Closer to Home: An Estimate-then-Optimize Approach to Improve Access to Healthcare Services

Fernanda Bravo, Ashvin Gandhi, Jingyuan Hu, Elisa F. Long
UCLA Anderson School of Management, Los Angeles CA 90095
{fernanda.bravo, ashvin.gandhi, jingyuan.hu.phd, elisa.long}@anderson.ucla.edu

Geographic inequalities in access to essential health services are well-documented, extending beyond a rural-urban divide to include socioeconomic, racial, and other disparities. Proximity to hospitals, clinics, healthcare providers, and pharmacies varies widely, posing a challenge in determining where to strategically locate such facilities. Demand for each service depends on the population health in the catchment area, individual preferences, provider capacity, and other factors. This study introduces a novel estimate-then-optimize framework combining structural (BLP) demand estimation with a choice-based optimal facility location model to maximize health service utilization. An advantage of this empirical approach is its reliance on aggregated data (*e.g.*, market share) rather than individual choices or outcomes.

We illustrate our proposed methodology by examining the Federal Retail Pharmacy Program—a historic public-private partnership that administered millions of COVID-19 vaccinations—in California. Our demand estimates reveal that residents of socioeconomically vulnerable communities are more sensitive to travel distances to pharmacy-based vaccination sites. Augmenting the existing set of pharmacies with 500 *strategically selected* retail stores serving lower-income communities could increase predicted vaccinations by 2.9 percent overall, translating to 770,000 additional vaccinations statewide, and by 5.4 percent in the least healthy neighborhoods. Even with just 100 optimally chosen stores, this partnership could achieve more than half a million additional vaccinations—five times the predicted gains associated with randomly selecting 100 new sites. By integrating a structural demand model with prescriptive analytics, our study provides a systematic, data-driven approach for policymakers and practitioners to identify where additional health resources are most needed.

Key words: structural estimation, BLP, choice model, facility location, healthcare access

1. Introduction

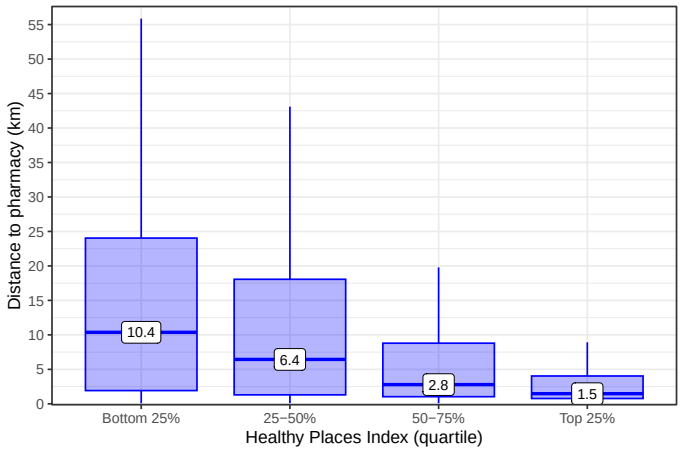
Across the United States, access to essential healthcare resources, such as acute care hospital beds, maternity units, primary care physicians, or pharmacy services, largely depends on where one lives (Fisher et al. 2008, Bronner et al. 2021). Numerous studies have established that geographical proximity to healthcare services influences utilization rates, expenditures and, ultimately, patient health outcomes. One study of Medicare claims data estimated that more than half of the geographic differences in healthcare utilization is explained by variations in the supply of services, with the remainder due to patient-related differences in demand (Finkelstein et al. 2016).

Proximity to a hospital is a determining factor not only for patients seeking emergent care, such as treatment following a heart attack (Tay 2003), but also for patients seeking routine inpatient services (Ho 2006, Gowrisankaran et al. 2015, Shepard 2022). The nursing home industry also reflects this pattern, where patients' facility choice is highly sensitive to distance, further affecting patient selection (Gandhi 2023) and local competition (Hackmann 2019). The Supreme Court's 2022 ruling overturning Roe v. Wade widened the country's existing spatial disparities in access to reproductive health services (Myers et al. 2019), particularly for low-income individuals and rural communities (Bearak et al. 2017).

Excessively long distances to health service providers pose greater challenges for some communities (Dingel et al. 2023). Residents of economically and socially vulnerable areas may face additional hurdles such as limited time off work, lower vehicle ownership, and limited access to public transit (Syed et al. 2013). Rural Americans generally face worse health outcomes and shorter life expectancy (Deryugina and Molitor 2021), owing in part to a lower quantity and quality of healthcare providers (Finkelstein et al. 2021). While cities have higher concentrations of physicians, many urban areas face shortages of primary care and specialist providers (Brown et al. 2016).

Outside of hospitals and clinics, pharmacists play a crucial role in dispensing essential medications, consulting with patients, and administering vaccinations. Nearly five percent of Americans live within a *pharmacy desert*—areas with no pharmacies within a one-mile radius (or within ten miles in rural areas)—and lack access to these essential services (Wittenauer et al. 2024). Pharmacy deserts are becoming increasingly ubiquitous as more retail pharmacy stores close and the industry consolidates (Salako et al. 2018). Medicaid reimbursements for a pharmacy's dispensing fee are often insufficient to cover the operating costs at low-volume pharmacies, leading to further closures (Ippolito et al. 2020). While pharmacy deserts are most common in rural areas, some densely populated regions also lack pharmacists, particularly in Black and Hispanic neighborhoods

Figure 1 **Variation in distance to a pharmacy offering COVID-19 vaccinations across zip-codes in California**



(Guadamuz et al. 2021). In California, 2.5 million residents (6 percent) live in a pharmacy desert (Wittenauer et al. 2024), including 400,000 residents of urban Los Angeles County (Guadamuz et al. 2021). Figure 1 illustrates the state's inequality in access to a pharmacy by the Healthy Places Index (HPI), a composite measure of community well-being based on education, income, housing, and other social determinants of health (Public Health Alliance of Southern California 2022). Residents in the lowest HPI quartile—the most economically and medically vulnerable—are twice as likely as those in the top quartile to lack access to an automobile, creating an additional burden to obtaining care.

Given the documented spatial inequalities in access to health services (Finkelstein et al. 2016), efforts to mitigate these disparities have urgent policy importance. One potential mechanism, when services are largely undifferentiated, is better positioning of medical facilities with a well-defined objective such as maximizing service utilization or minimizing inequities in distance traveled. This first requires an understanding of individuals' preferences in selecting where to obtain service. One notable empirical challenge, however, is that individual-level outcome data are often not available, given privacy concerns or constraints in obtaining patient health records. For example, disease prevalence is often reported at an aggregate geographic level, yet policymakers typically lack per-facility outcome data or detailed information on which patients visited each healthcare facility.

To help inform this policy question, we propose an estimate-then-optimize framework combining a structural demand model that estimates the effect of distance on the utilization of a specific healthcare service, with a prescriptive model that optimally selects locations to provide the service (Figure 2). Our framework addresses a class of problems in healthcare delivery: improving access to care by optimally selecting a network of providers offering a standard service. Our model aims to realign a mismatch between the supply of service providers and patient demand for such services. Consider an undifferentiated service, such as a routine eye examination. Not only are the needs for such services *heterogeneous* across patient groups, individual patient preferences (*e.g.*, willingness to travel) over the set of feasible providers can also vary. Moreover, the availability of these service

Figure 2 **Proposed estimate-then-optimize modeling framework**

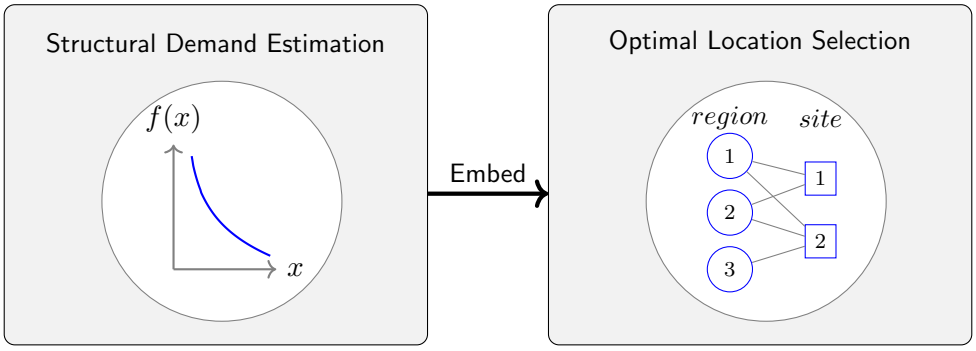


Table 1 Healthcare contexts for using an estimate-then-optimize framework to maximize access to service

Market characteristic	Description	Illustrative healthcare setting
Demand (by patients)	Only aggregate data on service utilization are available	Vaccination rates by county, cancer screening rates by state
	Heterogeneous patient preferences or need for health service	Socioeconomic differences in dialysis for treatment of end-stage renal disease
Supply (by providers)	Spatial variability in proximity to service providers	Pharmacy deserts, reproductive health clinic access, rural hospitals
	Unobserved utilization rates by service provider	Primary care visits by provider, prescription dispensing rates by pharmacy
	Limited service capacity per facility	Hospital bed capacity, appointment slots, hemodialysis machines
	Undifferentiated health service	Dental checkups, blood donations, mammography screenings, eye exams

providers can widely differ by geographic area, contributing to inequities in utilization across patient groups. Finally, health outcome data are often publicly available only at an *aggregate* level, posing a challenge for estimating patient demand for services, or measuring utilization by service provider. This class of problems can arise in many healthcare-related contexts, including public health policy-making and care delivery network design. Table 1 summarizes the key elements of our context and provides some illustrative examples.

A summary of our approach is as follows. In the first stage, we structurally estimate spatial demand for an undifferentiated healthcare service by employing a modified random-coefficient choice (BLP) model with a novel capacity constraint per facility. In the second stage, we integrate the structural model into a choice-based facility location problem, to identify the optimal set of locations that maximizes predicted utilization rates. The realized utilization rates are then estimated through a random sequential simulation. Our framework links structural estimation and optimization modeling, harnessing aggregate, observational data to enhance prescriptive modeling. Our approach could be adapted to studying demand for other types of service facilities and provide insights on policy design, spatial resource allocation, and firm expansion.

We illustrate our proposed framework with a case study on a widespread vaccination campaign. Utilizing cross-sectional data on COVID-19 vaccination rates by zip code in California, we estimate a heterogeneous distance sensitivity to vaccination uptake. We perform extensive counterfactual analyses and demonstrate how policymakers could augment the existing network of pharmacy partners offering vaccinations. To the best of our knowledge, this is the first work to document a significant, and heterogeneous, relationship between proximity to a vaccination site and uptake.

2. Related Literature

Our paper relates to prior work on the accessibility of health services, optimal facility location, and demand estimation. Each is a substantial body of work spanning multiple fields, and we highlight the most relevant studies in the following sections.

2.1. Accessibility of Health Services

The concept of healthcare access has been a subject of discussion for the past half-century (Aday and Andersen 1974). Penchansky and Thomas (1981) outline five dimensions that define the relationship between patients and health systems: (1) *Accessibility* concerns the geographical proximity of healthcare providers to patients, taking into account travel requirements; (2) *Availability* focuses on whether the volume and type of services offered adequately meet patient needs; (3) *Accommodation* pertains to features of the healthcare system such as appointment scheduling that align with patients' needs; (4) *Acceptability* considers whether patients find providers' attributes (*e.g.*, gender, race), acceptable and vice versa; and (5) *Affordability* addresses whether the costs of medical services are within the financial means of patients or their insurers. Ayer et al. (2014) provide an overview of research within operations management that examine one or more of these dimensions.

Disparities in healthcare access across socioeconomic groups are widely documented and evident in different healthcare contexts such as organ transplantation (Wang et al. 2022), use of decision support systems (Ganju et al. 2020), and medical appointment scheduling (Samorani et al. 2021). Equitable access to vaccinations has been examined in multiple studies with varying objectives (Enayati and Özaltın 2020, Rastegar et al. 2021). For instance, Bennouna et al. (2023) combine a predictive machine learning-based epidemic model with an optimization model to efficiently allocate vaccines to different populations. A recent review of the literature is available in Dey et al. (2024).

For patients with chronic diseases requiring ongoing treatment, distance to a health provider can be a major barrier to accessing care. Kelly et al. (2016) summarize research documenting the adverse effects of living farther from a healthcare facility. For instance, mammography use and radiation therapy for breast cancer are notably lower for women residing farther from a provider, even after adjusting for socioeconomic status. Survey data confirm that lack of transportation correlates with lower healthcare utilization, particularly in rural communities (Arcury et al. 2005).

During the COVID-19 pandemic, access to a nearby retail pharmacy was associated with higher vaccination rates (Brownstein et al. 2022). Despite a national COVID-19 vaccination campaign beginning in 2021, vaccine availability remained limited in rural and other socioeconomically disadvantaged communities, with a disproportionate impact on minority and elderly populations (Zhong et al. 2023, Rader et al. 2022). In examining one strategy to reduce such disparities, Chevalier et al. (2022) demonstrate that including all Dollar General stores in the federal vaccination program

could expand the share of the population within five miles of a vaccination site—with the largest benefits accruing to low-income and minority households.

Our case study extends these efforts by precisely estimating demand for a COVID-19 vaccine under service-provider capacity constraints, and formulating an optimization model to strategically choose vaccination delivery sites. To the best of our knowledge, this paper is the first to empirically estimate the heterogeneous vaccination elasticities to travel distance and we propose a quantitative methodology for evaluating potential policy solutions.

2.2. Optimal Facility Location

The optimization component of our framework relates to the extensive facility location literature, particularly the Maximal Covering Location Problem (MCLP) (Church and ReVelle 1974, Current and Storbeck 1988). Detailed reviews of facility location modeling in healthcare can be found in Daskin and Dean (2005) and Ahmadi-Javid et al. (2017), with further discussion on their application in emergency services provided by Jia et al. (2007).

Prior studies have integrated distance sensitivity into MCLP in various healthcare applications. In a primary care setting, for instance, Nobles et al. (2014) optimally assign patients to pediatricians, constrained by patients' travel willingness and provider capacity. Relatedly, McCoy and Johnson (2014) assign patients to HIV clinics, assuming a constant or linear decline in treatment adherence with increasing distance. Following the 2009 H1N1 influenza pandemic, Heier Stamm et al. (2017) model flu vaccine accessibility considering geographic factors and dose availability. Unlike these studies, which assume a homogeneous linear relationship between distance and service utilization, our approach estimates heterogeneous demand curves empirically.

In cases where demand is stochastic, MCLP models can allow for customers to queue at each location. Ekici et al. (2014) combine an MCLP with an epidemic model to project the spatial and temporal spread of an infectious disease. Deo and Sohoni (2015) model the assignment of HIV testing capacity assuming a declining probability of test result collection as turnaround time increases, and use simulation to evaluate different assignment policies. Relatedly, Jónasson et al. (2017) use a facility location model with congestion to assign capacity across geographically scattered HIV testing labs. The preceding studies focus on operational planning and logistics, whereas our framework pertains to more strategic decision-making surrounding equitable access to a healthcare service.

Most facility location modeling papers in healthcare assume the decision-maker is a central planner who can effectively assign individuals to different facilities. For example, Lee et al. (2013) create a decision support tool to optimally choose a set of mass medical dispensing sites during a widespread outbreak, but the authors note the computational challenges of performing a large-scale, multi-level optimization of selecting dispensing locations and assigning households to facilities.

During the COVID-19 pandemic, Bertsimas et al. (2022) combine a predictive compartmental epidemic model with an optimization model to select mass vaccination sites, assign populations to each site, and allocate vaccines by age group. Our study differs in the combinatorially greater number of potential vaccination sites, and the inclusion of empirically derived demand estimates.

Alternatively, *choice-based* facility location models examine how individual preferences, influenced by factors such as distance or service quality, can shape facility location strategies. Similar to prior studies in the healthcare context (Zhang et al. 2012, Denoyel et al. 2017, Krohn et al. 2021), we model individual choices using the multinomial-logit (MNL) discrete choice model (McFadden 1974). Our proposed formulation is essentially an assortment optimization problem under a mixture of MNL models, with customer segments having different distance elasticities, along with capacity and cardinality constraints. Bront et al. (2009) show that the problem is NP-hard if the number of customer types reaches the number of products. Others have proposed integer programming (*e.g.*, Méndez-Díaz et al. (2014)) and conic programming (*e.g.*, Sen et al. (2018)) solution approaches. While prior work focuses on the problem's theoretical properties and solution methodologies, we propose a real-world application that structurally estimates the MNL choice model from observational data, and uses it within a facility location modeling framework.

2.3. Demand Estimation

The primary input into a choice-based facility location model is the demand curve for obtaining the healthcare service, as a function of travel distance. To do so, we draw on the established literature on structural demand estimation in economics. More specifically, estimation follows a random-coefficient logit model using aggregate choice data, an approach referred to as "BLP" (Berry et al. 1995). The method has been widely used due to its ability to capture individual-level preference heterogeneity using only aggregate data, and to address endogeneity using instruments.

More advanced implementations of BLP incorporate techniques such as utilizing moments from disaggregated data (Petrin 2002, Berry et al. 2004), minimizing the impact of parametric error (Berry and Pakes 2007), or including spatial preferences (Davis 2006), among other variations. Many applications of BLP in the economics literature use structural models to measure industry behavior (Nevo 2001), or to quantify specific counterfactual scenarios (*e.g.*, Duch-Brown et al. (2023) studies online market integration). The operations management literature has gradually adopted BLP estimation for modeling demand, with applications to the fast-food industry (Allon et al. 2011), automobile sector (Guajardo et al. 2016), and online physician services (Xu et al. 2021). We expand on the classic BLP method by (1) incorporating capacity constraints via a fixed-point equation approach, and (2) embedding the estimated demand model into a prescriptive optimization model.

3. Estimate-then-optimize Framework

Our framework consists of a microfounded model of consumer demand for a healthcare service that relies only on aggregated data on service utilization. Demand curves represent the sum of decisions made by individuals with varying preferences and distances to a healthcare facility. Crucially, aggregating these individual-level decisions provides a structured way in which to estimate demand for a health service given the locations of facilities providing that service. Finally, we show how to incorporate the estimated demand model into an optimization framework to select facility locations to maximize use of the health service.

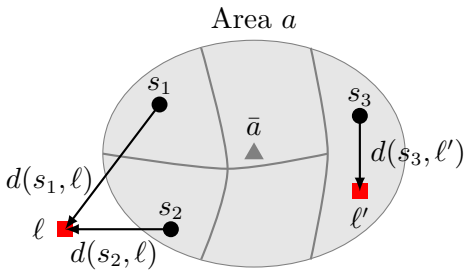
3.1. Demand Model

Consider a set of health facility locations $\mathcal{L}$ that provide an undifferentiated product or service, such as eye examinations, vaccinations, or blood collection. Each individual i can decide to receive service (referred to as an "appointment") at a location $\ell_i \in L_i$ from among the individual's feasible set of service locations, $L_i \subset \mathcal{L}$. The appointment location represents the facility that i most plausibly considers when choosing whether to receive the service and can vary across settings. For example, ℓ_i could be the most convenient location for i that has available service capacity. In contexts with limited insurance networks, ℓ_i might be the closest healthcare facility covered by i's insurance. In other contexts, such as a centrally organized blood drive or vaccination campaign, ℓ_i might be a service location assigned by a central planner.

A key feature of demand that we aim to capture is how consumers' decisions to use the health service depend on the proximity of where they live to service locations. Let $a_i \in \mathcal{A}$ denote the geographic area (*e.g.*, city or zip code) and $s_i \in a_i$ denote a smaller geographic subarea within a_i (*e.g.*, Census tract or block) where individual i resides. Let $d(s_i, \ell_i)$ denote a measure of travel distance between the centroid of s_i and ℓ_i. Figure 3 illustrates an example of an area a and depicts the distances from some of its subareas to healthcare facility locations ℓ and ℓ'. The distances shown represent how far individuals living in each subarea would need to travel for an appointment at their closest location.

The distance function $d()$ could be the geodesic distance, driving time, or transit time on public transportation. In addition, we allow preferences for the healthcare service to vary both systematically and idiosyncratically across geographic areas (*i.e.*, across $a \in \mathcal{A}$). For example, some communities have higher needs for the health service based on disease prevalence, or they may be particularly hesitant to receive medical services.

Figure 3 **Conceptual model of BLP demand estimation given aggregate outcome data by area a (with population centroid $\bar{a}$) and subarea-level (s_1, s_2 and s_3) variation in distance to a healthcare provider (ℓ and ℓ')**



3.1.1. Individual Preferences. We introduce the following utility-based binary choice decision model. Individual i may elect to obtain service at appointment location ℓ_i and receive additional utility u_i relative to no service:[1]

$$u_i = D_i + X_{a_i}\boldsymbol{\beta} + \xi_{a_i} + \epsilon_i. \tag{1}$$

The key term in this utility model is D_i, which represents the disutility that individual i experiences from traveling from her subarea s_i to her appointment at location ℓ_i. We define $D_i := d(s_i, \ell_i)\alpha_{a_i}$, where α_{a_i} represents the individual's area-specific sensitivity to travel distance. The variation we allow in α_{a_i} captures across-area variation in the difficulty of traveling for an appointment, such as differences in the availability of public transportation.[2] We denote the parameters characterizing distance sensitivity by the vector $\boldsymbol{\alpha}$. Note that the distance function $d()$ measures the individual's i distance from the centroid of s_i to location ℓ_i, a reasonable assumption given the small geographic size of subareas. The vector $\boldsymbol{\beta}$ captures the systematic relationship between the area's observable characteristics X_{a_i} (*e.g.*, demographics) and average preferences for obtaining the service. The term ξ_{a_i} represents the idiosyncratic preferences of area a_i due to unobservable factors, and ϵ_i is a logistic preference that is specific to i.

Under the utility model in equation (1), individual i's demand follows the classic binary logit discrete choice model where the probability that she obtains service is:

$$P_\epsilon\left(u_i \geq 0\right) = \frac{\exp\left(d(s_i, \ell_i)\alpha_{a_i} + X_{a_i}\boldsymbol{\beta} + \xi_{a_i}\right)}{1 + \exp\left(d(s_i, \ell_i)\alpha_{a_i} + X_{a_i}\boldsymbol{\beta} + \xi_{a_i}\right)}. \tag{2}$$

3.1.2. Demand Estimation. Given individual-level choice data with exogenous variation in s_i, l_i, and X_{a_i}, it is straightforward to estimate the demand model from Section 3.1 via maximum likelihood. However, such favorable settings for estimation are exceedingly uncommon. Notably,

[1] Note that to remain as consistent as possible with our empirical case study, equation (1) is not written in fullest generality. For example, it is straightforward to extend the model to allow β to vary with i (Berry et al. 1995).

[2] The model can be extended to allow variation across individuals to capture individual-level factors (*i.e.*, using α_i), such as vehicle ownership or job flexibility.

when consumer data are highly sensitive or proprietary—which occurs frequently in many healthcare settings—we may observe data only at an aggregate level (Table 1). Under such circumstances, we can estimate demand as a random coefficient logit model using aggregate data. This approach, known as "BLP" (Berry et al. 1995), is a common way to estimate demand in settings where individuals' choices are not directly observed—yet individual-level factors play a significant role in each individual's decision.

We consider the case where service utilization is only observed at the geographic area level. The BLP approach incorporates factors affecting demand that vary across individuals within a geographic area by treating these factors as *distributions*—*e.g.*, the income distribution within an area or the distribution of household locations within an area. These factors are commonly known as "random coefficients." Area-level demand curves are obtained by simply integrating individual-level demand curves over the distributions of random coefficients.

In our setting, the distance that each individual must travel for their appointment, $d(s_i, \ell_i)$, varies across individuals within a given area. Correspondingly, travel disutility $D_i := d(s_i, \ell_i)\alpha_{a_i}$ is a "random coefficient" within the BLP framework. The share of individuals in area $a \in \mathcal{A}$ receiving the service is computed by simply aggregating the demand from individuals living within a:

$$\begin{aligned}
\rho_a &:= \sum_{s_i \in a} \Big[\int P_\epsilon \left(u_i \geq 0 \mid D_i, s_i\right) dF\left(D_i \mid s_i\right) \Big] f(s_i \mid a) \\
&= \sum_{s_i \in a} \Big[\int \frac{\exp\left(D_i + X_a \boldsymbol{\beta} + \xi_a\right)}{1 + \exp\left(D_i + X_a \boldsymbol{\beta} + \xi_a\right)} dF\left(D_i \mid s_i\right) \Big] f(s_i \mid a) \\
&= \sum_{s_i \in a} \Big[\sum_{\ell_i \in L_i} \frac{\exp\left(d(s_i, \ell_i)\alpha_a + X_a \boldsymbol{\beta} + \xi_a\right)}{1 + \exp\left(d(s_i, \ell_i)\alpha_a + X_a \boldsymbol{\beta} + \xi_a\right)} f(\ell_i \mid s_i) \Big] f(s_i \mid a) \\
&= \sum_{s_i \in a} \sum_{\ell_i \in L_i} \frac{\exp\left(d(s_i, \ell_i)\alpha_a + X_a \boldsymbol{\beta} + \xi_a\right)}{1 + \exp\left(d(s_i, \ell_i)\alpha_a + X_a \boldsymbol{\beta} + \xi_a\right)} f(s_i, \ell_i \mid a),
\end{aligned} \tag{3}$$

where $F(\cdot)$ and $f(\cdot)$ denote cumulative distribution functions and probability density functions, respectively. The third equality follows from the definition of D_i and the distance measure $d(s_i, \ell_i)$ is defined by the set of discrete locations in L_i.

Equation (3) shows that, in computing the aggregate demand for area a, a key input is the density $f(s_i, \ell_i \mid a)$, which implies the distribution of travel cost. This density can be decomposed into a spatial distribution of the population of a and their appointment locations as follows:

$$f(s_i, \ell_i \mid a) := \underbrace{f(s_i \mid a)}_{\substack{\text{Spatial} \\ \text{population}}} \underbrace{f(\ell_i \mid s_i)}_{\text{Appointments}}. \tag{4}$$

We assume that the spatial distribution of the population across subareas within area a, $f(s_i \mid a)$, is observed. This assumption is reasonable in many settings given that publicly available data typically detail the spatial population distribution to an extremely fine level such as Census blocks.

The second term, $f(\ell_i \mid s_i)$,[3] characterizes the distribution of appointments, which is identified by a set of parameters, $\boldsymbol{\lambda}$, to be estimated. In some cases, the appointment process may be straightforward, such as a central planner assigning individuals to their closest facility. In other contexts, the appointment process may be decentralized or include additional complexities such as noise or capacity constraints. In our case study, the appointment location distribution $f(\ell_i \mid s_i)$ and its parameters $\boldsymbol{\lambda}$ are determined by a sequential appointment assignment process (discussed in the next section), allowing each individual to receive an appointment at her *preferred* service location with remaining capacity.

The parameters to be estimated are $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$. The BLP method exploits the insight that for any given $\boldsymbol{\theta}$, there is one unique set of idiosyncratic shocks $(\xi_a)_{a \in \mathcal{A}}$ that precisely rationalize the observed area-level service use (Berry 1994). Thus, in enforcing that the area-level service utilization implied by the demand model (ρ_a from equation 3) match the observed data, the unobserved area-level idiosyncratic preferences are an implicit function of $\boldsymbol{\theta}$. We indicate this relationship by denoting ξ_a as a function of $\boldsymbol{\theta}$.

Following Berry et al. (1995), we estimate $\boldsymbol{\theta}$ using orthogonality conditions:

$$E_a[Z_a(\boldsymbol{\theta})\xi_a(\boldsymbol{\theta})] = \vec{\mathbf{0}}. \tag{5}$$

This condition indicates that at the true parameter $\boldsymbol{\theta}$, the implied idiosyncratic preferences for each area, $\xi_a(\boldsymbol{\theta})$, are uncorrelated with a set of area-level instruments $Z_a(\boldsymbol{\theta})$. Since $\xi_a(\boldsymbol{\theta})$ is a utility residual from the demand model, equation (5) can be interpreted intuitively as enforcing that the residual variation in the utility model evaluated at the true $\boldsymbol{\theta}$ not be correlated with the instruments.

The instruments Z_a must be chosen in each setting so that the moment conditions are both informative and plausible. For instance, in our case study, Z_a includes the observable area-level characteristics X_a, as well as the average distance to service appointments in the area interacted with area-level demographics. Note that since service appointment locations vary with $\boldsymbol{\theta}$, these latter instruments vary with $\boldsymbol{\theta}$ as well. We therefore denote the instruments Z_a as a function of $\boldsymbol{\theta}$.

We leverage the moments in equation (5) via the generalized method of moments:[4]

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{|\mathcal{A}|} \sum_a Z_a(\boldsymbol{\theta})\xi_a(\boldsymbol{\theta}) \right)' \Phi \left(\frac{1}{|\mathcal{A}|} \sum_a Z_a(\boldsymbol{\theta})\xi_a(\boldsymbol{\theta}) \right), \tag{6}$$

[3] Note that we omit the conditioning on a because knowing s_i uniquely determines the individual's area, so $f(\ell_i \mid s_i, a) = f(\ell_i \mid s_i)$.

[4] Note that ξ_a is an implicit function of θ through the constraint that the model-implied market shares are equal to the empirical market shares, *i.e.*, $\rho_a(\boldsymbol{\theta}) = \rho_a^{Obs}$. This implicit function can be enforced at each step of the optimization either through the use of a contraction mapping (Berry et al. 1995) or solver (Reynaerts et al. 2012). Alternatively, in treating the equality as an equilibrium constraint, the GMM objective can be optimized using an MPEC approach as in (Dubé et al. 2012).

where optimal weighting matrix Φ is chosen via efficient two-step GMM (Hansen 1982). A description of the solution approach to obtain $\hat{\boldsymbol{\theta}}$ is provided in Appendix B.1.

3.1.3. Appointment Distribution. Thus far, we have presented the demand model and estimation quite generally to emphasize its flexibility. In this section, we present a specific model of sequential appointment assignment used in our estimation. The goal is to obtain an explicit structural model to connect appointment availability with individual choices.

We assume that an individual's choice of appointment location $\ell_i \in L_i$ is a function of its proximity $d(s_i, \ell_i)$ and her sensitivity to travel distance α_{a_i}. Under this assumption, no additional parameters are introduced in the model. In other words, to determine the distribution of appointments, $f(\ell_i \mid s_i)$, the only parameters needed are the sensitivity to distance $\boldsymbol{\alpha}$. Thus, in this setting no additional parameters ($\boldsymbol{\lambda}$) need to be estimated.

To model an individual's choice of appointment location, we assume a Type I extreme value (*i.e.*, Gumbel) distribution of idiosyncratic preferences over locations, resulting in an MNL model:

$$f(\ell_i \mid s_i, L_i) := \frac{\exp\left(d(s_i, \ell_i)\alpha_{a_i}\right)}{\sum_{\ell \in L_i} \exp\left(d(s_i, \ell)\alpha_{a_i}\right)}. \tag{7}$$

If an individual dislikes traveling longer distances (*i.e.*, $\alpha_{a_i} < 0$) then equation (7) implies that i's likelihood of choosing location ℓ is decreasing in distance. A reasonable alternative model that takes this to an extreme is to assume individual i chooses the closest site within L_i.[5] A key advantage to applying an MNL model instead is that it may better capture individuals' idiosyncratic preferences in their preferred location, such as preferring a site close to their workplace rather than home.

If all healthcare service locations had infinite capacity, then one could assume that individuals choose their appointment from among all sites (*i.e.*, that $L_i = \mathcal{L}$). In reality, however, healthcare service providers often have limited capacity so each individual's appointment options might depend on locations' available capacity. To capture this complexity in the model, we assume that appointments are fulfilled sequentially in a random order, where individual i's set of possible service locations ($L_i \subset \mathcal{L}$) are restricted to those *with remaining capacity*.[6] This sequential approach might capture a random arrival process of identical individual types requesting appointment slots, such as an online booking of vaccination appointments.

We implement the sequential approach by first generating a random order of all individuals in the population. Then, we iterate over the individuals and update each choice set L_i according to the appointment location choices (equation 7) and service choices (equation 2) made by individuals

[5] In our case study, we verify that our demand estimates are similar under this alternative model (see Appendix C.1).

[6] One could further limit an individual's choice set to locations within a maximum distance to avoid assigning appointments that are extremely far from each individual's subarea.

$1, ..., i-1$, so that only locations with available capacity are included in L_i. Thus, if a location ℓ reaches capacity at individual j's turn, then ℓ is removed from the choice set of all individuals making decisions after j, *i.e.*, $L_i \subset \mathcal{L} \backslash \{\ell\}\ \forall i > j$. Under this approach, $f(\ell_i \mid s_i)$ in equation (4) can be obtained by integrating $f(\ell_i \mid s_i, L_i)$ in equation (7) over the L_i for individuals who reside in subarea $s_i \in a$.

3.2. Optimization Model: Location Selection

We next formulate an optimization model to determine the optimal set of healthcare service locations to maximize the predicted demand served across all areas.

Let $L \subset \mathcal{L}$ be the decision variable corresponding to the set of locations where service is offered, termed *open locations*. Each candidate location $\ell \in \mathcal{L}$ has a capacity denoted by K_ℓ, and we assume that candidate locations are, otherwise, identical. We further assume that policymakers have a limited budget for the number of locations offering service and include a cardinality constraint on the number of open locations, *i.e.*, $|L| \leq N$.

We consider demand at the aggregate subarea level $s \in \mathcal{S}$. Let m_s be the size of the *eligible* population in subarea s who could receive service. Define $\rho_{s\ell}$ as the aggregate share of the population in subarea s (with parent area a), that obtains service at location ℓ. Based on the demand model in Section 3.1, we derive an expression for $\rho_{s\ell}$ as follows:

$$\begin{aligned}
\rho_{s\ell} &:= \frac{1}{m_s} \sum_{\{i \mid s_i = s\}} P_\epsilon(u_i > 0 \mid \ell, s)\ f(\ell \mid s), \quad \text{if } \ell \in L \\
&= \frac{1}{m_s} \sum_{\{i \mid s_i = s\}} \left[\frac{\exp\left(d(s,\ell)\alpha_a + X_a\boldsymbol{\beta} + \xi_a\right)}{1 + \exp\left(d(s,\ell)\alpha_a + X_a\boldsymbol{\beta} + \xi_a\right)} \right] \left[\frac{\exp(d(s,\ell)\alpha_a)}{\sum_{k \in L} \exp(d(s,k)\alpha_a)} \right], \quad \text{if } \ell \in L \\
&= \left[\frac{\exp\left(d(s,\ell)\alpha_a + X_a\boldsymbol{\beta} + \xi_a\right)}{1 + \exp\left(d(s,\ell)\alpha_a + X_a\boldsymbol{\beta} + \xi_a\right)} \right] \left[\frac{\exp(d(s,\ell)\alpha_a)}{\sum_{k \in L} \exp(d(s,k)\alpha_a)} \right], \quad \text{if } \ell \in L
\end{aligned} \tag{8}$$

and zero otherwise. Equation (8) results from equations (2) and (7). Note that the second term in brackets assumes that individuals select appointment locations without considering available capacity—*i.e.*, individuals can select any of the L open locations. This is in contrast to Section 3.1.3, where a sequential appointment fulfillment process that tracks remaining capacity at open locations is assumed. Instead, we implement a constraint that guarantees that the predicted number of served individuals at each open location ℓ does not exceed its capacity K_ℓ. The optimal set of open locations L is selected to ensure consistency with the assumption above, where, in expectation, the demand served at each location is kept below its capacity.

For all subarea-location pairs (s, ℓ), we compute $\rho_{s\ell}$ using equation (8) and the demand model estimates for sensitivity to travel distance $\hat{\alpha}_a$, demographic-based preferences $\hat{\boldsymbol{\beta}}$, and idiosyncratic preferences $\hat{\xi}_a$. To ease notation, let $\bar{u}(s,\ell) := d(s,\ell)\hat{\alpha}_a + X_a\hat{\boldsymbol{\beta}} + \hat{\xi}_a$. Equation (8) simplifies to:

$$\rho_{s\ell} = \left[\frac{\exp(\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))}\right]\left[\frac{\exp(\bar{u}(s,\ell)-X_a\hat{\boldsymbol{\beta}}-\hat{\xi}_a)}{\sum_{k\in L}\exp(\bar{u}(s,k)-X_a\hat{\boldsymbol{\beta}}-\hat{\xi}_a)}\right], \quad \text{if } \ell \in L \tag{9}$$
$$= \frac{\frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))}}{\sum_{k\in L}\exp(\bar{u}(s,k))}, \quad \text{if } \ell \in L$$

and zero otherwise. In equation (9), the term $\exp(-X_a\hat{\boldsymbol{\beta}}-\hat{\xi}_a)$ in the right bracket cancels out. Finally, the predicted demand from subarea s to open location ℓ is given by $m_s\rho_{s\ell}(L)$.

To formulate the optimization problem, we encode the main decision variable, the set of open locations $L \subset \mathcal{L}$, as a vector of binary decision variables $\mathbf{x} \in \{0,1\}^{|\mathcal{L}|}$: $x_\ell = 1$ if $\ell \in L$ and $x_\ell = 0$ otherwise. Thus, we can rewrite $\rho_{s\ell}$ as a function of $\mathbf{x}$ as follows:

$$\rho_{s\ell}(\mathbf{x}) = \frac{\frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))}x_\ell}{\sum_{k\in\mathcal{L}}\exp(\bar{u}(s,k))x_k}, \quad \text{if } \sum_{k\in\mathcal{L}} x_k \geq 1,$$

and zero otherwise. We introduce a decision variable $y_{s\ell} \in [0,1]$ to allow for partial fulfillment of demand, that is, $y_{s\ell}$ is the fraction of subarea s demand to location ℓ that is served at that location. We incorporate the partial fulfillment idea to add flexibility in how demand is served: a location ℓ can be opened even though its available capacity may be smaller than total demand, so only a fraction of demand may be served. The corresponding optimization problem is:

$$(\mathsf{P}_{\mathsf{loc}}) \qquad \max_{\mathbf{x},\mathbf{y}} \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}} m_s\rho_{s\ell}(\mathbf{x})y_{s\ell} \tag{10a}$$
$$\text{s.t.} \sum_{s\in\mathcal{S}} m_s\rho_{s\ell}(\mathbf{x})y_{s\ell} \leq x_\ell K_\ell, \ \forall \ell \in \mathcal{L} \tag{10b}$$
$$\sum_{\ell\in\mathcal{L}} x_\ell \leq N \tag{10c}$$
$$0 \leq y_{s\ell} \leq x_\ell, \ x_\ell \in \{0,1\}, \ \forall \ell \in \mathcal{L}, \ s \in \mathcal{S}. \tag{10d}$$

The objective (10a) maximizes the expected demand served across all subareas and locations. Constraints (10b) ensure that the expected total demand served by an open location does not exceed its capacity. Note that multiple locations may serve the demand of a subarea, which is determined by individuals' preferences reflected in the probabilities $\rho_{s\ell}(\mathbf{x})$. Constraint (10c) imposes a total budget of N open locations. Finally, constraints (10d) ensure that partial fulfillment of a subarea demand from a location is only possible if the corresponding location is open.

The problem $(\mathsf{P}_{\mathsf{loc}})$ is essentially an assortment optimization problem with cardinality, capacity, and partial fulfillment considerations. The formulation, however, is computationally challenging because of nonlinear objective function and constraints, which rely on the choice probabilities $\rho_{s\ell}(\mathbf{x})$. Using ideas from assortment optimization under an MNL choice model (*e.g.*, Bront et al. (2009)), we next propose an equivalent, and tractable, formulation.

We introduce an auxiliary variable $v_s \geq 0$ defined as:

$$v_s = \begin{cases} \frac{1}{\sum_{k\in\mathcal{L}} \exp(\bar{u}(s,k))x_k} & \text{if } \sum_{k\in\mathcal{L}} x_k \geq 1 \\ 0 & \text{o.w.} \end{cases},$$

and a binary decision variable $b_s \in \{0,1\}$, which together satisfy the following set of constraints:

$$v_s \sum_{k\in\mathcal{L}} \exp(\bar{u}(s,k))x_k = b_s \quad \forall s \in \mathcal{S}, \tag{11a}$$

$$v_s \leq M_s b_s \quad \forall s \in \mathcal{S}, \tag{11b}$$

$$b_s \geq x_\ell \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}. \tag{11c}$$

where $M_s = \bar{v}_s := \max_{k\in\mathcal{L}}[\exp(\bar{u}(s,k))^{-1}]$. We note that after incorporating these auxiliary variables, the objective function (10a) and constraints (10b) and (11a) are still nonlinear. Following the linearization approach proposed by Wu (1997), we define $w_{s\ell} := v_s x_\ell$, and introduce the set of linear constraints:

$$v_s - w_{s\ell} \leq M_s - M_s x_\ell \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}, \tag{12a}$$

$$w_{s\ell} \leq v_s \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}, \tag{12b}$$

$$w_{s\ell} \leq M_s x_\ell, \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}. \tag{12c}$$

The resulting (equivalent) optimization problem to ($\mathsf{P}_{\mathsf{loc}}$) is:

$$\begin{aligned} (\mathsf{P}'_{\mathsf{loc}}) \qquad \max_{\mathbf{x},\mathbf{y},\mathbf{b},\mathbf{v},\mathbf{w}} \ & \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}} m_s \frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))} w_{s\ell} y_{s\ell} \\ \text{s.t.} \ & \sum_{s\in\mathcal{S}} m_s \frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))} w_{s\ell} y_{s\ell} \leq x_\ell K_\ell, \ \forall \ell \in \mathcal{L} \\ & \sum_{\ell\in\mathcal{L}} x_\ell \leq N \\ & \text{Constraints: (11a)} - \text{(11c)} \text{ and (12a)} - \text{(12c)} \\ & 0 \leq y_{s\ell} \leq x_\ell, \ x_\ell, b_s \in \{0,1\}, \ v_s, w_{s\ell} \geq 0, \ \forall s \in \mathcal{S},\ \ell \in \mathcal{L}. \end{aligned}$$

For the bilinear terms $w_{s\ell}y_{s\ell}$, where $w_{s\ell} \in [0, M_s]$, $y_{s\ell} \in [0,1]$, we can further relax this non-convex problem using the standard McCormick inequalities (McCormick 1976). Let's define $z_{s\ell} := w_{s\ell}y_{s\ell}$, and include the following constraints:

$$z_{s\ell} \leq w_{s\ell} \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}, \tag{14a}$$

$$z_{s\ell} \leq M_s y_{s\ell} \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}, \tag{14b}$$

$$z_{s\ell} \geq w_{s\ell} + M_s y_{s\ell} - M_s \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}. \tag{14c}$$

Thus, we can approximate the optimization problem $(\mathsf{P}'_{\mathsf{loc}})$ via the following relaxation:

$$
\begin{aligned}
(\tilde{\mathsf{P}}_{\mathsf{loc}}) \qquad \max_{\mathbf{x},\mathbf{y},\mathbf{b},\mathbf{v},\mathbf{w},\mathbf{z}} \quad & \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}} m_s \frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))} z_{s\ell} \\
\text{s.t.} \quad & \sum_{s\in\mathcal{S}} m_s \frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))} z_{s\ell} \le x_\ell K_\ell, \ \ \forall \ell \in \mathcal{L} \\
& \sum_{\ell\in\mathcal{L}} x_\ell \le N \\
& \text{Constraints: (11a)} - \text{(11c)}, \ \text{(12a)} - \text{(12c)}, \ \text{and (14a)} - \text{(14c)} \\
& 0 \le y_{s\ell} \le x_\ell, \ \ x_\ell, b_s \in \{0,1\}, \ \ v_s, w_{s\ell}, z_{s\ell} \ge 0, \ \ \forall s \in \mathcal{S}, \ \ell \in \mathcal{L}.
\end{aligned}
$$

Here $\mathbf{b}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z}$ are auxiliary variables used to linearize and approximate the original problem $(\mathsf{P}_{\mathsf{loc}})$. The resulting formulation $(\tilde{\mathsf{P}}_{\mathsf{loc}})$ is a mixed-integer linear program, amenable to commercial optimization solvers.

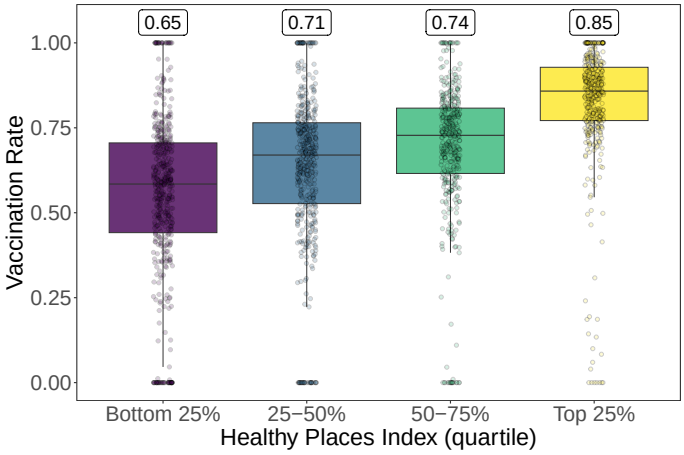
3.3. Counterfactual Evaluation

The primary outcome of $(\tilde{\mathsf{P}}_{\mathsf{loc}})$ is the set of optimally chosen locations, captured by $\tilde{\mathbf{x}}^*$ and the corresponding $\tilde{L}^* := \{\ell \in \mathcal{L} | \tilde{x}^*_\ell = 1\}$. This optimization problem, however, assumes individuals select service locations simultaneously while satisfying capacity constraints only in expectation. This contrasts with the sequential decision-making often observed in reality, where individuals take into account locations' remaining capacity when making their choices. While direct optimization of this sequential decision-making process is computationally intractable, we can evaluate the performance of the chosen locations, $\tilde{L}^*$, more realistically by using the sequential appointment assignment process used in the demand estimation (Section 3.1.3). To do so, we first randomly order all individuals in the population. Then, each individual selects their appointment location, and whether to obtain service, among the open locations $\tilde{L}^*$ with remaining capacity, according to the estimated MNL probabilities. Once a site reaches its capacity, it is removed from the set of open locations for all subsequent individuals. Hence, for a given $\tilde{L}^*$, we report the predicted demand served under the sequential appointment fulfillment process as the main outcome of our approach.

4. Case Study: COVID-19 Vaccine Distribution in California

Seeking to expand COVID-19 vaccination distribution points, in early 2021, the U.S. government launched the Federal Retail Pharmacy Program (FRPP), a partnership with 21 national pharmacy chains and independent pharmacy networks. By late-August 2023, nearly two years after the program launched, more than 300 million doses were administered through FRPP, across 41,000 retail locations nationwide (CDC 2023). Nearly half of all Americans live within one mile of a pharmacy

Figure 4 **Zip code-level COVID-19 vaccination rates in California, by HPI quartile (as of March 1, 2022)**



Note: Numbers above each boxplot are aggregated vaccination rates across all zip codes within each HPI quartile.

offering COVID-19 vaccinations and 89% live within five miles, yet travel distances vary widely—particularly for Black communities who often encounter distances in excess of ten miles (Berenbrok et al. 2021). This disparity extends beyond rural locales, with limited COVID-19 vaccination access points documented among communities of Color and those with chronic health conditions (Rader et al. 2022). Figure 1 depicts the inequalities in vaccination access within California, with the most vulnerable communities in the bottom HPI quartile encountering travel distances seven times greater than those in the top quartile. Vaccination rates are 20 percentage-points lower in the bottom HPI quartile (Figure 4), in part due to these structural barriers to vaccination.

Employing our two-stage estimate-then-optimize framework, we evaluate the effects of expanding FRPP to include partnerships with other organizations, to help alleviate a major barrier to vaccination access. Using an estimated spatial demand model (stage 1), we predict COVID-19 vaccination uptake under counterfactual scenarios of offering vaccines at these additional locations—which are optimally selected by the facility location model (stage 2). Our case study offers insights into how a national vaccine campaign might be strategically improved, underscoring the framework's utility in various healthcare settings.

Our primary analysis focuses on dollar store partnerships, but we also examine other private (*e.g.*, Starbucks) and public (*e.g.*, high schools) partners. Following the launch of FRPP in 2021, the CDC director and Dollar General confirmed reports of a potential partnership to administer COVID-19 vaccines at stores. Dollar General, known for its discounted prices and wide product selection, has thrived by locating in areas with little retail competition (Wolfrath et al. 2018). The chain operates over 17,000 stores in 46 states, nearly double the next largest private retailer offering COVID-19 vaccinations, and 75% of its stores serve rural communities (Bomey 2021, Roberts 2021). Two other chains, Dollar Tree and Family Dollar, each operate over 7,000 stores nationwide. These

discount stores often locate in communities underserved by retail pharmacies, creating a promising opportunity to improve vaccination access for low-income households (Chevalier et al. 2022).

4.1. Data

Our case study combines data across three domains: vaccination rates, spatial demographics, and store locations.

Vaccination rates. We obtain the most granular population-level COVID-19 vaccination data available in California: weekly vaccinations administered by zip code (California Department of Public Health 2021). Our main outcome variable is the proportion of the population aged 5+ who are fully vaccinated as of March 1, 2022.

Demographics. Demographic data are obtained from two sources. Our main demographic variable of interest, the Healthy Places Index (HPI), is a composite score constructed by the Public Health Alliance of Southern California (2022) that measures the well-being of a community. An area's HPI index is a weighted score across eight domains: socioeconomics, education, health care access, housing, neighborhood conditions, pollution/clean air, social factors, and transportation access (Maizlish et al. 2019). We group each region (*e.g.*, Census tract or zip code) into HPI quartiles, where the top 25% are the healthiest and the bottom 25% are the most vulnerable.

We use population estimates for residents aged 5+ for each of California's 377,023 Census blocks,[7] small geographic regions contained within a Census tract, with approximately 100 residents on average. We also obtain the geographic coordinates of blocks' population centroids, thus characterizing the entire spatial distribution of California's vaccine-eligible population (age 5+) of 37,336,716 (US Census Bureau 2019). Additional demographic data are obtained from the 2019 American Community Survey for 1,751 zip codes in California. Key variables include race/ethnicity, unemployment rate, poverty rate, college graduation rate, median household income, median home value, population, population density, and health insurance status. Appendix Table A1 gives summary statistics.

Vaccination locations. Location information including the geographic coordinates (latitude and longitude) of the 4,035 retail pharmacies in California participating in FRPP are from `https://www.vaccines.gov/` as of June 2021. Geodesic distances are computed between each vaccination site and the population centroid of each Census tract or block (US Census Bureau 2019). We do not include other dispensing sites such as mass vaccination centers or primary care offices because pharmacies accounted for 90% of COVID-19 vaccinations administered nationwide (IQVIA Institute for Human Data Science 2023).

[7] We only consider Census blocks with non-zero population.

Potential alternative vaccination sites include the three largest dollar store chains (Dollar General, Dollar Tree, and Family Dollar), which collectively operate 1,016 stores across California. Dollar store locations (latitude and longitude) are from ScrapeHero (2021). Public school locations are obtained from California Department of Education (2020). All other locations are obtained from Safegraph.

4.2. Implementing the Estimate-then-Optimize Framework

We describe the model parameters and include specific details on how to operationalize each step of the framework.

4.2.1. Estimating Demand for Vaccinations. Geographic areas ($\mathcal{A}$) correspond to zip codes because the primary outcome, COVID-19 vaccination rates (ρ_a^{obs}), is available only at the zip code-level. Subareas ($\mathcal{S}$) correspond to Census blocks, which are typically less than one acre in size, allowing us to account for variations in travel distance at an extremely granular level. The set of locations $\mathcal{L}$ corresponds to pharmacies participating in FRPP. We compute the geodesic distance $d(s,\ell)$ from the centroid of each block s to every vaccination location ℓ. This assumes that all residents in a block share the same travel distance to any given site, a reasonable assumption given Census blocks' small size.

Here, we assume the sensitivity to travel distance α_{a_i} varies by HPI quartile $\mathcal{H} := \{1,2,3,4\}$, *i.e.*, $\alpha_{a_i} = \alpha_{h(a_i)}$ where $h(a_i)$ denotes the HPI quartile of the zip code in which individual i resides. Thus, $\boldsymbol{\alpha}$ is a vector of four parameters representing the distance sensitivities for each HPI quartile. Consequently, the model allows travel distance disutility to vary with socioeconomic status. For example, proximity may play a greater role in the vaccination decisions of lower-HPI individuals who face additional barriers such as limited access to transportation, online appointment scheduling, or time off work.

An individual's utility from vaccination (equation 1) incorporates zip code-level demographic variables, X_{a_i}. These include race/ethnicity composition (Black, Asian, Hispanic, and Other), composition of health insurance (employer-provided, Medicare, Medicaid, and others), fraction of college graduates, unemployment rate, poverty rate, median household income, median home value, and population density. Including these factors in the utility model helps capture the influence that community demographics exert on latent preferences for vaccination.

In our case study, we limit each subarea's set of possible vaccination locations to be its M closest sites, to ease computation and avoid diluting the choice problem with excessively distant appointment locations. In both the estimation and optimization stages, individuals can only receive

an appointment at one of these M locations.[8] In our main analysis, we set $M = 5$, and as a robustness check, we vary M and report demand estimates (Appendix C.1) and optimization results (Appendix C.2).

Capacity per vaccination site is assumed to be constant across locations $K_\ell = 10,000 \quad \forall \ell \in \mathcal{L}$. We determine this value so the resulting supply (40.35 million vaccines) is sufficient to cover California's vaccine-eligible (aged 5+) population of 37,336,716. We also vary $K_\ell \in \{8,000,\ 12,000\}$ in sensitivity analyses (see Appendix C.1 for demand estimates and Appendix C.2 for optimization results).

Following Section 3.1.3, we model the allocation of appointments among individuals given capacity constraints. To do so, we first randomly order all 37.34 million individuals in California. Then, individuals are assigned appointments and make their decisions to vaccinate (or not) in that order. Each individual's set of possible vaccination locations L_i is determined by the available capacity at nearby sites after all previous individuals have had the chance to obtain a vaccination. We test the robustness of our demand estimates to the appointment assignment mechanism in Appendix C.1.

Our initial demand estimates measure travel costs using log-distance between each subarea and vaccination location. This functional form may cause the optimization model to prioritize reducing already small travel distances, while undervaluing the importance of reducing longer travel distances. To address this potential bias, we conduct a robustness check by considering an alternative distance function that treats all distances below a threshold $\bar{d}$ as constant. We explore various thresholds $\bar{d} = \{0.5 \text{ miles}, 1 \text{ mile}\}$ (see Appendix C.1 for demand estimates and Appendix C.2 for optimization results).

For computational tractability, we formulate the facility location problem at the Census tract-level. We obtain demand estimates for 8,000 Census tracts as follows. First, we compute $d(s, \ell)$ between all Census blocks and vaccination sites (FRPP pharmacies and partner locations). Second, we compute utility estimates at the block-level using the estimated demand model parameters. Finally, we aggregate demand to the Census tract-level and compute utility estimates at this level, which serve as inputs to the optimization model (see Appendix B.2 for details).

4.2.2. Optimizing Location Selection. Our primary counterfactual analysis evaluates predicted vaccination rates if vaccinations are offered at optimally selected dollar stores, out of the 1,016 possible locations statewide. Additionally, we assess partnering with selected high schools (out of 1,225 locations statewide) or Starbucks coffee houses (out of 3,172 locations statewide), as

[8] To estimate demand via BLP, all individuals must be assigned an appointment location. In the unusual case where all M closest locations are at capacity, we assign the individual to the Mth location. While this assumption does not strictly respect capacity constraints, we find that our demand estimates are similar as M increases (see Appendix C.1 for discussion).

detailed in Appendix C.2. We report the benefits of these partnerships relative to a benchmark "Pharmacy-only" policy, which utilizes only the $N = 4,035$ pharmacies participating in FRPP in California as of June 2021. For simplicity, we assume that partner locations have the same capacity as the existing pharmacies, though this assumption could be relaxed.

We let $\mathcal{L}$ include all FRPP pharmacies and any new candidate locations. The optimization model $(\tilde{\mathsf{P}}_{\mathsf{loc}})$ selects the optimal set of new vaccination sites to open, assuming all existing pharmacies remain operational; we refer to this as the *network expansion* scenario. Let A denote the number of additional partner locations to be selected. We consider $A = \{100,\ 200,\ \ldots,\ 1000\}$ locations, resulting in $N = 4,035 + A$ total vaccination sites. We let $x_\ell = 1$ if ℓ is a FRPP pharmacy. The primary outcome measure is the predicted increase in vaccinations compared to the Pharmacy-only benchmark. Evaluating the program's expansion in this manner provides valuable insights into budgetary or operational constraints (*e.g.*, staffing costs) and could inform the design of a pilot expansion program.

Finally, to maintain consistency we employ the same parameter values throughout the estimation and optimization stages. We perform extensive sensitivity analyses with respect to various parameter values and utility forms to assess the robustness of our findings, as detailed in Appendix C. Computationally, we solve $(\tilde{\mathsf{P}}_{\mathsf{loc}})$ using Python 3 and the Gurobi solver. The algorithm is configured to halt execution once the optimality gap falls below 0.1%, or runtime reaches six hours. We report the optimality gap between $(\tilde{\mathsf{P}}_{\mathsf{loc}})$ and $(\mathsf{P}_{\mathsf{loc}})$ in Appendix B.3.

4.3. Results

We first present our findings from the vaccination demand estimation model. We then utilize these estimates to predict demand at the optimal set of counterfactual locations, to illustrate the potential advantages of broadening FRPP through alternative partnerships.

4.3.1. Demand Estimates. The estimated parameters for the demand model, presented in Table 2, reveal key aspects of vaccine demand including variations linked to demographic factors. While these estimates are highly suggestive, individual coefficients should be interpreted with some caution due to the high correlation among many demographic variables.[9] Notably, the Healthy Places Index—a composite measure of health within a community—emerges as a strong predictor of vaccination uptake, with the least healthiest areas having conditionally lower vaccination rates. This finding suggests that the spatial inequalities observed in vaccination rates mirror those seen in other dimensions of health. Additionally, other forms of inequality are evident, such as more educated zip codes and those with higher employment rates exhibiting greater vaccine demand.

[9] For example, households with employer-provided health insurance typically have lower unemployment and poverty rates.

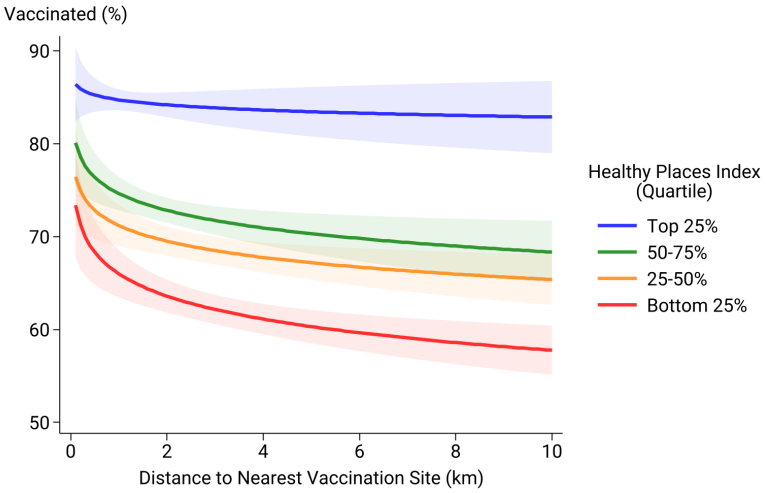
Table 2 Structural demand estimates

	Model Outcome: *Fraction Fully Vaccinated*	
Independent variable	Coef.	Std. error
HPI Quartile 4 (most healthy)	Ref.	
HPI Quartile 3	-0.304***	(0.084)
HPI Quartile 2	-0.451***	(0.112)
HPI Quartile 1 (least healthy)	-0.605***	(0.149)
Log-distance × HPI Quartile 4	-0.063	(0.070)
Log-distance × HPI Quartile 3	-0.142***	(0.050)
Log-distance × HPI Quartile 2	-0.124***	(0.043)
Log-distance × HPI Quartile 1	-0.161***	(0.043)
Race White	Ref.	
Race Black	-0.052	(0.342)
Race Asian	2.001***	(0.179)
Race Hispanic	0.987***	(0.167)
Race Other	7.077**	(3.049)
Health Insurance: Employer	-1.202*	(0.696)
Health Insurance: Medicare	0.197	(0.739)
Health Insurance: Medicaid	-0.973	(0.729)
Health Insurance: Other	-2.319***	(0.760)
College Graduation Rate	1.727***	(0.282)
Unemployment Rate	-1.777***	(0.686)
Poverty Level	-0.011	(0.472)
Log(Median Household Income)	-0.110*	(0.060)
Log(Median Home Value)	0.114**	(0.050)
Log(Population Density)	-0.078***	(0.022)
Constant	1.813**	(0.721)

Note. Estimates represent a demand model of the share of eligible individuals that are fully vaccinated for COVID-19 as of March 1, 2022. The parameters are estimated via the procedure described in Section 3.1.2. Healthy Places Index (HPI) is a composite measure of a community's health and well-being. Log-distance refers to the distance between the centroid of the individual's Census block and the location of their modeled vaccination appointment. Significance levels: *p<0.05; **p<0.01; ***p<0.001

Given our study's focus on optimizing healthcare facility locations, the key estimates for our analysis are the distance sensitivities: $(\alpha_h)_{h\in\mathcal{H}}$ for HPI quartiles $\mathcal{H} = \{1, 2, 3, 4\}$. The estimates reveal a distinct, statistically significant, and economically meaningful pattern: α_h is negative and statistically significant for all HPI quartiles except the top one, whose coefficient is much smaller in magnitude and not statistically distinguishable from zero. Importantly, these distance sensitivities remain significant after controlling for observable Census demographics, including race/ethnicity, health insurance coverage, college graduation rate, unemployment rate, poverty rate, and median household income, among others. Therefore, the estimated sensitivity of vaccine demand to distance

Figure 5 Predicted COVID-19 vaccination rates



Note. The figure plots the predicted vaccination rates under the estimated demand parameters at different distances to the offered vaccination site. Within each HPI quartile, we compute the mean vaccination rate assuming that all individuals are assigned to a location with the given distance.

cannot be solely explained by spatial variation in these demographics, and likely reflects a genuine sensitivity to distance.

One important implication of these estimates is that residents of less healthy communities experience greater reductions in vaccination rates as the distance to their vaccination site increases. For example, doubling the average distance from 1 km to 2 km results in a predicted 2.4 percentage-point decrease in vaccination rates for residents in the bottom HPI quartile. In contrast, those in the top HPI quartile experience a reduction of just 0.5 percentage points. These findings are consistent with the observation that traveling longer distances for vaccination appointments is most burdensome for individuals with lower socioeconomic status, who may lack access to transportation, work flexibility, etc. Figure 5 illustrates the demand estimates by plotting predicted vaccination rates at different travel distances, by HPI quartile. The sizeable gaps between curves highlight both the inequality in vaccination uptake across HPI quartiles, regardless of distance, as well as the considerably greater sensitivity of demand to distance for those in the bottom three HPI quartiles.

The preceding evidence demonstrates a strong correlation between distance to a vaccination site and vaccination uptake. One potential threat to this model specification is the possibility of selection bias among pharmacies included in FRPP, which might arise if the selection process is *ex-ante* correlated with both the propensity to get vaccinated and HPI—beyond what is captured by observable demographics. However, this seems unlikely given that many participating locations are part of national chains (*e.g.*, CVS, Walgreens), which had established store locations well in

advance of the COVID-19 vaccine gaining FDA approval. While retail pharmacies do tend to locate in higher income, urban communities, including additional demographics helps to control for this.

Appendix C.1 shows that our demand estimates remain qualitatively similar under a range of alternative specifications. These include alternative functional forms for distance, choice sets, capacity constraints, and appointment allocations. In all cases, the estimates continue to reflect significant inequality in vaccination by HPI, as well as distance sensitivities that are greater for low-HPI individuals than high-HPI individuals.

4.3.2. Optimization. Under the Pharmacy-only policy, our model predicts an overall COVID-19 vaccination rate of 71%, close to the observed rate of 70% in California. Predicted vaccination rates widely vary across HPI quartiles, with a 20 percentage-point difference between the top and bottom HPI quartiles (Table 3), which broadly aligns with empirical rates (Figure 4). This gap in predicted vaccination rates arises for two reasons. First, individuals in the top HPI quartile tend to live closer to pharmacies participating in FRPP (Figure 1). Second, these residents are less sensitive to long distances and are thus willing to travel farther to be vaccinated, as illustrated in Figure 5.

Of the 26.4 million predicted vaccinations statewide, nearly three-quarters (19.5 million) are obtained within walking distance of one's home Census tract (*i.e.*, within one mile or 1.6 km). Nevertheless, residents of the lowest HPI quartile who receive a vaccination are less likely to obtain it at a pharmacy within walking distance, compared to those in higher HPI communities.

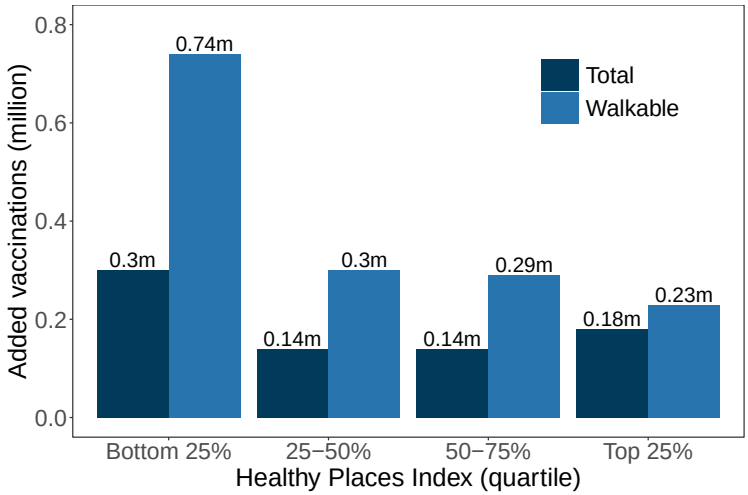
We evaluate a network expansion policy that optimally selects $A = \{100,\ 200,\ \ldots,\ 1000\}$ dollar stores from 1,016 possible locations in California. As expected, network expansion significantly improves predicted vaccination rates. For instance, augmenting the existing set of 4,035 FRPP pharmacies with 500 additional *optimally selected* dollar stores results in 0.77 million additional vaccinations statewide. This represents a 2.9% increase over the Pharmacy-only benchmark, equating to a marginal gain of approximately 1,500 vaccinations per added dollar store.

Table 3 Predicted vaccinations under Pharmacy-only policy

	Vaccinations in millions (% of population)				
		HPI quartile			
	All	Bottom 25%	25-50%	50-75%	Top 25%
Pharmacy-only					
Total	26.44 (71%)	5.62 (61%)	6.68 (69%)	6.83 (72%)	7.31 (81%)
Walkable	19.46 (52%)	3.98 (44%)	5.07 (53%)	4.96 (52%)	5.45 (60%)

Note. Walkable distances are defined as travel distances less than one mile (1.6 km).

Figure 6 **Added vaccinations relative to Pharmacy-only, assuming 500 dollar stores are added statewide**

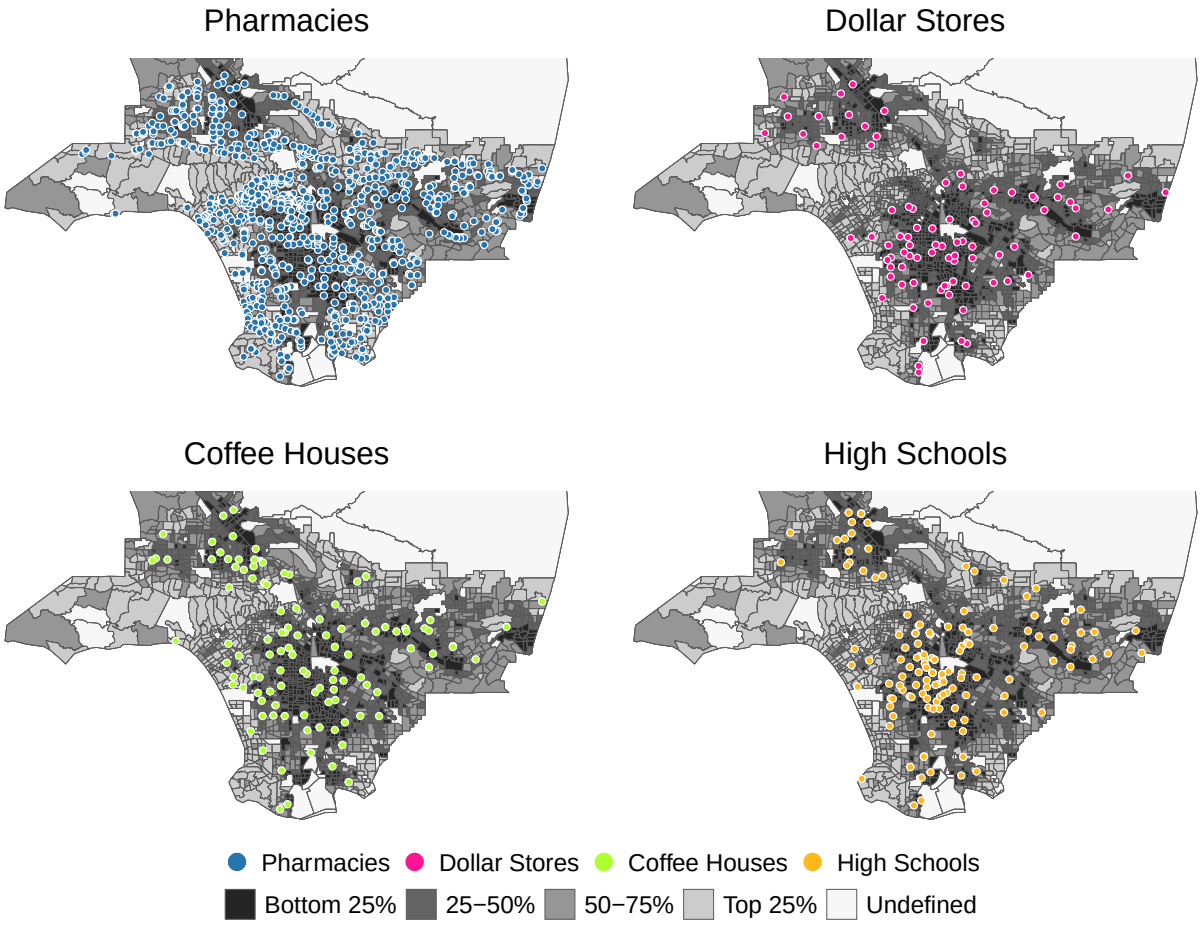


Notably, the benefits of adding dollar stores largely accrue to communities in the bottom HPI quartile, who experience a 5.4% increase in vaccinations compared to a 2.5% increase in the top quartile (Figure 6). A majority of the model-selected dollar stores fall within low HPI areas: 229 dollar stores open in tracts in the bottom HPI quartile. This boosts *walkable* vaccinations by 19% in precisely the communities with lower automobile access and higher shares of workers commuting by public transit, walking, or cycling (Public Health Alliance of Southern California 2022).

Maps depicting the geographic diversity of the existing FRPP pharmacies and selected dollar stores in Los Angeles County are shown in Figure 7, where darker regions correspond to lower HPI tracts. Fewer pharmacies are found in south Los Angeles, San Bernardino (east), and San Fernando Valley (north), some of the more economically disadvantaged areas of the county (US Census Bureau 2019). A network expansion policy of 500 additional dollar stores statewide optimally selects 98 dollar stores in LA County, with most found in lower HPI areas on the map. The tendency of dollar stores to be located in disadvantaged areas positions them particularly well to enhance vaccination rates in vulnerable communities. Although higher HPI communities may have many available pharmacies nearby, adding dollar stores can offer modest gains to these populations too, by allowing individuals across the HPI spectrum to receive a vaccination closer to home.

Beyond simply increasing capacity, incorporating partner stores as vaccine distribution points provides further benefits—especially when the partnership locations are strategically chosen. Compared to a naïve strategy that randomly selects A partner locations, our optimization-based approach yields substantially more predicted vaccinations (Figure 8). For $A = 500$, the set of optimally selected dollar stores generates more than 300,000 additional vaccinations versus a random strategy. In other words, 40% of the predicted gains are attributable to choosing the *best* locations.

Figure 7 **Maps of FRPP pharmacies in Los Angeles County, and optimal locations of partner stores assuming network expansion with A=500 locations statewide**

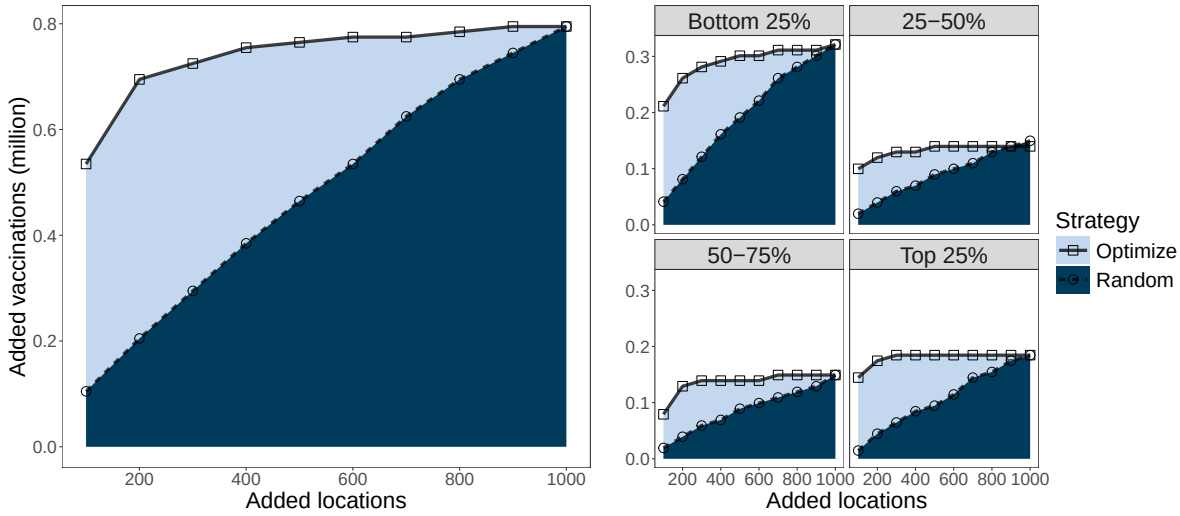


Note. Los Angeles County has 982 pharmacies participating in FRPP, 162 dollar stores, 788 Starbucks coffee houses, and 307 high schools. Of these potential locations, a network expansion policy ($A = 500$ statewide) selects 98 dollar stores, 115 coffee houses, or 154 high schools within the county. Tracts with fewer than 1,500 residents or where >50% of residents live in institutional settings (*e.g.*, dormitories, nursing homes, prisons) are Undefined.

We also observe rapidly diminishing returns as the number of partner locations A increases. With only 100 optimally selected dollar stores, nearly 540,000 added vaccinations statewide are predicted, reflecting 70% of the gains if 500 dollar store locations are employed. With 200 dollar stores, more than 700,000 added vaccinations are predicted—more than three times the vaccination gains achieved by random selection—with over 260,000 vaccinations accruing to the bottom HPI quartile (Figure 8). Efficiently selecting alternative vaccination locations could be particularly valuable during the initial phases of expansion or in resource-constrained settings.

4.3.3. Scenario Analysis. We test the robustness of our numerical results to changes in model parameters, key assumptions, and distance specifications. Additionally, we present an alter-

Figure 8 **Added vaccinations relative to Pharmacy-only, assuming network expansion with A={100, . . . , 1000} dollar stores statewide**



Note. Random strategy is reported as the average added vaccinations over the Pharmacy-only strategy; we computed the average over multiple random selections of 500-dollar stores.

native scheme that modifies the budget constraint, holding constant the total number of vaccination locations in operation to current FRPP levels. We refer to this as *network replacement*.

Parameter sensitivity. We consider, as a reference scenario, the network expansion scheme with $A = 500$ dollar stores statewide, and we vary the choice set size M, per-store capacity K, and travel distance function with threshold $\bar{d}$. All modifications are consistently applied across both the estimation and optimization stages. Results are provided in Appendix C.2.

Our modeling results remain qualitatively similar to our earlier findings. Dollar stores are especially beneficial if residents have few vaccination options (*i.e.*, small M), face limited vaccine supply (*i.e.*, small K), or have greater sensitivity to long travel distances (*i.e.*, indifference to distances below $\bar{d}$). An alternative distance function predicts larger vaccination gains for communities most distant from existing pharmacies. Overall, our findings are consistent across these settings, with dollar stores offering considerable benefits for lower HPI areas or those living in pharmacy deserts.

Alternative partnerships. We examine partnering with high schools or Starbucks coffee houses, instead of the previously considered dollar stores. We assume a similar network expansion with $A = 500$ selected locations statewide. Detailed results are provided in Appendix C.3.

High schools are distributed proportionally to local population levels, as they serve students from diverse socioeconomic backgrounds (Figure A1), making them suitable candidate locations for delivering vaccinations. Expanding the network to include high schools significantly increases predicted vaccinations within the bottom HPI quartile by providing more access points in under-

served communities. Figure 7 illustrates how high schools, similar to dollar stores, can effectively complement existing FRPP pharmacies as evidenced by the selection of schools in lower HPI areas.

Conversely, California has nearly three times as many Starbucks locations as dollar stores or high schools, with these coffee houses typically situated in more affluent, high-traffic areas (Figure A1). Expanding the vaccination network to include 500 Starbucks locations statewide (Figure 7) would primarily enhance vaccination rates in the top HPI quartile (see Figure C2).

Network replacement. In the network replacement scenario, the optimal mix of FRPP pharmacies and dollar stores is selected, effectively reallocating vaccination capacity from pharmacies to partner locations. Specifically, replacing 15% (626 out of 4,035) of existing pharmacies with dollar stores generates an additional 0.70 million predicted vaccinations, a 2.6% increase over the Pharmacy-only strategy. This reallocation involves removing pharmacy-based vaccination sites in neighborhoods with excess capacity and establishing new sites in underserved communities (Figure C4). Predicted vaccinations in the bottom HPI quartile increase by nearly 5%, thereby narrowing the gap between the highest and lowest HPI quartiles. This strategy yields more than 1.1 million additional vaccinations within walking distance (*i.e.*, less than one mile), including 650,000 vaccinations in the most vulnerable regions (Figure C3). This test case is particularly relevant to the design of a partnership program under budgetary or service capacity constraints.

4.4. Policy Insights

Our case study underscores the value of integrating structural estimation and optimization methodologies for policy design in revealing and alleviating disparities in healthcare access. Using only aggregated outcome data, we are able to effectively estimate the link between proximity to a vaccination site and uptake, showing that longer distances indeed result in lower vaccination rates. Importantly, distance does not equally affect all socioeconomic groups, suggesting that an effective vaccination campaign should consider travel distance sensitivity in selecting dispensing locations. More vulnerable neighborhoods should be prioritized and approached differently to achieve similar outcomes. For example, a priority scheduling system that reserves vaccination slots at locations near low HPI communities could help ensure that these underserved population can obtain a vaccination, given their greater sensitivity to travel distance.

We propose a complementary strategy to FRPP: forging partnerships with public sector or private businesses to expand the program and offer vaccinations at optimally selected sites. Beyond merely expanding vaccination capacity, a strategic partnership brings this capacity closer to the very populations least likely to seek vaccination when faced with longer distances. Offering vaccinations at just 100 *optimally chosen* dollar stores across California could theoretically boost vaccinations by more than half a million, or 2% of the predicted total. With 6% of California's population currently living in a pharmacy desert (Wittenauer et al. 2024), even this modest increase

could reduce vaccination disparities. Several studies have found minimal effects of targeted interventions on increasing vaccination coverage. For instance, one randomized controlled trial of vaccine education and messaging found that a host of interventions have no statistically significant effect on the vaccination rates of nursing home staff (Berry et al. 2022). Even interventions celebrated for their success in increasing short-run vaccination rates, such as nudges (Dai et al. 2021) and persuasive advertising (Larsen et al. 2023), typically achieve low single-digit effects.

The feasibility and costs of implementing a non-pharmacy-based vaccination campaign are critical factors that must be carefully considered. Starbucks locations, for example, are visited regularly and may offer more frequent opportunities for vaccination, but their smaller physical space may lead to operational constraints and privacy concerns. High schools, on the other hand, have ample indoor space but school districts are decentralized leading to organizational challenges. Evidence that dollar stores could make a compelling partnership for vaccination administration is demonstrated by their prior use for these services. Beginning in mid-2021, local health departments and jurisdictions across the U.S. collaborated with Dollar General to offer pop-up COVID-19 vaccination clinics (see Appendix Table A3 for a non-exhaustive list). These partnerships were ad-hoc and lacked any coordination at the state or national level. In contrast, our proposed framework proffers a more strategic view of existing vaccination shortages and identifies one efficient mechanism to re-balance supply and demand.

Administering vaccinations in a non-healthcare setting could pose staffing challenges. Some dollar store chains are currently expanding into the pharmacy business themselves (Katje 2021), and they could employ their own pharmacy services team on-site. Another operational strategy could involve using a rotating schedule of visiting nurses or other temporary healthcare staff who travel between sites. Similar programs have been implemented in several counties using the staffing agency SnapNurse.[10] We can approximate the associated staffing costs of a partnership strategy. The average hourly wage of SnapNurse workers in California is $30 (ZipRecruiter 2023). Suppose a vaccination campaign lasts three months, with one or two nurses assigned to each site for one 8-hour day per week. The staffing cost per location would amount to approximately $2,900 (with one nurse) to $5,800 (with two nurses) over the campaign duration. Partnering with 500 optimally selected dollar stores would therefore cost between $1.4 million and $2.9 million, garnering a predicted 0.77 million additional vaccinations, at a cost of about $2 to $4 per added vaccination. These additional vaccinations hold significant social value given the enormous protective effects for both vaccinated individuals and others in their community, particularly those with a low HPI index.

[10] SnapNurse launched in 2017 and deployed more than 10,000 nurses to over 1,000 healthcare facilities in 2020, increasing their revenue by more than $85 million from $3 million in 2019 (Barber 2021).

Despite these operational challenges, our analysis underscores the effectiveness of tailored location strategies that consider the distinct needs and circumstances of different community segments. This approach not only enhances public health outreach but also provides policymakers with evidence-based insights for selecting appropriate partners, both private and public, to systematically tackle structural inequalities inherent in existing health programs like the Federal Retail Pharmacy Program. While our case study is based on cross-sectional data for California due to limited data availability, our framework could be applied at any point in a vaccination campaign, potentially leading to a more proactive policy that mitigates disparities from the beginning.

5. Conclusions

Expanding access to essential services is crucial for eliminating socioeconomic disparities in health, yet few studies have rigorously quantified its impact on health outcomes or explored policy interventions through prescriptive analytics. The inherent complexity of healthcare needs, including patient heterogeneity and limited data availability, pose significant challenges to implementing a one-size-fits-all approach. This study introduces a data-driven framework that integrates a structural econometric model with a facility location model, offering a novel interdisciplinary perspective on the critical issue of health inequality.

We conclude our paper by suggesting several avenues for future research. Investigating disparities and distance elasticities across various domains, both within and beyond healthcare, offers fertile ground for empirical and analytical research in the operations management community. Future work could extend the prescriptive aspect of our framework by incorporating complex assumptions (*e.g.*, heterogeneous facilities), specifying alternative objectives (*e.g.*, cost, equity), exploring other solutions (*e.g.*, facility consolidation, mobile delivery), or considering context-specific constraints, while also developing scalable and efficient solutions for the resulting problem. Additionally, leveraging observational data in diverse formats to improve operational decision-making remains an understudied area. Our study highlights the value of combining empirical models with prescriptive analytics, underscoring their potential to inform and refine policy-making and operational strategies.

References

Aday LA, Andersen R (1974) A framework for the study of access to medical care. *Health Services Research* 9(3):208.

Ahmadi-Javid A, Seyedi P, Syam SS (2017) A Survey of Healthcare Facility Location. *Computers & Operations Research* 79:223–263.

Allon G, Federgruen A, Pierson M (2011) How much is a reduction of your customers' wait worth? an empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management* 13(4):489–507.

Arcury TA, Gesler WM, Preisser JS, Sherman J, Spencer J, Perin J (2005) The effects of geography and spatial behavior on health care utilization among the residents of a rural region. *Health Services Research* 40(1):135–156.

Ayer T, Keskinocak P, Swann J (2014) Research in public health for efficient, effective, and equitable outcomes. *Bridging Data and Decisions*, 216–239 (Institute for Operations Research and the Management Sciences).

Barber P (2021) Mission of mercy: Meet the traveling nurses vaccinating Sonoma County. *The Press Democrat*, URL https://www.pressdemocrat.com/article/news/mission-of-mercy-meet-the-traveling-nurses-vaccinating-sonoma-county/, Accessed May 2024.

Bearak JM, Burke KL, Jones RK (2017) Disparities and change over time in distance women would need to travel to have an abortion in the USA: A spatial analysis. *The Lancet Public Health* 2(11):e493–e500.

Bennouna A, Joseph J, Nze-Ndong D, Perakis G, Singhvi D, Lami OS, Spantidakis Y, Thayaparan L, Tsiourvas A (2023) COVID-19: Prediction, prevalence, and the operations of vaccine allocation. *Manufacturing & Service Operations Management* 25(3):1013–1032.

Berenbrok LA, Tang S, Coley KC, Boccuti C, Jingchuan G, Essien UR, Dickson S, Hernandez I (2021) Access to Potential COVID-19 Vaccine Administration Facilities: A Geographic Information Systems Analysis. URL https://s8637.pcdn.co/wp-content/uploads/2021/02/Access-to-Potential-COVID-19-Vaccine-Administration-Facilities-2-2-2021.pdf, Accessed May 2024.

Berry S, Levinsohn J, Pakes A (1995) Automobile Prices in Market Equilibrium. *Econometrica* 63(4):841–890.

Berry S, Levinsohn J, Pakes A (2004) Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy* 112(1):68–105.

Berry S, Pakes A (2007) The pure characteristics demand model. *International Economic Review* 48(4):1193–1225.

Berry SD, Goldfeld KS, McConeghy K, Gifford D, Davidson HE, Han L, Syme M, Gandhi A, Mitchell SL, Harrison J, et al. (2022) Evaluating the findings of the IMPACT-C randomized clinical trial to improve COVID-19 vaccine coverage in skilled nursing facilities. *JAMA Internal Medicine* 182(3):324–331.

Berry ST (1994) Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* 242–262.

Bertsimas D, Digalakis Jr V, Jacquillat A, Li ML, Previero A (2022) Where to Locate COVID-19 Mass Vaccination Facilities? *Naval Research Logistics* 69(2):179–200.

Bomey N (2021) Dollar General, CDC Exploring Partnership to Speed up COVID-19 vaccine rollout. *USA Today*, URL https://www.usatoday.com/story/money/2021/03/09/dollar-general-cdc-covid-vaccines/6925995002/, Accessed May 2024.

Bronner K, Eliassen SM, King A, Leggett C, Punjasthitkul S, Skinner J (2021) The Dartmouth Atlas of Health Care: 2018 Data Update .

Bront JJM, Méndez-Díaz I, Vulcano G (2009) A column generation algorithm for choice-based network revenue management. *Operations research* 57(3):769–784.

Brown EJ, Polsky D, Barbu CM, Seymour JW, Grande D (2016) Racial disparities in geographic access to primary care in Philadelphia. *Health Affairs* 35(8):1374–1381.

Brownstein J, Cantor JH, Rader B, Simon KI, Whaley CM (2022) If you build it, will they vaccinate? The impact of COVID-19 vaccine sites on vaccination rates and outcomes. Technical report, National Bureau of Economic Research.

California Department of Education (2020) Public Schools and Districts Data Files. URL https://www.cde.ca.gov/ds/si/ds/pubschls.asp, Accessed May 2024.

California Department of Public Health (2021) COVID-19 Vaccine Progress Dashboard Data by ZIP Code. URL https://data.chhs.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data-by-zip-code, Accessed May 2024.

Chevalier JA, Schwartz JL, Su Y, Williams KR (2022) Distributional Impacts of Retail Vaccine Availability. *Journal of Urban Economics* 127:103382.

Church R, ReVelle C (1974) The maximal covering location problem. *Papers of the Regional Science Association*, volume 32, 101–118.

Conlon C, Gortmaker J (2020) Best practices for differentiated products demand estimation with PyBLP. *The RAND Journal of Economics* 51(4):1108–1161.

Current JR, Storbeck JE (1988) Capacitated covering models. *Environment and Planning B: Planning and Design* 15(2):153–163.

Dai H, Saccardo S, Han MA, Roh L, Raja N, Vangala S, Modi H, Pandya S, Sloyan M, Croymans DM (2021) Behavioural nudges increase COVID-19 vaccinations. *Nature* 597(7876):404–409.

Daskin MS, Dean LK (2005) Location of Health Care Facilities. *Operations Research and Health Care* 43–76.

Davis P (2006) Spatial competition in retail markets: movie theaters. *The RAND Journal of Economics* 37(4):964–982.

Denoyel V, Alfandari L, Thiele A (2017) Optimizing healthcare network design under reference pricing and parameter uncertainty. *European Journal of Operational Research* 263(3):996–1006.

Deo S, Sohoni M (2015) Optimal decentralization of early infant diagnosis of HIV in resource-limited settings. *Manufacturing & Service Operations Management* 17(2):191–207.

Deryugina T, Molitor D (2021) The causal effects of place on health and longevity. *Journal of Economic Perspectives* 35(4):147–170.

Dey S, Kurbanzade AK, Gel ES, Mihaljevic J, Mehrotra S (2024) Optimization modeling for pandemic vaccine supply chain management: A review and future research opportunities. *Naval Research Logistics* 1–41.

Dingel JI, Gottlieb JD, Lozinski M, Mourot P (2023) Market size and trade in medical services. Technical report, National Bureau of Economic Research.

Dubé JP, Fox JT, Su CL (2012) Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80(5):2231–2267.

Duch-Brown N, Grzybowski L, Romahn A, Verboven F (2023) Evaluating the impact of online market integration—Evidence from the EU portable PC market. *American Economic Journal: Microeconomics* 15(4):268–305.

Ekici A, Keskinocak P, Swann JL (2014) Modeling influenza pandemic and planning food distribution. *Manufacturing & Service Operations Management* 16(1):11–27.

Enayati S, Özaltın OY (2020) Optimal Influenza Vaccine Distribution with Equity. *European Journal of Operational Research* 283(2):714–725.

Finkelstein A, Gentzkow M, Williams H (2016) Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics* 131(4):1681–1726.

Finkelstein A, Gentzkow M, Williams H (2021) Place-based drivers of mortality: Evidence from migration. *American Economic Review* 111(8):2697–2735.

Fisher ES, Goodman DC, Chandra A (2008) Disparities in health and health care among Medicare beneficiaries: A brief report of the Dartmouth Atlas Project .

Gandhi A (2023) Picking your patients: Selective admissions in the nursing home industry. *Available at SSRN* 3613950.

Ganju KK, Atasoy H, McCullough J, Greenwood B (2020) The role of decision support systems in attenuating racial biases in healthcare delivery. *Management Science* 66(11):5171–5181.

Gowrisankaran G, Nevo A, Town R (2015) Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review* 105(1):172–203.

Guadamuz JS, Wilder JR, Mouslim MC, Zenk SN, Alexander GC, Qato DM (2021) Fewer Pharmacies in Black and Hispanic/Latino Neighborhoods Compared with White or Diverse Neighborhoods, 2007–15. *Health Affairs* 40(5):802–811.

Guajardo JA, Cohen MA, Netessine S (2016) Service competition and product quality in the US automobile industry. *Management Science* 62(7):1860–1877.

Hackmann MB (2019) Incentivizing better quality of care: The role of Medicaid and competition in the nursing home industry. *American Economic Review* 109(5):1684–1716.

Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50(4):1029–1054.

Heier Stamm JL, Serban N, Swann J, Wortley P (2017) Quantifying and explaining accessibility with application to the 2009 H1N1 vaccination campaign. *Health Care Management Science* 20(1):76–93.

Ho K (2006) The welfare effects of restricted hospital choice in the US medical care market. *Journal of Applied Econometrics* 21(7):1039–1079.

Ippolito B, Levy JF, Anderson GF (2020) Abandoning List Prices In Medicaid Drug Reimbursement Did Not Affect Spending. *Health Affairs* 39(7):1202–1209.

IQVIA Institute for Human Data Science (2023) Trends in Vaccine Administration in the United States. URL https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/trends-in-vaccine-administration-in-the-united-states, Accessed May 2024.

Jia H, Ordóñez F, Dessouky M (2007) A modeling framework for facility location of medical services for large-scale emergencies. *IIE Transactions* 39(1):41–55.

Jónasson JO, Deo S, Gallien J (2017) Improving HIV early infant diagnosis supply chains in sub-Saharan Africa: Models and application to Mozambique. *Operations Research* 65(6):1479–1493.

Katje C (2021) Dollar General Makes Moves To Be The Next Corner Pharmacy. *Business Insider*, URL https://markets.businessinsider.com/news/stocks/dollar-general-makes-moves-to-be-the-next-corner-pharmacy-1030586485, Accessed May 2024.

Kelly C, Hulme C, Farragher T, Clarke G (2016) Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. *BMJ Open* 6(11):e013059.

Krohn R, Müller S, Haase K (2021) Preventive healthcare facility location planning with quality-conscious clients. *OR Spectrum* 43(1):59–87.

Larsen BJ, Ryan TJ, Greene S, Hetherington MJ, Maxwell R, Tadelis S (2023) Counter-stereotypical messaging and partisan cues: Moving the needle on vaccines in a polarized United States. *Science Advances* 9(29):eadg9434.

Lee EK, Pietz F, Benecke B, Mason J, Burel G (2013) Advancing public health and medical preparedness with operations research. *Interfaces* 43(1):79–98.

Maizlish N, Delaney T, Dowling H, Chapman DA, Sabo R, Woolf S, Orndahl C, Hill L, Snellings L (2019) California Healthy Places Index: Frames Matter. *Public Health Reports* 134(4):354–362.

McCormick GP (1976) Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming* 10(1):147–175.

McCoy JH, Johnson ME (2014) Clinic capacity management: Planning treatment programs that incorporate adherence. *Production and Operations Management* 23(1):1–18.

McFadden D (1974) Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* .

Méndez-Díaz I, Miranda-Bront JJ, Vulcano G, Zabala P (2014) A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics* 164:246–263.

Myers C, Jones R, Upadhyay U (2019) Predicted changes in abortion access and incidence in a post-Roe world. *Contraception* 100(5):367–373.

Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2):307–342.

Nobles M, Serban N, Swann J (2014) Spatial accessibility of pediatric primary healthcare: measurement and inference. *Annals of Applied Statistics* 8(4):1922–46.

Penchansky R, Thomas JW (1981) The concept of access: definition and relationship to consumer satisfaction. *Medical care* 127–140.

Petrin A (2002) Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy* 110(4):705–729.

Public Health Alliance of Southern California (2022) Healthy Places Index 2022. URL https://www.healthyplacesindex.org/, Accessed May 2024.

Rader B, Astley CM, Sewalk K, Delamater PL, Cordiano K, Wronski L, Rivera JM, Hallberg K, Pera MF, Cantor J, et al. (2022) Spatial modeling of vaccine deserts as barriers to controlling SARS-CoV-2. *Communications Medicine* 2(1):141.

Rastegar M, Tavana M, Meraj A, Mina H (2021) An Inventory-location Optimization Model for Equitable Influenza Vaccine Distribution in Developing Countries During the COVID-19 Pandemic. *Vaccine* 39(3):495–504.

Reynaerts J, Varadha R, Nash JC (2012) Enhancing the convergence properties of the BLP (1995) contraction mapping. *VIVES Discussion Paper* .

Roberts B (2021) Dollar General Stores Could Aid in Administering Vaccines. *Spectrum News 1*, URL https://spectrumnews1.com/ky/louisville/news/2021/03/16/dollar-general-could-assist-in-vaccine-distribution, Accessed May 2024.

Salako A, Ullrich F, Mueller KJ (2018) Update: Independently Owned Pharmacy Closures in Rural America, 2003-2018. *Rural Policy Brief* 2018(2):1–6.

Samorani M, Harris SL, Blount LG, Lu H, Santoro MA (2021) Overbooked and overlooked: Machine learning and racial bias in medical appointment scheduling. *Manufacturing & Service Operations Management* (in press).

ScrapeHero (2021) Dollar Store Locations in the USA. URL https://www.scrapehero.com/store/product/dollar-store-locations-in-the-usa/, Accessed May 2024.

Sen A, Atamtürk A, Kaminsky P (2018) A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model. *Operations Research* 66(4):994–1003.

Shepard M (2022) Hospital network competition and adverse selection: Evidence from the Massachusetts health insurance exchange. *American Economic Review* 112(2):578–615.

Syed ST, Gerber BS, Sharp LK (2013) Traveling towards disease: transportation barriers to health care access. *Journal of Community Health* 38:976–993.

Tay A (2003) Assessing competition in hospital care markets: the importance of accounting for quality differentiation. *RAND Journal of Economics* 786–814.

US Census Bureau (2019) Planning Database with 2010 Census and 2014 – 2018 American Community Survey Data. URL https://www.census.gov/topics/research/guidance/planning-databases.2021.html.

US Centers for Disease Control and Prevention (2023) Federal Retail Pharmacy Program. URL https://archive.cdc.gov/#/details?url=https://www.cdc.gov/vaccines/covid-19/retail-pharmacy-program/index.html, Accessed May 2024.

Wang G, Zheng R, Dai T (2022) Does transportation mean transplantation? Impact of new airline routes on sharing of cadaveric kidneys. *Management Science* 68(5):3660–3679.

Wittenauer R, Shah PD, Bacci JL, Stergachis A (2024) Locations and characteristics of pharmacy deserts in the United States: a geospatial study. *Health Affairs Scholar* 2(4):qxae035.

Wolfrath J, Ryan B, Nehring P (2018) Dollar Stores in Small Communities. URL https://fyi.extension.wisc.edu/downtowneconomics/files/2018/11/DE1218a.pdf, Accessed May 2024.

Wu TH (1997) A note on a global approach for general 0–1 fractional programming. *European Journal of Operational Research* 101(1):220–223.

Xu Y, Armony M, Ghose A (2021) The interplay between online reviews and physician demand: An empirical investigation. *Management Science* 67(12):7344–7361.

Zhang Y, Berman O, Verter V (2012) The impact of client choice on preventive healthcare facility network design. *OR Spectrum* 34:349–370.

Zhong H, Wang G, Dai T (2023) Wheels on the bus: Impact of vaccine rollouts on demand for public transportation. *Available at SSRN 3874150* .

ZipRecruiter (2023) What Is the Average Snap Nurse Salary by State. URL https://www.ziprecruiter.com/Salaries/What-Is-the-Average-Snap-Nurse-Salary-by-State, Accessed May 2024.

Appendix A: Case Study Data

Table A1 California summary statistics

	Mean	SD
Number of retail pharmacy vaccination sites	4,035	-
Number of dollar stores	1,016	-
State population	37,336,716	-
Number of zip-codes	1,751	-
Population (per zip-code)	22,265	22,608
Population density (residents per sq mi)	3,425	5,574
Race/ethnicity		
% White	0.51	0.28
% Black	0.04	0.07
% Asian	0.09	0.13
% Hispanic	0.30	0.25
% Other	0.05	0.07
Health insurance		
% Employer	0.41	0.17
% Medicare	0.15	0.11
% Medicaid	0.20	0.15
% Other	0.15	0.09
College graduation rate	0.31	0.21
Unemployment rate	0.07	0.06
Poverty level	0.10	0.11
Median household income ($000s)	70.2	40.1
Median home value ($000s)	501.1	398.3
HPI (continuous value in $[0, 1]$)	0.47	0.29
Distance to nearest vaccination site (km)		
HPI quartile 4 (most healthy)	3.8	5.7
HPI quartile 3	8.1	13.3
HPI quartile 2	12.0	13.5
HPI quartile 1 (least healthy)	15.3	16.2

Note. Zip code-level demographics include race/ethnicity, health insurance, college graduation rate, unemployment rate, poverty rate, median household income, median home value, population density, and population.

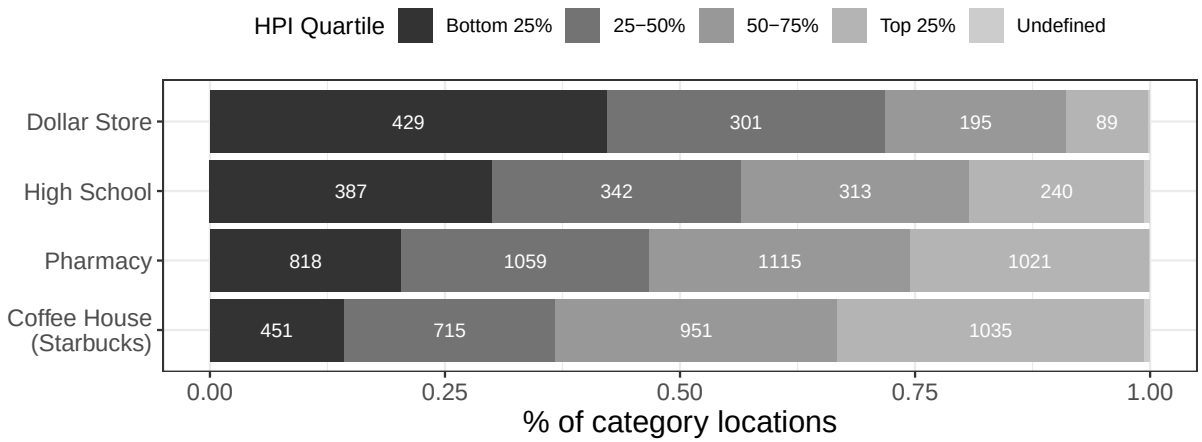
Table A2 Partnership categories in California

Category	Brand	Locations	Source
Pharmacy	CVS	1,092	vaccines.gov
	Walgreens	587	
	Rite Aid	538	
	Walmart	296	
	Safeway	161	
	Costco	124	
	Vons	109	
	Sav-On	91	
	Ralphs	75	
	Other	962	
Coffee House	Starbucks	3,172	Safegraph
Dollar Store	Dollar Tree	631	ScrapeHero
	Dollar General	238	
	Family Dollar	147	
High School	Public	1,225	CA Dept of Education

Table A3 Past COVID-19 vaccination or testing campaigns at Dollar General stores

Month	Locations	Services offered	Source
Aug 2021	9 counties in Michigan	COVID-19 vaccination by SnapNurse	Hall (2021)
Aug 2021	3 stores in Virginia	COVID-19 testing	Virginia Department of Health (2021)
Aug 2021	Sumter County, South Carolina	COVID-19 vaccination	Staff Reports (2021b)
Aug 2021	Jackson County, Indiana	COVID-19 vaccination	Banks (2021)
Sep 2021	6 counties in Kansas	COVID-19 vaccination with $100 incentive	Kansas Office of the Governor (2021)
Oct 2021	4 stores in Toledo, Ohio	COVID-19 vaccination by RiteAid pharmacists	WTVG News (2021)
Feb 2022	Shenandoah County, Virginia	COVID-19 vaccination	WHSV Newsroom (2022)
Feb 2022	Fairfax County, Virginia	COVID-19 vaccination	Staff Reports (2021a)
Feb 2022	3 counties in Virginia	COVID-19 vaccination	WFXR Fox (2022)
Mar 2022	Roanoke County, Virginia	COVID-19 vaccination	Cover Virginia (2022)
Jun 2022	San Joaquin County, California	COVID-19 vaccination	Health Plan of San Joaquin (2022)

Figure A1 Retail store types in California, by Healthy Places Index (HPI)



Note. Census tracts with fewer than 1,500 residents or where >50% of residents live in institutional settings (e.g., dormitories, nursing homes, prisons) are Undefined.

Appendix B: Implementation

B.1. Optimizing the GMM objective function

Optimizing the GMM objective given in equation (6) is particularly challenging because the distance parameters $\boldsymbol{\alpha}$ affect both the individuals' appointment location choice and their vaccination preferences. However, standard solvers can easily compute optimal demand parameters given a fixed appointment distribution $f(\ell_i \mid s_i)$. We propose an iterative approach for optimizing the GMM objective that consists of iterating between computing the optimal demand parameter values under a fixed appointment distribution and updating the appointment distribution based on the current parameter values. By iterating in this fashion, we eventually converge to demand parameter estimates—including the distance sensitivities $\boldsymbol{\alpha}$—that (locally) minimize the GMM objective given the appointment distribution that $\boldsymbol{\alpha}$ implies.

We denote $\tilde{f}^{(k)}(\ell_i \mid s_i)$ as the distribution of appointment locations in step k of the algorithm. In the initial step, we assume that locations have no capacity constraints, and that all individuals prefer their closest location.[11] Then, assuming this appointment distribution is fixed, we optimize the GMM objective function to obtain a new set of optimal parameters $\hat{\boldsymbol{\theta}}^{(k)}$, including new values for the distance sensitivities $\hat{\boldsymbol{\alpha}}^{(k)}$.

Given a fixed appointment distribution $\tilde{f}^{(k)}(\ell_i \mid s_i)$, we obtain $\hat{\boldsymbol{\theta}}^{(k)}$ using the fixed-point approach outlined by Berry et al. (1995) that optimizes the following GMM objective:

$$\hat{\boldsymbol{\theta}}^{(k)} = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{|\mathcal{A}|} \sum_a \tilde{Z}_a^{(k)}(\boldsymbol{\theta}) \tilde{\xi}_a^{(k)}(\boldsymbol{\theta}) \right)' \Phi \left(\frac{1}{|\mathcal{A}|} \sum_a \tilde{Z}_a^{(k)}(\boldsymbol{\theta}) \tilde{\xi}_a^{(k)}(\boldsymbol{\theta}) \right). \tag{B.1}$$

Note that the instruments $\tilde{Z}_a^{(k)}(\boldsymbol{\theta})$ are computed assuming the fixed appointment distribution $\tilde{f}^{(k)}(\ell_i \mid s_i)$. The term $\tilde{\xi}_a^{(k)}(\boldsymbol{\theta})$ is obtained by matching the implied aggregate demand $\tilde{\rho}_a^{(k)}(\boldsymbol{\theta})$ to the observed rates ρ_a^{obs}. Formally, the implied aggregate demand in area a is:

$$\tilde{\rho}_a^{(k)}(\boldsymbol{\theta}) = \sum_{s_i \in a} \Big[\sum_{\ell_i \in L_i} \frac{\exp\left(d(s_i, \ell_i)\alpha_a + X_a \boldsymbol{\beta} + \xi_a\right)}{1 + \exp\left(d(s_i, \ell_i)\alpha_a + X_a \boldsymbol{\beta} + \xi_a\right)} \tilde{f}^{(k)}(\ell_i \mid s_i) \Big] f(s_i \mid a). \tag{B.2}$$

We implement the BLP approach using the popular PyBLP package (Conlon and Gortmaker 2020). The output is $\hat{\boldsymbol{\theta}}^{(k)}$, which includes the distance sensitivities $\hat{\boldsymbol{\alpha}}^{(k)}$. These values are then used to update the appointment distribution in the next iteration according to equation (7) as follows:

$$\tilde{f}^{(k+1)}(\ell_i \mid s_i, L_i) = \frac{\exp\left(d(s_i, \ell_i)\hat{\alpha}_{a_i}^{(k)}\right)}{\sum_{\ell \in L_i} \exp\left(d(s_i, \ell)\hat{\alpha}_{a_i}^{(k)}\right)}, \tag{B.3}$$

where the distribution of L_i is determined by the sequential assignment process described in Section 3.1.3.

To determine convergence, we compute for each area the Wasserstein distance between the appointment location distributions in the current step and the previous step. We terminate the algorithm when the Wasserstein distance falls below a certain tolerance, that is, when $\tilde{f}^{(k+1)}(\ell_i \mid s_i) \approx \tilde{f}^{(k)}(\ell_i \mid s_i)$. At this point, our approach of iteratively updating appointment distributions and preferences converges to a $\hat{\boldsymbol{\theta}}$ that approximately satisfies the local optimality conditions for equation (6).[12]

[11] This is equivalent to having extreme distance sensitivities, *i.e.*, $\boldsymbol{\alpha} \to -\infty$.

[12] Note that all the fixed-point methods used are evaluated using multiple starting points.

B.2. Demand aggregation

For computational tractability, we formulate the case study optimization problem at the Census tract level. Solving this problem requires having tract-level utility values $\bar{u}(t,\ell)$, which determine the tract-level demand shares $\rho_{t\ell}$. However, these utility terms cannot be computed using the demand model parameters as these were estimated at the Census block level. Instead, we define $\hat{u}(t,\ell)$ the utility-equivalent individuals in Census tract t obtain from receiving service at location $\ell \in \mathcal{L}$, and define the tract-level demand share as:

$$\rho_{t\ell} := \frac{\frac{\exp(2\hat{u}(t,\ell))}{1+\exp(\hat{u}(t,\ell))}}{\sum_{k\in L_t}\exp(\hat{u}(t,k))}, \ \forall \ell \in L_t, \tag{B.4}$$

and zero otherwise, where L_t is the tract-specific choice set. We solve for these utility-equivalent values by solving a system of equations based on the estimated block-level demand shares $\rho_{b\ell}$ (*i.e.*, as a function of the block-level estimated utilities $\bar{u}(b,\ell)$) and the tract-level demand shares $\rho_{t\ell}(\{\hat{u}(t,\ell)\}_{t,\ell})$ in equation (B.4). The system of equations is derived from writing the tract-level demand shares as an aggregation of the block-level demand shares.

Let m_b and m_t be the eligible population of a Census block b and a Census tract t, respectively. Since blocks are perfectly nested within tracts (*i.e.*, each block belongs to one unique tract), we have $m_t = \sum_{b\in t} m_b$. The aggregate share of the population in tract t that obtains service at a specific location ℓ is given by the population-weighted share of all blocks b within the tract t:

$$\rho_{t\ell} = \sum_{b\in t}\frac{m_b}{m_t}\rho_{b\ell}, \quad \text{if } \ell \in L_t \tag{B.5}$$

The value of $\rho_{b\ell}$ is obtained following equation (9) with subareas defined as blocks, namely,

$$\rho_{b\ell} := \frac{\frac{\exp(2\bar{u}(b,\ell))}{1+\exp(\bar{u}(b,\ell))}}{\sum_{\ell\in L_b}\exp(\bar{u}(b,\ell))}, \quad \text{if } \ell \in L_b, \tag{B.6}$$

and zero otherwise, where L_b is the block-specific choice set. Then, we solve for $\{\hat{u}(t,\ell)\}_{t,\ell}$ by plugging in the definition of $\rho_{t\ell}$ in equation (B.4) into equation (B.5).

This system of non-linear equations is hard to solve, instead we provide a simple numerical approximation: we solve for $\hat{u}(t,\ell)$ assuming ℓ is the only location available to tract t and to all blocks $b \in t$ with $\ell \in L_b$. We then iteratively solve for each ℓ in L_t. This reduces the above system of equations to

$$\rho_{t\ell} := \frac{\exp(\hat{u}(t,\ell))}{1+\exp(\hat{u}(t,\ell))}, \text{ if } \ell \in L_t \text{ and zero otherwise;} \tag{B.7a}$$

$$\rho_{t\ell} = \sum_{b\in t}\frac{m_b}{m_t}\rho_{b\ell} = \sum_{b\in t}\frac{m_b}{m_t}\frac{\exp(\bar{u}(b,\ell))}{1+\exp(\bar{u}(b,\ell))}, \quad \text{if } \ell \in L_t \text{ and zero otherwise.} \tag{B.7b}$$

Let us denote the solution to equations (B.7a)-(B.7b) as $\{\tilde{u}(t,\ell)\}_{t,\ell}$, and the corresponding induced tract-level demand shares as $\{\tilde{\rho}_{t\ell}\}_{t,\ell}$ (following the definition in equation B.4).

While we cannot compare $\{\tilde{u}(t,\ell)\}_{t,\ell}$ with $\{\hat{u}(t,\ell)\}_{t,\ell}$ directly to assess the validity of our approximation (as the latter may not exist), we could compare (i) the induced demand shares, *i.e.*, the demand share implied by our approximation, $\tilde{\rho}_{t\ell}$ and (ii) the tract-level demand share implied by aggregating the actual block-level demand share, $\rho_{t\ell}$ from equations (B.5)-(B.6), both assuming $|L_t| = 5$. We report the relative gap $\Delta := |\tilde{\rho}_{t\ell} - \rho_{t\ell}|/\rho_{t\ell}$ across all tracts and candidate locations analyzed in our case study. The mean of Δ is 0.1%, with a maximum of 2%. This demonstrates the validity of our approximation. Notice that this is only done for the optimization stage to identify the optimal set of locations, then in the evaluation, we employ block-level demand estimates to determine the predicted vaccinations, requiring no approximation.

B.3. Optimality Gap of the MIP Relaxation

Under the equivalent formulation ($\mathsf{P}'_{\mathsf{loc}}$) of our original optimization problem, let $g(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w})$ denote the objective function where $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}$ are the decision variables, and $\mathcal{R}$ is the feasible region. Let the optimal solution to ($\mathsf{P}'_{\mathsf{loc}}$) be $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*)$, such that

$$g(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}) \leq g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*), \quad \forall\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}) \in \mathcal{R}.$$

Similarly, denote the optimal solution of the relaxation ($\tilde{\mathsf{P}}_{\mathsf{loc}}$) as $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}})$, with optimal objective value $\tilde{g}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}})$ and feasible region $\tilde{\mathcal{R}}$. Notice that for any feasible solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}) \in \mathcal{R}$ of the original problem ($\mathsf{P}'_{\mathsf{loc}}$), we can always construct a feasible solution for the relaxation ($\tilde{\mathsf{P}}_{\mathsf{loc}}$) by letting $\boldsymbol{z} = \boldsymbol{v}\boldsymbol{w}$, *i.e.*, $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v}\boldsymbol{w}) \in \tilde{\mathcal{R}}$. Then, we have

$$\begin{aligned} g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*) &= \tilde{g}(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*, \boldsymbol{v}^*\boldsymbol{w}^*) \\ &\leq \tilde{g}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}}), \end{aligned} \tag{B.8}$$

where the first equality follows from the following observation:

$$g(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}) = \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{L}} m_s \frac{\exp(2\bar{u}(s, \ell))}{1 + \exp(\bar{u}(s, \ell))} w_{s\ell} y_{s\ell} = \tilde{g}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v}\boldsymbol{w}),$$

and the second inequality follows from the definition of optimality for ($\tilde{\mathsf{P}}_{\mathsf{loc}}$).

To derive a lower bound on the optimal objective value of the original problem ($\mathsf{P}'_{\mathsf{loc}}$), we substitute the optimal set of locations $\tilde{\boldsymbol{x}}$ obtained from ($\tilde{\mathsf{P}}_{\mathsf{loc}}$) back into ($\mathsf{P}'_{\mathsf{loc}}$). In other words, we fix the set of locations $\boldsymbol{x} = \tilde{\boldsymbol{x}}$ in ($\mathsf{P}'_{\mathsf{loc}}$) and solve for the optimal values of other decision variables, and we denote the resulting optimal solution as $(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}')$, where

$$g(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}') \leq g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*). \tag{B.9}$$

Combining equations (B.8) and equation (B.9), we conclude the following relation holds:

$$g(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}') \leq g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*) \leq \tilde{g}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}}).$$

Then we can obtain a bound on the optimality gap of our relaxation as follows:

$$\begin{aligned} G_{opt} &:= \frac{g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*) - g(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}')}{g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*)} \\ &\leq \frac{\tilde{g}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}}) - g(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}')}{g(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{v}^*, \boldsymbol{w}^*)} \\ &\leq \frac{\tilde{g}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}}) - g(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}')}{g(\tilde{\boldsymbol{x}}, \boldsymbol{y}', \boldsymbol{b}', \boldsymbol{v}', \boldsymbol{w}')} := \overline{G}_{opt} \end{aligned}$$

In our case study's main results from Section 4.3.2, we obtain an optimality gap upper bound of 0.58%. Across all instances analyzed in Section C, the mean of $\overline{G}_{opt}$ is 0.58% (median = 0.60%) with a maximum of 0.81%. These results validate the strength of our relaxation. It's important to note that the optimality gap discussed here pertains to the relaxation of our optimization problem, not to the potential benefits of the selected locations.

Appendix C: Scenario Analysis

We test the robustness of both our estimation and optimization results to changes in model parameters and assumptions. In Section C.1, we provide the demand estimates under varying parameter values, distance functions, and the appointment assignment process. We then present our optimization model's results for select parameter values and alternative distance functions in Section C.2. We evaluate the potential gains of broadening FRPP by partnering with high schools or Starbucks coffee houses instead of dollar stores in Section C.3. Finally, in Section C.4, we examine an alternative *network replacement* strategy, where a combination of FRPP and dollar stores are jointly selected while holding constant the total number of vaccination distribution points.

C.1. Demand Estimates

In general, our estimated sensitivities to distance obtained from the structural demand estimation step are robust to assumptions in parameter values and our assumed appointment assignment process. Tables C4 and C5 give the demand estimation results for all variations considered.

Choice set. In our primary analysis, we assume that if all locations in an individual's choice set are at capacity, then individuals are assigned to their M^{th} location. This approximation is necessary because the observed vaccination rates in some zip codes are higher than the aggregate vaccination capacity available for those residents. In other words, lower values of M may not guarantee that all vaccinated individuals can obtain a vaccination within one of their M closest locations, which would result in non-convergence of the BLP methodology. If $M = 5$, then approximately 3.6% of individuals encounter no remaining capacity at all 5 locations, and are thus assigned to their 5^{th} closest location.

As a robustness check, we relax the assumption that $M = 5$ and we estimate demand under the MNL choice model with $M \in \{10,\ 300,\ 1{,}000\}$. The first four columns of Table C4 confirm that the demand estimates are not sensitive to the number of locations in the choice set. For instance, if $M = 10$, the distance sensitivity coefficient for HPI quartile 1 slightly attenuates to -0.158 ($p < 0.001$) compared to -0.161 ($p < 0.001$) when $M = 5$. Even when $M = 1{,}000$, which virtually guarantees that all individuals have some location in their choice set with remaining capacity, the sensitivity to distance for HPI quartile 1 only modestly changes to -0.180 ($p < 0.001$). Note, that the magnitude is greater because, under this assumption, some individuals will end up choosing a very distant vaccination site and the lower distance coefficient will nudge them to choose closer sites (under the MNL choice model), so as to match predicted demand with observed vaccination rates.

Capacity. Next, we consider varying the vaccination site capacity $K \in \{8{,}000,\ 12{,}000\}$, from our baseline assumption of $K = 10{,}000$. Similar to M, these adjustments affect the probability that an individual chooses a nearby vaccination site. For instance, with a reduced capacity per-site of $K = 8{,}000$, individuals may need to consider farther sites if their closest options reach capacity. Indeed, under limited capacity, demand estimates indicate that individuals are less sensitive to distance, suggesting a greater willingness to travel farther for vaccination (Table C5). The estimated coefficient on distance for HPI quartile 1 is -0.150 ($p < 0.001$) compared to -0.161 ($p < 0.001$) when $K = 10{,}000$. Conversely, with an increased capacity of $K = 12{,}000$,

individuals are more likely to secure a spot at their nearest site, aligning with their first preference under the MNL model. Under this scenario, the estimated coefficient on distance for HPI quartile 1 is -0.170 ($p < 0.001$), implying a slightly steeper disutility to traveling farther for vaccination.

Distance forms. We consider an alternative specification for the distance function $d(s, \ell)$ in which individuals' disutility of traveling is assumed to be zero for distances below a threshold $\bar{d}$. In other words, individuals are insensitive to variations in short distances. The functional form is:

$$d(s, \ell) = \log\left((d_{s\ell} - \bar{d})^+ + 1\right) \tag{C.10}$$

where $d_{s\ell}$ is the geodesic distance between s and ℓ, and $x^+ := \max\{x, 0\}$.

We consider two distance thresholds, $\bar{d} \in \{0.5, 1.0\}$ miles, and recalculate demand estimates. In both cases, the distance sensitivities increase slightly, to offset the flat utility values below $\bar{d}$ (Table C5). For instance, if $\bar{d} = 1.0$, the coefficient on distance for HPI quartile 1 is -0.174 ($p < 0.001$), as individuals become more sensitive to longer distances above $\bar{d}$. However, the signs and relative magnitudes of distance sensitivities remain unchanged.

Closest location. Finally, we test an alternative appointment process in lieu of the MNL-based choice model. Here, we assign individuals to their closest location with available capacity, out of the M nearest locations, and we vary $M \in \{300,\ 1,000\}$. The last two columns of Table C4 give the revised demand estimates. Distance sensitivity coefficients attenuate slightly as M increases, as expected. For instance, under this assignment regime with $M = 300$, the coefficient on distance for HPI quartile 1 is -0.130 ($p < 0.001$). Our general findings hold, confirming that distance to a pharmacy negatively correlates with vaccination rates and that individuals from lower HPI quartiles exhibit greater sensitivity to travel distances.

Table C4 Demand estimates under varying sizes of choice sets M and appointment assignment process

	MNL				Closest Location	
Independent Variable	$M=5$ (†)	$M=10$	$M=300$	$M=1000$	$M=300$	$M=1000$
HPI Quartile 4 (most healthy)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HPI Quartile 3	-0.304***	-0.306***	-0.303***	-0.306***	-0.299***	-0.299***
	(0.084)	(0.084)	(0.086)	(0.083)	(0.087)	(0.087)
HPI Quartile 2	-0.451***	-0.452***	-0.448***	-0.447***	-0.450***	-0.450***
	(0.112)	(0.112)	(0.113)	(0.111)	(0.114)	(0.114)
HPI Quartile 1 (least healthy)	-0.605***	-0.607***	-0.600***	-0.603***	-0.601***	-0.601***
	(0.149)	(0.150)	(0.150)	(0.148)	(0.152)	(0.152)
Log-distance × HPI Quartile 4	-0.063	-0.062	-0.051	-0.086	-0.027	-0.027
	(0.070)	(0.070)	(0.072)	(0.071)	(0.068)	(0.068)
Log-distance × HPI Quartile 3	-0.142***	-0.138***	-0.134***	-0.160***	-0.108**	-0.108**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.049)	(0.049)
Log-distance × HPI Quartile 2	-0.124***	-0.122***	-0.119***	-0.145***	-0.092**	-0.092**
	(0.043)	(0.043)	(0.043)	(0.045)	(0.042)	(0.042)
Log-distance × HPI Quartile 1	-0.161***	-0.158***	-0.157***	-0.180***	-0.130***	-0.130***
	(0.043)	(0.043)	(0.043)	(0.044)	(0.043)	(0.043)
Race White	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Race Black	-0.052	-0.049	-0.063	-0.104	0.007	0.007
	(0.342)	(0.343)	(0.344)	(0.342)	(0.345)	(0.345)
Race Asian	2.001***	2.004***	2.027***	1.968***	2.041***	2.041***
	(0.179)	(0.179)	(0.180)	(0.178)	(0.183)	(0.183)
Race Hispanic	0.987***	0.989***	0.981***	0.971***	1.016***	1.016***
	(0.167)	(0.167)	(0.167)	(0.167)	(0.166)	(0.166)
Race Other	7.077**	7.096**	7.258**	7.028**	7.217**	7.217**
	(3.049)	(3.051)	(3.088)	(3.052)	(3.054)	(3.054)
Health Insurance: Employer	-1.202*	-1.201*	-1.195*	-1.209*	-1.190*	-1.190*
	(0.696)	(0.697)	(0.696)	(0.695)	(0.700)	(0.700)
Health Insurance: Medicare	0.197	0.197	0.197	0.204	0.181	0.181
	(0.739)	(0.739)	(0.739)	(0.737)	(0.743)	(0.743)
Health Insurance: Medicaid	-0.973	-0.975	-0.967	-0.972	-0.977	-0.977
	(0.729)	(0.730)	(0.730)	(0.728)	(0.733)	(0.733)
Health Insurance: Other	-2.319***	-2.319***	-2.321***	-2.316***	-2.323***	-2.323***
	(0.760)	(0.760)	(0.761)	(0.758)	(0.765)	(0.765)
College Graduation Rate	1.727***	1.726***	1.737***	1.729***	1.732***	1.732***
	(0.282)	(0.282)	(0.282)	(0.281)	(0.283)	(0.283)
Unemployment Rate	-1.777***	-1.779***	-1.783***	-1.748**	-1.818***	-1.818***
	(0.686)	(0.686)	(0.686)	(0.684)	(0.689)	(0.689)
Poverty Level	-0.011	-0.011	-0.014	-0.017	-0.013	-0.013
	(0.472)	(0.472)	(0.472)	(0.473)	(0.474)	(0.474)
Log(Median Household Income)	-0.110*	-0.110*	-0.111*	-0.108*	-0.114*	-0.114*
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
Log(Median Home Value)	0.114**	0.115**	0.114**	0.113**	0.117**	0.117**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)
Log(Population Density)	-0.078***	-0.077***	-0.076***	-0.085***	-0.066***	-0.066***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)

Note. (†) denotes our main specification. M refers to the size of an individual's choice set. MNL estimates are under the multinomial-logit choice model. Closest Location refers to an alternative assignment process where individuals obtain a vaccination at their nearest pharmacy within M. Standard errors are in parentheses. Significance levels: *p<0.05; **p<0.01; ***p<0.001

Table C5 Demand estimates under varying per-store vaccination capacity and distance forms

		Capacity		Distance form	
Independent Variable	$K = 10,000$ (†)	$K = 8,000$	$K = 12,000$	$\bar{d} = 0.5$	$\bar{d} = 1$
HPI Quartile 4 (most healthy)	Ref.	Ref.	Ref.	Ref.	Ref.
HPI Quartile 3	-0.304***	-0.299***	-0.306***	-0.262**	-0.280***
	(0.084)	(0.086)	(0.084)	(0.108)	(0.097)
HPI Quartile 2	-0.451***	-0.448***	-0.451***	-0.416***	-0.424***
	(0.112)	(0.113)	(0.111)	(0.132)	(0.123)
HPI Quartile 1 (least healthy)	-0.605***	-0.597***	-0.607***	-0.553***	-0.569***
	(0.149)	(0.151)	(0.149)	(0.171)	(0.162)
Log-distance × HPI Quartile 4	-0.063	-0.045	-0.076	-0.087	-0.074
	(0.070)	(0.070)	(0.070)	(0.091)	(0.086)
Log-distance × HPI Quartile 3	-0.142***	-0.128**	-0.152***	-0.170***	-0.152***
	(0.050)	(0.050)	(0.050)	(0.060)	(0.056)
Log-distance × HPI Quartile 2	-0.124***	-0.111**	-0.135***	-0.149***	-0.136***
	(0.043)	(0.043)	(0.044)	(0.052)	(0.048)
Log-distance × HPI Quartile 1	-0.161***	-0.150***	-0.170***	-0.190***	-0.174***
	(0.043)	(0.044)	(0.044)	(0.051)	(0.048)
Race White	Ref.	Ref.	Ref.	Ref.	Ref.
Race Black	-0.052	-0.027	-0.074	-0.053	-0.065
	(0.342)	(0.343)	(0.342)	(0.343)	(0.342)
Race Asian	2.001***	2.021***	1.983***	2.020***	2.027***
	(0.179)	(0.180)	(0.178)	(0.179)	(0.179)
Race Hispanic	0.987***	0.998***	0.980***	0.983***	0.982***
	(0.167)	(0.166)	(0.167)	(0.166)	(0.167)
Race Other	7.077**	7.136**	7.036**	7.385**	7.415**
	(3.049)	(3.045)	(3.049)	(3.104)	(3.118)
Health Insurance: Employer	-1.202*	-1.195*	-1.207*	-1.258*	-1.262*
	(0.696)	(0.697)	(0.695)	(0.694)	(0.696)
Health Insurance: Medicare	0.197	0.189	0.200	0.151	0.144
	(0.739)	(0.740)	(0.738)	(0.734)	(0.736)
Health Insurance: Medicaid	-0.973	-0.976	-0.973	-1.026	-1.032
	(0.729)	(0.730)	(0.728)	(0.725)	(0.726)
Health Insurance: Other	-2.319***	-2.322***	-2.318***	-2.372***	-2.374***
	(0.760)	(0.762)	(0.759)	(0.757)	(0.759)
College Graduation Rate	1.727***	1.730***	1.727***	1.748***	1.754***
	(0.282)	(0.283)	(0.282)	(0.282)	(0.282)
Unemployment Rate	-1.777***	-1.796***	-1.763**	-1.787***	-1.792***
	(0.686)	(0.687)	(0.685)	(0.686)	(0.686)
Poverty Level	-0.011	-0.011	-0.012	-0.005	-0.002
	(0.472)	(0.473)	(0.471)	(0.475)	(0.476)
Log(Median Household Income)	-0.110*	-0.112*	-0.109*	-0.113*	-0.114*
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
Log(Median Home Value)	0.114**	0.115**	0.114**	0.115**	0.116**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)
Log(Population Density)	-0.078***	-0.073***	-0.082***	-0.078***	-0.076***
	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)

Note. (†) denotes our main specification. K refers to the vaccination capacity per location. Distance form refers to an alternative travel cost function where individuals are indifferent to distances below $\bar{d}$. Standard errors are in parentheses. Significance levels: *p<0.05; **p<0.01; ***p<0.001

C.2. Optimization Results

We further test the robustness of our optimal facility location results by varying the values of parameters M, K, and distance forms consistently across both the estimation and optimization stages. Under each scenario, we compute the change in predicted vaccinations relative to the Pharmacy-only strategy (also evaluated under the revised parameter values), and report results in Table C6.

Choice set. For computational tractability, we re-run the optimization model under one alternative value ($M = 10$) for the choice set size. With a larger choice set, the marginal benefit of adding dollar stores decreases as individuals can access more capacity, but at farther locations. Moreover, expanding the choice set affects HPI groups differentially, as some areas have a higher density of pharmacies so expanding the choice set has minimal impact on travel distance. High HPI areas maintain similar vaccination gains as before (0.16 million versus 0.18 million), yet low HPI areas show lower vaccination gains (0.19 million versus 0.30 million).

Capacity. In scenarios with limited per-store capacity, a network expansion strategy can offer substantially more benefits. When $K = 8,000$, the predicted vaccination gains with adding 500 dollar stores statewide exceed 1.3 million, compared to 0.77 million when $K = 10,000$. Although there is less supply in this setting, adding dollar stores as vaccination sites can increase vaccination rates across all HPIs even if they are somewhat distant, due to individuals' reduced sensitivity to distance (as estimated by the demand model in Table C5). The net effect is greater vaccination gains.

Our status quo Pharmacy-only policy with $K = 10,000$ and $N = 4,035$ pharmacies equates to 40.35 million possible vaccination slots, and achieves predicted vaccinations of 26.4 million. Under the reduced capacity scenario of $K = 8,000$ and the addition of 500 dollar stores, a total of 36.28 million vaccination appointments are available, and predicted vaccinations are 26.3 million—nearly identical to the Pharmacy-only policy—due to a more efficient use of appointment slots since dollar stores bring vaccinations closer to individuals' homes. These insights are relevant, for example, when deciding whether to add capacity to existing FRPP locations or to allocate additional capacity among new partner locations. Our analysis suggests the latter may be more effective in our context.

Distance forms. In general, an alternative distance function very modestly changes our optimal results. Total predicted added vaccinations with dollar stores assuming $\bar{d} = 1.0$ mile is 0.70 million, versus 0.77 million in our original analysis. The added walkable vaccinations also decrease (1.40 million versus 1.55 million) because the model does not prioritize reductions in very small distances. Thus, the vaccination gains mainly accrue to individuals who live farther from an FRPP pharmacy.

Table C6 Added vaccinations relative to Pharmacy-only, assuming 500 dollar stores are added statewide

			Vaccinations in millions				
				HPI quartile			
		Type	All	Bottom 25%	25-50%	50-75%	Top 25%
Choice set	$M=5$(†)	Total	0.77	0.30	0.14	0.14	0.18
		Walkable	1.55	0.74	0.30	0.29	0.23
	$M=10$	Total	0.59	0.19	0.10	0.15	0.16
		Walkable	1.31	0.59	0.22	0.26	0.23
Capacity	$K=8,000$	Total	1.31	0.47	0.30	0.26	0.27
		Walkable	1.88	0.80	0.44	0.35	0.30
	$K=12,000$	Total	0.38	0.17	0.07	0.04	0.10
		Walkable	1.20	0.60	0.22	0.20	0.17
Distance	$\bar{d}=0.5$	Total	0.72	0.26	0.12	0.15	0.19
		Walkable	1.38	0.62	0.27	0.26	0.23
	$\bar{d}=1.0$	Total	0.70	0.26	0.12	0.14	0.18
		Walkable	1.40	0.63	0.27	0.26	0.24

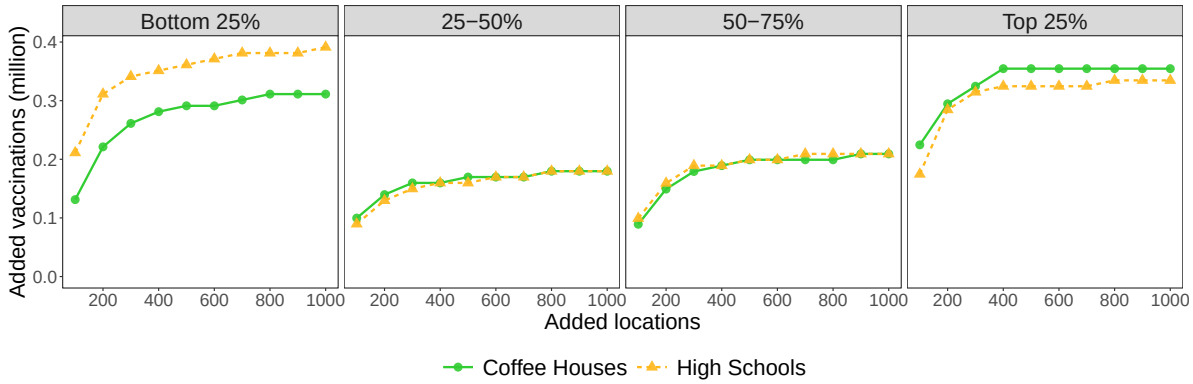
Note. (†) denotes our main scenario. All other partnerships report added vaccinations relative to Pharmacy-only, which is being evaluated under the same parameter setting.

C.3. Alternative partnerships

We consider expanding the FRPP network with an alternative private partner, Starbucks coffee houses, or public high schools. Table A2 summarizes the number of partner locations. Similar to the dollar store partnership, we assess the effects of network expansion with $A=\{100,\ldots,1,000\}$ added locations statewide.

Figure C2 shows the predicted increase in vaccinations by HPI quartile for each partnership. Adding 500 Starbucks locations yields approximately one million additional vaccinations, with 35% of the gains accruing to the top HPI quartile. In contrast, strategically adding 500 high schools achieves a predicted 1.05 million additional vaccinations, with 36% of the vaccinations occurring in the bottom HPI quartile. Notably, over 50% of these benefits are obtained by strategically adding only 100 high school locations.

Figure C2 Added vaccinations relative to Pharmacy-only, assuming network expansion with A={100, . . . , 1,000} partner locations statewide



C.4. Network replacement

We use the optimization model to strategically redistribute capacity within the network. Rather than increasing the total number of locations, the model helps us determine the most effective way to transfer capacity from existing FRPP pharmacies to selected partner locations. This strategy preserves the current size of the vaccine distribution network in terms of both the number of distribution points and the overall capacity (we set $N = 4,035$, the number of FRPP locations).

Under this capacity-neutral scenario, predicted vaccinations increase by 0.69 million statewide, a gain of 2.6% over the Pharmacy-only policy. Figure C3 summarizes the predicted added vaccinations by HPI. We note a sizeable increase in walkable vaccinations (0.65 million) among those in the lowest HPI quartile.

Overall, the model replaces 626 FRPP pharmacies with dollar stores. Figure C4 shows the changes in the number of vaccination locations within each HPI quartile. The lowest HPI areas see a net gain in vaccination sites, primarily through the addition of selected dollar stores. Higher HPI areas have fewer vaccination options nearby, because they currently have excess capacity and they are less sensitive to longer travel distances.

Figure C3 Added vaccinations relative to Pharmacy-only, under network replacement with dollar stores

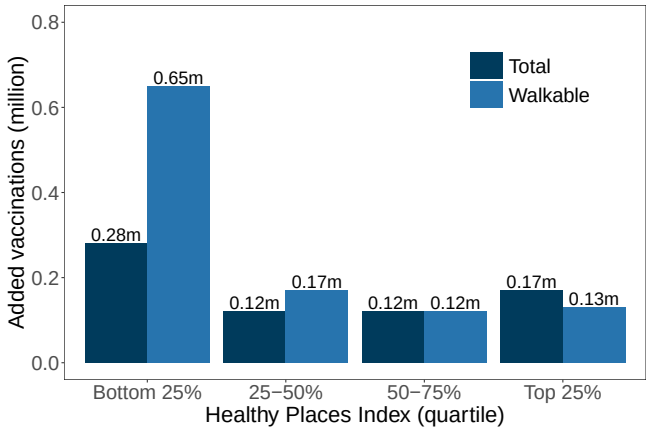
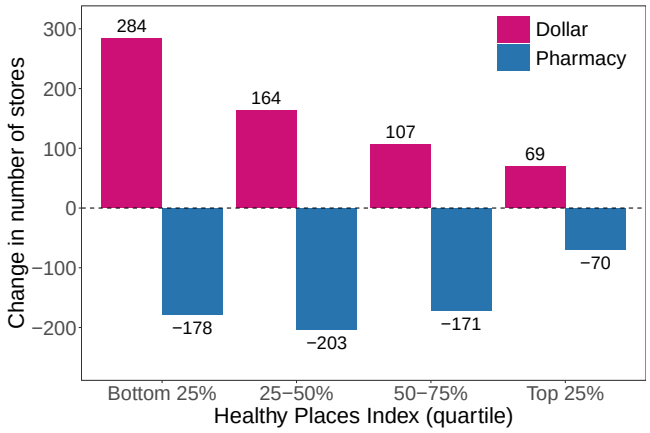


Figure C4 Change in number of stores by HPI quartile under network replacement



Appendix References

Banks M (2021) Vaccine clinic coming to Freetown Dollar General. *The Tribune*, URL http://tribtown.com/2021/06/09/vaccine_clinic_coming_to_freetown_dollar_general/, Accessed May 2024.

Cover Virginia (2022) It's About US: Get Your Free COVID-19 Vaccination at Family Dollar. *Twitter*, URL https://twitter.com/coverva/status/1501656007507906567, Accessed May 2024.

Hall C (2021) COVID-19 Vaccine Clinics to be in Dollar General Stores in 9 Michigan Counties. *Detroit Free Press*, URL https://www.freep.com/story/news/local/michigan/2021/08/06/covid-19-vaccine-clinics-dollar-general-stores/5513003001/, Accessed May 2024.

Health Plan of San Joaquin (2022) COVID-19 Vaccine Clinic—NAACP/Dollar General. URL https://www.hpsj.com/events/covid-19-vaccine-clinic-naacp-dollar-general/, Accessed May 2024.

Kansas Office of the Governor (2021) Governor Laura Kelly Announces Dollar General Stores Join Dillons Health in $100 Vaccine Incentive Program. URL https://governor.kansas.gov/governor-laura-kelly-announces-dollar-general-stores-join-dillons-health-in-100-vaccine-incentive-program/, Accessed May 2024.

Staff Reports (2021a) Free COVID vaccines at Dollar General stores. *The Winchester Star*, URL https://www.winchesterstar.com/winchester_star/free-covid-vaccines-at-dollar-general-stores/article_355be595-b4b2-5568-9f3b-eeb85b075d8c.html, Accessed May 2024.

Staff Reports (2021b) Where to get a free COVID-19 vaccine, test in Sumter, Lee, Clarendon counties: Aug. 24-29. *The Sumter Item*, URL https://www.theitem.com/stories/where-to-get-a-free-covid-19-vaccine-test-in-sumter-lee-clarendon-counties-aug-24-29,369251, Accessed May 2024.

Virginia Department of Health (2021) VDH partners with Dollar General to Expand Access to COVID-19 Testing. URL https://www.vdh.virginia.gov/news/2020-regional-news-releases/vdh-partners-with-dollar-general-to-expand-access-to-covid-19-testing/, Accessed May 2024.

WFXR Fox (2022) Dollar General to provide COVID vaccines at stores in Southside Health District, including Halifax County. URL https://www.wfxrtv.com/news/health/coronavirus/dollar-general-to-provide-covid-vaccines-at-stores-in-southside-health-district-including-halifax-county/, Accessed May 2024.

WHSV Newsroom (2022) CSHD hopes Dollar General partnership will increase area vaccination rates. URL https://www.whsv.com/2022/01/30/cshd-hopes-dollar-general-partnership-will-increase-area-vaccination-rates/, Accessed Mar 2024.

WTVG News (2021) Free COVID-19 vaccines offered at Toledo Dollar General stores. URL https://www.13abc.com/2021/10/19/free-covid-19-vaccines-offered-toledo-dollar-general-stores/, Accessed May 2024.