

We want to distribute COVID-19 vaccines efficiently across a region while balancing two important factors: prioritizing the individuals who are most at risk and minimizing the distance people need to travel to get vaccinated. As vaccines became available, governments faced the challenge of ensuring that the limited supply reached the right people in the right places. On one side, there is a need to ensure that the most vulnerable groups, such as the elderly or people with pre-existing health conditions, receive the vaccine first. On the other side, it is important that people do not have to travel far to get vaccinated. In this situation, we need to decide which healthcare centers will serve as vaccine distribution points. Not all available hospitals or clinics need to be used, but it is important to choose the ones that can serve the most people efficiently. Each of these centers has a limited number of staff who can administer vaccines, and the number of staff varies depending on the size of the facility.

Given these factors, the problem requires finding an optimal solution by defining key decision variables and ensuring certain constraints are satisfied. The first decision is to identify the set of vaccine distribution centers,  $H$ , from a larger set of available hospitals  $\tilde{H}$ . Using the K-medoids clustering algorithm, the optimal distribution centers are selected based on geographical proximity so that the centers can serve the surrounding population efficiently. Once the distribution centers are determined, the main decision variable is  $x_{i,j,k}$ , a binary variable that indicates whether a healthcare worker  $u_j$  from distribution center  $h_i$  is assigned to vaccinate an individual  $e_k$  (where  $x_{i,j,k} = 1$  if the assignment is made, and 0 otherwise). The variables  $U$  represent the set of available healthcare workers. The set  $E$  represents the individuals to be vaccinated, and  $P$  is used to denote the priority of each individual, with higher values indicating higher urgency for vaccination. Additionally, the distance between each individual and a distribution center is represented by  $D_{i,k}$ , which plays a crucial role in minimizing travel.

Several hard constraints must be satisfied to ensure a feasible solution. First, each healthcare worker can only vaccinate one person at a time, which ensures no worker is overloaded. Second, each individual must be assigned to only one distribution center, ensuring no person receives multiple vaccines from different centers. Third, the total number of vaccinations administered cannot exceed the available supply of vaccines. Finally, the decision variable  $x_{i,j,k}$  is binary, meaning it can only take values of 0 or 1, representing whether or not a vaccination occurs at a specific center by a specific healthcare worker.

In addition to the hard constraints, several soft constraints are incorporated to further refine the solution. The first soft constraint is maximizing the total number of people vaccinated. The second soft constraint focuses on prioritizing individuals based on their risk (priority) level. The third soft constraint seeks to minimize the travel distance for individuals. Based on these soft constraints, four different optimization models should be proposed: one that focuses purely on maximizing the number of vaccinations, another that prioritizes high-risk individuals, a third that minimizes travel distances, and finally a comprehensive model that balances both priority and distance.

For the **Basic Vaccine Distribution Model (B-VDM)**, the objective is to simply maximize the number of people vaccinated:

$$\text{Maximize} \quad \sum_{i,j,k} \alpha \cdot x_{i,j,k}$$

For the **Priority-based Vaccine Distribution Model (P-VDM)**, the objective is to prioritize higher-risk individuals by incorporating their priority level  $p_k$ :

$$\text{Maximize} \quad \sum_{i,j,k} x_{i,j,k} \cdot (\alpha + \beta \cdot p_k)$$

The **Distance-based Vaccine Distribution Model (D-VDM)** focuses on minimizing the total travel dis-

tance, adjusting the objective by penalizing longer distances  $D_{i,k}$ :

$$\text{Maximize} \quad \sum_{i,j,k} x_{i,j,k} \cdot (\alpha - \gamma \cdot D_{i,k})$$

Finally, the **Priority and Distance-based Vaccine Distribution Model (PD-VDM)** combines both priority and distance into a comprehensive objective function:

$$\text{Maximize} \quad \sum_{i,j,k} x_{i,j,k} \cdot (\alpha + \beta \cdot p_k - \gamma \cdot D_{i,k})$$

Here, the parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  are gain factors that balance different objectives in the vaccine distribution models.  $\alpha$  emphasizes maximizing the total number of vaccinated individuals, while  $\beta$  focuses on prioritizing higher-risk individuals.  $\gamma$  minimizes the travel distance for individuals by penalizing longer travel. The values for these factors are  $\alpha = 1/4 \times |\varepsilon|$  and  $\beta = 1/4 \times |\varepsilon|$ , where  $|\varepsilon|$  represents the total number of people to be vaccinated and  $\gamma$  is set to 1. Together, these gain factors allow the models to be fine-tuned to specific goals, such as maximizing coverage, prioritizing vulnerable groups, or reducing travel burdens.

## Task

- Provide a complete mathematical formulation of the problem, including decision variables, hard constraints, and four objective functions representing the soft constraints.
- Write a program that generates synthetic data and implements K-medoids clustering to identify the vaccine distribution centers (DCs). The synthetic data should include randomly generated priority levels, random locations (Cartesian coordinates between 0 and 100) for both individuals and hospitals, and random staff numbers for the hospitals.
- Implement all four optimization models using any LP solver (e.g., Gurobi, CPLEX, Google OR Tools, etc).
- Show results for three different combinations of the number of persons, hospitals, and available vaccines (e.g., 100 persons, 10 hospitals, 80 vaccines).
- For one of these combinations, analyze the effect of  $\alpha$ ,  $\beta$ , and  $\gamma$  by testing three different sets of values (e.g.,  $\alpha = \frac{1}{4} \times |\varepsilon|$ ,  $\beta = \frac{1}{4} \times |\varepsilon|$ ,  $\gamma = 1$ ).
- For each of these five runs, generate the following visualizations:
  - Number of vaccinated individuals from each priority level for each model.
  - Total number of vaccinated individuals for each model.
  - Total distance traveled by individuals for each model.

## Deliverables

- A PDF report detailing the mathematical formulation and all experimental results, including visualizations.
- A single code file (e.g., .py or .ipynb for Python) containing all the implementations.

## Rubrics

Criteria	Marks
Mathematical Formulation	15
Synthetic Data Generation and K-medoids Clustering	15
Implementation of Optimization Models	35
Analysis of Results (Different Combinations)	7.5
Effect of $\alpha, \beta, \gamma$ Analysis	7.5
Visualizations	20
<b>Total</b>	<b>100</b>

Table 1: Rubrics for Assignment Optimization

## Additional Materials

Further information related to the problem and clustering algorithm can be found in the following research paper. A PDF version of the paper is also attached.

[1] Emu, M., Chandrasekaran, D., Mago, V., & Choudhury, S. (2021, June). Validating optimal COVID-19 vaccine distribution models. In International Conference on Computational Science (pp. 352-366). Cham: Springer International Publishing.



# Validating Optimal COVID-19 Vaccine Distribution Models

Mahzabeen Emu<sup>✉</sup>, Dhivya Chandrasekaran<sup>✉</sup>, Vijay Mago<sup>✉</sup>,  
and Salimur Choudhury<sup>✉</sup>

Department of Computer Science, Lakehead University,  
Thunder Bay, ON P7B 5E1, Canada  
{memu,dchandra,vmago,schoudh1}@lakeheadu.ca

**Abstract.** With the approval of vaccines for the coronavirus disease by many countries worldwide, most developed nations have begun, and developing nations are gearing up for the vaccination process. This has created an urgent need to provide a solution to optimally distribute the available vaccines once they are received by the authorities. In this paper, we propose a clustering-based solution to select optimal distribution centers and a Constraint Satisfaction Problem framework to optimally distribute the vaccines taking into consideration two factors namely priority and distance. We demonstrate the efficiency of the proposed models using real-world data obtained from the district of Chennai, India. The model provides the decision making authorities with optimal distribution centers across the district and the optimal allocation of individuals across these distribution centers with the flexibility to accommodate a wide range of demographics.

**Keywords:** COVID-19 · Constraint satisfaction problem · Vaccine distribution · Operational research · Policy making

## 1 Introduction

The ongoing pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) called the coronavirus disease (COVID-19), has not only caused a public health crisis but also has significant social, political, and economic implications throughout the world [2, 13, 19]. More than 92.1 million cases and 1.9 million deaths have been reported worldwide as of January 2021 [6]. One of the widely used solutions, in preventing the spread of infectious diseases is vaccination. Vaccination is defined by WHO as “*A simple, safe, and effective way of protecting people against harmful diseases before they come into*

---

M. Emu and D. Chandrasekaran—Equal contribution.

© Springer Nature Switzerland AG 2021

M. Paszynski et al. (Eds.): ICCS 2021, LNCS 12742, pp. 352–366, 2021.

[https://doi.org/10.1007/978-3-030-77961-0\\_30](https://doi.org/10.1007/978-3-030-77961-0_30)

*contact with them. It uses the body's natural defenses to build resistance to specific infections and makes the immune system stronger"*<sup>1</sup>. Various researchers and pharmaceutical corporations began research work on identifying potential vaccines to combat the spread of COVID-19. Of the number of vaccines under trial, Tozinameran (BNT162b2) by Pfizer and BioNTech, and mRNA-1273 by Moderna have achieved an efficacy of more than 90% [12, 15]. Various nations have begun the process of approval of these vaccines for mass distribution and have placed orders to facilitate a continuous flow of supply of vaccines to meet the demands.

In this paper we propose a Constraint Satisfaction Programming (CSP) framework based model to optimize the distribution of vaccine in a given geographical region. We aim to maximize the distribution of vaccine among the group of the population with higher priority while minimizing the average distance travelled by any individual to obtain the vaccine. In order to justify the efficiency of the proposed model we compare the performance of the following four optimization models namely,

- Basic Vaccine Distribution Model (B-VDM)
- Priority-based Vaccine Distribution Model (P-VDM)
- Distance-based Vaccine Distribution Model (D-VDM)
- Priority in conjunction with Distance-based Vaccine Distribution Model (PD-VDM).

and present how the model PD-VDM provides the most optimal solution for the distribution of vaccines. We perform a Case Study using the demographic data obtained from Chennai - a well-renowned city in Southern India. This case study highlights how the model can be used by decision making authorities of a city with an population of 5.7 million. The model aids the decision making authorities to choose an optimal number of vaccine distribution centers (DCs), and to optimally assign an individual to a hospital such that the individuals in the priority groups are vaccinated first while minimizing the distance they travel to the vaccine DCs. In Sect. 2 we discuss the various steps and challenges involved in the process of vaccine distribution, in Sect. 3 we describe in detail the procedure followed to build the proposed models. The case-study is discussed in Sect. 4 followed by the discussion of results in Sect. 5. We conclude with future research directions in Sect. 6.

## 2 Related Work

In order to effectively distribute vaccines, it is necessary to understand the supply chain of vaccines. The supply chain of vaccines is divided into four major components namely the product, production, allocation, and finally the distribution [7]. The first concern of the decision making authorities is to decide on

<sup>1</sup> <https://www.who.int/news-room/q-a-detail/vaccines-and-immunization-what-is-vaccination>.

which vaccine to choose for distribution in their country, province, or region. For example, for countries in tropical regions, the storage temperature of the vaccine is an important factor. Similarly, while developed countries are able to afford vaccines at a higher price, most developing countries prefer the vaccine which has an affordable price. In the present scenario, three of the prominent vaccines in play are the BNT162b2, mRNA-1273, and AZD1222 [10]. They have storage temperatures of  $-70^{\circ}\text{C}$ ,  $-20^{\circ}\text{C}$ , and  $0^{\circ}\text{C}$  and cost USD 20, USD 50, and USD 4 respectively<sup>2</sup>. Based on the storage it is safe to assume that while countries in temperate regions like the United States, Canada, and Russia would have the option of purchasing any one of these vaccines while tropical countries like India, Bangladesh, Pakistan would prefer AZD1222. The cost of the said vaccines also has a significant impact on the decision-making process of developing nations which mostly cater to a greater number of people. Once the product is chosen, the production of vaccines has to be scheduled according to the demand. Factors that are taken into consideration at this stage include the production time, capacity for manufactures, supply-demand analysis, and so on [3]. Depending on the stage of the epidemic and the severity, the demand for the vaccine and the optimal timeline for the supply of vaccines may vary. Allocation and distribution stages go hand in hand in the vaccine supply chain. Depending on the distribution strategy, the allocation of vaccines at any level is tuned to achieve the best result. Allocation at a global level may depend on priority established through contracts and agreements among pharmaceuticals and governments. However, once a country receives the vaccines, further allocating the vaccines to provinces, states, or subgroups of the population is a critical decision that in turn has an impact on the distribution strategy. The distribution stage of the supply chain addresses the challenges of establishing an effective routing procedure, infrastructure of the vaccine distribution centers, inventory control, workforce, etc.

Operation Research (OR) involves the development and use of various statistical and mathematical problem-solving strategies to aid in the process of decision making. Various OR models are proposed over time to optimize the distribution of vaccines from the distribution centers. Ramirez-Nafarrate et al. [17] proposes a genetic algorithm to optimize the distribution of vaccines by minimizing the travel and waiting time. Huang et al. [9] formulated a vaccine deployment model for the influenza virus that ensures geographical priority based equity in Texas. However, their mathematical model might have generalization issues when applied to smaller or larger than state-level regions. Lee et al. [11] developed the RealOpt<sup>®</sup> a tool to aid in identifying optimal location for vaccine distribution centers, resource allocation and so on. Researchers have also provided models to accommodate specific locations, for example Aaby et al. [1] proposes a simulation model to optimize the allocation of vaccine distribution centers at Montgomery county, Maryland. This model aims to minimize the vaccination time and increase the number of vaccinations. While the above models

<sup>2</sup> <https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html>.

consider the distribution centers to be stationary, Halper et al. [8] and Rachaniotis et al. [16] consider the vaccine distribution centers to be mobile and address this as a routing problem. While the later model proposes that various mobile units serve different nodes in a network, the latter considers that a single mobile unit serves various areas with a goal to minimize the spread of infection. Some of the OR models used in epidemics' control include non-linear optimization, Quadratic Programming (QP), Integer Linear Programming (ILP) and Mixed Integer Linear Programming (MILP) [5]. ILP, MILP, and QP models are not suitable for many practical use cases due to its time expensive nature and infeasibility issues prevailing with irrational model designs. The mathematical design process and selection of pre-defined numerical bounds can lead to several technical glitches in the models. Despite the apparent similarities with ILP and MILP, CSP can eliminate all the aforementioned drawbacks and ensure sub-optimal solution in non-deterministic polynomial time by applying boolean satisfiability problem formulation. In the field of computer science, CSP is considered as a powerful mechanism to address research challenges regarding scheduling, logistics, resource allocation, and temporal reasoning [4]. Hence in this article we employ CSP to propose four models to optimize the distribution minimizing the traveling distance and maximizing the vaccination of high priority population.

### 3 Methodology

In this section, we initially determine the optimal number and location of vaccine DCs using K-medoids clustering algorithm. Provided with the various locations of possible vaccine DCs, the algorithm determines the optimal number of clusters into which the region can be divided into, in order to effectively distribute the vaccines across the chosen region. On selecting the number of clusters and the cluster heads, we further propose four different vaccine distribution simulation models to optimize vaccine distribution based on two factors namely distance and priority. The clustering algorithm and the simulation models are explained in the subsections below.

#### 3.1 Selection of Optimal Vaccine Distribution Center

Our proposed Algorithm 1 imitates the core logic of K-medoids clustering technique [14]. Firstly, the algorithm determines the optimal number of vaccine DCs to be selected from a set of hospitals  $\mathcal{H}$  based on silhouette score analysis [18], as mentioned in line number 1. We determine the silhouette score for 2 to  $\tilde{h}_{\tilde{H}-1}$  number of potential vaccine DCs, where  $\tilde{\mathcal{H}} = \{\tilde{h}_1, \tilde{h}_2, \tilde{h}_3, \dots, \tilde{h}_{\tilde{\mathcal{H}}}\}$ . Then, we select the optimal number of vaccine DCs as  $\eta$  that retains the highest silhouette score. As per the line number 2, we randomly select  $\eta$  hospitals as vaccine DCs into  $\mathcal{H}$ . Later, we initiate the clustering process. Each hospital is assigned to its closest vaccine DC to form  $\eta$  clusters, according to line numbers 5–8. The cluster information indicating which hospital is associated to which vaccine DC are recorded in  $\mathcal{C}$ . Next, the algorithm reassigns the vaccine DCs  $\mathcal{H}$  to the ones with the

**Algorithm 1:** K-medoids algorithm to choose vaccine DCs from a set of hospitals

---

**Input:**  $\tilde{\mathcal{H}}$ : A set of hospitals,  $dist$ : Squared matrix representing the distances of one hospital to every other hospital  
**Result:**  $\mathcal{H}$ : A set of COVID-19 vaccine DCs

```

1  $\eta \leftarrow$  Determine the number of optimal vaccine distribution centers using silhouette score
2  $\mathcal{H} \leftarrow$  Randomly select  $\eta$  hospitals from  $\tilde{\mathcal{H}}$ 
3  $\mathcal{C} \leftarrow \emptyset$ 
4 while there is no change in  $\mathcal{H}$  do
5   foreach  $\tilde{h}_a \in \tilde{\mathcal{H}}$  do
6     foreach  $h_b \in \mathcal{H}$  do
7        $h_b \leftarrow$  Find the closest  $h_b$  to  $\tilde{h}_a$  using  $dist$  matrix
8        $\mathcal{C} \leftarrow \mathcal{C} \cup (\tilde{h}_a, h_b)$ 
9   foreach  $h_b \in \mathcal{H}$  do
10     $temp \leftarrow \emptyset$ 
11    foreach  $\tilde{h}_a \in \tilde{\mathcal{H}}$  do
12      if  $(\tilde{h}_a, h_b) \in \mathcal{C}$  then
13         $temp \leftarrow temp \cup \tilde{h}_a \cup h_b$ 
14     $q \leftarrow \underset{\tilde{h}_a \in temp}{\operatorname{argmin}} \sum_{h_a^* \in temp} dist(\tilde{h}_a, h_a^*)$ 
15    Swap  $h_b$  with  $\tilde{h}_q$  in  $\mathcal{C}$ 
16    Update  $h_b$  by  $\tilde{h}_q$  in  $\mathcal{H}$ 

```

---

minimum total distance to all other hospitals under the same cluster, executed in line numbers 9–16. We let the algorithm repeat the entire clustering process until vaccine DC assignments do not change. Hence, the termination criteria depends on the stability of the clustering process. To summarize, we employ this algorithm with inputs of a set of hospitals/potential vaccine DCs,  $\tilde{\mathcal{H}}$  and a 2D  $dist$  matrix defining the distances of one hospital to every other hospital. The output of the algorithm is the set of optimally selected vaccine DCs  $\mathcal{H}$  based on the distance metric. The primary idea is to choose vaccine DCs optimally in a sparse manner to facilitate reachability for people living in any part of the considered region.

### 3.2 System Model for Vaccine Distribution

In this subsection, we proceed by explaining the system model for vaccine distribution. We denote  $\mathcal{H} = \{h_1, h_2, h_3, \dots, h_{|\mathcal{H}|}\}$  to be the set of COVID-19 vaccine distribution centers selected by Algorithm 1, where  $\mathcal{H} \subseteq \tilde{\mathcal{H}}$  and  $h_i$  is the  $i^{th}$  vaccine DC in  $\mathcal{H}$ . Moreover, we define  $\mathcal{U} = \{u_1, u_2, u_3, \dots, u_{|\mathcal{U}|}\}$  as the set of available staff, where every  $u_j \in \mathcal{U}_i$ . We denote  $\mathcal{U}_i$  as the available set of staff in vaccine DC  $h_i$  and  $\mathcal{U}_i \subseteq \mathcal{U}$ . Subsequently, we can infer that  $\mathcal{U} = \bigcup_{i=1}^{|\mathcal{H}|} \mathcal{U}_i$ . We further assume that  $\mathcal{E} = \{e_1, e_2, e_3, \dots, e_{|\mathcal{E}|}\}$  represents the set of people required to be vaccinated, and the  $k^{th}$  person is  $e_k$ . In order to specify the priority of people for vaccination purpose, we use the set  $\mathcal{P} = \{p_1, p_2, p_3, \dots, p_{|\mathcal{P}|}\}$ . Hence,  $p_k$  defines the priority level of a person  $e_k \in \mathcal{E}$ , and  $|\mathcal{E}| = |\mathcal{P}|$ . It is noteworthy that, the higher the priority level, the faster the vaccination service deployment is desired. The distance between a distribution center  $h_i \in \mathcal{H}$  and a specific person  $e_k \in \mathcal{E}$  is represented using  $\mathcal{D}_{i,k}$ . In this research, we consider the solution binary decision variable as  $x_{i,j,k} \in \{0, 1\}$ . The value of the decision variable  $x_{i,j,k}$  is 1, in case a



distribution center  $h_i \in \mathcal{H}$  allocates a staff  $u_j \in \mathcal{U}_i$  to vaccinate a person  $e_k \in \mathcal{E}$ , otherwise remains 0.

### 3.3 Problem Formulation

In this paper, we formulate the opted vaccine distribution research enigma as a CSP model. The CSP framework includes a set of aforementioned decision variables that should be assigned values in such a way that a set of hard constraints are satisfied. Hard constraints are essential to be satisfied for any model to reach a feasible solution. Suppose, our proposed model iterates over  $T = \{t_1, t_2, \dots, t_T\}$  times to complete vaccination, where each time instance  $t_n \in T$  refers to per time frame for vaccine deployment decision making. Finally,  $\mathcal{N}$  represents the total amount of available vaccine throughout the entire time periods  $T$ . All of our proposed CSP based models are subject to the following hard constraints that have been translated into integer inequalities, for any time frame  $t_n \in T$ :

$$\mathcal{C1} : \sum_{e_k \in \mathcal{E}} x_{i,j,k} \leq 1, \forall_{h_i \in \mathcal{H}}, \forall_{u_j \in \mathcal{U}_i} \quad (1)$$

$$\mathcal{C2} : \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} x_{i,j,k} \leq 1, \forall_{e_k \in \mathcal{E}} \quad (2)$$

$$\mathcal{C3} : x_{i,j,k} \in \{0, 1\}, \forall_{h_i \in \mathcal{H}}, \forall_{u_j \in \mathcal{U}_i}, \forall_{e_k \in \mathcal{E}} \quad (3)$$

The constraint  $\mathcal{C1}$  verifies that every staff from any distribution center can vaccinate at most one person at a time. Thus, for every staff  $u_j \in \mathcal{U}_i$  of distribution center  $h_i \in \mathcal{H}$ , either there is a unique person  $e_k \in \mathcal{E}$  assigned for vaccination, or the staff remains unassigned. Then, constraint  $\mathcal{C2}$  ensures that every person  $e_k \in \mathcal{E}$  is allocated at most one vaccine through a single staff  $u_j \in \mathcal{U}_i$  from a unique distribution center  $h_i \in \mathcal{H}$ . Finally,  $\mathcal{C3}$  is a binary constraint representing the value of decision variable to be 1, in case a staff  $u_j \in \mathcal{U}_i$  of distribution center  $h_i \in \mathcal{H}$  is assigned to vaccinate a person  $e_k \in \mathcal{E}$ , otherwise 0, as mentioned previously.

$$\mathcal{C4} : \sum_{t_n \in T} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \leq \mathcal{N} \quad (4)$$

The constraint  $\mathcal{C4}$  confirms that the total vaccine distribution should not be more than the available vaccine by any means throughout the entire periods  $T$  considered for vaccine deployment. Now, let us assume  $\Omega$  be the set of all the feasible solutions that satisfy all hard constraints.

$$\Omega = \{x_{i,j,k} \mid \mathcal{C1}, \mathcal{C2}, \mathcal{C3}, \mathcal{C4}\} \quad (5)$$

Apart from hard constraints, our proposed CSP formulation incorporates a set of soft constraints as well. Whilst hard constraints are modeled as inequalities, soft constraints are outlined through expressions intended to be eventually

358 M. Emu et al.

minimized or maximized. Soft constraints are not mandatory for finding a solution, rather highly desirable to improvise the quality of the solutions based on the application domain. The soft constraint  $\mathcal{C}5$  strives to maximize the number of overall vaccinated people. The focus of another soft constraint  $\mathcal{C}6$  remains to vaccinate the people with higher priority levels beforehand. In other words, this constraint leads to maximize the summation of priorities of all vaccinated people. Subsequently, the soft constraint  $\mathcal{C}7$  refers that every people should be vaccinated by staff from the nearest vaccine distribution center.

$$\mathcal{C}5 : \max_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \quad (6)$$

$$\mathcal{C}6 : \max_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \times p_k \quad (7)$$

$$\mathcal{C}7 : \min_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \times \mathcal{D}_{i,k} \quad (8)$$

### 3.4 Variations of Vaccine Distribution Models

By leveraging different combinations of soft constraints, we propose four different variations of vaccine distribution models: **a)** Basic - Vaccine Distribution Model (B-VDM), **b)** Priority based - Vaccine Distribution Model (P-VDM), **c)** Distance based - Vaccine Distribution Model (D-VDM), and **d)** Priority in conjunction with Distance based - Vaccine Distribution Model (PD-VDM). B-VDM is a rudimentary vaccine distribution model that solely concentrates on the soft constraint  $\mathcal{C}5$  to maximize the overall vaccine distribution, irrespective of any other factors. A gain co-efficient  $\alpha$  has been introduced to the ultimate objective function of the model in Eq. 9.

$$\mathcal{C}5 \iff \max_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} \alpha \times x_{i,j,k} \quad (9)$$

P-VDM ensures maximum vaccine distribution among the higher priority groups of people, by reducing soft constraints  $\mathcal{C}5$  and  $\mathcal{C}6$  into one objective function in Eq. 10. We denote  $\beta$  as the gain factor associated to soft constraint  $\mathcal{C}6$ .

$$\mathcal{C}5 \wedge \mathcal{C}6 \iff \max_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \times (\alpha + \beta \times p_k) \quad (10)$$

Contrarily, D-VDM encourages vaccination of the people located closer to distribution centers. This model can be specifically useful for rural regions, where distribution centers and people are sparsely located, including higher travelling expenses. For this model, we incorporate soft constraints  $\mathcal{C}5$  and  $\mathcal{C}7$ , by multiplying gain coefficients  $\alpha$  and  $\gamma$ , respectively.

$$\mathcal{C}5 \wedge \mathcal{C}7 \iff \max_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \times (\alpha - \gamma \times \mathcal{D}_{i,k}) \quad (11)$$

Finally, the PD-VDM merges all the soft constraints simultaneously. Our proposed PD-VDM considers maximization of vaccine distribution in priority groups and minimization of distance factored in transportation expenditure, collaboratively. Furthermore,  $\alpha$ ,  $\beta$ , and  $\gamma$  have been introduced as gain factors to equilibrate the combination of soft constraints and then presented as a multi-objective function in the Eq. 12. Hence, this model can optimize priority and distance concerns conjointly based on the adapted values of gain factors.

$$\mathcal{C5} \wedge \mathcal{C6} \wedge \mathcal{C7} \iff \max_{\Omega} \sum_{h_i \in \mathcal{H}} \sum_{u_j \in \mathcal{U}_i} \sum_{e_k \in \mathcal{E}} x_{i,j,k} \times (\alpha + \beta \times p_k - \gamma \times \mathcal{D}_{i,k}) \quad (12)$$

The gain parameters of all the proposed models can be tuned to balance or incline towards more focused convergent vaccine distribution solutions. For instance,  $\alpha$ ,  $\beta$ , and  $\gamma$  are individually responsible for maximum distribution, maximization of priorities, and minimization of distance focused vaccine distribution solutions, respectively. Moreover, the policymakers can exploit these models and adjust gain parameters according to the region specifics and domain knowledge of vaccine distribution centers and population density. The regulation of these gain factors can aid to analyze and figure out the applicability of our various proposed models relying on different contextual targets, environment settings, and demand-supply gaps.

## 4 Case Study - Chennai, India

We demonstrate the proposed models using real-world data obtained from one of the popular cities in the southern part of India - Chennai. As listed by Rachaniotis et al. [16] most of the articles in literature make various assumptions to demonstrate the performance of their models. Similarly, we make a few reasonable assumptions to accommodate the lack of crucial data required to implement the model. Various input parameters of the models and their method of estimation or assumptions made to reach the decisions are described below.

### 4.1 Clustering Phase

- The entire city is divided into 15 zones for administrative purposes and we assume the distribution of vaccines is carried out based on these 15 zones as well.
- To determine the vaccine distribution centers, we assume that the vaccines will be distributed from hospitals or primary health centers. While there are approximately 800 hospitals in Chennai, based on ‘on-the-ground’ knowledge, we select 45 hospitals (3 hospitals per zone) to enable us to determine the distance between the hospitals. The selected hospitals are classified as private (PVT) and public funded (PUB) based on their administration. At least one publicly funded hospital or primary health center is chosen per zone.

360 M. Emu et al.

- A  $45 \times 45$  distance matrix is constructed with each row and column representing the hospitals. The cells are populated with the geographic distance obtained from Google Maps<sup>3</sup>. Using these distance measures, the optimal vaccine DCs are chosen by implementing k-medoids clustering algorithm described in Sect. 3.1.

#### 4.2 Vaccine-Distribution Phase

- **Total population to be vaccinated ( $\mathcal{E}$ ):** As per the census records collected in 2011, Chennai has a population of 4.6 million distributed across an area of  $175 \text{ km}^2$  with a population density of 26,553 persons/ $\text{km}^2$ . Based on the growth in the overall population of India, we estimate the current population of Chennai to be 5,128,728 with a population density of 29,307/ $\text{km}^2$ .
- **Set of Vaccine Distribution centers ( $\mathcal{H}$ ):** The optimal vaccine DCs are chosen from using the clustering phase of the model using silhouette width such that the chosen DCs are evenly spread across the entire district.
- **Staff for vaccination per DC ( $\mathcal{U}$ ):** Based on the capacity of the hospital in terms of facilities, workforce, etc. the hospitals are classified as ‘SMALL’, ‘MED’, and ‘LARGE’. We assume that small, medium, and large hospitals allocate 5, 20, and 40 health-care workers for vaccination purposes respectively.
- **Priority levels ( $\mathcal{P}$ ):** While the priority levels can be decided by the authorities based on wide variety of parameters, for our simulation we assume that the priority depends on the age of the individual such that the elderly people are vaccinated earlier. The distribution of population across various age groups is calculated based on the age-wise distribution calculated during the census 2011. Table 1 shows the distribution of the population across six priority groups.

**Table 1.** Distribution of the population of Chennai across 6 age groups

Age group	0–9	10–19	20–54	55–64	65–74	75+
% of distribution	14.02	15.34	56.38	7.87	4.09	2.31

- **Time per vaccination ( $t$ ):** Based on the data provided during the recent trial dry run carried out in India, the government estimates to carry out 25 vaccinations in 2 h span, we approximate that the time taken for the vaccination of one person to be 5 min.
- **Time per vaccination ( $\mathcal{N}$ ):** Keeping in mind the deficit in the supply of vaccines in the early stages of vaccination we assume that only vaccine doses for 50% of the total population is available currently. However, the number can be increased based on increase in production<sup>4</sup>.

<sup>3</sup> <https://www.google.com/maps/>.

<sup>4</sup> <https://www.cnn.com/2020/12/18/asia/india-coronavirus-vaccine-intl-hnk/index.html>.

These described parameters can be tuned by the decision making authorities to accommodate the distribution at the area under consideration. For all the experimental settings, we consider the gain factors  $\alpha$ ,  $\beta$ , and  $\gamma$  as  $\frac{1}{4} \times |\epsilon|$ ,  $\frac{1}{4} \times \frac{1}{|\mathcal{P}|} \times |\epsilon|$ , and 1 respectively. Yet, as mentioned earlier, these gain factors can be explored and set according to the solutions desired by policy makers and decision making authorities.

## 5 Results and Discussion

We demonstrate the proposed models in two scenarios with randomly generated data and two scenarios with real world data from Chennai. We discuss in detail the inferences from each scenario in this section. In each scenario we compare the four models and highlight how PD-VDM optimizes the two parameters taken into consideration - priority and distance.

### 5.1 Random-Simulation

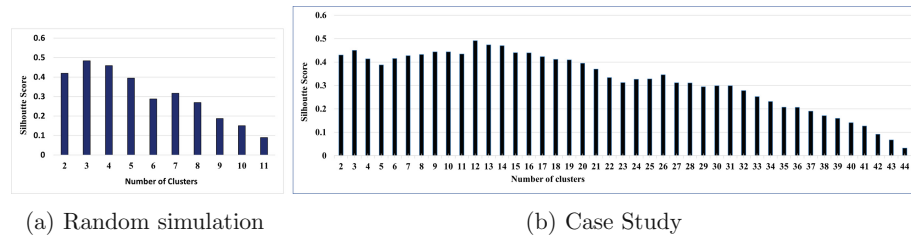
For the two random scenarios we consider that there are 12 vaccine DCs in total and based on the silhouette width measures as shown in Fig. 1a we select three DCs optimally distributed across the area. We assume the total population size ( $\mathcal{E}$ ) to be 200 and the total number of vaccines available ( $\mathcal{N}$ ) to be 85. Of the three chosen DCs we assume that each has a capacity ( $\mathcal{U}$ ) of 15, 30, 45. The model considers five different priority levels ( $\mathcal{P}$ ) with 1 being the least and 5 being the most. The population is distributed among these priority levels as shown in Table 2. To demonstrate the impact of the population distribution parameter, we run the simulation model under two different distributions namely,

**Table 2.** Distribution of population across priority groups in random simulation.

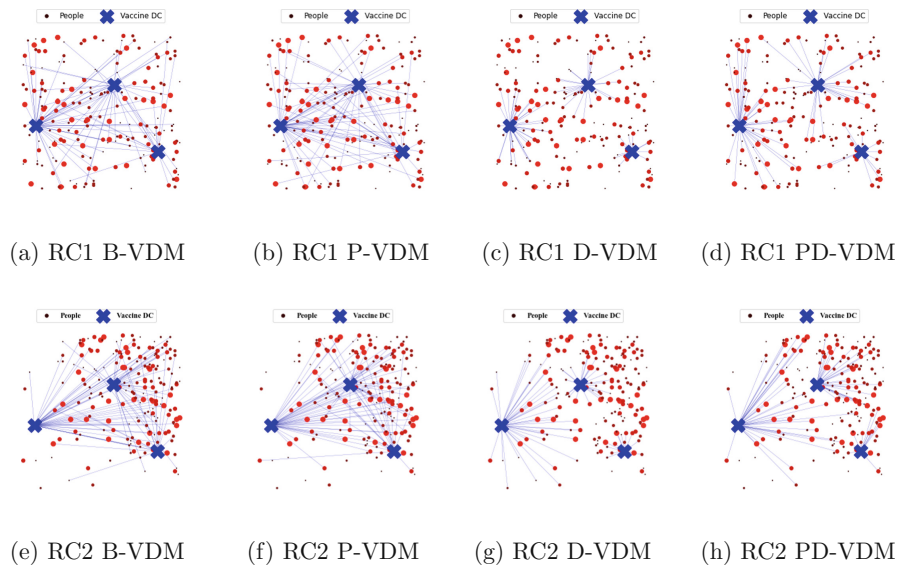
Priority group	1	2	3	4	5
Raw count	43	35	50	45	27

- Random-case - 1 (RC-1): Uniform random distribution
- Random-case - 2 (RC-2): Poisson like distribution where the population is dense at some regions and spares at others.

362 M. Emu et al.

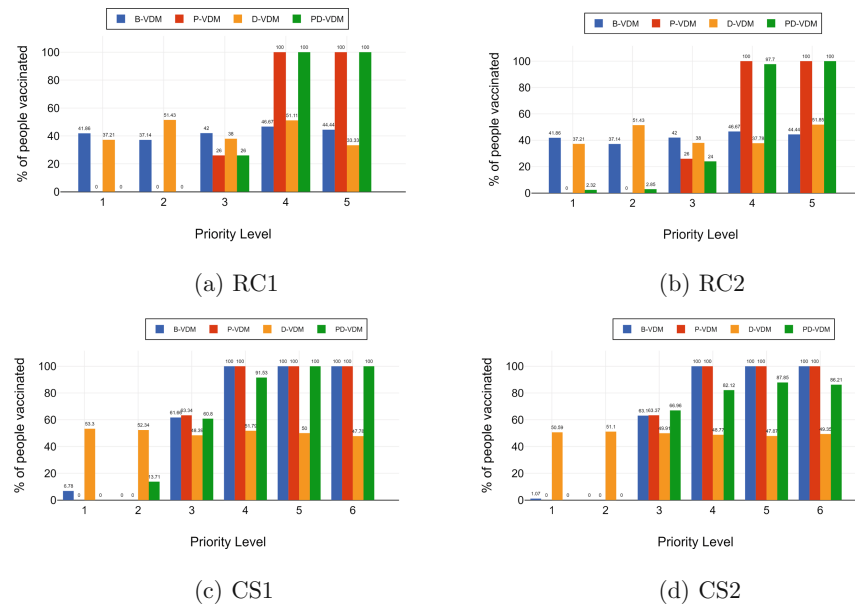
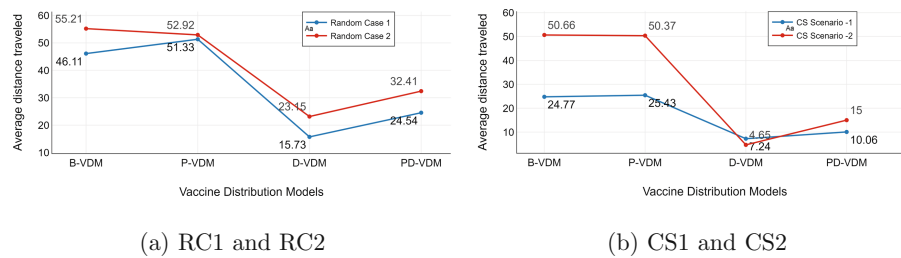
**Fig. 1.** Selection of optimum number of DCs based on Silhouette width

The distribution of vaccines at any time instance  $t_n$  by all four models, for both random cases is depicted in Fig. 2. The percentage of individuals vaccinated in each priority group in random case 1 and random case 2 are depicted in Fig. 3a and b respectively. The B-VDM vaccinates 41.86%, 37.14%, 42%, 46.67%, and 44.44% of individuals across priorities 1 to 5 respectively, in both RC1 and RC2. We can see that 55.56% of the individuals from the highest priority group are left out. The D-VDM optimizes only the distance parameter and vaccinates 37.21%, 51.43%, 38%, 51.11% and 33.33% of individuals across priorities 1 to 5 respectively in RC1 and almost the same results in RC2. Again we can notice, that a greater percentage of the individuals in the highest priority group are not vaccinated. While both P-VDM and PD-VDM attempt to vaccinate 100% of the high priority individuals in both RC1 and RC2, on studying the average distance travelled by each individual of the population as depicted in Fig. 4a,

**Fig. 2.** Snapshot of the simulation of the models in the random case studies

## Validating Optimal COVID-19 Vaccine Distribution Models

363

**Fig. 3.** Distribution of vaccine across various priority groups**Fig. 4.** Average distance travelled by the proposed models**Table 3.** Distribution of population across six priority groups for Case study scenarios

Priority group	1	2	3	4	5	6
Case-study scenario 1	546	598	2199	307	160	90
Case-study scenario 2	2817	3082	11331	1583	823	464

we can clearly identify that the PD-VDM reduces the distance parameter by more than 40% in both RC1 and RC2. We can also see that the ‘distribution of population’ parameter does not impact the performance of the models and PD-VDM is the most efficient across both Uniform and Poisson distribution. Although there is a slight increase in the average distance travelled in the second distribution the PD-VDM still achieves the most optimal results.

## 5.2 Case-Study Simulation

In both the case-study scenarios we consider that the vaccination process continues for five hours each day and as mentioned in Sect. 4 each vaccination takes ( $t_n$ ) 5 mins. Hence there are 60 vaccinations carried out by each health care worker for any given day. We compile and present the results of the vaccination process at the end of any given day. Although in the initial stages of the vaccination it is likely that individuals from one priority group will only be vaccinated, as time progresses people across various priority groups will need to be considered. Hence the population considered by the model for each day is taken as a stratified sample from the age-wise distribution of the overall population of Chennai provided in Table 1. Initially we identify the optimal number of DCs needed for effectively serving the district of Chennai based on the distance between the hospitals considered. On analyzing the silhouette width of various ‘number of clusters’ as depicted in Fig. 1b we present two different scenarios namely,

- Case study scenario - 1 (CS1): Three optimal DCs
- Case study scenario - 2 (CS2): Twelve optimal DCs

**Case Study Scenario - 1:** For this scenario, we consider 3 optimal vaccine DCs ( $\mathcal{H}$ ) with capacities ( $\mathcal{U}$ ) of 5, 20, 40 since the chosen DCs fall under ‘SMALL’, ‘MED’ and ‘LARGE’ respectively. Based on these factors the total population to be vaccinated ( $\mathcal{E}$ ) in any given data is estimated to be 3,900. As mentioned in Sect. 4 we assume, that the total vaccines ( $\mathcal{N}$ ) available to be half the total population which in this case sums upto 1,950. The entire population falls under six priority groups ( $\mathcal{P}$ ) based on their age. The actual number of people in each priority group is provided in Table 3.

**Case Study Scenario - 2:** For this scenario, we consider 12 optimal vaccine DCs ( $\mathcal{H}$ ) and among these 12 chosen DCs there are 3, 2 and 7 DCs with a staff capacity ( $\mathcal{U}$ ) of 5 (‘SMALL’), 20 (‘MED’), and 40 (‘LARGE’) respectively. Hence the total population ( $\mathcal{E}$ ) that can be vaccinated at any given 5 h day is 20100. Similar to scenario 1 we assume that the total vaccines ( $\mathcal{N}$ ) available to be half the total population which amounts to 10,050 and that the population is distributed across six different priority groups ( $\mathcal{P}$ ) based on their age as shown in Table 3.

The vaccination percentage across the priority groups for both case study scenarios are provided in Fig. 3c and d respectively. Unlike the random scenarios, it is interesting to note that in both CS1 and CS2, the B-VDM though



not optimized to satisfy a specific hard constraint, it produces results that are almost identical to P-VDM vaccinating 100% of the highest priority group individuals. This can be attributed to the similarity in the effect of  $\alpha$  and  $\beta$  gain parameters in both these models in CS1 and CS2. Though the PD-VDM model vaccinates less than 90% of the individuals in the three highest priority groups we can clearly see that it significantly reduces the ‘average distance travelled’ by an individual in the population by more than 70%, which makes it more efficient than all the other models. While D-VDM achieves the least ‘average distance travelled’ value it sacrifices vaccinating almost 50% of the high priority groups. Thus, we demonstrate how PD-VDM efficiently distributes the available vaccines by modifying various parameters like the distribution of population, total population, total number of available vaccines, the number of vaccine DCs and the capacity of each DCs. The model provides flexibility for the decision-making authorities of any given demographic to optimize the distribution of vaccine in the desired region. Moreover, the models developed using CSP are easier to design, adapt, and control compared to other existing combinatorial optimization techniques (e.g., ILP, MILP, and QP) that are often very computationally expensive in terms of memory and time with very rigid search space navigation restrictions.

## 6 Conclusion and Future Work

In this paper, we propose an optimization model (PD-VDM) based on Constraint Satisfaction Programming framework to find the most effective way to distribute vaccines in a given demographic region, in terms of distance and a priority (age, exposure, vulnerability, etc.). We compare the efficiency of the model with three other models which take into consideration either one or none of the two optimization constraints. While this model can be adapted across a wide variety of scenarios as demonstrated in our case studies, it is essential to understand that due to resource constraints we have demonstrated only two of the many available constraints. Expanding the scope of the model to allow optimizing a wide variety of parameters can be carried out in future research, along with an attempt to replace some of the assumptions made in our model with real world data.

## References

1. Aaby, K., Herrmann, J.W., Jordan, C.S., Treadwell, M., Wood, K.: Montgomery county’s public health service uses operations research to plan emergency mass dispensing and vaccination clinics. *Interfaces* **36**(6), 569–579 (2006)
2. Ali, I., Alharbi, O.M.: COVID-19: disease, management, treatment, and social impact. *Sci. Total Environ.* **728**, 138861 (2020)
3. Begen, M.A., Pun, H., Yan, X.: Supply and demand uncertainty reduction efforts and cost comparison. *Int. J. Prod. Econ.* **100**(180), 125–134 (2016)

366 M. Emu et al.

4. Bordeaux, L., Hamadi, Y., Zhang, L.: Propositional satisfiability and constraint programming: a comparative survey. *ACM Comput. Surv. (CSUR)* **38**(4), 12-es (2006)
5. Brandeau, M.L.: Allocating resources to control infectious diseases. *Operations Research and Health Care. ISORMS*, vol. 70, pp. 443–464. Springer, Boston, MA (2005). [https://doi.org/10.1007/1-4020-8066-2\\_17](https://doi.org/10.1007/1-4020-8066-2_17)
6. Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to track COVID-19 in real time. *Lancet. Infect. Dis* **20**(5), 533–534 (2020)
7. Duijzer, L.E., van Jaarsveld, W., Dekker, R.: Literature review: the vaccine supply chain. *Eur. J. Oper. Res.* **268**(1), 174–192 (2018)
8. Halper, R., Raghavan, S.: The mobile facility routing problem. *Transp. Sci.* **45**(3), 413–434 (2011)
9. Huang, H.C., Singh, B., Morton, D., Johnson, G., Clements, B., Meyers, L.: Equalizing access to pandemic influenza vaccines through optimal allocation to public health distribution points. *PLOS ONE* **12**, e0182720 (2017). <https://doi.org/10.1371/journal.pone.0182720>
10. Kaur, S.P., Gupta, V.: COVID-19 vaccine: a comprehensive status report. *Virus Res.* **288**, 198114 (2020)
11. Lee, E.K., Pietz, F., Benecke, B., Mason, J., Burel, G.: Advancing public health and medical preparedness with operations research. *Interfaces* **43**(1), 79–98 (2013)
12. Mahase, E.: COVID-19: Moderna vaccine is nearly 95% effective, trial involving high risk and elderly people shows. *Br. Med. J. (BMJ)* **371**, 4471 (2020)
13. Nicola, M., et al.: The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. *Int. J. Surg.* **78**, 185–193 (2020)
14. Park, H.S., Jun, C.H.: A simple and fast algorithm for K-medoids clustering. *Exp. Syst. Appl.* **36**(2), 3336–3341 (2009)
15. Polack, F.P., et al.: Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020)
16. Rachaniotis, N.P., Dasaklis, T.K., Pappis, C.P.: A deterministic resource scheduling model in epidemic control: a case study. *Eur. J. Oper. Res.* **216**(1), 225–231 (2012)
17. Ramirez-Nafarrate, A., Lyon, J.D., Fowler, J.W., Araz, O.M.: Point-of-dispensing location and capacity optimization via a decision support system. *Prod. Oper. Manag.* **24**(8), 1311–1328 (2015)
18. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987)
19. Xiong, J., et al.: Impact of COVID-19 pandemic on mental health in the general population: a systematic review. *J. Affect. Disord.* **277**, 55–64 (2020)