CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

A Set-Partitioning Formulation for Community Healthcare Network Design in Underserved Areas

Marilène Cherkesly
Marie-Ève Rancourt
Karen R. Smilowitz

May 2017

CIRRELT-2017-24

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

A Set-Partitioning Formulation for Community Healthcare Network Design in Underserved Areas

Marilène Cherkesly[1,*], Marie-Ève Rancourt[2], Karen R. Smilowitz[3]

[1] Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

[2] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Logistics and Operations Management, HEC Montréal, 3000 chemin de la Côte, Sainte-Catherine, Montréal, Canada H3T 2A7

[3] Department of Industrial Engineering & Management Science, Northwestern University, Technological Institute D239, 2145 Sheridan Road, Evanston, IL, USA 60208-3119

Abstract. In this paper, we develop tools to assist with the design of a community healthcare network to increase health coverage for underserved areas of Liberia. This study is a collaborative effort with Last Mile Health (LMH), a non-governmental organization that brings healthcare to underserved communities in Liberia. LMH trains community healthcare workers (CHWs) to prevent, diagnose, and treat the most common diseases in the region. CHWs are trained and supervised weekly by a community healthcare worker leader (CHWL). We introduce a variant of the location-routing problem to determine the number of CHWs and CHWLs, as well as the routing and scheduling of the CHWLs by taking into account LMH's operational constraints. We formulate the problem with a set-partitioning formulation with cycle variables. Because the number of variables is large, we propose an approach to generate only non-dominated variables and to break the symmetry between variables without loosing optimality. Computational results are presented for three districts of Liberia along with a discussion of the implementation progress of LMH for the recruitment of CHWs based on our recommendations.

Keywords. Mathematical models, humanitarian logistics, location-routing problems, branch-and-bound.

Acknowledgements. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the scholarship PDF-471388-2015. This support is gratefully acknowledged. We also thank Ross Feehan and Mallika Raghavan for their on-site input.

---

* Corresponding author: cherkesly.marilene@uqam.ca

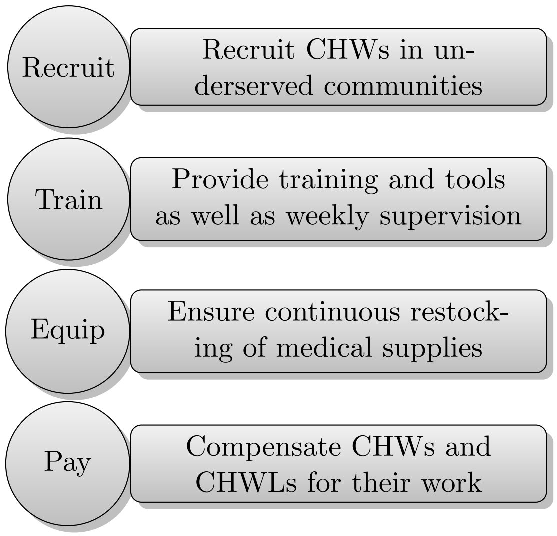
Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2017

1 Introduction

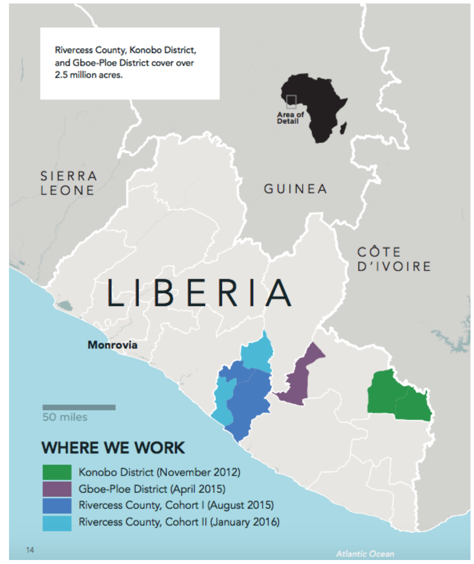
Millions of people around the world lack access to medical care. Traditional hospital-based services fail to reach those living in underserved areas. The World Health Organization (WHO) refers to underserved areas as "remote and rural areas, small or remote islands, urban slums, conflict and post-conflict zones, refugee camps, minority and indigenous communities, and any place that has been severely affected by a major natural or man-made disaster" (World Health Organization, 2010). This study was motivated by a collaboration research project with Last Mile Health (LMH), a non-governmental organization that provides health services to underserved communities in Liberia. In Liberia, in 2003, after 14 years of civil war, the Ministry of Health and Social Welfare partnered with donors and non-governmental organizations to rebuild the health system. According to a study conducted by Kruk et al. (2010), 60% of Nimba's county population lived within a two-hour walk of a health facility, which meant taking a full day to go to the clinic (two-hour walk each way and waiting time at the clinic). In addition, some health facilities are difficult to reach because of the road conditions and lack of accessible transportation options. In the National Health and Social Welfare Policy and Plan 2011-2021 produced by Liberia's Ministry of Health and Social Welfare (2011) one of the priorities is to make healthcare available to underserved communities at a cost affordable for the country. Carefully designing a healthcare network can help reduce the costs while providing a good health service level.

LMH's goal is to bring healthcare to the doorsteps of the underserved population. Figure 1 illustrates LMH's community healthcare model and LMH's current coverage area of Liberia. LMH recruits community health workers (CHWs) directly in underserved communities of Liberia and gives them necessary training, equipment, and support to become health workers. CHWs are trained to prevent, diagnose, and treat the most common health conditions in Liberia (e.g., malaria, tuberculosis, and pneumonia). CHWs are trained and supervised by Community Health Worker Leaders (CHWLs). Daily, CHWs walk towards communities to do routine visits, and, once a week, CHWs are trained by CHWLs. CHWLs ensure continuous one-on-one weekly training and bring medical supplies to CHWs. CHWLs have access to motorbikes to conduct their supervision visits. The design of regional healthcare networks for LMH consists of determining the density and location of CHWs as well as the density, the location and the supervision routes of the CHWLs.

This network design problem can be modeled as a location-routing covering problem (LRCP), a variant of the location-routing problem (LRP) and the covering tour problem (CTP). In the LRP, a set of facilities is available with a fixed opening cost and the problem consists of determining the set of opened facilities such that each customer is served at min-



(a) Last Mile Health's community healthcare model



(b) Where LMH works

Figure 1: LMH's community healthcare model and coverage area

imal cost. In LMH's context, opened facilities can be represented by CHWLs and customers can be represented by CHWs. The CTP consists of determining a least-cost feasible tour starting and ending from a known depot such that a subset of customers is visited and all unvisited customers are within a coverage radius of a visited customer. In LMH's context, the depot can be represented by a CHWL location, visited customers can be represented by CHWs, and unvisited customers can be represented by covered communities. Figure 2 illustrates the relationships among the problems.

In the LRCP emerging in LMH's context, we have a set of underserved communities in need of healthcare access and each community has an estimated population. CHWs and

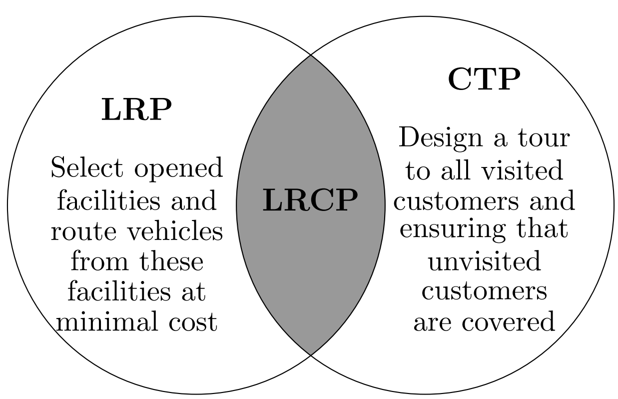


Figure 2: Relationship between the LRCP, the LRP, and the CTP

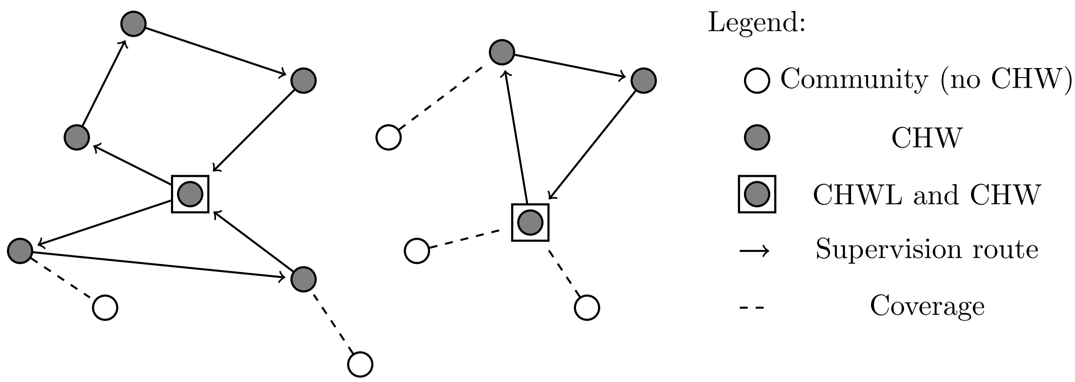


Figure 3: Example of a LRCP solution

CHWLs need to be located in those communities and are paid a fixed weekly salary, where each CHW is associated with a maximal coverage radius and with a maximal population coverage, and each CHWL has a maximum number of working hours per day and a maximum number of working days per week. The LRCP consists of determining the location and density of CHWs as well as the location, density and supervision routes of CHWLs.

Figure 3 illustrates an example of a LRCP solution, where each CHW is represented by a grey node, communities without CHWs are represented by a white node, and each CHWL is represented by a framed node. Each community without a CHW is covered by a CHW (dashed lines) and the supervision routes start and end from CHWL locations while visiting a subset of CHWs (black arrows). Note that, if there is more than one CHW in a community, the number of supervision visits will correspond to the number of CHWs and that these supervision visits can be conducted on different days and by a different CHWL. Figure 4 illustrates an example of a solution for LMH on Liberia's map.

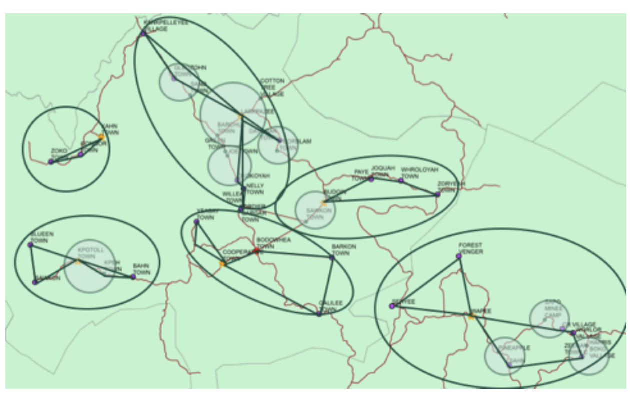


Figure 4: Example of a solution for LMH

1.1 Contributions and organization of this paper

The main objective is to propose tools to assist with the design of a community healthcare network in underserved areas. In terms of scientific contributions, a new variant of the LRP, combining the LRP and the CTP, is introduced. This variant emerges from a real-life application in healthcare supply chains in underserved areas. We formulate the problem with an integrated mathematical model. We provide strategies to reduce reduce the number of variables and implement those strategies to ease solving the problem. We discuss the quality of the data and analyze the impact of the different parameter values (e.g. maximum route distance and maximum population coverage) on the cost of the solution while delivering good quality health services.

This paper also has practical contributions. As pointed out by several experts in humanitarian aid (see Altay and Green, 2006; Van Wassenhove, 2006; Pedraza-Martinez et al., 2011, 2013; Rancourt et al., 2015; Von Achen et al., 2016), relevant and impactful methodological developments are created when the humanitarian context is properly understood. Thus, before modeling and solving our problem, we have first analyzed LMH's cost structure. In addition, data collection is often a challenge in developing countries like Liberia. In our context, one problem was the lack of real data. To ease the process, LMH conducted field data collection with geographical information system (GIS) specialists. Once all the data was gathered, we worked in collaboration with GIS specialists to fix some issues with the data such as disconnected network. In addition, the mathematical model developed has limited manual intervention. To increase the likelihood of successful adoption, we exploit dominance in the decision variables to maintain a reasonable problem size. Finally, we try to bridge the gap between practice and academia for the design of community healthcare network in underserved areas by providing optimal solutions that can be implemented in practice.

The remainder of the paper is organized as follows. Section 2 provides a literature review of the state-of-the-art algorithms proposed to solve the LRP and the CTP, and highlights the particularities of our problem. Section 3 proposes a set-partitioning formulation for the LRCP. Because the number of variables in a set-partitioning formulation is generally large, Section 4 presents strategies to decrease the number of variables while ensuring an optimal solution. Computational results are presented in Section 5 and conclusions are drawn in Section 6.

2 Literature Review

This work is an extension of Von Achen et al. (2016). These authors have proposed a two-step heuristic to optimize community healthcare coverage for Liberia's underserved areas in collaboration with LMH. The first stage consists of determining the number of CHWs per community and their catchment area. The second stage consists of first enumerating all feasible cycles respecting LMH's operational constraints, and then solving a LRP to visit exactly once each community containing at least one CHW. The authors show that in Liberia, using Euclidean distances often yields infeasible solutions. In our work, we integrate these two decisions and ensure that, if a community contains more than one CHW, each CHW will be visited exactly once.

To the best of our knowledge, the LRCP has not been previously studied. Sections 2.1 and 2.2 present algorithms proposed in the literature for the LRP and the CTP, respectively.

2.1 The location-routing problem

The LRP has been widely studied and several reviews on variants and applications of the LRP have been published (Prodhon and Prins, 2014; Albareda-Sambola, 2015; Drexl and Schneider, 2015). Contardo et al. (2013) proposed four different flow formulations for the capacitated LRP: a two-index vehicle flow formulation, a three-index vehicle flow formulation, a two-index capacity flow formulation, and a three-index commodity flow formulation. For each formulation, they developed a branch-and-cut algorithm and have developed new valid inequalities for the problem. On their set of artificial instances, compact formulations (two-index formulations), produce smaller optimality gaps. In addition, they show that, in general, branch-and-price algorithms (Baldacci et al., 2011; Belenguer et al., 2011; Contardo et al., 2014) outperform their branch-and-cut.

Yi and Özdamar (2007) modified the classical LRP for logistics support and evacuation operations in disaster response activities. In their problem, two types of commodities need to be transported: emergency supplies and wounded people. For each type of commodity, different priority levels are given according to the type of supply needed or the injuries. Their objective consists in minimizing the delay for the arrival of supplies at aid centers and the delay in provided healthcare to the injured people. They propose a two-stage algorithm to solve this problem.

2.2 The covering tour problem

Covering location problems have been widely studied in the literature (see Farahani et al., 2012; García and Marín, 2015, for surveys). More recently, attention has been given to integrating covering location problems to routing problems. In this section, we first highlight the algorithms developed for the CTP and the CSP, and, then, we present contributions related to real-life humanitarian problems.

The covering salesman problem (CSP) was introduced by Current (1981) and was formulated in Current and Schilling (1989). In the CSP, all customers must be covered and there is no specific set of customers that must be visited. Current and Schilling (1994) later proposed two variants of the CSP: the median tour problem and the maximal covering tour problem. In both problems, they propose a two-objective formulation which consists of minimizing the sum of the total routing costs and of maximizing accessibility. In the median tour problem, accessibility is defined as the total demand of unvisited customers multiplied by the travel distance to their nearest visited customer. In the maximal covering tour problem, accessibility is defined as the total demand of customers who are not within a maximal coverage radius of a visited customer. Gendreau et al. (1997) developed the first exact algorithm for the CTP, a variant of the CSP. In the CTP, the set of customers is divided in three sets: a set of customers that must be visited, a set of customers that can be visited and a set of customers that must be covered. This differs from the CSP where there is no specific set of customers that need to be visited. The authors implemented a branch-and-cut algorithm and introduced several families of valid inequalities for the problem. Hachicha et al. (2000) introduced and formulated the multi-vehicle CTP. Three heuristics were implemented to solve this problem; the first one based on a savings criterion, the second one based on a sweep algorithm, and the third one based on a route-first, cluster-second algorithm. Doerner et al. (2007) have extended the model proposed by Hodgson et al. (1998) and Hachicha et al. (2000) to multiple objectives: effectiveness of workforce employment, average accessibility, and coverage. To find Pareto-efficient solutions, three heuristics are proposed and implemented. Jozefowiez et al. (2007) proposed a matheuristic algorithm which combines an evolutionary algorithm with the branch-and-cut algorithm introduced by Gendreau et al. (1997) to solve a bi-objective CTP. The first objective consists of minimizing the total routing cost while the second objective consists of maximizing the total coverage. Golden et al. (2012) proposed three variants of the generalized CSP where each customer needs to be covered a fixed number of times. In the first variant, a tour can visit each customer at most once and the remainder of the demand must be covered by other visited customers. In the second variant, a customer can be visited more than once, but the same customer cannot be

Table 1: Summary of the CTPs treated in the literature

Authors	Facility location	Multiple vehicles	Mutiple visits	Min. distance	Other objectives	Exact algo.
Current and Schilling (1989)				•		
Current and Schilling (1994)				•	•	
Gendreau et al. (1997)				•		•
Hodgson et al. (1998)				•	•	•
Hachicha et al. (2000)		•		•		
Doerner et al. (2007)					•	
Jozefowiez et al. (2007)				•	•	
Nolz et al. (2010)		•		•	•	
Golden et al. (2012)			•	•	•	
Naji-Azimi et al. (2012)		•		•		
Our contribution	•	•	•	•	•	•

visited consecutively. In the third variant, a customer can be visited more than once, and the visits can be done consecutively. Two local search heuristics are implemented to solve each of the variants.

In addition to the previous articles, many applications of the CTP are for healthcare problems arising in the context of humanitarian logistics. We highlight three contributions. Hodgson et al. (1998) adapted the CTP to a case study for primary healthcare in Suhum District, Ghana. In this problem, healthcare facilities are located in communities and communities without a healthcare facility must walk towards the nearest health facility. The authors developed a multi-objective model which aimed at minimizing the total travel cost of the tour, at maximizing the number of covered communities and at maximizing the total population covered by the tour. Nolz et al. (2010) developed a metaheuristic for a bi-objective CTP. Their work is in collaboration with the Austria Red Cross and is applied to the delivery of water to the people living in affected areas after a natural disaster such as an earthquake, a flood, or a tsunami. Their first objective consists of minimizing the sum of the distances between the population and the nearest facility and the total population outside the coverage radius. The second objective aims at minimizing the routing and can take two forms: (1) minimizing the tour length or (2) minimizing the latest arrival at each customer. Naji-Azimi et al. (2012) have adapted the CTP to the distribution of survival goods in disaster areas and propose a multi-start heuristic to solve this problem. In their problem, one central depot is open and multiple satellite distribution centers can be used. Each victim must travel from home to a satellite distribution center, and each victim must be within a covering distance of a satellite distribution center. Multiple types of items can be sent to the victims and each victim has a demand for each item type. Their problem consists of selecting the location of the satellite distribution centers, and determining how to supply them for the central depot using the available fleet at its best.

Table 1 presents the different characteristics of CTP related articles and shows the contribution of our paper. One of the biggest difference between our problem and the different problems studied in the literature is that we integrate facility location decisions. In the CTP literature, there is only one facility and its location is known. In our context, we have multiple potential facilities locations (CHWL location) and we need to determine which facilities will be opened. Unlike multiple problems, we also allow multiple visits to each community, i.e., one per CHW, and we propose an exact algorithm to solve the problem.

3 Mathematical formulation

The network is defined on a directed graph $G(N, A)$, where N is the set of underserved communities and A is the set of arcs (roads in Liberia's context). Each underserved community $i \in N$ is associated with a known population p_i. Each arc $(i, j) \in A$ is associated with a total distance d_{ij} and a total cost c_{ij} which is a function of the total distance. Note that in the road network of Liberia, $d_{ij} = d_{ji}, \forall (i, j) \in A$.

We propose a compact formulation and assume that all cycles can be generated. As in Von Achen et al. (2016), the supervision routes are modeled with cycle variables. Figure 5 illustrates a cycle visiting communities $i, j, k \in N, i \neq j \neq k$. A cycle is represented by a set of communities and a set of edges. In Liberia's context, for each arc $(i, j) \in A$, an edge (i, j) is generated if $i < j$ and its distance will be the same as the one of the arc. This is done because $d_{ij} = d_{ji}, \forall (i, j) \in A$. Let Ω be the set of all feasible cycles. A feasible cycle is a cycle starting and ending in a community, visiting a set of CHWs to conduct one-on-one supervision visits while respecting the maximal cycle distance and the maximal supervision time. Each cycle $c \in \Omega$ is associated with c_c its total cost and with d_c its total distance. Each combination of cycle $c \in \Omega$ and community $i \in N$ is associated with a_{ic} a binary parameter equal to one if cycle c visits community i. Each combination of cycle $c \in \Omega$ and community $i \in N$ is associated with b_{ic} a integer parameter representing the number of supervision visits conducted in community i in cycle c. Section 4 presents the strategies and the detailed algorithm implemented to generate all variables.

Each CHW is associated with a weekly salary α. Due to Liberia's geography, i.e., its remoteness, and due to the fact that CHWs need to walk to conduct healthcare visits, a maximal coverage radius of σ km is imposed. That is, CHWs can only cover communities in their maximal coverage radius to limit the time spent walking. In addition, to ensure

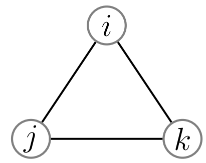


Figure 5: Cycle visiting communities $i, j, k \in N$

healthcare access, LMH has determined that CHWs cannot cover more than ρ people also known as the maximal population coverage. LMH pays its CHWLs a fixed weekly salary β and CHWLs are required to work at most δ days per week to conduct training.

The variables are defined as follows. Let y_i be the number of CHWs located in community $i \in N$ and w_i a binary variable equal to one if a CHWL is located in community $i \in N$. Let x_{ij} be a binary variable equal to one if a CHW in community $i \in N$ covers the population of community $j \in N$; x_{ij} is defined for $\{(i,j) \in A : (d_{ij} \leq \sigma) \wedge (p_j \leq p_i)\}$. This definition imposes that each CHW can only cover towns in the maximal radius σ and that each CHW must cover communities with smaller population than its own community. Let z_c represent the number of times cycle $c \in \Omega$ is used in the solution, and μ_{ic} is an integer variable representing the number of times cycle c starts and ends at community i. The weekly community healthcare network design problem can be modeled as follows.

$$\text{Minimize} \quad \sum_{i \in N} \alpha y_i + \sum_{i \in N} \beta w_i + \sum_{c \in \Omega} c_c z_c \tag{1}$$

$$\text{s.t.} \quad \sum_{i \in N} x_{ij} = 1, \quad \forall j \in N, \tag{2}$$

$$\sum_{j \in N} p_j x_{ij} \leq \rho y_i, \quad \forall i \in N, \tag{3}$$

$$w_i \leq y_i \quad \forall i \in N, \tag{4}$$

$$\sum_{c \in \Omega} b_{ic} z_c \geq y_i, \quad \forall i \in N, \tag{5}$$

$$\mu_{ic} \leq a_{ic} z_c, \quad \forall i \in N, c \in \Omega, \tag{6}$$

$$\sum_{c \in \Omega} \mu_{ic} \leq \delta w_i, \quad \forall i \in N, \tag{7}$$

$$\sum_{i \in N} \mu_{ic} = z_c, \quad \forall c \in \Omega, \tag{8}$$

$$y_i \in \mathbb{N}, \quad \forall i \in N, \tag{9}$$

$$w_i \in \mathbb{B}, \quad \forall i \in N, \tag{10}$$

$$x_{ij} \in \mathbb{B}, \quad \forall i, j \in N, \tag{11}$$

$$z_c \in \mathbb{N}, \qquad \forall c \in \Omega, \tag{12}$$

$$\mu_{ic} \in \mathbb{N}, \qquad \forall i \in N, c \in \Omega. \tag{13}$$

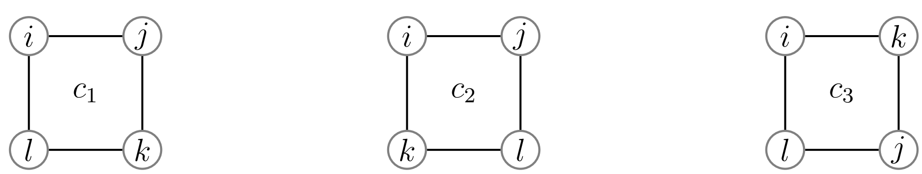
The objective function (1) minimizes the total weekly costs computed as the sum of the weekly CHW and CHWL salaries, and the routing costs. Constraints (2) impose that each community must be covered by exactly one CHW. Constraints (3) impose the maximum population coverage for each CHW. Constraints (4) impose that a CHWL can only be located in a community where there is a CHW. Constraints (5) impose that each CHW must have a weekly one-on-one supervision visit, i.e., if community i contains y_i CHW, then the number of weekly supervision visits in community i have to be at least y_i. Constraints (6) impose that a community can be a CHWL location for a given cycle only if it is visited in that cycle and if the cycle is used in the optimal solution. Constraints (7) impose the maximal number of work days per CHWL during a week. Constraints (8) link the cycle variables to the number of times a cycle starts and ends from one of the visited communities. Constraints (9)–(13) define the set of variables.

4 Cycle generation

In the following section, we propose a methodology to generate all non-dominated cycles. This strategy is necessary to help decrease the number of generated cycles and to solve the model with a MIP solver. We present three strategies to reduce the number of cycles and explain how they are applied. We then present the algorithm to generate all non-dominated variables.

4.1 Strategy 1: distance dominance

For cycles visiting at least four different communities, there exists more than one order to visit all these communities in a cycle. To decrease the number of generated cycles, only the



$d_{ij} + d_{jk} + d_{kl} + d_{li} =$ $1 + 2 + 3 + 4 = 10$ $\quad d_{ij} + d_{jl} + d_{lk} + d_{ki} =$ $1 + 5 + 3 + 6 = 15$ $\quad d_{ik} + d_{kj} + d_{jl} + d_{li} =$ $6 + 2 + 5 + 4 = 17$

Figure 6: All possible cycles visiting four communities $i, j, k, l \in N$

cycle with the smallest distance is kept. Figure 6 presents all possible cycles visiting a set of four communities $i, j, k, l \in N$ such that $i \neq j \neq k \neq l$. The minimal distance cycle is c_1 and it will be the only cycle generated.

4.2 Strategy 2: visit dominance

In our problem, each CHWL can work at most $\mathcal{T}$ hours per day. This work time consists of the driving time and supervision time. Thus, for a given cycle $c \in \Omega$, we can compute its maximal number of supervision visits as

$$\overline{v_c} = \left\lfloor \frac{\mathcal{T} - s * d_c}{\tau} \right\rfloor,$$

where $\overline{v_c}$ is the maximal number of supervision visits in cycle c, s is the travel speed, d_c is the distance of cycle c, and τ is the time needed to conduct one supervision visit. Thus, we can use dominance on the number of supervision visits. For example, let us have a cycle $c \in \Omega$ for which the maximum number of visits is 3 ($\overline{v_c} = 3$). Table 2 presents all ten supervision visit discretization possibilities for a cycle visiting two communities $i, j \in N, i \neq j$. In that case, only four cycles are non-dominated, namely c_4, c_7, c_9, and c_{10}, because for each of these cycles the total number of visits is equal to the maximal number of supervision visits ($\overline{v_c} = 3$). Thus, at most four cycles will be generated. In order to further decrease the number of generated cycles, Section 4.3 details which cycles will be generated.

4.3 Strategy 3: maximal number of CHWs dominance

We can further reduce the number variables by considering the population in each community and an upper bound on the number of CHWs needed in each community. For each community $i \in N$, the total population in its maximal coverage radius σ is computed while ensuring that CHWs can only cover smaller communities. This is divided by the maximal population coverage ρ and serves as an upper bound on the number of CHWs needed in each community.

Table 2: The possible number of supervision visits for a cycle visiting two CHWs and with $\overline{v_c} = 3$

Cycle	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
b_{ic}	0	0	0	0	1	1	1	2	2	3
b_{jc}	0	1	2	3	0	1	2	0	1	0
Total number of visits per cycle	0	1	2	3	1	2	3	2	3	3

Table 3: Generated cycles according to the values of $\overline{y_i}$ and $\overline{y_j}$

Values of $\overline{y_i}$ and $\overline{y_j}$	Generated cycles
$\overline{y_i} + \overline{y_j} \leq \overline{v_c}$	One cycle where $b_{ic} = \overline{y_i}$ and $b_{jc} = \overline{y_j}$
$\overline{y_i} = 1, \overline{y_j} \geq 3$	Two cycles: c_4 and c_7
$\overline{y_i} = 2, \overline{y_j} = 2$	Two cycles: c_7 and c_9
$\overline{y_i} = 2, \overline{y_j} \geq 3$	Three cycles: c_4, c_7, and c_9
$\overline{y_i} \geq 3, \overline{y_j} = 1$	Two cycles: c_9 and c_{10}
$\overline{y_i} \geq 3, \overline{y_j} = 2$	Three cycles: c_7, c_9, and c_{10}
$\overline{y_i} \geq 3, \overline{y_j} \geq 3$	Four cycles: c_4, c_7, c_9, and c_{10}

For a community $i \in N$, this upper bound is computed as follows:

$$\overline{y_i} = \left\lceil \frac{p_i + \sum_{j \in N | (i,j) \in A, (d_{ij} \leq \sigma) \wedge (p_j \leq p_i)} p_j}{\rho} \right\rceil, \forall i \in N.$$

Table 3 presents the generated cycles according to the upper bound on the number of CHWs in communities i and j for the example presented in Section 4.2. We can realize that when when the sum of the maximum number of CHWs is at most equal to the maximal number of supervision visits in a cycle ($\overline{y_i} + \overline{y_j} \leq \overline{v_c}$), then the cycle for which the number of supervision visits in each community is equal to its maximal number of CHWs ($b_{ic} = \overline{y_i}$ and $b_{jc} = \overline{y_j}$) will be non-dominated. Thus, it is unnecessary to generate all other cycles which implies that only one cycle will be generated. If $\overline{y_i} + \overline{y_j} > \overline{v_c}$, more than one cycle needs to be generated and the choice of generated cycles depends on the maximal number of CHWs for each visited community.

4.4 Algorithm

Algorithm 1 details how the set Ω is generated. The algorithm is in three parts. In the first part (lines 11–15), cycles with only one community ($|N(c)| = 1, c \in \Omega$) are created and the number of supervision visits will be equal to the minimum between the upper bound on the number of CHWs for that community ($\overline{y_i}, i \in N(c)$) or the maximum number of supervision visits in the cycle ($\overline{v_c}$). In the second part, (lines 17–21), cycles with more than one community such that the sum of the upper bounds of the number of CHWs is at most the maximum number of supervision visits ($\sum_{i \in N(c)} \overline{y_i} \leq \overline{v_c}$) are created. For those cycles, we set the number of supervision visits in each community as the upper bound on the number of CHWs in that community ($b_{ic} = \overline{y_i}, \forall i \in N(c)$). Finally, in the third part, (lines 23–35), cycles with more than one stop such that the sum of the upper bounds of the

Algorithm 1 Generation of the set Ω

Considering all feasible minimal distance cycles $C = \{c_1, c_2, ...\}$ where each cycle $c = (i, j, ..., k), c \in C$ is a cycle visiting customers $i, j, ..., k \in N$
Define $N(c) = \{i, j, ..., k\}$ as the set of nodes in cycle $c \in C$
Define $A(c) = \{(i, j), (j, ...)..., (..., k), (k, i)\}$ as the set of arcs in cycle $c \in C$
Define β as the cost of traveling one unit of distance
Initialize $\Omega \leftarrow \emptyset$, $\omega \leftarrow \emptyset$
for $c \in C$ **do**
    $d_c \leftarrow \sum_{(i,j)\in A(c)} d_{ij}$
    $c_c \leftarrow \beta d_c$
    $a_{ic} \leftarrow 1, \forall i \in N(c)$
    $a_{ic} \leftarrow 0, \forall i \in N \backslash N(c)$
    if $|N(c)| = 1$ **then**
        $b_{ic} \leftarrow \min\{\overline{y_i}, \overline{v_c}\}, \forall i \in N(c)$
        $b_{ic} \leftarrow 0, \forall i \in N \backslash N(c)$
        $\omega \leftarrow \{c\}$
        $\Omega \leftarrow \Omega \cup \omega$
    else
        if $\sum_{i\in N(c)} \overline{y_i} \leq \overline{v_c}$ **then**
            $b_{ic} \leftarrow \overline{y_i}, \forall i \in N(c)$
            $b_{ic} \leftarrow 0, \forall i \in N \backslash N(c)$
            $\omega \leftarrow \{c\}$
            $\Omega \leftarrow \Omega \cup \omega$
        else
            for $i \in N(c)$ **do**
                for $b_{ic} = \min\left\{\overline{y_i}, \left\lfloor \frac{\overline{v_c}}{|N(c)|} \right\rfloor\right\}$ s.t. $b_{ic} \leq \min\{\overline{y_i}, \overline{v_c} - |N(c)| + 2\}$ **do**
                    Let Φ be the set of all possible values of $b_{jc}, \forall j \in N(c)\backslash\{i\}$ such that:
                    (1) $|\{j \in N(c)|b_{jc} = 0\}|\ 1 \leq 1$;
                    (2) $b_{jc} \leq \overline{y_j}$, $\forall j \in N(c)\backslash\{i\}$;
                    (3) $\sum_{j\in N(c)} b_{jc} = \overline{v_c}$.
                    for $\phi \in \Phi$ **do**
                        $b_{ic} \leftarrow b_{ic}$
                        $b_{jc} \leftarrow b_{jc}, \forall j \in N(c)\backslash\{i\}$
                        $b_{jc} \leftarrow 0, \forall j \in N \backslash N(c)$
                        $\omega_\phi \leftarrow \{c\}$
                        $\Omega \leftarrow \Omega \cup \omega_\phi$
                        $b_{ic} \leftarrow b_{ic} + 1$
                    end for
                end for
            end for
        end if
    end if
end for

number of CHWs is greater than the maximum number of supervision visits in the cycle ($\sum_{i \in N(c)} \overline{y_i} > \overline{v_c}$) are created. For those cycles, multiple feasible visit cycles will be created. Our algorithm ensures that (i) for each of those visit feasible cycles created, at most one node will have no supervision visit (line 26), (ii) that the number of supervision visits at each node does not exceed its upper bound on the number of CHWs (line 27), and (iii) that the total number of supervision visits on that cycle will be exactly the maximum number of supervision visits of the cycle (line 28). Our algorithm also ensures that each generated cycle will not be dominated or equivalent to another cycle. Thus, each cycle is unique and non-dominated.

5 Results

Our cycle generation algorithm was implemented in C++ and our mathematical model was solved with CPLEX. All tests were performed on a Linux computer equipped with an Intel(R) Core(TM) i7-3770 processor (3.4 GHz). In order to find solutions in a reasonable amount of time, we allow a time limit of one hour and we report the results only when the instance was solved within the prescribed time limit. This time limit does not include the time needed to generate cycles as, for all tested instances, this took less than a second. In this section, we present the characteristics of LMH's instances. We also present the impact of using the our exact mathematical model with the two-step heuristic of Von Achen et al. (2016). Finally, we conduct sensitivity analysis on key parameters.

5.1 Data collection and characteristics of the instances

This problem was solved for real-life instances of Liberia. As explained by Von Achen et al. (2016), using Euclidean distances in the context of Liberia often yields infeasible solutions. In fact, the road network is sparse. In order to provide feasible solutions, we have used the real road network distances. To obtain this real road network, LMH sent a team of enumerators on the field to conduct GIS data collection. For each district of Liberia, this process can take several months. Once this data was collected, three issues were encountered: (i) some communities were disconnected from the roads, (ii) some roads that needed to be connected were not always connected, and (iii) some road segments were repeated. With the help of a GIS specialist, we were able to fix these problems. Finally, the real road network distance matrix was extracted by using tools in ArcGIS.

The problem was solved for three districts of Liberia where LMH is currently expanding its operations. For confidentiality reasons, these districts are named Districts 1, 2 and 3.

Table 4: Descriptive statistics per district

District	$\|N\|$	Underserved (%)	$\overline{\text{Pop.}}$	St. dev. population	$\overline{d_{ij}}$	max d_{ij}	St. dev. d_{ij}
1	58	74.4	166.7	255.3	49.8	186.3	34.5
2	34	70.6	189.4	245.5	25.9	68.3	16.2
3	41	70.7	80.5	57.6	27.4	129.0	27.9

Table 5: Average proportion of communities (in %) within a specific radius of another community

	Radius (in km)								
District	2.5	5	10	15	20	25	50	100	200
1	3.27	4.99	8.15	13.32	19.56	25.98	57.97	93.70	100.00
2	6.60	9.72	19.79	29.86	37.85	49.65	91.67	100.00	100.00
3	5.89	11.72	24.57	37.18	48.96	61.09	88.70	94.53	100.00

Table 4 presents descriptive statistics. In Liberia's healthcare delivery context, a community is underserved if its closest health facility is located at more than five kilometers. We present the following information: the number of underserved communities ($|N|$); the percentage of underserved communities computed as $\frac{|N|}{\#\text{ communities}}$, where $|N|$ is the number of underserved communities per district and # communities is the total number of communities per district (*Underserved %*); the average population per underserved community ($\overline{Pop.}$); the standard deviation on the population per underserved community (*St. dev. population*); the average distance in kilometers between two communities ($\overline{d_{ij}}$); the maximal distance in kilometers between two communities (*max* d_{ij}); the standard deviation on the distance in kilometers between two communities (*St. dev.* d_{ij}). Table 5 presents additional characteristics for each district. It presents the average proportion of communities (in percentage) within a specific radius of another community.

By taking a closer look at both tables, we can realize that, on average, more than 70% of the number of communities in Liberia are further than five kilometers from their nearest health facility. We can also determine that District 2 is the smallest district and District 1 is the biggest. District 1 is the district with the largest communities, while District 3 has the smallest communities. In addition, District 3 seems to be the densest and, District 1, the sparsest.

Considering LMH's reality and Liberia's geography, in LMH's basic model, each CHW has a maximal coverage radius of 2.50 km, i.e., $\sigma = 2.50$. In practice, in a normal work day, CHWs walk towards one community to conduct healthcare visits and walk back to their

Table 6: Number of cycles with our variable generation technique

District	D	$\|\Omega\|$	$\overline{\|\Omega\|}$	$\downarrow$ (%)
1	30	1,943	13,643	85.8
1	40	5,405	39,487	86.3
1	50	11,916	90,566	86.8
1	60	23,395	149,246	84.3
2	30	702	4,994	85.9
2	40	1,348	13,609	90.1
2	50	3,040	25,922	88.3
2	60	6,308	25,922	75.7
3	30	10,555	57,606	81.7
3	40	24,908	100,596	75.2
3	50	49,128	123,136	60.1
3	60	97,313	135,720	28.3

base after their day. In Liberia, the average walking speed is 5 km per hour and can be lower because of the geography of the region. Thus, allowing a maximal coverage radius of 2.50 km to conduct healthcare visit implies that each CHW can walk at most 30 minutes each way (one hour per day), which is reasonable in practice. In LMH's basic model, the maximal population coverage is 250, i.e., $\rho = 250$. In fact, in LMH's context, CHWs are part-time workers and cannot visit as many communities as full-time workers. In addition, communities are small and CHWs would have to cover communities quite far away if ρ would be larger. In addition, each CHWL can work at most eight hours per day, i.e., $\mathcal{T} = 8$, and each supervision visit lasts two hours, i.e., $\tau = 2$. The maximal number of working days per week for each CHWL has been set to four, i.e., $\delta = 4$. In fact, CHWLs spend four days a week to conduct supervision visits and their fifth day of work is spent to get additional training at health facilities. Finally, LMH wanted to test different values for the maximal distance of the CHWL cycle denoted with D. In Liberia, the average motorbike speed is of 30 km per hour. Thus, in their basic model, they wanted to have a maximal cycle distance of 30 km, i.e., $D = 30$, which means that each CHWL could be driving at most one hour per day. With further discussions, we have determined that this limit could be increased to a maximal cycle distance of 60 km. Thus, we have tested our instances for values of $D = \{30, 40, 50, 60\}$.

Table 7: Summary of the obtained results with the exact model

District	D	CHWs	CHWLs	z^*
1	30	60	11	19.18
1	40	59	8	22.93
2	30	28	4	8.57
2	40	28	3	9.04
2	50	28	3	9.04
3	30	24	4	9.51
3	40	24	4	8.83

5.2 Number of generated cycles

Because the proposed set-partitioning formulation needs all variables to be generated, we first analyze the impact of our variable generation algorithm. Table 6 compares the impact of generating cycles using our variable generation technique with all possible feasible cycles. We present the following information: the name of the district (*District*); the maximal cycle distance (in km) (D); the number of generated cycles with our variable generation technique ($|\Omega|$); the number of generated cycles without our variable generation technique ($\overline{|\Omega|}$); and the reduction in percentage of cycles generated with our variable generation technique computed as $1-|\Omega|/\overline{|\Omega|}$ ($\downarrow$). On average, we are able to reduce the number of generated cycles by 86%, 85% and 62% of cycles for Districts 1, 2, and 3, respectively. Thus, our variable generation technique is useful to reduce the number of variables.

5.3 Results with the exact model

Table 7 presents a summary of the results obtained with the exact model. We present the following information: the name of the district (*District*); the maximal cycle distance in km (D); the number of CHWs in the solution (*CHWs*); the number of CHWLs in the solution (*CHWLs*); and the total routing costs in the solution (z^*). The results for District 1 with $D \geq 50$, District 2 with $D \geq 60$, and District 3 with $D \geq 50$ are not reported because the total computation time exceeded the one-hour limit.

By looking at the results, one can realize that increasing the maximal cycle distance has almost no impact in decreasing the number of CHWs in the final solution. On the other hand, the number of CHWLs decreases. This is drastic, in particular for District 1, where increasing the maximal distance from 30 km to 40 km decreases the number of CHWLs from 11 to 8. Thus, even if increasing the maximal distance makes the problem harder to solve because the number of generated cycles becomes larger, the impact on the number of

Table 8: Impact of the two-step heuristic on solution attributes compared with our exact model

District	D	$\Delta\lvert\Omega\rvert$ (%)	Δ CHWs (%)	Δ CHWLs (%)	Δz (%)
1	30	12.5	–1.67	9.09	–7.90
1	40	13.3	0.00	0.00	1.50
2	30	22.7	0.00	0.00	0.00
2	40	20.9	0.00	0.00	0.00
2	50	18.7	0.00	0.00	0.00
3	30	4.8	0.00	0.00	0.68
3	40	6.3	0.00	0.00	1.30

CHWLs is important at it represents the highest cost for LMH.

5.4 Comparison between the exact model and the two-step heuristic

Table 8 presents summarized results with an adapted version of the two-step heuristic proposed by Von Achen et al. (2016) where training times are considered when generating cycles. We present the following information: the name of the district (*District*); the maximal cycle distance in km (D); the proportion in percentage of cycles generated with the two-step heuristic compared with the exact model $|\Omega|_2/|\Omega|$, where $|\Omega|_2$ and $|\Omega|$ are the number of cycles generated with the two-step heuristic and with the exact model ($\Delta|\Omega|$); the variation in percentage on the optimal number of CHWs computed as $(CHW_2 - CHW^*)/CHW^*$, where CHW_2 and CHW^* are the number of CHWs in the solution obtained with the two-step heuristic and the exact model (Δ *CHWs*); the variation in percentage on the optimal number of CHWLs computed as $(CHWL_2 - CHWL^*)/CHWL^*$, where $CHWL_2$ and $CHWL^*$ are the number of CHWLs in the solution obtained with the two-step heuristic and the exact model (Δ *CHWLs*); and the variation in percentage on the optimal routing costs computed as $(z_2 - z^*)/z^*$, where z_2 and z^* are the routing costs in the solution obtained with the two-step heuristic and the exact model (Δz).

Solving the problem with the two-step heuristic does not yield the optimal solution but provides a heuristic solution. Except for District 1 with $D = 30$, the number of optimal CHWs and optimal CHWLs are obtained with the two-step heuristic. On the other hand, the two-step heuristic often yields solutions with higher routing costs. Because the two-step heuristic first solves a covering problem for CHWs and then solves a LRP to determine where to locate CHWLs and their supervision cycles, we can determine that if the CHW cost would increase greatly, the two-step heuristic would probably provide optimal solutions. On the other hand, if the routing costs would increase greatly, the exact model would provide better solutions.

5.5 Sensitivity analysis

After discussions with LMH, we have determined that four parameters could be modified (the maximal cycle distance, D, the maximal coverage radius, σ, the maximal population coverage, ρ, and the supervision training time, τ) in order to reduce the total costs while ensuring that it remains appropriate to ensure good healthcare access in Liberia's context. These parameters have been increased with the help of LMH to determine which values would remain appropriate for Liberia's context. The other parameters (the maximum number of working hours per day, $\mathcal{T}$, and the maximal number of work days, δ) could not be increased as LMH did not wish to overload its CHWLs and wanted to ensure a weekly supervision visit instead of every two weeks or every month for example. This section presents our sensitivity analysis for values of D, σ, ρ, and τ.

5.5.1 Maximal cycle distance

The maximal cycle distance had initially been set to 30 km. Table 5 shows that the average distance between two communities is always more than 25 km and that, in some cases, some smaller communities are so remote that its closest community is at more than 15 km away which implies that setting $D = 30$ meant locating a CHWL in a very remote community. Thus, by discussing with LMH, we determined that increasing the value of D to at most 60 km would increase the maximal driving time from one to two hours, but would decrease the number of very remote communities with a CHWL. In our sensitivity analysis, we have considered values of $D = \{30, 40, 50, 60\}$.

In practice, the maximal cycle distance only has an impact on the total routing costs and the number of CHWLs needed. Thus, in this section, we present the impact of the maximal cycle distance on the CHWL workload. Tables 9–11 present for each district descriptive statistics on the CHWLs depending on the maximum cycle distance. We present the following information: the ID of the CHWL (*CHWL*); the number of work days for each CHWL, i.e., the number of cycles assigned to each CHWL (*# days*); the total number of km traveled per week by each CHWL (*KM*); the total supervision training time done by each CHWL (*TT*); the average number of km each CHWL spends on the road each day (*KM/day*); the average supervision training time each CHWL spends each day (*TT/day*); and the total work time per week (*Work time*).

Table 9: Summary statistics on CHWLs for District 1

CHWL	# days	KM	TT	KM/day	TT/day	Work time
$D = 30$ km						
1	4	88.43	16	22.11	4.00	18.95
2	4	51.54	26	12.89	6.50	27.72
3	1	16.41	4	16.41	4.00	4.55
4	1	0.00	4	0.00	4.00	4.00
5	4	78.19	18	19.55	4.50	20.61
6	2	15.09	8	7.54	4.00	8.50
7	2	44.16	12	22.08	6.00	13.47
8	1	7.75	6	7.75	6.00	6.26
9	1	8.00	6	8.00	6.00	6.27
10	4	85.40	16	21.35	4.00	18.85
11	1	0.00	4	0.00	4.00	4.00
Average	2.27	35.91	10.91	12.52	4.82	12.11
$D = 40$ km						
1	4	111.99	22	28.00	5.50	25.73
2	3	51.35	26	17.12	8.67	27.71
3	1	16.41	4	16.41	4.00	4.55
4	1	0.00	4	0.00	4.00	4.00
5	4	116.97	22	29.24	5.50	25.90
6	4	76.30	16	19.08	4.00	18.54
7	4	76.23	16	19.06	4.00	18.54
8	2	15.09	8	7.54	4.00	8.50
Average	2.88	58.04	14.75	17.06	4.96	16.68

In Liberia, CHWLs can work at most eight hours per day and four days a week (32 hours per week) and receive a fixed weekly salary. By taking a look at the results, we observe that increasing the maximal cycle distance increases the total work time of CHWLs. This is particularly true for District 1, because when the maximal cycle distance is set to 30 km, five CHWLs work only one day a week. Thus, when increasing the maximal cycle distance to 40 km, the number of CHWLs decreases from eleven to eight and the number of CHWLs working only one day a week decreases from five to two. On the other hand, in District 3, the number of CHWLs does not decrease when the cycle distance increases. Interestingly, even though the cycle distance increases, the average work time decreases as the average time spent on the road decreases. The practical implications of increasing the maximal cycle distance are to decrease the number of CHWLs working only one or two days per week, while each CHWL spends at most four hours per week in District 1, three hours per week in District 2, and two and a half hours per week in District 3 to travel to conduct their supervision training visits.

Table 10: Summary statistics on CHWLs for District 2

CHWL	# days	KM	TT	KM/day	TT/day	Work time
$D = 30$ km						
1	3	79.00	14	26.33	4.67	16.63
2	4	57.52	24	14.38	6.00	25.92
3	3	37.07	16	12.36	5.33	17.24
4	1	0.00	2	0.00	2.00	2.00
Average	2.75	43.40	14.00	13.27	4.50	15.45
$D = 40$ km						
1	3	59.52	14	19.84	4.67	15.98
2	4	86.49	26	21.62	6.50	28.88
3	3	37.07	16	12.36	5.33	17.24
Average	3.33	61.03	18.67	17.94	5.50	20.70
$D = 50$ km						
1	3	59.52	14	19.84	4.67	15.98
2	4	86.49	26	21.62	6.50	28.88
3	3	37.07	16	12.36	5.33	17.24
Average	3.33	61.03	18.67	17.94	5.50	20.70

Table 11: Summary statistics on CHWLs for District 3

CHWL	# days	KM	TT	KM/day	TT/day	Work time
$D = 30$ km						
1	3	51.03	16	17.01	5.33	17.70
2	1	0.00	4	0.00	4.00	4.00
3	2	55.10	8	27.55	4.00	9.84
4	4	86.42	20	21.60	5.00	22.88
Average	2.50	48.14	12.00	16.54	4.58	13.60
$D = 40$ km						
1	3	51.03	16	17.01	5.33	17.70
2	1	0.00	4	0.00	4.00	4.00
4	2	60.03	10	30.02	5.00	12.00
5	3	67.72	18	22.57	6.00	20.26
Average	2.25	44.69	12.00	17.40	5.08	13.49

5.5.2 Maximal coverage radius

By discussing with LMH, we have determined that increasing the value of the maximal coverage radius from 2.5 km to 3 km would be feasible in practice. As explained previously, the average walking speed for CHWs is about 5 km per hour in Liberia. Thus, increasing the maximal coverage radius from 2.5 km to 3 km would imply increase the maximal walking time

Table 12: Impact of maximal coverage radius on solution attributes

District	D	Δ CHWLs (%)	Δ CHWs (%)	Δz (%)
Increasing maximal coverage radius from 2.5 km to 2.75 km				
1	30	0.00	0.00	–1.68
1	40	0.00	0.00	0.00
2	30	0.00	–3.57	–14.22
2	40	0.00	–3.57	–3.39
2	50	0.00	–3.57	–3.39
3	30	0.00	0.00	0.00
3	40	0.00	0.00	0.00
Increasing maximal coverage radius from 2.5 km to 3.0 km				
1	30	0.00	0.00	–1.68
1	40	0.00	0.00	0.00
2	30	0.00	–7.14	–22.58
2	40	0.00	–7.14	–10.82
2	50	0.00	–7.14	–10.82
3	30	0.00	0.00	0.00
3	40	0.00	0.00	0.00

from 60 minutes to 72 minutes which would remain adequate in practice. In our sensitivity analysis, we have tested values of $\sigma = \{2.75, 3.00\}$. Table 12 presents the impact of increasing the maximum coverage radius from 2.5 km to 2.75 km and from 2.5 km to 3 km on the solution attributes. We present the following information: the name of the district (*District*); the maximal cycle distance in km (D); the variation in percentage on the number of CHWs with the increased maximum coverage radius computed as $(CHW_\sigma - CHW_{2.5})/CHW_{2.5}$, where CHW_σ and $CHW_{2.5}$ are the number of CHWs in the solution where the maximal coverage radius is σ, $\sigma = \{2.75, 3.0\}$ and 2.5 (Δ *CHWs*); the variation in percentage on the number of CHWLs with the increased maximum coverage radius computed as $(CHWL_\sigma - CHWL_{2.5})/CHWL_{2.5}$, where $CHWL_\sigma$ and $CHWL_{2.5}$ are the number of CHWLs in the solution where the maximal coverage radius is σ, $\sigma = \{2.75, 3\}$ and 250 (Δ *CHWLs*); and the variation in percentage on the total routing costs with the increased maximum coverage radius computed as $(z^*_\sigma - z^*_{2.5})/z^*_{2.5}$, where z^*_σ and $z^*_{2.5}$ are the total routing costs in the solution where the maximal coverage radius is σ, $\sigma = \{2.75, 3\}$ and 2.5 (Δz).

One can observe that, in general, increasing the maximal coverage radius decreases the costs by decreasing the number of CHWs and by decreasing the routing costs. In particular, for District 2, the number of CHWs decreases from 28 to 27 when increasing the maximal coverage radius from 2.5 km to 2.75 km, and decreases to 26 when increasing the maximal coverage radius to 3.0 km. On the other hand, this impact is limited because, as we can see

Table 13: Average proportion of communities (in %) within a specific radius of another community

	Radius (in km)		
District	2.50	2.75	3.00
1	3.27	3.45	3.57
2	6.60	6.94	7.99
3	5.89	6.25	6.48

in Districts 1 and 3, sometimes increasing the maximal coverage radius has no impact on the solution and sometimes it decreases the routing costs of at most 1.68%. Because of Liberia's sparse road network, allowing an additional 0.5 km of coverage radius has little impact on the number of potential communities that can be covered by a CHW. As reported in Table 13, the proportion of communities within a 3 km radius of another community is slightly larger for Districts 1 and 3 than the proportion of communities within a 2.5 km radius of another community. This helps explain the results found when increasing the maximal coverage radius. In summary, in practice, in denser areas, for example District 2, increasing the maximal coverage radius of 20% decreases the number of CHWs by more than 7% and the total routing costs by more than 10%. On the other hand, in sparser areas, increasing the maximal coverage radius has no impact on the total solution cost.

5.5.3 Maximal population coverage

In our sensitivity analysis, we have tested values of $\rho = \{300, 350\}$. These values have been chosen in collaboration with LMH to ensure that the workload of CHWs would remain appropriate as they are part-time workers. In addition, by taking a look at different community healthcare programs in other countries (The World Bank, 2017), we have determined that the average ratio is one CHW per 1,656 people, but in some countries this ratio can go as low as one CHW per 118 people. Thus, it seems that these ratios depend on each countries geography and increasing our ratio to 350 would remain appropriate for Liberia's context. Table 14 presents the impact of increasing the maximum population coverage from 250 to 300 and from 250 to 350 on the solution attributes. We present the following information: the name of the district (*District*); the maximal cycle distance in km (D); the variation in percentage on the number of CHWs with the increased maximum population coverage computed as $(CHW_\rho - CHW_{250})/CHW_{250}$, where CHW_ρ and CHW_{250} are the number of CHWs in the solution where the maximal population coverage is ρ, $\rho = \{300, 350\}$ and 250 (Δ *CHWs*); the variation in percentage on the number of CHWLs with the increased maxi-

Table 14: Impact of maximal population coverage on solution attributes

District	D	Δ CHWLs (%)	Δ CHWs (%)	Δz(%)
Increasing maximal population coverage from 250 to 300				
1	30	0.00	–6.67	–2.92
1	40	0.00	–6.78	–0.29
2	30	0.00	–3.57	–1.15
2	40	0.00	–3.57	–2.32
2	50	0.00	–3.57	–2.32
3	30	0.00	–8.33	–3.01
3	40	0.00	–8.33	–3.85
Increasing maximal population coverage from 250 to 350				
1	30	0.00	–13.33	–4.37
1	40	–12.50	–13.56	8.07
2	30	0.00	–21.43	–10.96
2	40	0.00	–21.43	–7.55
2	50	–33.33	–21.43	46.62
3	30	0.00	–8.33	–3.58
3	40	0.00	–8.33	–3.85

Table 15: Average population statistics within a 2.5 km radius of each community

District	Min	Max	Average	St. dev.
1	30	1,520	241	247
2	55	1,390	308	339
3	10	365	177	78

mum population coverage computed as $(CHWL_\rho - CHWL_{250})/CHWL_{250}$, where $CHWL_\rho$ and $CHWL_{250}$ are the number of CHWLs in the solution where the maximal population coverage is ρ, $\rho = \{300, 350\}$ and 250 (Δ $CHWLs$); and, the variation in percentage on the total routing costs with the increased maximum population coverage computed as $(z^*_\rho - z^*_{250})/z^*_{250}$, where z^*_ρ and z^*_{250} are the total routing costs in the solution where the maximal population coverage is ρ, $\rho = \{300, 350\}$ and 250 (Δz).

One can realize that, in general, increasing the maximal population coverage helps decrease the costs by decreasing the number of CHWs and by decreasing the routing costs. In addition, in two cases, the number of CHWLs also decreases which has an even greater impact on the total costs as CHWLs represent the highest costs. Table 15 reports the statistics concerning the total population within a 2.5 km radius of each community. Thus, we can see that in Districts 1 and 2, where the average are the highest and the standard deviation is high, increasing the maximal population coverage has a greater impact than in District 3 where the average is lower and the maximum is also much lower. Practically, increasing the maximal population coverage of 40% yields in a decrease between 8% and 20% on the number of CHWs, and either yields a decrease on the number of CHWLs or on the total routing costs.

By comparing Tables 12 and 14 our results show that in Liberia's context, increasing the maximal population coverage has a greater impact on reducing the total costs than increasing the maximal coverage radius. This is due to the geography of Liberia where communities are quite far from each other and to the fact that often CHWs will cover at most one or two communities.

5.5.4 Supervision training time

In LMH's basic model, the conducted supervision training time (τ) was two hours. Following discussions with LMH, we thought that it could be interesting to see the impact on the total costs of decreasing the supervision training time. On the other hand, it was important to ensure that this supervision training time lasted long enough to cover all key points. Thus, by taking a closer look at LMH's operations, we determined that it was best to have at least an hour and a half of training per week. In our sensitivity analysis, we have tested a value of $\tau = 1.5$. Table 16 presents the impact of decreasing the supervision training time from 2 hours to 1.5 hours on the solution attributes. We present the following information: the name of the district (*District*); the maximal cycle distance in km (D); the variation in percentage on the number of CHWs with the decreased supervision training time computed as $(CHW_{1.5} - CHW_2)/CHW_2$, where $CHW_{1.5}$ and CHW_2 are the number of CHWs in the solution where the supervision training time is 1.5 hours and 2 hours (Δ *CHWs*); the variation in percentage on the number of CHWLs with the decreased supervision training time computed as $(CHWL_{1.5} - CHWL_2)/CHWL_2$, where $CHWL_{1.5}$ and $CHWL_2$ are the number of CHWLs in the solution where the supervision training time is 1.5 hours and 2 hours (Δ *CHWLs*); and the variation in percentage on the total routing costs with the decreased supervision training time computed as $(z^*_2 - z^*_{1.5})/z^*_2$, where $z^*_{1.5}$ and z^*_2 are the total routing costs in the solution where the supervision training time is 1.5 hours and 2 hours (Δz).

One can realize that, in general, decreasing the supervision training time has a positive impact on reducing the total costs as the number of CHWLs can be reduced by up to 33.33% (District 2 with $D = 50$). In addition, for all instances where the number of CHWLs remains the same, the number of CHWs also remains unchanged, but the total routing costs

Table 16: Impact of supervision training time on solution attributes

District	D	Δ CHWLs (%)	Δ CHWs (%)	Δz(%)
Decreasing the supervision training time from 2 hours to 1.5 hours				
1	30	0.00	0.00	–2.90
1	40	–12.50	0.00	–0.63
2	30	0.00	0.00	–7.96
2	40	0.00	0.00	–7.60
2	50	–33.33	0.00	57.92
3	30	0.00	0.00	–5.39
3	40	–25.00	0.00	22.19

decrease up to 7.96%. Thus, because the CHWLs represent the highest cost for LMH, we can determine that decreasing the training time will have a positive impact on the total costs. Thus, this solution is interesting as it decreases the total costs and, if an organization determines that the quality of the training decreases, it could be combined with an additional monthly group supervision training to catch-up. This would ensure the appropriate training while decreasing the total costs and the number of required CHWLs.

6 Conclusions

In this section, we first highlight the contributions of our paper and, then, we provide future research ideas.

6.1 Highlights of our contributions

To conclude, we propose an integrated set-partitioning model to solve the LRCP for a real-life application in the context of a development program in underserved areas of Liberia. Because the number of variables is large, we developed a tool to generate non-dominated variables. We have shown that our variable generation technique helps reduce on average the number of variables by 62%. Thus, a MIP-solver can be used to solve the problem.

Our results show that solving the problem with the two-step heuristic yields the optimal solutions for District 2, but yields sub-optimal decisions for Districts 1 and 3. This is due to the geography and socio-demographics of all three districts. In addition, depending on the cost structure, the two-step heuristic could provide optimal solutions or could provide even worse solutions than the exact model. In fact, if the CHW costs are the highest costs, the two-step heuristic will most likely provide the optimal solution as it first minimizes the number of CHWs and then the sum of CHWL costs and routing costs. On the other hand, if the routing costs increase thus making routing more expensive than locating CHWs, then the two-step heuristic would provide much worse solutions than the exact model. In addition, the two-step heuristic requires more manual intervention than the exact model. Thus, our new model allows to reduce potential manual mistakes and is easier to use. In practice, because both the exact model and the two-step heuristic can provide solutions within one hour of computational time, we believe that the exact model is more interesting than the two-step heuristic as it guarantees optimality and requires less manual intervention.

We have conducted for the first time sensitivity analysis on various parameters for community healthcare network design in underserved areas. Our results show that in the context of Liberia increasing the maximal cycle distance helps decrease the total number of CHWLs needed as well as the total routing costs. In addition, increasing the maximal population coverage and the maximal coverage radius helps decrease the number of CHWs, CHWLs and the total routing costs. As shown in Table 13, the average proportion of communities within a given radius of a given community increases of at most 1.39% when increasing the radius from 2.5 km to 3 km. On the other hand, as shown in Table 15, the average population within a 2.5 km radius of each community is more than 241 for Districts 1 and 2, and some communities are much larger (1,520 people for the largest community of District 1). Thus, in Liberia, increasing the maximal population coverage has a greater impact on decreasing the total costs than increasing the maximal coverage radius. Finally, decreasing the supervision training time while ensuring appropriate training reduces the total costs by reducing the number of CHWLs and the routing costs when the number of CHWLs remains the same.

6.2 Future research

In terms of scientific future contributions, we believe that a first step would be to work on branching strategies to solve instances for District 1 with $D \geq 50$, District 2 with $D \geq 60$, and District 3 with $D \geq 50$. In fact, we have realized that we were not able to solve those instances due to excessive branching. Thus, we believe that working on branching strategies could help reduce the total solution time. A second step would be to develop valid inequalities for the proposed mathematical model. We think that developing those inequalities could help reduce the total solution time and solve additional instances. A third step would be to propose alternative mathematical formulations by changing the sets of variables and of constraints to determine improvements to our mathematical model. We would like to test different variable types such as route variables instead of cycle variables and eliminating the training time in the variables to add it as a variable of the model. We believe that these changes could have an impact on the number of generated variables as well as the lower bound quality. A fourth step would be to develop robust or stochastic mathematical models to include realities of the road network (e.g., road accessibility during the rainy and the dry seasons) as well as potential parameter changes (e.g., population per community and disease rate per community). These would help to have a solution that remains feasible even when some roads cannot be accessed or if the population was not well estimated.

In terms of practical future contributions, a first step would be to implement our model in an open source solver as this would increase the adaptation of such a tool in non-governmental organizations. We also believe that it would be useful to develop a tool that is more user-friendly. Currently, the user has to run CPLEX on a command line to solve the problem. Developing a graphical interface would be best to ease the use of this tool. A second step would be to propose a systematic method to determine the best parameter values (e.g., maximal cycle length, maximal coverage radius, maximal population coverage, and supervision training time) in the context of community healthcare in underserved areas. In practice, we think that integrating the notion of quality of healthcare to determine the best parameter values would be useful is this context. Currently, it is unclear, for example, how an increase on the maximal population coverage from 250 to 300 would impact the quality of healthcare.To determine appropriate ratios, qualitative data collecting needs to be conducted to better understand the impact of each operational constraint on the quality of healthcare provided. Data will need to be collected to determine how CHWs and CHWLs perceive their workload and capabilities to provide appropriate healthcare when those ratios change, and we should also collect data concerning different mortality rates and disease rates of the regions. The collected data will then need to be analyzed to identify key performance indicators for healthcare in underserved areas. Our model could then be adapted to a multi-criteria optimization method to systematically determine the best parameter values. A third step would be to define an objective where the workload equity is considered. In fact, while taking a look at our results, we have realized that one CHWL would be working less than a quarter of the time another CHWL in the same district would be working. This problem also arises for CHWs. Because CHWLs and CHWs received fixed weekly salaries, these disparities could yield to frustration from those working more. Thus, we believe that it would be important to add workload equity as a secondary objective and to develop a multi-criteria optimization method to provide alternative solutions. In order to reduce the total costs while ensuring appropriate healthcare, a fourth step would be to discuss with LMH to determine if it would be possible to have different types of CHWs. In fact, the training of CHWs is divided in four stages, but CHWs can start working once the first stage is completed. Thus, it could be interesting to see the impact of having different types of CHWs: some more experimented and some less experimented. In this model, less experimented CHWs could have a smaller coverage radius and population coverage, while those more experimented could have a larger coverage radius and population coverage. The salaries could also be different according to their responsibilities.

References

M. Albareda-Sambola. Location-routing and location-arc routing. In G. Laporte, S. Nickel, and F. Saldanha da Gama, editors, *Location Science*, chapter 15, pages 399–418. Springer International Publishing, 2015.

N. Altay and W. G. Green. OR/MS research in disaster operations management. *European Journal of Operational Research*, 175(1):475–493, 2006.

R. Baldacci, A. Mingozzi, and R. Wolfer Calvo. An exact method for the capacitated location-routing problem. *Operations Research*, 59(5):1284–1296, 2011.

J.-M. Belenguer, E. Benavent, C. Prins, C. Prodhon, and R. Wolfler Calvo. A branch-and-cut method for the capacitated location-routing problem. *Computers & Operations Research*, 38(6):931–941, 2011.

C. Contardo, J.-F. Cordeau, and B. Gendron. A computational comparison of flow formulations for the capacitated location-routing problem. *Discrete Optimization*, 10(4):263–295, 2013.

C. Contardo, J.-F. Cordeau, and B. Gendron. An exact algorithm based on cut-and-column generation for the capacitated location-routing problem. *INFORMS Journal on Computing*, 26(1):88–102, 2014.

J. R. Current. *Multiobjective design of transportation networks*. PhD thesis, Department of Geography and Environmental Engineering, The Johns Hopkins University, 1981.

J. R. Current and D. A. Schilling. The covering salesman problem. *Transportation Science*, 23(3):208–213, 1989.

J. R. Current and D. A. Schilling. The median tour and maximal covering tour problems: Formulations and heuristics. *European Journal of Operational Research*, 73(1):114–126, 1994.

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

M. Drexl and M. Schneider. A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research*, 241(2):283–308, 2015.

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

S. García and A. Marín. Covering location problems. In *Location Science*, pages 93–114. Springer, 2015.

M. Gendreau, G. Laporte, and F. Semet. The covering tour problem. *Operations Research*, 45(4):568–576, 1997.

B. Golden, Z. Naji-Azimi, S. Raghavan, M. Salari, and P. Toth. The generalized covering salesman problem. *INFORMS Journal on Computing*, 24(4):534–553, 2012.

M. Hachicha, M. J. Hodgson, G. Laporte, and F. Semet. Heuristics for the multi-vehicle covering tour problem. *Computers & Operations Research*, 27(1):29–42, 2000.

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

N. Jozefowiez, F. Semet, and E.-G. Talbi. The bi-objective covering tour problem. *Computers & Operations Research*, 34(7):1929–1942, 2007.

M. E. Kruk, P. C. Rockers, E. H. Williams, S. T. Varpilah, R. Macauley, G. Saydee, and S. Galea. Availability of essential health services in post-conflict liberia. *Bulletin of the World Health Organization*, 88:527–534, 2010.

Ministry of Health and Social Welfare, Republic of Liberia. National health and social welfare policy and plan 2011-2021. `http://www.mohsw.gov.lr/documents/Final%20NHPP%20(high%20res).pdf`, 2011.

Z. Naji-Azimi, J. Renaud, A. Ruiz, and M. Salari. A covering tour approach to the location of satellite distribution centers to supply humanitarian aid. *European Journal of Operational Research*, 222(3):596–605, 2012.

P. C. Nolz, K. F. Doerner, W. J. Gutjahr, and R. F. Hartl. A bi-objective metaheuristic for disaster relief operation planning. In *Advances in multi-objective nature inspired computing*, pages 167–187. Springer, 2010.

A. J. Pedraza-Martinez, O. Stapleton, and L. N. Van Wassenhove. Field vehicle fleet management in humanitarian operations: a case-based approach. *Journal of Operations Management*, 29(5):404–421, 2011.

A. J. Pedraza-Martinez, O. Stapleton, and L. N. Van Wassenhove. On the use of evidence in humanitarian logistics research. *Disasters*, 37(1):51–67, 2013.

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

M.-È. Rancourt, J.-F. Cordeau, G. Laporte, and B. Watkins. Tactical network planning for food aid distribution in kenya. *Computers & Operations Research*, 56:68–83, 2015.

The World Bank. Community health workers (per 1,000 people). `http://data.worldbank.org/indicator/SH.MED.CMHW.P3`, accessed January 2017.

L. N. Van Wassenhove. Humanitarian aid logistics: supply chain management in high gear. *Journal of the Operational Research Society*, 57(5):475–489, 2006.

P. Von Achen, K. Smilowitz, M. Raghavan, and R. Feehan. Optimizing community healthcare coverage in remote liberia. *Journal of Humanitarian Logistics and Supply Chain Management*, 6(3):352–371, 2016.

World Health Organization. *Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations*. World Health Organization, 2010.

W. Yi and L. Özdamar. A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research*, 179(3):1177–1193, 2007.