Data-Driven Real-Time Strategic Placement of Mobile Vaccine Distribution Sites

Zakaria Mehrab[1,2], Mandy L. Wilson[1], Serina Chang[3], Galen Harrison[1,2], Bryan Lewis[1], Alex Telionis[1,4], Justin Crow[4], Dennis Kim[4], Scott Spillmann[4], Kate Peters[5] Jure Leskovec[3], Madhav V. Marathe[1,2]

[1]Biocomplexity Institute & Initiative, University of Virginia, Charlottesville, VA, USA
[2]Department of Computer Science, University of Virginia, Charlottesville, VA, USA
[3]Department of Computer Science, Stanford University, Stanford, CA, USA
[4]Virginia Department of Health, P.O. Box 2448, Richmond, Virginia 23218-2448
[5]Deloitte Consulting LLP, 555 W 5th St #2700, Los Angeles, CA 90013

Abstract

The deployment of vaccines across the US provides significant defense against serious illness and death from COVID-19. Over 70% of vaccine-eligible Americans are at least partially vaccinated, but there are pockets of the population that are under-vaccinated, such as in rural areas and some demographic groups (e.g. age, race, ethnicity). These pockets are extremely susceptible to the Delta variant, exacerbating the healthcare crisis and increasing the risk of new variants. In this paper, we describe a data-driven model that provides real-time support to Virginia public health officials by recommending mobile vaccination site placement in order to target under-vaccinated populations. Our strategy uses fine-grained mobility data, along with US Census and vaccination uptake data, to identify locations that are most likely to be visited by unvaccinated individuals. We further extend our model to choose locations that maximize vaccine uptake among hesitant groups. We show that the top recommended sites vary substantially across some demographics, demonstrating the value of developing customized recommendation models that integrate fine-grained, heterogeneous data sources. We also validate our recommendations by analyzing the success rates of deployed vaccine sites, and show that sites placed closer to our recommended areas administered higher numbers of doses. Our model is the first of its kind to consider evolving mobility patterns in real-time for suggesting placement strategies customized for different targeted demographic groups.

1 Introduction

As of August 2, 2021, at least 70% of American adults aged 18 and older had received at least one dose of a COVID-19 vaccine (Reuters 2021). However, in many subpopulations, including young people, Black people, people of Latinx ethnicity, and in rural areas, the vaccination rate runs far below that (UCSF 2021; PBS 2021). Strategies have been devised to address vaccine accessibility – free child care, paid time off for employees, and other financial incentives – but these measures have not proven effective for these under-vaccinated demographic groups.

In order to increase the rate of vaccination among the under-vaccinated or less vaccine-enthusiastic populations, the Virginia Department of Health (VDH) has begun deployment of mobile vaccine distribution sites. These mobile units distribute the one-dose Johnson & Johnson vaccine in order to simplify scheduling and encourage "impulse" vaccinations (i.e., seizing the opportunity to get vaccinated when presented). When VDH started this program, deployment was driven primarily by intuition or educated guesses by local public health officials. However, two important factors were not well-addressed in this deployment strategy: (i) each demographic group has its own mobility patterns ; and (ii) mobility patterns have been evolving as the lockdown has eased. Therefore, a more methodical, real-time deployment plan was needed to address these deficits.

The success of mobile site placements depends on (i) the **accessibility** of these sites for the target populations, and (ii) their willingness to get vaccinated (**acceptance/hesitancy**). To address the **accessibility** aspect, we propose that areas with high foot traffic from the target demographic groups would be productive locations to place mobile vaccination units. To this end, we employ a *dynamic*, *data-driven* recommendation model, using heterogeneous data sources, that can recommend locations with high probability of vaccination uptake success in *real-time*.

Our model works as follows. *First*, it uses aggregated and anonymized mobility data from SafeGraph to identify areas with high mobility concentrations and the Census Block Groups (CBGs) that contribute to that traffic. Such data has been used extensively as a means to study the spread of COVID-19 and to track the degree of compliance with social distancing directives (Badr et al. 2020; Buckee et al. 2020; Warren and Skillman 2020; Wellenius et al. 2020; Chang et al. 2021a; Wang et al. 2020). The candidate areas are defined as tessellations indexed by Google's S2 Geometry. *Second*, it leverages multiple data sources containing the demographic profile of each CBG to adjust the previously computed mobility for target demographic groups using a set of equations. Based on this adjusted mobility, it ranks the tessellations separately for each group. *Third*, it is equipped with a module to estimate vaccine **acceptance** across different demographic groups to refine the previous rankings.

Overall, our model can be described as a rule-based system which consists of a set of rules (equations) used to process the heterogeneous knowledge graph data sources (SafeGraph + Census data) and takes subsequent actions (recommendations). The model also champions fairness and equity, which are lingering issues in AI. Machine learning models can suffer from bias, serving certain demographic groups better than others (Sweeney et al. 2019), but our model mitigates this issue by employing the rules after taking into account the racial heterogeneity of CBGs.

To evaluate how our placement strategy compared with actual placements made by public health officials, we review existing mobile distribution site placements from VDH-provided data and observe vaccines administered at these locations. Furthermore, we analyze the robustness of our model for various demographic groups and different weeks.

In line with the continuing support our group has provided to various local, state, and federal public health authorities since the onset of the pandemic, we presented two prototypes of our model to VDH and received valuable guidance integral to the current implementation of our model and the selection of demographic groups for deployment. Our model has been operational since the beginning of June 2021, continuously providing real-time placement recommendations. Although we have not received quantifiable data reflecting the effectiveness of these sites, VDH does rely on these recommendations for mobile vaccination site planning. Francisco Diaz, the Vaccine Administration Support Supervisor for VDH, has stated that this program allows VDH to identify where vaccines are needed. This improves their ability to focus their efforts on reaching target demographic groups in locations that are accessible and convenient to them. His complete statement is available at (Diaz 2021).

Our work is the first of its kind to develop a data-driven model that considers evolving mobility patterns and finds a real-time placement strategy that is accessible to different targeted demographic groups. Through its simplicity and interpretability, our model also outshines difficult-to-solve location-theory algorithms and hard-to-interpret deep learning models, something that is sought by policymakers due to its safety-critical use case. From this aspect, our model advocates interpretable AI.

The rest of this paper is organized as follows. We describe similar previous works relying on mobility data in Section 2. In Section 3, we describe the incremental design and implementation of our model to create a ranking of areas for placement of vaccine distribution sites. Salient insights about mobility of different demographic groups and comparison between our recommendations and existing sites are presented in Section 4. Lastly, we explore the utility and implications of our model in Section 5.

2 Related Work

Mobility and Vaccination A first set of works (Saldaña et al. 2021; Chen et al. 2021; Buckner et al. 2021; Jentsch et al. 2021) simulated different vaccination strategies for various scenarios and studied their effectiveness. (Jentsch et al. 2021) found that if there is a delay in vaccine availability, it is more effective to target individuals with high social contact instead of focusing on the elderly; (Chen et al. 2021) reached similar conclusions, except they compared between high contact people and essential workers. (Saldaña et al. 2021) evaluated the effectiveness of different vaccination strategies by simulating a meta-population model across several scenarios. (Buckner et al. 2021) used a mathematical model that indicated that the prioritization strategy should vary depending on the objective; for example, targeting essential workers minimizes infection, but targeting older individuals minimizes the number of deaths. In terms of vaccination, these works focus mainly on *who* but not *where*, whereas in our work we focus on both aspects.

COVID-19 systems to aid policymakers Similar to how our system is designed to provide data-driven, real-time support to policymakers, quite a few systems have been developed to aid policymakers during the pandemic. Some were designed for surveillance purposes, e.g. visualizing infection rates and trends at different spatial resolutions (Dong et al. 2020; Peddireddy et al. 2020; Wissel et al. 2020), identifying anomalous hotspots (Hohl et al. 2020), and informing policymakers about the necessary levels of restriction in a timely fashion (Qiu 2021). Another set of systems was developed to help policymakers observe the effects of different non-pharmaceutical interventions in order to help them make informed decisions (Barrett et al. 2007; Beckman et al. 2014; Chang et al. 2021b).

What sets our work apart is that we are the first to develop an operational system that provides weekly updates to policymakers regarding placement strategies for mobile vaccine distribution sites across different demographic groups. Also, unlike other systems which largely focus on surveillance and retrospective analyses, our system provides real-time support to policymakers for mitigating disease transmission.

3 Methodology

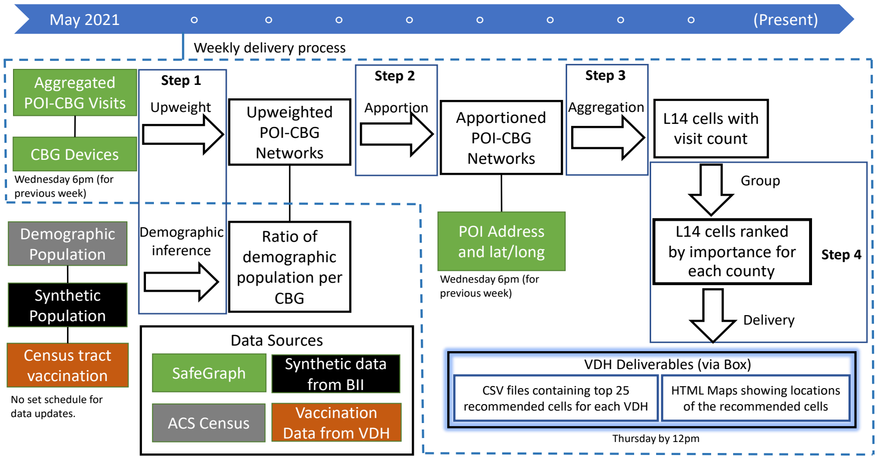
3.1 Datasets

Fine-grained mobility data (SafeGraph) Mobility data can reveal important information about populations, such as where people are visiting and how this behavior evolves over time. For our model, we use anonymized and aggregated data from SafeGraph (SafeGraph 2018). It provides detailed information about non-residential locations visited by individuals (e.g. grocery stores, parks), also referred to as *points of interest* (POIs). SafeGraph's Weekly Patterns dataset[1], released on Wednesdays with the data for the previous week (Monday through Sunday), includes weekly estimates of visits from CBGs to these POIs. The dataset can be naturally viewed as a bipartite graph, as described later in 3.3. For our work here, we focus on POIs and CBGs in the Commonwealth of Virginia, where visits from 5,293 CBGs to 74,535 POIs were compiled in the latest release.

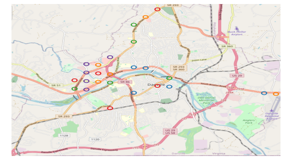
However, SafeGraph has some limitations. For instance, it does not cover all POIs or populations (e.g., children). Furthermore, depending on the number of devices carried by a user, visits may be underreported or overreported. The GPS signal itself can also be noisy. We describe how we account for some of these limitations in Sections 3.2 and 3.3.

Demographic Data As the SafeGraph dataset does not include demographic information, we use demographic data from the US Census American Community Survey (ACS) in conjunction with the visits from the CBGs to POIs data.

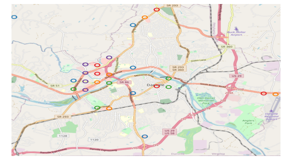
[1]https://docs.safegraph.com/v4.0/docs/weekly-patterns



(a) Data-to-deployment pipeline.



(b) Danville HTML map (B)



(c) Danville HTML map (L)

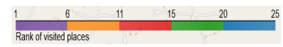


Figure 1: Detailed breakdown of the pipeline of our recommendation model (1a). The pipeline is run weekly on Wednesdays with the updated SafeGraph data. Sample HTML deliverables generated by the model are shown for Danville City for the week of June 21-June 27 for two demographic groups (1b,1c). The deliverables to VDH are generated by noon on Thursday.

The 2015-2019 release of the 5-year Census data contains the population breakdown in each CBG by different demographic groups (e.g. Black, Latinx). This adapts our model to estimate visits from each demographic group to each POI, or which POIs are frequented by each group.

Although the census data provides numbers for individual demographics, such as race or age group, it does not include fine-grained information about combined demographics. We describe our approach for handling combined demographics in the extended version of this paper (Mehrab et al. 2021).

Vaccination Data Our final dataset, obtained from VDH, contains the number of individuals per census tract who have received vaccine doses. However, SafeGraph and ACS Census data are provided at the CBG level. In order to maintain the same level of resolution across all data sources, we estimate the number of unvaccinated individuals at the CBG level, assuming that the number of doses in a census tract is distributed proportionately across its underlying CBGs.

3.2 S2 Geometry

Although SafeGraph provides the number of visits to specific POIs, the resolution level is quite dense for calculating the placement of mobile vaccination sites. Furthermore, the signals picked up for a specific POI do not necessarily indicate that the device was present in that specific POI at that time of collection due to the noise associated with Global Positioning Systems (GPS) signals. In order to address these issues, instead of considering individual POIs with high foot traffic for potential site placement, we identify geographical areas with high foot traffic as follows. (i) We divide each county in Virginia into much smaller *areas*, each of which contains a group of POIs. (ii) For each area, we aggregate the foot traffic for all of the POIs in that area, then rank each area inside a region based on this aggregated foot traffic.

To partition Virginia into regions, we use Google's S2 geometry[2], which divides the world map into nested cells of decreasing size (L0 - L30). Level L0 is one cell representing the entire map; it contains 4 L1 cells, each of which contains 4 L2 cells, each of which contains 4 L3 cells, and so on.

3.3 Placement model

Our data-driven model was guided by two preceding proof-of-concept prototypes, described in more detail in the extended version of this paper (Mehrab et al. 2021). Lessons learned from these prototypes helped shape our final model, which is described below.

Mobility Network: One input to our placement model is a dynamic mobility network which can be represented as a bipartite graph $G(V, E)$, where V is the set of nodes and E is the set of time-varying edges. V is the union of two disjoint sets $C = \{c_1, ..., c_m\}$ and $P = \{P_1, ..., P_n\}$. Here C represents the set of CBGs, and P is the set of POIs in the dataset. Each edge (c_i, p_j) is associated with a weight w_{ij}^t, the number of people from c_i visiting p_j at time t.

Let $D = \{d_1, d_2, ...\}$ be a set of demographic groups of interest. Each CBG c_i is associated with its overall population N_i and the population N_i^k, which is specific to demographic group d_k (where $d_k \in D$). Each CBG c_i is also associated with M_i, the number of mobile devices from c_i captured by SafeGraph. Subsequently, each POI p_j is also associated with a small geographic area denoted by S_j and a large region indicated by L_j. In our final model described below, each small geographic area is an L14 cell of Google's S2 geometry as described in Section 3.2 and each large region is a county. Therefore, we use these notations for denoting the L14 cell and county also in subsequent paragraphs.

Our placement model takes the dynamic bipartite graph generated from SafeGraph and a set of demographic groups of interest, and generates as output a ranked list of areas as

[2]https://s2geometry.io/

potential candidates for setting up mobile vaccination units within each larger region. This is performed as follows:

1. First, we adjust the weights along the edges of the graph to estimate the number of actual visits to each POI to mitigate the under-reporting issue discussed in 3.1. Due to this issue, the actual number of visits is higher than the reported number of visits. This issue is addressed by *upweighting* the visits along each edge as : $U_{ij}^t = w_{ij}^t \frac{N_i}{M_i}$
2. Second, we adjust the weights for each demographic group d_k by obtaining an apportioned estimate of how many people of that group from CBG c_i visited a POI p_j. In other words, we *apportion* the visit along the edge for a particular demographic d_k as : $A_{ij}^{tk} = U_{ij}^t \frac{N_i^k}{N_i}$
3. Third, each area (L14 cell) within a region (county) is ranked based on the apportioned visits to the POIs within that cell. Let $\mathcal{S} = \{s_1, s_2, ...\}$ be the set of all distinct areas. Then, we *aggregate* the apportioned number of visits to the POIs of a particular L14 cell s_L and calculate its *visit count* at time t for a particular demographic group d_k as: $I(s_L, t, k) = \sum_{p_j \in P} \sum_{c_i \in C} A_{ij}^{tk} \mathbb{1}_{s_L}(S_j)$ where $\mathbb{1}_{s_L}(S_j)$ is 1 when $S_j = s_L$ and 0 otherwise.
4. Finally, we *group* each area (L14 cell) S_j by their corresponding region (county) L_j, then sort by their *visit counts*. For each county, the 25 top-ranked areas are reported as candidate sites. The conceptual pipeline for the model is presented in Figure 1a.

3.4 Demographic Acceptance/Hesitancy

Even if vaccination units are placed in easily accessible locations, some individuals may be unwilling or afraid to get vaccinated (CNN 2021; PBS 2021). We employ a simulation-based approach to infer hesitancy levels of different demographic groups at the county level; this approach is described in the extended version of this paper (Mehrab et al. 2021). After the simulation, for each timestamp t, each CBG C_i and a demographic group d, we have the population count, number of individuals who had at least one dose of vaccine V_{id}^t, and the estimated number of hesitant individuals as per the calibrated model H_{id}^t. In this section, we describe how we incorporate these values into our model.

Overcoming the hesitancy threshold in areas where vaccine acceptance is low can be a challenge; for this reason, placing mobile sites in these areas may not be the best strategy for maximizing uptake. However, hesitant people may be influenced to take the vaccine if they see their peers getting vaccinated (i.e. peer pressure). Therefore, our approach is to improve accessibility to the vaccine-accepting in areas that may also be frequented by the hesitant. Our method to update the apportioned value in Step 2 of Section 3.3 using this new dataset is described as follows:

1. We still pick locations with high foot traffic (emphasis on the original value of the apportioned visits A_{ij}^t)
2. We also give higher priority to areas with greater numbers of unvaccinated individuals.
3. If vaccine-hesitant people can be influenced by peer pressure, then we want to lower priority in areas where most of the unvaccinated are hesitant.

Based on this, we calculate an updated apportioned weight $\mathcal{A}_{ij}^t$ by incorporating hesitancy data from the original apportioned weight A_{ij}^t as: $\mathcal{A}_{ij}^t = A_{ij}^t * R_{id}^t * X_{id}^t$

Here, R_{id}^t is the ratio of the unvaccinated individuals to the total population of demographic group d in CBG C_i at time t. X_{id}^t is the ratio of vaccine-accepting individuals to vaccine-hesitant individuals in the population.

Updating the apportioned weight in this approach ensures that places with high mobility are still given importance A_{ij}^t, priority is reduced for CBGs where a majority of the population is vaccinated R_{id}^t, and a good balance between accepting and hesitant populations is maintained.

3.5 Implementation

Due to the large volume of the SafeGraph Weekly Patterns data, *upweight* and *apportion* are done on separate chunks of the complete dataset in parallel jobs. We also filter out CBGs and POIs outside of Virginia and associate the remaining POIs with an L14 cell during this step. Afterwards, the output from each parallel job is concatenated into a single dataset, and the *visit counts* to each POI are aggregated by their corresponding L14 cells. Then, we generate a CSV file for each demographic group containing the S2 identifier of each L14 cell, latitude and longitude of the centroid of that cell, and its visit count. The L14 cells are ranked based on their *visit count* for each county. The top 25 cell locations per county are also shown in HTML maps, one for each demographic group, for ease of visualization. In Figure 1b and 1c, we display sample HTMLs generated by our model for the week of June 21 - June 27 showing some of the recommended locations for the deployment of mobile vaccination sites in Danville City for two demographic groups.

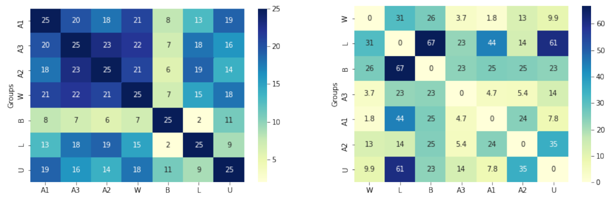
Our model is operationalized to deliver these CSV and HTML files to VDH weekly, which they are using currently to target the 10 health districts with the lowest vaccination rates. Based on guidance from VDH, we are currently delivering recommendations for the target demographic groups Ethnicity Latinx (L), Race Black (B), and populations within the ages of 20 to 39 (A1), 20 to 29 (A2), and 30 to 39 (A3). We also use cumulative vaccination data from VDH to generate recommendations targeting unvaccinated (U) people. These, along with the population group "Whole population" (W) which considers foot traffic only, means we deliver recommendations for a total of seven population groups.

4 Analysis

4.1 Comparison of model recommendations across demographics and temporality

This section presents some key insights obtained from analyses of our model output conducted at state level and across each health district. Detailed discussion on these results are covered in the extended version (Mehrab et al. 2021).

First, we present comparative analyses across different demographic groups at the state level. Figure 2a tabulates the number of common recommendations across the top 25 recommended sites in Virginia for pairs of demographic groups for the week of July 19 - July 25. We observe only two common recommended areas between *L* and *B*, indicating



(a) Number of common areas. (b) AMD (km).

Figure 2: Statewide comparison of recommended places across demographic groups for the July 28 delivery.

that frequently visited locations by these two demographic groups differ quite a bit. We also find that the locations frequented by *B* are quite different compared to those of the other demographic groups. Therefore, it is worthwhile to consider customized recommendations for different groups.

Over the course of our weekly deliveries to VDH, we also noticed that recommended sites varied across different weeks. Figure 3a and 3b show variations for the top 25 recommended areas over a two-month delivery period for demographic groups *W* and *L*, respectively. Interestingly, the recommended areas are largely similar for the four weeks of June, then again across the four weeks of July. But if we compare any week in June to any week in July, the recommendations differ significantly. This suggests that mobility patterns changed after June going into the month of July. The state of emergency mandated by Virginia ending on June 30 may be a possible cause for this change in mobility pattern.

Upon further study, we found that the "different" L14 cells often shared a border, suggesting that aggregated foot traffic in those cells came from adjacent POIs. Therefore, instead of looking at common areas, we examined how far apart highly-ranked areas were relative to each other by calculating the average minimum distance (AMD) between two sets of areas. The two sets can either be the top 25 areas for two groups in a week or the top 25 areas for two different weeks for a particular group, depending upon the analysis involved. We calculated AMD by matching each area in one set with its closest area in other set and taking average haversine distance of the matched areas based on their centroids.

We find that while the monthly pattern is still observable for *W* (Figure 3c), it is not the case for *L* (Figure 3d). This is contrary to the similarity observed in Figure 3a and Figure 3b. Furthermore, the AMD between areas are quite concentrated for *L* in Figure 3d while the AMD is high between areas recommended for *L* and other demographic groups (Figure 2b). This indicates that while the frequently visited locations by Latinx individuals are comparatively far away from other demographic groups, the locations themselves visited by Latinx individuals remain relatively close to each other across different weeks.

4.2 Acceptance/Hesitancy

In this section, we analyze the effect of incorporating hesitancy into our model by examining the recommended mobile site placements for *B* and *L*. We refer to our base model as M_B and the hesitancy-incorporated model as M_H.

First, we observe the number of common recommendations between the two implementations of the model, M_B and M_H. We compare from the top 1 to the top 100 areas recommended by both models and find there are more differences across the models between the recommended areas for the group *B* than group *L* (left y-axis of Figure 4a). For example, M_B and M_H have only two common areas between their recommended top 25 areas for group *B*, while there are 17 common areas for *L*. This indicates that many of the areas recommended by the base model for the Black race see high percentages of vaccine-hesitant individuals, while this phenomenon is less evident for the Latinx ethnicity group.

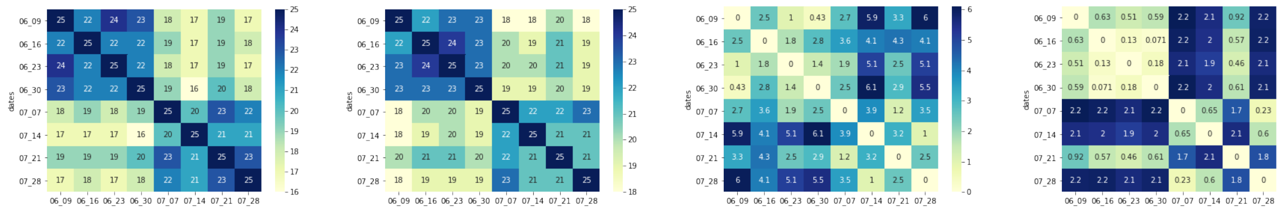
To explore this further, we looked at the AMD of the two sets of recommended areas considering from the top 1 to top 100 areas. We see that the two top areas for *B* in the two models are highly disparate (right y-axis of Figure 4a). For example, the top recommended area by M_B is in Virginia Beach, whereas the top recommended area by M_H is in Stafford, which is about 200km from Virginia Beach.

Finally, we visualize the top 100 areas recommended by both models on a map. For *B*, we see that while many areas in the southern part of Virginia are recommended by M_B, the areas recommended by M_H are mostly concentrated within the Central and Northern parts of Virginia. It is also interesting that M_H does not recommend any area in Virginia Beach even within its top 100 recommendations, while the area was recommended highly by M_B (Figure 4). The model is much less sensitive to the Ethnicity Latinx group, as the map tells us that the recommended areas are mostly the same in both versions of the model (Figure 4c).

4.3 Validation study

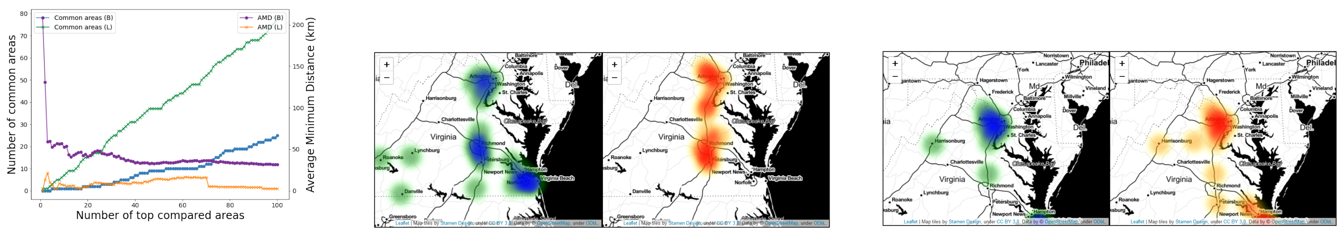
The absence of ground truth data about the accessibility of our recommended areas made the evaluation of our strategy a bit involved. However, VDH provided a list of 147 mobile sites deployed from May 19 to June 30, and the number of daily doses administered at those sites, which we used to retrospectively analyze the effectiveness of our strategy. We looked at the corroboration of our recommendations with these deployed sites using the recommendations for *B* over the month of June. The results were also similar for *L*, which we show in the extended version (Mehrab et al. 2021). Specifically, for each deployed site, we looked at our closest recommended area and found that most of the deployed sites were within 1km of at least one of our recommended areas, suggesting that either our placement strategy led to selection of these sites, or the original selection strategy and our strategy corroborated each other to some extent.

To evaluate the effectiveness of these conforming sites with respect to vaccine administration, we compared the sites within 1km of at least one of our recommended areas with the remaining sites in terms of doses administered. We find that, in general, sites close to our recommended areas administered higher numbers of doses on average. (Figure 5b). This indicates that our recommended areas are indeed more accessible, and it is probable that using our placement strategy could help policymakers to increase uptake among targeted demographic groups.



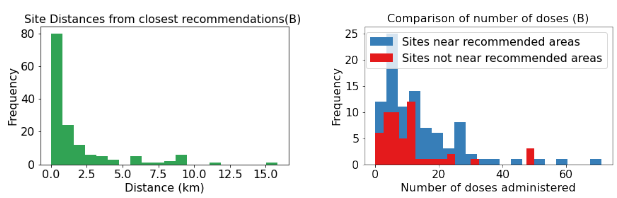
(a) Number of common areas across delivery weeks for W. (b) Number of common areas across delivery weeks for L. (c) AMD (km) across delivery weeks for W. (d) AMD (km) across delivery weeks for L.

Figure 3: Statewide comparison of recommended places across delivery weeks.



(a) Comparison of top 100 places. (b) Heatmap for *B* by M_B (left) and M_H (right). (c) Heatmap for *L* by M_B (left) and M_H (right).

Figure 4: Number of common areas and AMD for areas recommended by M_B and M_H (Figure 4a) and heatmaps of areas recommended by M_B and M_H for Race Black (*B*) and Ethnicity Latinx (*L*) (Figures 4b and 4c).



(a) Conformity of our recommendations with deployed sites. (b) Comparison of doses administered in sites.

Figure 5: Effectiveness of our placement strategy.

5 Discussion

In this work, we have devised a data-driven, equity supportive, dynamic, rule-based recommendation model for the social good that considers evolving mobility patterns to find a real-time placement strategy for making vaccines more accessible to targeted demographic groups. This model has been operational since early June 2021, and, since then, we have delivered mobile vaccination recommendations to VDH every week. Our strategy is simple and transparent, yet effective. There are some additional takeaways from this experience that are worth noting, however.

First, the target demographic groups are divergent in their mobility, alluding to the necessity of a targeted ranking of locations. Similarly, depending on the target group, considering hesitancy may be necessary. *Second*, trends at the state level do not carry through at finer resolutions. *Third*, while mobility may differ across demographic groups, it stays comparatively stable over time for a particular group. *Fourth*, although our research focuses on the Commonwealth of Virginia, this approach can be generalized for any other US state, which we leave as future work.

There are some limitations to our work. The absence of data made it difficult to validate our work against actual vaccination rates at the recommended sites; however, our site recommendations compared favorably with the sites placed by VDH where data was available. Our model also assumes that people from a given CBG who frequent these L14 locations are demographically similar to the population of that CBG; this may not hold true in all cases. In the future, we want to factor this into our model by taking into consideration that the mobility of a particular demographic group from a CBG to a POI not only depends on the demographic distribution of the CBG, but also the category of the POI.

Acknowledgments

This work was partially supported by the National Institutes of Health (NIH) Grant R01GM109718, VDH Grant PV-BII VDH COVID-19 Modeling Program VDH-21-501-0135, NSF Grant No.: OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819, NSF RAPID CNS-2028004, NSF RAPID OAC-2027541, US Centers for Disease Control and Prevention 75D30119C05935, DTRA subcontract/ARA S-D00189-15-TO-01-UVA.

References

Badr, H. S.; et al. 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*.

Barrett, C. L.; et al. 2007. A scalable data management tool to support epidemiological modeling of large urban regions. In *Proc. of TPDL*.

Beckman, R.; et al. 2014. Isis: A networked-epidemiology based pervasive web app for infectious disease pandemic planning and response. In *Proc. of ACM SIGKDD*.

Buckee, C. O.; et al. 2020. Aggregated mobility data could help fight COVID-19. *Science*.

Buckner, J. H.; et al. 2021. Dynamic prioritization of COVID-19 vaccines when social distancing is limited for essential workers. *PNAS*.

Chang, S.; et al. 2021a. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*.

Chang, S.; et al. 2021b. Supporting COVID-19 Policy Response with Large-Scale Mobility-Based Modeling. In *Proc. of ACM SIGKDD*.

Chen, J.; et al. 2021. Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness. *medRxiv*.

CNN. 2021. New data shows many Black Americans remain hesitant to get Covid-19 vaccine. https://cnn.it/3FQ7Zbs.

Diaz, F. 2021. VDH's Comments about UVA Mobile Site Recommendations. https://bit.ly/2ZdDOus.

Dong, E.; et al. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*.

Hohl, A.; et al. 2020. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-temporal Epidemiology*.

Jentsch, P. C.; et al. 2021. Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *The Lancet Infectious Diseases*.

Mehrab, Z.; et al. 2021. Data Driven Real-Time Strategic Placement of Mobile Vaccine Distribution Sites - Extended Version. https://bit.ly/3dhNDeR.

PBS. 2021. Black and Latino communities often have low vaccination rates – but blaming vaccine hesitancy misses the mark. https://to.pbs.org/3cPTffX.

Peddireddy, A. S.; et al. 2020. From 5vs to 6cs: Operationalizing epidemic data management with covid-19 surveillance. In *Proc. of IEEE BigData*.

Qiu, R. G. a. 2021. Data-Driven Modeling to Facilitate Policymaking in Fighting to Contain the COVID-19 Pandemic. *Procedia Computer Science*.

Reuters. 2021. 70% U.S. adults have taken at least one dose of COVID-19 vaccine - CDC. https://reut.rs/3l9mq2i.

SafeGraph. 2018. SafeGraph Data for Academics. https://www.safegraph.com/academics.

Saldaña, F.; et al. 2021. The trade-off between mobility and vaccination for COVID-19 control: a metapopulation modelling approach. *Royal Society open science*.

Sweeney, C.; et al. 2019. *Understanding and mitigating unintended demographic bias in machine learning systems*. Ph.D. thesis, Massachusetts Institute of Technology.

UCSF. 2021. Vaccine Hesitancy in Young Adults May Hamper Herd Immunity. https://bit.ly/3CMdUfH.

Wang, H.; et al. 2020. Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona. *arXiv preprint*.

Warren, M. S.; and Skillman, S. W. 2020. Mobility changes in response to COVID-19. *arXiv preprint*.

Wellenius, G. A.; et al. 2020. Impacts of state-level policies on social distancing in the United States using aggregated mobility data during the COVID-19 pandemic. *arXiv e-prints*.

Wissel, B. D.; et al. 2020. An interactive online dashboard for tracking COVID-19 in US counties, cities, and states in real time. *Journal of the American Medical Informatics Association*.