



Deploying vaccine distribution sites for improved accessibility and equity to support pandemic response

George Z. Li¹ · Ann Li^{2,3} · Madhav Marathe^{2,3} · Aravind Srinivasan¹ · Leonidas Tsepenekas¹ · Anil Vullikanti^{2,3}

Accepted: 9 July 2023 / Published online: 2 August 2023
© Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

In response to COVID-19, many countries have mandated social distancing and banned large group gatherings in order to slow down the spread of SARS-CoV-2. These social interventions along with vaccines remain the best way forward to reduce the spread of SARS CoV-2. In order to increase vaccine accessibility, states such as Virginia have deployed mobile vaccination centers to distribute vaccines across the state. When choosing where to place these sites, there are two important factors to take into account: accessibility and equity. We formulate a combinatorial problem that captures these factors and then develop efficient algorithms with theoretical guarantees on both of these aspects. Furthermore, we study the inherent hardness of the problem, and demonstrate strong impossibility results. Finally, we run computational experiments on real-world data to show the efficacy of our methods.

Keywords COVID-19 · Approximation algorithms · Facility location

✉ George Z. Li
gzli929@umd.edu

Ann Li
al7gc@virginia.edu

Madhav Marathe
marathe@virginia.edu

Aravind Srinivasan
srin@cs.umd.edu

Leonidas Tsepenekas
ltsepene@umd.edu

Anil Vullikanti
asv9v@virginia.edu

¹ Computer Science, University of Maryland, College Park, USA

² Computer Science, University of Virginia, Charlottesville, USA

³ Biocomplexity Institute, University of Virginia, Charlottesville, USA

1 Introduction

The COVID-19 pandemic continues to cause immense social, health, and economic impact globally. As of writing this paper, the U.S. alone has seen over 850,000 deaths and over 65 million confirmed cases; see [1] for the latest numbers. Vaccines have proven to be very effective in reducing the health burden of the pandemic and continue to be the best strategy to control disease spread and potentially end the pandemic in its current form. Despite the effectiveness, administering COVID-19 vaccines to all eligible individuals in the population continues to be a challenge. As of February 2022, only 64% of the eligible population is fully vaccinated in the United States [2]. Furthermore, there is a significant disparity in vaccination rates between demographics—the rate among Whites was 1.2 times that of African Americans and 1.1 times that of Hispanic people. The reasons why some people have not been vaccinated include distrust and skepticism regarding COVID-19, accessibility issues, and concerns about the cost [3]. Lottery schemes, mandates, vaccine clinics, and other strategies have been implemented to increase the vaccination rate with varying levels of success. Since cost and accessibility remain a challenge for a fraction of the population, especially minorities and people in poorer neighborhoods, mobile vaccine clinics have been an important part of the public health response strategy of government agencies. In this paper, we study the problem of deploying mobile vaccine administration sites with the goal of improving the accessibility of vaccines to individuals.

Deploying vaccination clinics is a form of a facility location problem [4, 5], which is a family of problems where we have largely autonomous agents—people in a population—who desire high-quality service for themselves (e.g., in the form of easily-accessible vaccination services); we balance this with the global constraint of a budget on the number of vaccination facilities, as well as through the notion of equity across the agents and the demographic groups they belong to. Our problem is referred to as the k -supplier problem, in which a limited set of k facilities needs to be placed so that every person (i.e., a client) is “close” to a facility; a common metric to measure closeness is the maximum distance between a client and their closest facility, though many other notions have been studied. Facility location problems are well understood, and efficient approximation algorithms and practical heuristics exist. However, deploying vaccine clinics leads to a novel facility location problem (referred to as the *MOBILEVACCCLINIC* problem) since people (clients) are mobile rather than stationary. Suppose each person p visits a set S_p of locations during the day; then it suffices to deploy a clinic close to at least one location in S_p . Our contributions are the following:

- We formalize the *MOBILEVACCCLINIC* problem for modeling the deployment of mobile vaccine clinics in a way that takes into account human mobility patterns (by considering the distance to a facility from any of the locations visited by a person), fairness (by requiring that at least a fraction of people in each demographic group have a nearby clinic), outliers (by allowing partial coverage), capacity constraints (by restricting the number of people assigned to each clinic), and fault tolerance (by requiring each client to be served by multiple facilities). We show that this problem is much harder than the standard k -supplier problem and getting any bounded polynomial-time approximation to the minimum distance is not possible, thus motivating bicriteria and fixed-parameter tractable algorithms.
- We design two approximation algorithms. The first is a fixed-parameter tractable algorithm that gives a 3-approximation, where the parameterization is on the number u of

locations where people travel. Note that even this is non-trivial, since the possible locations $\mathbf{S}$ where we can place facilities is still variable, so there are still $\binom{n}{k}$ possible solutions. The second algorithm, based on covering problems, is a $(1, \log n + 1)$ -bicriteria one, where n is the number of people. This means that if we violate the budget on the number of vaccine centers by a $\log n + 1$ multiplicative factor, we can find a solution that is optimal. Finally, we extend both of our algorithms to satisfy many practical considerations related to vaccine distribution including outliers, fairness, capacity; we also extend our bicriteria algorithm to be fault tolerant.

- We evaluate our algorithms for a realistic population of a county in Virginia. We find that our algorithms generally give a significant improvement over natural baselines, even when limiting the amount of travel information available to the algorithms. In particular, we see many shortcomings of only considering a client's home (rather than their entire travelling route), emphasizing the importance of our problem formulation. Additionally, our algorithms allow us to compute a tradeoff between the maximum distance to a clinic and the number of clinics; this naturally enables us to give a recommendation to the government on the most cost-effective budget policy. Finally, we find that the solutions computed by our algorithms have a useful “kernel” property—as the budget is increased, the locations which were picked for a lower budget are still part of the solution. This implies that an incrementally constructed solution (which is how such facilities would be deployed in practice since the budget is not known ahead of time) will still be good.

We remark that though our framework is motivated by the current COVID-19 pandemic, it can be generally applied to both epidemiological and non-epidemiological settings. Examples within healthcare include placing testing and treatment units (as deployed during the Ebola crisis) and delivering healthcare in rural settings for resource-limited countries. Beyond healthcare, the placement of mobile distribution centers arises in disaster-management settings. For instance, shelters need to be set up for individuals evacuating during a hurricane or forest fire, who might need food and other basic survival kits. During such large events, mobile sites are also used to place security posts and information kiosks.

2 Preliminaries

Recall that we wish to place vaccination centers such that vaccines are more accessible to the population. This question is often formulated as an appropriate variant of the facility location problem, which is well-studied in the operations research literature (see Related Work). In our paper, we introduce a new variant that follows a recent line of work on integrating the mobility patterns of the population into disease models [6–8]. As is standard, we will use the distance from a vaccination center as the metric for defining accessibility. The key change, however, is that clients will be represented by a set of locations that they visit (within a time period) instead of just one point. Though this will make the problem much harder to solve efficiently, it will more strongly correlate with the likelihood of a person going to a vaccine center (Fig. 1).

Problem Statement We are given a set of locations $\mathbf{C}$ in a metric space characterized by the distance function $d : \mathbf{C} \times \mathbf{C} \mapsto \mathbb{R}_{\geq 0}$. We additionally have a set of n individuals/clients $\mathbf{P}$. Each individual $p \in \mathbf{P}$ is associated with a set $S_p \subseteq \mathbf{C}$, which we can interpret

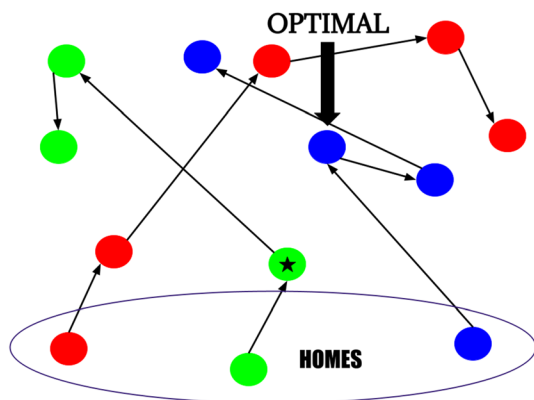


Fig. 1 An example of MOBILEVACCLINIC. The different colors represent different people and the circles represent the locations they visit (with the bottom three being their homes). In this case, the blue location in the middle is an optimal location to place a vaccination center. If we instead only considered homes in the problem formulation, we would place the vaccination center in the green circle marked with a star, which would require people to deviate from their normal travels much more when getting a vaccine (Color figure online)

as the set of locations p visits throughout the day. Finally, the input also includes a positive integer k constraining the number of facilities we can place, and a set $S \subseteq C$ containing the locations where we are allowed to place facilities. The goal of MOBILEVACCLINIC is to choose a set $F \subseteq S$ with $|F| \leq k$ to place facilities, such that for every $p \in P$ we have $d(S_p, F) \leq R$, for the minimum R possible. Here, we use the standard notation where $d(S, F) = \min_{j \in S, j' \in F} d(j, j')$. Intuitively, this objective tries to minimize the maximum distance between the set of facilities placed and the locations visited by any client. Since agents have varied preferences and behaviors, we also consider a few natural extensions in order to take these into account:

- *Outliers* in order to achieve herd immunity, we only need to vaccinate a large portion of the population (rather than every single person). Additionally, a few outliers in client set may have an disproportionate impact on the facility placement. In order to counteract this, we can take as input a parameter q , and seek to provide for only $\lfloor qn \rfloor$ of the clients, ignoring the remaining ones. Formally, the new objective is to minimize R such that $|\{p \in P : d(S_p, F) \leq R\}| \geq \lfloor qn \rfloor$.
- *Fairness* many studies have shown that COVID-19 disproportionately affects some demographic groups [9]. To counteract this, we seek to guarantee that different demographic groups have similar accessibilities to vaccines. As an example, when we solve the outliers formulation, we can guarantee that we are covering the same proportion of each demographic group when deciding the facility placements.
- *Capacity* it is natural to assume that the number of vaccines that can be stored in each mobile facility is limited. To take this into account, we consider the capacitated MOBILEVACCLINIC where take in as input the capacity L and need to guarantee that every chosen facility will have at most L people assigned to it.
- *Priority* some people have cars while some people can only walk or use public transportation, so traveling the same distance may require significantly different travel times. If we use the same notion of distance for every person, this would treat

people without cars unfairly. To counteract this, we consider a weighted (or priority) variant of the problem where our objective is to minimize $\max_{p \in P} w_p \cdot d(S_p, F)$. For people p who don't have cars, we can set their weight w_p to be higher to force facilities to be placed closer to these people.

- *Fault Tolerance* unpredictable emergencies may cause some facilities to be unable to provide vaccines (such as traffic accidents or family emergencies). In order to prevent such unforeseen circumstances from effecting whether or not the clients can be vaccinated, we can take an input α and require that each client is served by at least α facilities within radius R , for the minimum R possible.

3 Related work

Due to its applications in a large number of domains, facility location and broader location theory is a very well-studied area; see, e.g., the surveys [4, 5, 10]. The general goal in this family of problems is to deploy facilities to provide the best possible service to a set of clients. A huge number of objectives have been considered, along with a plethora of variations such as fairness variants and online or stochastic versions. The MOBILEVACCCLINIC problem we study here is a generalization of the well-known k -center problem, where the goal is to open at most k centers while minimizing the maximum distance of a point to its closest center. For this simple clustering setting, there exist efficient 2-approximation algorithms [11, 12]. Furthermore, it is shown that unless $P=NP$ this is the best achievable approximation ratio [13].

In a recent related work [8], we studied the problem from a practical perspective and gave data driven heuristics for real-time placement of distribution sites. The key idea in that work is the use of near-real-time mobility data to decide dynamic placement of the sites. The focus of this work was to provide strategies that take into account weekly changes in traffic patterns. One of the baselines we compare to in Sect. 6 is inspired by the work in this paper. Recently, other authors have also studied the facility location problem for vaccine distribution; see [14–18]. The paper by [16] considers a possible placement of the distribution sites at CVS stores across a region. The work by [14] considers the problem at a national level and is based on simulating the disease progression using a compartmental model combined with placement algorithm. Finally, the work of [17] is similar to our work in [8] in that it develops data driven strategies for allocating vaccine distribution sites.

Location theory problems have also been considered in the more general area of healthcare, e.g., [10, 19–21]. A lot of this work has been focused on placing mobile clinics or temporary facilities to ensure good service, especially in resource-poor countries. As mentioned in [10], the healthcare domain poses new challenges for location theory, such as uncertainty, reliability, operation efficiency, patient safety, and cost-effectiveness. Overall, prior work has generally not considered the mobility of clients at a detailed scale, which provides more flexibility in deploying facilities. Our formulation of MOBILEVACCCLINIC explicitly models human mobility, thus providing a realistic framework for public health agencies in their response efforts. In concurrent and independent work, [15] formulate a related problem to MOBILEVACCCLINIC with a slightly different objective; however, they approach the problem heuristically via a mixed-integer linear programming approach so our results are largely complementary.

4 Hardness result

For our hardness result, we use the following problem studied in [22], called γ -Colorful k -Center or γ CkC for short. This problem is a generalization of the outliers version of k -center — in addition to the classical constraints, colors (representing demographic groups) are assigned to each client and the problem requires that a sufficient number of points of each color is covered. The formal definition is given below:

Definition 1 Let $\gamma \in \mathbb{Z}_{\geq 1}$ be the number of colors, $k \in \mathbb{Z}_{\geq 1}$ be the budget, $\mathbf{C}$ be a set of points in a metric space, and $d : \mathbf{C} \times \mathbf{C} \rightarrow \mathbb{R}_{\geq 0}$ be the distance function on $\mathbf{C}$. For each $\ell \in [\gamma]$, let $\mathbf{C}_\ell \subseteq \mathbf{C}$ be the points with color ℓ and let $m_\ell \in \mathbb{Z}_{\geq 1}$ be the number of points with color ℓ which need to be covered. γ CkC asks for the minimum radius R together with a set $F \subseteq \mathbf{C}$ with $|F| \leq k$, such that at least m_ℓ points of $\mathbf{C}_\ell$ are covered within distance R by F . Formally, if $B(F, R) = \{j \in \mathbf{C} : d(j, F) \leq R\}$ then we want $|B(F, R) \cap \mathbf{C}_\ell| \geq m_\ell$ for every $\ell \in [\gamma]$.

In [22] the authors prove the following hardness result, which we use to prove a hardness result for our problem later on.

Lemma 1 When γ is not a constant, there exist instances of γ CkC with $m_\ell = 1$ for all $\ell \in [\gamma]$, such that if R^* is the optimal value of the instance, the following hold:

- For any $\rho > 0$, it is NP-hard to find $F \subseteq \mathbf{C}$ with $|F| \leq k$ and $|B(F, \rho R^*) \cap \mathbf{C}_\ell| \geq m_\ell$ for all ℓ . In words, it is NP-hard to devise any approximation algorithm for γ CkC.
- For any $\rho > 0$ and $\epsilon \in (0, 1)$, it is NP-hard to find $F \subseteq \mathbf{C}$ with $|F| \leq (1 - \epsilon) \ln \gamma \cdot k$ and $|B(F, \rho R^*) \cap \mathbf{C}_\ell| \geq m_\ell$ for every $\ell \in [\gamma]$ (i.e., it is NP-hard to devise any bicriteria approximation for γ CkC, whose chosen centers will be at most $(1 - \epsilon) \ln \gamma \cdot k$).
- These problematic instances consist of points on a line.

Remark 1 Regarding the second statement in Lemma 1, the authors of [22] show a bicriteria hardness result in terms of $\log |\mathbf{C}|$ and not $\ln \gamma$. However, a closer look into their proof reveals that the claim mentioned above follows trivially. We choose to present this form of the claim because it better fits our narrative later on in the paper.

Theorem 2 There exists a bicriteria preserving reduction of the problematic instances of γ CkC described in Lemma 1 to instances of MOBILEVACCCLINIC with $|\mathbf{P}| = \gamma$. Specifically, any (ρ, α) -bicriteria approximation for MOBILEVACCCLINIC translates to a (ρ, α) -bicriteria approximation for the problematic instances of γ CkC.

Proof Let $(\mathbf{C}, \mathbf{C}_1, \dots, \mathbf{C}_\gamma, k, m_1, \dots, m_\gamma)$ be a problematic instance of γ CkC as described in Lemma 1, and recall that this instance has $m_\ell = 1$ for all $\ell \in [\gamma]$. We will now construct an instance of MOBILEVACCCLINIC as follows. The metric space for MOBILEVACCCLINIC will be the same as in the γ CkC problem. That is, we assume we have points $\mathbf{C}$ with a distance function d on them. For every $\ell \in [\gamma]$ construct a client p_ℓ , and set $S_{p_\ell} = \mathbf{C}_\ell$. The set of locations $\mathbf{S}$ for MOBILEVACCCLINIC where we can place facilities will be the set of locations $\mathbf{C}$ of γ CkC, and the value k will stay the same for the two problems.

Consider now the optimal solution F^* of the γCkC instance and its corresponding value R^* . We claim that F^* is a feasible solution for the constructed MOBILEVACCLINIC instance, and its value for that is exactly the same. This is easy to see because $|F^*| \leq k$, $|B(F^*, R^*) \cap C_\ell| \geq 1$ for every $\ell \in [\gamma]$, and C_ℓ are exactly the locations visited by client p_ℓ . Hence if R_{OPT} is the value of the optimal solution to the the constructed MOBILEVACCLINIC instance, we have $R_{OPT} \leq R^*$.

Take now any (ρ, α) -bicriteria solution F for MOBILEVACCLINIC . At first we trivially have $|F| \leq \alpha k$. Moreover, for every ℓ we can express $d(F, S_{p_\ell}) \leq \rho R_{OPT}$ (the condition guaranteed by the (ρ, α) -bicriteria solution F for MOBILEVACCLINIC) since we know $|B(F, \rho R_{OPT}) \cap C_\ell| \geq 1$. Finally, because $R_{OPT} \leq R^*$, we have $B(F, \rho R_{OPT}) \subseteq B(F, \rho R^*)$. Hence, we can conclude $|B(F, \rho R^*) \cap C_\ell| \geq |B(F, \rho R_{OPT}) \cap C_\ell| \geq 1$ for every $\ell \in [\gamma]$. The latter completes the bicriteria preserving reduction. $\square$

Corollary 3 *Even when the metric space is the Euclidean line, we have the following for MOBILEVACCLINIC (unless $P=NP$):*

1. *No approximation algorithm exists.*
2. *Any bicriteria approximation algorithm must use at least $k \ln n$ facilities.*

5 Algorithms

In this section, we introduce efficient methods which give (approximately) optimal facility placements, despite the hardness results. We also show how to extend each of our algorithms to incorporate outliers, fairness constraints, capacity constraints, and fault-tolerance.

5.1 Fixed-parameter tractability

Let $U = \bigcup_{p \in \mathbf{P}} S_p$ denote the set of all the locations visited by the set of clients and $u = |U|$ be the number of locations in this set. Due to potential privacy concerns, we can assume that the client locations we have access to only include large public areas in the county such as malls, shopping centers, etc. Hence, it is reasonable to conclude that u is a fixed parameter, which we assume ranges from 15 – 30. Given this fixed parameter, we develop an efficient algorithm for our problem.

The main observation here is the following: consider an instance of MOBILEVACCLINIC and let F^* be its optimal solution, whose maximum radius we denote by R^* . For each $p \in \mathbf{P}$, we know that $d(F^*, S_p) \leq R^*$, and hence there must exist a location $i_p \in S_p$ with $d(i_p, F^*) \leq R^*$. See now that $\{i_p \mid p \in \mathbf{P}\} \subseteq U$, and therefore $|\{i_p \mid p \in \mathbf{P}\}| \leq u$. The latter implies that we can guess, via an exhaustive search, the set $\{i_p \mid p \in \mathbf{P}\}$ in time at most 2^u (recall that since u is considered a fixed parameter, 2^u is thought of as a small constant). Let A be the correct guess for that set; we can think of A as the set of locations through which the optimal solution covers every client within distance R^* . Given A , we see that the problem of computing F^* reduces in a straightforward manner to the well-known k -supplier problem [11].

In k -supplier we have a set of points $\mathbf{X}$ and a set of locations $\mathbf{Y}$ in a metric space with distance function d . The goal is to choose $C \subseteq \mathbf{Y}$ with $|C| \leq k$, such that the maximum distance of any point in $\mathbf{X}$ to its closest location of S is minimized. Hence, after correctly

guessing A , we create an instance of k -supplier where the points are the ones in A , and the set of locations $\mathbf{Y}$ is $\mathbf{S}$. The previous discussion shows that F^* is a solution for this k -supplier instance, and its maximum radius will again be R^* . Moreover, any ρ -approximate solution to the k -supplier instance will trivially be a ρ -approximate solution for MOBILEVACCCLINIC. Using the 3-approximation algorithm from [11] proves the following theorem.

Theorem 4 *Algorithm 1 yields a 3-approximation algorithm for MOBILEVACCCLINIC and runs in time $2^u \text{poly}(n, |\mathbf{C}|)$.*

Algorithm 1 FPT

- 1: **for** $A \in 2^U : |A \cap S_p| \neq 0, \forall p \in \mathbf{P}$ **do**
 - 2: Obtain locations F_A by running the k -supplier algorithm on the appropriate instance discussed above.
 - 3: Calculate the objective value for F_A .
 - 4: **end for**
 - 5: Pick the F_A with the smallest objective value.
-

Moving forward, we see that the same approach of guessing the correct set of client locations A will also apply in different settings. In fact, the only thing that may differ is the need for an alternative k -supplier algorithm that can incorporate the specific constraints of each unique setting; we survey some of these settings below.

Outliers To modify our algorithm so that it only considers some fraction q of the population, we only need to change the objective value evaluated in line 3 of Algorithm 1. To improve efficiency, we can also only consider guesses A that contain locations from at least $\lfloor qn \rfloor$ clients since the correct guess A contains locations from at least $\lfloor qn \rfloor$ clients. If we then feed A to the k -supplier algorithm in the exact same manner as before, we will get a 3-approximation.

Corollary 5 *After changing the objective evaluated in line 3 to the partial objective, Algorithm 1 gives a 3-approximation for MOBILEVACCCLINIC with outliers.*

Fairness Although our algorithm provides an upper bound guarantee for the maximum distance to a facility, the facility placement may significantly differ between individuals, with some having a facility right next to them, while others need to travel the whole $3R^*$ guarantee. Luckily, the vaccine centers can vary from week to week or even day to day. Thus, we can use a randomized algorithm such as the one given in [23], to guarantee that the re-provisioning of facilities over the course of many tries will provide an improved per-point guarantee on expectation. Hence, we treat the clients stochastically fairly.

Corollary 6 *When using the algorithm from [23] for k -supplier, Algorithm 1 outputs a distribution $\tilde{\Omega}$ such that $\forall p \in \mathbf{P}$, we have $\mathbb{E}_{F \sim \tilde{\Omega}}[d(S_p, F)] \leq (1 + 2/e)R^*$ and $\Pr[d(S_p, F) \leq 3R^*] = 1$.*

Capacity In this case, we assume that each facility we use has a capacity L , i.e., at most L clients can be assigned to it in any solution. Once again, the FPT process we described earlier suffices to solve the problem. Specifically, we can think of the set A as the locations through which the clients of $\mathbf{P}$ receive their service. Hence, as we did for the regular case, we can create an instance of k -supplier where the points requiring service are those of A , but this time each location of $\mathbf{Y}$ for the k -supplier instance will have a capacity L . In other words, this will be an instance of capacitated k -supplier. Furthermore, the optimal solution of capacitated MOBILEVACCCLINIC will be a solution of the same value for the capacitated k -supplier instance. Finally, it is also trivial to see that any ρ -approximate solution for capacitated k -supplier instance, will yield a ρ -approximate solution to capacitated MOBILEVACCCLINIC.

Corollary 7 *When using the algorithm from [24] instead of a simple k -supplier algorithm, Algorithm 1 is an 11-approximation for capacitated MOBILEVACCCLINIC.*

5.2 Covering algorithm

In Corollary 3, we show that any bicriteria algorithm needs to open at least $k \ln n$ facilities in order to give a bounded approximation guarantee. Here, we show that this is essentially tight: we give an algorithm that outputs a set of locations of size at most $k(\ln n + 1)$, while guaranteeing that the objective value is at most that of an optimal solution.

Consider the related problem, which we call CLIENTCOVER, in which instead of optimizing the radius R given a budget k , we are given a target radius R and want to choose a set $F \subseteq \mathbf{S}$ which minimizes $|F|$ and guarantees that $d(S_p, F) \leq R$ for each $p \in \mathbf{P}$. Notice that this is just a standard Set Cover problem, where the sets are $\{p \in \mathbf{P} : d(S_p, j) \leq R\}$ for each $j \in \mathbf{S}$ and the universe consists of the clients $\mathbf{P}$. Using a known greedy algorithm for Set Cover [25], we have an H_n -approximation algorithm for CLIENTCOVER, where $H_n \leq \ln n + 1$ is the n -th harmonic number.

For generality, we will show how any α -approximation algorithm for Set Cover yields an $(1, \alpha)$ -bicriteria algorithm for MOBILEVACCCLINIC via a reduction to CLIENTCOVER. First, note that the optimal radius R^* for an instance of MOBILEVACCCLINIC is always the distance between some $j \in \mathbf{C}$ and some $i \in \mathbf{S}$. Hence, there are at most polynomially many options for it, specifically $|\mathbf{C}| \cdot |\mathbf{S}|$. For each such option R , we create the corresponding instance of CLIENTCOVER and run the set cover algorithm on it. The final guarantees follow from the iteration when $R = R^*$. Observe at this point that we can speed up the whole process by performing a binary search in order to find R^* , and thus avoid the previously described exhaustive search.

Algorithm 2 CLIENTCOVER Search

- 1: Binary search on the sorted list $\{d(i, j) : j \in \mathbf{C}, i \in \mathbf{S}\}$, and let the current guess be R :
 - 2: Use R to create the proper instance of CLIENTCOVER.
 - 3: Obtain α -approximate solution F_R for that instance.
 - 4: If $|F_R| > \alpha \cdot k$, increase R ; else, decrease R .
 - 5: Output F_R for the minimum R such that $F_R \leq \alpha \cdot k$.
-

Theorem 8 *Given an α -approximation algorithm for set cover, Algorithm 2 gives an $(1, \alpha)$ -bicriteria algorithm for MOBILEVACCCLINIC.*

Proof Let R^* be the objective value for the optimal solution $F^* \subseteq \mathbf{S}$ of MOBILEVACCCLINIC, where $|F^*| \leq k$. We wish to show that CLIENTCOVER Search finds a radius R in the list such that $|F_R| \leq \alpha \cdot k$ and $R \leq R^*$. Consider an iteration of the binary search where the radius guess is R . Suppose $R \geq R^*$; then there must exist a solution of CLIENTCOVER of size at most k . The α -approximation algorithm will therefore output a set F_R with $|F_R| \leq \alpha \cdot k$ and R will decrease. If $R < R^*$, then we either find a solution with $|F_R| \leq \alpha \cdot k$, or we increase R and move closer to R^* . Finally, since R^* is in the list $\{d(i, j) : j \in \mathbf{C}, i \in \mathbf{S}\}$, the binary search necessarily finds some $R \leq R^*$ with $|F_R| \leq \alpha \cdot k$. $\square$

As in the case of our FPT algorithm, we can easily extend Algorithm 2 in order to accommodate different settings. The only difference here lies at step 3, where instead of a classic Set Cover algorithm we can run a different algorithm.

Outliers In order to modify our algorithm to only consider some fraction $q \in (0, 1)$ of the population, we can use some α -approximation algorithm for the Partial Set Cover problem, where the goal is to cover at least a q -fraction of the universe elements. Hence, we naturally consider a variant of CLIENTCOVER, which we call Partial CLIENTCOVER, that requires only $\lfloor qn \rfloor$ points to be covered by balls of radius R^* . Trivially, Partial CLIENTCOVER is a special case of Partial Set Cover. Then the approach we described previously remains the same: we can guess the optimal radius R^* and obtain an α -approximate solution F_{R^*} for the corresponding Partial CLIENTCOVER instance. This solution will be optimal for MOBILEVACCCLINIC with outliers, while placing at most αk facilities. In particular, we have the following corollary of Theorem 8.

Corollary 9 *When using the greedy algorithm for Partial Set Cover [25], Algorithm 2 gives a $(1, H_{\lfloor qn \rfloor})$ -bicriteria algorithm for MOBILEVACCCLINIC with outliers.*

Fairness When solving MOBILEVACCCLINIC with outliers, the algorithm may view some demographic groups as outliers more often than others. To mitigate such possibilities, we can use an algorithm for the Partition Set Cover problem [26] to guarantee that a large proportion of each demographic group gets coverage. Formally, the Partition Set Cover problem takes in as input a Set Cover instance $(\mathbf{U}, \mathbf{S})$ and r (not necessarily disjoint) color classes and requires that p_i elements in each color class is covered. For example, we can guarantee that the algorithm considers a proportional number of people from each (demographic) group when choosing the vaccine center locations. The following approximation guarantee will then follow directly from [26] and the outliers reduction before.

Corollary 10 *Let $C_t \subseteq \mathbf{P}$ for $t \in [r]$ be (not necessarily disjoint) demographic classes and let $0 \leq p_t \leq |C_t|$ be the coverage requirements for each class. Using the algorithm of [26] at step 3, Algorithm 2 gives a $(1, O(\log n) + \log r)$ -bicriteria algorithm while satisfying the coverage constraints.*

Capacity As before, we assume that each facility we use has capacity L . We see that our general framework is still applicable: we can modify our algorithm to satisfy these capacity constraints by replacing the Set Cover algorithm in CLIENTCOVER with an algorithm for Capacitated Set Cover, which requires each set in the universe to only be assigned to cover

L elements. It is clear that the capacity constraints on the sets in Capacitated Set Cover corresponds to the capacity constraints on facilities in our problem. Thus, the greedy algorithm given in [27], which gives a H_n -approximation algorithm in this more general case, directly yields to the following corollary:

Corollary 11 *When using the greedy algorithm given in [27] for Capacitated Set Cover, Algorithm 2 gives a $(1, H_n)$ -bicriteria algorithm while satisfying the capacity constraints.*

Priority To take into account the weights for the priority version of our problem, we can still use the general framework of CLIENTCOVER. In Line 1 of Algorithm 2, we now need to binary search on $\{d(i, j) \cdot w_p : j \in \mathbf{C}, i \in \mathbf{S}, p \in \mathbf{P}\}$ to guess the optimal (weighted) radius. Given a guess R , we now construct the corresponding set system F_R as follows: for each facility location $i \in \mathbf{S}$, add the set $\{p \in \mathbf{P} : d(S_p, j) \leq R/w_p\}$ to the set system. It is easy to see that when $R = R^*$ is guessed correctly, we can cover the clients using exactly k facilities. Thus, we have the following corollary with a proof analogous to that of Theorem 8:

Corollary 12 *When creating the set cover instance as described above, Algorithm 1 is a $(1, H_n)$ -approximation for priority MOBILEVACCLINIC.*

Fault Tolerance To make sure each client is served at $\alpha \leq k$ different facilities during its travels, we can continue using the general CLIENTCOVER framework. We create the set cover instance as before, but now our requirement is that each element is covered at least α times by some facility. This corresponds to the problem of Set Multi-Cover, which additionally takes in as input the number of times each element needs to be covered. Using the standard greedy algorithm for this problem [28], we get the same guarantees for Fault-Tolerant MOBILEVACCLINIC:

Corollary 13 *When using the greedy algorithm for Set Multi-cover [28], Algorithm 1 is a $(1, H_n)$ -approximation for fault-tolerant MOBILEVACCLINIC.*

Budget In the previous algorithms which solve CLIENTCOVER as a subroutine, we violate the budget constraint k on the number of facilities by a non-trivial multiplicative factor, which is a practical consideration that needs to be addressed. Luckily, it has been shown that the greedy algorithm and other heuristics for Set Cover have very small approximation ratios in practice [29]. In fact, many real-life instances of Set Cover are solved optimally or near optimally by the greedy algorithm [30]. Given this empirical result (which we also validate for our instances of the Set Cover problem), we get $\alpha = 1$ in our experiments when running CLIENTCOVER Search. In particular, if we solve the CLIENTCOVER problem using a commercial mixed-integer linear program (MILP) solver [31, 32], we can solve the original problem to optimality. We emphasize that this is a non-trivial contribution: directly formulating MOBILEVACCLINIC as an MILP requires $\Theta(n^3)$ constraints, and we cannot even initialize the solver using or-tools. In contrast, the Set Cover MILP only has $\Theta(n)$ constraints, which can be solved efficiently using commercial solvers. Hence, when using an MILP solver to solve CLIENTCOVER, Algorithm 2 yields a practical solution for solving MOBILEVACCLINIC optimally.

Table 1 Network information

	Clients	Activity locations	Residential locations	Maximum activity	Measured diameter (km)
Charlottesville City	33,156	5660	10,038	9952	8.12
Albemarle County	74,253	9619	32,981	24506	61.62

6 Experiments

6.1 Experimental setup

Data We run our experiments on the mobility data from Charlottesville City and Albemarle County in Virginia. For these counties, we use synthetic data constructed from the 2019 U.S. population pipeline (see [33, 34] for details). This dataset was constructed by tracking the week-long activity of county residents. Each resident is represented by a sequence of activities, where each activity is described by duration, type, and location in the county. The locations are given in geodetic coordinates and are categorized as either a residential or activity (non-residential) location. From this dataset, we can extract the locations visited by individuals residing in the county and set all activity locations as potential facility placements. A summary of the dataset is given in Table 1.

Baselines We compare our algorithms with two heuristics: HOMECENTERS and MOSTACTIVE. In MOSTACTIVE, we open vaccination centers at the k most visited locations. We set MOSTACTIVE as the baseline because it is related to the current heuristic used by the Virginia Department of Health. In HOMECENTERS, we run k -supplier to place facilities at locations that minimize the maximum distance from client homes. We compare with this baseline in order to show the importance of considering mobility when placing the vaccination centers.

Objective Recall that our objective is to minimize the maximum distance any client needs to deviate from their path to reach some facility. Since our location data is given in the geographic coordinate system, we approximate the Earth as a sphere and use geodesic distance as our metric. In Sect. 6.2, we notice that there is a sharp drop in the objective value if we only consider 99% of the population. As a result, we also evaluate the objective value of our algorithms when 5% of the people are considered outliers.

FPT details When using FPT in our experiments, we pick $u = 15$ locations that cover the largest portion of the population (as given by the greedy algorithm for the Maximum Coverage problem). We then run FPT using only knowledge of these u locations. The locations chosen are all popular public activity locations, so we have a limited amount of privacy violation. As a result, the performance of FPT is weaker on the full objective, but remains strong on partial coverage (the outliers formulation). It is important to note that even though we limit the knowledge of client-visited locations, FPT can still choose to place facilities at any activity location in the dataset. For more details on our implementation of FPT and the experiments, see our GitHub¹.

¹ <https://github.com/gzli929/MobileVaccClinic>.

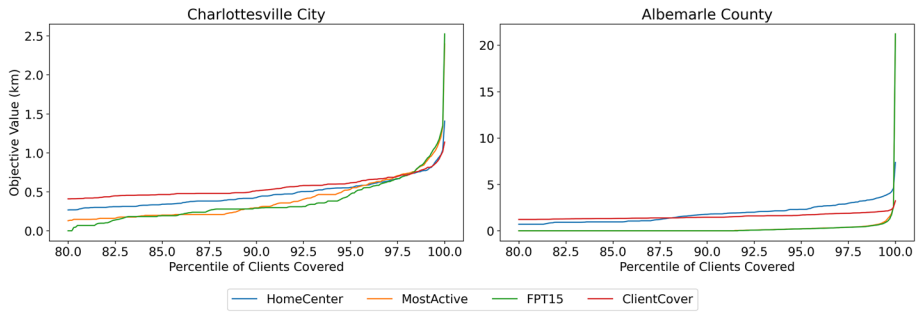


Fig. 2 A comparison of the algorithms and baseline heuristics for various coverage requirements $p \in [0.8, 1.0]$ in the outliers objective of our problem. The budget is set to 10 for Charlottesville and 20 for Albemarle

6.2 Client coverage performance analysis

We begin by directly comparing the performances between our algorithms and the baselines. Because our objective value is defined by the maximum distance any client must travel to reach their closest facility, it does not give good insight into the distribution of travel distances for the entire set of clients. For this reason, we also assess how the service cost for the closest p -fraction of the clients (i.e., the outliers objective for our solution) varies when we let p range between 0.8 and 1.0.

As seen in Fig. 2, facility placements from HOMECENTERS and CLIENTCOVER result in better full objectives while facility placements from MOSTACTIVE and FPT result in better partial objectives. This has a simple explanation: the former two algorithms are forced to provide service to outliers since they directly optimize the objective while the latter two optimize over only a portion of the population by design. As a result, it makes sense to compare HOMECENTERS with CLIENTCOVER and FPT with MOSTACTIVE in the subsequent experiments. We note that if we instead used the outliers version of HOMECENTERS and CLIENTCOVER, we may not see this disparity.

6.3 Tradeoff between radius and budget

In addition to evaluating the performance of our algorithms at the current budget, it is important to evaluate the sensitivity of our algorithms to an increase in budget. That is, we want to know how much the objective value would decrease if the county allocated more resources to deploy a greater number of mobile facilities. This knowledge can influence policy decisions: when an increase budget yields a sharp decrease in objective, the government has more incentive to fund additional vaccination centers.

As seen in Fig. 3, there is generally a sharp decrease in the objective value when the budget is less than 6 facilities for Charlottesville and 9 facilities for Albemarle. As the budget increases past those thresholds, the marginal returns become so diminished that increasing the budget hardly changes the objective value. This is especially prominent in

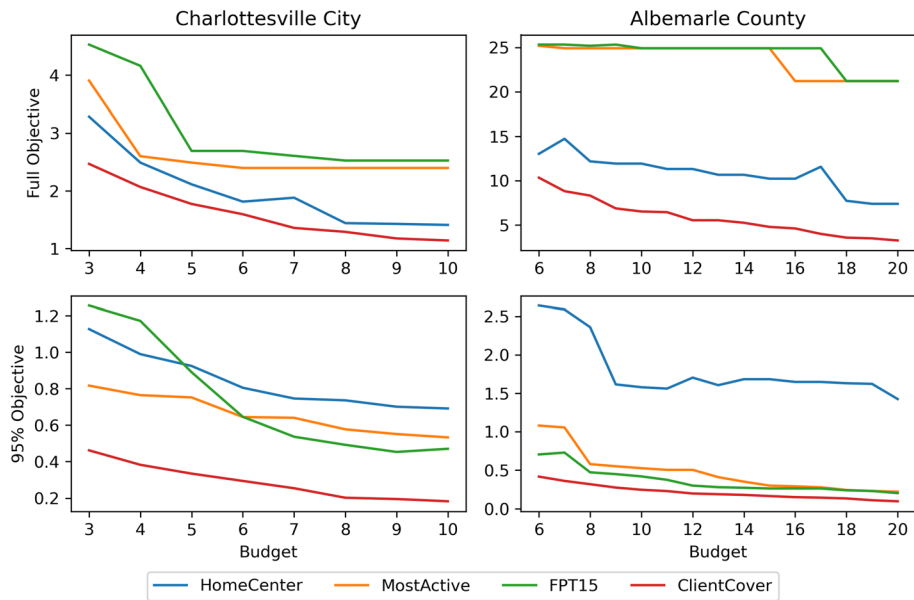


Fig. 3 Tradeoff between maximum distance needed to travel and the number of vaccine centers placed. In the 95% objective plots, CLIENTCOVER and HOMECENTERS are run with the outliers formulation

the full objective performance of FPT and MOSTACTIVE. Hence, it is natural to recommend budgets of 6 and 9 facilities to the Charlottesville and Albemarle government, respectively.

A seemingly weird result from the experiment is the tradeoff curve for HOMECENTERS. Though there is a general downward trend in the objective value as the budget increases, there are cases in each county where increasing the budget results in an increase in the objective value. This contradictory phenomenon is caused by the limited correlation between the distance to homes and our objective; as a result, noise/luck has a considerable effect. The noisiness of HOMECENTERS emphasizes the importance of our work of modeling mobile populations.

6.4 A kernel property

Through our experiments, we notice a nice (empirical) property of the vaccine center locations selected by some of our algorithms. Imagine a case where we (the government) have the funds to place five mobile vaccine centers and we use our algorithms to pick the five locations to place them. Then, after two weeks, the government decides that the disease is causing too much economic devastation and, in turn, funds three more mobile vaccine centers. When we ask our algorithms to place the eight vaccine centers (approximately) optimally, it turns out that the eight chosen locations will often contain the original five chosen locations as a subset. The original five locations are then called a kernel.

In order to determine the presence of a kernel for each of our algorithms, we calculate the number of facilities chosen with budget $k - 1$ that are not also chosen with budget k . These values populate Tables 2 and 3, where the leftmost column denotes the budgets compared. By definition, MOSTACTIVE has the kernel property since it is a greedy algorithm.

Table 2 Investigating the kernel property for the four algorithms on Charlottesville

	Most active	Home center	TPF	Client cover
3 → 4	0	1	1	3
4 → 5	0	1	1	3
5 → 6	0	2	1	5
6 → 7	0	3	2	6
7 → 8	0	0	1	5
8 → 9	0	0	1	8
9 → 10	0	3	1	6

Table 3 Investigating the kernel property for the four algorithms on Albemarle

	Most active	Home center	FPT	Client cover
6 → 7	0	3	1	5
7 → 8	0	1	2	6
8 → 9	0	1	2	7
9 → 10	0	1	1	8
10 → 11	0	4	3	5
11 → 12	0	1	1	9
12 → 13	0	1	2	4
13 → 14	0	0	1	7
14 → 15	0	5	2	7
15 → 16	0	1	1	8
16 → 17	0	1	0	12
17 → 18	0	7	0	8
18 → 19	0	2	1	7
19 → 20	0	0	0	8

Our FPT algorithm also (approximately) satisfies the kernel property while maintaining a stronger performance than the baseline. The remaining two algorithms do not exhibit the property: both HOMECENTERS and CLIENTCOVER pick (almost) completely different locations upon increasing the budget. Because they require less relocation, MOSTACTIVE and FPT have advantageous properties when the budget is adaptive and vaccine distribution is time-consuming.

We recognize that this is not necessarily applicable to our experimental setting, COVID-19, since transportation of vaccines is (relatively) easy in Virginia. However, for the Ebola outbreak in 2014, the kernel property was recognized as an important property to have since vaccine distribution was a much more costly process. Furthermore, we note that our algorithms are not explicitly designed to have this property; it is only empirically verified.

6.5 Information constraints

In our previous experiments with CLIENTCOVER, we assumed that we had full knowledge of the locations each person visited throughout a day. Next, we wish to understand how fine-grained this data needs to be in order for CLIENTCOVER to outperform our other algorithms; this also addresses privacy concerns raised when using the exact mobility data of

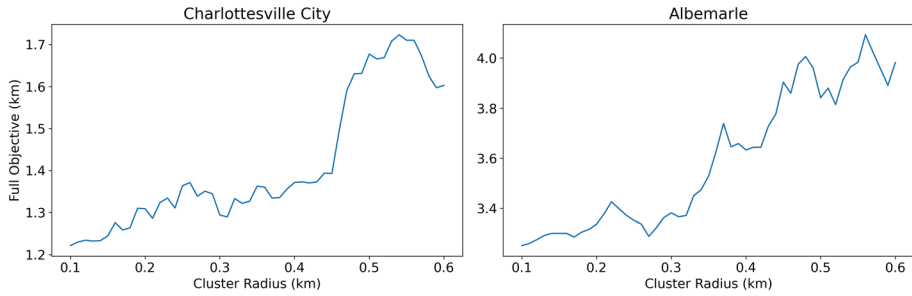


Fig. 4 Analyzing how CLIENTCOVER degrades as we lose fine-grained information about travel patterns. The budget is set to 10 for Charlottesville and 20 for Albemarle

individuals. In order to model loss of detailed movement patterns, we cluster the locations within a given radius r together and apply CLIENTCOVER on the resulting cluster centers. The details of the clustering algorithm can be found in our code, but the general idea is to define each location to be a potential cluster center and then use the greedy set cover algorithm to pick a minimum set of clusters centers that cover all original locations with radius r . Using this clustering method for both Charlottesville and Albemarle, we vary the clustering radius r between 100 and 600 ms to see how much privacy CLIENTCOVER can preserve while maintaining a superior performance over the baselines.

As we see in Fig. 4, our algorithms gradually degrade in effectiveness as the data becomes more and more coarse-grained with a sharp loss in objective for Charlottesville and Albemarle at a clustering radius of 0.48km and 0.35km, respectively. Even with the approximate travel data, the solutions produced have significantly smaller objective value than that of MOSTACTIVE at 2.39km and FPT at 2.52km. Furthermore, even by clustering the data with a radius that is 45% of the original CLIENTCOVER objective, we can still perform better than the HOMECENTERS baseline. From this experiment, we can conclude that even when giving reasonable privacy to individuals, CLIENTCOVER still performs much better than FPT, MOSTACTIVE, and HOMECENTERS.

6.6 Only using home and work

In the previous subsection, we observed that CLIENTCOVER was successful in preserving a reasonable amount of client privacy while maintaining a strong performance in objective value. We will now explore if this is still true in the extreme case when only the home and work location data are available when running CLIENTCOVER. In general, the location of your home and work are known to the government so this doesn't constitute a strong privacy violation. Surprisingly, we see in Fig. 5 that CLIENTCOVER still (relatively) consistently outperforms HOMECENTERS even when using very little additional information. This suggests that any additional mobility information can play a large role in reducing the distance clients need to travel to obtain a vaccine. In particular, this greatly highlights the importance of our problem formulation: our CLIENTCOVER algorithm can take advantage of any amount of additional travel data to outperform the standard k -supplier algorithm.

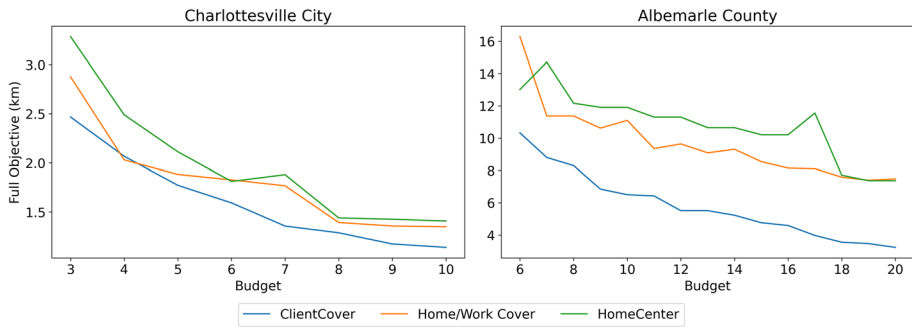


Fig. 5 Observing the performance of CLIENTCOVER when only using home and work locations of clients. We compare this with the full information CLIENTCOVER and HOMECENTERS under various budgets Charlottesville and Albemarle

7 Conclusion

We introduced a generalization of the classical k -supplier problem where we consider the mobility of populations when placing facilities. We showed that designing an approximation algorithm for this variant is NP-Hard. Fixed-parameter tractability and bicriteria approximation algorithms were studied to overcome the intractability results. Finally, we carried out computational experiments to demonstrate that the algorithms do well in practice as compared to natural baseline algorithms.

We conclude with a few directions for further research. First, having demonstrated the importance of modeling mobile populations, a natural next step is to extend other variants of the facility location problem to the mobile population setting as well. Second, in practice individuals have weekly schedule and thus the visits to the locations follow temporal pattern; formulations that capture that naturally and obtain provable approximations will be of interest. Finally, it would be interesting to explore how the framework can be extended to make more realistic choices for agents. For example, agents can decide which distribution centers to visit based on their personal profiles and the policymakers, with limited knowledge of the agents' preferences, needs to choose the facility locations to maximize the number of vaccinations.

Acknowledgements We express our sincere thanks to the referees for suggesting the experiments in Sect. 6.5 and the extension of capacity constraints. We also thank members of the Biocomplexity COVID-19 Response Team and the Network Systems Science and Advanced Computing (NSSAC) Division for their thoughtful comments and suggestions related to epidemic modeling and response support. George Li, Aravind Srinivasan, and Leonidas Tsepenekas were supported in part by NSF award number CCF-1918749. Ann Li, Madhav Marathe, and Anil Vullikanti were supported by DTRA (Contract HDTRA1-19-D-0007), University of Virginia Strategic Investment Fund award number SIF160, National Institutes of Health (NIH) Grants 1R01GM109718, 2R01GM109718, OAC-1916805 (CINES), CCF-1918656 (Expeditions), CNS-2028004 (RAPID), OAC-2027541 (RAPID), IIS-1908530, IIS-1955797, and IIS-2027848. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Author Contributions AS and MM and AV came up with the problem formulation. GL and LT designed the algorithms. LT proved the hardness results. GL and AL wrote the code and generated the figures. All authors contributed to writing and reviewing the manuscript.

Funding George Li, Aravind Srinivasan, and Leonidas Tsepenekas were supported in part by NSF award number CCF-1918749. Ann Li, Madhav Marathe, and Anil Vullikanti were supported by DTRA (Contract

HDTRA1-19-D-0007), University of Virginia Strategic Investment Fund award number SIF160, National Institutes of Health (NIH) Grants 1R01GM109718, 2R01GM109718, OAC-1916805 (CINES), CCF-1918656 (Expeditions), CNS-2028004 (RAPID), OAC-2027541 (RAPID), IIS-1908530, IIS-1955797, and IIS-2027848.

Data Availability All data and code for the experiments can be found at <https://github.com/Ann924/MobileFacility>.

Declarations

Conflict of interest The authors have no financial or proprietary interests in any material discussed in this article.

Ethical Approval Not applicable.

References

1. COVID-19 Surveillance Dashboard. <https://nssac.bii.virginia.edu/covid-19/dashboard/>. Accessed 23 Jan 2022
2. See How Vaccinations Are Going in Your County and State. <https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html>. Accessed 23 Jan 2022
3. Meet the Four Kinds of People Holding Us Back From Full Vaccination, New York Times. <https://www.nytimes.com/interactive/2021/05/18/opinion/covid-19-vaccine-hesitancy.html>
4. Celik Turkoglu, D., & Erol Genevois, M. (2020). A comparative survey of service facility location problems. *Annals of Operations Research* **292**. <https://doi.org/10.1007/s10479-019-03385-x>
5. Drezner, Z., & Hamacher, H.W. (2002). *Facility Location - Applications and Theory*. Springer, ???
<http://www.springer.com/computer/swe/book/978-3-540-42172-6>
6. Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of covid-19 explain inequities and inform reopening. *Nature* **589**
7. Wang, L., Ben, X., Adiga, A., Sadilek, A., Tendulkar, A., Ventatramanan, S., Vullikanti, A., Tambe, M., & Marathe, M. (2020). Using mobility data to understand and forecast covid 19 dynamics. medRxiv
8. Mehrab, Z., Wilson, M. L., Chang, S., Harrison, G., Lewis, B., Telionis, A., et al. (2022). Data-driven real-time strategic placement of mobile vaccine distribution sites. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*, 12573–12579.
9. Tai, D.B.G., Shah, A., Doubeni, C.A., Sia, I.G., Wieland, M.L. (2020). The disproportionate impact of covid-19 on racial and ethnic minorities in the united states. *Infectious Diseases Society of America*
10. Afshari, H., & Peng, Q. (2014). Challenges and solutions for location of healthcare facilities. *Industrial Engineering and Management*, *2014*, 1–12.
11. Hochbaum, D. S., & Shmoys, D. B. (1985). A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, *10*(2), 180–184. <https://doi.org/10.1287/moor.10.2.180>.
12. Gonzalez, T. F. (1985). Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, *38*, 293–306.
13. Hochbaum, D. S., & Shmoys, D. B. (1986). A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM*, *33*(3), 533–550. <https://doi.org/10.1145/5925.5933>.
14. Bertsimas, D., Digalakis, V., Jr., Jacquillat, A., Li, M. L., & Previero, A. (2022). Where to locate covid-19 mass vaccination facilities? *Naval Research Logistics (NRL)*, *69*(2), 179–200.
15. Bravo, F., Hu, J., Long, E. (2022). Optimal covid-19 vaccination facility location. Available at SSRN 4008669
16. Shukla, S., Fressin, F., Un, M., Coetzer, H., & Chaguturu, S. K. (2022). Optimizing vaccine distribution via mobile clinics: a case study on covid-19 vaccine distribution to long-term care facilities. *Vaccine*, *40*(5), 734–741.
17. Rader, B., Astley, C.M., Sewalk, K., Delamater, P.L., Cordiano, K., Wronski, L., Rivera, J.M., Hallberg, K., Pera, M.F., Cantor, J., et al. (2021). Spatial accessibility modeling of vaccine deserts as barriers to controlling sars-cov-2. medRxiv

18. Roy, S., Dutta, P., Ghosh, P. (2021). Generalizable multi-vaccine distribution strategy based on demographic and behavioral heterogeneity. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1495–1498. IEEE
19. Nedjati, A., & Valipour, M. (2012). Solving health care facility location problems with new heuristic algorithm method. *International Journal of Modeling and Optimization*. <https://doi.org/10.7763/IJMO.2013.V3.225>.
20. Meskarian, R., Penn, M.L., Williams, S., Monks, T.A. (2017). Facility location model for analysis of current and future demand for sexual health services. *PLoS One*
21. de Vries, H., van de Klundert, J., & Wagelmans, A. P. M. (2020). The roadside healthcare facility location problem a managerial network design challenge. *Production and Operations Management*, 29(5), 1165–1187.
22. Anegg, G., Angelidakis, H., Kurpisz, A., Zenklusen, R. (2020). A Technique for Obtaining True Approximations for k -Center with Covering Constraints
23. Harris, D. G., Li, S., Pensyl, T. W., Srinivasan, A., & Trinh, K. (2019). Approximation algorithms for stochastic clustering. *Journal of Machine Learning Research*, 20, 153–115333.
24. An, H.-C., Bhaskara, A., Svensson, O. (2013). Centrality of Trees for Capacitated k -Center
25. Williamson, D. P., & Shmoys, D. B. (2011). *The Design of Approximation Algorithms* (1st ed.). Cambridge: Cambridge University Press.
26. Inamdar, T., Varadarajan, K. (2018). On the Partition Set Cover Problem
27. Wolsey, L. A. (1982). An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4), 385–393. <https://doi.org/10.1007/BF02579435>.
28. Rajagopalan, S., & Vazirani, V. V. (1998). Primal-dual RNC approximation algorithms for set cover and covering integer programs. *SIAM Journal of Computing*, 28(2), 525–540. <https://doi.org/10.1137/S0097539793260763>.
29. Lan, G., DePuy, G.W., Whitehouse, G.E. (2007). An effective and simple heuristic for the set covering problem. *European Journal of Operational Research* 176
30. Grossman, T., & Wool, A. (1997). Computational experience with approximation algorithms for the set covering problem. *European Journal of Operational Research*, 101(1), 81–92. [https://doi.org/10.1016/S0377-2217\(96\)00161-0](https://doi.org/10.1016/S0377-2217(96)00161-0).
31. Gurobi Optimization, L. (2021). Gurobi Optimizer Reference Manual. <http://www.gurobi.com>
32. Perron, L., Furnon, V.: OR-Tools. <https://developers.google.com/optimization/>
33. Chen, J., Hoops, S., Marathe, A., Mortveit, H., Lewis, B., Venkatramanan, S., Haddadan, A., Bhattacharya, P., Adiga, A., Vullikanti, A., Srinivasan, A., Wilson, M.L., Ehrlich, G., Fenster, M., Eubank, S., Barrett, C., Marathe, M. (2021). Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness
34. Machi, D., Bhattacharya, P., Hoops, S., Chen, J., Mortveit, H., Venkatramanan, S., Wilson, B.L.M., Fadikar, A., Maiden, T., Barrett, C.L., Marathe, M.V. (2021). Scalable Epidemiological Workflows to Support COVID-19 Planning and Response

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.