proved this to be the case for price- and time-sensitive commodities when decision makers chose to invest more in their supply network to drive higher demand (Russell, Jauffred, & Goentzel, 2019). The impact of these findings can go beyond this single chemical firm. This previously developed model has the potential to be applied to other cases, including non-profit health facilities to increase vaccine access.

1.2 Problem statement

Given the significance of stagnated vaccine coverage in Africa and the organizational mandate of UNICEF, it is important to design a more efficient vaccine network. The goal is to increase immunization access with responsible and efficient use of budgets. The proposed analysis helps local governments to optimize the locations and operation of outreach sites. To this end, a mathematical optimization model can be a useful tool. However, it needs to be well formulated enough to consider the unique characteristics of vaccine supply chains. The two most important characteristics specified for such a model concern its objective function and demand modelling:

1. **Objective function:** Instead of minimizing cost or maximizing profit, this model should aim to maximize vaccination access
2. **Demand function**: Rather than having demand as fixed and exogenous, demand is endogenous and based on the proximity of health centers.

It is also important to mention that our project scope, based on the discussion with UNICEF, focused on the strategic decision of utilizing only the existing fixed health centers. The model considered adding resources to existing fixed health centers and adding new outreach sites. The model did not consider the establishment of new fixed health centers.

1.3 Research questions

Based on the problem statement, the research questions of this project are as follows:

1. How should the efficient vaccine network model be formulated in order to maximize access to those seeking immunization services under certain constraints?
2. How does the proximity of population to vaccination health centers affect the vaccine coverage?
3. How can other factors that influence the demand for vaccinations be included in the model?

1.4 Study overview

In order to answer the above research questions, we started by performing a literature review of pertinent studies in vaccination network optimization, described in Section 2. Based on the results of this study and discussions with UNICEF, a set of characteristics for a simplified network optimization model were defined in Section 3.1. More in-depth discussions with stakeholders resulted in the identification of new relevant aspects that were incorporated into the formulation and resulted in a final optimization model, shown in Section 3.2. This model's efficacy was then tested in a case study of The Gambia Vaccination Network with data provided by UNICEF, discussed in Section 4. The results of this study were explained in Section 5 together with insights and recommendations on how to improve vaccination access. Finally, the study conclusions are presented in Section 6.

Previous research papers typically used approximate and aggregated data for model validation without presenting details on how the input parameters were derived. Therefore, the applicability and efficiency of such models in a professional environment are extremely limited because other users are not able to derive the necessary data input of the model.

Working with real data is often challenging and requires some assumptions and data treatment procedures. With the objective of guaranteeing the reproducibility of the developed model and its applicability in real case scenarios, the case study in Section 4 described in detail the procedures and assumptions that were required to treat the real dataset. Whenever possible data was derived from publicly available sources and its extraction procedure was explained. By doing this, we expect to assist other professionals facing the issue of handling data on their own network design models and increase the practical application of the theoretical works found in the literature.

2 Literature review

This section is organized as follows: In Section 2.1, corresponding to research question 1, we conducted a thorough research on previous vaccination network design models, and we compared them based on a set of criteria relevant to the problem. This comparison allowed us to identify some gaps and define preliminary characteristics of our model. In Section 2.2, we answered research question 2 and described how previous works have incorporated the negative impact of distance on demand. To answer the last research question, in Section 2.3, we outlined what other factors have been found in the literature that could have potential impacts on demand and how this topic could be further explored. Finally, as a conclusion, we summarized the literature review, and the identified research gaps.

2.1 Vaccine network design model

In this section, we investigated previous works in the field of vaccination network design. First, in Section 2.1.1, we identified that a wide variety of work has been done in the broad field of vaccination networks and narrowed down our detailed review to the papers with the greatest similarity to our problem. Then, In Section 2.1.2, we conducted a more detailed review of six papers and compared them based on a set of self-established criteria in order to highlight their models' characteristics.

2.1.1 Previous work on vaccination networks

We found two past studies that revised previous optimization models of vaccination networks. Both papers presented their own way of clustering the research they reviewed. Initially, Lemmens et al. (2016) clustered the past studies of network design based on the type of problem they aimed to solve, giving the following framework: (1) demand allocation; (2) vaccination location; (3) production capacities, and (4) batch sizes. A more recent work by Duijzer, Van Jaarsveld, and Dekker (2018) used four other components for classification, similarly focusing on the problem to solve. Their categorization is: (1) product; (2) production; (3) allocation; and (4) distribution (Figure 1: Classification of the vaccine supply chain and overview of characteristics

Source: Duijzer, L. E., van Jaarsveld, W., & Dekker, R. (2018). Literature review: The vaccine supply chain. European Journal of Operational Research, 268(1), 174–192.Figure 1). Although the naming of the clusters is different, we can find similarity amongst the components presented by both papers. For instance, the vaccination location in Lemmens et al. (2016) is similar to the distribution in Duijzer et al. (2018) since the decision of where the vaccination center should be located is intrinsically connected to how vaccines are distributed.

	Product	Production	Allocation	Distribution
	What kind of vaccine should be used?	**How many doses should be produced and when?**	**Who should be vaccinated?**	**How should the vaccines be distributed?**
	Right product (decision)	*Right product (realization), Right time*	*Right place (decision)*	*Right place (realization), Right time*
Similarities	- Product development (R&D)	- Long production time - Uncertain demand - Pull process: initiated by the customer (i.e., public health organisation) - Uncertain yields		- Inventory control - Facility location - Routing - Supply chain design - Perishable product - Temperature controlled chain
Unique characteristics	- Decentralized decisions: product is determined by public health organizations, not by the supplier - Public health organizations are non-profit, whereas supplier is for-profit - Product changes very frequently (yearly for annual influenza vaccine) - Product decision is made under time pressure and high demand uncertainty	- Demand externalities due to disease dynamics and the protective power of vaccinations for non-vaccinated people	- Complex decision making: political interests, equity considerations - End customer (i.e., `patient') does not pay for the product in most cases - Push process: initiated and performed in anticipation of end customer need - Decentralized decisions: end customer has no power in this phase	- Mass distribution under time pressure

Figure 1: Classification of the vaccine supply chain and overview of characteristics
Source: Duijzer, L. E., van Jaarsveld, W., & Dekker, R. (2018). Literature review: The vaccine supply chain. European Journal of Operational Research, 268(1), 174–192.

Further exploring the literature, we concluded that our literature review narrows down to two groups of papers - the distribution and vaccination location. We focused on location-distribution papers because the work with UNICEF specifically focuses on low-income countries. For vaccination networks in low-income countries, it is reasonable to assume that the decisions on product and production, such as who will be the supplying countries for the vaccines and how many doses of what type will be made available, are done by centralized planning arms of the governments (Chen, Norman, Rajgopal, Assi, Lee, & Brown, 2014). However, our model is not limited to developing countries and can be applied to any country with centralized planning of the health system. While previous papers fail to differentiate their studies based on the degree of development of the country, we believe this is an important consideration. Moreover, the decisions on allocation, such as who should receive the vaccine, are not pertinent to our objective of research since our project is to maximize the access to immunization and not to decide who should or should not receive vaccines. With the product, production and allocation decisions out of scope, our literature review narrows down to distribution aspects of the vaccination.

Furthermore, according to Chen et al. (2014), the basic structure of vaccine supply chains in these countries follows a similar structure, with a central storage location that serves as a supplier to a multi-layered chain formed by stores/hubs and clinics.

Focusing on vaccination-location issues in the low-income countries, we narrowed down the relevant studies to six papers.

2.1.2 Distribution design models in the literature

Defining the scope of further literature review, as distribution issues in low-income countries, we identified six papers which are highly relevant to the specified topic. In order to establish a methodology of comparing these six works, we propose the following criteria for classification:

- Type of objective: Minimize costs, maximize access to vaccine
- Decisions: Locations, flow between nodes, inventory level, vehicle types, storage types, and other relevant characteristics.
- Network structure: Central distribution, hubs, clinics, outreach sites
- Demand: Fixed exogenous, stochastic, causal endogenous
- Product: One vaccine, multiple vaccines
- Period: Single period, multi-period
- Specific constraints: Vehicle/storage capacities, maximal distance

Our classification criteria are similar to what has been proposed by Duijzer et al. (2018). According to them, there are three types of decisions when designing vaccination distribution networks: (1) the design of the supply chain – number of layers and its locations; (2) The inventory control policy – size and location of stock; and (3) The distribution to end users from fixed points of dispensing (clinics for the distribution of vaccines) or via mobile facilities. Although this segmentation makes sense, the greater level of detail in criteria is crucial in order to differentiate the existing models more precisely. It ultimately enables us to identify gaps and define specific characteristics of our model.

At the end of this section, a summary matrix which illustrates the different characteristics of the literature review is presented (Table 1). For the clarity of analysis, we divided the six papers into two groups based on the strong connection among the works.

2.1.2.1 Non-outreach models

In this first group, we clustered three papers which do not consider outreach sites. Chen et al. (2014) builds a general mathematical model to optimize operations in vaccination networks in low-income countries. They proposed a formulation that considers a multi-period capacitated network model with different types of storage devices, vehicles, and vaccines types. The model's main decisions are the inventory levels and flows in a multi-period time horizon. From a vaccination perspective, they only consider the possibility to fulfill the demand from fixed health centers. The most innovative aspect of its formulation is the introduction of the vaccination regimen concept that determines how many doses of each vaccine type a child must receive to be considered fully immunized. Apart from its complexity, their model is not making strategic decisions regarding location opening. It means that the main purpose of this work is to define the optimal way to operate a vaccination network that is already established under the scenario of several specific constraints, which differs a bit from our strategic level research objective.

Building upon the work done by Chen et al. (2014) and Lim (2016) proposed a model to redesign vaccine distribution networks in low- and middle-income countries, incorporating strategic location opening decisions. To this end, a sophisticated Mixed Integer Programming (MIP) model was proposed. The objective of this model is the minimization of costs. The model's main decision variables are the opening of distribution hubs for a given fixed source (national distribution center) and clinics. In addition to the hub opening, the other operational decisions of the model are the flows between nodes, types and quantity of storage devices in each facility, types of vehicles used, and the number of trips. Compared to the work developed by Chen et al. (2014), the multiple vaccines complexity was removed, but the addition of the location opening, combined with other more operational aspects, made the problem complex. Due to this complexity, the authors had to develop a novel hybrid algorithm based on MIP and an evolutionary strategy.

The most recent work addressing the optimization of vaccine distribution networks operation in low-income countries was conducted by Yang and Rajgopal (2019). This research was built upon Lim (2016) and added some complexities to the model such as flexible sizes of storage devices and a single trip constraint on deliveries. Once again, the problem complexity and size made it necessary to create a disaggregation-and-merging algorithm because standard commercial software would not be able to solve the problem directly.

Although these three papers can be a great source of the potential enhancement of our model with additional operational constraints, they considered exogenous demand and required a complex solving methodology due to the high level of operational decisions being made. Most critically, none of them incorporated outreach sites. As we describe in the next section, based on the input from interviews with UNICEF experts, we learned that the vaccination through outreach sites is highly relevant factor for the immunization efforts in low-income nations, and hence incorporating outreach sites becomes is defined as premise of our model.

2.1.2.2 Outreach models

In this section, we present three works in the second group which considered outreach sites in the vaccination network modelling. As defined by Yang and Rajgopal (2019), an outreach vaccination operation is characterized by health care professionals who are sent from fixed clinics to villages where one-day clinics are set to vaccinate the nearby population. Outreach sites are extremely important way to increase vaccination access in remote areas. This effort is aligned with the Reaching Every District strategy (RED) established by UNICEF and its partners since one of the operational components of RED is to reach target population through the development of different vaccination delivery strategies (Vandelaer, Bilous, & Nshimirimana, 2008).

The first quantitative model to determine optimal outreach trips and policies to maximize coverage was presented by Lim, Claypool, Norman, and Rajgopal (2016). This work, in addition to be the first one to consider outreach sites, also considered the distance impact in coverage – a key aspect of our research. The main decisions made by the model are where outreach sites should be placed, and fixed clinics should serve as a base for the outreach operations. An important assumption of this modelling is that each outreach must be located at a maximum distance from fixed clinics.

Building upon the work developed by Lim et al. (2016), a new formulation was created to find the design of fixed and outreach vaccination services in Hasanzadeh-Mofrad (2016). An innovative aspect of this work lies in the combination of health centers location and outreach planning by including the fact that population could also be vaccinated in health centers. Using the different approach from Lim et al. (2016), this model treats the outreach planning as a vehicle routing problem. In practice, this means that there is also a limitation on the number of vaccines that each outreach location can receive from a single trip in addition to the distance constraint already presented by Lim et al. (2016). The formulation proposed by Hasanzadeh-Mofrad (2016) is quite complex, and it includes stochasticity to the demand. However, one downside of the research is that it does not consider the effect of causal demand that is reduced with the increase in distance.

The latest work focusing on outreach planning was presented by Yang and Rajgopal (2019). Their modelling approach is very similar to Hasanzadeh-Mofrad (2016) with the main distinction of incorporating the time aspect into the outreach trip planning. In their modelling, each trip is allowed to visit multiple destinations, but it is subjected to specific time windows constraints. In this sense, the work by Yang and Rajgopal (2019) can be seen also as vehicle routing problem with time windows (VRPTW) combined with a set-covering problem (SCP).

The works proposed by Hasanzadeh-Mofrad (2016) and Yang and Rajgopal (2019) added significant complexity to the modelling of Lim et al. (2016) with the incorporation of routing aspects to outreach trips, the decision of clinics location, and service levels. However, both papers lost an important aspect of Lim et al. (2016) formulation, which is the incorporation of the negative effect of distance on immunization demand. In addition to that, the modelling approach of considering the outreach sites as vehicle routing problems, which can be performed by different vehicle types, is already in an operational level of detail which is not the scope of this research. Finally, the health center opening decision is irrelevant to our model since the health centers locations and opening will be treated as a fixed input parameter.

We concluded that the work with most similarity to the objective of our research is Lim et al. (2016).

2.1.2.3 Models summary and research gap

Based on the criteria presented in the beginning of this section, Table 1 shows our classification of the literature.

To build upon Lim et al. (2016), the following gaps were addressed in our initial formulation:

- Incorporation of the fact that population can also be served by fixed clinics
- Incorporation of a more flexible demand function instead of the stepwise or binary solution proposed in the reference paper.
- Incorporation of flexible capacity in fixed clinics as decision variable (e.g., varying number of employees or working hours)
- Incorporation of variable costs in fixed clinics depending on number of employees

Following the criteria established in the beginning of this section, and presented in Table 1, we can define our model preliminary characteristics as follows:

- Objective: Maximize access to vaccine
- Decisions: Open outreach; service level at fixed clinics.
- Structure: Clinics, and outreach sites.
- Demand: Causal
- Product: Single vaccine
- Period: Single period
- Constraints: Maximum distance between outreach and fixed clinics

Table 1: Comparison of key characteristics of previous vaccination network models

	Group 1 - No outreach Locations			Group 2 - Outreach Locations		
	1- (Chen et. Al 2014) A planning model for the WHO-EPI	**2 -(Lim , 2016)** Improving the design and (...)	**3 - (Yang, Bidhkori & Rajgopal, 2019)** Optimizing vaccine distribution (...)	**4- (Lim et. al 2016)** Coverage models to (...)	**5- (Mofrad 2016)** Optimizing vaccine clinic (...)	**6- (Yang and Rajgopal,2019)** Outreach Strategies for Vaccine (...)
Objective						
Maximize coverage	Maximize coverage			Maximize coverage		
Minimize costs		Minimize costs	Minimize costs		Minimize costs	Minimize costs
Decisions						
Open locations		Open hub	Open hub	Open outreach	Open outreach and clinic	Open outreach and clinic
Flow between nodes	Flow between nodes	Vehicle types and n°	Vehicle types and n°			
Vehicle types	Vehicle types	Vehicle types	Vehicle types		Vehicle types of outreachs	
Storage type	Storage type	Storage type	Storage type			
Inventory levels	Inventory levels	Inventory levels	Inventory levels			
Outreach routes						
Capacity			variable device capacity			
					n° outreach trips	n° outreach trips
Network structure						
Central distribution	Central distribution	Central distribution	Central distribution			
Hubs	Hubs	Hubs (variable)	Hubs (variable)			
Clinics	Clinics	Clinics	Clinics	Clinics	Clinics (variable)	Clinics (variable)
Outreaches				Outreachs	Outreachs (variable)	Outreachs (variable)
Demand						
Fixed exogenous	Fixed	Fixed				
Sthocastic			Stochastic		Stochastic	Stochastic
Causal (distance)				Causal (distance)		
Product						
Single Vaccine		Single Vaccine	Single Vaccine	Single Vaccine	Single Vaccine	Single Vaccine
Multiple vaccines	Multiple vaccines					
Period						
Single period				Single Period	Single Period	Single Period
Multiple Period	Multiple Period	Multiple Period	Multiple Period			
Constraints						
Capacity on facility	Capacity on facility	Capacity on facility	variable capacity			
Capacity on vehicles	Capacity on vehicles	Capacity on vehicles	Capacity on vehicles		capacity on vehicles	Vehicle cappacity
Other					Daily trips, one dest trip	Maximum travel time, multiple dest
Other			single trip delivereis	Dist max outreach to IHC	Patient and outreach Max travel distance	Patient Max travel distance

2.2 Distance effect on vaccination coverage

As we presented our second research question, one of the goals of this project is to understand the impact of distance on vaccination coverage and how to incorporate it into the model. The relevance of including this factor has been quantitatively assessed by Blanford, Kumar, Luo, and MacEachren (2012). After conducting study in Niger, their results showed with a 95% level of significance that the probability of children living at an hour distance to vaccination facilities to be fully immunized at one year old are 1.88 higher than children living in more distant areas. Because the distance is directly correlated to time, this is a strong indication that the distance effect should be considered when designing demand access to health care.

Some papers, in a health care context, modelled the negative impacts on coverage as distance from clinics to patients increases. The work developed by Verter and Lapierre, (2002) was the first to consider this aspect. In their maximal coverage location of preventive health care facilities problem, they assumed that the coverage reduced linearly with an increase of distance from patients to health facilities. After that, Tanser (2006) and Gu, Wang, and McGregor (2010) also incorporated this effect in their health care facilities location researches with similar approaches.

The first paper to incorporate this specific effect in their optimization vaccination networks models was Lim et al. (2016). They presented four different models that explored the effect of distance increase on reduction of coverage. Different from Verter and Lapierre (2002), however, they consider the reduction of coverage to be either binary or stepwise. In our modelling, we will initially assume a linear decrease in demand function based on the distance, similar to what was done by Verter and Lapierre (2002).

In addition, some papers studied the distance-decay effect quantitatively. For instance, Feikin, Nguyen, Adazu, Ombok, Audi, Slutsker, and Lindblade (2009) presented the impact of distance of residence from a peripheral health facility on pediatric health utilization in rural western Kenya. Distance-decay effect was confirmed by their research statistically. The rate of clinic visits by residents decreased linearly at 0.5 km intervals up to 4 km, after which the rate stabilized. Using

Poisson regression, for every 1 km increase in distance of residence from a clinic, the rate of clinic visits decreased by 34% (Table 2). The differences from our research are that their study was not applied to the immunization context and the target was specifically sick children. Ibnouf, Van den Borne, and Maarse (2007) studied time factor influencing immunization demand among children under five years of age in Khartoum State, Sudan. Walking time to the nearest place of vaccination strongly influenced the correct vaccination status of the child. Children of mothers who have better access to vaccine services (less than 30 minutes walking time to the nearest place of vaccination) were about 3.4 times more likely to have had the correct vaccinations than were children of mothers who have to walk 30 minutes or longer (Table 3).

Table 2: Rates of visits to peripheral health facilities among children living in the western Kenya by distance lived from the nearest clinic 2003–2004

Distance (km)	No. children who visited clinic based on distance lived from clinic	Child-Time (years) of children by distance lived to clinic	Rate of clinic visits by distance lived from clinic, visits/child-year (95% C.I.)	Rate Ratio of clinic visits by distance lived from clinic (95% C.I.)
0.000–0.50	402	482	0.83 (0.76–0.92)	Ref
0.501–1.00	507	832	0.61 (0.56–0.66)	0.73 (0.64–0.83)
1.001–1.50	654	1179	0.55 (0.51–0.60)	0.67 (0.59–0.75)
1.501–2.00	710	1515	0.47 (0.44–0.50)	0.56 (0.50–0.63)
2.001–2.50	607	1612	0.38 (0.35–0.41)	0.45 (0.40–0.51)
2.501–3.00	441	1240	0.36 (0.32–0.39)	0.43 (0.37–0.49)
3.001–3.50	133	555	0.24 (0.20–0.28)	0.29 (0.24–0.35)
3.501–4.00	27	374	0.072 (0.049–0.10)	0.087 (0.059–0.13)
4.001–4.50	24	268	0.090 (0.060–0.13)	0.11 (0.071–0.16)
4.501–5.00	26	224	0.12 (0.079–0.17)	0.14 (0.094–0.21)
5.001–5.50	11	159	0.069 (0.038–0.13)	0.083 (0.046–0.15)
5.501–6.00	1	16	0.063 (0.008–0.44)	0.075 (0.011–0.55)

Source: Feikin, D. R., Nguyen, L. M., Adazu, K., Ombok, M., Audi, A., Slutsker, L., & Lindblade, K. A. (2009). The impact of distance of residence from a peripheral health facility on pediatric health utilization in rural western Kenya. Tropical Medicine and International Health, 14(1), 54-61. https://doi.org/10.1111/j.1365-3156.2008.02193.x

Table 3: Percent of correct vaccination depending on walk-time

Percentage vaccination					
			Correct	Incorrect	
Overall	**(n)**	**%**	**75.1**	**24.9%**	**chi-square**
Walk time					
Less than 30 min	368	89.8	78.3	21.7	18.939***
More than 30 min	42	10.2	47.6	52.4	

Variable	OR	95.0% CI	P value
Walk-time			
30 minutes and more	1*		
29 minutes and less	3.36	1.61-7.02	0.001

Source: Ibnouf, A.H., Van den Borne, H.W. & Maarse, J.A.M. (2007). Utilization of family planning services by married Sudanese women of reproductive age. EMHJ - Eastern Mediterranean Health Journal, 13 (6), 1372-1381, 2007 https://apps.who.int/iris/handle/10665/117388

2.3 Other factors affecting the demand

The last research question, and literature strand, is to understand how other factors apart from long distances negatively affect vaccination demand and how they have been incorporated to vaccination network models.

Only one paper performed an assessment of the factors impacting access to immunization services. The case study by Echakan, Oluoch, and Osuga (2018), quantitatively and qualitatively assessed through interviews of the population and health records analysis, that distance/time is a main factor in access to immunization. One additional aspect mentioned in their work was the importance of advertisement of outreach operations so that the target population is aware of when and where outreach sites will be opened. This is an intuitive piece of information that no models have considered since it goes beyond supply chain aspects.

In addition, it can be assumed from the discussion with UNICEF experts that the quality of human resources, especially at outreach locations, significantly affects the demand of vaccination. Human resource includes both the skilled health professionals and unskilled volunteers. If their morale as well as their quality of work are high, more people will tend to get vaccinated. It is not yet clear, however, how these effects can be incorporated into optimization models.

2.4 Literature review summary and gaps

From the first objective of our literature review, to understand the characteristics of previous work done on vaccination network design, we showed that a total of six papers had similar characteristics to the object scope of our study. Of those, it was shown that in only one the demand was treated as endogenous to the model. This was the main reason for defining Lim et al. (2016) as the basis formulation for our model. The matrix comparison allowed us to compare all the works in a methodological way and to define how our model characteristics can be compared to previous research. In Section 2.2, we showed that in addition to Lim et al. (2016), other papers also incorporated effects of demand decrease based on distance, but in general we found a lack of a structured approach for defining a demand function shape and its parameters. Finally, in Section 2.3, it was shown that only one paper referenced an additional factor that could have an effect on demand which is the necessity of advertisement campaigns for the outreach sites. However, this is a factor difficult to quantify and will hardly be incorporated to supply chain optimization models.

By defining Lim et al. (2016) as the model to be used as reference to our work, we showed that other aspects can be incorporated to the formulation to increase its complexity and adherence to the case study from this project. It is important to point out that the final model characteristics will be defined in conjunction with UNICEF to achieve the required level of adherence to our specific case of study. Some papers found in the literature raise awareness for the importance to keep the formulation simple enough to allow flexibility and efficiency on the solution. Our model should not include operational level decisions such as vehicle type or storage devices types but should focus on strategic location facilities decisions. This will allow decisions makers to build the sufficient level of trust and intuition necessary for the design of efficient vaccination supply chain networks using our model.

Even though our initial toy model was based on Lim et al. (2016), the developed model was not simply be a replication of the previous research. The following gaps have been identified in their formulation and will be addressed in our model:

- Incorporation of the fact that population can also be served by fixed health centers
- Incorporation of more flexible demand function instead of the stepwise or binary solution proposed by the previous study
- Incorporation of flexible capacity in fixed health centers
- Incorporation of variable and fixed costs in fixed health centers and outreach sites

3 Methodology

The key objective of this research is to develop a network design optimization model that maximizes access to vaccination in low-income countries through the incorporation of an endogenous demand function into the model. To achieve this end, the first step is to define what characteristics the model should possess. Specifically, this entails the selection of the decision variables, demand function, constraints, and cost factors of the problem. Initially, the toy model, a simplified version of the model, was formulated to understand the basic behavior of the model, as discussed in Section 3.1.

With the toy model validated, additional discussion with UNICEF allowed us to define the characteristics of the final model, which is outlined in Section 3.2. The final formulation incorporated the fact that the capacity of the outreach sites would no longer be fixed, but instead is dependent on the amount of resources sent from the fixed health centers.

3.1 Toy model

Based on a literature review and discussion with UNICEF, we developed our initial problem and state its characteristics in Section 3.1.1 (detailed formulation is presented in Appendix A). This initial model, called toy model, had an objective of understanding the basic factors that would influence the optimal network design and served as a tool to guide our discussions with stakeholders. The toy model was tested in a small-scale problem with fictitious data. The main results and conclusions obtained with this exercise are shown in Section 3.1.2 (further details are presented in Appendix B). Furthermore, Python formulation for the toy model can be found in the GitHub repository[1]

3.1.1 Toy Model characteristics

The objective function of the toy model is to maximize the immunization access in a certain country where the vaccine is not sufficiently distributed and the potential demand for immunization services exists. The immunization can be done in either fixed health centers or outreach sites. Fixed health centers are established health centers that offer routine immunization services in the facility. Outreach sites are single-day clinics conducted in a building in a more remote community where the vaccination operation for local residents is conducted by people sent from fixed health centers. The developed optimization model must decide the number of employees necessary at the fixed health centers and which candidate outreach sites should open.

Our toy model considers that each fixed health centers has a minimum operating capacity and that the maximum capacity of each fixed health center is dependent on the number of workers of the optimal solution. Outreach sites are considered to have a fixed capacity.

Regarding costs, three types of costs are taken into consideration in the toy model: (1) direct labor cost at fixed health centers; (2) fixed implementation cost of outreach sites; and (3) the vaccine administration service cost. This last cost factor is proportional to the distance between population and immunization center. The model adopts a steady state operation.

Following our project objective, the demand for immunization was modeled as an endogenous function. This causal demand accounts for the fact that people's willingness or capacity to travel for immunization decreases as the distance from population center to immunization center increases. This approach is in line with other previous studies identified, as described in Section 2.2.

Details of the toy model formulation are presented in Appendix A, but the key characteristics of it can be summarized as follows:

---

[1] GitHub repositor URL: https://github.com/optimization-network/vaccination-networks

- Objective:
    - Maximize access to vaccine
- Immunization facilities types:
    - Fixed health centers (always open)
    - Outreach sites
- Decisions:
    - Number of employees at fixed health centers
    - Open outreach sites candidates
- Constraints:
    - Maximum distance between outreach sites and fixed health centers
    - Minimum number of employees in fixed health centers
    - Maximum number of employees in fixed health centers
    - Vaccination capacity of outreach sites
    - Total cost must be lower than available budget
- Costs:
    - Fixed cost for outreach implementation
    - Fixed health center employees' cost
    - Cost to serve, per vaccine, at fixed health center
    - Cost to serve, per vaccine, at outreach sites
- Period:
    - Single period (steady state operation)
- Product:
    - Single vaccine
- Demand:
    - Linear decay function with 100% access level at 0 km and 0% at 100km.

3.1.2 Toy model validation and conclusions

The details of the validation procedure of the toy model are presented in Appendix B. In summary, the validation was done by varying three input variables: (1) health center employee costs; (2) fixed outreach implementation costs; and (3) the linear distance coefficient of the demand function. The impacts of the input variation were evaluated in three outputs: (1) the number of employees necessary in the fixed health centers; (2) the number of outreach sites open; and (3) the total vaccination access levels obtained.

In the first scenario tested, it was observed that an extreme increase in outreach implementation costs led to an increase in the utilization of the fixed health centers. Consequently, the number of employees in fixed health centers was increased while the number of outreach sites opened was decreased. In the second scenario tested, the opposite model behavior was witnessed when we increased the cost of fixed health center employees and obtained a final solution with less participation by the fixed health centers. Finally, we demonstrated in the last scenario that the optimal solution tends to have more outreach sites opened as the effect of distance on demand increases.

Since the results from these three tests matched the expectations, we conclude that the toy model works as expected, and that the validation process was successful. Running the different scenarios allowed us to identify some gaps in the formulation and address them in the final model.

3.2 Final model

The toy model formulation was fundamental to validating the basic dynamic of the network design problem and the impact of parameter variations in the final network configuration. However, one important factor that the toy model failed to incorporate is the fact that the size of the outreach sites may vary depending on the amount of resources used, and so its capacity should vary. Another point for improvement is that the costs of the outreach operation should not also be fixed; rather, they depend on how large and distant the outreach site is from its supplying facility. In this section, is presented how those, and other factors were implemented in the model, together with its final mathematical formulation.

3.2.1 The addition of outreach resource supply variable

In the toy model formulation, the cost of an outreach operation would be the sum of a fixed implementation cost *Co*, plus the cost of vaccinating each population center *j* from the outreach site *o*, c_{oj}. This last factor would be impacted by the distance from each center *j* to the outreach *o,* with more distant centers being more expensive to serve.

This model was useful for the initial formulation of the problem and to guide the discussions with stakeholders. However, after further research and inputs from experts, three factors became clear:

1. The immunization cost from an outreach site is not dependent on the distance from the patient to the outreach site, since the fact that a person has to dislocate does not impact the cost of the operation. It undoubtedly impacts the amount of people a facility can reach, but not the costs.
2. The cost of an outreach operation is not fixed. Rather, it is driven by the amount of resources (vehicles, nurses, and vaccines) that need to be supplied from the health centers, and by their distances to the supply source.
3. The vaccination capacity of an outreach site is not fixed, but instead is dependent on the amount of resources sent.

Considering these three factors, the new formulation added a resource flow from the fixed health center *f* to the outreach site o. These resources are divided in units, called bundles, and are basically everything that is key for implementing an outreach operation: nurses, vaccines doses, cold boxes, and vehicles. The amount of resource bundles and the distance between supply point and outreach destination will be the factors used to determine the outreach capacity.

3.2.2 Resource bundle concept

A resource bundle will be defined as a unitary package of the essential resources necessary to implement a certain amount of capacity in the outreach sites. The amount of resource bundles sent will impact the size of the outreach in terms of its immunization potential (in number of vaccine doses), and ultimately its costs.

After the interviews with UNICEF experts, five key resources necessary to guarantee the outreach implementation have been identified: (1) vaccines doses; (2) cold boxes and ice packs; (3) the vehicle used in the transport; (4) nurses responsible to administer the drug; and (5) other basic equipment, such as needles. Figure 2 illustrates the resource bundle concept.

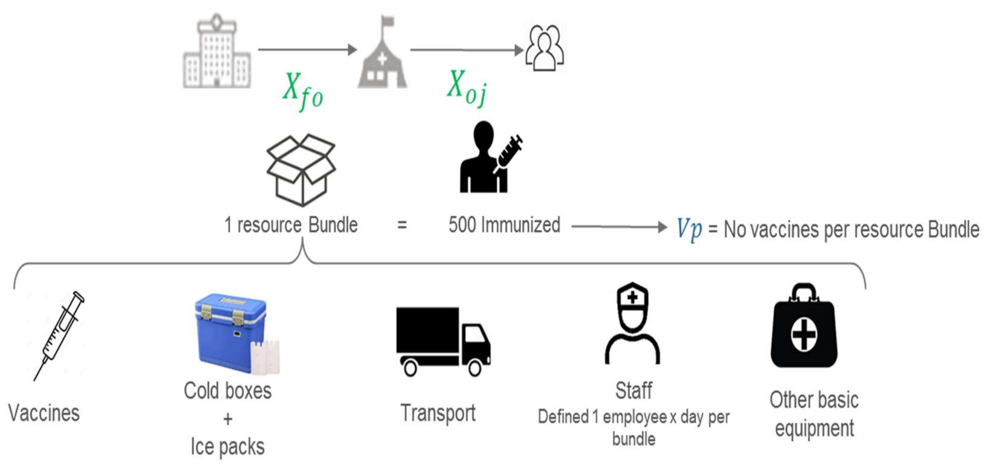


Figure 2: Resource bundle illustration

The model will consider that just an integer amount of resource bundles can be allocated to each operation. The main assumption behind the resource bundle concept is that it is possible to derive correlation parameters between the quantities of each of the resources inside each bundle. For example: one vehicle may fit five people and 12 cold boxes.

For each use case of our model, the network modeler should define which of the five key resources will have a 1:1 correlation to the bundle—in other words, which of the five resources available will have one unit included in each bundle. The definition of what will be the unitary resource bundle will depend on what is the factor that is most relevant to the country-specific circumstances and its operational bottleneck. We present two examples to illustrate this concept:

- In a specific network, the number of nurses could be fixed, and there is unlimited access to transportation. In this example, the best way would be to define one resource bundle as one nurse, and then derive how much of the other factors each nurse demands to fulfill the immunization capacity. In this case, the model would be allocating nurses.
- In a different network the limiting factor might be the number of available vehicles. In this scenario, the unitary resource bundle should be defined by the vehicle.

For our case study application, presented in Section 4, the vehicles were defined as the unitary resource inside the bundle. It is important to highlight that we kept the formulation flexible in case a different unitary resource bundle is required to meet different countries' needs. More details on the resource bundle definition for our case study can be found in Section 4.2.10.

Once the unitary resource has been defined, it is necessary to find the conversion factors that allow the derivation of the relative amount of key resources per bundle. The conversion factors will depend on the specific characteristics of each country's immunization network logistics, such as vehicle types used and cold box sizes. Table 4 shows a summary of the conversion factors necessary to derive the amount of resources in each bundle.

Table 4: Resource bundle parameters

Input variable	Unit
V_{cv}	Number of cold boxes per vehicle
Ve	Maximum number of employees per vehicle
V_{dc}	Number of doses per cold box
PE_o	Employee productivity in outreach operation [Doses/employee.day]

The key takeaway from this section is the importance for the model to allow flexible sizes of outreach operations. This was accomplished through the modelling of their outreach site capacity through resource bundles. These bundles can be seen as packages of the key resources needed in each operation (nurses, vehicles, and vaccines). The decision on what aspects to consider inside the resource bundle may vary depending on each country's specific characteristics.

3.2.3 Final model formulation

In this section, we presented the entire mathematical formulation of the problem. We defined sets, variables parameters, objective function and constraints.

3.2.3.1 Sets

Our network problem consists of three types of nodes, the fixed health centers, outreach sites and discrete populations. Their naming follows the same convention used in the toy model and is reproduced as follows:

- *F* = set of fixed health centers *f*
- *O* = set of potential outreach sites *o*
- *J* = set of population in regions *j*

3.2.3.2 Variables

The variables are the mathematical representation of the decision being made by the model. There are three key decisions in the final model: (1) which facilities will be responsible for each population; (2) how many doses will be supplied to each population region; and (3) how many resource bundles will be sent from each health center to each outreach site. The decision of sending bundle quantity to an outreach site indirectly defines whether the candidate location is used or not. Figure 3 illustrated how each of these variables is positioned in the network flow.

1. Variables (already in toy model):
   - X_{fj} = number of doses administrated to population in region *j* by fixed health centers *f* [Continuous]
   - X_{oj} = number of doses administrated to population in region *j* by outreach site *o*, [Continuous]
   - Y_{fj} = control if there are any people from region *j* vaccinated by fixed health center *f*, [binary]
   - Y_{oj} = control if there are any people from region *j* vaccinated by outreach site *o*, [binary]
2. New Variables:

- X_{fo} = Total number of resource bundles sent from fixed health center *f* to outreach location *o* [integer]

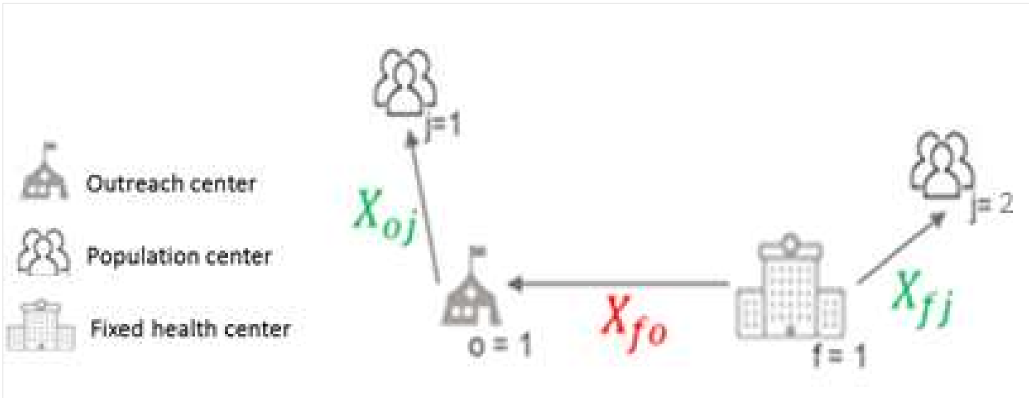


Figure 3: Illustrative representation of the model variables

3.2.3.3 Parameters

The parameters used for constraints and objective function formulation are presented below and are grouped in four categories. The model incorporates a demand that varies depending on the distance that each patient of region *j* is to its service sites. From this perspective, the total demand of a region j, D_j, initially a property of the node, can be converted to a arch demand , d_{fj} or $d_{oj,}$ though its multiplication to a percentage vaccination coverage factor, α_{fj} or α_{oj}, that decreases as the distance between region *j* to the outreach site *o* or health center *f* increases. A more detailed explanation of the demand modelling is presented in Section 3.2.3.7.

1. Fixed health centers
    - PE_f = Daily employee productivity at health center *f* [doses/day]
    - CE_f = Employee cost per day at fixed health center *f* [$/day]
2. Demand
    - α_{fj} = vaccination coverage factor between fixed health center *f* to region *j*
    - α_{oj} = vaccination coverage factor between outreach site *o* to region *j*
    - D_j = total demand for immunization, number of doses, in region *j*
    - d_{fj} = demand for immunization, in number of doses, which fixed health center *f* can capture in region *j*

- d_{oj} = demand for immunization, in number of doses, that outreach facility *o* can capture in region *j*

3. Resource bundle (outreach)

    - PE_o = Daily employee productivity at outreach *o* [doses/day]
    - V_{dc} = Average number of doses per cold box
    - Ve = Maximum number of employees per vehicle
    - V_{cv} = Maximum number of cold boxes per vehicle
    - c_{fo} = Total resource bundle cost
    - V_p= Number of doses per resource bundle
    - CE_o = Employee cost per day at outreach site *o* [$/day]
    - C_{ov} = Vehicle operation cost [$/km]
    - $\mathrm{Dist_{fo}}$ = Road distance between *f* and *o* [km]

4. General

    - c_v= Total cost to have vaccines available at health centers [$/dose]
    - B = total available budget [$]

3.2.3.4 Objective function and cost equation

Our objective function is to maximize the total amount of doses distributed which will, consequently, increase our vaccination access levels. The total amount of doses is obtained through the sum of the doses distributed from fixed health centers X_{fj} plus the ones from outreach sites X_{oj} (Equation 1).

$$max \sum_{j \in J} \left(\sum_{\mathrm{f} \in F} X_{fj} + \sum_{o \in O} X_{oj} \right) \qquad (1)$$

The most complex constraint, Equation 2, in our modelling is the budget. Due to the complexity of the cost equation, In order to simplify its explanation the total cost was broken down into three factors: (A) the vehicle operation costs; (B) the vaccination cost at fixed health centers; and (C) the vaccination cost at outreach sites.

$$\overbrace{\sum_{f\in F}\sum_{o\in O} X_{fo} * c_{fo}}^{\textbf{A}} + \overbrace{\sum_{f\in F}\sum_{o\in O} c_v * (X_{fj} + X_{oj})}^{\textbf{B}} + \overbrace{\sum_{j\in J}\left(\sum_{f\in F} X_{fj} * \left(\frac{CE_f}{PE_o}\right) + \sum_{o\in O} X_{oj}\left(\frac{CE_o}{PE_o}\right)\right)}^{\textbf{C}} \leq B \quad (2)$$

A. Resource bundle costs

The first cost component of the budget constraint refers to the cost of sending the resource bundles. As previously defined, each resource bundle in our scenario will be equal to one vehicle. The total cost of the resource bundles will be equal to the number of resource bundles sent, X_{fo} , multiplied by the unitary resource bundle cost.

For our case study application, the key resources inside a bundle unit are: (1) vehicle; (2) direct labor used on outreach; and (3) the vaccines doses.

- The vehicle operation costs are given by the multiplication of the vehicle operation cost per km, C_{ov}, by the distance between each fixed health center *f* and the outreach site o, $\mathrm{Dist}_{\mathrm{fo}}$.
- For the staff costs in the outreach, is assumed that each vehicle must always have a minimum of one employee. This minimum amount will incur a cost that is accounted by the single term, CE_o, inside cost Equation 2.A. Since each vehicle might transport more employees than only the assumed minimum, the cost of the additional employees, proportional to the vaccination needs and their productivity, are included in the third component of the cost equation (2.C). If the minimum quantity was not set, then in scenarios of very low demand, where staff requirements are less than one, the staff costs would be artificially smaller than the reality. The tradeoff of setting this minimum quantity is the fact that, at some circumstances, the planning of staff requirements will be considering that there will be one employee that is not directly involved in the vaccination need, which can actually be a pretty reasonable assumption. If, for example, the staff requirements for vaccination $\left(\frac{X_{oj}}{PE_o}\right)$ is of 1.3 employees, the total cost will be of 2.3 employees. This a

reasonable conservative simplification, since the employee productivity is often a difficult number to precise.

- The costs of the vaccine doses used in outreach sites are accounted for in the second component of the cost equation (2.B).

$$\sum_{f \in F} \sum_{o \in O} c_{fo} * \; X_{fo} \qquad (2.A)$$

Where: $c_{fo} = C_{ov} \text{ x } Dist_{fo} + \; CE_o$

B. Cost of vaccines doses

The cost of having the vaccination network divided by the number of doses should give the cost per dose C_v. This should represent the total purchase cost + the total logistic costs (transportation, storage, and spoilage) to have the vaccine available at the health centers, excluding the last-mile distribution (outreach trips) and the direct labor used to administer each vaccine.

$$\sum_{f \in F} \sum_{o \in O} C_v \; (X_{fj} + \; X_{oj}) \qquad (2.B)$$

C. Cost of direct labor for vaccine administration

The direct labor costs for the vaccination can be calculated by multiplying the daily costs of an employee by the total amount of employees required to administer the doses. The number of employees required is obtained by dividing the number of doses X_{fj} by the employee productivity Pe.

$$\sum_{j \in J} \left(\sum_{f \in F} X_{fj} * \left(\frac{CE_f}{PE_f} \right) \; + \sum_{o \in O} X_{oj} \left(\frac{CE_o}{PE_o} \right) \right) \leq B \qquad (2.C)$$

3.2.3.5 Constraints

Looking at a high level, the final model formulation has only three other constraints in addition to the budget, presented in the previous section: (1) the amount of doses distributed must be lower than the demand possible to capture by each facility, given a certain demand function shape; (2) the vaccination capacity of an outreach site is limited by the resource bundles; and (3) single sourcing constraint, which means each population is served by only one facility.

1. Demand constraint

Demand constraints control that the number of doses distributed to region *j* from a fixed health center *f* or outreach site o must be lower than the amount of demand that these facilities can capture in each region. Details on how to convert the total demand for vaccines of a region j, D_j, in an arch demands d_{fj}, and d_{oj} can be found in Section 3.2.3.7. The equations for demand constraints, (3) and (4), are simply the multiplication of the arch demand by the binary variable Y that determines whether that population center is fulfilled by the respective facility or not.

Endogenous demand functions

$$X_{fj} \leq d_{fj} * Y_{fj} \qquad \forall f \in \text{F}, \forall \text{j} \in \text{J} \qquad (3)$$

$$X_{oj} \leq d_{oj} * Y_{oj} \qquad \forall o \in \text{O}, \forall j \in \text{J} \qquad (4)$$

2. Resource bundle constraint – Outreach size operations

The following constraint guarantees that the amount of vaccines administered at each outreach site does not exceed the capacity of vaccines of each resource bundle.

$$\sum_f X_{fo} * Vp \geq \sum_j X_{oj} \qquad \forall \text{o} \in \text{O} \qquad (5)$$

This is a generic constraint under which the key factor of the equation V_p, the number of doses per bundle unit, will depend on how the resource bundle is defined. For the case application of this project, it was defined that each resource bundle would be represented as a vehicle; more information is presented in Section 4.2.10.

Given the resource bundle definition, it is possible to break down Equation 5 into two sub-equations, 5.1 and 5.2. Assume that each vehicle can carry a maximum Ve number of employees, and that each employee will be considered to have an average productivity at the outreach of PE_o doses per day. Based on that, and assuming each outreach trip must start and end in the same day, it is possible to say that with the available employees in each vehicle the number of doses per resource bundle cannot exceed PE_o x Ve doses per day. Consequently, the aggregated human resources constraint can be given by Equation 5.1.

$$\sum_{f} X_{fo} * PE_o * Ve \geq \sum_{j} X_{oj} \quad \forall \text{ o} \in \text{O} \qquad (5.1)$$

A similar approach is taken to calculate the maximum vaccination capability at an outreach due to vehicle transportation capacity for the vaccine doses. In this case, each cold box has a maximum number of doses per unit V_{dc} and each vehicle has a maximum number of cold boxes it can fit V_{cv}. The vehicle capacity constraint in terms of vaccine doses is given by Equation 5.2.

$$\sum_{f} X_{fo} * V_{cv} * V_{dc} \geq \sum_{j} X_{oj} \quad \forall \text{ o} \in \text{O} \qquad (5.2)$$

3. Single source constraint

The last constraint, presented in Equation 6, guarantees that each population region *j* is serviced by one and only one facility. This constraint is important to guarantee that the total flow of vaccine dose for region *j* do not exceed the rejoin total demand, D_j.

$$\left(\sum_{\text{f} \in F} Y_{fj} + \sum_{o \in O} Y_{oj} \right) \leq 1 \qquad \forall\, j \in \text{J} \qquad (6)$$

3.2.3.6 Limiting the amount of resource – Potential model modification with additional constraints

The introduction of the resource bundle concept allows for great flexibility in modifying the model according to the scenario and the reality of each use-case situation. The addition of constraints to the model to control the amount of resources available can be easily done with the application of correlation factors between the number of doses applied and productivity factors.

Some examples of new potential constraints that could be incorporated are presented below, but they were not included in our model case study application.

1. Limit staff capacity in health centers

If the maximum number of workers in a specific health center *f*, Ae_f , is a known hard constraint, the following equation could be incorporated to the model.

$$\sum_{j \in J} \left(\frac{X_{fj}}{Pe}\right) \leq Ae_f \qquad \forall f \in \mathrm{F} \quad (7)$$

Where:

Pe = Employee productivity [doses per day]

Ae_f = Available employees at location *f*

2. Limit the number of vehicles from location *f*

If the maximum number of vehicles available in a specific health center *f*, Av_f, is a known hard constraint, the following equation could be incorporated to the model.

$$\sum_{\mathrm{o} \in \mathrm{O}} \left(\frac{X_{fo}}{V_b.}\right) \leq Av_f \qquad \forall f \in \mathrm{F} \ (8)$$

Where:

V_b = Number of vehicles per resource bundle (equals to 1 in our application)

Av_f = Available vehicles at location *f*

3.2.3.7 Endogenous demand functions

It is one of the objectives of our model to incorporate an endogenous demand function. Based on our literature review and wide agreement, we assume that the demand that a fixed health center *f* can capture from population center j, d_{fj}, would decrease as the distance between *f* and *j* increases. This captures the effect that people would be less willingly to participate in immunization services as the location for immunization becomes further away to them. Mathematically, this can be formulated by multiplying the demand D_j by factor α_{fj}, as shown in Equation 9. This factor, called a vaccination coverage, is a function of the distance between *f* and j, α_{fj}= f ($Dist_{fj}$). The same logic can be extended for outreach sites, as Equation 10 shows.

$$d_{fj} = D_j \cdot \alpha_{fj} \qquad \forall\, j \in \text{J}, \forall\, f \in \text{F} \qquad (9)$$

$$d_{oj} = D_j \cdot \alpha_{oj} \qquad \forall\, j \in \text{J}, \forall\, \text{o} \in \text{O} \qquad (10)$$

The key question is what the shape of the function α_{fj}= f ($Dist_{fj}$) is. As we showed in our literature review, there is no consensus on what the shapes and coefficients of such function should be, and the derivation of a demand function for a specific country requires a lot of detailed data. However, the literature review allowed us to arrive at reasonable estimated numbers, boundaries on which it would be reasonable to test our model. In Section 4.2.4, we detail the three shapes tested in our use case and later, in Section 5.1, the model sensitivity to the different shapes is evaluated.

Regardless of the shape of the function—linear, exponentiation, or polynomial decay—the applicability of the Equations (9) and (10) remains. For the sake of illustration, Equations (11) and (12) show how a linear decay would be, assuming a decrease of 20% every 100 km.

$$\alpha_{fj} = 100\% - 20\% \cdot \frac{Dist_{fj}}{100} \qquad \forall\, j \in \text{J}, \forall\, f \in \text{F} \qquad (11)$$

$$\alpha_{Oj} = 100\% - 20\% \cdot \frac{Dist_{fj}}{100} \qquad \forall\, j \in \text{J}, \forall\, \text{o} \in \text{O} \qquad (12)$$

In order to further explain how Equations 9to 12are applied in practice, Figure 4 illustrates an example of the potential demand captured by a fixed health center with regard to two distinct population centers. In this example, fixed health center f1 is at a distance of $Dist_{12}$ = 200 km to population in Region 2 and at a distance $Dist_{13} = 300\text{ km}$ to population in Region 3. By applying Equation 11, it is possible to derive that demand factors will be $\alpha_{12} = 60\%$ and $\alpha_{13} = 40\%$. Since the total demand in the two regions is the same ($D_2 = D_3 = 500$), it is possible to substitute this value and the demand factors in Equation 9 to arrive at the final demands for immunization in the arches $d_{12} = 300$ and $d_{13} = 200$.

With the incorporation of the demand function shown in this section, the demand becomes a property of the arch and not the node, as it is typically represented in network optimization problems. In our case, it means that the amount of demand for immunization that one region may request will depend on how close it is to the immunization center selected by the optimization to serve this region.

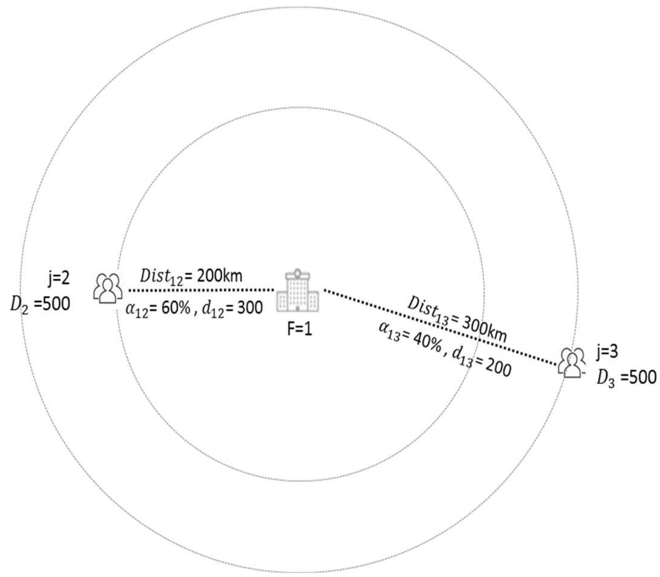


Figure 4: Example of distance effect on demand

4 Case study of The Gambia - Final model application

One of the objectives of this project is to validate the developed optimization model in a real case scenario. UNICEF provided us with immunization data from The Gambia, which was used to validate the effectiveness of our model to provide useful insights in a real-world context and the applicability of our solution procedure.

This chapter is organized as follows: First, in Section 4.1, we presented an overview of The Gambia's current vaccination network and define the scope of our analysis. After that, in Section 4.2, we described an extensive list of how all necessary inputs to the model were derived. To solve the model, a Python code in Jupyter Notebook was created; details of the solution procedure are shown in Section 4.3. Finally, in Section 4.4, the baseline scenario is created and optimization scenarios are defined.

In this chapter, simplifying assumptions were made when deriving the input data for our model. Some of these assumptions may seem simplistic. However, our main objective with this exercise is not to reproduce or provide a specific number for the cost of The Gambia Network. Rather, the purpose is to determine whether our suggested modeling approach and solution procedure are applicable to a real scale scenario. All costs and parameters are proportional to each other and overall were derived from reasonable assumptions. However, the exercise will be successful if it provides the capacity to derive insights and suggestions that allow the vaccination network structure to achieve greater vaccination access levels.

We derived as much data as possible from publicly available sources. We also thoroughly detail the data extraction procedure. This enables the model to be easily reproduced in any country, even if limited data are available.

4.1 The Gambia network description

The Gambia was the country selected to validate and test the performance of our model. This country's network follows the typical vaccination supply chain design of developing countries: a national store, regional distribution centers, health centers, and outreach sites (Figure 5). The different levels of the network are shown together with the number of facilities on each level. One important aspect to consider is that the fixed facilities (regional stores, national stores, and health centers) are not only used for immunization services but are also part of the logistics networks of other health services.

The flow of immunization commodities in The Gambia starts in the ports and moves from there to the national store. From the centralized distribution centers, the products are distributed to the seven regional stores, to finally flow down to health centers and outreach sites. The regional stores are supplied on a quarterly basis by trucks coming from the national stores. In general, each health center is responsible for performing the commodities pickups in the regional stores every month, using the vehicle it has available at its sites. This same vehicle is used for the outreach operations.

The outreach treks depart from the health centers, where medical staff leave in a vehicle at the beginning of the day with all the necessary equipment, go to the pre-specified outreach sites, perform the vaccination and return to the health center at the end of the same day. There is a pre-determined monthly schedule that defines the specific days for the immunization service at each outreach site and health center.

Our network design model focuses on the last-mile distribution of the above described vaccination network. This means that we will focus on defining the optimal schedule of distribution to the outreach sites , the location of the optimal outreach sites (i.e., which should be closed, and which should be opened), and the requested personnel capacity at the fixed health centers.

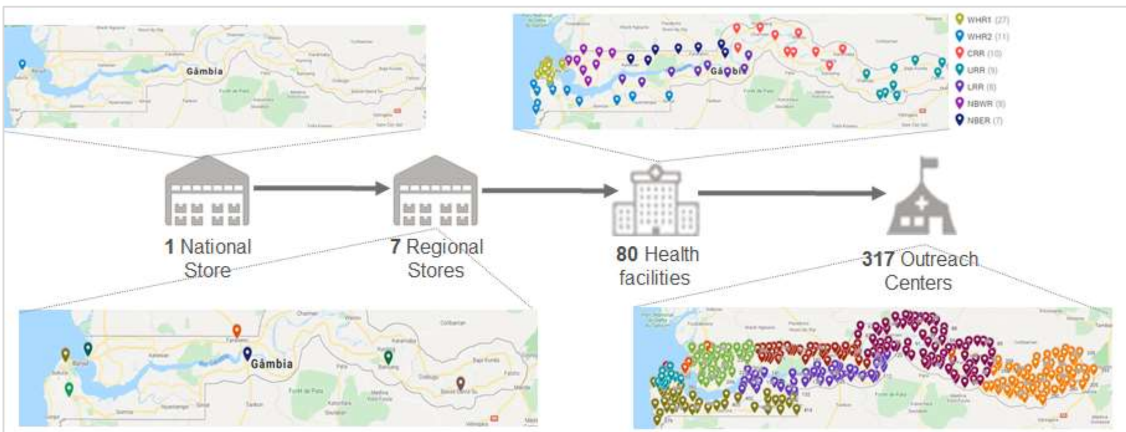


Figure 5: The Gambia vaccination supply chain schematic view

It is important to highlight that our model does not suggest the opening or closure of health centers, nor does it evaluate changes in the network at a regional or national stores level. This limitation has no particularly big impacts in this case, since the upstream fixed facility structure is also used for providing other health services. Thus, changing health centers footprint would require a different approach that would need to consider other operational constraints, services, and products flow that are part of these distinct health services.

4.2 Model inputs

One of the most challenging tasks in any network design project is to derive the necessary data required for the optimization model. This section discusses a detailed view of how the optimization model inputs were derived, including what data sources were used, how they were cleaned, and what assumptions were made.

In general terms, a network design problem has inputs that can be associated to three elements: nodes, demand, and costs. Accordingly, this section is organized based on those three elements:

- The **nodes** in our model consist of three components: the population regions, the outreach sites, and the health centers. Data are presented in Section 4.2.1, 4.2.2, 4.2.5, and 4.2.7.
- The relationship between vaccination **demand** and total population of a region is shown in Section 4.2.3. The demand function, that reflects the impact of distance in vaccination coverage, is presented in Section 4.2.4.
- Finally, the **costs** will depend on the distances presented in Section 4.2.6, together with vaccine cost in Section 4.2.8, the fixed health center costs in Section 4.2.9 and outreach resource bundles values in Section 4.2.10.

4.2.1 Population data from UNICEF

UNICEF provided us with population data for The Gambia. The data contains population information together with a list of all health centers, outreach sites and their locations. The total population of 2.26 million people was calculated by the summation of population in each site catchment area. We validated this population information by comparing it with the statistics from the World Bank (2.28 million), which revealed a 99.1% match.

Even though the reliability of the total population was high, the downside of the population data provided by UNICEF was that it allocated each site to a population in a specific region, according to the current allocation rules. Since there is low clarity on how the population allocation to each site was derived, and there was no way to trace back the original population location, we researched other population data sources to identify the actual location of each population segment. This information is fundamental for exploring candidate locations for the outreach sites. The representation of UNICEF population data is shown in Figure 6, from this map it is possible to see that most people reside in the capitol city of Banjul in the west end of the nation.

Figure 6: Distribution of total population by site catchment area
Data Source: Socioeconomic Data and Applications Center. Retrieved from https://sedac.ciesin.columbia.edu

4.2.2 Population data from the Socioeconomic Data and Applications Center

Population data play an essential role in detecting where the demand exists which helps to identify the best candidate locations for outreach sites. The concept of the model is people-centric, which means that we select the outreach sites' locations in the places people reside instead of setting the locations first to gather people. In this sense, knowing the distribution of population density is of great importance.

We extracted the population dataset from the Socioeconomic Data and Applications Center (SEDAC) which has 1-km downscaled population projections from 2010 to 2100 for every 10 years. The scope of this dataset is the entire globe. In addition, there are three kinds of data: total, urban, and local, and two types of data, NetCDF[2] and GeoTIFF[3], are available. We selected data for 2020 total with GeoTIFF data type (file name: ssp2_total_2020.tif). We then extracted the data for The Gambia from this dataset, as per the following process (Figure 7).

Step 1: Extract by Mask for The Gambia

---

[2] "NetCDF (Network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data". Source: https://www.unidata.ucar.edu/software/netcdf/

[3] "GeoTIFF is based on the TIFF format and is used as an interchange format for georeferenced raster imagery. GeoTIFF is in wide use in NASA Earth science data systems". Source: https://earthdata.nasa.gov/esdis/eso/standards-and-references/geotiff

The selected dataset contains population within one square kilometer all over the world. To extract only the data for The Gambia, we set the boundary of the country by using the GADM database (file name: gadm36_GMB_1). Then, we used the extract by mask function in ArcPy[4].

Step 2: Import ArcPy

```
# Replace a layer/table view name with a path to a dataset (which can be a layer file) or create the layer/table view within the script
# The following inputs are layers or table views: "gadm36_GMB_1"
arcpy.gp.ExtractByMask_sa("L:/UserFiles/yuto0726/Raw data/ssp2_total_2020.tif", "gadm36_GMB_1", "E:/UserFiles/yuto0726/OutputData/gambiaunp2020")
```

Step 3: Create point data

We obtained the output data of the previous process (file name: gambiaunp2020). The next step was to convert raster data[5] into point data for the smooth use of the model input, using the raster to point function in ArcPy.

```
# Replace a layer/table view name with a path to a dataset (which can be a layer file) or
create the layer/table view within the script
# The following inputs are layers or table views: "gambiaunp2020"
arcpy.RasterToPoint_conversion(in_raster="gambiaunp2020",
out_point_features="E:/UserFiles/yuto0726/OutputData/gambia2020_points_unp.shp",
raster_field="Value")
```

We obtained the point data of The Gambia (file name: gambia2020_points_unp.shp). Adding latitude and longitude information to this data by using the function in ArcGIS, we output this as a csv file type and used it as input for our final model. The total number of grid centroids resulted

[4] "ArcPy is a Python site package that provides a useful and productive way to perform geographic data analysis, data conversion, data management, and map automation with Python".
Source: https://pro.arcgis.com/en/pro-app/arcpy/get-started/what-is-arcpy-.htm

[5] "a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information".
Source: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm

from this procedure was 12,796. Because the large number of nodes required excessive computational capacity in the optimization runs, the grid was changed to a 2 x 2 km granularity, which resulted in a total of 3,952 grid centroids. The sum of all point population is 2.05 million. This total number is 10% lower than the information provided by UNICEF, but further inspections showed that the population distribution in the two datasets followed a similar distribution. This gave us confidence to use the SEDAC dataset as the basis for our model, keeping in mind that we may need to offset our costs by 10% at the end of the optimization.

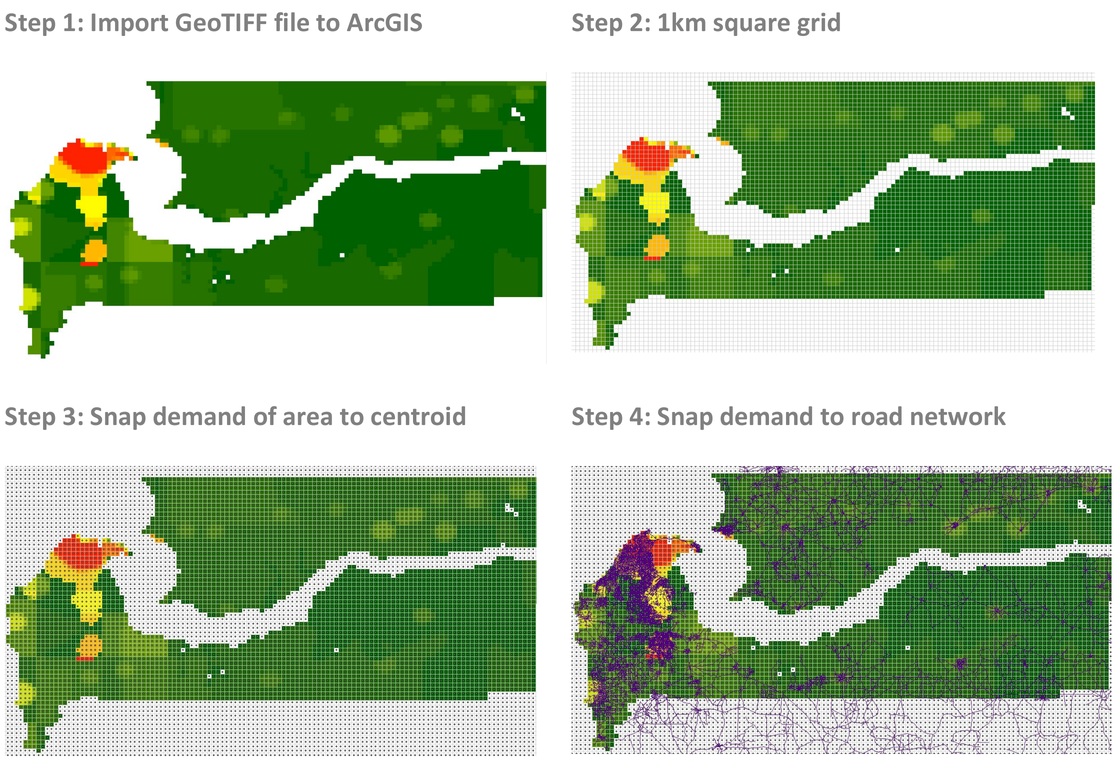


Figure 7: Illustration of step by step procedure to derive the population data from SEDAC

4.2.3 Vaccination doses demand

With an understanding of how the population is distributed in the country, one important question arises – how to translate the population data into the need for vaccine doses? Based on interviews with UNICEF experts, the standard procedure to derive the vaccination needs can be summarized in Figure 8.

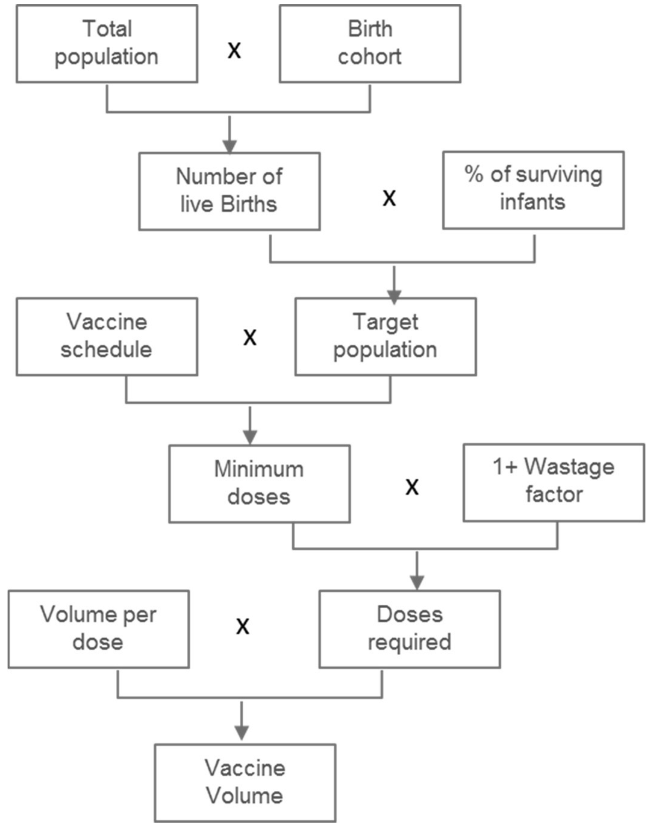


Figure 8: Calculation from population to vaccine doses.
Source: Interviews with UNICEF experts

UNICEF provided the inputs for the calculation, shown in Table 5. Based on this information, it is possible to derive a factor that correlates the total population of the country and the number of doses required to provide the target population with 100% coverage. The multiplication of the "target population" column, by (1) the total population of The Gambia; and (2) the "schedule" column (number of doses per year), resulted in the number of doses required per vaccine per year. In the case of BCG, for example, doses were given by the multiplication of 2.28 m (total population) by 4% (target population) and 1 (schedule) leading to a total requirement of 91 thousand doses per year. By repeating this procedure that for every vaccine on Table 5, and summing the results, it is possible to arrive in the total number of doses required across all vaccines, 2.29 million doses per year. If this number is divided by the total population of country, 2.28 million, a factor of 1.002 doses per person per year is calculated. Indeed, this does not mean that every person receives one dose - newborns will probably receive the full schedule while adults may not receive any doses. However, since the population data do not distinguish the age range of the population, this approximation to convert the total population to dose demand is important for our modeling. A similar approach allows us to derive the average dose volume for The Gambia of 5.63 cm3/dose.

Table 5: The Gambia vaccination schedule, target, wastage

Vaccine	Schedule [doses /year]	Target Population	Vaccine Wastage	Volume vaccines [cm^3]	Volume Diluents [cm^3]
BCG	1	4%	50%	2	1
bOPV	5	4%	10%	5	0
DTP-HepB-Hib	3	4%	10%	9	0
MR	2	4%	15%	6	9
PCV-13	3	4%	5%	9	0
HepB	1	4%	10%	5	0
TT	4	4%	15%	14	0
IPV	1	4%	10%	16	0
Rota_liq	2	4%	5%	36	0
YF	1	4%	15%	4	4
DTP	1	4%	10%	3	0
Men_A	1	4%	10%	3	4
HPV	2	2%	5%	28	0

Source: UNICEF - Cold chain inventory 2018 XLS

4.2.4 Demand coverage function

The result of our literature review on how the distance to the facility impacts the access levels (Section 2.2), did not specify a conclusive shape for the function. We observed, however, that typically the chosen shape is a linear decay, and maximum distances ranging between 5 to 10 km appear reasonable.

Based on these findings, we drew three different demand functions with different parameters and ran the model for all of them. These functions are named: moderate access (Figure 9), high access (Figure 10), and low access (Figure 11). All these demand functions have as common factor the assumption of linear decline in demand as the distance between population and vaccine facilities increases. The main difference lies in the slope of decrease in demand.

1.Moderate access demand function (1, 5)

The first demand function, shown in Figure 9,considered moderate access, is the linear decrease from 1 km to 5 km. This means that no one is willing to walk more than 5 km for immunization, and 100% of the population would accept walking up to 1 km. People's willingness to walk decreases linearly between 1 km and 5 km. Also, we assumed that most people in rural areas in The Gambia do not have access to vehicles, and therefore they would walk. The moderate access function relates to the results obtained by Feikin et al. (2009), as presented in Section 2.2. The study showed how no children were reached for distances greater than 6 km and very low access levels (6% - 9%) for distances from 4 to 6 km.

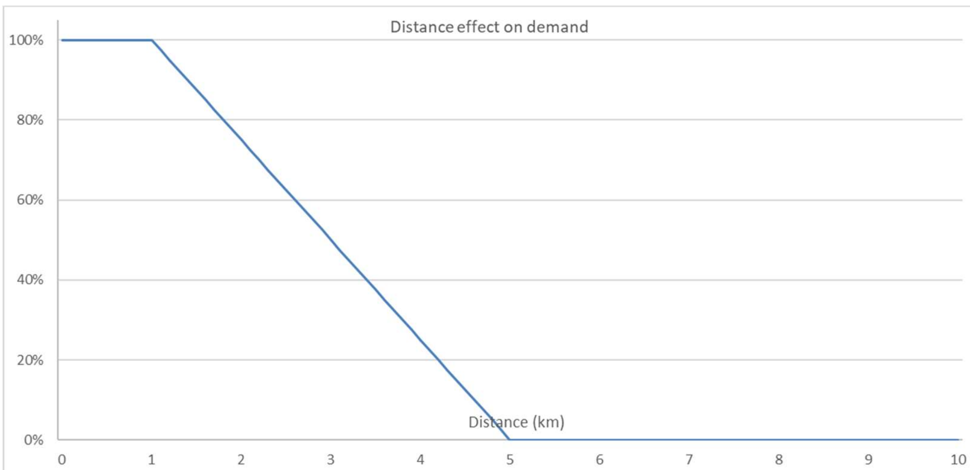


Figure 9: Moderate access demand function (1, 5)

2. High access demand function (2, 10)

The second demand function, shown in Figure 10, is the linear decay from 2 km to 10 km. This is the optimistic demand function. It assumes that most people are not willing to walk more than 10 km for immunization, but would be willing to walk up to 2 km. People's willingness to walk decreases linearly between 2 km and 10 km.

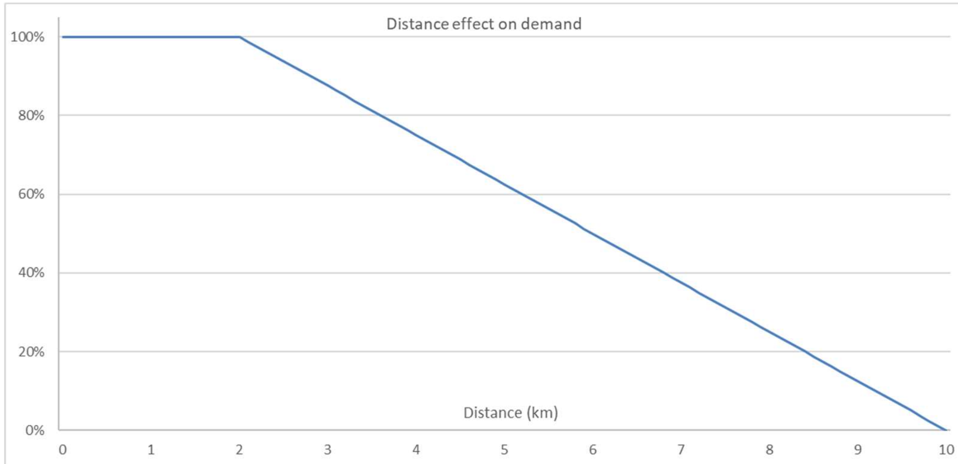


Figure 10: High access demand function (1, 10)

3. Low access demand function (0,2.5)

The third demand function, shown in Figure 11, is the linear decrease from 0 km to 2.5 km. This is the pessimistic demand function. It assumes that most people are not willing to walk more than 2.5 km for immunization. People's willingness to walk decreases linearly and immediately from 0 km 2.5 km. This is based on the research paper that proved the people's willingness to use local health facilities significantly dropped after 30 minutes of walking (approximately 2.5 km).

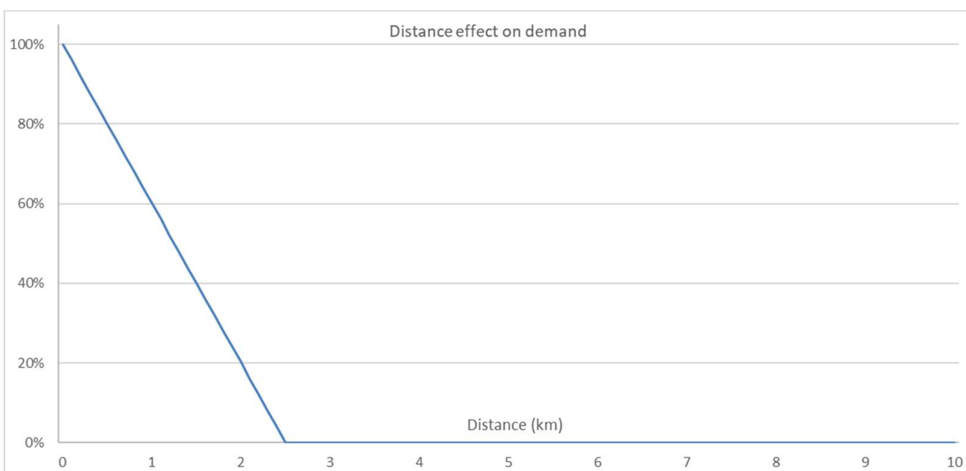


Figure 11: Low access demand function (0, 2.5)

4.2.5 Current facilities -Baseline

In order to build the baseline a key piece of information was the current GPS position of the current health centers and outreach sites. This data was provided in an Excel spread sheet named "master_data xls. The main information provided in this data file is shown in Table 6.

Table 6: Key columns name and description of Master data xls.

Column name	Description
Site_ID	Numeric single identifier
Site_Name	String text (mostly unique)
Region	Region that facility belongs to (7 in total)
Site_Category	Health facility, outreach site, regional store, national store
Longitude	Geolocalization
Latitude	Geolocalization
Total_Pop_for_Site_Catchment_Area	UNICEF calculation that estimates total population covered by each facility

In order to evaluate the quality of the GPS information provided, all points for the outreach sites were plotted using "Google My Maps". Total number of health centers is 80, and total number of outreach sites is 331. The results, illustrated in Figure 12 and Figure 13, show a good distribution of the points and that all nodes are located inside The Gambia borders. These two characteristics lead us to the conclusion that the data has a sufficiently good quality.

Figure 12: Plot of health care facilities in The Gambia network (colored by region)

Figure 13: Plot of outreach sites in The Gambia network (colored by region)

4.2.6 Distances

The calculation of distances is a key component of the model formulation, since it impacts both the demand coverage function and the outreach distribution costs. For the estimation of the distances between health centers and outreach sites the shortest road path distance was used, while in calculating the distance for the population to get to the outreach site a linear distance was calculated. Figure 14 shows in a schematic way the different distance calculation approaches used in the model.

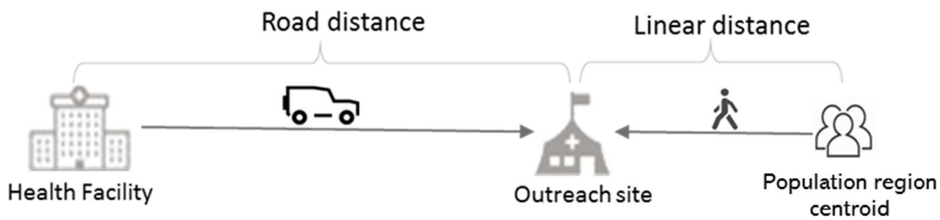


Figure 14: Distance calculation method

It is a common approach when doing network design problems to work with linear Euclidian distances instead of road distance. The main reason is that calculating road network distances can be quite challenging when using only open-source data and for large networks.

However, for some specific geographies, the distances between the road network and the linear distance differ dramatically, and unfortunately this was the case for The Gambia. Figure 15, generated with ArcGIS and Open Street Map (OSM), demonstrates the large part of the country without road connections due to the river. As a long and narrow country, crossed by a large river, the linear distance between two points in opposite margins of the river will differ significantly when compared to the road distance. For this reason, we calculated the road distances between health facilities and outreach sites candidates.

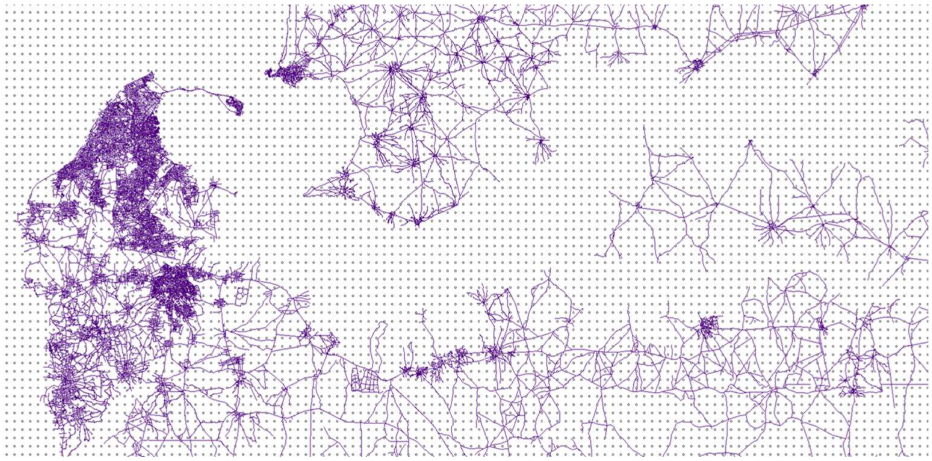


Figure 15: Extract of The Gambia road network

In order to estimate the distances between the defined nodes, a Python code was developed. The full python code developed can be found in the following GitHub repository[6]. Open Street Map[7] was the source of the road graph. This open-source library provides the road network of the entire world in a graph format; that is, as a collection of nodes connected by arches with their respective distances. The road-network distance calculation procedures we applied are shown in Figure 16.

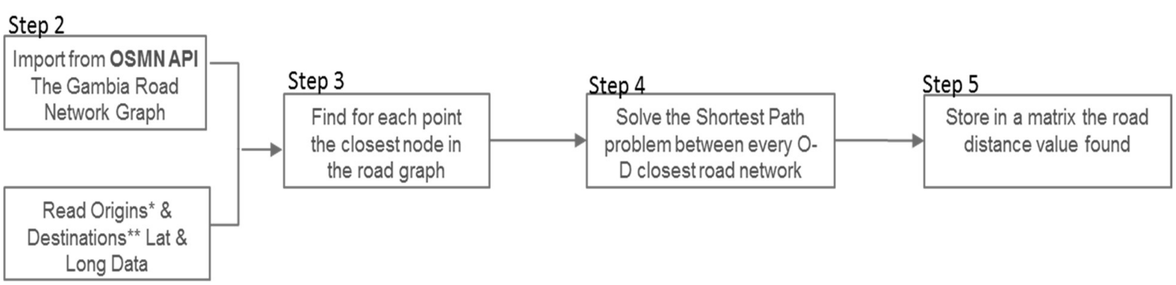


Figure 16: Schematic view for the calculation of road distances using OSM

The impact on the model of the distance between the population region centroid and the outreach site is defined by the demand coverage function. Our modelling uses linear distances between the population centers to the facilities for two main reasons: (1) many of population

[6] GitHub repositor URL: https://github.com/optimization-network/vaccination-networks

[7] "Open Street Map is a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license. The OpenStreetMap License allows free (or almost free) access to our map images and all of our underlying map data. The project aims to promote new and interesting uses of this data."
Source : https://wiki.openstreetmap.org/wiki/About_OpenStreetMap

centers in the designed grid lay in points that have no access to road network, therefore the calculation of road distance is not possible; and (2) we assume that most of the population in the rural areas will walk directly to the outreach site rather than take the road.

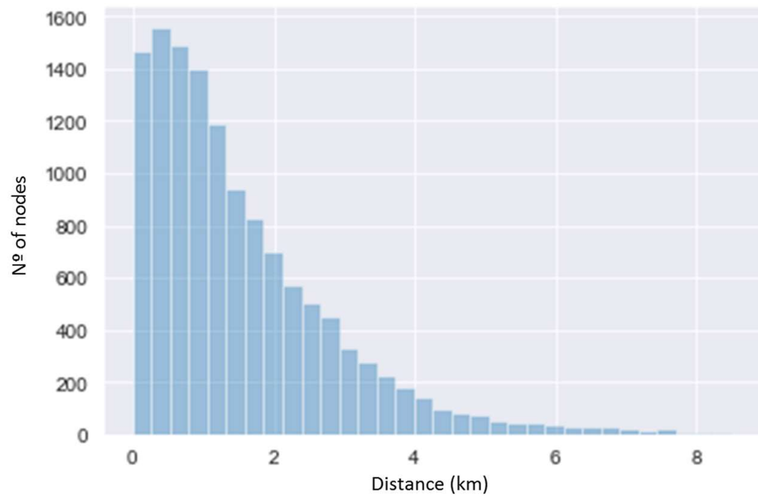
In order to calculate the walking distance a Haversine Python library was used. This library has a function that calculates the distances between two GPS coordinates, not only by determining the Euclidian distance, but also by considering the effects of the Earth's curvature. It is true that the curvature effect might be negligible for the small distances considered in this application, but implementing it prepares our code for future potential applications in other regions where distances are greater.

4.2.7 Outreach site candidates using road network

One of the objectives of the optimization model is to explore new potential locations for the outreach operation. The starting point for detecting candidate locations is to set the centroid of the 1 km x 1 km grid to discretize the population. The Gambia map was initially divided into 12,796 grids, with the population discretized in the centroid of each of these grids. As previously explained, this number was reduced after a modification on the granularity level to a 2 x 2 km grid, resulting in 3,952 grid centroids.

Initially one could make the natural choice of considering all 3.952 grid centroids as potential candidates for the outreach operation. The main problem with this approach is that this number of nodes would lead to millions of decision variables in the problem, which would be unfeasible to solve due to limits on computational capacity.

In order to reduce the amount of candidate sites, we asssumed that many of these points would fall in remote areas, such as in the middle of a forest, farmland with low population density, or places far away from roads. Potentially, all these points could be excluded from the grid, since it is unreasonable to locate an outreach site in an area far from road access. In order to validate this approach, the distances between all centroids of the grid to the closest node in the road network were calculated. The result can be seen in Figure 17.



Descriptive statistics

Total nº of nodes	12796
mean	1.50 km
std	1.30 km
min	0.00 km
25%	0.56 km
50%	1.16 km
75%	2.13 km
max	8.49 km

Figure 17: Histogram showing distance between the centroids of the grid and closest node in OSM road network

Two main conclusions can be derived from this histogram: (1) the quality of the road network is quite good, since 75% of the points on the grid have a distance to the closest node in the network of less than 2 km; and (2) cutting out the points located at a large distance (i.e., 4 km or more) will have a minor impact on the number of candidate outreach sites.

Expanding on the idea of discarding points located far from the road network raised an interesting point: whether we should only consider nodes that are located directly on the road network. This is a particularly good idea, not only because the road distance is calculated based on the nodes of the road network, but also because it is more likely for the outreach sites to be in the road, since they need to be accessible by vehicle. However, the entire road network of The Gambia has 39,000 nodes, so if we considered all single nodes, we would increase the number of sites instead of reducing them. This impasse was solved through a mixed approach for selecting the outreach sites candidates:

- Step 1: Generate the 2 x 2km grid of with 3,952 centroids.
- Step 2: Find the closest node in the road network for each of the 3,952 nodes.
- Step 3: Get all the unique nodes in the road network found in Step 2 and use them as candidate sites.

Through the application of the above steps it was possible to reduce the number of candidate sites , from 3,952 to 2,331, as plotted on the map in Figure 18. Figure 19 shows how this reduction could be obtained. In this illustrative example, each of the three colored nodes represents a node of the road network, and the black points indicates the 1 x 1 grid centroid. We can observe how dramatically this procedure reduces the number of candidate locations in the remote areas.

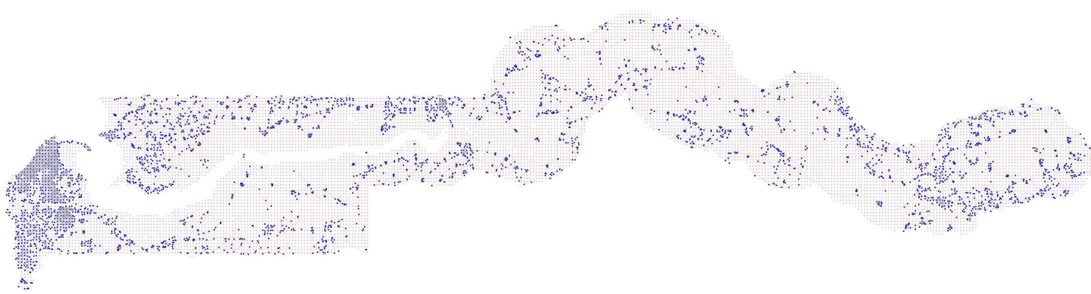


Figure 18: Outreach candidate sites using OSM road network approach

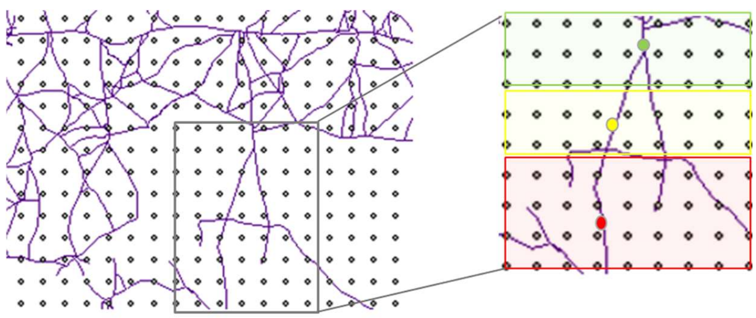


Figure 19: Candidate sites selection based on OSM graph closest nodes to grid

4.2.8 Vaccine cost

Based on discussion with UNICEF, we determined that a good approximation for the total cost for having each vaccine dose available at the fixed health center would be C_v = \$0.02 per dose. The precision of this number is not so relevant for our model, since our focus is to optimize the last-mile operations of the network. From this perspective, it is important to clarify that this cost does not include the costs to provide the administration of the vaccines, either in outreach sites or in fixed health centers. In the next two sections is presented the cost components to be considered in predicting the last-mile and customer service costs at both fixed facilities and outreach sites.

4.2.9 Fixed health center vaccine administration direct labor cost C_{ef} and productivity P_{ef}

This analysis considers that the only relevant cost to administer the dose at the health center is the direct labor cost. In order to estimate this cost C_{ef}, we need to derive two parameters: (1) the hourly wage of an average nurse in The Gambia; and (2) the average number of doses a nurse can administer per hour PE.

According to Table 7, the average hourly wage of a nurse in The Gambia, can range from 0.66 $/hour, for a senior nursing officer, to 0.24 $/hour for a community health nurse in a simplistic analysis, if the average hourly wage is $0.49. As we expect to have more employees at the base positions than at the top, it seems reasonable to estimate the salary of the community health nurse midwives, 0.46 $/hour, as the standard for the direct labor involved in the immunization. Assuming eight hours of workload per day the total cost C_{ef} = 3.68 $/day.

Table 7: Nurse salaries in The Gambia, inflation adjusted

Nurse salaries The Gambia	Annual salary 2015 [dalasis]	Accumulated inflation 2015 - 2020	Annual salary 2020 [dalasis] *	Currency exchange [$/dalasi][8]	Annual salary	$/hour
Senior Nursing Officers	48,000	41%	67,680	49	1,381.22	0.66
Registered Nurse Midwives	42,000	41%	59,220	49	1,208.57	0.58
State Registered Nurses	36,000	41%	50,760	49	1,035.92	0.50
Community Health Nurse Midwives	33,000	41%	46,530	49	949.59	0.46
Community Health Nurses	17,000	41%	23,970	49	489.18	0.24

Source: Federal Reserve Bank of St. Louis. Retrieved on April 17, 2020 from https://fred.stlouisfed.org/series/FPCPITOTLZGGMB

[8] Based on Online Currency Converter (https://freecurrencyrates.com/en/exchange-rate-history/USD-GMD/2015/yahoo)

In order to estimate the average time for vaccine dose administration, a literature review was conducted. Table 8, created by Mokiou, Standaert, Li, and De Cock (2017), outlines the typical activities involved in the vaccine administration process.

It is necessary to have in mind the difference between clinic in the UK, where the study was conducted, and a health clinic in The Gambia. From the activities listed, we estimate that only the vaccine preparation, reconstitution and administration would not change. Based on the values for these activities we estimate that a nurse could potentially vaccinate 20 doses per hour, an average of three minutes per dose.

Assuming that each working day at the fixed health center has eight hours, we can conclude that the daily productivity of a nurse in the fixed health center, P_{ef}, is equal to 20 x 8 = 160 doses per day.

Table 8: Active HCP time by activity per single vaccine administration process

Time spent by HCP per activity per visit (minutes)	Pooled	UK-01	UK-02	UK-03	UK-04	UK-05	UK-06
Guardian/subject arrival and registration at GP surgery	0.3	0.8	0.0	0.4	0.3	0.0	0.5
Invite and welcome guardian/subject in the vaccination (consultation) room	0.4	0.6	1.1	0.0	0.5	0.0	0.4
Nurse consultation visit	3.1	2.9	3.2	2.2	3.2	3.0	3.9
Vaccine preparation	0.5	0.5	0.7	0.5	0.4	0.5	0.4
Vaccine reconstitution	1.0	1.2	1.5	0.7	0.8	0.9	1.1
Vaccine (s) administration and consumables disposal	1.4	0.9	2.2	1.0	1.3	1.5	1.3
Record keeping and post-administration monitoring	2.8	1.9	2.6	2.2	2.0	4.4	3.7
Total active HCP time	9.5	8.8	11.3	7.0	8.5	10.3	11.3
Lower 95% CI	7.7	8.2	9.9	6.1	7.4	9.5	9.9
Higher 95% CI	11.3	9.4	12.8	7.9	9.6	11.0	12.6
Total time spent by HCP type (min)							
RN & AN	8.6	8.0	11.3	6.6	8.2	5.7	10.7
RE/AS	0.9	0.8	0.0	0.4	0.3	4.5	0.6

AN: Auxiliary Nurse; AS: Administrative Staff; CI: Confidence Intervals; GP: General Practitioner; HCP: Health Care Professional; RE: Receptionist; RN: Registered Nurse; UK: United Kingdom.

Source: Mokiou, S., Standaert, B., Li, X., De Cock, E. (2017). Measuring the cost of a pediatric vaccine administration in the UKS.

4.2.10 Resource bundle definition

Based on the interviews, we drew a conclusion that one key asset for providing the vaccination services in The Gambia is the vehicles. Each health center has only one vehicle available for performing both the pick-ups from storage and the outreach distributions. Because each outreach round trip takes one day, the vehicles are a scarce resource that must be closely monitored. For this reason, we define the vehicle as the unitary resource in the bundle.

- **Vehicle capacity and operating costs**

  The vehicles used in The Gambia network are Ford Everests, Nissan Patrols, or Toyota Land Cruisers (Figure 20). Internal studies performed by UNICEF concluded that a total of 12 cold boxes could reasonably fit in the back of one of those vehicles, in three rows of two boxes side-by-side, stacked two high. However, this transport capacity assumes the entire back of the truck is transporting cold boxes. By considering the average vehicle size, box sizes, and the photos, we conclude that is reasonable to assume that each vehicle is capable of taking a maximum of V_e =5 person and V_{cv} =3 cold boxes at the same time. The same study estimated that the average vehicle operating costs for the vehicles mentioned above averaged $C_{ov} = 0.6\ \$/km$.

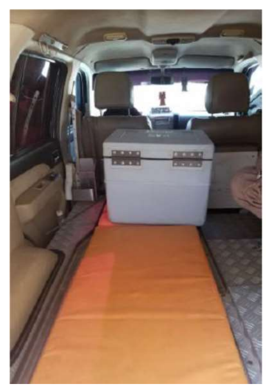


Figure 20: Vehicles used in the outreach treks
Source: Manual, C., Mckinnon J. (2019). Supply Chain Design: The Gambia. Assessment of the Vaccine Supply Chain Design in The Gambia. UNICEF Internal report.

- **Employee productivity**

  The productivity of the nurses in the outreach sites will be affected by the traveling time and the time needed to set up the vaccination structure at the outreach site. Based on the interviews with UNICEF, we concluded that is reasonable to assume that each of the employees will have available, out of the eight working hours per day, only six hours to perform the vaccination services. Assuming the same 20 doses per hour productivity, the average daily productivity of an employee at an outreach trip can be estimated as PE_o = 120 (20 x 6) doses per day.

- **Cold box capacity**

The cold box most commonly used in The Gambia is the Blowkings 7-liter unit measuring 49 x 44 x 49 cm (Figure 21).

Figure 21: Blowkings 7-liter vaccine cold box
Source: Manual, C., Mckinnon J. (2019). Supply Chain Design: The Gambia. Assessment of the Vaccine Supply Chain Design in The Gambia. UNICEF Internal report.

To derive the carrying capacity in number of doses of a cold box, we determined the "packed volume" of each vaccine dose vial. We defined packed volume as the average volume occupied by a dose plus its package and surrounding empty spaces in the box. The packed volume of each dose was calculated by dividing the total reported throughput of vaccines in The Gambia by the total amount of doses administered. Based on the calculation procedure shown in Figure 22, we concluded that for each dose, the unitary vaccine volume is 26.20 cm^3/dose.

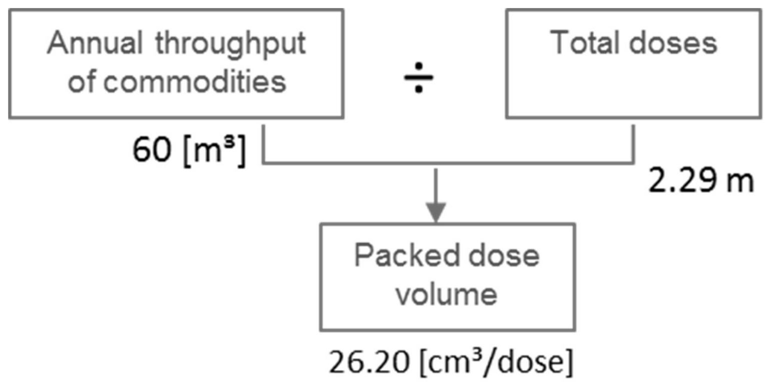


Figure 22: Calculation procedure to derive average volume of a dose including package and empty spaces

Next, the number of doses per cold box was determined by dividing the cold box volume by the unitary dose volume, which resulted in to $V_{dc} = 267$ doses per cold box.

$$V_{dc} = \frac{Cold\ box\ Volume}{Single\ dose\ volume} = \frac{7000}{26.2} = \text{ 267 doses per cold box}$$

By combining the vehicle's transportation capacity for the cold boxes, employee productivity, and the cold boxes' capacity, we calculated the capacity for doses per vehicle as:

1. Vaccination capacity per vehicle in terms of employee = 120 x 5 = 600 doses
2. Vaccination capacity per vehicle in terms of cold boxes = 267 x 3 = 802 doses

As the above numbers indicate, the bottleneck of the operation is not the capacity to transport the vaccines or the precise number of doses each cold box can hold, but rather the number of employees. The constraining factor of an outreach operation is the number of nurses available and their productivity. This point could be challenged with regard to productivity; however, the nurses often have the flexibility to extend their working day in case there is still demand at the outreach site. For this reason, even if our number of doses per hour is not precise, this can be offset by the flexibility in the working hours per day.

4.2.11 Summary of model inputs

In the past three sections we presented the reasoning to derive the input parameters of our optimization model. In this section, we summarized all inputs in Table 9.

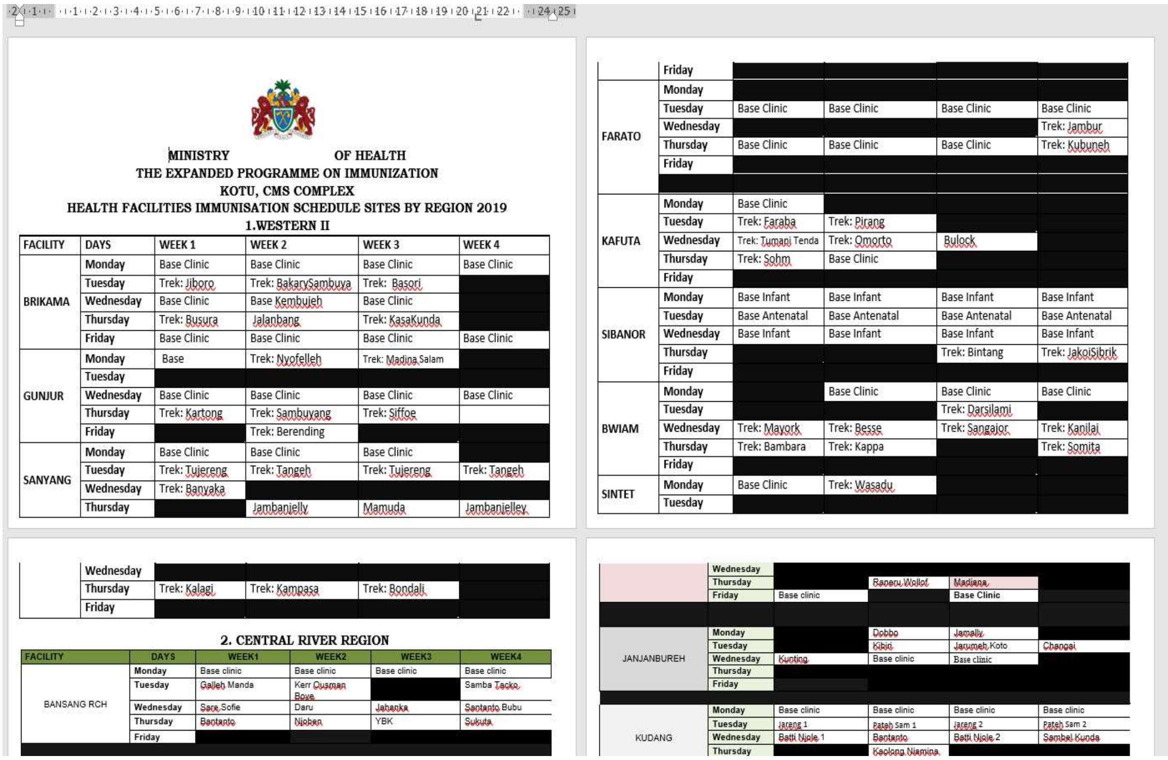
Table 9: Summary of model input parameters

Input variable	Description	Value	Unit
C_{ov}	Vehicle operation cost	0.6	$/km
C_v	Total cost to have a dose available at fixed health center	0.02	$ / dose
CE_o	Employee cost per day at fixed outreach site	3.68	$/emp.day
CE_f	Employee cost per hour at fixed health center f	3.68	$/emp.day
PE_o	Daily employee productivity at outreach o	120	Doses/emp.day
PE_f	Daily employee productivity at health center f	160	Doses/emp.day
V_{cv}	Number of cold boxes per vehicle	3	Cold boxes
Ve	Maximum number of employees per vehicle	5	Nurses
V_{dc}	Number of doses per cold box	267	Doses

4.2.12 Current Schedule

Each population region can be serviced either by a fixed health center or by an outreach site. According to the UNICEF interviews, and the data we had access to, the allocation of immunization professionals follow a fixed monthly schedule that determines for each day of the week where the team of each health center should go. The immunization schedule for the year 2019 was provided in a word file called "National Trekking Sites". An extract of the document provided is shown in Figure 23. As can be seen from the schedule, on some days of the week the team is allocated to neither in the health facility, referred to in the document as base clinic, nor to any of the outreach treks. This means that the team is located at the base clinic but is providing different services.

Our main objective with this dataset is to derive the information of how often an outreach site is visited and which health center is used as a base. To derive this information a data cleaning and standardization was performed, as detailed in the rest of this section.



MINISTRY OF HEALTH
THE EXPANDED PROGRAMME ON IMMUNIZATION
KOTU, CMS COMPLEX
HEALTH FACILITIES IMMUNISATION SCHEDULE SITES BY REGION 2019
1.WESTERN II

FACILITY	DAYS	WEEK 1	WEEK 2	WEEK 3	WEEK 4
BRIKAMA	Monday	Base Clinic	Base Clinic	Base Clinic	Base Clinic
	Tuesday	Trek: Jiboro	Trek: BakarySambuya	Trek: Basori	
	Wednesday	Base Clinic	Base Kembujeh	Base Clinic	
	Thursday	Trek: Busura	Jalanbang	Trek: KasaKunda	
	Friday	Base Clinic	Base Clinic	Base Clinic	Base Clinic
GUNJUR	Monday	Base	Trek: Nyofelleh	Trek: Madina Salam	
	Tuesday				
	Wednesday	Base Clinic	Base Clinic	Base Clinic	Base Clinic
	Thursday	Trek: Kartong	Trek: Sambuyang	Trek: Siffoe	
	Friday		Trek: Berending		
SANYANG	Monday	Base Clinic	Base Clinic	Base Clinic	
	Tuesday	Trek: Tujereng	Trek: Tangeh	Trek: Tujereng	Trek: Tangeh
	Wednesday	Trek: Banyaka			
	Thursday		Jambanjelly	Mamuda	Jambanjelley

	Wednesday				
	Thursday	Trek: Kalagi	Trek: Kampasa	Trek: Bondali	
	Friday				

	Friday				
FARATO	Monday				
	Tuesday	Base Clinic	Base Clinic	Base Clinic	Base Clinic
	Wednesday				Trek: Jambur
	Thursday	Base Clinic	Base Clinic	Base Clinic	Trek: Kubuneh
	Friday				
KAFUTA	Monday	Base Clinic			
	Tuesday	Trek: Faraba	Trek: Pirang		
	Wednesday	Trek: Tumani Tenda	Trek: Omorto	Bulock	
	Thursday	Trek: Sohm	Base Clinic		
	Friday				
SIBANOR	Monday	Base Infant	Base Infant	Base Infant	Base Infant
	Tuesday	Base Antenatal	Base Antenatal	Base Antenatal	Base Antenatal
	Wednesday	Base Infant	Base Infant	Base Infant	Base Infant
	Thursday			Trek: Bintang	Trek: JakoiSibrik
	Friday				
BWIAM	Monday		Base Clinic	Base Clinic	Base Clinic
	Tuesday			Trek: Darsilami	
	Wednesday	Trek: Mayork	Trek: Besse	Trek: Sangajor	Trek: Kanilai
	Thursday	Trek: Bambara	Trek: Kappa		Trek: Somita
	Friday				
SINTET	Monday	Base Clinic	Trek: Wasadu		
	Tuesday				

2. CENTRAL RIVER REGION

FACILITY	DAYS	WEEK1	WEEK2	WEEK3	WEEK4
BANSANG RCH	Monday	Base clinic	Base clinic	Base clinic	Base clinic
	Tuesday	Galleh Manda	Kerr Ousman Boye		Samba Tacko
	Wednesday	Sare Sofie	Daru	Jahanka	Santanto Bubu
	Thursday	Bantanto	Njoben	YBK	Sukuta
	Friday				

	Wednesday				
	Thursday		Raneru Wollof	Madiana	
	Friday	Base clinic		Base Clinic	
JANJANBUREH	Monday		Dobbo	Jamally	
	Tuesday		Kibiri	Jarumeh Koto	Changai
	Wednesday	Kunting	Base clinic	Base clinic	
	Thursday				
	Friday				
KUDANG	Monday	Base clinic	Base clinic	Base clinic	Base clinic
	Tuesday	Jareng 1	Pateh Sam 1	Jareng 2	Pateh Sam 2
	Wednesday	Batti Njole 1	Bantanto	Batti Njole 2	Sambel Kunda
	Thursday		Kaolong Niamina		

Figure 23: The immunization schedule for the year 2019.
Source: UNICEF, "National Trekking Sites"

- Data cleaning

In the National Trekking file, the facilities are designated by name and not by site ID. As we showed in Section 4.2.5, the geocoding of each health center and outreach site has been provided, but they are associated to the location's IDs. In order to find the addresses of the facilities in the schedule file, we first had to derive their IDs based on their names.

Before matching names and IDs, the first step in the data cleaning process was to transform data in the National Trekking sites, which is provided in Word format. Once the table was generated, it was necessary to normalize it and look in the master file for the ID of each of the facilities named. As expected, some facilities were not found.

After a manual check of each mismatch, we found out that the main reasons for no match were typing mistakes or the use of abbreviations. Table 10 shows as example of the corrections performed for the 11 health centers that originally were not found in the master file.

Table 10: Data cleaning procedure example in National Trekking file

Original name	Corrected name	ID
SAMI	Sami Karataba	89
Leman Street Clinic	Leman St. Clinic	254
Serre Kunda H/C	Serrekunda	256
Sukuta H/C	Sukuta	11
Faji Kunda H/C	Faji-kunda	257
Banjulinding H/C	Banjulinding	265
New Jeshwan	New Jeshwang	288
New Yundum H/C	New Yundum	289
BMCH Hospital	Bmchh	272
FARRAFENNI	Farafenni	156
JALLANGBEREH	Jalambereh Health Centre	155
Baja kunda	Bajakunda	336

In the schedule file there were a total of 61 health centers. After making corrections, we were able to link all of their names to facilities from the master file. Of these 61 health centers, 12 were in the schedule as only being used for the base clinic, leaving a total of 49 potential origins for the outreach treks. However, three of these were outreach sites, leaving 46 potential origins for the outreach treks.

However, the cleaning procedure could not solve the matching issue for all outreach sites. The schedule file presented 306 distinct outreach site names. Of those, 24 outreach sites could not have their name identified and crossed to the master file, leaving 282 outreach sites. Of the 282 outreach sites, eight were classified in the master file as health facilities and not outreach sites: removing those produces (282 – 8) 274 outreach sites. Finally, of those, nine were identified as being served by other outreach sites and were also removed. This left a final number of outreach candidates of (274 – 9) 263 sites. This will be the outreach sites considered in the baseline.

- Number of visits per month

Finally, after cleaning the schedule file, we analyzed the number of visits per month at health center and outreach site. The results are shown in Figure 24.

.

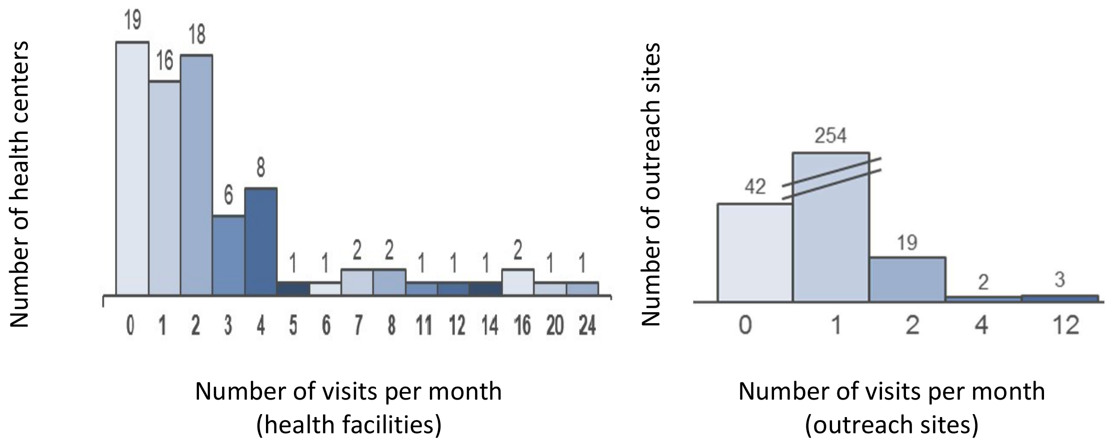


Figure 24: Histogram of number of number of facilities divided by their frequency visit
Source: 2019 National trekking file

The first aspect that calls attention in the health centers histogram are the 19 facilities with no visits. Those are health centers that are in the master file but do not have any trek in the schedule file. For these facilities we assumed that the reason why they were not in the outreach trek file is because they do not have outreach treks originated from them, which does not mean it does not provide immunizations services at the facility itself. As so, in the baseline, they will be considered as candidate locations to provide immunization services. A different situation applies for the 42 outreach sites that are in the master file but are not present in the schedule file. In this situation, we assumed that these outreach sites were not being used. This assumption was supported by the fact that UNICEF representatives reported that the national trekking sites information was more reliable than the master file.

Since each facility has the amount of the target population that is theoretically assigned, it is possible to generate a graph that evaluates the frequency that each population has access to immunization in their assigned site at a monthly basis. The result of this crossing is presented in Figure 25.

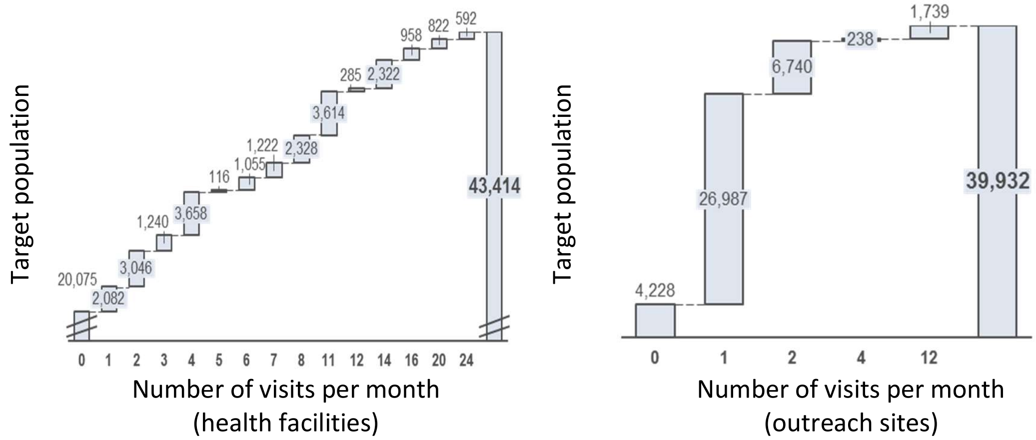


Figure 25: Histogram of total target population for each facility, dived by frequency of visits

As we can see from Figure 25, a total target population of 4,228 thousand people, roughly 5.07% of the target population, is serviced from outreach sites that theoretically are closed (0 visits). This means that, if the current population allocation to outreach sites was enforced in the baseline model, we should expect a maximum immunization access of 94.93%. That is one of the reasons why the population allocation decision for the baseline construction was not enforced.

4.3 Solution procedure – Code overview

The solution procedure of the optimization model consisted of three main steps: the data pre-processing, followed by the optimization routine and the result analysis. The full code developed is presented in GitHub, but a schematic view of it is presented in Figure 26. The code was developed using Python language embedded in a Jupyter notebook compiler. The solver used in the MIP optimization was the Gurobi. During the code development, a set of useful libraries were used, the most important ones are listed in Table 11.

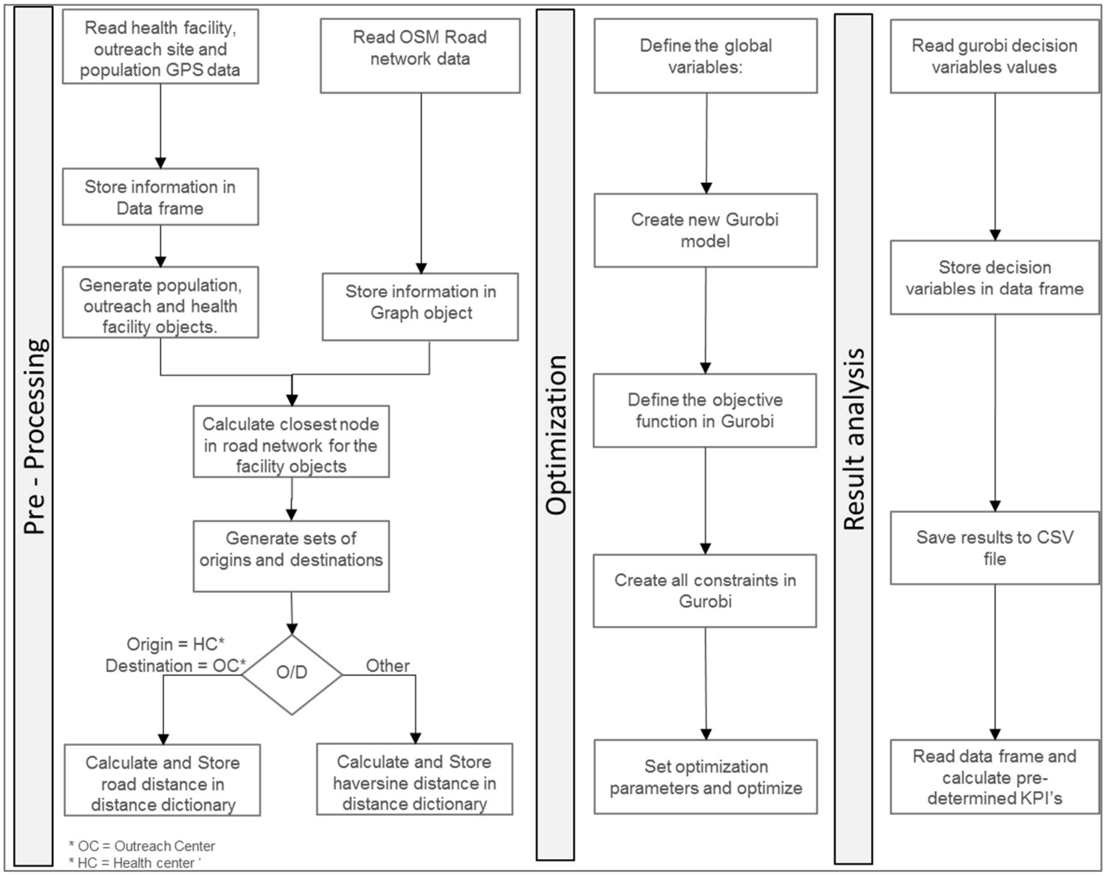


Figure 26: Solution procedure schema

Table 11: List of main python libraries used and their application

Library	Description
Pandas	Data processing
Gurobi	Optimization solver
Haversine	Calculation of earth circle distances
OSMNX	Import road network from OpenStreetMap
Networkx	Calculation of the road distances between od-s (solves shortest path problem)
seaborn	Data visualization

4.4 Optimization scenarios

In order to assess the effectiveness of the developed model, three different scenarios were run. In the first scenario, the model baseline is defined by running it with the current site visit constraints and only considering the current existing outreach sites. In the second scenario, the current number of visits to the outreach constraint is eliminated but only the existing outreach sites are considered. Finally, in the third scenario, both the current outreach sites location and visit schedule are optimized (Table 12).

Table 12: Overview of the differences between the three optimization scenarios

	1. Optimization baseline	2 Outreach Schedule optimization	3. Outreach schedule and location optimization
Population allocation	**Optimized**	**Optimized**	**Optimized**
Health center location	Current	Current	**Current**
Outreach site location	Current	Current	**Optimized**
Outreach allocation & schedule	**Current**	**Optimized**	**Optimized**

4.4.1 Scenario 1: Optimized baseline

The optimization baseline is the basis which the optimized scenarios will be compared to. This section presents how all relevant aspects of the vaccination network were derived.

- Health center and outreach site location

  As explained in Section 4.2.5 , the current location of health facilities and outreach sites was provided by UNICEF. However, since we are used the road network for calculating the distances between the facilities, both the health center and the outreach sites location were individually shifted to the respective closest node of the road network (Figure 27).

- Population data (demand for vaccines)

  As we presented in Section 4.2.1 and 4.2.2, we had two sources of population data available. The first, provided by UNICEF, had the information of total population allocated to each site. The second, derived from SEDAC, had the total population plotted in a map. As expected, some divergencies between the two datasets were identified. The SEDAC dataset must be the one used in scenarios optimized scenarios (2,3) since we needed a granularity greater than the one obtained with the current facilities. To keep the consistency between baseline and optimized scenarios we also used the SEDAC dataset in the baseline.

- Population allocation (demand for vaccines)

The decision of what will be the facility responsible for servicing each node of the network will be optimized already in the baseline.

- Outreach schedule (Resource bundles)

  In the baseline, the number of monthly visits to each outreach site, or in other words, the number of resource bundles sent to each outreach site, will follow the current schedule provided by UNICEF and presented in detail in Section 4.2.12.

- Demand function

  The optimization baseline results were calculated for the three different demand coverage functions presented in Section 4.2.4.

- Budget

  The optimization baseline was run for different monthly budgets, in steps of one thousand dollars until the unconstrained level is reached. As the number of bundles is being enforced, we should expect for the model to be infeasible until the minimum budget necessary to operate the current network is reached.

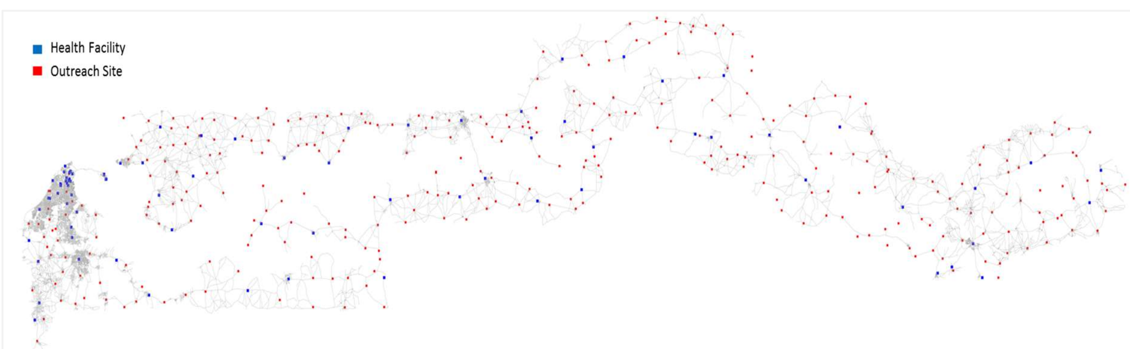


Figure 27: Outreach and fixed health centers of baseline scenario

4.4.2 Scenario 2: Schedule and allocation optimization with current outreach sites

The schedule optimization scenario had the objective to evaluate whether the current outreach site candidates can be used in a more efficient way. For the purpose of our model, efficiency can be defined on whether it is possible to achieve better vaccine access levels without changing the current outreach sites footprint. The increase in operational efficiency will be achieved through the variation of the following variables:

- Number of bundles sent from a health center *f* to outreach site *o* – X_{oj}
- What is the health center *f*, or outreach site o, servicing each population center - Y_{oj} , Y_{fj}
- Number of health facilities open
- Number of outreach sites used

Based on the results of the optimization it was possible to assess the requested number of vehicles necessary at each facility and number of employees. This can lead to suggestion of better allocation of the existing resources.

The same criteria used in the baseline for defining the number of scenario configurations due to different demand functions and budget was used in this schedule optimization scenarios.

- Budget: Starts from 0 and increases by 1,000 dollars until access levels reach a plateau
- Demand function: three different scenarios (conservative, moderate, and aggressive)

4.4.3 Scenario 3: Schedule, allocation and outreach location optimization

The last scenario, in addition to the population allocation optimization (Scenario 1) and the schedule optimization (Scenario 2) also optimizes the location of the outreach sites. To do so, all outreach sites candidates presented on Section 4.2.7 were considered.

5 Results and discussion

In the first part of this chapter, Section 5.1, the access levels and overall configuration of the network for three optimization scenarios in described in Section 4.4 are presented: (1) optimization baseline; (2) Schedule and allocation optimization with current outreach sites; and (3) the schedule and outreach sites location optimization. After that, in Section 5.2, a more detailed analysis was conducted, focusing on the budget and demand function that best represent current situation in The Gambia. Based on this detailed analysis, insights and recommendations were proposed.

5.1 Gambia case scenarios optimization results

This section presents the most important results obtained with the model optimization of each scenario under different demand functions and budget constraints. The primary results presented for each scenario are the immunization access levels and the number of open facilities.

5.1.1 Results Scenario 1: Optimization baseline

The optimization baseline is the scenario on which we keep the current outreach allocation and schedule. The schedule was enforced through a constraint on the resource bundle flows between the health centers and outreach sites. There are three demand functions: high (Figure 28 and Figure 29), moderate(Figure 30 and Figure 31), and low(Figure 32 and Figure 33).

- High reach demand function (2,10)

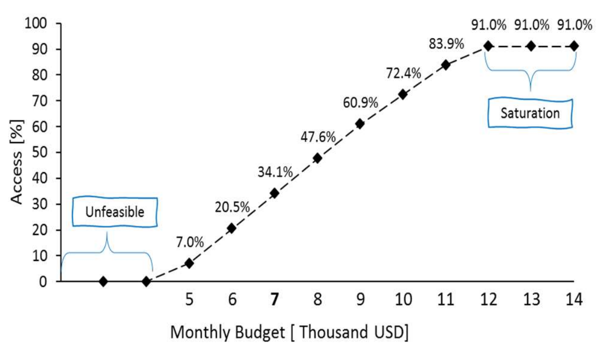


Figure 28: Vaccination access level for different budgets. High coverage demand function

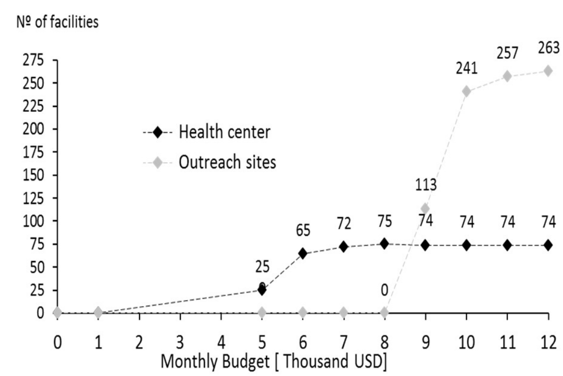


Figure 29: Number of used locations. High coverage demand function.

- Moderate reach demand function (1,5)

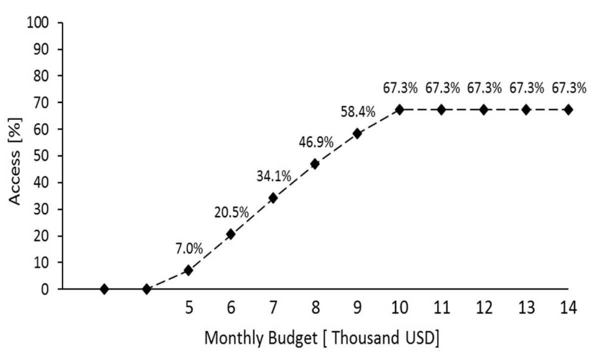


Figure 30: Vaccination access level for different budgets. Medium coverage demand function

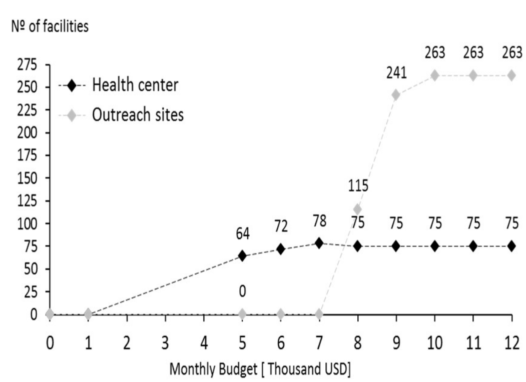


Figure 31: Number of used locations. Medium coverage demand function

- Low coverage Demand function (0,2.5)

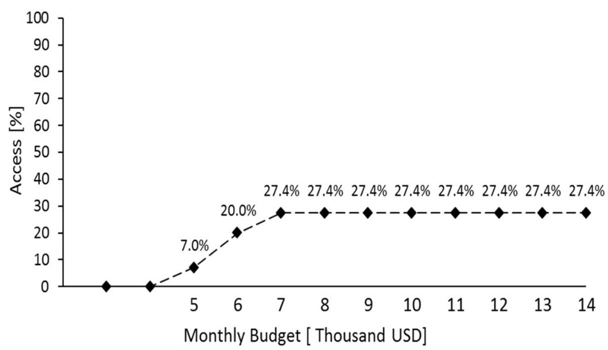


Figure 32: Vaccination access level for different budgets. Low coverage demand function

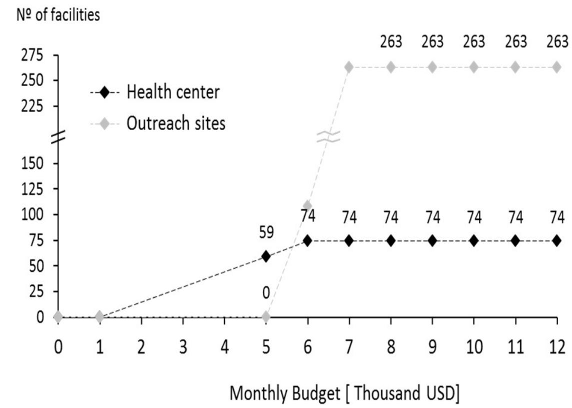


Figure 33: Number of used locations. Low coverage demand function

First, we analyzed the shape of the immunization access graph. In Figure 28, we highlighted two different areas of the graph, the Infeasibility and the saturation zones. The infeasibility happens because in the baseline we enforced the frequency of the outreach visits, which means there exists a lower bound on the amount of money necessary to guarantee all 263 outreach sites are opened. As the budget increases, more health centers are opened and the access levels increase at a proportional rate until it reaches the saturation point. The point of saturation represents the budget from which the access levels stop to increase regardless of how much money is available. This means that the maximum population that can be reached given the assumed demand function and existing outreach sites has been reached. We will define the optimal budget as the minimum amount of money necessary to reach the saturation point for a given demand function.

As expected, the point of network saturation happens at lower budget levels as the demand functions get stricter. As we can see from Figure 28, Figure 30, and

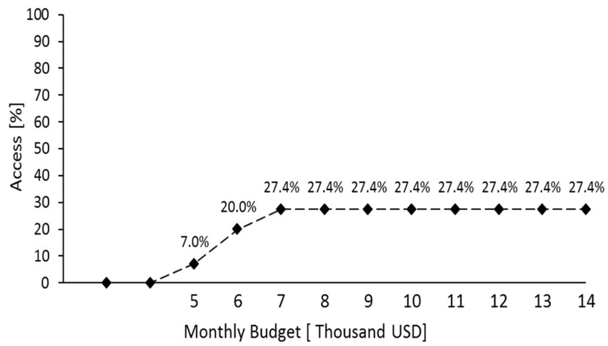


Figure 32, the points of saturation happen at 12 thousand US dollars, 10 thousand US dollars, and seven thousand US dollars for the high, moderate, and low coverage demand functions respectively.

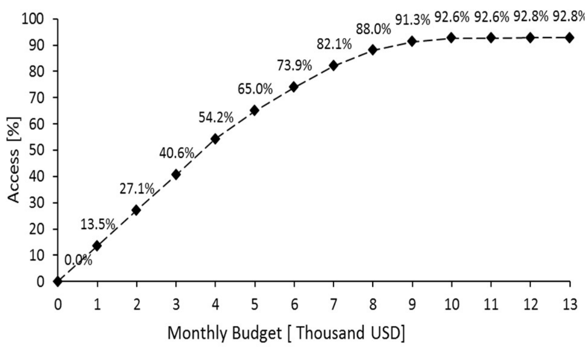
One important point to note is that the maximum access levels vary quite drastically depending on the demand function. As we can see from the graphs, the maximum access level is 91% at the high coverage function and is reduced to 67.3% and 27.4% at the medium and low coverage demand functions, respectively. A 27.4% access is a completely unrealistic number, indicating that this demand function is too strict. For this reason, we discarded this shape and only perform the optimization for high (2,10) and moderate (1,5) demand functions.

5.1.2 Results Scenario 2: Schedule and allocation optimization with current outreach sites

The results of schedule and allocation optimization with current outreach sites are presented based on two demand functions: high (
Figure 34 and

Figure 35) and moderate(Figure 36 and Figure 37).

- High coverage demand function (2,10)



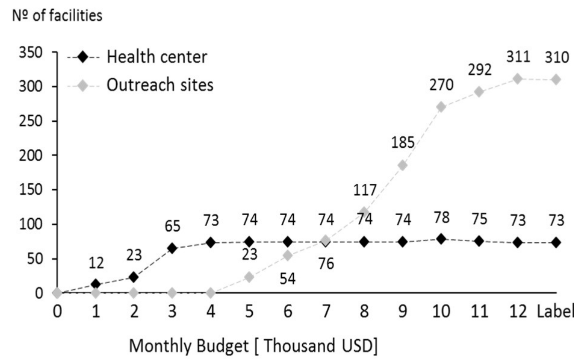
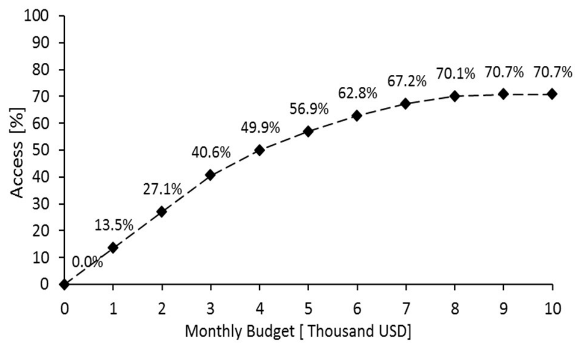


Figure 34: Scenario 2 – Vaccination access level and budget. High coverage demand function

Figure 35: Scenario 2 - High coverage demand function

- Moderate coverage Demand function (1,5)



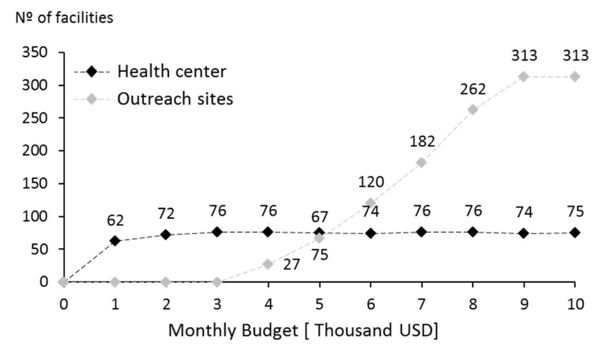


Figure 36: Scenario 2 - Vaccination access level and budget. Moderate coverage demand function

Figure 37: Scenario 2 - Number of open facilities. Moderate coverage demand function

The removal of the scheduling constraint eliminated the infeasibility zones. The saturation points occur now at 12 thousand US dollars and 9 thousand US dollars at access levels of 92.8% and 70.7% respectively. Those numbers represent a small increase in the access for the same amount of money. This is an indication that there is a margin for improvement in the asset utilization.

With regard to the evolution of the number of facilities we can observe that initially, the health centers are opened, and their numbers remain stable with small fluctuations. The stability of the health centers number in the different budget levels is a consequence of cheaper vaccination costs in health centers. For this reason, in case the budget is limited, the initial behavior is to open all health centers and vaccinate the nearby population.

5.1.3 Results Scenario 3: Schedule, allocation and outreach location optimization

The results of Schedule, allocation and outreach location optimization are presented based on two demand functions: high (Figure 38 and Figure 39) and moderate(Figure 40 and Figure 41).

- High coverage demand function (2,10)

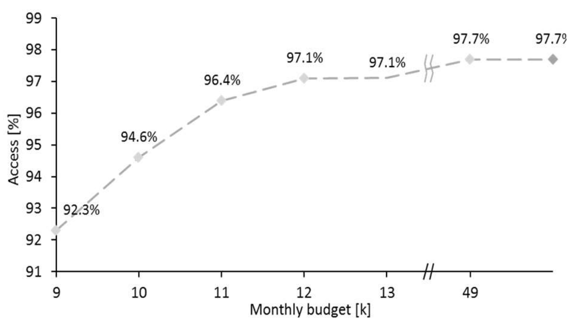


Figure 38: Scenario 3 - Vaccination access level and budget. Demand (2,10)

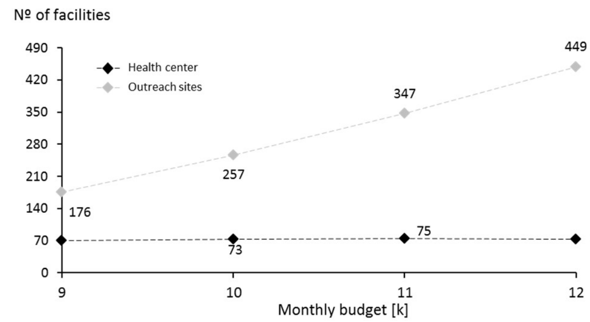


Figure 39: Scenario 3 - Number of open facilities. Demand (2,10)

- Moderate coverage demand function (1,5)

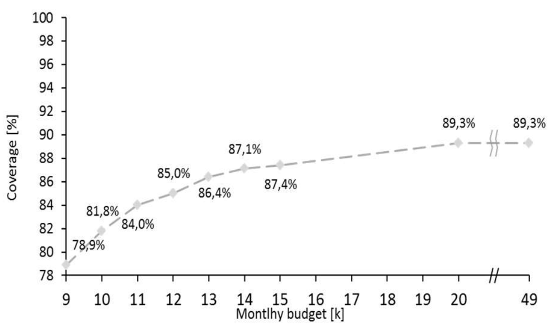


Figure 40: Scenario 3 - Vaccination access level and budget. Demand (1,5)

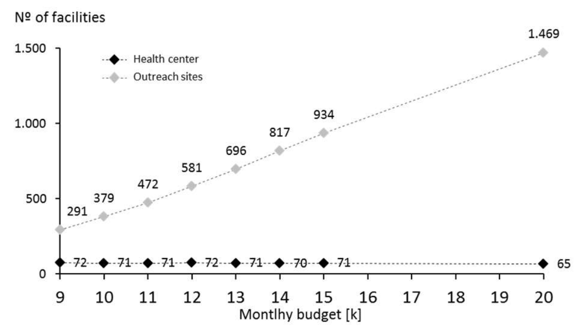


Figure 41: Scenario 3 - Number of open facilities. Demand (1,5)

Once we already had defined the region where the saturations happened, we focused our analysis on the range of budgets higher the nine thousand US dollars. Given the higher number of possibilities (around three thousand candidate outreach sites), the processing time of each scenario can take up to two hours. Therefore, we narrowed down the scope.

Two aspects worth highlight in the immunization access graphs. The first one is the significant larger access levels obtained in the location optimization scenarios when compared to the schedule optimization. For the high demand coverage this increase in access at 12 thousand US dollars was from 91% to 97.1%, and for the medium coverage demand function at 15 thousand US dollars access increased from 70.7% to 87.4%. These great increases are an indication of the effectiveness on maximizing the coverage when more outreach sites are opened and better located. A more detailed analysis is provided in the following section.

5.2 Gambia case analysis and recommendation

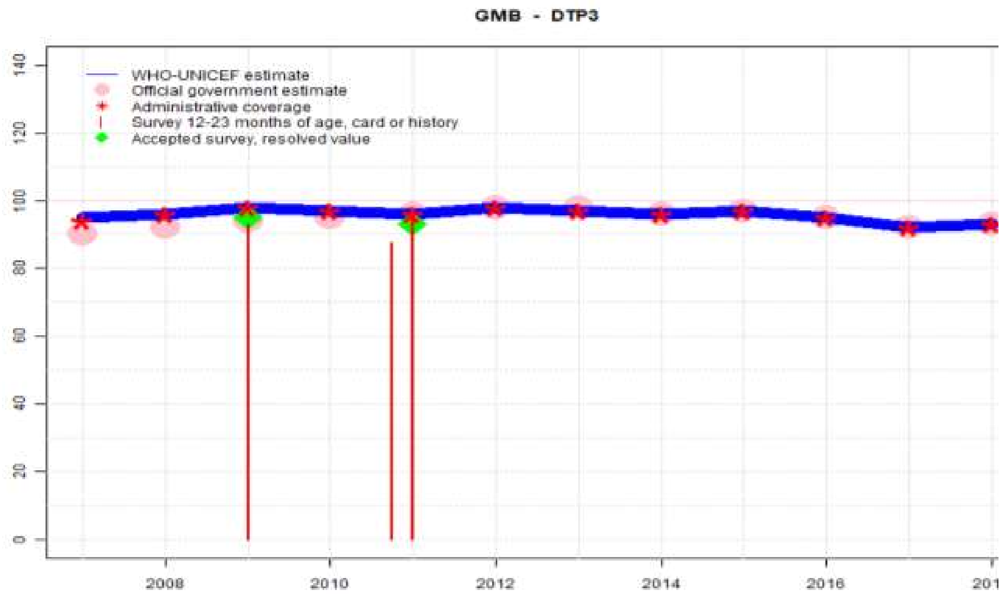
With the overall results presented for the different scenarios at different demand functions and cost budgets, we understand that the basic behavior of the model was validated. In this section is analyzed the network's configurations at a specific cost level and demand function. The objective is to better understand what is driving the increases on access levels as discussed in Section 5.1. This analysis suggests actionable measures to capture these gains, presented in Section 5.2.3.

5.2.1 Budget and demand function parameters definition

This section presents the reasoning behind the selection of the demand function parameter and budget level for performing the detailed analysis of the network.

STEP 1: Selecting the most suitable demand function parameters

In Section 4.2.4, we have established three different shapes for the demand function. The levels of vaccination access obtained for each scenario showed that at an unconstrained budget situation, three different demand functions (low, moderate, and high) would provide access levels of 27.4%, 67.3%, and 91% respectively. Current coverage estimates for a basic vaccine of the immunization schedule in The Gambia, the DTP3, was estimated at 93% in 2018 (Figure 42). Based on that, the demand function that most closely reflects the current network configuration is the high access shape.



	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Estimate	95	96	98	97	96	98	97	96	97	95	92	93
Estimate GoC	•	•	•••	•••	•••	•••	•••	••	••	••	••	••
Official	90	92	94	95	96	98	98	96	97	95	92	93
Administrative	94	96	98	97	96	98	97	96	97	95	92	93
Survey	NA	NA	93.2	NA	*	NA	NA	NA	NA	NA	NA	NA

Figure 42: The Gambia DTP3 coverage levels
Source: WHO - UNICEF estimate

STEP 2: Finding the network saturation budget level.

For each demand function the optimal network configuration was found under a set of different cost constraints. The observation of the vaccination access level graphs showed that at some point the access stops increasing regardless of how much money is available on both Scenarios 1 and 2, where the number of outreach sites is limited. In order to evaluate the effectiveness of our optimization algorithm, we compared the access level for the point on which the saturation point of the network is found in the baseline scenario, that is 12 thousand US dollars (Figure 43).

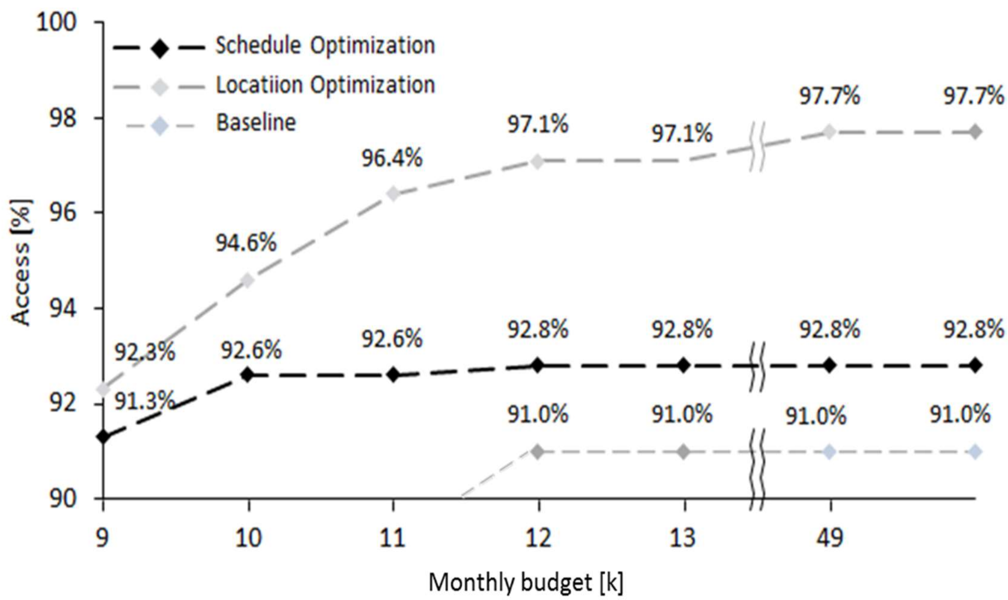


Figure 43: Comparison of access levels of different optimization approaches, demand function (2,10)

5.2.2 Understanding the results: analysis of The Gambia immunization network optimization scenario

In Section 5.2.1, we detailed the reasoning for selecting the budget level of 12 thousand US dollars and the high access demand function as the most reasonable boundary conditions to reflect the current vaccination network in The Gambia. Now our focus is to explore the configuration network of each of the solutions to derive intuition on how the increase on vaccination access is being obtained. It is important to notice that the drivers for the gains presented will be the same regardless of the parameters considered as the most likely for the current network. Our main objective here is not a precise number on how much access can be increased. As we showed previously, this is a difficult number to precisely estimate because we do not know exactly what the shape of the demand function is. However, we do believe our assumptions generated a good representation of the reality that suggests insights on the drivers for increasing the current vaccination level. With that in mind the following five factors will be analyzed in detail:

- Number of doses administrated per facility type
- Number of outreach sites open
- Average vaccination doses per trip

- Health center utilization: Total number of bundles (trips) sent from each health center
- Cost break down: Total cost with fixed facility X Outreach sites

For the sake of clarity, the results will be presented in two phases. First the schedule optimization, Scenario 2, will be compared with the baseline, Scenario 1. In a second moment, the location optimization model, Scenario 3, will be compared against the schedule optimization, Scenario 2 (Figure 44).

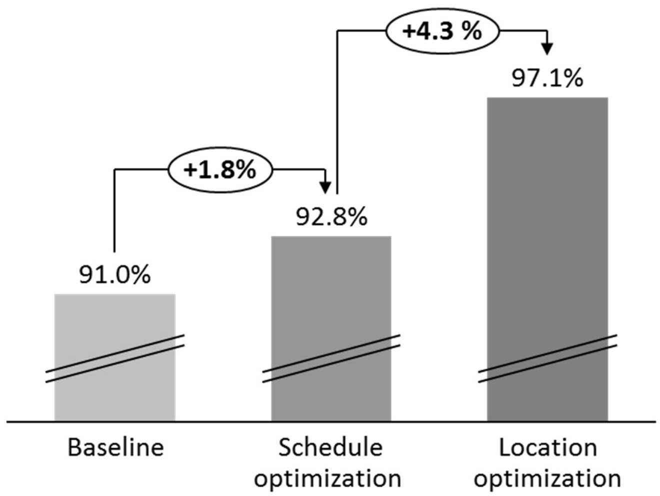


Figure 44: Access levels of different optimization approaches

5.2.2.1 Schedule optimization (Scenario 2) and Baseline (Scenario 1)

The first aspect analyzed was how the network configuration changed in terms of total number of health facilities and outreach sites locations. We expected that the removal of the arch constraints would lead to an increase in the number of outreach sites, and potentially to an unconstrained configuration where all 318 outreach sites would be used. As we can see from Figure 45, the schedule optimization scenario at a 12-thousand-dollar constraint uses 311 outreach sites. The possibility to increase the number of outreach sites while keeping the budget is a potential indication that there were inefficiencies in the Baseline.

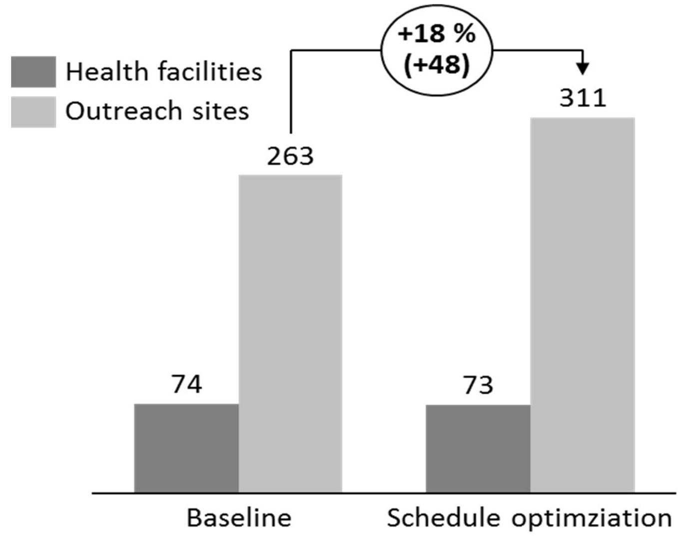


Figure 45: Number of facilities

To find the source of inefficiencies, we compared how the population were being serviced in the baseline versus the optimized scenario. The first interesting effect to highlight is a 17% increase in the amount of population serviced by outreach sites and a 11% decrease in the population serviced by health center (Figure 46 and Figure 47). It is cheaper to service someone from a fixed health center because no costs with the vehicle incur. Therefore, it could be expected that if a population center was serviced by a fixed health center in the baseline, that it would continue to be serviced by it in the optimized scenario. This would be true, if demand was a fixed property of the node and not varied with the distance. However, as Figure 49 shows, as new outreach locations open, populations that were previously only partially serviced by a single facility, due to its distance, may now have a closer location to visit, which would explain the partial reduction in the population service in health centers. In other words, the opening of new outreach centers increases the alternatives of service for some population centers and reduces service in fixed health centers.

Another aspect, illustrated in Figure 48, is the fact that the cost to serve from each outreach site reduced by 8%, from 11 cents to 10 cents when compared to the baseline scenario. The reason is that in the baseline the outreach sites are not allocated to the closest available health centers. When the allocation is optimized, there Is a drop in the average distance and consequently in the cost. Figure 50 shows that the average distance is reduced by 57% in the schedule optimized scenario in comparison with the baseline.

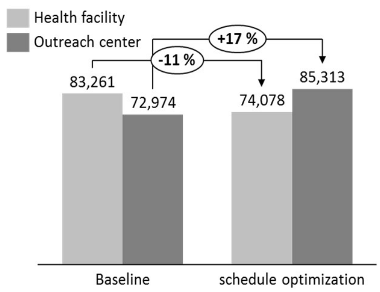


Figure 46: Total number of dose per facility type

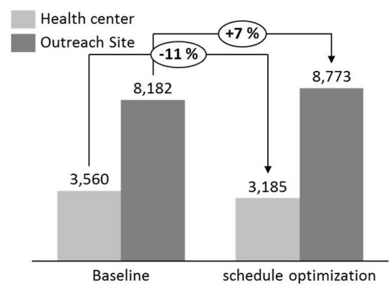


Figure 47: Total monthly cost per facility

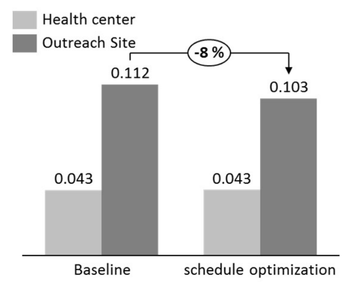


Figure 48: Cost per dose per facility type

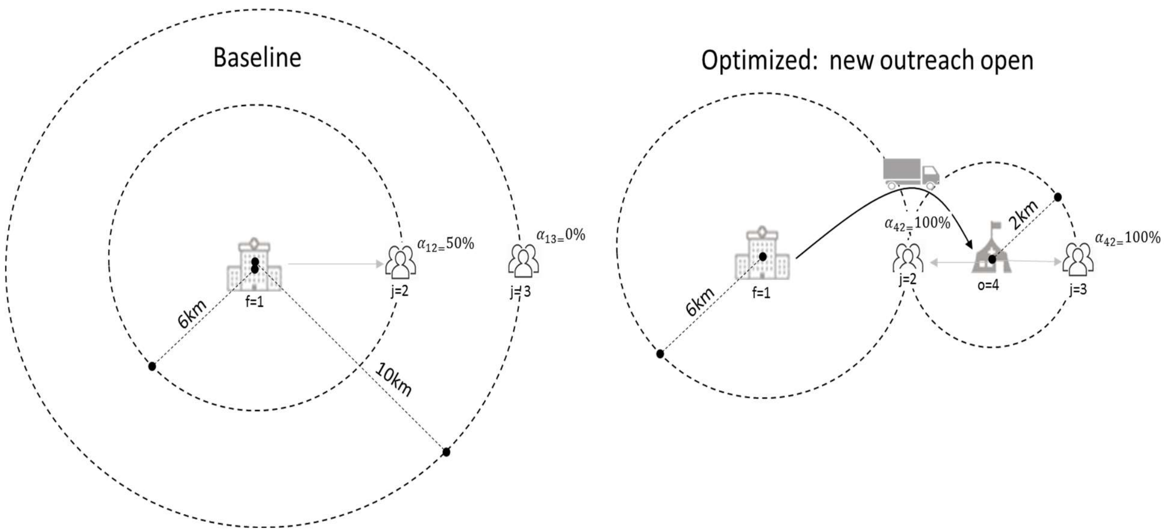


Figure 49: Mechanism of shifting to outreach sites

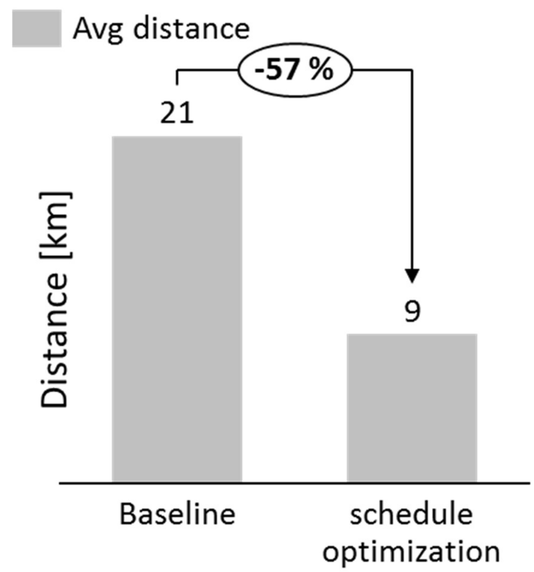


Figure 50: Average distance from health facilities to outreach sites

In this section we showed that the access to vaccination increased by 1.8% when the available set of outreach sites was optimized. This was obtained through a better allocation of the outreach sites to the fixed health centers with a significant reduction of the average distance traveled per outreach trip.

5.2.2.2 New outreach locations optimization (Scenario 3) and schedule optimizations (Scenario 2)

Following the same approach performed in the previous section, the first aspect of comparison between the resulting networks of the schedule x location optimization concerns the number of facilities. As Figure 51 shows, when the network is optimized with a large number of candidate outreach sites, the total number of outreach sites increased 44%.

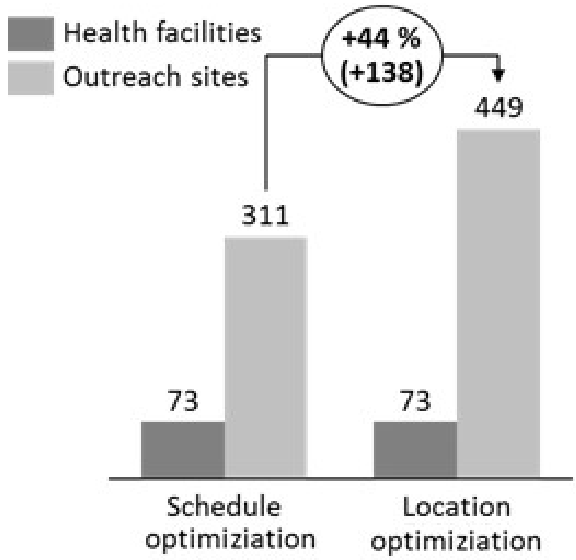


Figure 51: Number of sites per facility type

The increase in the number of outreach sites provides access to immunization for a larger amount of the population. Figure 52 shows that an increase of 29% in the total population service through outreach sites can be observed in the location optimization, together with a decrease of 23% of the population serviced by health centers. Just as in the previous optimization, it is also possible to observe a shift on the amount of population serviced from fixed health centers to outreach sites. As shown in Figure 53 the cost comparison shows a proportional reduce on costs for the health centers vaccinated population, but the increase on 29% of the population vaccinated through outreach sites only incurred an increase in costs of 9%. This was only possible due to a 16% reduction on the unitary vaccination cost through outreach sites (Figure 54).

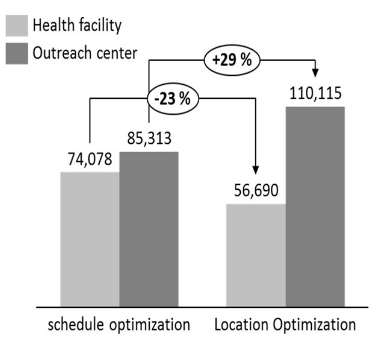


Figure 52: Total Number of doses per facility type

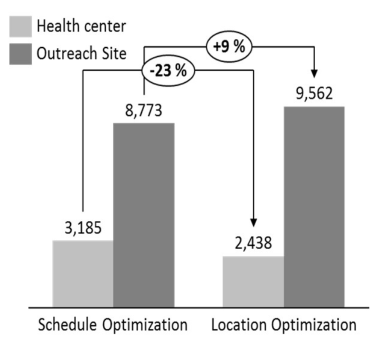


Figure 53: Total monthly cost per facility type

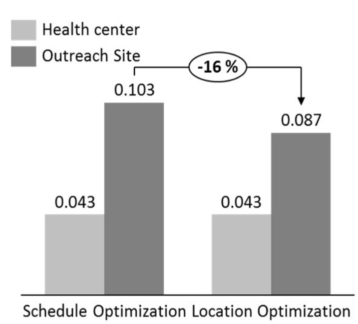


Figure 54: Cost per dose per facility type

In the baseline optimization we showed how the reason behind the unitary cost reduction of vaccination through outreach sites was associated to a decrease on the average distance from health centers to outreach sites. The increase on the number of outreach sites would intuitively drive to a network with closer distances between health center to outreach sites, and as Figure 55 showed, this intuition is confirmed with a reduction on 11% on the average distance.

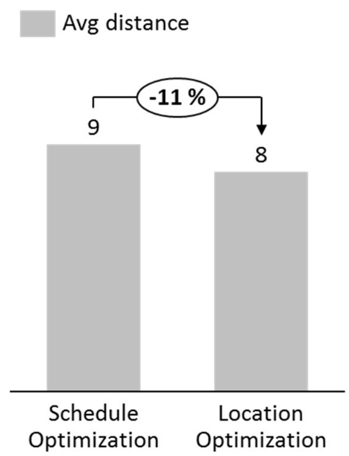


Figure 55: Average distance from health facilities to outreach sites

However, the reduction of the average distance is not the only factor contributing to a lower unitary cost for the outreach sites. A comparison of the schedule optimization scenario versus the location optimization scenario shows a 10.2% reduction in the number of trips (Figure 56). This reduction is reflected on a higher number of doses per trip, which leads to a better dilution of the fixed vehicle costs on each dose (Figure 57). In other words, asset utilization increases.

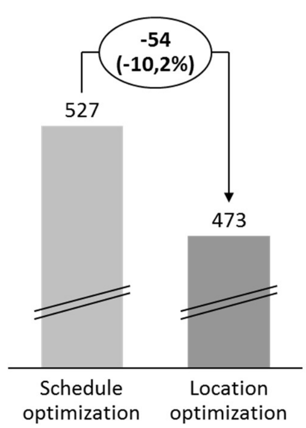


Figure 56: Number of outreach trips

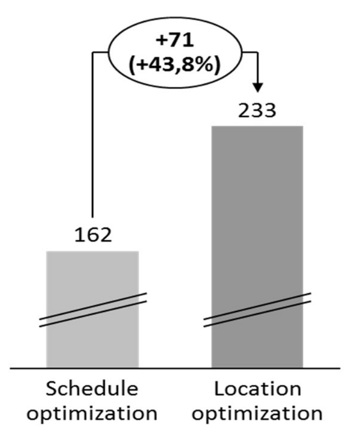


Figure 57: Average number of doses per outreach trip

5.2.2.3 Operational aspects: Number of vehicles and staff requirements

We ran the optimizations without the number of employees or vehicle constraints on health centers. The reason is that our objective was to explore the potentials of the network without the limitation of the current operational conditions. As previously pointed out, there are two aspects that could be a potential constraint to the new configurations: (1) The number of vehicles required to perform the outreach trips; and (2) the number of employees required at each health facility. This section presents in more details the resource consumption of the optimized solutions. For the sake of simplicity, we focused on the analysis in a comparison of the baseline (Scenario 1) versus the location optimization (Scenario 3).

- **Resource bundles (number of vehicles)**

  Initially, the analysis focused on the number of vehicles. Figure 58 showed the total number of trips sent per month from each locality (the number of bundles) at a health center level. As expected, the total number of trips leaving the health center in the baseline is higher in the optimized scenario. However, it is important to notice that no facilities have over 20 trips per month. This is an especially important result because it guarantees that the optimized solution is feasible even if each facility only has one vehicle available per day.

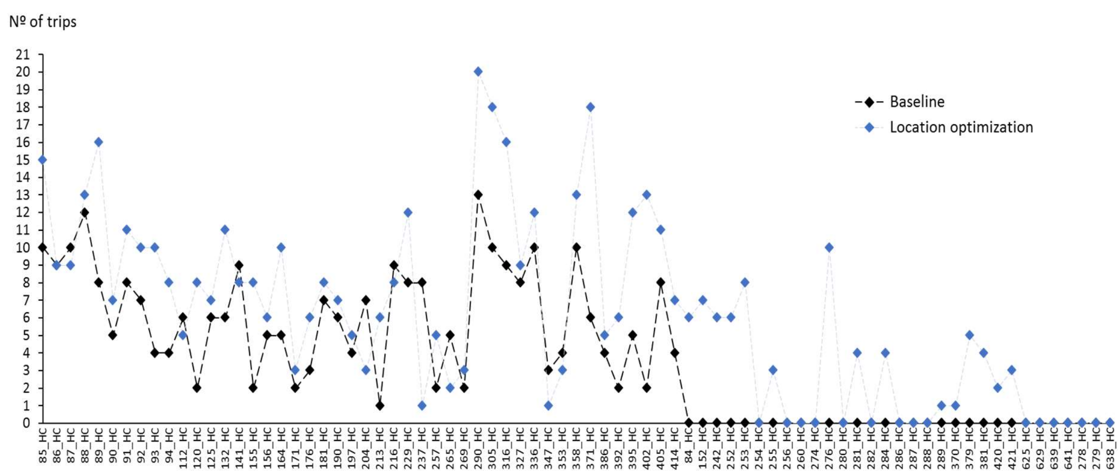


Figure 58: Number of vehicles required per month, per health center ID, for outreach trips

One important aspect to notice is that in the baseline only 46 out of the 80 health centers available have outreach trips originating from it, with each health center sending on average 3.6 trips per month. In the optimized scenario, is possible to observe from the graph that 15 extra health centers were used to send outreach trips, and also the average numbers of trips increase to 6.1 per month.

The profile of the trips also differs in terms of number of nurses required in each trip. Figure 59 showed the average number of car occupants in the trips leaving each health facility. The average value for the optimized scenarios 1.6 whereas in the baseline is 1.3.

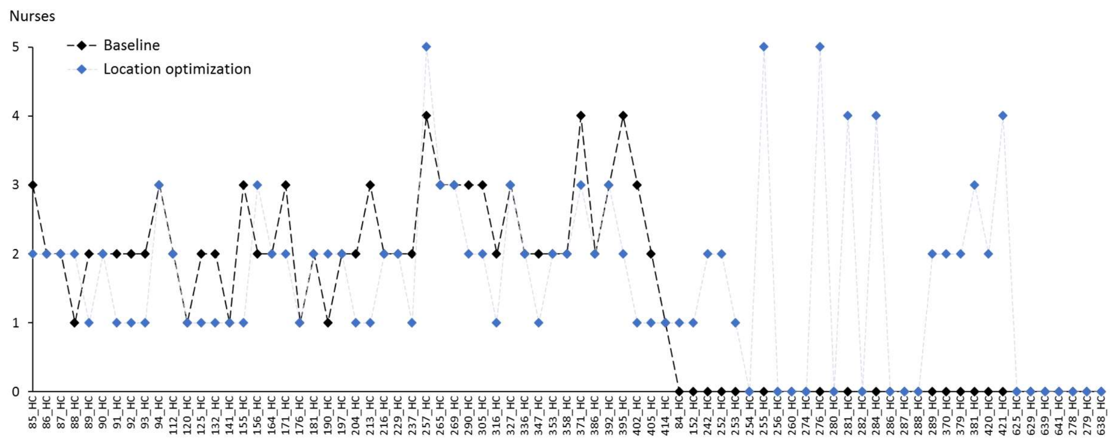


Figure 59: Average number of nurses per outreach trip

It is possible to conclude the optimal solution sends more trips, from more origins with more employees at a lower distance, and that the current available assets are enough to fulfill the increase in the size of the operation if vehicle sharing is allowed.

- **Labor in health centers**

Based on the total amount of vaccines administrated in the health centers, it is also possible to arrive in the requirement of employees-day required in each health center to fulfill the demand of people vaccinated in the health center. This does not account for the employee requirements for the outreach sites. This information can be used, in conjunction with the planned trips, to determine the staffing in each location. Overall, the average number of employees-day in the baseline is 7.3 while in the optimized solution is 4.3 (Figure 60).

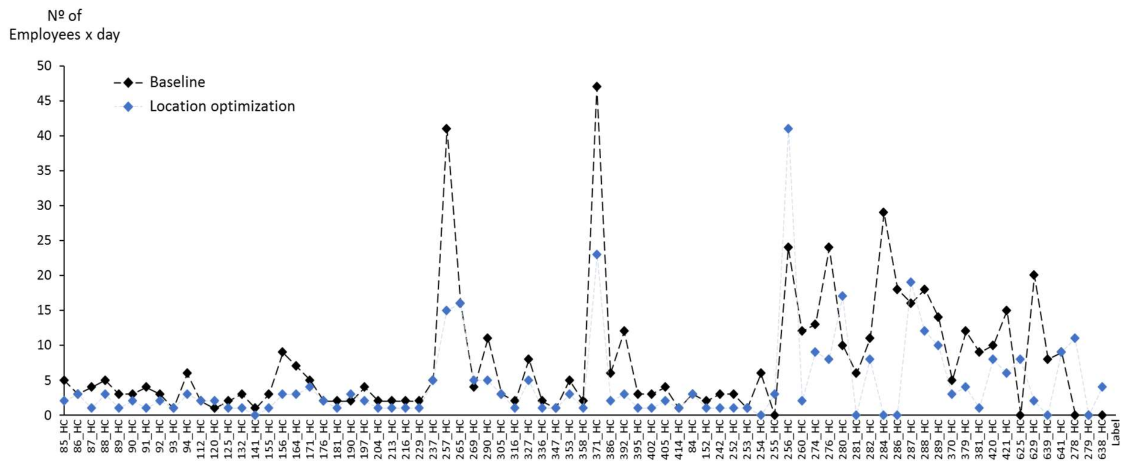


Figure 60: Number of "employees-day" requirements at health centers

5.2.3 Insights and recommendation

The main purpose of this case study is to derive insights and lay out ideas and recommendations for how the vaccination access levels can be increased in The Gambia network. This exercise does have its limitations, especially regarding the demand function parameters, but the underlying factors that could drive the increase in access should not change.

After defining a reasonable working baseline with total access level of 91% we showed how the access levels could be increased in a two-step approach: First by optimizing the frequency of visits to the outreach sites (schedule optimization), and later by opening new outreach sites. The main insights obtained are:

- Outreach allocation: Serve from the closest health center to reduce transport costs.

  The baseline construction showed that an average distance of 21 km is the current practice in The Gambia network. After an optimization of the allocation it was showed that by only considering the current set of health facilities it would be possible to reduce the average distance to only 9 km. Our recommendation is that whenever it is operationally feasible each outreach site should be serviced from the closest health center base.

- Outreach site visits frequency is important.

  The number of times each outreach site is visited is an aspect that should be paid close attention to. The tool developed here helps to guarantee that outreach sites are not visited more than they should; an asset that is used to visit the same location two times could be used to reach a new area previously not serviced. We also observed that the current schedule only uses 82% of the outreach sites listed in the master file. This could be a problem in the dataset and should be further investigated.

- There is potential to open new outreach sites, and their location can be optimized with our tool.

  The main finding of the exercise is how much the opening of new outreach sites can increase access levels. We showed how sensitive the optimal number of outreach sites is to the demand function parameters selected. We also showed that for the high access demand function the number of outreach sites that maximizes the immunization access would be 445.

- Use more health centers as origins for the outreach operations.

  We showed that as the number of outreach sites increases, the number of bases for these trips should also increase. In the current operation, only over half of the health centers are used as origins for the outreach trips. Increasing the number of origins can lead to lower vehicle operating costs. If necessary, vehicles could be shared between two health centers.

- Optimize the number of nurses per outreach trip.

  We do not have data on the current number of nurses sent on each outreach trip, but we showed that the optimized number is far from the total vehicle capacity. The number of nurses should not be fixed and may vary depending on the size of the outreach campaign.

- The optimal number of outreach sites as well as the access levels are heavily dependent on the shape of the demand function, therefore it is necessary to collect further immunization coverage data.

  In the different runs, we also showed how the access levels and number of outreach sites at the unconstrained network changed significantly depending on the demand function selected. We also showed that there is no wide agreement in the past studies in the shape of demand function.

To derive a demand function, data are necessary. Ultimately, what is needed would be information at a grid level (1 x 1 km, for example) regarding the closest vaccination point and the current immunization access level. These two pieces of information combined would allow us to derive a demand function influenced only by distance. In addition, we could potentially include the amount of time it takes to get to the facility, the region income level, and the access to transportation levels.

Knowing the importance of the demand function allows us to conclude that this is an important future research field. To arrive at this demand function, a machine learning algorithm or statistical regressions could be performed. Regardless of approach, data are essential. With that information, the Expanded Program on Immunization team could work to increase the visibility of the current coverage levels at a more granular level, and also enhance this information with auxiliary data that could drive access to immunization. We suggest that these information could be collected through interviews during the current outreach trips that are already serving the communities.

5.2.4 Summary of analysis and recommendation

We discussed in this chapter that there are opportunities to increase the access levels. To this end, the immediate actions that could be taken are a revision of the current outreach to health center allocation and a change in the frequency of the outreach trips. We also discussed how our model can suggest optimal outreach locations and how the number of optimal locations is highly dependent on the demand function shape. A recommended approach would be to increase the visibility of the current immunization access by obtaining more granular data on demand coverage function. Our model is flexible to incorporate any shapes of demand function and it could be used to optimize the number of outreach sites.

6 Conclusion

This study had as its main objective the elaboration of an efficient vaccine network model that incorporated, through an endogenous demand function, the distance impact in vaccination coverage. Through a series of discussions with UNICEF experts, MIT researchers and a literature review, we developed a unique optimization model that proved its efficiency through a successful case study application with data from The Gambia.

Through a structured approach to treating the provided data in order to derive the necessary parameters for the model, a baseline was created, and a series of optimization scenarios were run for different budget levels and demand coverage functions. Based on the analysis of the results, managerial insights were generated on how to increase the performance of the existing vaccination network. This performance was quantified through an increase of immunization access under certain budget constraints. It was showed that there is a lack in the current literature of a universal vaccination coverage demand function or a procedure to derive one. Since the shape of the demand function proved to have a high impact on the results of the model, we highlighted the importance of future research to further detail a specific demand coverage function for The Gambia. From this perspective, we suggested potential aspects that could be incorporated in such a function and the importance of gathering new data for this purpose.

Our developed model's main contribution is its flexibility. This flexibility is present in both the way the resource constraints were created and in the use of publicly available sources to derive demographic and distance data. Another key concern and important contribution of our study was the detailing of how the main inputs of the optimization model could be derived. As identified in our literature review, there are many good models available for vaccine network optimization, but little work has been done on how these models can be applied and how to obtain the data needed for them. We believe this often limits their applicability in real use cases.

Through the creation of the concept of a resource bundle, we developed a structured approach for formulating constraints that control the resources necessary for establishing outreach sites. The resource bundle concept is a creative way to quantify the outreach vaccination capacity and more precisely control the total cost of the operation, without the incorporation of many complexities to the model.

In terms of data gathering, we showed how to obtain the necessary population data from a publicly available source, the SEDAC, and, moreover, how demographic information could be converted into vaccine demand based on a country vaccination schedule. Another key contribution of our model is its connection to an online Open Street Map API, allowing the import of road distances to estimate the transportation costs of the outreach operations. To the best of our knowledge this is the first study to incorporate such a practical matter, which is of specific and great importance for countries like The Gambia that have an unusual geography that makes linear approximations less precise.

The results of The Gambia case study showed the ability of the model to increase immunization access. Through the opening of new outreach sites and an optimization of the outreach allocation and scheduling we would be able to increase immunization access from 91% to 97.1%. We showed that the magnitude of the increase relies on a lot of assumptions. Specifically, we highlighted the importance of having a more precise demand function, and we identify this as a promising area for future work. Regardless of the magnitude of the increase, the model has proved its value, since it generated good managerial insights to increase immunization access.

7 References

Blanford, J. I., Kumar, S., Luo, W., & MacEachren, A. M. (2012). It's a long, long walk: Accessibility to hospitals, maternity and integrated health centers in Niger. International Journal of Health Geographics, 11(1), 24. https://doi.org/10.1186/1476-072X-11-24Chen, S.-I., Norman, B. A., Rajgopal, J., Assi, T. M., Lee, B. Y., & Brown, S. T. (2014). A planning model for the WHO-EPI vaccine distribution network in developing countries. IIE Transactions, 46(8), 853–865. https://doi.org/10.1080/0740817X.2013.813094

Chen, S.-I., Norman, B. A., Rajgopal, J., Assi, T. M., Lee, B. Y., & Brown, S. T. (2014). A planning model for the WHO-EPI vaccine distribution network in developing countries. IIE Transactions, 46(8), 853–865. https://doi.org/10.1080/0740817X.2013.813094

Church, R., ReVelle C. (1974). The Maximal covering location problem. Papers in regional science. Daskin, M. S., & Dean, L. K. (2004). Location of Health Care Facilities. In M. L. Brandeau, F. Sainfort, & W. P. Pierskalla (Eds.), Operations Research and Health Care: A Handbook of Methods and Applications (pp. 43–76). https://doi.org/10.1007/1-4020-8066-2_3

Duijzer, L. E., van Jaarsveld, W., & Dekker, R. (2018). Literature review: The vaccine supply chain. European Journal of Operational Research, 268(1), 174–192. https://doi.org/10.1016/j.ejor.2018.01.015

Echakan, D. E., Oluoch, M. M., & Osuga, M. B. (2018). Factors Affecting Access to Immunization Services in the Integrated Outreach Model: A Case Study of Loima Sub County. Journal of Health, Medicine and Nursing, 3(2), 37–58.

Federal Reserve Bank of St. Louis. Retrieved on April 17, 2020 from https://fred.stlouisfed.org/series/FPCPITOTLZGGMB

Feikin, D. R., Nguyen, L. M., Adazu, K., Ombok, M., Audi, A., Slutsker, L., & Lindblade, K. A. (2009). The impact of distance of residence from a peripheral health facility on pediatric health utilisation in rural western Kenya. Tropical Medicine and International Health, 14(1), 54-61. https://doi.org/10.1111/j.1365-3156.2008.02193.x

Foroyaa Newspaper. (2019). Salaries in the Health Sector. Retrieved on April 17, 2020 from https://foroyaa.gm/salaries-in-the-health-sector/

Gu, W., Wang, X., & McGregor, S. E. (2010). Optimization of preventive health care facility locations. International Journal of Health Geographics, 9(1), 17. https://doi.org/10.1186/1476-072X-9-17

Hasanzadeh-Mofrad, M. (2016, June 15). Optimizing Vaccine Clinic Operations in Low- and Middle-Income Countries [University of Pittsburgh ETD]. Retrieved on November 5, 2019, from http://d-scholarship.pitt.edu/27262/

Ibnouf, A.H., Van den Borne, H.W. & Maarse, J.A.M. (2007). Utilization of family planning services by married Sudanese women of reproductive age. EMHJ - Eastern Mediterranean Health Journal, 13 (6), 1372-1381, 2007 https://apps.who.int/iris/handle/10665/117388

Lemmens, S., Decouttere, C., Vandaele, N., & Bernuzzi, M. (2016). A review of integrated supply chain network design models: Key issues for vaccine supply chains. Chemical Engineering Research and Design, 109, 366–384. https://doi.org/10.1016/j.cherd.2016.02.015

Lim, J. (2016, September 20). Improving the design and operation of WHO-EPI vaccine distribution networks [University of Pittsburgh ETD]. Retrieved on November 11, 2019, from http://d-scholarship.pitt.edu/28696/

Lim, J., Claypool, E., Norman, B. A., & Rajgopal, J. (2016). Coverage models to determine outreach vaccination center locations in low- and middle-income countries. Operations Research for Health Care, 9, 40–48. https://doi.org/10.1016/j.orhc.2016.02.003

Manual, C., Mckinnon J. (2019). Supply Chain Design: The Gambia. Assessment of the Vaccine Supply Chain Design in The Gambia. UNICEF Internal report.

Mokiou, S., Standaert, B., Li, X., De Cock, E. (2017). Measuring the cost of a pediatric vaccine administration in the UKS.

Russell, T., & Jauffred, F., & Goentzel, J. (2019, March). MIT-BASF Research Report: Network Competitive Advantage. MIT Center for Transportation and Logistics.

Schilling, D. A. (1993). A Review of Covering Problems in Facility Location. Location Science, 1, 25–55.

Socioeconomic Data and Applications Center. Retrieved from https://sedac.ciesin.columbia.edu/

Tanser, F. (2006). Methodology for optimising location of new primary health care facilities in rural communities: A case study in KwaZulu-Natal, South Africa. Journal of Epidemiology & Community Health, 60(10), 846–850. https://doi.org/10.1136/jech.2005.043265

UNICEF. (2019). UNICEF annual report 2018. Retrieved on May 4, 2020 from https://www.unicef.org/reports/annual-report-2018

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

Verter, V., & Lapierre, S. D. (2002). Location of Preventive Health Care Facilities. Annals of Operations Research, 110(1), 123–132. https://doi.org/10.1023/A:1020767501233

World Health Organization (WHO). Retrieved on April 17, 2020 from http://apps.who.int/immunization_monitoring/globalsummary/wucoveragecountrylist.html

WHO Regional Office for Africa. (2019). Experts caution against stagnation of immunization coverage in Africa. Retrieved from https://www.afro.who.int/news/experts-caution-against-stagnation-immunization-coverage-africa

Yang, Y., & Rajgopal, J. (2019). Outreach Strategies for Vaccine Distribution: A Two-Period Robust Approach. ArXiv:1908.10465 [Math, Stat]. Retrieved from http://arxiv.org/abs/1908.10465

Yang, Bidkhori H., & Rajgopal, J. (2019). Optimizing vaccine distribution networks in low and middle-income countries. ArXiv:1907.13434 [Math, Stat]. Retrieved from ttp://arxiv.org/abs/1907.1343

Appendix A: Toy model formulation

As discussed in Section 3.1, Appendix A summarizes the mathematical formulation of the toy model. The objective of this section is to present a set of equations that comprise the mathematical formulation of the Toy model.

First, sets of variables in this optimization problem are defined. From a modelling perspective, they can also be seen as the nodes of the vaccination network. We defined fixed health centers, outreach sites, and population.

Sets – Network nodes

- F = set of fixed health centers f
- O = set of potential outreach sites o
- J = set of population-regions j

Second, the problem variables are defined. They represent the optimization problem decisions. We defined the number of people vaccinated by fixed health centers and by outreach sites, the number of employees working in fixed health centers, controls if outreach site is open or not, controls if there are any people vaccinated by fixed health centers, and controls if there are any people vaccinated by outreach sites.

Variables

- X_{fj} = number of people from region *j* vaccinated by fixed health centers *f* [Continuous]
- X_{oj} = number of people from region *j* vaccinated by outreach site o, [Continuous]
- E_f = number of employees working in fixed health center F, [integer]
- Z_o = controls if outreach site *o* is open or not, [binary]

- Y_{fj} = controls if there are any people from region *j* vaccinated by fixed health centers f, [binary]
- Y_{oj} = controls if there are any people from region *j* vaccinated by outreach sites o, [binary]

$$X_{fj}, X_{oj}, E_F, Z_o \geq 0 \; \forall f \in \mathrm{F}, \forall o \in \mathrm{O}, \forall j \in \mathrm{J}$$

Third, the parameters of the models are defined. They include the minimum number of employees in fixed health centers, employee cost in fixed health centers, the number of vaccines which a single employee can serve, fixed cost of setting outreach sites, the number of people which an outreach site can immunize, demand factor for fixed health centers, demand factor for outreach sites, total demand for immunization, demand for immunization which fixed health centers can capture, demand for immunization outreach sites can capture, total budget, cost per vaccine in fixed health centers, and cost per vaccine in outreach sites.

Fixed health centers

- $Emin_f$ = minimum number of employees which a fixed health center must have
- *Ce* = employee cost in fixed health centers
- *Ve* = number of vaccines which a single employee can serve

Outreach sites

- *Co* = fixed cost of setting outreach site
- *Vo* = number of people which an outreach site can immunize

Demand

- α_{fj} = vaccination coverage between fixed health center f to region j

- α_{oj} = vaccination coverage outreach site *o* to region j
- D_j = total demand for immunization in region j
- d_{fj} = demand for immunization which fixed health center *f* can capture in region *j*
- d_{oj} = demand for immunization outreach facility *o* can capture in region *j*

General

- B = total available budget
- c_{fj} = cost per vaccine applied from fixed health center *f* to region j
- c_{oj} = cost per vaccine applied from outreach site *o* to region j

Fourth, the objective function is determined. In general, the objective function means what the problem aims to optimize. Our objective function is to maximize the total immunized population, which is the sum of the population vaccinated by fixed health centers and the population vaccinated by outreach sites.

$$max \sum_{j \in J} \left(\sum_{\mathrm{f} \in F} X_{fj} + \sum_{o \in O} X_{oj} \right)$$

Fifth, seven constraints are set. Constraints 1 and 2 control that the amount of people vaccinated in region *j* from a fixed health center *f* or outreach site *o* is lower than the amount of demand that fixed health center or outreach site can capture in those regions. Constraint 3 guarantees that the minimum capacity of fixed health center *f* is respected. Constraints 4 and 5 guarantee that the maximum capacity of fixed health center *f* and outreach site is respected. Constraint 6 guarantees that the total budget is higher than the total cost. The costs are given by the sum of three components: Fixed health center operation, outreach site implementation, and cost to serve (costs impacted by distance such as transportation). Constraint 7 guarantees that each region is serviced by one, and only one, facility.

$$X_{fj} \leq d_{fj} * Y_{fj} \qquad \forall f \in \mathrm{F}, \forall j \in \mathrm{J} \qquad (1)$$

$$X_{oj} \leq d_{oj} * Y_{oj} \qquad \forall o \in \mathrm{O}, \forall j \in \mathrm{J} \qquad (2)$$

$$E_f > Emin_f \qquad \forall f \in \mathrm{F} \qquad (3)$$

$$\sum_{j \in \mathrm{J}} X_{fj} < E_f \cdot \mathrm{Ve} \qquad \forall f \in \mathrm{F} \qquad (4)$$

$$\sum_{o \in \mathrm{O}} X_{oj} < \mathrm{Vo} \cdot \mathrm{Zo} \qquad \forall o \in \mathrm{O} \qquad (5)$$

$$\sum_{f \in F} E_f \cdot \mathrm{Ce} + \sum_{o \in O} Z_o \cdot \mathrm{Co} + \sum_{j \in J} \left(\sum_{f \in F} c_{fj} \cdot X_{fj} + \sum_{o \in O} c_{oj} \cdot X_{oj} \right) \leq B \qquad (6)$$

$$\left(\sum_{f \in F} Y_{fj} + \sum_{o \in O} Y_{oj} \right) \leq 1 \qquad \forall j \in \mathrm{J} \qquad (7)$$

Sixth, the consideration of the endogenous demand function is a key aspect of this formulation. In this initial model, the demand for immunization reduces linearly as distance between immunization center and the population demanding vaccines increases. For instance, the demand that a fixed health center f can capture a population center j, d_{fj}, is given by the multiplication of a vaccination coverage α_{fj} and the total population of region j D_j, as Equation 8 shows. Same logic is applied for outreach sites, as demonstrated in Equation 9. In Equations 10 and 11, we presented how the vaccination coverage is calculated for a given distance. In this toy model, the vaccination coverage reduction is linear, with a decrease of 20% every 100 km. This is the same for outreach or fixed health centers.

$$d_{fj} = D_j \cdot \alpha_{fj} \qquad \forall j \in \mathrm{J}, \forall f \in \mathrm{F} \qquad (8)$$

$$d_{oj} = D_j \cdot \alpha_{oj} \qquad \forall j \in \mathrm{J}, \forall \mathrm{o} \in \mathrm{O} \qquad (9)$$

$$\alpha_{fj} = 100\% - 20\% \cdot \frac{Dist_{fj}}{100} \qquad \forall j \in \mathrm{J}, \forall f \in \mathrm{F} \qquad (10)$$

$$\alpha_{Oj} = 100\% - 20\% \cdot \frac{Dist_{fj}}{100} \qquad \forall j \in \mathrm{J}, \forall o \in \mathrm{O} \qquad (11)$$

For this toy model exercise, we created a fictitious country with 10 discrete population centers, with a fictitious amount of demand for vaccines in each of this center.

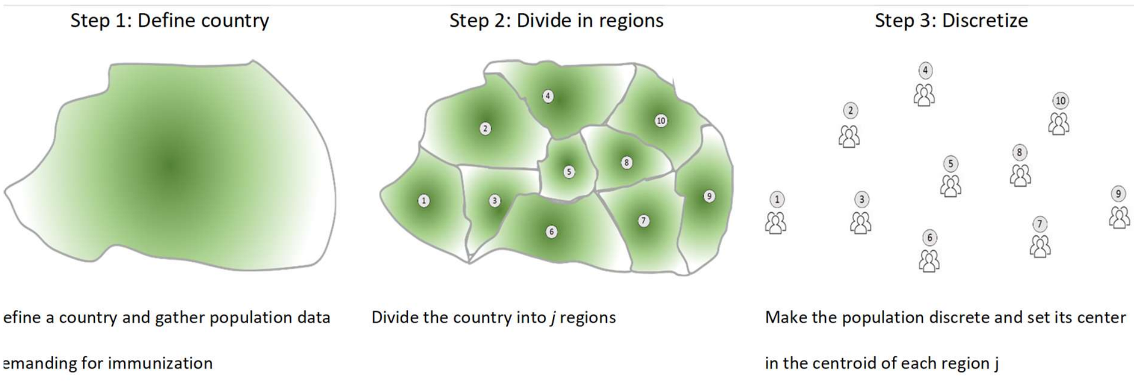


Figure 61 represents how this demand can be derived based on the population data from a country.

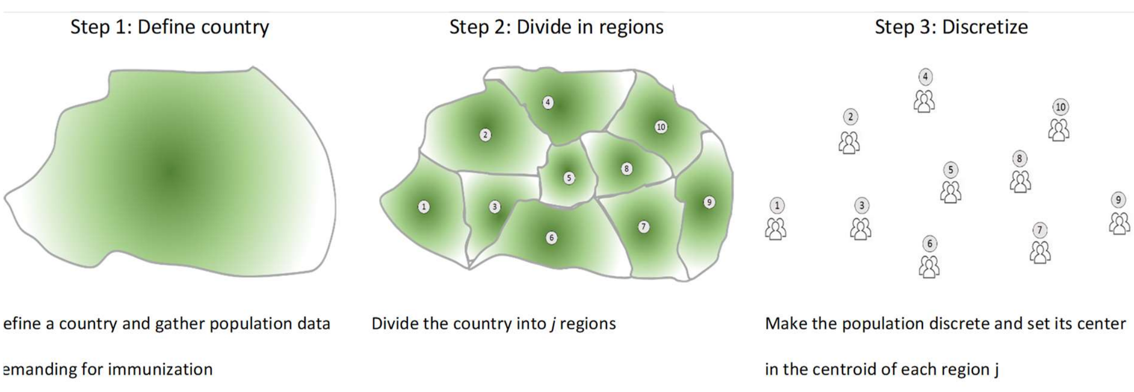


Figure 61: Example of population sets creation

In summary, the main characteristics of the population and facilities presented in this section are as follows:

- Fixed health centers quantity and location are an input of the model and will not change
- Each region either has either a fixed health center or a potential outreach site, not both
- Every region without fixed health center will be a potential place for outreach sites
- The population is discrete and located in the centroid of each region
- The location of the potential outreach site of region *j* is also at the centroid of the area

Appendix B: Toy model validation

In this section, we tested the toy model formulated in the previous section. The main objective of this model is to play with the input variables in order to validate if the model behaves as expected. In addition to validating the model behavior, this tool was used to show the model characteristics to stakeholders in order to align expectations and potentially identify new constraints or decision variables that could be incorporated into the final model.

To test the model, we first drew a baseline scenario by building a small-scale network. In the following sections, we analyzed different scenarios where only one main input variable was altered. We then checked how the outputs from the model in the four different scenarios differed from the baseline and conducted a qualitative analysis of the result. Finally, we summarize the findings of the different scenarios. Table 13 summarizes the three input variables and outputs that will be evaluated in the toy model validation.

Table 13: Summary of inputs and outputs used in toy model validation

Main Inputs	Outputs
Cost of setting up outreach facility c_o	Total fixed health center employees $\sum Ef$
Fixed health center employee cost c_e	Number of outreach facilities $\sum Z_o$
Distance effect on demand $\sum \alpha_{fj}, \alpha_{oj}$	Total coverage

Scenario 1: Baseline

As we previously explained, the network used in the toy model had nine different regions, and consequently nine different populations to be served j=9. Two out of nine regions have fixed health center, F=2, and set of potential outreach sites *o* comprises 7 (9-2) regions. Figure 62 illustrates the network structure of the toy model.

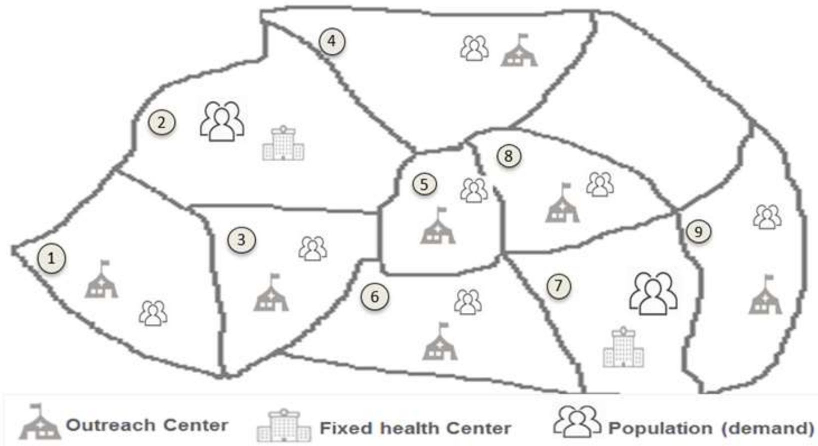


Figure 62: Toy model network structure

All the input parameters were set to represent reasonable values. The values used are not reflected here because they have no influence in our objective with this exercise. We are not interested in the absolute output values in the different scenarios, but instead in the relative changes in the result.

The optimal baseline network that resulted from the model optimization is shown in Table 14 and Figure 4Figure 63. As we can see from the table and the figure, in the optimal solution, five out of seven potential outreach sites are opened, and the two regions that do not have outreach sites are fulfilled respectively by outreach sites in the nearby regions.

Table 14: Toy model baseline outputs

Output	Scenario 1 Baseline
Total fixed health center employees ∑Ef	45
Number of outreach facilities -∑Zo	5
Total coverage	97%

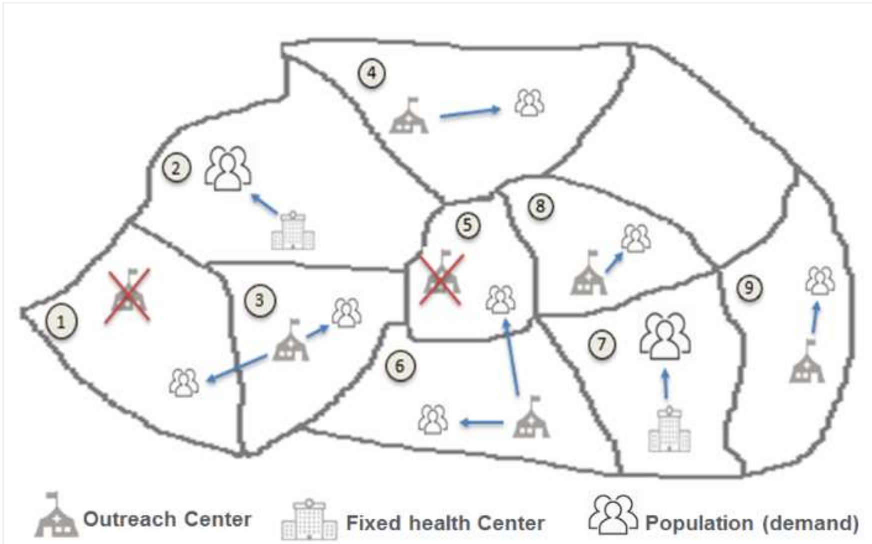


Figure 63: Toy model Scenario 1 optimal network

Scenario 2: High outreach cost

In Scenario 2, the outreach implementation cost - c_o was increased by 100% compared with baseline scenario. Due to the increase in the outreach implementation costs, it was expected that the number of open outreach sites would be reduced and that more employees would be hired in order to attend the regions that are no longer serviced by the outreach sites. The reduction in the number of outreach sites would increase the average distance between population and immunization facilities, leading to a reduction in the immunization coverage.

As we can observe from Table 15, all the expected behaviors for the output variables were obtained. With regard to the increase in the importance of fixed health centers, the optimal solution showed a 40% increase in the number of fixed health center employees. This is a consequence of the fact that a higher share of the regions is now fulfilled by the fixed health centers, since there was a reduction in the outreach sites opened, from five to only one. Finally, we can also see that the expected reduction in coverage was obtained, with a decrease of 9%. The optimal network configuration of Scenario 2 is shown in Figure 64.

Table 15: Comparison of toy model scenarios outputs - baseline and high outreach costs

Output	Scenario 1 Baseline	Scenario 2 High outreach cost
Total fixed health center employees ∑Ef	45	63 (+ 40%)
Number of outreach facilities -∑Zo	5	1
Total coverage	97%	89%

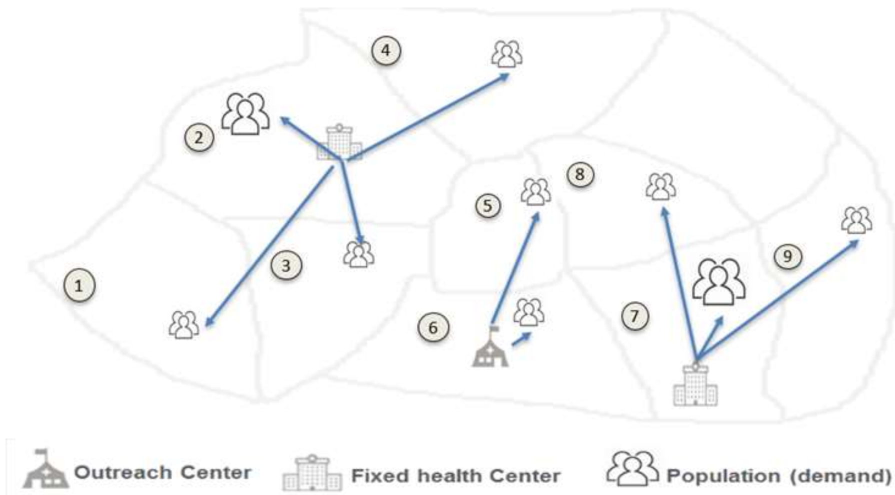


Figure 64: Toy model Scenario 2 optimal network

Scenario 3: High fixed health center employee cost

In Scenario 3 the employee costs increased by 150%. With the increase in the fixed health center cost, we should see the opposite behavior compared to Scenario 2, with a decrease in the participation of the fixed health center in the optimal solution. This effect could be quantified by a reduction in the number of employees at fixed health centers. The increase in costs should also be reflected in a reduction of coverage.

As shown in Table 16, compared to the baseline, the expected decrease in the number of employees was obtained, with a 55% reduction. It is worthwhile noting that, as some additional tests showed, the opening of one of the fixed health centers was only presented in the final solution due to the minimum health center capacity. The expected reduction in immunization demand was also observed with a 20% decrease. The optimal network configuration of Scenario 2 is shown in Figure 65.

Table 16: Comparison of toy model scenarios outputs - baseline and high fixed health center employee costs

Output	Scenario 1 Baseline	Scenario 3 High fixed health center employee cost
Total fixed health center employees ∑Ef	45	24
Number of outreach facilities -∑Zo	5	5
Total coverage	97%	77%

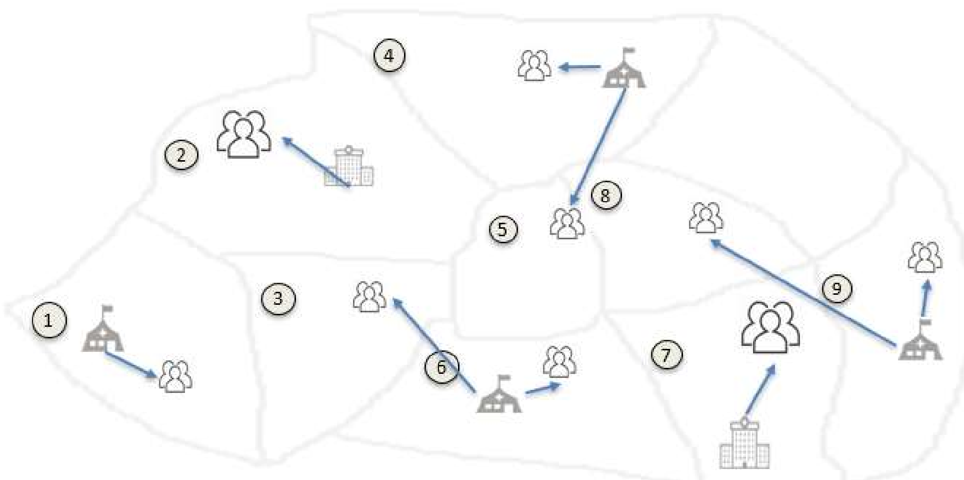


Figure 65: Toy model Scenario 3 optimal network

Scenario 4: Increase effect of distance in demand

The demand for immunization reduces by a linear factor as vaccination centers' distance to population increases. The magnitude of the decline is given by the demand function. In Scenario 4, we evaluated how the model behaves with a change in the demand function parameters. It was expected that increasing the distance effect on demand would result in optimal networks that have more facilities. It is because this would reduce the average distance of population to immunization facilities, increasing the captured demand and potentially the coverage. To validate this expected behavior, we doubled the original distance effect on demand of a 20% decrease every 100 km to a 40% decrease every 100 km. Since the baseline scenario already has many facilities opened, instead of comparing the network of Scenario 4 with the baseline, we compared it with Scenario 2 which had only one outreach site opened. All the other parameters apart from the demand function remained the same to be consistent with the high outreach cost of Scenario 2.

As can be seen in Table 17, the expected behavior of an increase in the number of facilities was achieved, since the number of outreach sites of the optimal solution increase from one in Scenario 2 to two in Scenario 4. Also, the total coverage was reduced by 6%.

Table 17: Comparison of toy model scenarios outputs - baseline and high fixed health canter employee costs

Output	Scenario 2 High outreach cost	Scenario 4 Increase effect of distance in demand
Total fixed health center employees $\sum Ef$	63	54
Number of outreach facilities -$\sum Zo$	1	2
Total coverage	89%	83%