Title

Closer to Home: A Structural Estimate-Then-Optimize Approach to Improve Access to Healthcare Services

Authors

Bravo, Fernanda
Gandhi, Ashvin
Hu, Jingyuan
et al.

Publication Date

2025

DOI

10.1287/mnsc.2024.06274

Closer to Home: A Structural 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 healthcare access extend beyond rural-urban divides to include socioeconomic, racial, and other disparities. Proximity to hospitals, clinics, healthcare providers, and pharmacies varies widely, posing a challenge in deciding where to strategically locate such facilities. Demand for each service depends on local population health, individual preferences, provider capacity, and other factors. This study introduces a novel structural estimate-then-optimize (SETO) framework, combining structural demand estimation using a modified Berry-Levinsohn-Pakes (BLP) approach that accounts for provider capacity with a choice-based optimal facility location model to maximize health service utilization.

Our methodology is illustrated with a case study on the Federal Retail Pharmacy Program in California, a public-private partnership that administered millions of COVID-19 vaccinations. Demand estimates indicate that residents of socioeconomically vulnerable communities are more sensitive to travel distances to pharmacy-based vaccination sites. Strategically adding 500 retail stores serving lower-income communities increases predicted vaccinations by 2.9% overall (770,000 additional vaccinations statewide) and by 5.3% in the least healthy neighborhoods. Our integrative SETO approach outperforms heuristics that allocate resources based on current vaccination rates, existing service gaps, population density, or predicted demand.

The case study demonstrates the importance of (1) accounting for heterogeneity in estimating demand and (2) selecting partnerships to complement existing networks with spatially heterogeneous supply and efficiently fill service gaps. Our study provides a systematic approach to optimize healthcare delivery networks, using publicly available aggregate data while accounting for individuals' preferences, highlighting the value of combining a structural demand model with prescriptive analytics.

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

1. Introduction

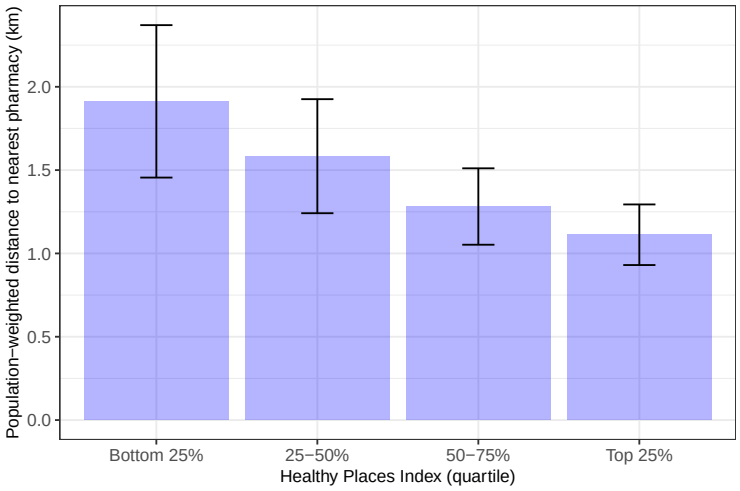
Across the United States, access to essential healthcare resources, such as acute care hospital beds, maternity units, primary care physicians, or pharmacy services, often depends on location (Bronner et al. 2021). Numerous studies show that proximity to a health service affects utilization, expenditures, and patient health outcomes. More than half of the geographic variation in healthcare utilization are explained by supply-side differences, with the rest attributed to patient-related demand (Finkelstein et al. 2016). Proximity to a hospital affects utilization of emergency care, such as treatment following a heart attack (Tay 2003), and routine inpatient services (Gowrisankaran

et al. 2015, Shepard 2022). Similarly, nursing home facility choice is highly sensitive to distance, affecting both patient selection (Gandhi 2023) and local competition (Hackmann 2019).

Long distances to health service providers pose greater challenges for some communities. Economically and socially vulnerable individuals face additional hurdles such as limited time off work, lower vehicle ownership, and limited public transit access (Syed et al. 2013). Rural Americans generally have worse health outcomes and shorter life expectancy (Deryugina and Molitor 2021), owing in part to fewer and lower quality 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).

Pharmacists play a crucial role in dispensing medications, consulting with patients, and administering vaccinations. Five percent of Americans live in *pharmacy deserts*—areas without a pharmacy in a one-mile radius (ten miles in rural areas)—and lack access to essential services (Wittenauer et al. 2024). Pharmacy deserts increasingly arise as retail pharmacies close and the industry consolidates (Salako et al. 2018). At low-volume pharmacies, Medicaid reimbursements are often insufficient to cover operating costs, leading to further closures (Ippolito et al. 2020). While pharmacy deserts are common in rural areas, some densely populated regions, particularly Black and Hispanic neighborhoods, also face pharmacist shortages (Guadamuz et al. 2021). In California, 2.5 million residents (6%) live in pharmacy deserts, including 400,000 in urban Los Angeles (Wittenauer et al. 2024). Figure 1 highlights the inequality in pharmacy access by the Healthy Places Index (HPI), a composite measure of community well-being (Public Health Alliance of Southern California 2022).

Figure 1 Average distance to a pharmacy offering COVID-19 vaccinations in California by HPI quartile



Note: Bars correspond to the population-weighted average distances by zip code within each HPI quartile and error bars are 95% confidence intervals.

Table 1 Healthcare features that support a structural estimate-then-optimize framework

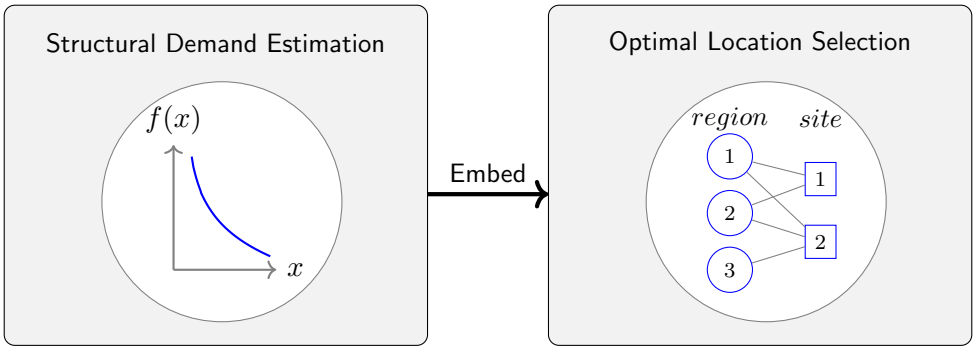
Demand-side (by patients)	Only aggregate data on service utilization (no individual-level outcomes)
	Heterogeneous sensitivity to provider's proximity
Supply-side (by providers)	Spatial inequality in the distribution of service providers
	Unobserved utilization rates by provider
	Limited service capacity per provider
	Undifferentiated health service

Mitigating spatial inequalities in healthcare access has urgent policy importance. One potential mechanism is better positioning of medical providers, given a well-defined objective (*e.g.*, maximize service utilization, minimize average distance traveled). However, this requires an understanding of individuals' preferences in selecting a provider. One empirical challenge is the lack of individual outcome data due to privacy concerns or constraints in obtaining patient health records. For example, disease prevalence is reported at an aggregate geographic level, and policymakers typically lack provider-level outcome data or information on which patients visited each healthcare facility.

Consider an undifferentiated service, such as routine mammography, where service needs are *heterogeneous* across patient groups and individual preferences for a service provider vary. Provider availability differs by geographic area, contributing to inequities in access. Table 1 summarizes the key features often present in the healthcare contexts that motivate our work. Given these challenges, our study addresses an important question in healthcare delivery: *How to optimally design a network of providers offering a standard service, to improve access to care?* This design could involve adding, closing, or replacing facilities. We propose a data-driven approach to optimally design the healthcare delivery network, with an aim to realign a mismatch between a supply of service providers and patient demand for such service (Figure 2). Our contributions are summarized as follows:

- *Structural estimate-then-optimize (SETO) approach.* We develop a novel two-stage approach where the first stage corresponds to *structural demand estimation* that relies only on aggregated, publicly available healthcare utilization data. We employ a modified Berry-Levinsohn-Pakes (BLP) approach that, beyond allowing for spatial demand heterogeneity, also incorporates facility capacity limits. The second stage is an *optimization approach* that integrates the demand estimates into a choice-based facility location problem to identify the optimal set of locations that maximize predicted uptake. Our framework links an established technique from economics—structural demand estimation—with optimization, harnessing observational aggregated data to inform prescriptive modeling.
- *Case study on vaccine distribution.* Our study demonstrates the potential value of integrating structural demand estimation with prescriptive analytics to design more equitable and efficient vaccination networks. In this context, heterogeneity exists both on the demand side as

Figure 2 **Proposed structural estimate-then-optimize (SETO) modeling framework**



socioeconomically vulnerable communities show greater sensitivity to travel distances, and on the supply side as existing pharmacy sites fail to reach underserved populations. Using only aggregated data, our first stage estimates these heterogeneous demand elasticities by HPI, and our second stage optimally selects locations that complement the existing network, reducing service gaps and improving access. Our integrated SETO approach delivers more predicted gains than heuristic approaches that rely on demand estimation or optimization, but not both.

2. Related Literature

Our study builds on extensive prior work in healthcare accessibility, optimal facility location, and demand estimation. Each topic spans multiple fields, and we highlight the most relevant studies.

2.1. Accessibility of Health Services

Healthcare access has been a topic 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: *accessibility* concerns the geographical proximity of healthcare providers; *availability* relates to services offered meeting patient needs; *accommodation* concerns health system features such as appointment scheduling; *acceptability* considers whether patients find providers' attributes (*e.g.*, gender, race) acceptable; and *affordability* addresses whether costs are within the financial means of patients or their insurers. Ayer et al. (2014) summarize research within operations management that examine several of these dimensions.

Socioeconomic disparities in healthcare access are evident in various healthcare contexts including organ transplantation (Wang et al. 2022), decision support systems (Ganju et al. 2020), and appointment scheduling (Samorani et al. 2022). For patients needing treatment for chronic diseases, distance to a provider can be a major barrier to accessing care (Kelly et al. 2016). For instance, use of radiation therapy for breast cancer is 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).

Prior studies show a link between proximity to influenza vaccination providers and uptake (Beshears et al. 2016). During the COVID pandemic, access to a retail pharmacy was associated with higher vaccination rates (Brownstein et al. 2022), including for residents of California, regardless of political background (Mazar et al. 2023). Although these studies report a correlation between distance and vaccine uptake, our demand estimation is more formal and economically micro-founded. Despite a national vaccination campaign launching in 2021, vaccine availability was limited in rural and socioeconomically disadvantaged areas (Zhong et al. 2023, Rader et al. 2022). Chevalier et al. (2022) find that offering COVID vaccinations at Dollar General stores could substantially benefit low-income and minority households. Improving vaccination access has been studied under various objectives (Enayati and Özaltın 2020, Rastegar et al. 2021). Bennouna et al. (2023) use a machine learning-based epidemic model and optimization to efficiently allocate vaccines to different populations. Dey et al. (2024) provide a recent literature review.

Our case study extends these efforts by precisely estimating demand for a COVID vaccine under service provider capacity constraints and formulating an optimization model to strategically choose vaccination delivery sites. Our study is the first to *structurally estimate* different vaccination elasticities to travel distance—a heterogeneous demand curve—using only aggregate data. Our methodology could extend to other contexts with varying distance sensitivities and unequal provider availability, such as cancer screenings, kidney dialysis, and substance abuse treatment.

2.2. Optimal Facility Location

A wide literature exists on optimal facility location, particularly the Maximal Covering Location Problem (MCLP) (Church and ReVelle 1974), including extensive applications in healthcare (Daskin and Dean 2005, Ahmadi-Javid et al. 2017, Jia et al. 2007). Under stochastic demand, MCLP models allow for customer queuing at each location. Ekici et al. (2014) formulate an MCLP with underlying epidemic model to project the spread of an infectious disease. Deo and Sohoni (2015) and Jónasson et al. (2017) assign HIV testing capacity across geographically distant testing facilities, while allowing for congestion and the test collection rate to decline in turnaround time. While these studies focus on optimal capacity allocation, our study relates to *strategic network design* (*i.e.*, service locations) for equitable healthcare access. Prior studies have integrated distance sensitivity into MCLPs in healthcare settings including primary care (Nobles et al. 2014), influenza vaccination (Heier Stamm et al. 2017), and HIV care (McCoy and Johnson 2014). These studies assume a homogeneous linear relationship between distance and service utilization, whereas our model incorporates empirically estimated heterogeneous demand curves, resulting in more equitable resource allocation than under a homogeneous demand model.

In contexts with uncertain travel distances and costly service delays, robust optimization can be used. Boutilier and Chan (2020) use predict-then-optimize to predict daily demand for emergency medical services and then optimally locate and route ambulances using robust optimization. While they consider demand to be exogenous, we estimate a demand curve that is *endogenously* determined based on the set of service providers. In other words, we model demand for service as changing in response to the underlying cost (*e.g.*, travel distance).

Most healthcare facility location studies assume a central planner assigns individuals to facilities. Lee et al. (2013) optimally select mass medical dispensing sites during an outbreak and assign households to facilities. Bertsimas et al. (2022) combine a dynamic compartmental model of COVID-19 with optimization to forecast the epidemic's trajectory, optimally select mass vaccination sites, assign populations to each site, and allocate vaccines by age group. While their resulting strategy is driven by epidemic conditions, we embed individual preferences into a facility location problem via a choice model. *Choice-based* location models examine how individual preferences, influenced by factors such as distance or service quality, shape facility location strategies. Such preferences may be captured using a multinomial logit (MNL) discrete choice model, and then subsequently incorporated in an optimization model. For example, in choosing a healthcare provider, studies vary in modeling patient utility based on price and quality (Denoyel et al. 2017), time sensitivity (Zhang et al. 2012), waiting time (Krohn et al. 2021), or travel time (Hwang et al. 2022). Notably, the first three studies do not include empirical demand estimation but instead perform calibration exercises. In a study of maternity care, Hwang et al. (2022) use aggregate geographic covariates to estimate value function parameters. In contrast, our utility function explicitly depends on proximity to a provider and allows for heterogeneous distance elasticities. Without individual-level choice data, we rely on the BLP approach to structurally estimate demand.

Our approach 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 (Méndez-Díaz et al. 2014) and conic programming (Sen et al. 2018) solution techniques. Similarly, we provide a mixed-integer problem that reformulates nonlinear terms to achieve a tractable facility location problem.

2.3. Structural Demand Estimation

The primary input into a choice-based facility location model is the demand curve for obtaining service, as a function of travel distance. We draw on the established literature from economics on structural demand estimation, which follows a random-coefficient logit model using aggregate

choice data. This approach, known as "BLP" (Berry et al. 1995), has been widely used due to its ability to capture individual-level preferences 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 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 in 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.

2.4. Predict-then-Optimize

This approach is widely used in decision-making problems with uncertain parameters. First, machine learning is used to estimate these parameters by minimizing prediction error through a loss function (*e.g.*, mean squared error in regression). Then, the estimated parameters are incorporated into an optimization model to determine the best decision based on an objective function and relevant constraints.

One potential concern of this method is that small prediction errors can lead to suboptimal decisions, as noted by Tan and Frazier (2022), who analyze regret bounds for unconstrained optimization problems with smooth objectives. Robust optimization can mitigate sensitivity to estimation errors, leading to more conservative yet stable decisions (Bertsimas and Sim 2004). Recent integrative approaches guide parameter estimation using a decision-aware loss function, optimizing decision quality rather than predictive accuracy (Elmachtoub and Grigas 2022, Chung et al. 2022).

Our framework shares similarities with predict-then-optimize but differs in a key aspect: we employ an econometric approach to estimate demand using only aggregated service utilization data. Notably, our aim is not to maximize predictive accuracy, but to obtain unbiased estimates that rationalize the observed utilization rates.

3. Structural Estimate-then-Optimize Framework

Our framework consists of a micro-founded 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 provider. Crucially, aggregating these individual-level decisions offers a structured way to estimate demand for a service

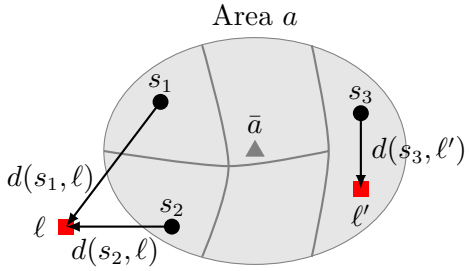
given the locations of service providers. Lastly, we incorporate the estimated demand curve into an optimization model to select facility locations to maximize predicted service utilization.

3.1. Demand Model

Consider a set of locations $\mathcal{L}$ that provide an undifferentiated health service, such as eye exams, vaccinations, blood collections, etc. Individual i decides whether to receive service at a location $\ell_i \in L_i$ (referred to as "provider location" or "service location") from among i's feasible set of provider locations, $L_i \subset \mathcal{L}$. The location represents the provider that i most plausibly considers when choosing whether to obtain service and can vary across settings. For example, ℓ_i could be the most convenient location with available capacity, or ℓ_i might be the closest provider 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 obtain the health service depend on the proximity of available providers. 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 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 (*e.g.*, geodesic distance, travel time, public transit time) between the centroid of s_i and ℓ_i. Figure 3 illustrates an area a, with arrows representing the distance between each subarea s and the closest service location, ℓ or ℓ'.

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 a utility-based binary choice model. Individual i may choose to obtain service at location ℓ_i and receive additional utility u_i relative to no service:[1]

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

A key term in this model is D_i, the disutility that individual i experiences from traveling from home subarea s_i to service location ℓ_i. We define $D_i := d(s_i, \ell_i)\alpha_{a_i}$, where α_{a_i} represents the individual's

[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).

area-specific sensitivity to travel distance. The variation in α_{a_i} captures across-area differences in the travel cost to a service location (*e.g.*, due to the availability of public transit).[2] We denote the parameters characterizing distance sensitivity by the vector $\boldsymbol{\alpha}$. The distance function $d()$ measures individual i's 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 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. With individual-level choice data and exogenous variation in s_i, l_i, and X_{a_i}, it is straightforward to estimate the demand model from Section 3.1.1 via maximum likelihood. However, such favorable settings are exceedingly uncommon. When consumer data are highly sensitive or proprietary, a frequent occurrence in healthcare, we may observe only aggregate data (Table 1). We instead estimate demand using the BLP approach (Berry et al. 1995), which captures individual-level factors influencing decisions without directly observing individual choices. We consider the case where service utilization is observed by geographic area. BLP incorporates factors affecting demand that vary across individuals within a geographic area by treating these factors as *distributions* (*e.g.*, the distribution of income or household locations within an area). These factors are commonly known as "random coefficients." Area-level demand curves are obtained by integrating individual-level demand curves over the distributions of random coefficients.

In our setting, the distance that each individual travels to obtain service, $d(s_i,\ell_i)$, 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 aggregating demand from all individuals living in a:

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

[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.

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.

In computing the aggregate demand for area a in equation (3), a key input is the distribution over travel costs $f(s_i, \ell_i \mid a)$, which can be decomposed as follows:

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

The first term, $f(s_i \mid a)$, is the observed spatial distribution of the population across subareas within area a. This is a reasonable assumption given that public data typically include an extremely granular spatial distribution of the population (*e.g.*, Census blocks). The second term, $f(\ell_i \mid s_i)$,[3] is a distribution over service locations, characterized by a set of parameters, $\boldsymbol{\lambda}$, to be estimated. In some cases, formulating this term is straightforward (*e.g.*, if a central planner assigns individuals to the closest facility). In our case study, $f(\ell_i \mid s_i)$ is determined sequentially (discussed in section 3.1.3), allowing each individual to receive service at her *preferred* location with remaining capacity.

The parameters to be estimated are $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$. BLP exploits the idea that for any given $\boldsymbol{\theta}$, exactly one set of idiosyncratic shocks $(\xi_a)_{a \in \mathcal{A}}$ will precisely rationalize the observed utilization rates (Berry 1994). Thus, in enforcing that these area-level utilization rates implied by the demand model in 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 value $\boldsymbol{\theta}$, the implied idiosyncratic preferences for each area, $\xi_a(\boldsymbol{\theta})$, are uncorrelated with the area-level instruments $Z_a(\boldsymbol{\theta})$. Since $\xi_a(\boldsymbol{\theta})$ is a utility residual from the demand model, equation (5) enforces 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 so that the moment conditions are both informative and plausible. In our case study, Z_a includes the observable area-level characteristics X_a, and the average distance to service providers in the area interacted with area-level demographics. Since service 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 $\boldsymbol{\theta}$.

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

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

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

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. Service Location Distribution. We have presented the demand model and estimation quite generally to emphasize its flexibility. Here, we describe a specific model of sequential location selection used in estimation, connecting service provider availability to individual choices.

We assume that an individual's choice of service 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 probability distribution over service locations, $f(\ell_i \mid s_i)$, the only parameters needed are the sensitivity to distance $\boldsymbol{\alpha}$. We assume a Type I extreme value (*i.e.*, Gumbel) distribution of idiosyncratic preferences over locations, resulting in a multinomial logit (MNL) discrete choice model over the set of service locations L_i:

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

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 of using 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 providers have infinite capacity, individuals could choose their service location from among all sites (*i.e.*, that $L_i = \mathcal{L}$). In reality, healthcare providers have finite capacity, limiting an individual's set of available providers. To capture this, we assume that individuals' choice of service provider is fulfilled sequentially in a random order, where individual i's set of possible service locations ($L_i \subset \mathcal{L}$) is restricted to those *with remaining capacity*.[6] A sequential approach could reflect a random arrival process of identical individual types requesting service slots, such as an online booking of vaccination appointments. To implement this, we first randomly generate an ordering

[4] Note that ξ_a is an implicit function of θ through the constraint that the model-implied market shares are equal to the observed market shares, *i.e.*, $\rho_a(\boldsymbol{\theta}) = \rho_a^{obs}$. This implicit function is enforced at each step of the optimization either through 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 (Dubé et al. 2012).

[5] In our case study, we verify that our demand estimates are similar under this alternative model (see Appendix C.3).

[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 home.

of all individuals in the population. Then, we iterate over the individuals and update each choice set L_i according to the location choices (equation 7) and service choices (equation 2) made by individuals $1, ..., i-1$, so that only locations with remaining capacity are included in L_i. Thus, if location ℓ reaches capacity at individual j's turn, then ℓ is removed from the choice sets of all individuals following j (*i.e.*, $L_i \subset \mathcal{L} \backslash \{\ell\}\ \forall i > j$). Finally, $f(\ell_i \mid s_i)$ in equation (4) is 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 a facility location model to determine the optimal set of healthcare service locations to maximize the predicted demand served across all areas.

Consider $\mathcal{L}^+$ (with $\mathcal{L} \subseteq \mathcal{L}^+$) an augmented set of locations in the consideration set. 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 K_ℓ, and we assume that candidate locations are otherwise identical. We allow for a limited budget on 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 subarea level $s \in \mathcal{S}$ (with parent area a). Let m_s be the service-eligible population in subarea s, and let $\rho_{s\ell}$ be the share of the population in subarea s that obtains service at location ℓ. Using equations (2) and (7), 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. Note that the second term in brackets assumes that individuals can choose any of the L open service locations without considering available capacity. This is different than in Section 3.1.3, where sequential service fulfillment tracks remaining capacity at open locations. Here, we incorporate a constraint to ensure that the predicted number of individuals served at each open location ℓ does not exceed its capacity K_ℓ.

For all subarea-location pairs (s, ℓ), we compute $\rho_{s\ell}$ using equation (8) and the estimated distance sensitivities $\hat{\alpha}_a$, 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:

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

and zero otherwise. In equation (9), the term $\exp(-X_a\hat{\boldsymbol{\beta}} - \hat{\xi}_a)$ in the right bracket cancels out. Thus, the predicted demand from subarea s served at 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. To account for partial demand fulfillment, we introduce a decision variable $y_{s\ell} \leq \rho_{s\ell}(\mathbf{x})$ as the share of the eligible population from subarea s that is served at location ℓ. Hence, a location ℓ may be opened to serve a fraction of the overall demand from s to ℓ. The resulting optimization problem is:

$$(\mathsf{P}_{\mathsf{loc}}) \qquad \max_{\mathbf{x},\mathbf{y}} \ \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}} m_s y_{s\ell} \tag{10a}$$

$$\text{s.t.} \ \sum_{s\in\mathcal{S}} m_s y_{s\ell} \leq x_\ell K_\ell, \ \forall \ell \in \mathcal{L}^+ \tag{10b}$$

$$\sum_{\ell\in\mathcal{L}} x_\ell \leq N \tag{10c}$$

$$y_{s\ell} \leq \rho_{s\ell}(\mathbf{x}), \ \forall \ell \in \mathcal{L}^+, \ s \in \mathcal{S} \tag{10d}$$

$$x_\ell \in \{0,1\}, \ y_{s\ell} \geq 0, \ \forall \ell \in \mathcal{L}^+, \ s \in \mathcal{S}. \tag{10e}$$

The objective (10a) maximizes the total expected demand served, where each vaccination is accounted for equally. Constraints (10b) ensure that the predicted demand served by an open location does not exceed its capacity. Note that multiple locations may serve the demand of a subarea, as determined by individuals' preferences reflected in $\rho_{s\ell}(\mathbf{x})$ through $y_{s\ell}$. Constraint (10c) imposes a total budget of N open locations. Constraints (10d) ensure that partial fulfillment of a subarea's demand is bounded by the total induced demand $\rho_{s\ell}(\mathbf{x})$. This formulation can be adapted to include other considerations, such as linear costs for selecting less desirable locations (*e.g.*, stores more distant to public transit), limiting travel distances, or imposing equity constraints regarding distance traveled or demand served across groups.

($\mathsf{P}_{\mathsf{loc}}$) is essentially an assortment optimization problem with cardinality, capacity, and partial fulfillment considerations, but is computationally challenging because of its nonlinear constraints, which rely on the choice probabilities $\rho_{s\ell}(\mathbf{x})$. Using techniques from assortment optimization under an MNL choice model (Bront et al. 2009), we 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:

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

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

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

Here, $M_s = \bar{v}_s := \max_{k\in\mathcal{L}^+}[\exp(\bar{u}(s,k))^{-1}]$. Even after incorporating these auxiliary variables, the constraints (10d) and (11a) remain nonlinear. Following the linearization approach from Wu (1997), we define $w_{s\ell} := v_s x_\ell$, and introduce a set of linear constraints:

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

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

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

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

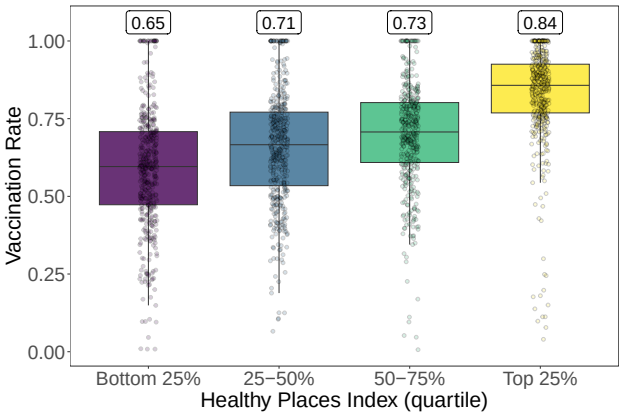
$$\begin{aligned}
(\tilde{\mathsf{P}}_{\mathsf{loc}}) \qquad \max_{\mathbf{x},\mathbf{y},\mathbf{b},\mathbf{v},\mathbf{w}} \ & \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}^+} m_s y_{s\ell} \\
\text{s.t.} \ & \sum_{s\in\mathcal{S}} m_s y_{s\ell} \le x_\ell K_\ell, \ \ \forall \ell \in \mathcal{L}^+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

Here, $\mathbf{b}$, $\mathbf{v}$, and $\mathbf{w}$ are auxiliary variables used to linearize 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. Policy 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

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.

order all individuals in the population. 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.[7] A detailed description of the entire estimate-then-optimize approach is provided in Appendix B.3.

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

In early 2021, the U.S. government launched the Federal Retail Pharmacy Program (FRPP) to expand COVID vaccination distribution points through a partnership with 21 national chains and independent pharmacy networks. By August 2023, more than 300 million doses were administered through FRPP, across 41,000 locations nationwide (CDC 2023). Nearly half of all Americans live within one mile of a pharmacy offering COVID vaccinations and 89% live within five miles. However, travel distances vary widely, particularly for Black communities, who often face distances exceeding ten miles (Berenbrok et al. 2021). This disparity extends beyond rural areas, with limited vaccination options among communities of Color and those with chronic health conditions (Rader et al. 2022). In California, residents in the bottom HPI quartile face average travel distances 72% greater than those in the top quartile (Figure 1). Median vaccination rates are 19 percentage-points lower in the bottom HPI quartile (Figure 4), in part due to these structural barriers to vaccination.

Using our SETO approach, we evaluate an FRPP expansion to include other partnerships to help alleviate a major barrier to vaccination access. Using an estimated spatial demand model (stage 1), we predict COVID vaccination uptake under counterfactual scenarios of offering vaccines at these

[7] The main results are reported for a single ordering of the population. In Appendix C.7, we perform sensitivity analysis for several population orderings.

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 main analysis focuses on dollar stores, but we also examine partnering with Starbucks or public high schools. After FRPP's 2021 launch, the CDC director and Dollar General confirmed discussions about administering COVID vaccines in stores (Roberts 2021). Dollar General operates over 17,000 locations in 46 states—nearly double the next largest private vaccine provider—with 75% serving rural communities (Bomey 2021). Two other chains, Dollar Tree and Family Dollar, each operate over 7,000 stores nationwide. Collectively, these discount retailers offer a promising way to expand vaccination access in low-income communities (Chevalier et al. 2022).

4.1. Data

Our case study integrates data on vaccination rates, spatial demographics, and store locations.

Vaccination rates. We obtain the most granular population-level COVID vaccination data available in California: 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, although we consider alternative time points in Appendix C.5.

Vaccination locations. Geographic coordinates 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. We exclude other dispensing sites such as mass vaccination centers or physician offices because pharmacies accounted for 90% of COVID vaccinations administered nationwide (IQVIA 2023).

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 are obtained from ScrapeHero (2021), Starbucks locations are from Safegraph, and public school locations are from the California Department of Education (2020).

Demographics. 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 weighted 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 into HPI quartiles, where the top 25% are the healthiest and the bottom 25% are the most vulnerable.

Population estimates of California's residents aged 5+ for each of the 376,762 Census blocks are used, totaling 37,336,716.[8] The geographic coordinates of each block's population centroid

[8] Census blocks are small geographic regions within a tract, with approximately 100 residents on average. We only consider Census blocks with non-zero population.

characterizes the spatial distribution of the vaccine-eligible population (age 5+) (US Census Bureau 2019). Demographic data are obtained from the 2019 American Community Survey for 1,751 zip codes in the state[9]. 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. Summary statistics are available in Appendix Table A1.

4.2. Implementing the Estimate-then-Optimize Framework

We describe the model parameters and include specific details on how to operationalize each step of our framework. We also provide a small working example at `https://github.com/juanhu96/Demo`.

4.2.1. Estimating Vaccination Demand. Areas ($\mathcal{A}$) correspond to zip codes, the most granular geographic region that reports our primary outcome, COVID vaccination rates (ρ_a^{obs}). 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 $\mathcal{L}$ include pharmacies participating in FRPP. We compute the geodesic distance $d(s, \ell)$ from the centroid of each block s to every location ℓ. This implies that all residents of a block have identical travel distance to any given site, a reasonable assumption given the small size of Census blocks.

We allow sensitivity to travel distance α_{a_i} to vary 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 individual i's zip code, resulting in a vector $\boldsymbol{\alpha}$ consisting of the distance sensitivities for each HPI quartile. This permits heterogeneity in travel distance disutility across socioeconomic status (*e.g.*, long travel distances may be a larger deterrent for lower-HPI individuals who have limited transportation access or time off work). Appendix C.6 examines *household income* as a single dimension to measure heterogeneity in distance sensitivity.

An individual's utility from vaccination (equation 1) incorporates zip code-level demographic variables, X_{a_i}, including 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.

For each subarea, we consider the set of possible vaccination locations to be the M closest sites, to avoid diluting the choice probabilities with excessively distant locations and to ease computation. In both the estimation and optimization stages, individuals can obtain service at one of these M locations.[10] We initially set $M = 5$ but we vary this in sensitivity analysis. Capacity per vaccination

[9] The analysis includes 1,659 zip codes with non-zero populations.

[10] To estimate demand via BLP, all individuals must be assigned a service 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).

location is assumed to be constant, $K_\ell = 10,000 \ \ \forall \ell \in \mathcal{L}$, but we consider $K_\ell \in \{8,000,\ 12,000\}$ in sensitivity analysis. We compute K_ℓ such that total capacity (40.35 million vaccines) sufficiently covers California's vaccine-eligible population.

Following Section 3.1.3, we model individuals' choice of service locations under capacity constraints. We first randomly order all 37.34 million individuals in California. Then, individuals choose a location and decide to vaccinate (or not) in that order. An individual's set of possible vaccination locations L_i is determined by the remaining capacity at nearby sites after all previous individuals have made their location and vaccination decisions. We test the robustness of our demand estimates to the assignment mechanism in Appendices C.3 and C.7.

Our main demand estimates measure travel costs using log-distance between each subarea and vaccination location. This functional form can result in the optimization model seeking to reduce small travel distances, while neglecting reductions in longer travel distances. We also consider an alternative distance function that treats all distances below a threshold $\bar{d}$ as constant. We explore various thresholds $\bar{d} = \{0.5$ miles, 1 mile$\}$ in Appendix C.4.

With the need to assign more than 376,000 Census blocks in California to one or more vaccination locations, our mixed-integer facility location problem is computationally intractable. We therefore formulate this portion (the second stage of our framework) at the Census tract-level as follows. First, we compute block-level utility estimates using the estimated demand model parameters. Then, we aggregate demand up to the Census tract-level and compute approximate 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. We let $\mathcal{L}^+$ include the existing FRPP pharmacies ($N = 4,035$) and any new candidate locations. The model ($\tilde{\mathsf{P}}_{\mathsf{loc}}$) selects the optimal set of new sites assuming all FRPP pharmacies remain operational, referred to as *network expansion*. We consider partnering with dollar stores, assuming network expansion to 100 to 1,000 locations (out of 1,016 stores statewide). We also consider partnering with public high schools (out of 1,225 locations statewide) or Starbucks (out of 3,172 locations statewide), as detailed in Appendix E.[11]

Our primary outcome measure is the predicted gain in vaccinations compared to a *Pharmacy-only* policy, which utilizes only the $N = 4,035$ pharmacies participating in FRPP in California as of June 2021. To maintain consistency, we employ the same parameter values throughout the estimation and optimization stages. We assume that partner locations have equal capacity to existing pharmacies, but we test a scenario where the capacity is reduced by one-half (Appendix C.2).

We solve ($\tilde{\mathsf{P}}_{\mathsf{loc}}$) using Python 3.9.12 and the Gurobi 10.0.1 solver. The algorithm is configured to halt execution once the optimality gap falls below 0.1%, or runtime reaches six hours.

[11] The partner's locations refer to the geographic coordinates of the vaccination distribution points and the specific deployment strategy could be on-site, pop-up, or mobile clinic.

4.3. Results

We first present empirical demand estimates under *Pharmacy-only*, and we validate the choice-based model's predictions through policy evaluation. We then conduct counterfactual analyses to assess network expansion and compare results to various benchmarks. We explore site closures and replacement strategies in Appendix G.

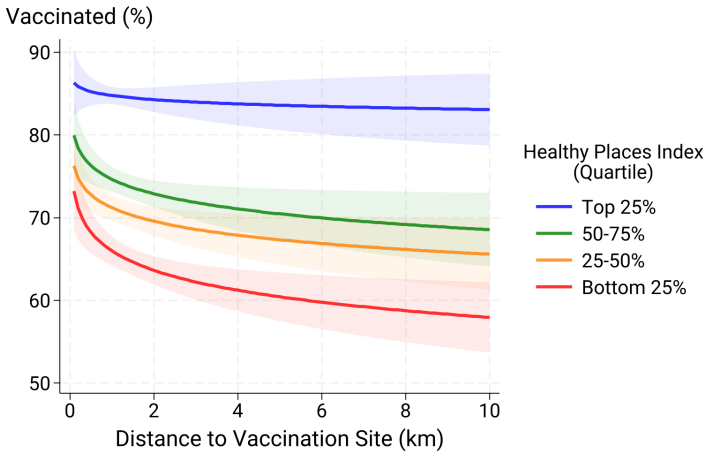
4.3.1. Demand Estimates. Table 2 gives our main estimates and key demographic predictors of vaccine demand. HPI strongly predicts vaccination uptake, with less healthy areas showing conditionally lower vaccination rates, suggesting that the observed spatial inequalities in vaccinations mirror those seen in other dimensions of health. Regions with higher employment 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 population fully vaccinated for COVID as of March 1, 2022. Parameter estimation is described in Section 3.1.2. Healthy Places Index (HPI) is a composite measure of a community's health and well-being. Log-distance is the distance between an individual's Census block centroid and the selected vaccination location. Significance levels: *p<0.05; **p<0.01; ***p<0.001

Figure 5 Predicted COVID-19 vaccination rates



Note. Within each HPI quartile, we compute the mean vaccination rate, assuming that all individuals are assigned to a location with the given distance (x-axis).

and more college-educated residents also show greater vaccine demand. While these estimates are suggestive, individual coefficients should be interpreted with some caution due to high correlations among demographic covariates.[12]

The estimates of interest are the distance sensitivities by HPI quartile, α_h, which reveal a clear and economically meaningful pattern: α_h is negative and statistically significant for all HPI quartiles except the top one, whose coefficient is statistically indistinguishable from zero. Less healthy communities experience greater declines in vaccination rates as distance increases. Doubling the average distance from 1 km to 2 km equates to a 2.4 percentage-point decrease in vaccinations for the bottom HPI quartile, compared to a 0.5 percentage-point reduction for the top HPI quartile. This aligns with the observation that longer travel distances are more burdensome for individuals with lower socioeconomic status. The coefficients remain significant after controlling for observable demographics, suggesting that the results cannot be explained by spatial demographics alone, but indicate a genuine sensitivity to distance. Figure 5 shows predicted vaccination rates by HPI quartile, highlighting the unequal uptake and greater sensitivities for the bottom quartiles.

The preceding evidence underscores the strong correlation between distance to a vaccination site and uptake. One potential threat to this model specification is selection bias among pharmacies in FRPP, which might arise if the selection process is *ex-ante* correlated with both vaccination propensity and HPI—beyond what is captured by observable demographics. However, this is unlikely since most participating locations are national chains (*e.g.*, CVS, Walgreens) with established store locations before the COVID vaccine gained FDA approval. Including additional demographics helps control for the tendency of retail pharmacies to locate in higher-income, urban areas.

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

Table 3 Predicted vaccinations under network expansion with 500 additional dollar stores statewide

			HPI quartile			
Strategy		All	Bottom 25%	25-50%	50-75%	Top 25%
Pharmacy-only	Total	26.44	5.62	6.68	6.83	7.31
	Walkable	19.46	3.98	5.07	4.96	5.45
Pharmacy + Dollar	Total	+0.77	+0.30	+0.14	+0.14	+0.19
	Walkable	+1.54	+0.73	+0.29	+0.29	+0.24

Note. All values are reported in millions of vaccinations. The Pharmacy + Dollar strategy is the change in vaccinations compared to Pharmacy-only. Walkable vaccinations refer to travel distances of less than one mile.

We conduct extensive robustness checks, as detailed in Appendix C. Demand estimates remain qualitatively similar across various specifications, including alternative choice sets (Table C11), capacity constraints (Table C12), functional forms for distance (Table C12), cross-sectional time periods (Table C13), pandemic controls (Table C13), and demand fulfillment (Table C15). In all cases, the estimates consistently reflect significant vaccination inequalities by HPI and greater distance sensitivities for low-HPI individuals.

4.3.2. Baseline Network Validation. Under the baseline Pharmacy-only scenario, predicted vaccinations vary widely between the top and bottom HPI quartiles (Table 3). This gap arises because lower HPI residents live farther from FRPP pharmacies (Figure 1) and have steeper travel distance elasticities (Figure 5). Of the 26.4 million predicted vaccinations, three-quarters (19.5 million) occur at pharmacies within a 1-mile walking distance, though fewer are walkable for residents of the lowest HPI quartile. Our model's predicted vaccinations broadly align with observed zip code-level vaccination rates (Figure 4). As expected with BLP estimation, goodness-of-fit measures confirm strong model performance (Appendix Table A4). At the county-level, the model slightly underpredicts vaccination rates in some highly vaccinated counties (Figure A2), likely due to more restrictive choice sets and capacities imposed during the policy evaluation stage.[10]

4.3.3. Network Expansion. Augmenting the FRPP network to include 500 *optimally selected* dollar stores statewide generates 0.77 million additional predicted vaccinations (a 2.9% increase), a marginal gain of approximately 1,500 vaccinations per added dollar store. While residents of the bottom HPI quartile accrue much of the vaccination benefits (a 5.3% increase), adding dollar stores does benefit high HPI communities (a 2.6% increase) by expanding capacity and allowing individuals across the HPI spectrum to obtain a vaccination closer to home. Notably, in the lower HPI communities that have lower automobile ownership and access to public transit, *walkable* vaccinations increase by 18% (Table 3, bottom panel). We observe diminishing returns as the number of partner locations increases, as shown by the SETO curve in Figure 7. With just

Table 4 Characteristics of selected dollar stores under heterogeneous and homogeneous demand models, assuming network expansion with 500 locations statewide

		Selected dollar stores	
	All dollar stores	Heterogeneous demand(†)	Homogeneous demand
No. of stores (%)	1,016	500	500
in HPI quartile 4 (most healthy)	89 (9%)	56 (11%)	74 (15%)
in HPI quartile 3	195 (19%)	91 (18%)	104 (21%)
in HPI quartile 2	301 (30%)	127 (26%)	141 (28%)
in HPI quartile 1 (least healthy)	431 (42%)	226 (45%)	181 (36%)
No. of stores (%) in pharmacy deserts	71 (7%)	56 (11%)	36 (7%)
Avg. distance to the nearest pharmacy (km)	2.1	2.8	1.4

Note. (†) denotes our main model results. Homogeneous results are based on the model with constant preferences across locations within the choice set.

100 optimally selected dollar stores, a gain of 540,000 vaccinations is predicted, representing 70% of the gains with 500 dollar stores. With 200 dollar stores, more than 700,000 added vaccinations are predicted, with over 260,000 vaccinations accruing to the bottom HPI quartile.

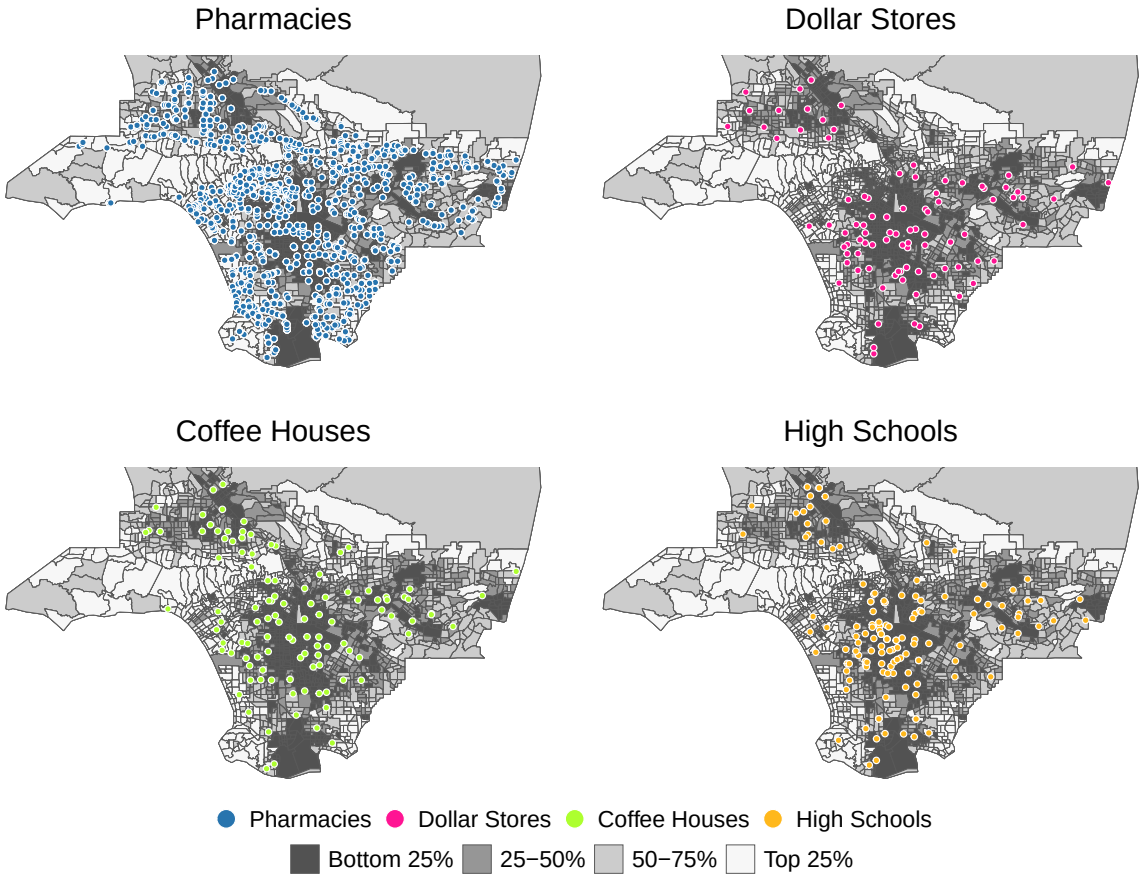
Dollar stores often locate in lower income areas: 431 out of 1,016 stores statewide are in the bottom HPI quartile. The optimal expansion policy indeed prioritizes these locations: out of 500 stores selected, 226 are in the bottom HPI quartile (Table 4, heterogeneous demand). The selected dollar stores tend to be located farther from existing FRPP pharmacies (average distance of 2.8 km), and a sizable proportion of dollar stores inside pharmacy deserts are selected (56 out of 71).

We illustrate the geographic variation in pharmacies and potential partners for Los Angeles County in Figure 6, with darker regions indicating lower HPI tracts. Out of 500 optimally selected dollar stores statewide, 102 are located in LA County, primarily in lower HPI areas. Alternatively, an optimal network expansion policy using public high schools selects 151 schools in LA County, providing more access points in underserved communities, and boosting predicted vaccinations by approximately one million statewide. Finally, a Starbucks partnership selects 117 locations in LA County and also generates about one million additional vaccinations statewide, although the gain primarily accrues to the top HPI quartile. Additional details are given in Appendix E.

We conduct several robustness checks using the revised demand estimates obtained previously. We perform the subsequent optimization and policy evaluation for alternative distance forms, choice sets, capacity constraints, demand fulfillment, cross-sectional time periods, and pandemic controls. All results are reported in Appendix C.

We also examine to what extent misspecification of the demand model affects its performance. If the estimation step ignores demand *heterogeneity* (Appendix Table C16), then the optimization model prioritizes different populations in selecting dollar stores (Table 4). More dollar stores are

Figure 6 **Maps of FRPP pharmacies in Los Angeles County, and optimal locations of partner stores assuming network expansion with 500 locations statewide**

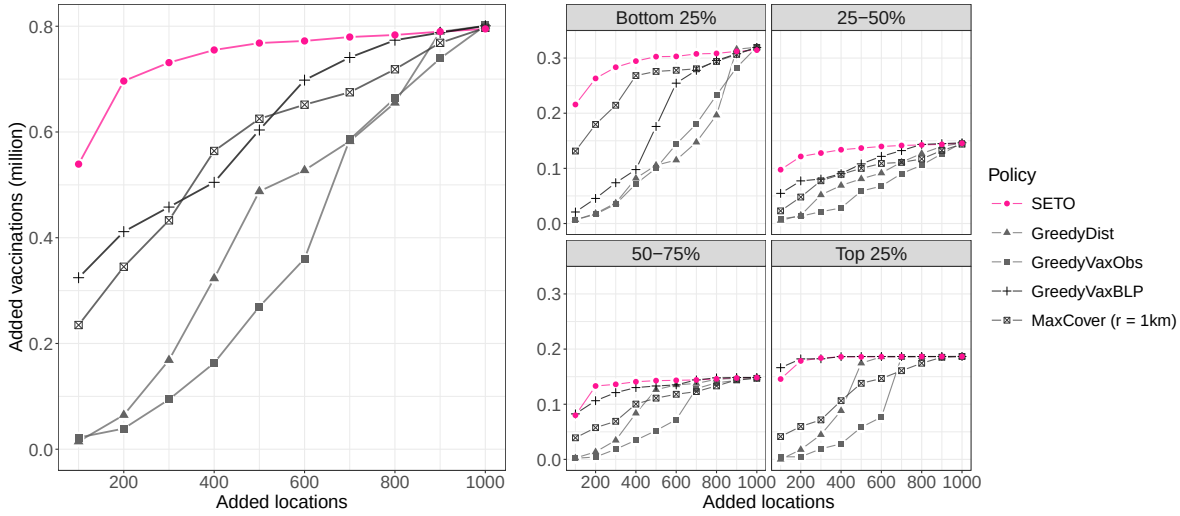


Note. LA County has 982 FRPP pharmacies, 162 dollar stores, 788 Starbucks, and 307 high schools. Network expansion with 500 locations statewide selects 102 dollar stores, 117 Starbucks, or 151 high schools in LA County. Tracts with <1,500 residents or >50% in institutional settings (*e.g.*, dorms, nursing homes, prisons) are Undefined.

selected in the top HPI quartile (74 vs. 56 with heterogeneous demand), with fewer in the bottom quartile (181 vs. 226 with heterogeneous demand). We also observe a shift in selecting dollar stores closer to existing pharmacies (average distance of 1.4 km vs. 2.8 km with heterogeneous demand), with fewer selected inside pharmacy deserts (7% vs. 11% under heterogeneous demand). Altogether, these results highlight the importance of flexibly accounting for heterogeneity while estimating demand, and failing to do so results in a more regressive allocation of resources. We further discuss the effect of ignoring demand heterogeneity on predicted vaccinations in Appendix D.1.

We also assess the effect of ignoring provider *capacity* in demand estimation, and we observe steeper distance elasticities, suggesting a lower willingness to travel for vaccination (Table C12). This occurs because, with unlimited capacity, most individuals can be served at their closest pharmacy. To match the observed vaccination rates, the demand model estimates distance elasticities

Figure 7 Predicted added vaccinations with network expansion of 100 to 1,000 dollar stores statewide



Note. Added vaccinations are relative to Pharmacy-only.

that are greater in magnitude than under the original capacitated model. Intuitively, failing to account for capacity results in the BLP demand model incorrectly rationalizing individuals as highly vaccine-hesitant or distance-sensitive, when in reality, they simply could not find capacity at nearby sites. The effect on predicted vaccinations is detailed in Appendix D.2.

4.3.4. Alternative Benchmarks. We compare the performance of our SETO approach to benchmark policies. The Maximal Covering Location Problem (MaxCover) does not rely on demand estimation and simply aims to maximize the population within radius r of a service location. GreedyVax$^{\text{Obs}}$ prioritizes areas with the lowest observed vaccination rates, while GreedyDist focuses on areas farthest from existing pharmacies. We also include a heuristic policy, GreedyVax$^{\text{BLP}}$, based on the BLP-based demand estimates that prioritizes new locations that result in the most predicted vaccinations, assuming an area's demand is served at the closest new location. Figure 7 shows predicted vaccination gains under each benchmark with 100 to 1,000 dollar stores, using the policy evaluation approach from Section 3.3. Additional benchmark details are in Appendix F.

SETO delivers superior performance in predicted vaccinations, especially when fewer new locations are added. For instance, if 200 new locations are added, MaxCover ($r = 1$ km)—one of the best-performing benchmarks—generates only half the vaccination gains as SETO, amounting to 350,000 fewer vaccinations. This performance gap persists across HPI quartiles, with our approach yielding nearly 100,000 more vaccinations in the bottom HPI quartile compared to MaxCover. Unaware of demand sensitivities, MaxCover prioritizes densely populated areas that often have pharmacies nearby and overlooks potential sites in suburban and rural neighborhoods. Interestingly, GreedyVax$^{\text{BLP}}$ performs similarly to MaxCover, despite not relying on optimization. This

Figure 8 Predicted vaccination gains under different approaches

	Observational Data	BLP Estimation
No Optimization	GreedyDist +0.49 million GreedyVax$^{\text{Obs}}$ +0.27 million	GreedyVax$^{\text{BLP}}$ +0.60 million
With Optimization	MaxCover +0.62 million	SETO +0.77 million

Note. All scenarios assume network expansion with 500 additional dollar stores statewide.

heuristic is demand-aware, favoring areas with high predicted vaccination uptake, which tend to be high HPI and densely populated. However, this policy fails to account for the existence of other vaccination sites and it ultimately selects new locations that overlap with existing pharmacies. Simpler heuristics like GreedyDist and GreedyVax$^{\text{Obs}}$ perform poorly as they ignore distance sensitivities and the spillover effects of opening a new location. To illustrate the limitations of a greedy policy, we examine three California counties (Appendix F.2). In areas with more homogeneous demand (*e.g.*, Orange County), GreedyVax$^{\text{Obs}}$ performs competitively. However, in regions with heterogeneous demand (*e.g.*, Los Angeles County), this myopic policy fails to prioritize communities with average vaccination rates but highly distance-sensitive populations.

Notably, SETO outperforms both the optimization-only and BLP estimation-only policies (Figure 8). For example, under network expansion with 500 new locations, the best heuristics provide 77–80% of the predicted benefits of our SETO approach. This underscores that SETO's advantage lies in integrating the estimated demand model within an optimization framework. By doing so, it ensures that resources are allocated where they are most likely to be used, while also considering the distribution of existing service locations to effectively address spatial gaps in service providers.

4.4. Feasibility and Operational Challenges

Despite vaccinating millions of Americans, FRPP falls short in reaching vulnerable communities. We propose a data-driven, transparent approach to expand vaccination access by partnering with non-healthcare organizations and strategically select new locations to complement the existing network. Adding 100 optimally selected dollar stores in California could boost vaccinations by over half a million—a 2% gain, comparable to prior interventions. Unlike vaccine education, which showed no impact among healthcare staff (Berry et al. 2022), or nudges (Dai et al. 2021) and persuasive advertising (Larsen et al. 2023), which yield modest gains, our model offers a scalable, cost-efficient solution that reduces geographic inequities.

Beyond only vaccination gains, the cost and feasibility of non-pharmacy partnerships are key considerations. Each option has trade-offs as summarized in Appendix Table E20. Starbucks attracts

repeat visitors, enabling frequent vaccinations, but physical space and patient privacy are concerns. High schools offer ample indoor space but face organizational challenges due to decentralized districts. Dollar stores previously hosted pop-up COVID vaccination sites in 2021 (Appendix Table A3), but efforts were ad-hoc and lacked coordination. Our approach helps identify the most effective partnerships, and policymakers can weigh these factors against potential vaccination gains.

Administering vaccinations in non-healthcare settings presents staffing challenges. Some dollar store chains may soon offer pharmacy services themselves (Katje 2021), while another option could entail visiting nurses via agencies like SnapNurse.[13] Staffing costs can be approximated using SnapNurse's average California wage of $30/hour (ZipRecruiter 2023). A three-month campaign with one or two nurses per site for an 8-hour shift per week costs $2,900–$5,800 per location. Partnering with 500 optimally chosen dollar stores would cost $1.4–$2.9 million, yielding 0.77 million vaccinations at $2–$4 per dose. Even a six-month campaign would cost about $5 million—just 0.002% of California's $300 billion budget. Given the protective benefits, these additional vaccinations offer significant social value.

Our case study has several limitations. First, we perform a static analysis to determine vaccination locations, but time-varying vaccination uptake and inventory availability may warrant a dynamic policy for site selection. Second, operational considerations, including the costs of opening new locations and varying capacities across sites, are not incorporated into the baseline model. Lastly, our analysis does not account for immunizations administered outside of pharmacies nor the age- or occupation-based prioritization implemented early in the vaccination campaign.

4.5. Discussion

Integrating structural estimation and optimization into policy design can uncover and mitigate disparities in healthcare access, extending beyond COVID vaccination to other healthcare services. Our findings highlight the importance of accounting for *heterogeneity* in both demand and supply. First, we show that demand elasticities vary by socioeconomic status, with vulnerable communities being more sensitive to distance (Table 2), a pattern likely applicable to other health services. Other barriers like limited transportation may further restrict access to infrequent care (*e.g.*, annual vaccinations, cancer screenings), while frequent care needs (*e.g.*, dialysis, substance abuse treatment) pose even greater challenges. Ignoring demand heterogeneity can lead to inefficient or inequitable service networks (Table 4 and Appendix D.1). Second, given a spatially heterogeneous supply of service providers, strategically selecting new sites to complement the existing network helps eliminate geographic service gaps (Figure 6). In contrast, when demand is more uniform, simpler policies

[13] Founded in 2017, SnapNurse deployed over 10,000 nurses during COVID, growing revenue from $3 million in 2019 to nearly $90 million in 2020 (Barber 2021).

like a greedy approach may perform well (Appendix F.2). Targeting distance-sensitive underserved populations yields greater benefits than prioritizing high-density areas (as with MaxCover) or low-utilization regions (as with GreedyVax$^{\text{Obs}}$). Finally, embedding demand estimation within an optimization model can adequately account for network spillover effects—something not achievable with demand estimation alone (Figure 8).

To illustrate the generalizability of our approach in a different healthcare setting, consider a policymaker deciding where to expand opioid use disorder (OUD) treatment centers. In 2021, the California Department of Health Care Services (CDHCS) initiated the Behavioral Health Continuum Infrastructure Program (BHCIP) with a budget of $2.2 billion to "construct, acquire, and expand properties" for behavioral health services. BHCIP awarded grants to for-profit and non-profit entities, including inpatient and outpatient treatment facilities for OUD. In one funding round, $518 million was awarded to 45 partners to build 1,292 inpatient beds and serve 130,000 patients (CDHCS 2024). However, the geographic distribution of grants does not match populations or estimated needs. For example, a single facility near Sacramento (population 1.6 million) received funds for 36,000 outpatient treatments per year, while Los Angeles County (population 9.7 million) received funds for only 28,000 treatments, despite having significantly higher opioid prescriptions and overdose deaths (California Department of Public Health 2024). In 2024, California passed Proposition 1, a $6.4 billion behavioral health bond measure, including $4.4 billion for substance abuse treatment facilities, appropriated similarly to BHCIP.

Given the financial stakes and potential health impact of such programs, it is essential for state leadership to allocate funding effectively. This requires examining questions like: *Which communities are expected to face high rates of OUD? What is the existing infrastructure available to treat OUD? How far can individuals travel to receive outpatient treatment?* Our SETO framework can address these questions, relying on publicly available data on treatment facilities, opioid-related hospitalizations, overdose deaths, and treatment rates. A related study combined a predictive epidemic model of OUD with optimization for facility allocation (Luo and Stellato 2024). Our approach complements this study by endogenously linking individuals' decision-making (*i.e.*, whether and where to initiate OUD treatment) with policy decisions (*i.e.*, locations of treatment facilities) to predict outcomes (*i.e.*, OUD treatment completion). Our approach is not only *data-driven* but also *demand-driven*, better matching public funds to the needs of the most vulnerable patients.

5. Conclusions

Expanding access to essential health services is vital for addressing disparities in care, yet few studies rigorously examine policy interventions through prescriptive analytics. This study contributes to the literature by developing a data-driven framework combining structural econometrics and

facility location optimization, offering an interdisciplinary approach to tackling health inequality. Heterogeneity in the demand for and supply of health services can pose significant challenges, but they also present opportunities to measure how utilization depends on several factors including distance and socioeconomics, a necessary component to determine where additional resources are most needed.

Optimizing healthcare delivery networks pose challenges to both the private and public sectors, and our structural estimate-then-optimize approach is applicable across diverse settings. In the private sector, major pharmacy chains (Walgreens, CVS, and Rite Aid) have announced over 1,000 store closures based on low utilization and local market dynamics (Roeloffs 2024). Many health systems pursue cross-market mergers and hospital acquisitions to expand their geographic footprint, affecting access to care and costs, particularly when hospitals are geographically close (Cooper et al. 2019). In the public sector, optimizing healthcare networks is equally critical. For example, the UK's National Health Service (NHS) launched the Community Diagnostic Centres program in 2021 (NHS 2024), establishing over 160 diagnostic sites in shopping centers, universities, and stadiums to alleviate wait times and improve access to care. With a single policymaker overseeing service expansion, our framework can systematically compare potential testing sites based on patient needs, facility proximity, and space availability. By predicting utilization, our approach supports cost-effectiveness analyses and investment decisions, demonstrating its adaptability across both private and public healthcare systems.

We conclude our paper by suggesting avenues for future research. Investigating disparities and distance elasticities across various healthcare domains offers fertile ground for empirical and analytical research in the operations management community. Future work could extend the prescriptive aspect of our framework by relaxing some assumptions (*e.g.*, heterogeneous capacity per facility), specifying alternative objectives (*e.g.*, cost, equity), or considering context-specific constraints (*e.g.*, allowing repeat visits), 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 econometrics models with prescriptive analytics, underscoring their potential to inform and refine policy-making and operational strategies.

Acknowledgments

The authors thank Zhijian Li for his excellent research assistance and Peter Rossi for his valuable insights during the early stages of this work. We are also grateful to the Department Editor, Stefan Scholtes, the Associate Editor, and the three anonymous reviewers for their constructive feedback throughout the review process, which significantly improved the manuscript.

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 (INFORMS).

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.

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.

Bertsimas D, Sim M (2004) The price of robustness. *Operations research* 52(1):35–53.

Beshears J, Choi JJ, Laibson DI, Madrian BC, Reynolds GI (2016) Vaccination rates are associated with functional proximity but not base proximity of vaccination clinics. *Medical Care* 54(6):578–583.

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.

Boutilier JJ, Chan TC (2020) Ambulance emergency response optimization in developing countries. *Operations Research* 68(5):1315–1334.

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 Health Care Services (2024) Behavioral Health Continuum Infrastructure Program. URL https://www.infrastructure.buildingcalhhs.com/, Accessed Feb 2025.

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.

California Department of Public Health (2024) Overdose Surveillance Dashboard. URL https://skylab.cdph.ca.gov/ODdash/, Accessed Feb 2025.

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

Chung TH, Rostami V, Bastani H, Bastani O (2022) Decision-aware learning for optimizing health supply chains. *arXiv preprint arXiv:2211.08507* .

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

Cooper Z, Craig SV, Gaynor M, Van Reenen J (2019) The price ain't right? Hospital prices and health spending on the privately insured. *The Quarterly Journal of Economics* 134(1):51–107.

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.

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.

Elmachtoub AN, Grigas P (2022) Smart "predict, then optimize". *Management Science* 68(1):9–26.

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.

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.

Hwang K, Asif TB, Lee T (2022) Choice-driven location-allocation model for healthcare facility location problem. *Flexible Services and Manufacturing Journal* 34(4):1040–1065.

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.

Luo J, Stellato B (2024) Frontiers in Operations: Equitable Data-Driven Facility Location and Resource Allocation to Fight the Opioid Epidemic. *Manufacturing & Service Operations Management* .

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.

Mazar A, Jaro D, Tomaino G, Carmon Z, Wood W (2023) Distance to vaccine sites is tied to decreased covid-19 vaccine uptake. *PNAS nexus* 2(12):pgad411.

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

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.

National Health Services England (2024) Community diagnostic centres. URL https://www.england.nhs.uk/long-read/community-diagnostic-centres//, Accessed Feb 2025.

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.

Roeloffs MW (2024) Here's Why Drug Stores Are Closing In Minority Neighborhoods: Walgreens, CVS And Rite Aid Shutter More Than 1,000. *Forbes*, URL https://www.forbes.com/sites/maryroeloffs/2024/01/14/heres-why-drug-stores-are-closing-in-minority-neighborhoods-walgreens-cvs-and-rite-aid-shutter-more-than-1000/, Accessed Oct 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 (2022) Overbooked and overlooked: Machine learning and racial bias in medical appointment scheduling. *Manufacturing & Service Operations Management* 24(6):2825–2842.

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.

Tan S, Frazier PI (2022) Regret bounds and experimental design for estimate-then-optimize. *arXiv preprint arXiv:2210.15576* .

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) American Community Survey, 2015 – 2019. 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.

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.

Online Appendix

Closer to Home: A Structural 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

Appendix A: Case Study Data and Performance

Table A1 California summary statistics

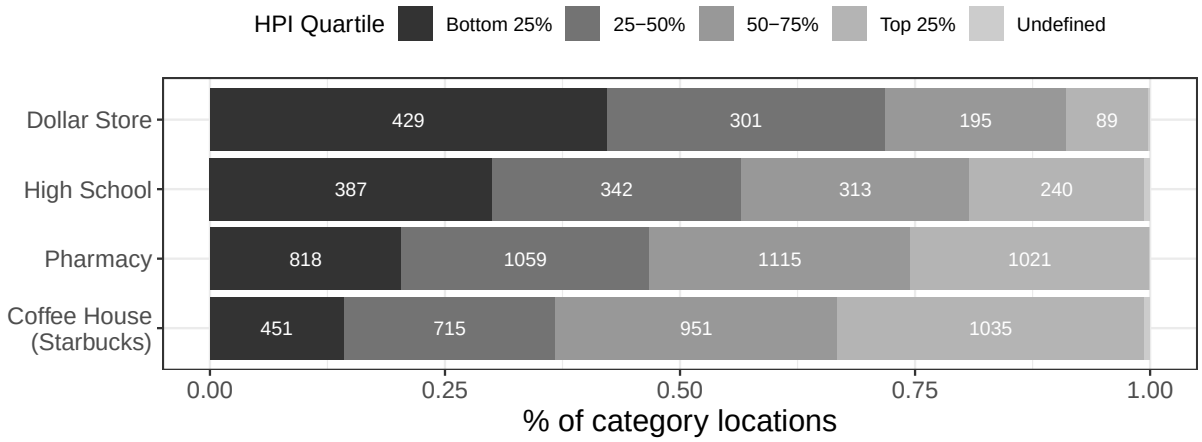
	Mean	SD
Number of retail pharmacy vaccination sites	4,035	-
Number of dollar stores	1,016	-
State population	37,336,716	-
Number of zip-codes (non-zero pop.)	1,659	-
Population (per zip-code)	23,637	22,629
Population density (residents per sq mi)	3,588	5,665
Race/ethnicity		
% White	0.55	0.28
% Black	0.04	0.07
% Asian	0.10	0.13
% Hispanic	0.31	0.25
% Other	<0.01	0.01
Health insurance		
% Employer	0.42	0.16
% Medicare	0.15	0.10
% Medicaid	0.20	0.14
% Other	0.15	0.08
College graduation rate	0.32	0.20
Unemployment rate	0.07	0.06
Poverty level	0.10	0.10
Median household income ($000s)	73.0	38.7
Median home value ($000s)	522.9	392.7
HPI (continuous value in $[0, 1]$)	0.49	0.28
Distance to nearest vaccination site (km)		
HPI quartile 4 (most healthy)	3.2	7.5
HPI quartile 3	6.5	12.7
HPI quartile 2	8.8	12.7
HPI quartile 1 (least healthy)	11.8	17.4

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

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.

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)

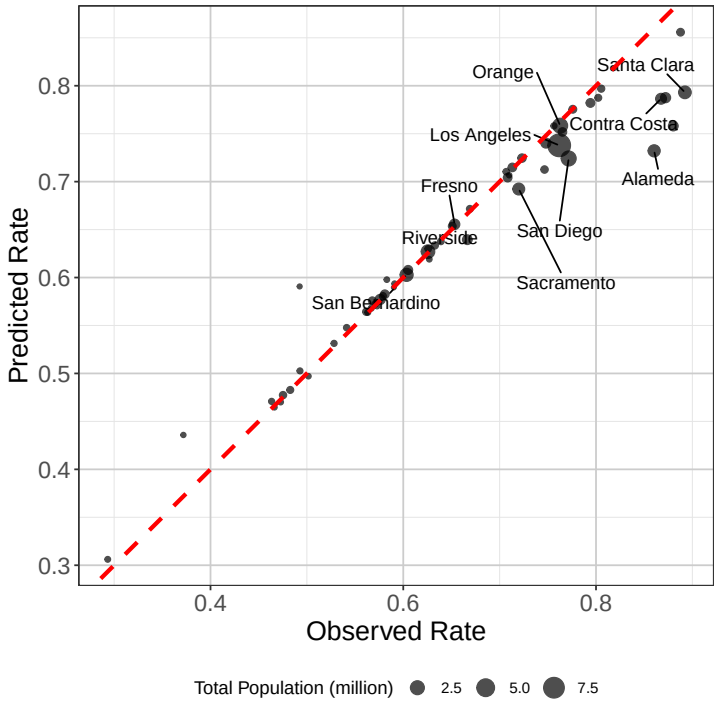
To assess model performance, we compare prediction errors from the Pharmacy-only policy with observed vaccination rates at the zip code-level. Goodness-of-fit measures (Table A4) show the model performs well, with a mean squared error (MSE) of 0.006 and $R^2 = 0.991$, consistent across HPIs. A good performance is expected, given that under the BLP estimation, we perfectly rationalize the observed demand rates. Figure A2 illustrates the model's performance at the county level, where it slightly underpredicts vaccination rates in a few highly vaccinated counties. This occurs because, in the policy evaluation step, we impose strict capacity constraints, which attenuate predicted vaccinations in some densely populated areas.

Table A4 Population-weighted errors between predicted vs. observed vaccination rates at zip code-level

	MSE	RMSE	MAE	MAPE	R^2
Total	0.006	0.074	0.033	0.043	0.991
HPI quartile 4 (most healthy)	0.009	0.093	0.048	0.053	0.992
HPI quartile 3	0.004	0.064	0.025	0.032	0.987
HPI quartile 2	0.003	0.058	0.021	0.026	0.990
HPI quartile 1 (least healthy)	0.006	0.080	0.040	0.062	0.986

Note. MSE (Mean Squared Error) $= \frac{1}{|\mathcal{A}|}\sum_a \left(\rho_a^{obs} - \hat{\rho}_a\right)^2$
RMSE (Root Mean Squared Error) $= \sqrt{\text{MSE}}$
MAE (Mean Absolute Error) $= \frac{1}{|\mathcal{A}|}\sum_a \left|\rho_a^{obs} - \hat{\rho}_a\right|$
MAPE (Mean Absolute Percentage Error) $= \frac{1}{|\mathcal{A}|}\sum_a \left|\frac{\rho_a^{obs} - \hat{\rho}_a}{\rho_a^{obs}}\right| \times 100$

Figure A2 Predicted vs. observed vaccination rates by county in California



Note. The 10 counties with the largest populations are labeled in the plot.

Appendix B: Implementation details

B.1. Optimizing the GMM objective function

Optimizing the GMM objective in equation (6) is challenging because the distance parameters $\boldsymbol{\alpha}$ influence both individuals' location probabilities and vaccination preferences.

We propose an iterative approach leveraging the fact that standard solvers can compute optimal demand parameters $\hat{\boldsymbol{\theta}}$ for a fixed location distribution $f(\ell_i \mid s_i)$. The method alternates between calculating demand parameters under a fixed location distribution and updating the distribution based on the current demand parameters. At convergence, this yields demand parameter estimates—including distance sensitivities $\boldsymbol{\alpha}$—that locally minimize the GMM objective for the corresponding location distribution (implied by $\boldsymbol{\alpha}$).

Let $\tilde{f}^{(k)}(\ell_i \mid s_i)$ represent the location probability distribution in step k of the algorithm. Initially, we assume no capacity constraints and that all individuals prefer their nearest location,[14] yielding the initial location distribution. With the location distribution fixed, we optimize the GMM objective function to update the parameters $\hat{\boldsymbol{\theta}}^{(k)}$, including new distance sensitivities $\hat{\boldsymbol{\alpha}}^{(k)}$. This optimization uses the fixed-point method from Berry et al. (1995) to minimize the following GMM objective:

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

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 PyBLP package by Conlon and Gortmaker (2020). The outputs $\hat{\boldsymbol{\theta}}^{(k)}$, which include the distance sensitivities $\hat{\boldsymbol{\alpha}}^{(k)}$, are then used to update the location 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 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 iterative approach converges to a $\hat{\boldsymbol{\theta}}$ that approximately satisfies the local optimality conditions for equation (6).[15] An algorithmic description of the approach is provided in Appendix B.3.

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

[15] Note that all the fixed-point methods 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. Algorithmic description of SETO approach

Experiments are implemented in Python 3.9.12, and optimization uses Gurobi 10.0.1. The numerical study is carried out on a Linux server with a 64-Core AMD EPYC 7742 processor and 2TB of RAM. We provide a description of the entire approach in Algorithm 1.

Algorithm 1 Estimate-then-Optimize

INITIALIZE:

Create population ordering: $\Omega \leftarrow$ random() ▷ Each i resides in a subarea s_i

procedure ESTIMATE(Ω)

Define individual choice set: $L_i \leftarrow \mathcal{L}\ \forall i \in \Omega$ ▷ It can include specific constraints on the choice set

Initialize location distribution: ▷ Service location chosen at closest site

$$\tilde{f}^{(0)}(\arg\min_{\ell \in L_i} d(s_i,\ell) \mid s_i) \leftarrow 1; \quad \tilde{f}^{(0)}(\ell' \mid s_i) \leftarrow 0\ \forall \ell' \neq \arg\min_{\ell \in L_i} d(s_i,\ell)$$

Define converge parameter: $W \leftarrow \infty$

while $W > \epsilon$ **do**

Increment step count: $k \leftarrow k+1$.

Construct instruments: $\tilde{\boldsymbol{Z}}_a^{(k)} \leftarrow \boldsymbol{Z}_a\left(\tilde{f}^{(k)}(\ell_i \mid s_i)\right)\ \forall a \in \mathcal{A}$

Obtain parameter estimates $\hat{\boldsymbol{\theta}}^{(k)} = [\hat{\boldsymbol{\alpha}}^{(k)}, \hat{\boldsymbol{\beta}}^{(k)}]$ and $\hat{\boldsymbol{\xi}}^{(k)}$, using the nested fixed point algorithm proposed by Berry et al. (1995), namely,

$$\hat{\boldsymbol{\theta}}^{(k)},\ \hat{\boldsymbol{\xi}}^{(k)} \leftarrow \arg\min_{\boldsymbol{\theta},\boldsymbol{\xi}} \left(\frac{1}{|\mathcal{A}|}\sum_a \tilde{\boldsymbol{Z}}_a^{(k)}\xi_a\right)' \Phi \left(\frac{1}{|\mathcal{A}|}\sum_a \tilde{\boldsymbol{Z}}_a^{(k)}\xi_a\right),$$
$$s.t.\ \tilde{\rho}_a^{(k)}(\boldsymbol{\theta}) := \sum_{s_i \in a}\ \Big[\sum_{\ell_i \in L_i} \frac{\exp\left(d(s_i,\ell_i)\alpha_a + X_a\boldsymbol{\beta} + \xi_a\right)}{1+\exp\left(d(s_i,\ell_i)\alpha_a + X_a\boldsymbol{\beta} + \xi_a\right)} \tilde{f}^{(k)}\left(\ell_i \mid s_i\right)\Big] f(s_i \mid a) = \rho_a^{obs} \quad \forall a \in \mathcal{A}$$

Update individual service location: $\{\ell_i\}_{i\in\Omega} \leftarrow$ SEQUENTIAL_LOC($\Omega, \mathcal{L}, \hat{\boldsymbol{\theta}}^{(k)}, \hat{\boldsymbol{\xi}}^{(k)}$) ▷ See function SEQUENTIAL_LOC in procedure COUNTERFACTUAL EVALUATION step below

Update service location distribution:

$$\tilde{f}^{(k+1)}(\ell \mid s) \leftarrow \sum_{i \in s} \mathbb{1}[\ell_i = \ell] / \sum_{i \in s} \mathbb{1}[s_i = s] \quad \forall s \in \mathcal{S},\ \ell \in \mathcal{L}$$

Update converge metric:

$$W \leftarrow \min_a Wasserstein\left(\sum_{s_i}\sum_{\ell} \tilde{f}^{(k+1)}(\ell \mid s_i) f(s_i|a)\, \delta\big(d - d(s_i,\ell)\big), \sum_{s_i}\sum_{\ell} \tilde{f}^{(k)}(\ell \mid s_i) f(s_i|a)\, \delta\big(d - d(s_i,\ell)\big)\right),$$

where $\delta(\cdot)$ is the Dirac delta function for a specific distance d and *Wasserstein* is a measure comparing the area distribution of distances resulting from the sequential service fulfillment process between iterations.

end while

Update coefficients: $\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^{(H)}$ and $\hat{\boldsymbol{\xi}} \leftarrow \hat{\boldsymbol{\xi}}^{(H)}$ ▷ H is the last iteration in the *while* loop

return $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\xi}}$

end procedure

Algorithm 1 Estimate-then-Optimize (Continued)

procedure OPTIMIZE($\mathcal{L}^+$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\xi}}$)

Compute utility: $\overline{u}(s,\ell) \leftarrow d(s,\ell)\hat{\alpha}_a + X_a\hat{\beta}_a + \hat{\xi}_a,\ \forall s \in \mathcal{S},\ \ell \in \mathcal{L}^+,$ and $s \in a$

Solve optimization problem $\tilde{\mathsf{P}}_{\mathsf{loc}}$:

$$
\begin{aligned}
(\tilde{x},\tilde{y},\tilde{b},\tilde{v},\tilde{w}) \leftarrow \underset{\mathbf{x},\mathbf{y},\mathbf{b},\mathbf{v},\mathbf{w}}{\arg\max}\ & \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}^+} m_s y_{s\ell} \\
\text{s.t.}\ & \sum_{s\in\mathcal{S}} m_s y_{s\ell} \le x_\ell K_\ell,\ \forall \ell \in \mathcal{L}^+ \\
& \sum_{\ell\in\mathcal{L}^+} x_\ell \le N \\
& \text{Constraints: (11a)} - \text{(11c)},\ \text{(12a)} - \text{(12c)} \\
& y_{s\ell} \le \frac{\exp(2\bar{u}(s,\ell))}{1+\exp(\bar{u}(s,\ell))} w_{s\ell} \\
& x_\ell, b_s \in \{0,1\},\ v_s, w_{s\ell}, y_{s\ell} \ge 0,\ \forall s \in \mathcal{S},\ \ell \in \mathcal{L}^+
\end{aligned}
$$

Update set of locations: $\tilde{L} \leftarrow \{\ell \in \mathcal{L}^+ | \tilde{x}_\ell = 1\}$

return $\tilde{L}$

end procedure

procedure COUNTERFACTUAL EVALUATION(Ω, $\tilde{L}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\xi}}$)

function SEQUENTIAL_LOC(Ω, $\tilde{L}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\xi}}$)

Initialize total demand served: $D \leftarrow 0$

Initialize remaining capacity: $C_\ell \leftarrow K_\ell,\ \forall \ell \in \tilde{L}$

Initialize set of locations with remaining capacity: $L_C \leftarrow \{\ell \in \tilde{L}\ s.t\ C_\ell > 0\}$

Compute utility: $\overline{u}(s,\ell) \leftarrow d(s,\ell)\hat{\alpha}_a + X_a\hat{\beta}_a + \hat{\xi}_a,\ \forall s \in \mathcal{S},\ \ell \in \mathcal{L}^+,$ and $s \in a$ ▷ It may use aggregation approach in Appendix B.2

for each individual $i \in \Omega$ **do**

Initialize individual choice set: $L_i \leftarrow L_C$ ▷ It can include other constraints on the choice set

Sample individual choice of provider from the service location distribution:

$$
\ell_i \sim \hat{f}(\ell_i \mid s_i) = \frac{\exp\left(d(s_i,\ell_i)\hat{\alpha}_{a_i}\right)}{\sum_{\ell\in L_i}\exp\left(d(s_i,\ell)\hat{\alpha}_{a_i}\right)}\ \mathbb{1}[\ell_i \in L_i]
$$

Sample individual's idiosyncratic error relative to outside option: $\epsilon_0 - \epsilon_i \sim \text{Logistic}(0,1)$

Compute individual's decision to seek service: $b_i \leftarrow \mathbb{1}[\overline{u}(s_i,\ell_i) > \epsilon_0 - \epsilon_i]$

if $b_i = 1$ **then**

$C_{\ell_i} \leftarrow C_{\ell_i} - 1$ and $D \leftarrow D + 1$

Update set of locations with remaining capacity: $L_C \leftarrow \{\ell \in L_C\ s.t\ C_\ell > 0\}$

end if

end for

return D, $\{\ell_i\}_{i\in\Omega}$

end function

end procedure

Appendix C: Robustness Checks

We test the robustness of our estimation and optimization results to changes in parameters and assumptions.

C.1. Choice set

In our primary analysis, we assume that if no available capacity remains at any location in an individual's choice, then she is assigned to the M^{th} closest location. This approximation is necessary because observed vaccination rates in some zip codes exceed the available capacity. A lower value of M may not guarantee that all individuals can obtain a vaccination within one of their M closest locations, which would result in non-convergence of the BLP methodology. For $M = 5$, approximately 3.6% of individuals encounter no remaining capacity at all 5 locations, and are thus assigned to their 5^{th} closest location.

We relax the assumption of $M = 5$ and estimate demand under the MNL choice model with $M \in \{10,\ 300,\ 1{,}000\}$. Table C11 shows that demand estimates are not sensitive to the choice set size. For instance, with $M = 10$, the distance sensitivity coefficient for HPI quartile 1 is -0.158 ($p < 0.001$) compared to -0.161 ($p < 0.001$) when $M = 5$. Even with $M = 1{,}000$, which guarantees that everyone has some location in their choice set with available capacity, the coefficient for HPI quartile 1 only modestly changes to -0.180 ($p < 0.001$). Under the MNL choice model, some individuals will choose distance sites, and a lower distance coefficient will nudge them to closer sites to match observed vaccination rates.

For computational tractability, we re-run the optimization model with $M = 10$ (Table C5). With a larger choice set, the marginal benefit of adding dollar stores decreases as individuals can access more pharmacy capacity nearby. The choice set size differentially affects HPI groups, and areas with more pharmacies see minimal impact on travel distances. High HPI areas maintain similar vaccination gains (0.16 million vs. 0.19 million with $M = 5$), while low HPI areas show lower vaccination gains (0.19 million vs. 0.30 million with $M = 5$).

C.2. Capacity

We vary the per-site vaccination capacity $K \in \{8{,}000,\ 12{,}000\}$ from our baseline of $K = 10{,}000$. Similar to M, these adjustments affect the probability of choosing a nearby vaccination site. Table C12 shows that a 20% variation in capacity results in modest changes in demand sensitivities. If $K = 8{,}000$, individuals must consider farther sites if the closest options reach capacity. Under this scenario, the estimated distance sensitivities are flatter, suggesting a greater willingness to travel farther for vaccination. The distance coefficient for HPI quartile 1 is -0.150 ($p < 0.001$) compared to -0.161 ($p < 0.001$) when $K = 10{,}000$. Conversely, with $K = 12{,}000$, individuals are more likely to secure a spot at their nearest site, aligning with their preferred choice under the MNL model. In this case, the distance coefficient for HPI quartile 1 is -0.170 ($p < 0.001$), implying a slightly steeper disutility of traveling farther for vaccination. Altogether, these results suggest that correct site-level capacities could modestly improve distance sensitivity estimates, but the slight differences in estimated coefficients imply that lacking such data does not seriously invalidate our estimates.

Varying capacity by 20% does meaningfully affect our counterfactual results, as shown in Table C5. Our status quo Pharmacy-only policy with $K = 10{,}000$ and $N = 4{,}035$ pharmacies amounts to 40.35 million vaccination slots, and achieves 26.44 million predicted vaccinations. Adding 500 dollar stores statewide predicts

a gain of nearly 0.8 million vaccinations. With $K = 8,000$ and the same number of dollar stores, predicted vaccination gains exceed 1.3 million. Although there is less aggregate supply, vaccination rates increase across all HPIs—especially at more distant locations—due to individuals' reduced distance sensitivity, as discussed above. With $K = 12,000$, dollar stores provide only 0.38 million additional vaccinations, because more individuals can secure vaccination spots at nearby pharmacies.

We further validate our observation of decreasing marginal returns as new service locations are added. Figure C3 shows the robustness of this insight for various combinations of M and K. Settings with limited

Table C5 Added vaccinations (million) relative to Pharmacy-only, assuming 500 dollar stores are added statewide

				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	0.19
		Walkable	1.54	0.73	0.29	0.29	0.24
	$M = 10$	Total	0.59	0.19	0.10	0.14	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	1.89	0.81	0.43	0.35	0.30
	$K = 12,000$	Total	0.38	0.17	0.07	0.04	0.10
		Walkable	1.19	0.60	0.22	0.20	0.17
Distance	$\bar{d} = 0.5$	Total	0.72	0.26	0.13	0.14	0.19
		Walkable	1.40	0.63	0.26	0.27	0.25
	$\bar{d} = 1.0$	Total	0.70	0.25	0.12	0.14	0.18
		Walkable	1.37	0.59	0.26	0.27	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.

Figure C3 Added vaccinations under varying M and K with network expansion of dollar stores

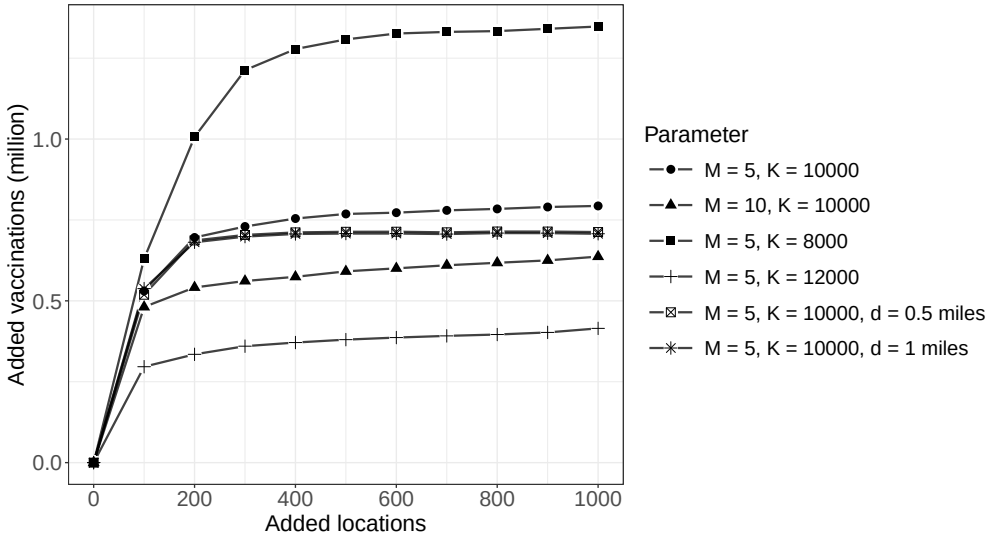


Table C6 Sensitivity analysis of dollar store capacity on added vaccinations (million) with network expansion of 500 dollar stores statewide

			HPI quartile			
Dollar store capacity		All	Bottom 25%	25-50%	50-75%	Top 25%
$K = 10,000$(†)	Total	0.77	0.30	0.14	0.14	0.19
	Walkable	1.54	0.73	0.29	0.29	0.24
$K = 5,000$	Total	0.41	0.17	0.08	0.07	0.09
	Walkable	0.99	0.53	0.18	0.18	0.11

Note. (†) denotes our main scenario. All values are reported in millions of vaccinations. Added vaccinations are relative to Pharmacy-only. Walkable vaccinations refer to travel distances of less than one mile.

capacity ($K = 8,000$) show larger marginal effects and benefit more from adding new locations. These insights are relevant for deciding whether to add capacity to existing pharmacies or allocate it among new partner locations. Our analysis suggests the latter may be more effective.

The mathematical formulation of our demand estimation and optimization model can accommodate *heterogeneous* store capacities. Due to a lack of store-level vaccination data in our case study, we use sensitivity analyses to examine the impact of heterogeneous capacity. We consider a conservative case where dollar stores operate at 50% of the capacity of established pharmacies. Under this scenario, adding 500 dollar stores increases predicted vaccinations by 410,000, 47% less than in our main analysis (Table C6), with gains distributed similarly across HPI quartiles. The selected dollar stores are located farther from their nearest pharmacy (average distance of 3.1 km vs. 2.8 km with full capacity), and fewer dollar stores are selected within the top HPI quartile (40 vs. 56 stores with full capacity). With reduced capacity in each dollar store, the model suggests opening more stores in remote locations, particularly in lower HPI quartiles (Table C7).

Table C7 Comparison of selected vs. not selected locations with network expansion of 500 dollar stores statewide, each with 50% lower capacity than pharmacies

		Selected	
	All dollar stores	Full capacity(†)	Half capacity
No. of stores (%)	1,016	500	500
in HPI quartile 4 (most healthy)	89 (9%)	56 (11%)	40 (8%)
in HPI quartile 3	195 (19%)	91 (18%)	86 (17%)
in HPI quartile 2	301 (30%)	127 (26%)	130 (26%)
in HPI quartile 1 (least healthy)	431 (42%)	226 (45%)	244 (49%)
No. of stores (%) in pharmacy deserts	71 (7%)	56 (11%)	49 (10%)
Avg. distance to the nearest pharmacy (km)	2.1	2.8	3.1

Note. (†) denotes our main model results.

C.3. Closest location

We test an alternative appointment process instead of the MNL-based choice model. Here, individuals are assigned 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 C11 show the revised demand estimates. Distance sensitivity coefficients attenuate slightly as M increases, as expected. For instance, with $M = 300$, the distance coefficient for HPI quartile 1 is -0.130 ($p < 0.001$). Our findings confirm that distance to a pharmacy negatively correlates with vaccination rates and that individuals from lower HPI quartiles are more sensitive to travel distances.

C.4. Distance forms

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

$$d(s, \ell) = \log\left((d_{s\ell} - \bar{d})^+ + 1\right) \tag{C.9}$$

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

We evaluate two distance thresholds, $\bar{d} \in \{0.5, 1.0\}$ miles, and recalculate demand estimates. In both cases, distance sensitivities increase slightly, to offset the flat utility values below $\bar{d}$ (Table C12). For instance, if $\bar{d} = 1.0$, the coefficient on distance for HPI quartile 1 is -0.165 ($p < 0.001$), as individuals become more sensitive to longer distances above $\bar{d}$. However, the signs and relative magnitudes of distance sensitivities remain unchanged.

An alternative distance function very modestly changes our optimal results (see Table C5). Total predicted added vaccinations with dollar stores assuming $\bar{d} = 1.0$ mile is 0.70 million, vs. 0.77 million in our original analysis. The added walkable vaccinations also decrease (1.37 million vs. 1.54 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.

C.5. Pandemic dynamics

In our main analysis, we use cross-sectional data on fully vaccinated individuals as of March 1, 2022, to estimate distance elasticities. We consider how epidemiological factors, such as pandemic phase or recent spikes in disease spread, might influence our numerical results using two robustness checks. First, we replicate our demand estimation using cross-sectional vaccination data from three additional periods: September 2021, December 2021, and June 2022. These periods capture different phases of the pandemic, during which state and federal guidelines—and likely public attention toward COVID—changed. Second, we include additional controls for local epidemic conditions, such as cumulative COVID cases and deaths within the county, and COVID cases reported in the previous week.[16] These metrics capture the cumulative and recent impact of COVID in a community. Finally, we examine how these alternative demand models affect our counterfactual evaluations, using different cross-sectional data, with and without local pandemic controls.

The purpose of these additional analyses is twofold. By comparing estimates across time periods (with and without controls for disease spread), we infer how distance preferences might have changed over time and

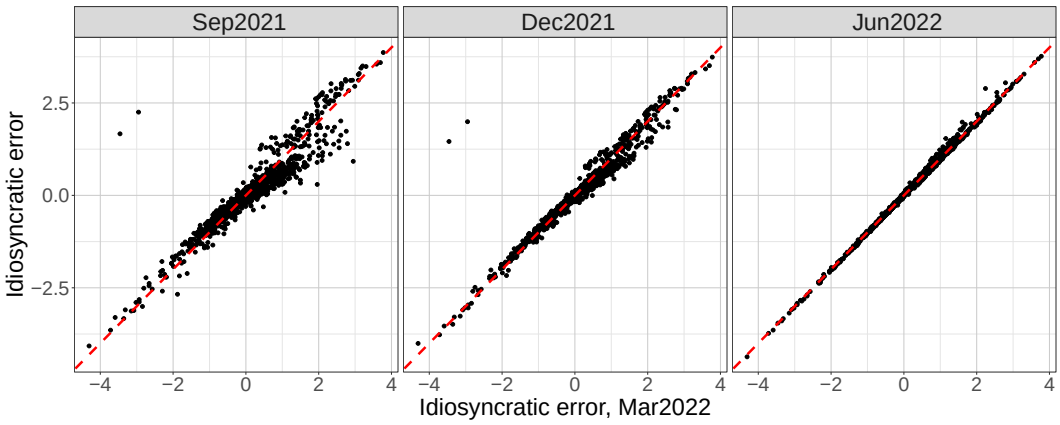
[16] California does not publish statistics on cases and deaths at levels smaller than counties.

with recent outbreaks. Additionally, we produce distinct demand functions for each period to assess whether our policy recommendations from the optimization problem differ by pandemic phase.

Results for these alternative specifications are shown in Table C13. Both the pandemic phase and the recent disease spread meaningfully affect vaccination behaviors. The pandemic phase's impact is evident from the increase in vaccination rates over time, while disease spread's impact is shown by the statistically significant coefficients. Other patterns, such as changes in racial, socioeconomic, and insurance gaps, are also evident. COVID spread is conditionally correlated with vaccination rates.[17] Importantly, the distance sensitivity coefficients (*i.e.*, those on Log-distance × HPI)—the principal inputs to our counterfactual optimization—remain statistically significant, economically meaningful, and relatively consistent across all timeframes, even with the inclusion of disease outbreak controls. This indicates that structural preferences over travel distance remain significant throughout the pandemic phases, regardless of specific epidemic conditions.

As another robustness test, we examine the idiosyncratic errors $\boldsymbol{\xi}$ estimated from our structural demand model. The errors remain stable over time, as evidenced by the high correlation between errors from March 2022 and other time periods (Figure C4). This suggests that unobserved variations in vaccination preferences (*e.g.*, due to hesitancy or misinformation)—beyond what we can capture with HPI and other demographic covariates—do not change meaningfully over time.

Figure C4 Idiosyncratic errors for different estimation periods



Under each of the eight alternative demand models, we run the optimization model assuming our baseline scenario of adding 500 dollar stores statewide to the vaccination network, with results shown in Table C8. The pattern of predicted vaccination gains across HPI quartiles remains consistent, with the bottom quartile gaining the most in all cases. Predicted gains also increase over time, likely reflecting a growing willingness to vaccinate. The gains are slightly larger (*i.e.*, 0.79 million vs. 0.77 million) when including controls for disease spread. While our overall conclusion remains unchanged, the modest difference in predicted gains likely stems from the slight difference in distance sensitivities when including controls for disease spread.

[17] One should interpret these coefficients with caution. Cases and deaths may be correlated with vaccination coverage for a number of reasons. In fact, the direction of causality may even be reversed: the protective effects of vaccines mean that higher vaccination rates in a community likely result in fewer cases and deaths.

Table C8 Predicted added vaccinations (million) under network expansion with 500 additional dollar stores statewide for different cross-sectional data

				HPI quartile			
Pandemic controls	Date		All	Bottom 25%	25-50%	50-75%	Top 25%
No	Sep 2021	Total	0.35	0.13	0.06	0.05	0.11
		Walkable	0.99	0.45	0.19	0.19	0.16
	Dec 2021	Total	0.57	0.22	0.10	0.10	0.15
		Walkable	1.29	0.60	0.25	0.24	0.21
	Mar 2022(†)	Total	0.77	0.30	0.14	0.14	0.19
		Walkable	1.54	0.73	0.29	0.29	0.24
	Jun 2022	Total	0.82	0.33	0.15	0.15	0.19
		Walkable	1.59	0.74	0.31	0.30	0.25
Yes	Sep 2021	Total	0.35	0.13	0.06	0.05	0.11
		Walkable	1.00	0.46	0.19	0.19	0.17
	Dec 2021	Total	0.59	0.23	0.11	0.10	0.15
		Walkable	1.30	0.61	0.24	0.24	0.21
	Mar 2022	Total	0.79	0.31	0.14	0.15	0.19
		Walkable	1.56	0.73	0.30	0.28	0.25
	Jun 2022	Total	0.84	0.33	0.16	0.15	0.20
		Walkable	1.61	0.76	0.32	0.29	0.25

Note. (†) denotes our main scenario. All values are reported in millions of vaccinations. Added vaccinations are relative to Pharmacy-only. Walkable vaccinations refer to travel distances of less than one mile.

C.6. Distance sensitivity by income

Many factors relevant to distance sensitivity are undoubtedly highly correlated. For example, high-income zip codes likely have better local job opportunities, more educated residents, and better transportation access. This motivates our decision to model heterogeneity in distance sensitivity using a simple index, like the Healthy Places Index—which is specifically designed to capture a host of relevant correlates. It is also an index currently used by the relevant state and local regulators—including the California Department of Public Health and counties such as Los Angeles and San Francisco—to allocate public health resources such as vaccine allocations, educational outreach programs, and food resources.

We estimate an alternative demand model where heterogeneity in distance sensitivity is instead measured along a single dimension: *household income*. Results are shown in Table C14. Low-income zip codes have lower vaccination rates, but the relationship between income and distance sensitivity is not straightforward. While the second income quartile is markedly more sensitive than the top quartiles, the bottom quartile is not. This non-monotonicity could result from unobserved outreach efforts in low-income neighborhoods (*e.g.*, pop-up clinics) or indicate that other correlated factors are more important drivers of distance sensitivity than income alone. This underscores the value of allowing heterogeneity along a composite index, such as HPI, which more plausibly captures most of the variation. Results of the counterfactual analysis (Table C9) under this demand specification indicate that our main findings hold: adding dollar stores to the network could improve vaccination coverage considerably. However, the gains are less concentrated in the bottom income quartile compared to the bottom HPI quartile.

Table C9 Predicted vaccinations (million) by median income, under network expansion with 500 additional dollar stores statewide

			Median income quartile			
Strategy		All	Bottom 25%	25-50%	50-75%	Top 25%
Pharmacy-only	Total	26.43	2.78	6.41	8.34	8.91
	Walkable	19.47	2.02	4.79	6.22	6.45
Pharmacy + Dollar	Total	+0.77	+0.12	+0.24	+0.16	+0.24
	Walkable	+1.53	+0.29	+0.56	+0.37	+0.31

Note. The Pharmacy + Dollar strategy is the change in vaccinations compared to Pharmacy-only. Walkable vaccinations refer to travel distances of less than one mile.

C.7. Population order sequence

We assume that service requests are fulfilled according to a random ordering of all vaccine-eligible California residents. This intends to capture a random arrival process of service requests, such as during online booking. In our implementation, we fix a single random ordering, for both demand estimation and counterfactual optimization. To evaluate the sensitivity of our main findings, we perform the following analyses.

First, we consider 100 alternative random sequences and estimate demand for each sequence. The distribution of distance sensitivity estimates by HPI quartile is shown in Figure C5. The coefficients do not meaningfully vary, giving us confidence that our conclusions are highly robust to the choice of random sequence. Table C15 gives the demand estimates for the 5th- and 95th-percentiles of the distance sensitivity coefficient for HPI quartile 1—the most relevant for our study—which again shows minimal variation.

Second, we explore the effect of prioritizing some subpopulations, which occurred widely with COVID-19. We consider an order sequence giving individuals age 65 or older priority access to vaccinations before the general population. We generate the order sequence by sampling individuals' ages from the Census-block age distribution. For a given random sequence of the population, we simply re-order individuals based on their sampled age, giving priority to those 65 and older. We observe no significant differences in demand estimates, shown in Table C15, again confirming that the sequence in which individuals choose their preferred location has minimal impact on the key distance sensitivity estimates.

Third, we evaluate the impact of the ordering sequence on predicted vaccination gains in our counterfactual analysis. To do so, we optimize and predict vaccinations using the demand estimates in Table C15 and the corresponding random sequences. Predicted vaccinations are quite robust to the random order used in the estimation and counterfactuals, with differences of less than 0.01 million, as shown in Table C10.

Figure C5 Distribution of distance sensitivities by HPI for 100 random population order sequences

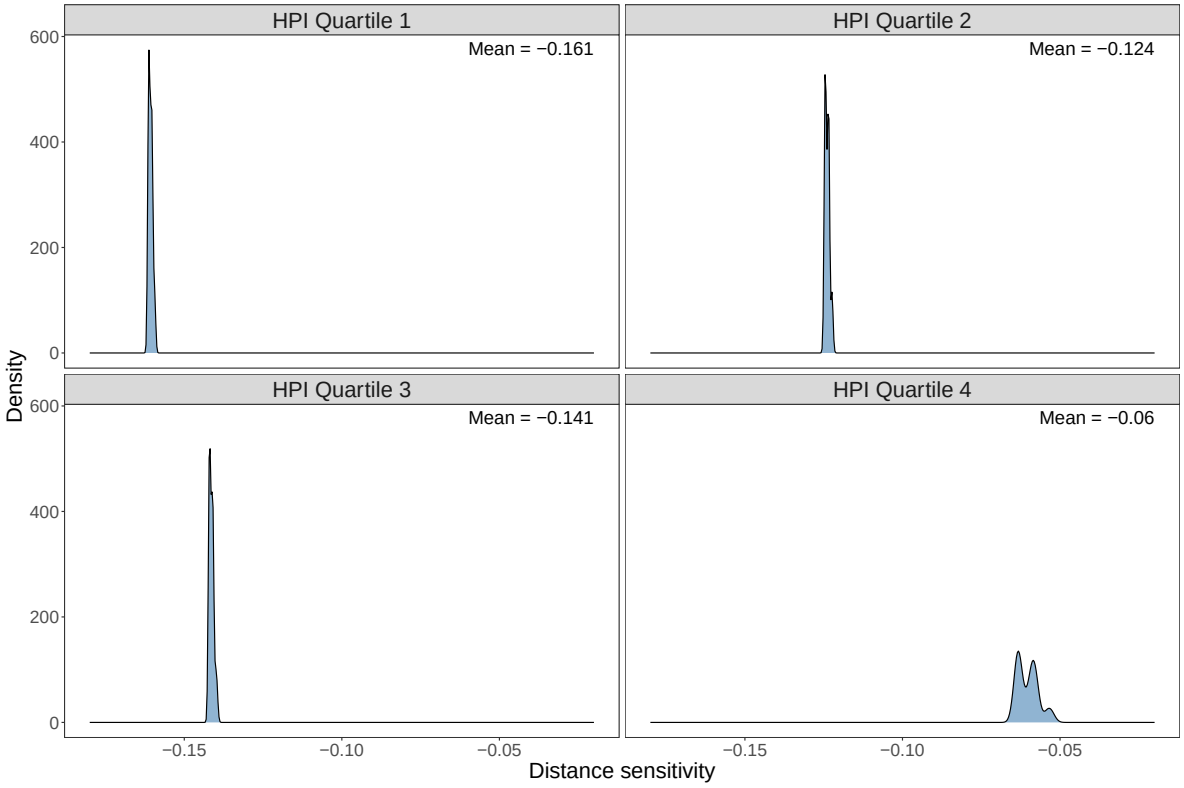


Table C10 Predicted added vaccinations (million) under network expansion with 500 additional dollar stores statewide for different population order sequences

			HPI quartile			
Population ordering		All	Bottom 25%	25-50%	50-75%	Top 25%
Random(†)	Total	0.77	0.30	0.14	0.14	0.19
	Walkable	1.54	0.73	0.29	0.29	0.24
HPI Quartile 1 (5th-percentile)	Total	0.77	0.30	0.14	0.14	0.19
	Walkable	1.54	0.73	0.29	0.29	0.24
HPI Quartile 1 (95th-percentile)	Total	0.77	0.30	0.14	0.14	0.18
	Walkable	1.54	0.72	0.30	0.29	0.24
Priority 65+	Total	0.77	0.30	0.14	0.14	0.18
	Walkable	1.54	0.73	0.29	0.28	0.24

Note: (†) denotes our main scenario. HPI Quartile 1 (x-percentile) gives the predicted vaccinations under the 5th and 95th percentiles of the demand estimate for HPI quartile 1 across 100 random orders.

Table C11 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 C12 Demand estimates under varying per-store vaccination capacity and distance forms

	Capacity				Distance form	
Independent Variable	$K = 10{,}000$ (†)	$K = 8{,}000$	$K = 12{,}000$	$K \to \infty$	$\bar{d} = 0.5$	$\bar{d} = 1$
HPI Quartile 4 (most healthy)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HPI Quartile 3	-0.304***	-0.299***	-0.306***	-0.307***	-0.275***	-0.285***
	(0.084)	(0.086)	(0.084)	(0.083)	(0.100)	(0.092)
HPI Quartile 2	-0.451***	-0.448***	-0.451***	-0.448***	-0.422***	-0.426***
	(0.112)	(0.113)	(0.111)	(0.110)	(0.126)	(0.119)
HPI Quartile 1 (least healthy)	-0.605***	-0.597***	-0.607***	-0.605***	-0.563***	-0.584***
	(0.149)	(0.151)	(0.149)	(0.148)	(0.164)	(0.158)
Log-distance × HPI Quartile 4	-0.063	-0.045	-0.076	-0.086	-0.078	-0.066
	(0.070)	(0.070)	(0.070)	(0.070)	(0.087)	(0.085)
Log-distance × HPI Quartile 3	-0.142***	-0.128**	-0.152***	-0.159***	-0.157***	-0.148***
	(0.050)	(0.050)	(0.050)	(0.050)	(0.057)	(0.053)
Log-distance × HPI Quartile 2	-0.124***	-0.111**	-0.135***	-0.144***	-0.140***	-0.131***
	(0.043)	(0.043)	(0.044)	(0.044)	(0.049)	(0.046)
Log-distance × HPI Quartile 1	-0.161***	-0.150***	-0.170***	-0.178***	-0.180***	-0.165***
	(0.043)	(0.044)	(0.044)	(0.044)	(0.049)	(0.046)
Race White	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Race Black	-0.052	-0.027	-0.074	-0.105	-0.060	-0.076
	(0.342)	(0.343)	(0.342)	(0.342)	(0.342)	(0.343)
Race Asian	2.001***	2.021***	1.983***	1.966***	2.025***	2.032***
	(0.179)	(0.180)	(0.178)	(0.178)	(0.179)	(0.179)
Race Hispanic	0.987***	0.998***	0.980***	0.973***	0.982***	0.983***
	(0.167)	(0.166)	(0.167)	(0.167)	(0.167)	(0.166)
Race Other	7.077**	7.136**	7.036**	7.022**	7.404**	7.415**
	(3.049)	(3.045)	(3.049)	(3.052)	(3.112)	(3.128)
Health Insurance: Employer	-1.202*	-1.195*	-1.207*	-1.210*	-1.259*	-1.274*
	(0.696)	(0.697)	(0.695)	(0.695)	(0.695)	(0.698)
Health Insurance: Medicare	0.197	0.189	0.200	0.203	0.148	0.135
	(0.739)	(0.740)	(0.738)	(0.737)	(0.735)	(0.737)
Health Insurance: Medicaid	-0.973	-0.976	-0.973	-0.974	-1.029	-1.039
	(0.729)	(0.730)	(0.728)	(0.728)	(0.726)	(0.727)
Health Insurance: Other	-2.319***	-2.322***	-2.318***	-2.316***	-2.372***	-2.388***
	(0.760)	(0.762)	(0.759)	(0.758)	(0.758)	(0.761)
College Graduation Rate	1.727***	1.730***	1.727***	1.730***	1.751***	1.763***
	(0.282)	(0.283)	(0.282)	(0.281)	(0.282)	(0.282)
Unemployment Rate	-1.777***	-1.796***	-1.763**	-1.748**	-1.791***	-1.793***
	(0.686)	(0.687)	(0.685)	(0.684)	(0.687)	(0.686)
Poverty Level	-0.011	-0.011	-0.012	-0.016	-0.003	0.000
	(0.472)	(0.473)	(0.471)	(0.473)	(0.476)	(0.478)
Log(Median Household Income)	-0.110*	-0.112*	-0.109*	-0.108*	-0.113*	-0.115*
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
Log(Median Home Value)	0.114**	0.115**	0.114**	0.113**	0.115**	0.116**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)
Log(Population Density)	-0.078***	-0.073***	-0.082***	-0.085***	-0.077***	-0.075***
	(0.022)	(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

Table C13 Demand estimates for different dates of cross-sectional data with and without pandemic controls

	No pandemic controls				With pandemic controls			
Independent Variable	Sep 2021	Dec 2021	Mar 2022(†)	Jun 2022	Sep 2021	Dec 2021	Mar 2022	Jun 2022
HPI Quartile 4 (most healthy)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HPI Quartile 3	-0.252***	-0.300***	-0.304***	-0.308***	-0.181**	-0.252***	-0.262***	-0.196**
	(0.077)	(0.079)	(0.084)	(0.085)	(0.078)	(0.080)	(0.085)	(0.083)
HPI Quartile 2	-0.403***	-0.438***	-0.451***	-0.460***	-0.289***	-0.364***	-0.367***	-0.256**
	(0.104)	(0.106)	(0.112)	(0.113)	(0.107)	(0.107)	(0.113)	(0.112)
HPI Quartile 1 (least healthy)	-0.576***	-0.614***	-0.605***	-0.613***	-0.436***	-0.518***	-0.489***	-0.383***
	(0.142)	(0.143)	(0.149)	(0.151)	(0.144)	(0.144)	(0.151)	(0.148)
Log-distance × HPI Quartile 4	-0.030	-0.053	-0.063	-0.068	-0.059	-0.100	-0.108	-0.120*
	(0.068)	(0.068)	(0.070)	(0.070)	(0.066)	(0.069)	(0.072)	(0.069)
Log-distance × HPI Quartile 3	-0.148***	-0.148***	-0.142***	-0.145***	-0.163***	-0.187***	-0.160***	-0.169***
	(0.048)	(0.048)	(0.050)	(0.050)	(0.047)	(0.048)	(0.050)	(0.050)
Log-distance × HPI Quartile 2	-0.140***	-0.135***	-0.124***	-0.124***	-0.148***	-0.158***	-0.145***	-0.161***
	(0.042)	(0.042)	(0.043)	(0.044)	(0.041)	(0.042)	(0.044)	(0.043)
Log-distance × HPI Quartile 1	-0.169***	-0.170***	-0.161***	-0.161***	-0.183***	-0.193***	-0.181***	-0.183***
	(0.041)	(0.042)	(0.043)	(0.044)	(0.040)	(0.041)	(0.043)	(0.042)
Race White	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Race Black	-0.409	-0.089	-0.052	0.004	-0.484	0.061	-0.068	-0.461
	(0.322)	(0.326)	(0.342)	(0.346)	(0.334)	(0.328)	(0.350)	(0.362)
Race Asian	1.266***	1.478***	2.001***	2.070***	1.104***	1.500***	1.987***	1.616***
	(0.167)	(0.170)	(0.179)	(0.180)	(0.175)	(0.176)	(0.180)	(0.190)
Race Hispanic	0.637***	0.779***	0.987***	1.038***	0.593***	0.880***	1.011***	0.900***
	(0.155)	(0.159)	(0.167)	(0.169)	(0.174)	(0.162)	(0.178)	(0.177)
Race Other	8.270**	8.019**	7.077**	6.943**	8.109***	9.085***	6.009*	3.759
	(3.436)	(3.243)	(3.049)	(3.020)	(2.968)	(2.760)	(3.146)	(2.474)
Health Insurance: Employer	-0.915	-1.165*	-1.202*	-1.223*	-0.778	-1.067	-1.178*	-1.356**
	(0.652)	(0.686)	(0.696)	(0.694)	(0.662)	(0.694)	(0.703)	(0.691)
Health Insurance: Medicare	0.665	0.296	0.197	0.197	0.753	0.409	0.288	0.190
	(0.693)	(0.727)	(0.739)	(0.736)	(0.697)	(0.728)	(0.733)	(0.721)
Health Insurance: Medicaid	-0.723	-0.954	-0.973	-0.955	-0.631	-1.038	-0.948	-0.953
	(0.684)	(0.713)	(0.729)	(0.728)	(0.705)	(0.715)	(0.746)	(0.724)
Health Insurance: Other	-1.819**	-2.187***	-2.319***	-2.343***	-1.739**	-2.264***	-2.391***	-2.403***
	(0.716)	(0.749)	(0.760)	(0.760)	(0.731)	(0.758)	(0.770)	(0.767)
College Graduation Rate	1.595***	1.630***	1.727***	1.733***	1.493***	1.589***	1.741***	1.606***
	(0.279)	(0.278)	(0.282)	(0.283)	(0.281)	(0.277)	(0.283)	(0.280)
Unemployment Rate	-1.940***	-1.918***	-1.777***	-1.742**	-1.807***	-1.802***	-1.585**	-1.497**
	(0.663)	(0.665)	(0.686)	(0.695)	(0.651)	(0.661)	(0.689)	(0.669)
Poverty Level	-0.105	-0.065	-0.011	-0.019	0.041	0.130	-0.005	0.087
	(0.471)	(0.471)	(0.472)	(0.473)	(0.474)	(0.478)	(0.480)	(0.471)
Log(Median Household Income)	-0.170***	-0.145**	-0.110*	-0.106*	-0.149**	-0.130**	-0.102*	-0.071
	(0.058)	(0.059)	(0.060)	(0.060)	(0.059)	(0.058)	(0.061)	(0.061)
Log(Median Home Value)	0.076	0.083*	0.114**	0.119**	0.052	0.067	0.105**	0.094*
	(0.048)	(0.049)	(0.050)	(0.050)	(0.049)	(0.049)	(0.050)	(0.049)
Log(Population Density)	-0.070***	-0.074***	-0.078***	-0.077***	-0.079***	-0.079***	-0.084***	-0.105***
	(0.021)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)	(0.022)	(0.022)
Cumulative Cases per 1000					-0.003**	-0.005***	0.002	-0.002*
					(0.001)	(0.001)	(0.001)	(0.001)
Cumulative Deaths per 1000					0.025	-0.022	-0.200***	-0.012
					(0.063)	(0.056)	(0.058)	(0.054)
Previous Week Cases per 1000					-0.107***	0.305***	-0.024	0.368***
					(0.033)	(0.079)	(0.093)	(0.049)

Note. (†) denotes our main specification. Standard errors are in parentheses. Significance levels: *p<0.05; **p<0.01; ***p<0.001.

Table C14 Demand estimates with heterogeneity in distance sensitivity by median income

	Quartile variable	
Independent Variable	HPI (†)	Income
Quartile 4	Refs.	Refs.
Quartile 3	-0.304***	-0.255***
	(0.084)	(0.068)
Quartile 2	-0.451***	-0.389***
	(0.112)	(0.112)
Quartile 1	-0.605***	-0.634***
	(0.149)	(0.160)
Log-distance × Quartile 4	-0.063	-0.098
	(0.070)	(0.066)
Log-distance × Quartile 3	-0.142***	-0.127**
	(0.050)	(0.050)
Log-distance × Quartile 2	-0.124***	-0.200***
	(0.043)	(0.044)
Log-distance × Quartile 1	-0.161***	-0.114**
	(0.043)	(0.048)
Race White	Refs.	Refs.
Race Black	-0.052	-0.099
	(0.342)	(0.334)
Race Asian	2.001***	1.886***
	(0.179)	(0.184)
Race Hispanic	0.987***	0.859***
	(0.167)	(0.167)
Race Other	7.077**	6.014*
	(3.049)	(3.536)
Health Insurance: Employer	-1.202*	-1.147
	(0.696)	(0.718)
Health Insurance: Medicare	0.197	0.275
	(0.739)	(0.750)
Health Insurance: Medicaid	-0.973	-0.741
	(0.729)	(0.741)
Health Insurance: Other	-2.319***	-2.307***
	(0.760)	(0.770)
College Graduation Rate	1.727***	1.990***
	(0.282)	(0.258)
Unemployment Rate	-1.777***	-1.635**
	(0.686)	(0.683)
Poverty Level	-0.011	-0.065
	(0.472)	(0.487)
Log(Median Household Income)	-0.110*	-0.209***
	(0.060)	(0.069)
Log(Median Home Value)	0.114**	0.133***
	(0.050)	(0.049)
Log(Population Density)	-0.078***	-0.078***
	(0.022)	(0.022)

Note. (†) denotes our main specification. Standard errors are in parentheses. Significance levels: *p<0.05; **p<0.01; ***p<0.001.

Table C15 Demand estimates for different population orderings

		Distance Sensitivity HPI quartile 1		
Independent Variable	Main (†)	5th-percentile	95th-percentile	Priority 65+
HPI Quartile 4 (most healthy)	Ref.	Ref.	Ref.	Ref.
HPI Quartile 3	-0.304***	-0.304***	-0.302***	-0.302***
	(0.084)	(0.084)	(0.084)	(0.084)
HPI Quartile 2	-0.451***	-0.451***	-0.449***	-0.449***
	(0.112)	(0.112)	(0.111)	(0.111)
HPI Quartile 1 (least healthy)	-0.605***	-0.604***	-0.602***	-0.602***
	(0.149)	(0.149)	(0.149)	(0.149)
Log-distance × HPI Quartile 4	-0.063	-0.063	-0.054	-0.059
	(0.070)	(0.070)	(0.070)	(0.070)
Log-distance × HPI Quartile 3	-0.142***	-0.142***	-0.141***	-0.142***
	(0.050)	(0.050)	(0.050)	(0.050)
Log-distance × HPI Quartile 2	-0.124***	-0.125***	-0.123***	-0.124***
	(0.043)	(0.044)	(0.043)	(0.043)
Log-distance × HPI Quartile 1	-0.161***	-0.161***	-0.160***	-0.161***
	(0.043)	(0.043)	(0.043)	(0.043)
Race White	Ref.	Ref.	Ref.	Ref.
Race Black	-0.052	-0.052	-0.053	-0.054
	(0.342)	(0.342)	(0.342)	(0.342)
Race Asian	2.001***	2.001***	2.002***	2.002***
	(0.179)	(0.179)	(0.179)	(0.179)
Race Hispanic	0.987***	0.987***	0.987***	0.986***
	(0.167)	(0.167)	(0.167)	(0.167)
Race Other	7.077**	7.078**	7.070**	7.067**
	(3.049)	(3.048)	(3.046)	(3.046)
Health Insurance: Employer	-1.202*	-1.201*	-1.200*	-1.200*
	(0.696)	(0.696)	(0.696)	(0.696)
Health Insurance: Medicare	0.197	0.197	0.196	0.196
	(0.739)	(0.739)	(0.739)	(0.739)
Health Insurance: Medicaid	-0.973	-0.973	-0.972	-0.972
	(0.729)	(0.729)	(0.729)	(0.729)
Health Insurance: Other	-2.319***	-2.319***	-2.319***	-2.319***
	(0.760)	(0.760)	(0.760)	(0.760)
College Graduation Rate	1.727***	1.727***	1.728***	1.729***
	(0.282)	(0.282)	(0.282)	(0.282)
Unemployment Rate	-1.777***	-1.777***	-1.776***	-1.776***
	(0.686)	(0.686)	(0.686)	(0.686)
Poverty Level	-0.011	-0.012	-0.013	-0.012
	(0.472)	(0.472)	(0.472)	(0.472)
Log(Median Household Income)	-0.110*	-0.110*	-0.110*	-0.110*
	(0.060)	(0.060)	(0.060)	(0.060)
Log(Median Home Value)	0.114**	0.114**	0.114**	0.114**
	(0.050)	(0.050)	(0.050)	(0.050)
Log(Population Density)	-0.078***	-0.078***	-0.078***	-0.078***
	(0.022)	(0.022)	(0.022)	(0.022)

Note. (†) denotes our main specification. Standard errors are in parentheses. Significance levels: *p<0.05; **p<0.01; ***p<0.001.

Table C16 Demand estimates with and without heterogeneity by HPI quartile

	HPI		
Independent Variable	Heterogeneous(†)	Homogeneous (with distance)	Homogeneous (without distance)
HPI Quartile 4 (most healthy)	Refs.		
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		-0.144***	
		(0.037)	
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			
Race Black	-0.052	-0.225	-0.176
	(0.342)	(0.329)	(0.350)
Race Asian	2.001***	1.988***	2.076***
	(0.179)	(0.186)	(0.204)
Race Hispanic	0.987***	0.884***	0.958***
	(0.167)	(0.169)	(0.136)
Race Other	7.077**	8.234**	8.703**
	(3.049)	(3.295)	(3.261)
Health Insurance: Employer	-1.202*	-0.777	-0.696
	(0.696)	(0.694)	(0.379)
Health Insurance: Medicare	0.197	0.376	0.338
	(0.739)	(0.747)	(0.408)
Health Insurance: Medicaid	-0.973	-0.906	-0.903*
	(0.729)	(0.752)	(0.421)
Health Insurance: Other	-2.319***	-2.246***	-2.240***
	(0.760)	(0.768)	(0.434)
College Graduation Rate	1.727***	2.225***	2.265***
	(0.282)	(0.239)	(0.174)
Unemployment Rate	-1.777***	-1.816***	-1.972***
	(0.686)	(0.704)	(0.419)
Poverty Level	-0.011	-0.210	-0.240
	(0.472)	(0.484)	(0.266)
Log(Median Household Income)	-0.110*	-0.116*	-0.130***
	(0.060)	(0.061)	(0.033)
Log(Median Home Value)	0.114**	0.138***	0.146***
	(0.050)	(0.050)	(0.029)
Log(Population Density)	-0.078***	-0.081***	-0.028*
	(0.022)	(0.022)	(0.012)

Note. (†) denotes our main specification. Standard errors are in parentheses. Significance levels: *p<0.05; **p<0.01; ***p<0.001.

Appendix D: Demand Model Misspecifications

We analyze how demand model misspecification affects the optimal selection of stores and predicted vaccinations across two scenarios: assuming homogeneous demand estimates, and assuming providers have unlimited capacity. Under each scenario, we re-run the optimization model to optimally select 500 dollar stores statewide. To ensure an equivalent comparison, we compute predicted vaccination gains under the original demand model with capacity and heterogeneous distance sensitivities based on HPI.

D.1. Homogeneous demand

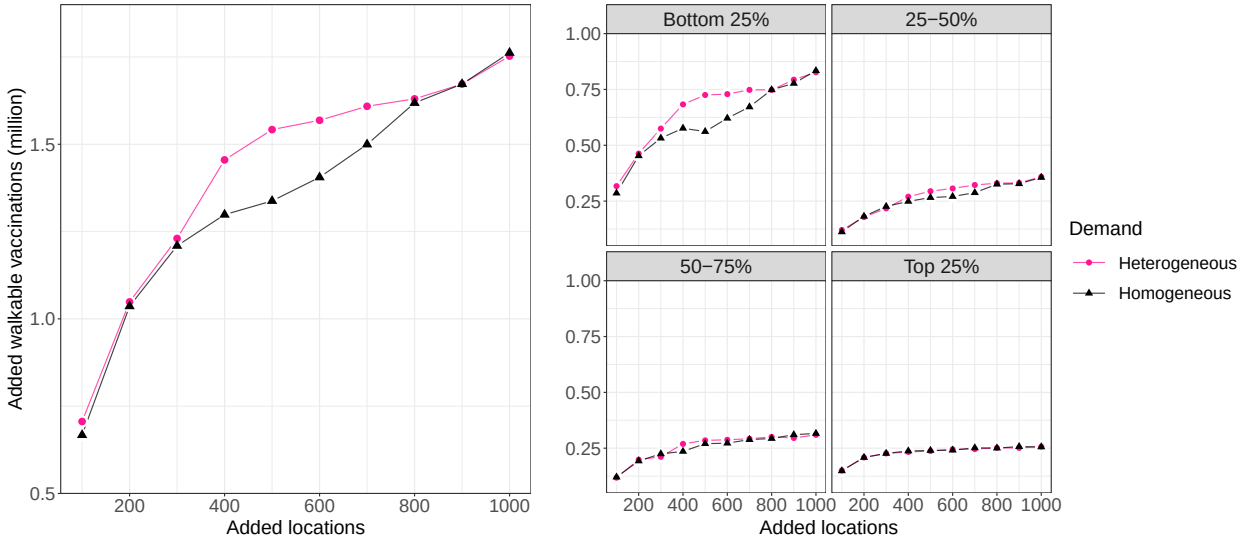
We evaluate two alternative homogeneous demand models. The first model (*homogeneous with distance*) estimates a constant distance elasticity across HPI quartiles. The distance coefficient of -0.144 ($p < 0.001$) is both statistically significant and economically meaningful in magnitude (Table C16). The optimal site selection under this demand model differs slightly from the baseline, with a small resource shift towards high HPI areas. The second model (*homogeneous without distance*) allows for constant preferences across vaccination locations within the choice set. Here, the predicted gain in vaccinations is 30,000 less than with heterogeneous demand, but the impact on low-HPI populations is striking (Table D17). Approximately 200,000 fewer vaccinations are within walking distance, including 190,000 vaccinations among individuals in the bottom two HPI quartiles. As discussed in Section 4.3, a homogeneous demand model results in a more regressive policy that favors high-HPI areas. Figure D6 shows that this resource shift disproportionately affects low-HPI communities, especially for medium-sized network expansions.

Table D17 Predicted added vaccinations (million) under homogeneous demand model, assuming network expansion with 500 additional dollar stores statewide

			HPI quartile			
Demand		All	Bottom 25%	25-50%	50-75%	Top 25%
Heterogeneous(†)	Total	0.77	0.30	0.14	0.14	0.19
	Walkable	1.54	0.73	0.29	0.29	0.24
Homogeneous	Total	0.74	0.28	0.13	0.14	0.19
	Walkable	1.34	0.56	0.27	0.27	0.24

Note. (†) denotes our main model results. Homogeneous results are based on the model with constant preferences across locations within the choice set. Predicted vaccinations are calculated under heterogeneous demand.

Figure D6 Predicted added walkable vaccinations under homogeneous vs. heterogeneous store selection



Note: Walkable vaccinations added relative to Pharmacy-only under heterogeneous demand model. Homogeneous results are based on the model with constant preferences across locations within the choice set.

D.2. Uncapacitated providers

Using the revised demand estimates where vaccination providers have unlimited capacity (Table C12), we re-run the optimization model ignoring capacity constraints to select 500 dollar stores statewide. To ensure an equivalent comparison, we compute predicted vaccinations under the original demand model with capacity constraints. Total added vaccinations decrease by 250,000 and walkable vaccinations decrease by 330,000 when capacity is ignored in both the estimation and optimization stages. Although the model selects stores at similar distances to existing pharmacies as in the baseline, ignoring capacity shifts more resources to lower-HPI communities. Intuitively, network expansion prioritizes proximity for all communities without considering the capacity availabile at nearby stores (Table D18). However, because the network expansion does not add capacity where is most needed and some low-HPI stores are in low-density areas, walkable vaccinations in low-HPI areas drop by approximately 20%, from 0.73 to 0.60 million (Table D19). Deviations in walkable vaccinations are most notable for medium-size network expansions (Figure D7).

Table D18 Characteristics of selected dollar stores under heterogeneous and uncapacitated demand models, assuming network expansion with 500 locations statewide

		Selected dollar stores	
	All dollar stores	Capacitated model(†)	Uncapacitated model
No. of stores (%)	1,016	500	500
in HPI quartile 4 (most healthy)	89 (9%)	56 (11%)	33 (7%)
in HPI quartile 3	195 (19%)	91 (18%)	93 (19%)
in HPI quartile 2	301 (30%)	127 (26%)	152 (31%)
in HPI quartile 1 (least healthy)	431 (42%)	226 (45%)	222 (44%)
No. of stores (%) in pharmacy deserts	71 (7%)	56 (11%)	52 (10%)
Avg. distance to the nearest pharmacy (km)	2.13	2.81	2.76

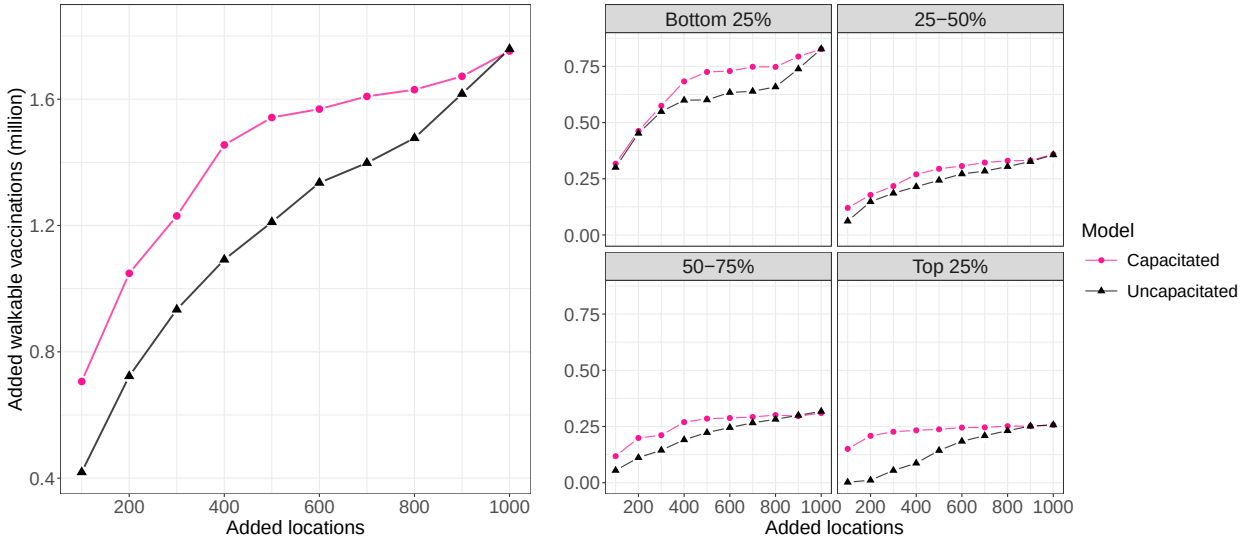
Note. (†) denotes our main model results.

Table D19 Predicted added vaccinations (million) under uncapacitated demand model, under network expansion with 500 additional dollar stores statewide

			HPI quartile			
Model		All	Bottom 25%	25-50%	50-75%	Top 25%
Capacitated(†)	Total	0.77	0.30	0.14	0.14	0.19
	Walkable	1.54	0.73	0.29	0.29	0.24
Uncapacitated	Total	0.52	0.23	0.10	0.10	0.10
	Walkable	1.21	0.60	0.24	0.22	0.14

Note. (†) denotes our main model results. Predicted vaccinations are calculated under the capacitated model.

Figure D7 Predicted added walkable vaccinations under capacity vs. uncapacitated model

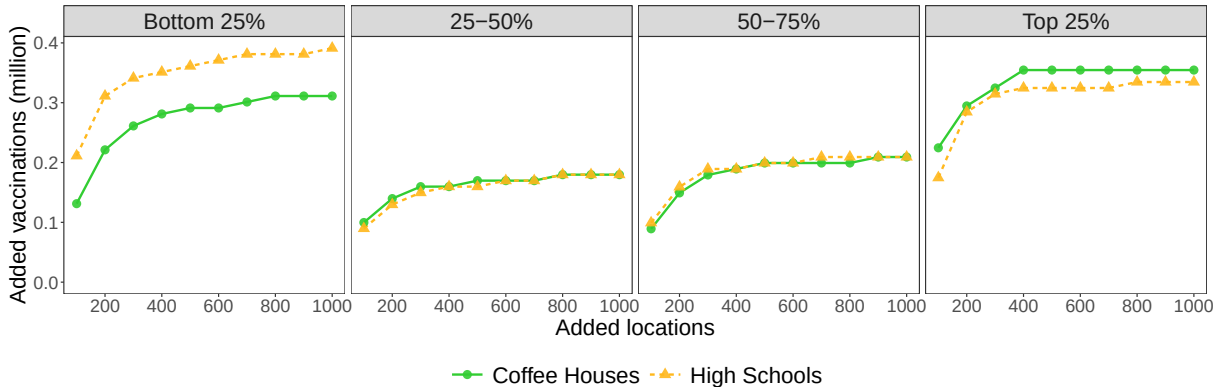


Note: Walkable vaccinations added relative to Pharmacy-only under baseline model with capacity limits.

Appendix E: Other Partnerships

We evaluate a network expansion policy with an alternative private partner, Starbucks coffee houses, or public high schools. Table A2 summarizes the number of partner locations. Figure E8 shows the predicted gain in vaccinations by HPI quartile for each partnership assuming network expansion with $A = \{100, \ldots, 1,000\}$ added locations statewide. Added vaccinations exhibit decreasing marginal gains, similar to those observed with dollar stores. 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 1.06 million additional vaccinations, with 37% occurring in the bottom HPI quartile. Notably, over 50% of these benefits are obtained by strategically adding only 100 high school locations.

Figure E8 Predicted added vaccinations with network expansion of 100 to 1,000 partner locations statewide



We also measure how each partnership could reduce the prevalence of pharmacy deserts (Figure E9). This metric, independent of our vaccination demand estimates, simply measures the population living in a *service desert*.[18] With a network expansion of 500 locations statewide, dollar stores and high schools each reduce the population living in a service desert by over 25%, while Starbucks offers a more modest reduction of 16%.

We qualitatively compare feasibility considerations in Table E20. One key factor is whether a location has adequate space and suitable hours for a pop-up vaccination clinic. High schools offer the most indoor space, while dollar stores also provide ample room. Smaller venues like coffee houses may lack sufficient space to ensure patient privacy. In all cases, mobile clinics could be set up in parking areas. The essential service status of a venue is important, as non-essential businesses and public venues faced closures during the COVID pandemic. Dollar stores, which remained open, could be prudent partners in future pandemics. Additionally, frequent customer visits to these businesses may increase vaccination opportunities. Lastly, the degree of centralization in ownership structure is relevant. Coffee houses and dollar stores would require coordination across multiple entities (*e.g.*, licensing stores). In the public sector, more than 1,200 high schools operate across 421 districts, each with its own leadership and accountability to school boards.

[18] A service desert is defined as an area where the distance to the closest service location (pharmacy, dollar store, high school or Starbucks offering vaccination) exceeds 1 mile in urban areas and 10 miles in rural areas.

Figure E9 **Residents living in service deserts for all three partnerships**

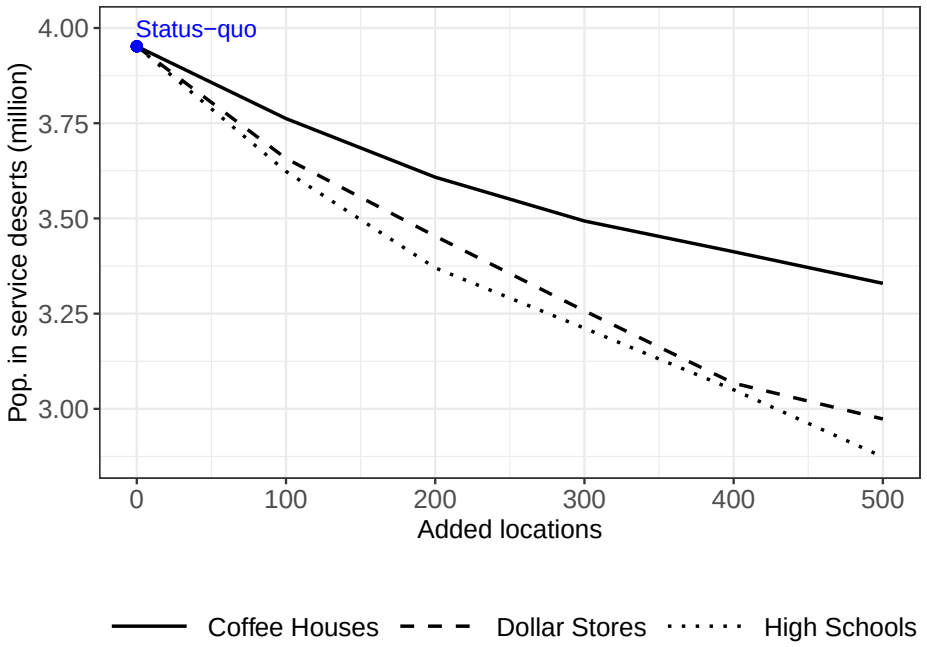


Table E20 **Summary of each partnership's effectiveness and key features**

Category	Added Vaccinations per Location	Adequate Space	Essential Service	Centralized Management	Equity Considerations
Coffee House	2,013	×	×	✓	×
Dollar Store	1,537	✓	✓	✓	✓
High School	2,113	✓	×	×	✓

Note. Added vaccinations per location computed under network expansion with $A = 500$ locations.

Appendix F: Alternative Benchmarks

F.1. Definitions

To test the performance of our approach, we consider the following benchmarks:

1. **Maximum coverage**: An optimization problem where the goal is to select at most N facilities to maximize the number of individuals living within r miles of a facility. We consider subarea level (tracts) as the unit of analysis and formulate the problem as a mixed integer program:

$$(\mathsf{P}_{\mathsf{cover}}) \quad \max_{\mathbf{x},\mathbf{y}} \sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{L}} m_s y_{s\ell} \tag{F.10a}$$

$$\text{s.t.} \sum_{s\in\mathcal{S}} m_s y_{s\ell} \le x_\ell K_\ell, \ \forall \ell \in \mathcal{L} \tag{F.10b}$$

$$\sum_{\ell\in\mathcal{L}} y_{s\ell} \le 1, \ \forall s \in \mathcal{S} \tag{F.10c}$$

$$y_{s\ell} \le a_{s\ell} x_\ell, \ y_{s\ell} \le a_{s\ell}, \ \forall \ell \in \mathcal{L}, s \in \mathcal{S} \tag{F.10d}$$

$$\sum_{\ell\in\mathcal{L}} x_\ell \le N \tag{F.10e}$$

$$0 \le y_{s\ell} \le x_\ell, \ x_\ell \in \{0,1\}, \ \forall \ell \in \mathcal{L}, \ s \in \mathcal{S}. \tag{F.10f}$$

   where $a_{s\ell} = I[d_{s\ell} \le r]$ indicates whether location ℓ is within a distance r to subarea s.
   Decision variable x_ℓ represents whether facility ℓ is chosen, and $y_{s\ell}$ represents the fraction of subarea s population that could be served at location ℓ. Parameter m_s represents the population of subarea s. Constraints (F.10b) ensure that the total demand covered by each location does not exceed its capacity, and constraints (F.10d) ensure that demand from subarea s to location ℓ can only be covered if the subarea is within the maximum allowed radius r. We consider $r \in \{1 \text{ km}, \ 5 \text{ km}\}$, but only report the most competitive model, $r = 1$ km.

2. **GreedyVax$^{\mathbf{Obs}}$**: All areas (zip codes) are ranked in ascending order based on *observed* vaccination rates. Starting with the areas with the lowest vaccination rates, we select the closest facility to each area centroid and continue down the list until we have opened the target number of new facilities. If a location has already been selected as the closest for a different area, we choose the next closest location that has not yet been selected.

3. **GreedyDist**: All areas (zip codes) are ranked in descending order based on distance to the nearest existing pharmacy. Starting with the areas with the longest distance, we select the closest new location to the area centroid and continue down the list until the target number of new facilities has been selected. If a location has already been selected as the closest for a different area, we choose the next closest location that has not yet been selected.

4. **GreedyVax$^{\mathbf{BLP}}$**: All areas (zip codes) are ranked in descending order based on *predicted* vaccinations under the BLP demand model, assuming demand is served at the area's nearest new facility $\tilde{\ell}$. We denote predicted vaccinations by $\hat{\rho}_a(d_{a\tilde{\ell}})m_a$, where $\hat{\rho}_a(d_{a\tilde{\ell}})$ is the predicted population share in area a with travel distance equal to $d_{a\tilde{\ell}}$ and m_a is the eligible population in area a. Starting with the areas with the largest predicted vaccinations, we select the nearest new facility to the area and continue down the list until we have opened the target number of new facilities. If a location has already been selected as the closest for a different area, we choose the next closest location that has not yet been selected.

F.2. Illustration of benchmark limitations

We compare the performance of our SETO model against a greedy policy (GreedyVax$^{\text{Obs}}$) that does not incorporate demand estimates for three example counties in California: Kern, Los Angeles, and Orange. These counties differ along three dimensions: (1) demand heterogeneity, (2) proximity to an existing location, and (3) underlying vaccination rates (Figures F10 and F11).

Figure F10 Census tract distribution by HPI

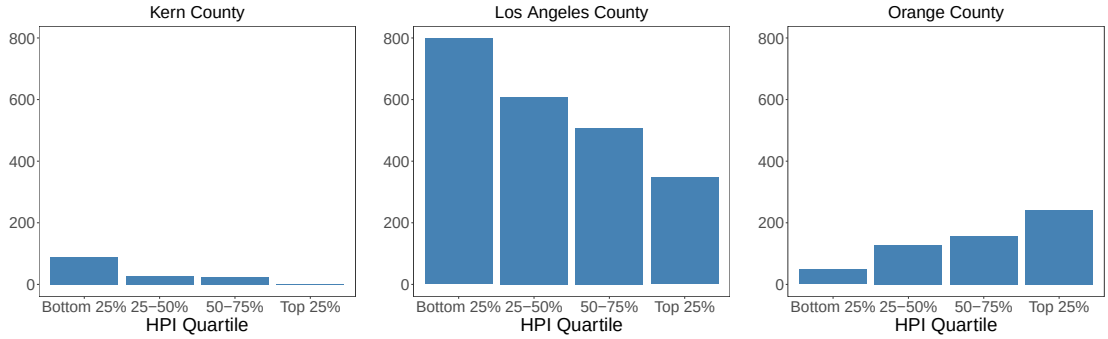


Figure F11 Population weighted density of distance to nearest pharmacy (left) and vaccination rate (right)

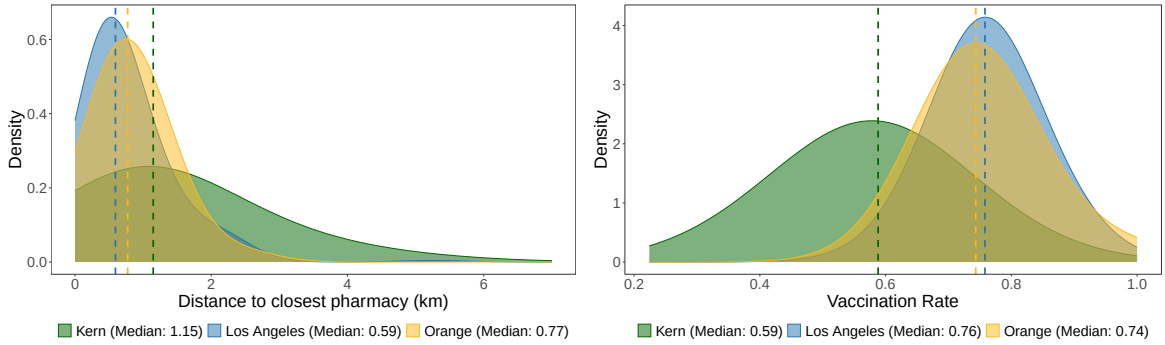


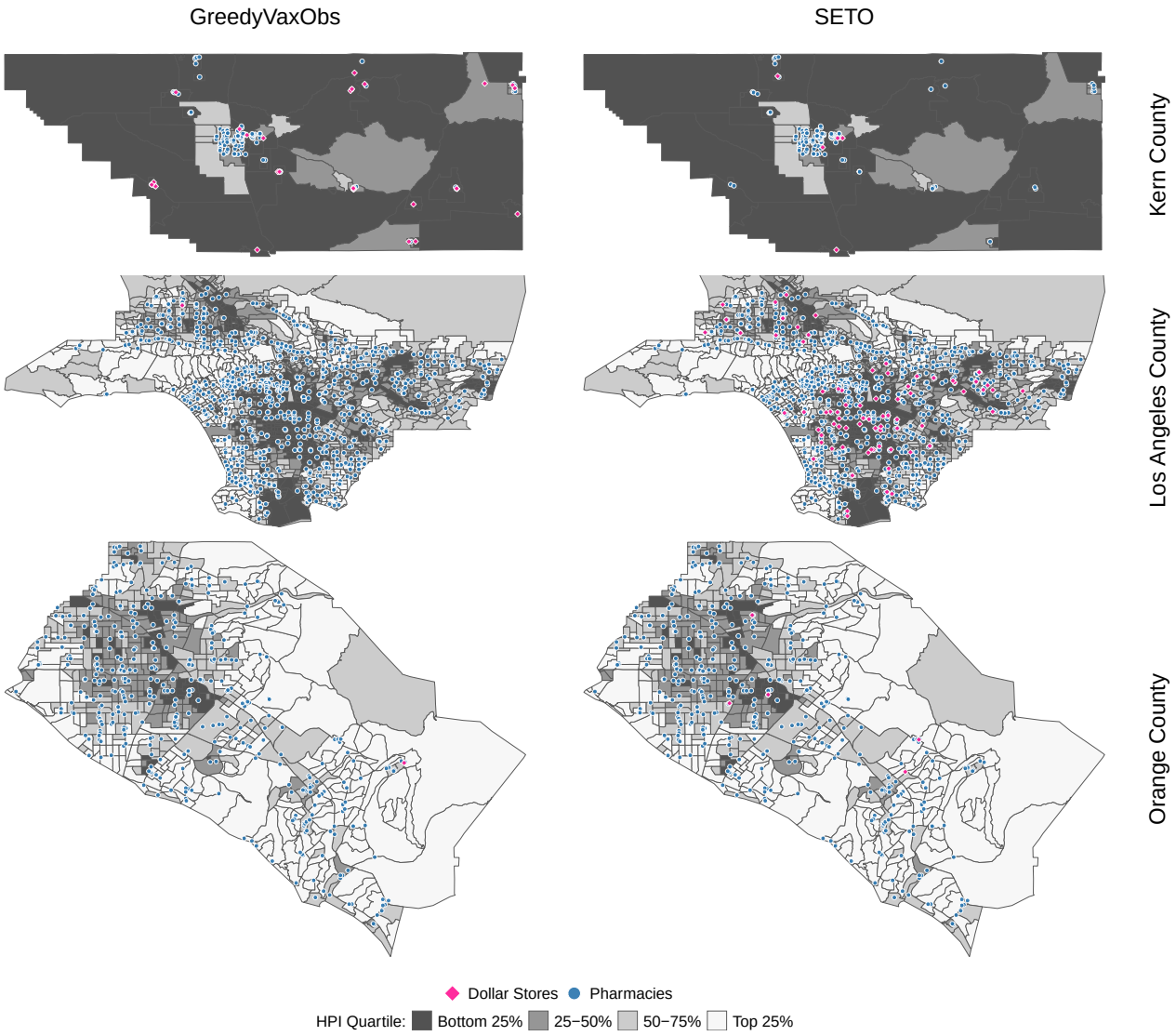
Figure F12 illustrates the new locations selected under each policy. We observe that GreedyVax$^{\text{Obs}}$ opens *too many* new locations in Kern County, *too few* locations in Los Angeles County, and *similar* number of locations in Orange County relative to the SETO policy.

- *Kern County: Low utilization areas with high distance sensitivity communities*
  Under Pharmacy-only, Kern has a mean vaccination rate of 59%. GreedyVax$^{\text{Obs}}$ prioritizes these areas, opening 26 locations. Despite benefiting low HPI individuals facing long distances, this policy ignores the opportunity cost of adding stores. In contrast, SETO opens only 5 locations in Kern and shifts resources to communities that benefit the most (*e.g.*, low HPI areas in Los Angeles).

- *Los Angeles County: High utilization areas with heterogeneous distance sensitivity communities*
  Despite achieving a mean vaccination rate of 76% under Pharmacy-only, Los Angeles includes highly variable HPIs, leading to heterogeneous demand sensitivities. GreedyVax$^{\text{Obs}}$ ignores these demand sensitivities and opens only 3 locations, increasing vaccinations by <0.1%. SETO is demand-aware and identifies service gaps, opening 82 locations and increasing vaccinations by 2.6%. Figure F12 shows the stark contrast between the policies and demonstrates how our demand-aware policy brings vaccination closer to underserved, densely populated, and distance-sensitive areas of Los Angeles County.

- *Orange County: High utilization areas with low distance sensitivity communities*
  The county is characterized by homogeneous areas with low distance sensitivity (mostly high HPI) and closer proximity to existing pharmacies. Demand estimation offers little benefit, and GreedyVax$^{\text{Obs}}$ performs comparably to our SETO approach. Both policies report similar, modest benefits because of the wide availability of pharmacies.

Figure F12 **SETO and GreedyVax$^{\text{Obs}}$ policies added stores in Kern, Los Angeles, and Orange counties, assuming network expansion with 300 dollar stores statewide**



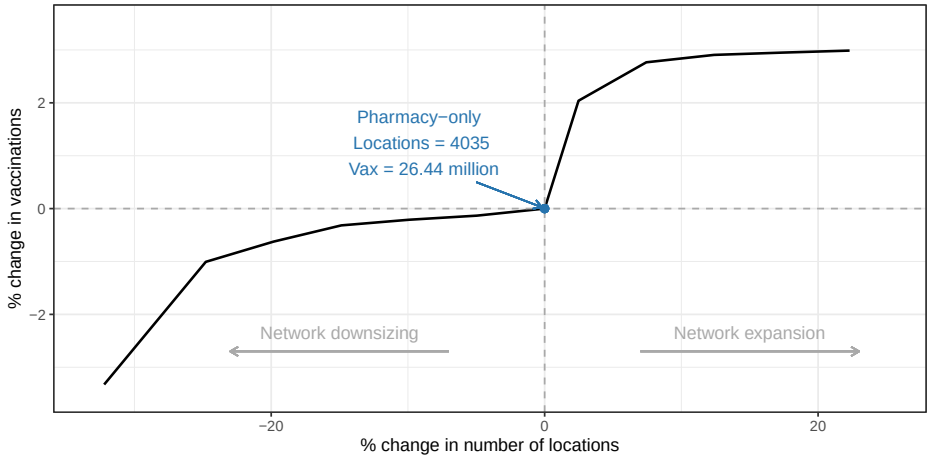
Note: Potential dollar store locations are: 65 (Kern), 162 (Los Angeles), and 41 (Orange).
GreedyVax$^{\text{Obs}}$ selects 26 (Kern), 3 (Los Angeles), and 1 (Orange).
SETO selects 5 (Kern), 82 (Los Angeles), and 5 (Orange).

Appendix G: Other Applications

G.1. Network Downsizing

Our approach could be used to strategically reduce network size, which may be valuable during economic downturns or if operational costs exceed benefits. Simply closing low-utilization locations without considering the broader effects could result in a suboptimal reduced network. For instance, closing 20% of existing pharmacies (approx. 800 stores) based solely on utilization results in 1.1 million fewer vaccinations, with 15% in vulnerable communities. Our optimization-based approach more efficiently selects locations for closure, minimizing the impact on demand fulfillment: with 20% fewer pharmacies, predicted vaccinations drop by only 165,000. If more pharmacies are closed, the marginal loss in vaccinations increases, but strategically selecting store closures helps mitigate the impact of reduced capacity. Figure G13 shows predicted vaccinations under both *network expansion* and *network downsizing*.

Figure G13 Impact of efficient network expansion or network downsizing on predicted vaccinations



G.2. Network Replacement

As an alternative to *network expansion*, we consider a *network replacement* strategy, where a combination of FRPP and dollar stores are jointly selected, holding constant both the total number of vaccination sites (at current FRPP levels, $N = 4,035$), and overall capacity. We use the optimization model to strategically redistribute capacity from existing FRPP pharmacies to selected partner locations.

Assuming dollar stores have identical capacity as pharmacies, replacing 17% (694 out of 4,035) of existing pharmacies with dollar stores generates 0.70 million additional vaccinations, a 2.6% increase, and boosts vaccinations within walking distance by nearly 1.1 million, including 610,000 vaccinations in the most vulnerable regions (Table G21). This reallocation removes pharmacy-based vaccination sites in neighborhoods with excess capacity and establishes new sites in underserved communities (Figure G14). This test case is particularly relevant to the design of a partnership program under budgetary or service capacity constraints. If the capacity of dollar stores is only half that of pharmacies, around 25% of the vaccination gains are

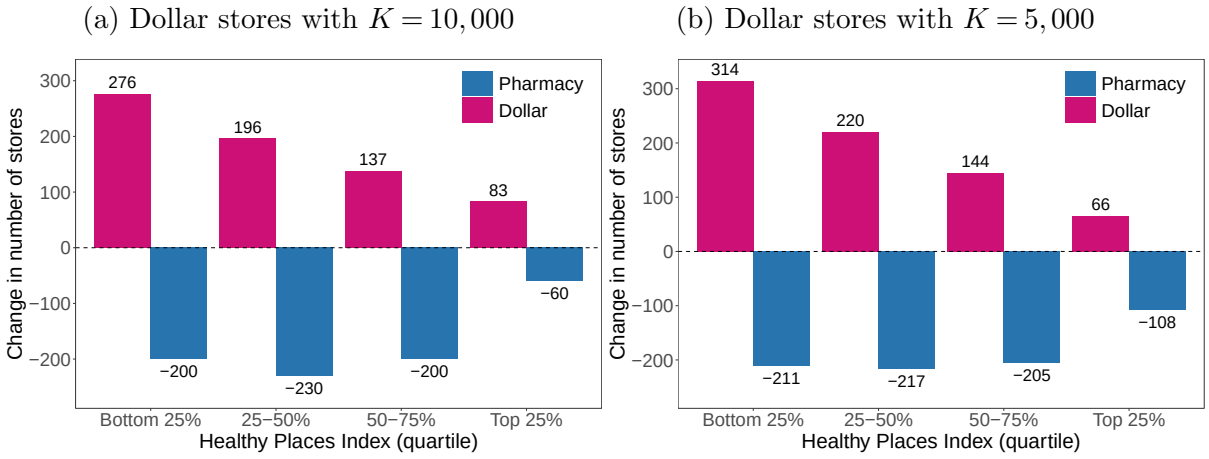
Table G21 Predicted vaccinations (million) under network replacement

Strategy	Dollar store capacity		All	HPI quartile			
				Bottom 25%	25-50%	50-75%	Top 25%
Pharmacy-only		Total	26.44	5.62	6.68	6.83	7.31
		Walkable	19.46	3.98	5.07	4.96	5.45
Pharmacy + Dollar	$K = 10,000$	Total	+0.70	+0.28	+0.12	+0.12	+0.17
		Walkable	+1.07	+0.61	+0.18	+0.14	+0.15
Pharmacy + Dollar	$K = 5,000$	Total	+0.17	+0.09	+0.02	+0.00	+0.06
		Walkable	+0.26	+0.36	+0.01	-0.10	-0.01

Note. All values are reported in millions of vaccinations. The Pharmacy + Dollar strategy is the change in vaccinations compared to Pharmacy-only. Walkable vaccinations refer to travel distances of less than one mile.

predicted. Notably, despite their lower capacity, the model selects a larger number of dollar stores compared to pharmacies replaced (746 vs. 626), with more dollar stores selected in lower HPI areas.

Figure G14 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.

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

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

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.