Incorporating Equity into Vaccine Access

by

Matthias Schumm
Bachelor of Arts in Commerce, 2014, Cooperative State University Baden-Württemberg
and
Mehdi Tagorti
Bachelor of Science in Hautes Etudes Commerciales, 2015, IHEC Carthage and Master of Science in Supply Chain Management, 2017, Normandy Business School

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Signature of Author: ______________________________
Matthias Schumm
Department of Supply Chain Management
May 12, 2023

Signature of Author: ______________________________
Mehdi Tagorti
Department of Supply Chain Management
May 12, 2023

Certified by: ______________________________
Dr. Jarrod Goentzel
Principal Research Scientist, Center for Transportation and Logistics
Capstone Advisor

Certified by: ______________________________
Timothy Russell
Research Engineer, Center for Transportation and Logistics
Capstone Co-Advisor

Accepted by: ______________________________
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

Incorporating Equity into Vaccine Access

by

Matthias Schumm

and

Mehdi Tagorti

Submitted to the Program in Supply Chain Management on

May 12, 2023, in Partial Fulfillment of the

Requirements for the Degree of Master of Applied Science in

Supply Chain Management

ABSTRACT

COVID-19 vaccine access inequity was a major challenge during the pandemic. This inequity was present between countries and regions and within cities. We developed a novel approach to measure and improve vaccine access equity to address this issue. Our approach first created a vaccination attainment index based on CDC COVID-19 data. We then selected the most relevant spatial, e. g., the density of medical facilities, and socioeconomic factors, to train an XGBoost model on the county level of the United States. Using this model on census tracts within counties, we used the Gini and Theil indices to measure equity. We identified the main drivers of vaccine access based on SHAP values. With the main drivers identified (percentage of American Indian and Alaska native population, health insurance coverage, and transportation options), we conducted a case study on Cambridge, MA. We improved the short-term access equity by adjusting each census tract's density of medical facilities (from Gini 0.14 to 0.13). Our novel approach provides decision-makers with a tool to identify and address drivers of vaccine access equity in their region and predict vaccination attainment on the tract level. These insights are crucial to ensuring equal access to vaccines and other essential healthcare services for everyone.

Capstone Advisor: Dr. Jarrod Goentzel
Title: Director of the MIT Humanitarian Supply Chain Lab, Principal Research Scientist, Center for Transportation and Logistics

Capstone Co-Advisor: Timothy Russell
Title: Research Engineer, Center for Transportation and Logistics

Acknowledgments

We would like to express our deepest gratitude and appreciation to our advisors, Dr. Jarrod Goentzel and Tim Russell, for their invaluable guidance, support, and encouragement throughout our capstone project on equity in vaccination access. Their expertise and insights have been instrumental in shaping our understanding and have significantly contributed to the success of this project.

We would also like to extend our heartfelt thanks to our professors Drs. Elenna R. Dugundji, Ilya Jackson, Thomas Koch, and Jafar Namdar, who have generously shared their knowledge and expertise in the quantitative aspects of this project. Their assistance in the application of machine learning techniques and data analysis has been crucial in addressing the complex challenges associated with the study.

A special note of appreciation goes to our writing coach, Pamela Siska, who has provided invaluable assistance in refining our writing skills and helping us effectively communicate our research findings. Her guidance and support have been essential in crafting a well-structured and coherent capstone project.

We are grateful to Daniel Sheehan, who has diligently worked on gathering and processing the geographical and census data that form the foundation of this project. His expertise in handling complex data sets and commitment to ensuring data accuracy have contributed significantly to the quality of this research.

Finally, we would like to acknowledge the support of our peers, friends, and family, who have provided us with a nurturing environment allowing us to focus on our research and grow as researchers. Their unwavering support and encouragement have been invaluable throughout this journey.

This capstone project would not have been possible without these remarkable individuals' guidance, support, and contributions, and we are immensely grateful for their assistance and encouragement.

Matthias Schumm
Mehdi Tagorti

Table of Contents

List of Tables

List of Figures

1 Introduction

Despite private companies' expansion into new areas, public direction is still necessary for certain fields. For instance, privatized healthcare, schools, and social security have all been linked to increased inequality (Basu et al., 2012). Access to health care is one of the cardinal issues of our era (United Nations, 2022). Despite the advancements in the health field, most of the world's population suffers from inadequate access to primary medical care (World Bank & World Health Organization, 2017).

1.1 Motivation

As we observed recently, when the COVID-19 pandemic struck in 2020, those issues of inequality became even more pressing. For instance, we noted that different geographical regions showed a tremendous difference in how they were affected by the pandemic (Brown, 2021). While some countries had sufficient resources to combat the crisis and care for their people, others, e. g. Nigeria, lacked even essential medical supplies like fever thermometers (Ugochukwu et al., 2020). Surprisingly, we noticed severe disparities between regions or even districts within individual countries, even within the United States (Dukhovnov & Barbieri, 2021). A closer look revealed that these discrepancies were mainly between different socioeconomic groups (Dukhovnov & Barbieri, 2021). The socioeconomic factors distinguishing these groups were related to income, race, education, and occupation (Kilinc et al., 2018; J. T. Chen & Krieger, 2021).

Regardless of background, race, and social class, all humans have the right to medical welfare, declared by the United Nations in 1948 (1948). Vaccines have proven to be one of the essential pillars in providing medical welfare (Killeen, 2007), as we have seen, from eradicating smallpox to mitigating the consequences of a COVID-19 infection (Lazarus et al., 2022; National Institute of Allergy and Infectious Diseases, 2019). Therefore, vaccines and their distribution are of the utmost importance for making this right a reality. Even though many countries are developed, the COVID-19 pandemic proved most were unprepared to deal with such an unprecedented event (Organisation for Economic Co-operation and Development, 2020; Richter, 2021). Risk management stipulates that prevention is better than firefighting an acute crisis (Greenberger, 2018). However, there was insufficient preparation from these counties.

The United States is no exception (Wilensky, 2021). In the United States, an established network of vaccination sites exists to provide children of all ages with the required vaccines following recommended immunization plans (Centers for Disease Control and Prevention, 2021). Due to a lack of preparation of policymakers, this network and its infrastructure were used but proved inefficient (O'Donnell & Erman, 2021; Simmons-Duffin & Huang, 2021) and ineffective concerning the entire population's welfare and also concerning equity considerations (Bardosh et al., 2022; Rydland et al., 2022).

This outcome is unsurprising since the existing infrastructure was built to serve continuous demand. It was not meant to serve a larger number of people within a short time frame, while the supply of vaccines could not even be guaranteed. Consequently, immunization coverage was low, and targets were constantly behind plans.

This experience emphasizes the need to prepare for the next pandemic-like event. While current supply-chain network distribution models have proven to improve immunization coverage under the same budgets, they neglected the aspects of equity (Goentzel et al., 2022). Therefore, our objective is to incorporate equity and establish a plan to mitigate the effects when the next pandemic hits. Equity in healthcare can be defined as the absence of avoidable or remediable differences between groups of people, whether these groups are socially, economically, demographically, or geographically defined (World Health Organization, 2022).

1.2 Problem Statement and Research Question

Considering socioeconomic factors like race, age, ethnicity, education, income, and many more, we developed an equity indicator for vaccine distribution networks to attain maximum equitable immunization coverage. Therefore, our capstone proposes the density of medical facilities used for administering vaccines as a possible short-term trigger to maximize equity by targeting communities with access barriers.

The problem of vaccine distribution and the non-acquired herd immunity against certain diseases in many developing countries became a topic of interest with the arrival of the COVID-19 pandemic. Many research projects approached this issue. For instance, the UNICEF model attempted to achieve maximum immunization coverage against COVID-19 by relating access to vaccines to the distance between individuals and vaccination centers (Goentzel et al., 2022). While this model showed satisfactory results in the field, for example, in Gambia, the approach remains superficial because it fails to account for the characteristics of communities and ignores the resulting disparities. This failure to consider socioeconomic factors was also noted during the recent pandemic, where vaccine coverage was not equitable across regions and countries. The resulting inequity occurred due to several factors: unavailability of vaccines at the beginning of the virus' spread, the urgency of distributing them once available, lack of resources, logistical difficulties, and many more. For all those reasons, it did not allow decision-makers to take the necessary time for essential preparation and to get the whole picture, particularly considering social inequalities. Hastily, politicians decided to use the existing network infrastructure for vaccine distribution to children at schools, health centers, and pediatricians. Their approach led to inequitable vaccine distribution.

Based on our motivation, our capstone answers the following questions:

1. How can we define measures of equity in the context of vaccine accessibility?
2. How can we use these vaccine equity measures to improve the equity of vaccine access?

1.3 Project hypotheses and outcomes

This project proposes an equity measure to serve as an input into a supply-chain network design model to provide decision-makers and healthcare providers with a tool to identify the number of required vaccination sites under equity aspects.

First, we hypothesized that equity can be proxied through socioeconomic factors, which can be proxied by geospatial data. We conducted a literature review of the common definitions of equity and their variables and how to measure them. We could confirm it through our empirical study on the county level of the United States by selecting the socioeconomic features that impact vaccination attainment.

We hypothesized that a data-driven supply-chain network design considering socioeconomic criteria would result in maximally equitable immunization coverage. To consider equity in a distribution network, our model can serve as an input, suggesting the number of vaccination sites required in local geographic areas: sites per census tract. Running our scenarios by adjusting the density of medical facilities confirms this hypothesis in improving equity measures on the tract level in the United States or Cambridge, MA case.

Following the WHO's equity definition, we finally hypothesized that when we can isolate spatial effects from socioeconomic factors, the outcome solely based on socioeconomic factors should be equal to attain equity. Only spatial factors should be the cause of differences between vaccine access. Consequently, changing spatial factors, like the density of medical facilities, could be used to improve equity by compensating for disadvantages. For ease of reading, when we talk about socioeconomic factors, we mean to include demographic and cultural aspects describing the population living in a region. Spatial factors, on the other hand, describe the location where people live. It is important to note that the WHO definition also refers to the *geographic* attributes that describe a population. This reference must not be confused with the spatial factors that describe a region.

2 State of the Art

This section discusses how equity is defined (2.1) and distinguishes between equity, equality, justice, and fairness. We focus on two equity dimensions, social equity (2.1.1) and spatial equity (2.1.2), and how they are linked. We then connect them to establish equity in health systems (2.2) and apply these concepts to vaccine distribution (2.3).

2.1 Definition of equity

Our capstone has now discussed the abstract idea of equity without discussing its definitions further. This section discusses how equity and equality are linked and identifies the difference between both concepts. They are manifestations of what we perceive as just or fair, i. e., right. Therefore, equality usually means that the outcomes or procedures are the same under the same conditions (Cook & Hegtvedt, 1983; Mehrabi et al., 2020).

On the other hand, equity is sometimes referred to as *relative equality* and accounts for different needs, which will not necessarily lead to the same outcome (Cook & Hegtvedt, 1983; Mehrabi et al., 2020). Cook and Hegtvedt (1983) point out that it depends on the situation and which concept of equity stakeholders prefer. The Centers for Disease Control (Centers for Disease Control and Prevention, 2020b) defines health equity as “when everyone has the opportunity to be as healthy as possible” on its website and sets it as its objective to attain health equity. We believe that the concept of equity used to access vaccine equity highly depends on the context and requires different processes or procedures.

In the following sections, we look at social and spatial equity concepts to relate them to the health field later.

2.1.1 Social equity

Many theories evolved around social equity (utilitarianism, intuitionism, egalitarianism, Rawls' theory of justice, and Sadr's theory of justice), which Dastgoshade et al. (2022) considered when developing their view on social equity. Although these theories are derived from alternative doctrines, they share the primary goal of practicing social justice in a distribution network to ensure that no population group or area will be deprived of COVID-19 vaccines. In the following, we will give a short overview of utilitarianism (2.1.1.1), Rawls' theory of justice (2.1.1.2), and Sadr's theory of justice (2.1.1.3) as the main relevant streams for our capstone project.

2.1.1.1 Utilitarianism

Various philosophers studied the theory of utilitarianism over the years, initially thought by Jeremy Bentham and John Stuart Mill. It is based on three values (Pereira et al., 2017):

- Human welfare: this value should be the basis of equity.
- Equal respect: mutual respect that should be practiced.
- Ethical judgment about actions should be based solely on consequences, particularly how well-being can be enhanced.

For utilitarians, what matters is how resources are distributed in the final state, and they do not care if that means only a select few receive them or if everyone's rights are violated in the process. Utilitarianism is concerned only with maximizing whatever social outcome is deemed most important, regardless of the state of society at the time. Among utilitarians, moral equality is defined as the inclusion of all people in calculating social welfare. Forcing this to its logical conclusion would mean that no one has the right to life if their death would result in the best possible overall outcome for society (Lewis et al., 2021).

2.1.1.2 Rawls' theory of justice

Justice as fairness is also a central tenet of Rawls' "A Theory of Justice." Rawls assumes that people are rational, self-interested actors trying to maximize their claims to income and wealth, which are the primary goods of society. It also points out that everyone has the right to the same minimum level of freedom, also known as the principle of greatest equal liberty (Lewis et al., 2021).

When developing the principle of justice theory, Rawls introduced the "original position" as an artificial device. The device created a hypothetical situation in which members of the population could reach a contractual agreement on the distribution of resources without one part of the peers appearing to be more advantaged. However, individuals could align the principles to their advantage without further control (Lewis et al., 2021).

John Rawls created this framework to reflect the principles of justice that could prevail based on free and fair social interactions. Absent a 'veil of ignorance,' specific individuals, such as the privileged and talented, would exert pressure on the vulnerable, weak, and handicapped. Coercing vulnerable members of the population renders any contractual arrangement that may exist, in nature, null and void (Norton, 1989).

2.1.1.3 Sadr's theory of justice

According to Sadr's theory of justice, social equity is the equitable distribution of benefits among social groups (Reda, 2014). As a result, a fair system enables people from all social classes to become wealthy (Fahlevi, 2019). Sadr's theory also holds that the economic climate should be such that poverty and deprivation are eradicated while total benefits are maximized. Sadr believed that achieving these two goals would create a social balance among a society's various classes, thereby reducing class differences (Behbahani et al., 2019). Social balance and mutual responsibility for support are the two pillars of Sadr's theory (Askari & Mirakhor, 2020, p. 22).

Al-Sadr supports limited economic liberties and the joint ownership of assets (private and public). He identifies these as the conditions necessary to realize a just society wherein the distribution of resources is equitable and poverty eradicated. This theory also argues that the ultra-rich should not be allowed to hoard their wealth, as doing so hurts both the rich and the poor (Lewis et al., 2021).

In summary, utilitarianism is more focused on the final result and how to maximize the social outcome without considering the limited resources or capacities as constraints, or the equitable sharing of resources among groups and individuals. It can be assimilated to maximizing accessibility to vaccine centers, not considering the disparities between the individuals (e. g., minimizing distance, allocating centers everywhere). In contrast, Rawls' and Sadr's theories focus on compensating for inequalities and offering a fair opportunity. The common point of those theories is that they try to privilege the least advantaged people. However, both theories do not guide how to define the neediest proportions of society.

2.1.2 Spatial equity

While social equity looks at the interaction between individuals and society, it makes sense to decrease the granularity and investigate how those interactions occur spatially (Whitehead et al., 2019). We explore the different dimensions of spatial equity (horizontal and vertical) in section 2.1.2.1 and review the most frequently used measures of spatial equity in section 2.1.2.2.

2.1.2.1 Horizontal and vertical Spatial Equity

We can identify horizontal and vertical spatial equity to assess spatial equity in equipment distribution. On the one hand, horizontal spatial equity refers to the equitable distribution of facilities among residents regardless of location or socioeconomic status (Ashik et al., 2020). In other words, this strategy aims to improve residents' access to facilities disregarding their needs. Vertical spatial equity, on the other hand, refers to an equitable distribution of facilities over space based on the population's need or demand (Ashik et al., 2020). This approach aims to reduce inequalities in equipment distribution by ensuring unequal treatment with inequalities by prioritizing the needy people.

Lucy (1981), in his research papers, developed five facets of equity: “equality” (each resident consumes equal services); “need” (consumption level based on residents’ needs); “demand” (residents actively benefiting from equipment should receive a higher allocation); “preferences” (services equal to resident preferences); and “willingness to pay” (the level of consumption depends on the willingness to pay). According to the *equity needs approach*, most socially disadvantaged populations should receive more services to provide these groups with favorable conditions they would not otherwise have (Nicholls, 2001). The concept of “equality” adheres to the principles of horizontal spatial equity, whereas “needs” and “demand” adhere to the concept of vertical spatial equity.

To create a long-term distribution system of urban amenities, ensure adequate and easy access for city dwellers, and improve residents’ quality of life, planners must consider spatial equity in the distribution of urban amenities. In the context of the geographical supply of urban equipment, spatial equity helps planners assess the consequences of current strategies for allocating urban resources and study “the inter- and intra-effects of all urban amenities on townspeople” (Ashik et al., 2020).

The measure of accessibility has been widely used in the literature for the quantitative assessment of spatial equity (Talen & Anselin, 1998; Tsou et al., 2005; Liao et al., 2009; Mashrur et al., 2015). However, because these studies have examined equality rather than equity of access, most accessibility-based studies consider only spatial accessibility in assessing spatial equity and measure the accessibility of horizontal spatial equity. These studies assume that an equitable distribution of urban amenities occurs when all urban residents, regardless of location or socioeconomic status, can experience equal ease and convenience in arriving at a service in a given travel time or distance (Talen & Anselin, 1998; Nicholls & Shafer, 2001; Tsou et al., 2005; Liao et al., 2009; Mashrur et al., 2015). It is apparent and reasonable that the level of spatial accessibility will vary across space (Dear, 1974). Furthermore, there may be inequity in the distribution of urban amenities over space due to allocating priority to socially disadvantaged or highly needed areas, resulting in inequality but equity in access to this equipment. In this regard, Rawls proposes that inequalities can be distributed to benefit the most disadvantaged (Ashik et al., 2020).

These studies, however, could not assess vertical spatial equity in terms of needs, constraints, and demands because it is impossible to uncover the needs of urban and socially disadvantaged residents without considering spatial accessibility.

As a result, these studies did not contribute to identifying whether and to what extent better spatial accessibility to urban facilities corresponds to the needs of socially disadvantaged populations or the demand of city dwellers. Although few studies have examined vertical spatial equity, these studies have several limitations. Most researchers evaluate spatial equity studies in the context of a specific facility (Smoyer-Tomic et al., 2004; Boone et al., 2009; Chang & Liao, 2011; Yuan et al., 2017; Kelobonye et al., 2019). If researchers argue that within the framework of spatial equity, the scarcity of one urban facility can be compensated by the abundance of another, this implies that the technique for measuring spatial accessibility should include all types of urban facilities in a unique setting (Mashrur et al., 2015; Tsou et al., 2005; Dadashpoor et al., 2016). Moreover, these studies only consider a single indicator to characterize spatial accessibility in terms of local demand or social disadvantage.

All urban residents should have equal spatial access to urban amenities to assess the horizontal spatial equity of urban amenities (Liao et al., 2009; Nicholls & Shafer, 2001; Mashrur et al., 2015). In contrast, for vertical spatial equity, there should be an inequality of spatial access arranged according to the needs of the socially disadvantaged population and the demand of urban residents. It is necessary to identify needs and demands before implementing 'needs' and 'demand-based approaches. Low-income, poor, and minority groups, among others, can be identified as needs (Lucy, 1981), and demand can be measured by identifying the actual users of a facility (Kelobonye et al., 2019). As the measure of spatial accessibility addresses the socioeconomic, demographic, and cultural constraints of the user population ("acceptability," "affordability," and "awareness" of a facility), it can be used to assess the "needs and "demand" based approaches to equity.

For example, in the transport sector, accessibility is a critical indicator in assessing the spatial distribution of transport facilities as well as the operational efficiency of the transport system. Transportation planners can assess the vertical and horizontal equity of transportation benefits by revealing the level of accessibility between regions and groups of people (Jin et al., 2022).

A study conducted by Zhao et al. (2020) used SHAP values in combination with the Gini index, to extract the drivers of horizontal equity in healthcare in China and decompose them into "need-factors" and "non-need" factors (household registration, region, work status, education, income, insurance, and marital status). The value they base their equity calculations on are the healthcare expenses for the groups subject to the study (see 3.4 on SHAP).

In summary, horizontal and vertical equity are key pillars for our research. On the one hand, we intend to maximize the populations' access to vaccine distribution centers based on spatial factors. On the other hand, we must consider different socioeconomic factors to characterize the different social groups and attain equitable access based on spatial factors.

2.1.2.2 Most frequent measures of spatial equity

In a systematic review of 75 papers, Whitehead et al. (2019) synthesize which measures were used in the past to measure spatial equity. Table 1 shows the ten most frequent measures, the definitions of equity, and the count of papers using them. The most common measures were the Gini Index, regression, and spatial autocorrelation or local indicators of spatial autocorrelation (LISA). Most papers used a needs-based definition of equity; however, there is no consensus on how need is defined and whether the need is viewed at an individual's level or the entire population. Regression and correlation analysis mainly was used to identify the most relevant socioeconomic factors and their relation to the investigated equity measure.

***Table 1**:*
The ten most common measures of spatial equity analysis are classified by the associated definitions used by each article.

Spatial equity measure	Definition (1) Distributional fairness ($n = 11$)	(2) Needs-based distribution ($n = 45$)	(3) Outcomes or causes ($n = 6$)	(4) Not provided ($n = 14$)	Total ($n = 75$)
Gini Index	2	6	n/a	9	17
Regression	2	6	2	4	14
Spatial autocorrelation or LISA	1	10	2	1	14
Correlation	4	7	1	1	13
Concentration Index	n/a	8	n/a	1	9
Lorenz curve	n/a	3	n/a	2	5
Population analysis of accessibility	1	2	n/a	1	4
ANOVA	2	1	1	n/a	3[a]
Atkinson Index	n/a	n/a	n/a	1	2
Robbin Hood Index	n/a	2	n/a	n/a	2
Service ratios	n/a	n/a	n/a	1	1

[a]Some papers included more than one definition, and the total number of papers using this method has been adjusted accordingly and does not equal the sum of the row. Furthermore, 26 papers used more than one method, and so the total number of papers in each definitional category does not equal the sum of the column.

Note. From "How can the spatial equity of health services be defined and measured? A systematic review of spatial equity definitions and methods" by Whitehead, J., L. Pearson, A., Lawrenson, R., & Atatoa-Carr, P., 2019, Journal of Health Services Research & Policy, 24(4), 270–278.(https://doi.org/10.1177/1355819619837292). Copyright 2019 by Whitehead, J., L. Pearson, A., Lawrenson, R., & Atatoa-Carr, P.

Remarkably, eight papers investigated by Whitehead et al. (2019) assessed equity without defining it. It also concludes that vertical equity definitions may be more applicable when discussing health-related issues. When using the Gini coefficient, the authors conclude that it does not require an explicit definition of spatial equity. However, they alert that if not used carefully, it may only lead to equal outcomes, which are not necessarily equitable.

The Gini coefficient is an inequality measure, not a resource allocation measure. Also, it adheres to anonymity and scalability. It is anonymous which subgroups have high and low vaccination coverage. Considered "scale independent," the Gini coefficient is unaffected by the vaccine supply size or population size. Finally, the Gini coefficient can compare vaccination distributions among subgroups (Enayati & Özaltın, 2020).

2.2 Equity in health systems

After exploring the general concepts of equity, this section discusses how researchers proxied equity and its different measures in the medical field, especially health services distribution. We start by defining spatial factors used in assessing medical needs (2.2.1) and then detail the factors influencing health facilities' attractiveness (2.2.2). That explained, we highlight the importance of efficiency in equity (2.2.3) and how factors other than distance can be an advantage or barrier to accessibility (2.2.4).

2.2.1 Non-Spatial factors in assessing medical needs

Certain factors remain essential to study equity within groups to characterize a society. Field (2000) developed a relative advantage index after compiling a list of factors affecting access to health care. He identified the following factors based on year 2000 census data:

- **Demographic variables (sex and age):** Seniors over 65, children 0-4, and women 15-44 are among the demographic variables influencing health needs (considering all three population groups to have high needs for primary medical care services).
- **Socioeconomic status**: poverty rate, female-headed households, home ownership, and median income. Low socioeconomic status can be an obstacle to healthcare access.
- **Environment**: households with more than one person per room on average and residences lacking basic amenities can contribute to higher levels of poor health.
- **Linguistic barrier and service awareness**: Non-white minorities, uneducated people, and linguistically isolated households face language barriers and lack awareness of services. Minorities, low levels of education, and linguistic isolation may be significant barriers to accessing health care.
- **Transportation mobility**: households without vehicles. People who rely solely on public transport may have less mobility and significantly limited access to physicians.

While Field (2000) operationalized the factors to identify the vulnerable groups and highlight the disparities between groups, Kilinc et al. (2018) used different socioeconomic factors to assess the distribution of healthcare services:

- **Racial/ethnic population composition:** Data was collected on the Black, Hispanic, Native Hawaiian, and other Pacific Islander groups.
- **Rural/urban status:** The percentage of the rural population and population density are two variables chosen to reflect population locations' rural/urban status.
- **Income:** The income per capita, unemployment rate, and poverty rate were chosen as income variables, with the latter defined as the percentage of the population living below the federal poverty line.
- **Education:** The percentage of the population with a high school diploma or higher is chosen to reflect the level of education. This data is obtained at the census tract level from the ACS and converted to the postal code level.

As socio-demographic variables are frequently correlated, Wang and Luo (2005) studied the non-spatial factors to aggregate them into three indicators:

- **Socioeconomic disadvantages:** Socioeconomic disadvantage is a crucial factor, and it includes six variables: female-headed households, poor population, non-white minorities, vehicle-less households, home ownership, and homes lacking basic amenities. Except for the homeownership variable, all loading coefficients are positive, indicating that a lower percentage of owner-occupied housing units is generally associated with a more deprived neighborhood. This variable is a general indicator of socio-economic disadvantage. According to the field's research, high-scoring areas (low-access areas) are concentrated in urban areas, while low-scoring areas (good-access areas) are found primarily in suburban and rural areas. As Wang and Luo (2005) point out, this factor is unquestionably the most significant.

- **Sociocultural barriers:** The second factor includes linguistically isolated households, households with more than one person per room, and those without a high school diploma. Positive loading coefficients exist for all three variables, forming a comprehensive indicator of sociocultural barriers. As noted earlier, language isolation and low educational attainment may be linked to a lack of awareness of services, a significant barrier to accessing health care. Immigrants are less educated and more likely to live in overcrowded housing. High-scoring (hard-to-reach) areas are mainly concentrated in and around Chicago and its suburbs.
- **High healthcare needs:** The third factor includes two variables: the population with high needs and median income are the least important. Note that the income variable is split almost evenly between the first and third factors (negative correlations). The focus on the first factor is justified, while the focus on the third factor is less intuitive but not entirely unexpected.

Using these indicators, Wang and Luo (2005) identified four types of shortage areas that need to be prioritized while allocating resources:

- **For geographical groups:** “Areas of poor spatial access” and “Areas of marginally poor spatial access with high needs.”
- **For demographic groups:** “Disadvantaged population” and “Marginally disadvantaged population with sociocultural barriers.”

2.2.2 Health Facilities attractiveness and its effect on equity

In the previous section, we characterized the demand in the health sector based on socioeconomic factors. In comparison, Tian et al. (2022) studied the accessibility and equity of the multi-tiered medical system in Shenzhen from another perspective. Having the facilities close geographically (minimum distance) to the inhabitants cannot influence the patients’ willingness to go to health centers. The quality of the service offered and the transportation time and mode could be a barrier or an advantage for center accessibility. Tian et al. (2022) introduced the concept of the “attractiveness of health facilities” and visitor “tolerance of health facilities.”

According to Tian et al. (2022), the attractiveness of health facilities is proxied by the hospital’s planned capacity. A hospital’s planned capacity is determined by the potential demand of the surrounding neighborhoods, which is reflected in the hospital’s final size. This includes the number of beds, doctors, nurses, and health professionals. The attractiveness of the hospital can be expressed using an equation based on the proportion of effective consultations compared to the number of doctors in the hospital for outpatient visits. Thus, visitor tolerance was defined as the overall assessment of residents visiting a healthcare facility, which included two factors: the population’s distance from the healthcare facility (based on travel time: drive and walk time) and the attractiveness of the healthcare facility.

From this study, we concluded that distance, while important, is not the only factor in assessing equitable access to health centers. Other factors related to the facility features must be considered when evaluating optimal locations for vaccine distribution, like the number of physicians, service quality, and crowding effects. Also, Tian et al. (2022) used the Gini index to assess the spatial equity of health facilities at the sub-district level, and the results showed that regions with higher accessibility generally experienced better equity among different groups. We, therefore, concluded that higher accessibility could realize better equity results.

2.2.3 Equity-efficiency Trade-off in medical care facilities

Whereas treating equity in health distribution services is vital, we need to consider the efficiency of the allocated resources, especially costs. Our study provides a framework for decision-makers, so we cannot omit the budget constraint when deploying equity.

Cho’s (1998) equity-efficiency trade-off model uses three factors to determine its objective function: consumer welfare, producer welfare, and the opportunity to receive service. System efficiency is assessed by the total of consumers’ and producers’ welfare, while system equity is reflected in the ability of all people to obtain medical care regardless of their ability to pay for that care per capita.

The welfare accrued to providers of medical services can be measured by the concept of producer surplus, i. e., the welfare changes to suppliers because of price changes. The difference between revenues (to the service providers) and total variable costs (incurred costs to provide the service) is equivalent to “producer surplus,” conceptually measured by the area between the supply curve and price.

The per capita transportation cost incurred to reach medical facilities is used to define system equity, which is an objective function in this equity-efficiency trade-off model. However, existing allocations are often well outside the optimal range for a given set of equity-efficiency weights. Despite the preferences encompassing both some efficiency and equity gains may be possible (Cho, 1998).

So accordingly, having an equitable vaccine distribution network is related to the nature of the vaccine, as the social groups do not have the same needs for the different vaccines (e. g., MMR, COVID-19, Varicella). The population needs vary and must be addressed based on the vaccine and target group. Also, the transportation cost (public or private car) could be an excellent proxy for evaluating accessibility besides distance and travel time.

While this study considered the producer surplus to evaluate efficiency, which can be the right approach for choosing private providers, we can only examine the allocation cost (budget) for opening a location in the case of vaccines, proxied by the cost of the density of medical facilities.

2.2.4 Means of transport as an equity driver

Many studies argue that distance can be a good proxy to assess accessibility, and we discussed how equity is affected by travel time to facilities in the preceding sections. In this section, we look closer at the impact of the means of transport on the equity of access.

2.2.4.1 Relative accessibility difference between private cars and public transit

Jin et al. (2022) concluded that the higher the level of health care, the larger the proportion of people with relative access to public transportation. High-income groups are more accessible by public transportation than by private car, while low-income groups are less accessible by public transportation. Furthermore, high-income groups have the highest absolute values for all health services, implying that high-income groups benefit more from public transportation than other groups. High-income neighborhoods are concentrated in central urban areas with well-developed bus and metro systems, whereas low- and middle-income neighborhoods are concentrated in the suburbs with less developed public transportation, particularly for metro services.

2.2.4.2 Spatial accessibility inequity by private cars and public transportation

Using the Theil index, Jin et al. (2022) found that relative accessibility by private cars (RAPC) to tertiary healthcare services (THC) is the highest, indicating the greatest inequality in the relative accessibility of tertiary healthcare services by private cars. The value of RAPC services to primary healthcare services (PHC) is the lowest, implying that relative accessibility to PHC services by private car is balanced. In terms of total inequity, the inequity of access by private car increases with the level of health services, whereas the greatest inequity of access by public transportation exists in secondary services, followed by tertiary and primary services. Within-group inequality shows that high-income households have the best access to tertiary health care, while low-income households have the best access to primary health care. For low-income groups, access by private car is more equitable than access by public transportation, while the opposite trend was observed for high-income groups. In terms of intergroup inequality, the level of health services increases intergroup inequality.

Furthermore, intergroup inequity in public transportation is greater than in private cars, implying that access to public transportation is more unequally distributed among different groups. As we can see from these findings, having access to public transportation or owning a private, depending on social group, can penalize access to health services based on the social group. Thus, adding these factors to our scope is relevant to assess accessibility and the attractiveness of the locations. Another essential aspect to consider when creating subgroups based on socioeconomic factors is population density. According to Jin et al. (2022), Geographical correlations between population density and relative accessibility are generally positive. Public transport accessibility and population density are correlated stronger than private car accessibility and population density. Correlations between population density and access to higher-level health services are strong.

In conclusion, while many studies have examined inequality in spatial accessibility to health services, few have looked at inequalities between and within groups. Furthermore, most approaches attempt to conceive the distance-dependent interactions between the supply of health services and the demand of populations and the representation of competition between populations for limited resources. The results of the accessibility assessment can be used to identify disadvantaged areas and reallocate facilities accordingly for populations. The absence of avoidable or remediable differences between groups of people, whether these groups are socially, economically, demographically, or geographically defined, is defined as equity (World Health Organization, 2022). Inequalities can be categorized as horizontal/ geographical or vertical/ socioeconomic. It is important to note that the WHO definition also refers to the *geographic* attributes that describe a population. This reference must not be confused with the spatial factors that describe a region.

2.3 Vaccine distribution approaches

This section examines what has been done recently on vaccine distribution. We will first see how the scenarios would have performed in past pandemics (2.3.1), especially the effects on equity. Then, we identify factors besides socioeconomic ones in the context of vaccines used in mixed integer linear programming (2.3.2). Finally, we investigate how social equity theories perform in a real-world situation (2.3.3) and emphasize the key takeaways.

2.3.1 Equitable Influenza vaccine distribution

Researchers tried to study how vaccine distribution networks performed and what would happen if we applied an equity lens in the past pandemic. For instance, Enayati & Özaltın (2020) used a segmented model for influenza transmission and formulated a mathematical program to minimize the number of vaccine doses distributed in the case of the 1957 Asian influenza pandemic. They propose an equity constraint to assist public health officials in considering equity when making vaccine distribution decisions. They performed a sensitivity analysis to demonstrate the application of the proposed model on key epidemic parameters and devised a precise solution approach that yields a vaccine distribution policy with guaranteed solution quality.

2.3.1.1 Model Formulation

Enayati and Özaltın (2020) developed a fractional model to characterize influenza spread through interacting subgroups. They developed a mathematical program to reduce the number of vaccines distributed to contain a possible epidemic. Finally, they suggested a constraint to ensure subgroup coverage equity.

The population is divided into four compartments based on disease stage in classic SEIR epidemic models: susceptible, exposed, infected, and cured (or suppressed). Each compartment is assumed to be homogeneous. Each disease stage portion is further subdivided.

They calibrated their models by age and zone: Five age groups are mutually exclusive: preschool (0-4), school (5-18), young people (19-29), middle-aged (30-64), and seniors (65+). Although each pandemic virus has unique illness-specific properties, they used parameter values from the 1957 Asian influenza pandemic to calibrate disease spread models.

2.3.1.2 Principal findings

Adding the equity constraint to the model, they introduced an auxiliary variable representing uniform vaccine coverage across all regions and enforced an adapted Gini Index (GR) constraint. By constraining model equity in regional vaccine allocation, the proposed solution to the same problem with GR ≤ 0.04 and GR ≤ 0.02 shifted vaccine allocation from other regions to a particular region when coverage inequity tolerance was reduced – knowing that GR equals 0.1 when coverage inequity is not controlled.

Vaccine uptake increases when the equity constraint is combined with a tolerance for inequality – greater vaccine efficacy results in a higher percentage increase. After applying the equity constraint, vaccine efficacy improves, so vaccine usage increases even more to meet the equity constraint. As they considered more regions, the difference between the minimum and maximum regional coverage grew.

We can infer from this study that vaccine distribution equity can be a driver to meeting maximum vaccine coverage and so better immunization to stop virus spread. Based on past pandemics, scenario analysis can help assess our real-world scenarios, as the models used in the past did not meet equity requirements.

2.3.2 Immunization coverage maximization through an optimized vaccine network

Other researchers tried to maximize immunization coverage by enhancing accessibility geographically. In this sense, Goentzel et al. (2022) developed a model to support a government's strategic decisions regarding vaccination campaign budgeting with an optimized method of determining vaccination point location. Using a Mixed Integer Programming (MIP) optimization, the authors used a multi-configuration approach to develop this model, first defining an exploratory model to facilitate interaction with UNICEF in formulating the formal model for use with empirical data from Gambia.

2.3.2.1 Model formulation

They developed an exploratory model to maximize vaccinations in fixed health centers or outreach sites with a limited budget. Fixed health centers were established locations that offered routine immunization services. Outreach sites were one-day clinics held in more remote communities by people sent from fixed health centers. The demand function was modeled as an endogenous function that reflects people's willingness or ability to travel for vaccination services. The authors concluded that as the vaccination location became further away from them, people would be less likely to travel for vaccination services. Besides the physicians and allocated resources, these properties can affect the center's attractiveness, which can be a determinant factor of equity and enhancing the accessibility of vaccine locations.

2.3.2.2 Principal Findings

Goentzel et al. (2022) could reduce the distance between vaccine services and the population by adding outreach sites under a fixed budget. With more candidate proximity sites, location optimization reduced the average distance by more than 10% and increased vehicle utilization. No health center had more than twenty monthly trips, meaning the outreach plan was feasible even if a health center only had one vehicle. With no vehicle costs and more productive employee days, service cost was lower at a fixed health center.

2.3.3 Distribution networks designed for the COVID-19 vaccine

Operationalizing social theories and incorporating them into modeling helped to see how they would perform in vaccine distribution. Dastgoshade et al. (2022) developed a new stochastic distribution network design model based on social equity and solved it using a multi-objective approach in the case of two large provinces in Iran. Researchers discussed the context of the COVID-19 pandemic in central provinces that have been heavily influenced.

Dastgoshade et al. (2022) applied three social equity theories to COVID-19 vaccine distribution to build their model. They aimed to help decision-makers and experts choose the best theoretical framework for equity while considering all relevant factors. Firstly, applying utilitarianism seeks to maximize COVID-19 vaccines for each social group. Secondly, Rawls' theory was used because it prioritizes poor groups. Sadr's theory of justice was the third one used with the same objective as Rawls' theory. The calculations and equations proposed by the authors serve to reduce class differences and maximize total benefits. They used the Gini index in the calculations to calculate inequity.

The following section will summarize how Dastgoshade et al. (2022) approached their network design for vaccine distribution equity.

2.3.3.1 Model Formulation

For their distribution network, Dastgoshade et al. (2022) tried to minimize the Distribution Centers' (DC) opening costs, transportation resource allocation costs, and vaccine delivery costs from DCs to social groups to express the efficiency of the designed distribution network.

The authors considered vaccine requirements, cold chain and equipped facilities, and the predefined delivery points at social group locations (the public health authorities predefine the set of potential DCs in certain places in the territory, e. g., vaccination centers and hospitals). Consequently, the study examined the role of social equity theories on the COVID-19 vaccine distribution network in low- and middle-income countries, housing in large cities, dispersed populations, lack of health access in small towns, and poor logistics, health services, and infrastructure in rural communities. Characterizing social vulnerabilities is vital to distributing COVID-19 vaccines fairly to developing countries and vulnerable groups.

In order to identify social vulnerabilities and ensure equitable distribution, the study considers three social groups. Human settlements are divided into cities, towns, suburbs, and rural areas based on population density (Dijkstra & Poelman, 2014; Dijkstra et al., 2021). Applying social theories to vaccine distribution requires knowledge of local conditions. Before calculations are made, each social group is assumed to be a specific type.

Dastgoshade et al. (2022) assumed that social groups are rated hierarchically as low, moderate, and highly (poorest) vulnerable. They also state that supply shortages correlate positively with the ranking of the social groups: the more vulnerable a group, the more uncertain the supply. Thereby, it adds complexity to serve those groups that need it most. In the first phase of the vaccination program, those most vulnerable to COVID-19 would be vaccinated first. Thus, the highest complexity occurs at the beginning of a vaccination campaign.

2.3.3.2 Principal Findings

Dastgoshade et al.'s (2022) solution approach is based on the Lexicographic Goal Programming (LGP) method, which is appropriate for multi-objective problems with established priorities. The LGP method prioritizes different objectives in lexicographical order to reduce undesirable slack variables.

The study presents nine alternative design solutions that are created and compared using three social equity theories applied (utilitarianism, Sadr's theory, and Rawls' theory) to a bi-objective model and two single-objective models (the first is used to maximize allocation, while the second goal is to minimize the total costs).

We concluded from the study of (Dastgoshade et al., 2022) that each theory proposes a different solution and does not satisfy the same objective regarding cost efficiency, population coverage, and equity between groups. Also, each theory's applicability varies from the specific context and cannot be generalized to all groups as each theory focuses on other aspects of equity.

Table 2:
Social Theories Impact on Vaccine Distribution

	Utilitarianism	***Sadr's theory***	***Rawls' theory***
Objective	Maximize welfare across society as a whole	Social balance among different classes of society and maximize welfare	Prioritize underprivileged groups (equal opportunity)
Coverage in Cities	93%	95%	94%
Coverage in Suburbs and towns	92%	89%	90%
Coverage in Rural Areas	97%	98%	100%
Cost and Allocation	Most Expensive (Maximize Allocation)	Most Cost-efficient	Least Efficient (Expensive with low allocation)
Adapted Context	Developed Countries	Countries seeking balanced solutions	Countries with the highest rural population

Note. The numbers in this table are compiled from "Social equity-based distribution networks design for the COVID-19 vaccine" by Dastgoshade, S., Shafiee, M., Klibi, W., & Shishebori, D., 2022, International Journal of Production Economics, (https://doi.org/10.1016/j.ijpe.2022.108684).

These results show that not all the factors can be integrated and applied in the same way to proxy equity in different contexts. The approach to establishing equitable distribution networks for vaccines can differ based on the attributes to evaluate and the objectives to attain. In addition, we must account for uncertainty as it influences the performance of each social equity theory. For example, it will make Rawls' theory the least reliable solution for the equitable vaccine allocation problem when the rate of arrivals is relatively low (Dastgoshade et al., 2022).

Dastgoshade et al. (2022) and Mohammadi et al. (2022) provided an overview of some papers conducting network distribution planning and incorporating equity. The authors compare equity measures, distribution context, objective, model characteristics, and decision level. As shown in Table 3, only two of the seven models were studied, and the proposed model uses a multi-objective design; all of the models except for the study of Dastgoshade et al. are deterministic, and only three of them provide a strategical component.

Table 3:
Reviewed vaccine distribution networks literature

Author (year)	Equity	Equity Measure	Decisions				Objective		Case Study
			Allocation	*Location*	*Inventory*	*Capacity Planning*	*Single-Objective*	*Multiple-Objective*	
Chen et al. (2014)		-			✓		✓		Multiple
Yarmand et al. (2014)		-	✓				✓		Influenza
Dai et al. (2016)		-	✓		✓		✓		Influenza
Lemmens et al. (2016)		-			✓	✓	✓		General
Sadjadi et al. (2019)		-		✓		✓	✓		General
Lim et al. (2019)		-		✓	✓		✓		General
De-Carvalho et al. (2019)		-			✓			✓	MMR
Bertsimas et al. (2020)		-	✓				✓		Covid 19
Yang et al. (2020)		-			✓		✓		General
Chen et al. (2020)	✓	Gini Index	✓				✓		Covid 19
Enayati and Özaltın (2020)	✓	Gini Index	✓				✓		Influenza
Rastegar et al. (2021)		-		✓	✓		✓		Influenza
Sinha et al. (2021)		-			✓		✓		Encephalitis
Thul and Powell (2021)		-	✓				✓		Covid 19
Balcik et al. (2022)	✓	Minimizing the deviation from the fair coverage levels	✓				✓		Covid 19
Goentzel et al. (2022)		-	✓	✓			✓		General
Dastgoshade et al. (2022)	✓	Social equity theories	✓	✓				✓	Covid 19
Khodaee et al. (2022)	✓	Minimizing the deprivation Cost	✓		✓		✓		Covid 19

Note. Data for this table was compiled from "Social equity-based distribution net-works design for the COVID-19 vaccine" by Dastgoshade, S., Shafiee, M., Klibi, W., & Shishebori, D., 2022, International Journal *of Production Economics, (https://doi.org/10.1016/j.ijpe.2022.108684), "Bi-objective optimization of a stochastic resilient vaccine distribution network in the context of the COVID-19 pandemic" by Mohammadi, M., Dehghan, M., Pirayesh, A., & Dolgui, A., 2022, Omega, 113 (https://doi.org/10.1016/j.omega.2022.102725) and "Vaccine network design to maximize immunization coverage" by Goentzel, J., Russell, T., Carretti, H. R., & Hashimoto, Y., 2022, Journal of Humanitarian Logistics and Supply Chain Management (https://doi.org/10.1108/JHLSCM-10-2021-0101). Copyright 2022 by Elsevier B.V, copyright 2022 by Elsevier Ltd. and copyright 2022 by Emerald Publishing Limited.*

2.4 Conclusion

Our review shows that equity has many parallel definitions that sometimes overlap or contradict each other. Which theory to favor highly depends on the situation, the individual's equity perception congruence, and the situation to judge. We looked at the spatial expansion of equity theories comprising horizontal and vertical equity. While one is close to being equal to equality, the other focuses on an individual's, a group's, or an area's needs. Both theories applied can lead to outcomes being considered equitable. Much, sometimes contradicting, literature exists on which factors should be used. For example, the United States census data also defines a vulnerability index based on other socioeconomic factors.

Regarding health systems, we found it essential to consider non-spatial factors when discussing spatial equity. Those need to be proxied by other means (e. g., distance or time) and then translated into an equity goal. Furthermore, we found that there is always a balance between efficiency (of cost and other resources) and equity in health systems.

We also found the WHO's equity definition most suitable for our capstone. It provides a clear definition and subsumes other ideas, like the *equity needs approach* or *horizontal equity,* into a single clear definition. Thus, we oriented our study mainly on this definition.

3 Methodology

In our capstone, we investigated how we can incorporate different equity measures to assess vaccine access based on Covid-19 historical data in the United States. Whitehead et al. (2019) identified that "[f]uture research should examine the impact of different measures of accessibility and need on the results of spatial equity." We, therefore, selected different sets of relevant socioeconomic and spatial factors from publicly available datasets (3.1). We then processed the data collected to treat the missing values and calculate a measure for vaccination assessment (3.2). Before running a model, we ran a correlation analysis to reduce the number of features further. With a reduced feature set, we built a model to explain vaccination attainment using an index we developed (3.3). From the model built, we then used Shapley values to extract the most significant socioeconomic and spatial factors driving equity of access (3.4). We then validated our model statistically and used a bottom-up approach to ensure we could infer meaningful insights in census tract areas (3.5). We then calculated different equity metrics (Gini and Theil indices) to measure the equity among tracts (3.6). Finally, we ran scenarios to assess the as-is situation in the United States first and in Cambridge, MA, second and propose a more equitable solution by influencing the density of medical facilities as the only (short-term) endogenous factor (3.7).

3.1 Data Collection and Modelling

A significant part of our methodology entailed selecting the best socioeconomic and spatial variables to calculate the equitability of a specific vaccine distribution network on a microscopic scale, i. e., towns or municipalities based on census tracts. We selected the data from the American Community Survey (ACS, 2017) and the Social Vulnerability Index (Centers for Disease Control and Prevention, 2020a) as these two sources regroup all the needed socioeconomic factors based on the most recently conducted surveys. Those publicly available datasets offer adequate socioeconomic information on county and tract levels, which we used to distill the meaningful drivers for vaccination attainment. In addition to these two datasets, we used the ArcGIS Business Analyst to get the number of medical facilities and public transportation stops (Bureau of Transportation Statistics, 2023) in a census tract or county, respectively. We selected medical facilities that were most likely to have administered COVID-19 shots based on their Standard Industrial Classification (SIC) codes:

- Clinics,
- Pharmacies,
- Pharmaceutical kiosk,
- Pulmonary and respiratory diseases,
- Medical centers,
- Preventive medicine,
- Hospitals,
- Health screening and vaccinations,
- Allergy physicians.

However, vaccination attainment over time for COVID-19 vaccine data is publicly available only at the county level (Centers for Disease Control and Prevention, 2023). We, therefore, needed to calibrate our model based on county levels and assume that those factors relevant to the aggregated county level have the same impact on individual tracts' vaccination rates. As the dependent variable for our model, we wanted to consider the highest vaccination rate a county achieved at some point in time and how long it took to reach this level.

Therefore, considering those two factors, we built the vaccination attainment index, VAIc, for each county (Equation 6). The index is calculated as the natural logarithm of the time in weeks it took to achieve the maximum vaccination level (Equation 3) over the highest vaccination rate (Equation 2).

$$\mathrm{C} = \text{Set of Counties c} \tag{1}$$

$$v_c = \text{maximum vaccination rate (i. e, percentage of people fully vaccinated)}\ \forall\, c\ \in C \tag{2}$$

$$\mathrm{w}_c = \text{weeks to reach the maximum vaccination rate } v_c\ \forall c \in C \tag{3}$$

$$a = \min w_c\ - 1 \tag{4}$$

$$b = \max w_c \tag{5}$$

$$VAI_c = \ln \frac{w_c - a}{v_c \cdot (b - a)}\ \ \forall\, c \in C \tag{6}$$

We excluded all counties where the maximum vaccination rate was zero as outliers since we assume that it is practically impossible that no vaccination took place over the last two years. The natural logarithm was applied to reduce the strong skew of the data. *Weeks to Max* was scaled to achieve comparable ranges between *Weeks to Max* and the *max vaccination rate*. The scaling was offset to avoid the argument of the natural logarithm becoming zero. The vaccination attainment index measures the speed required to attain the achieved vaccination rate. Our index indicates that a geographical region performs better (higher vaccination rate in a shorter time) when the index is lower.

We then verified that the results obtained at the county level could be applicable at the tract level by aggregating tract-level predictions to the respective counties and comparing them to the actual outcomes (see section 3.5). This aggregation was performed by calculating the weighted average based on the population of each tract within each county. It is essential to highlight that we assumed in our model that the vaccination attainment is a result of only the selected socioeconomic and spatial factors: This assumption implies, for instance, that site attractivity, availability of a particular vaccine brand, or vaccine hesitancy only play a minor role in the equitability of vaccine access. We made this assumption because (1) we wanted to see the effect of allocating the facilities on equity, keeping all other factors equal, and (2) because it includes attitudinal factors, which are very hard to measure, and we did not have data readily available.

3.2 Data Processing

Once we compiled all the required data for our study, we processed the data into a format usable in our model. We checked for potential outliers and missing values and dropped those rows as they did not represent much of the dataset. It was crucial to be careful for two reasons. First, a technical one: census data is sometimes masked to protect individuals' identities. This masking is not apparent in all cases and may lead to outcome bias. The main concern is when conducting the bottom-up analysis on tract-level data since counties are large enough not to require masking.

Additionally, an ethical reason: we aimed to propose a consensual equity indicator, which can potentially affect the well-being of thousands of people. Consequently, to measure the effectiveness of vaccination campaigns during COVID-19, we calculated an index that combines the percentage of vaccinated people compared to the eligible ones and the time needed to attain this percentage. For the spatial factors, we calculated the densities of medical facilities and public transportation stops based on each county's population and area. We then normalized all features of both datasets before proceeding.

3.3 Model selection

The datasets we selected provide significant insights into the socioeconomic and spatial aspects of the individual geographic regions. However, we assumed that neither socioeconomic nor spatial features individually could explain the accessibility of vaccines sufficiently. We tested our assumption by running a linear regression baseline model to explain the county-level vaccination attainment solely using the unprocessed features. We could confirm our assumption by a poorly performing initial model with an R^2 of 0.14.

Instead of using such a simple model, we expected a high level of interaction between the socioeconomic and spatial feature classes and engineered the interaction accordingly. We continued to engineer interaction terms between the spatial and the socioeconomic features for better results. However, the explanatory power only improved slightly, resulting in an R^2 of 0.15. We, therefore, decided to consider more advanced machine learning models: extreme gradient boosting (XGBoost). Chen and Guestrin (2016) describe it as a state-of-the-art ensemble model using tree-boosting. XGBoost led us to an R^2 of 0.69 using the raw features and performed considerably better than the linear baseline model in explaining the data's variance.

The main ideas behind XGBoost can be summarized as follows:

Gradient Boosting: XGBoost is an implementation of gradient boosting, a technique for optimizing a differentiable loss function by iteratively fitting weak learners to the residuals of the previous learners. The main idea is to minimize the loss function by updating the model with the sum of the predictions of the weak learners. Gradient boosting calculates the gradients of the loss function concerning the predictions of the previous learners and updates the model accordingly (Friedman, 2001).

Regularization: XGBoost incorporates regularization to control the complexity of the individual decision trees and prevent overfitting. It uses both L1 (Lasso) and L2 (Ridge) regularization terms ($\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$) in the objective function, which penalizes the complexity of the trees, reducing overfitting and improving the generalization of the model (T. Chen & Guestrin, 2016).

3.4 Feature selection and explainability using SHAP

Both primary datasets, the ACS and the SVI, have many features in common. We thus had to disregard some of the features to avoid duplicates and ease achieving the explainability of our model. We, therefore, need to make a pre-selection of potentially relevant factors. Researchers and practitioners can use various methods to tackle multicollinearity, including eliminating highly correlated variables or using dimensionality reduction techniques such as Principal Component Analysis (Gujarati & Porter, 2009, pp. 320–351). For selecting the relevant variables of the raw datasets, we decided first to eliminate highly correlated variables based on their correlation coefficients (exceeding a threshold of 0.7). Eliminating highly correlated variables and addressing multicollinearity before modeling is an essential step in data preprocessing, as it can improve model performance and interpretability (Dormann et al., 2013). These techniques help reduce the risk of overfitting, enhance feature importance interpretation, and stabilize model estimates.

After reducing the potential features to consider, we assessed the essential features against common sense to switch variables for others with a more intuitive meaning. For example, we added the number of transportation stops per medical facility and the density of medical facilities per area instead of respectively the density of transportation stops per population and the population density as it is more tailored for our study. This selection was made only on socioeconomic factors because we found that spatial variables did not suffer from high correlation. Additionally, we restricted our selection to percentages and densities only to maintain generalizability on tract data.

For further refining the selection, we needed to assess the importance of the features on our model's outcome. Since we switched from an easily explainable linear regression model to an XGBoost model, an ensemble model, we cannot easily explain the importance of features nor their impact on the outcome of a county or census tract. Thus, we use *SHapley Additive exPlanations* (SHAP) values to (1) identify the most critical features driving equitable access to vaccines and (2) how they affect each tract, helping us to identify those independent variables we need to adjust.

The main ideas behind SHAP can be summarized as follows:

Shapley Values: Shapley values are a way to distribute each player's contributions in a cooperative game fairly. They are derived from the marginal contributions of each player to all possible coalitions. The Shapley value for each player is the average of their marginal contributions across all possible coalitions (Shapley, 1953). Zhao et al. (2020) used SHAP values to decompose their study's contribution factors to inequity (see 2.1.2.1).

Connection to Machine Learning: SHAP values connect Shapley values from cooperative game theory to machine learning by considering each feature as a player in the game and the prediction as the outcome. The SHAP value for each feature measures the average contribution of that feature toward the prediction across all possible feature combinations (Lundberg & Lee, 2017).

Additivity: SHAP values are additive, meaning the sum of SHAP values for all features in a given instance should equal the difference between the model's prediction for that instance and the expected prediction across all instances. This property makes SHAP values easily interpretable, providing a consistent and locally accurate representation of the model's behavior (Lundberg & Lee, 2017; Shapley, 1953).

Model Agnostic: SHAP is a model-agnostic method, which means it can be applied to explain the output of any machine learning model, including linear models, tree-based models, and deep learning models. The method can be adapted to different models by computing each feature's marginal contributions in the context of that specific model (Lundberg & Lee, 2017).

KernelExplainer and TreeExplainer: The SHAP library provides multiple explainers for different types of models. The *KernelExplainer* approximates the SHAP values using a weighted linear regression and is suitable for any model. The *TreeExplainer* is explicitly designed for tree-based models, such as decision trees, random forests, and gradient boosting machines (including XGBoost), and computes exact SHAP values for these models more efficiently (Lundberg & Lee, 2017). We thus used TreeExplainer in our model since it is the only suitable approach for the XGBoost model we used.

After defining the main features impacting vaccination attainment, we studied the top 10 interactions based on the SHAP values. We calculated the following correlation coefficients to study these interactions: Pearson, Kendall's Tau, and Spearman. Pearson, Kendall's Tau, and Spearman coefficients are all correlation coefficients used to measure the strength and direction of a relationship between two variables.

The Pearson Correlation Coefficient measures the linear relationship between two continuous variables, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation) (Chok, 2010). Kendall's Tau and Spearman's rank correlation coefficients are non-parametric measures of association based on the ordinal ranking of data points (Chok, 2010).

Kendall's Tau measures the similarity of the orderings of two variables, while Spearman's rank correlation measures the strength and direction of the relationship between two ranked variables. Kendall's Tau and Spearman's coefficients range from -1 to 1, with -1 indicating a perfect negative correlation and 1 indicating a perfect positive correlation (Chok, 2010). These coefficients offer different perspectives on the relationships between variables, with Pearson being most appropriate for linear relationships and Kendall's Tau and Spearman being more robust for non-linear or non-parametric relationships.

The final selection of parameters for the proposed consensual equity measure was derived from the selected features running the correlation and SHAP analysis. In Table 4, we show the remaining features:

Table 4:
Remaining features after feature selection

Factor type	Subgroup	Remaining Feature	Variable Name in the model
Demographic	*Age*	Percentage of persons aged 65 and older estimate	% Persons aged 65 and older
		Percentage of persons aged 17 and younger estimate	% Persons aged 17 and younger
	Gender	Percentage of persons Male	% Male
Social	*Ethnicity*	Percentage of American Indian and Alaska Native Alone persons	% American Indian and Alaska Native Alone
		Percentage of Black or African American Alone persons	% Black or African American Alone
		Percentage of Asian Alone persons	% Asian
		Percentage of persons with Two or More Races	% Two or more races
Economic	*Occupation*	Percentage Population 16 Years and Over in Labor Force	% in Labor Force

Factor type	Subgroup	Remaining Feature	Variable Name in the model
	Mobility	Percentage of households with no vehicle available estimate	% Households with no vehicle available
		Percentage of Workers 16 Years and Over that Walk to Work	% commute walking
		Percentage Workers 16 Years and Over that use Motorcycle for transport	% commute w/ motorcycle
		Percentage Workers 16 Years and Over that use Car, Truck, or Van for transport	% commute w/ car, truck, or van
		Average time to commute to Work (in minutes)	average commute (in min)
	Housing	Percentage of occupied housing units with more people than rooms estimate	% more people than rooms
		Percentage of mobile homes estimate	% Mobile homes estimate
Geographic / Environment	*Rurality*	Percentage of housing in structures with ten or more units estimate	% Housing in structures with 10 or more units
	Facility Access	The density of Medical Facilities per Area Not attractive	Density Medical Facilities per Area
Other	*Health Status*	Percentage of persons without Health Insurance Coverage	% w/o health insurance
		Percentage of persons with Private Health Insurance Coverage	% with private health insurance
	Education	Percentage of persons (age 5+) who speak English "less than well."	% Persons who speak English 'less than well

Note. Compiled by the authors. Copyright 2023 by the authors.

3.5 Model validation

After selecting the best features, we trained the model on county-level data based on the preprocessed data. To validate our model, we proceeded with a double-validation approach. The first is statistical, splitting the data on training and testing using a cross-validation technique to ensure our model is neither over- nor underfitting. In cross-validation, *folds* refer to the number of equally sized subsets into which the dataset is divided. By dividing the dataset into folds, the model can be trained on different subsets of the data and then tested on the remaining unseen portion. The second is geographical, by consolidating the data on the tract level and seeing whether the results match the calculation obtained on the county level.

First, we calculated different metrics: Mean Average Error (MAE) and Mean Squared Error (MSE), to evaluate the fit of our model and its ability to explain the variability in our dataset. From the tracts, we consolidate our prediction to the county level to validate that our model predicts well the As-Is situation on the disaggregated level since we only have county data available. The predicted index was aggregated using the weighted average by populations of the tracts within the county.

3.6 Equity calculation

We initially aimed to assess vaccine access equity based on our model's prediction for census tracts. Our central assumption was that if we remove all the effects of the spatial factors from the outcome, we get the effect socioeconomic factors have on the outcome. According to the WHO equity definition, we propose that the outcome based solely on socioeconomic factors should be equal (see). We could then apply the equity measures introduced earlier based on the outcome.

However, we could not remove spatial factor effects numerically because we shifted to a non-linear XGBoost-model, and the result of the SHAP analysis showed that depending on the county or census tract, the effects of those variables vary. The individual SHAP force plots on the tract level highlight the drivers for the vaccination attainment index; spatial and socioeconomic factors significantly contribute to the vaccination attainment index.

While only used in one paper subject to Whitehead et al. (2019), we think it could be helpful for our study to use the Theil index. This index is considered a better tool for analyzing regional inequalities than other indicators because it can also report inequalities between and within different population subgroups (Jin et al., 2022). We, therefore, decided to focus on using both the Gini and the Theil index and calculated their values based on the As-Is situation in the United States.

3.7 Scenario analysis on the United States and Cambridge, MA

Finally, we conducted a scenario analysis to analyze the impact of different strategies for increasing the density of medical facilities on vaccination rates. For all our scenarios, on the entire United States and Cambridge, MA, we considered the density of medical facilities as the only endogenous variable we can influence to impact the equity of vaccine distribution. The first scenario involved a uniform increase in medical facilities across all tracts in the United States, while the second and third scenarios involved increasing medical facilities proportionally to population density and average commute time, respectively. These scenarios were then applied on a smaller scale, focusing on tracts within Cambridge, MA. The results of these scenarios were analyzed in terms of their effects on the Gini and the Theil index. Based on Theil contributions calculated on the tracts within Cambridge, MA, we clustered the tracts using the K-means and Elbow method.

The K-means algorithm is a clustering method used to partition a dataset into K distinct, non-overlapping clusters based on the similarity of their features. As an unsupervised machine learning algorithm, it works iteratively to assign each data point to one of the K groups based on the features provided (Cui, 2020). The algorithm tries to make the intra-cluster points as similar as possible while keeping the clusters as different (far) as possible. For each cluster, the algorithm places data points so that the sum of their squared distances from the centroid (the arithmetic mean of all the points in that cluster) is minimal (Nainggolan et al., 2019). In addition to K-means, we used the *Elbow* method to determine the optimal number of clusters. The Elbow method is a heuristic used to determine the optimal number of clusters in a dataset. It involves running the K-means algorithm for a range of K values and plotting the total within-cluster variation or the sum of squared distances between the data points and their respective cluster centroids against the number of clusters (K) (Nainggolan et al., 2019). As the number of clusters increases, the within-cluster variation will decrease since data points will be closer to their respective centroids. An elbow point can be identified by visually inspecting the resulting plot where the decrease in within-cluster variation slows.

To summarize, the K-means algorithm is used for clustering data points into distinct groups based on their features, while the Elbow method helps to estimate the optimal number of clusters to use in the K-means algorithm by examining the within-cluster variation as a function of the number of clusters. We performed the clustering step to identify the similarities between the tracts in small scopes contributing to inequity in vaccine attainment.

4 Results

After having looked at how equity can be defined (2.1) and how to apply those definitions in the health sector (2.2), we looked at vaccine distribution in particular (2.3). We continued to discuss the methodology we used to derive a suitable equity measure in chapter 3. In this chapter, we analyze the situation on the county level and distill the meaningful features from the data collected (4.1). We then use those features to create a model to predict the vaccination attainment index created and validate it (4.2). Based on the model, we then evaluate the equity measures discussed (4.3). Using those measures, we conduct a scenario analysis, showing the effect of tweaking features to improve equity (4.4) and conclude with a Cambridge, MA, case study (4.5).

All models and data can be accessed on the Humanitarian Supply Chain Lab GitHub: https://github.com/MIT-HSCL/VaxEquity

4.1 Results at the county level

This study aimed to identify the most relevant spatial and socioeconomic factors influencing vaccine distribution equity at the county level in the United States. We started by calculating an index representing the vaccine distribution pace, which was obtained by dividing the number of weeks it needs to attain the maximum percentage of people vaccinated by the maximum percentage and then taking the natural logarithm of the resulting number (see also 3.1). We excluded the Hawaii and Alaska counties for consistency in our analysis because these two states are geographically isolated from the contiguous United States and have distinct demographic and economic characteristics compared to the other states. This simplified the analysis focused on the contiguous states, which share more similarities regarding geography, demographics, and economic structure.

4.1.1 Disparities on the county level

For the entire United States, we calculated a Gini index of 0.27 at the county level on the vaccination attainment index, indicating that while there is not extreme inequality in vaccine distribution, disparities still exist between counties. Generally, if the Gini-Index is far below 0.5, a distribution is considered equitable (0 is considered perfect equity, and 1 indicates the highest degree of inequity). These disparities could be driven by differences in access to healthcare services, socioeconomic conditions, transportation infrastructure, and local policies affecting vaccination efforts.

4.1.2 Relevant factors

Using the SHAP method in combination with the XGBoost model, we analyzed various features and their impact on this index. This analysis identified several key factors influencing vaccination attainment, including racial and ethnic attributes, health insurance coverage, language proficiency, and transportation access. In the following sections, we provide a qualitative analysis of these factors, examining the underlying reasons for their significance. Table 5 shows all the features ranked by their importance according to the weighted absolute SHAP values.

Table 5:

Feature importance according to SHAP

	feature	importance
1	% American Indian and Alaska Native Alone	6.28E-02
2	% w/o health insurance	5.93E-02
3	% Persons who speak English 'less than well'	3.56E-02
4	% with private health insurance	3.50E-02
5	% commute walking	3.18E-02
6	% Male	3.01E-02
7	% Black or African American Alone	2.94E-02
8	% Housing in structures with 10 or more units	2.60E-02
9	Density Medical Facilities per Area	2.56E-02
10	% Persons aged 65 and older	2.42E-02
11	% commute w/ motorcycle	2.29E-02
12	% commute w/ car, truck, or van	2.13E-02
13	average commute (in min)	2.10E-02
14	% Persons aged 17 and younger	2.06E-02
15	% Households with no vehicle available	2.04E-02
16	% more people than rooms	2.03E-02
17	% Mobile homes estimate	2.01E-02
18	% Asian	2.00E-02
19	% Two or more races	1.98E-02
20	% in Labor Force	1.90E-02
21	% household income between 50 to 75 (k USD)	1.89E-02
22	% commute w/ carpooling	1.87E-02
23	% household income between 25 to 50 (k USD)	1.76E-02
24	% Persons in group quarters	1.75E-02
25	% commute w/ public transport	1.74E-02

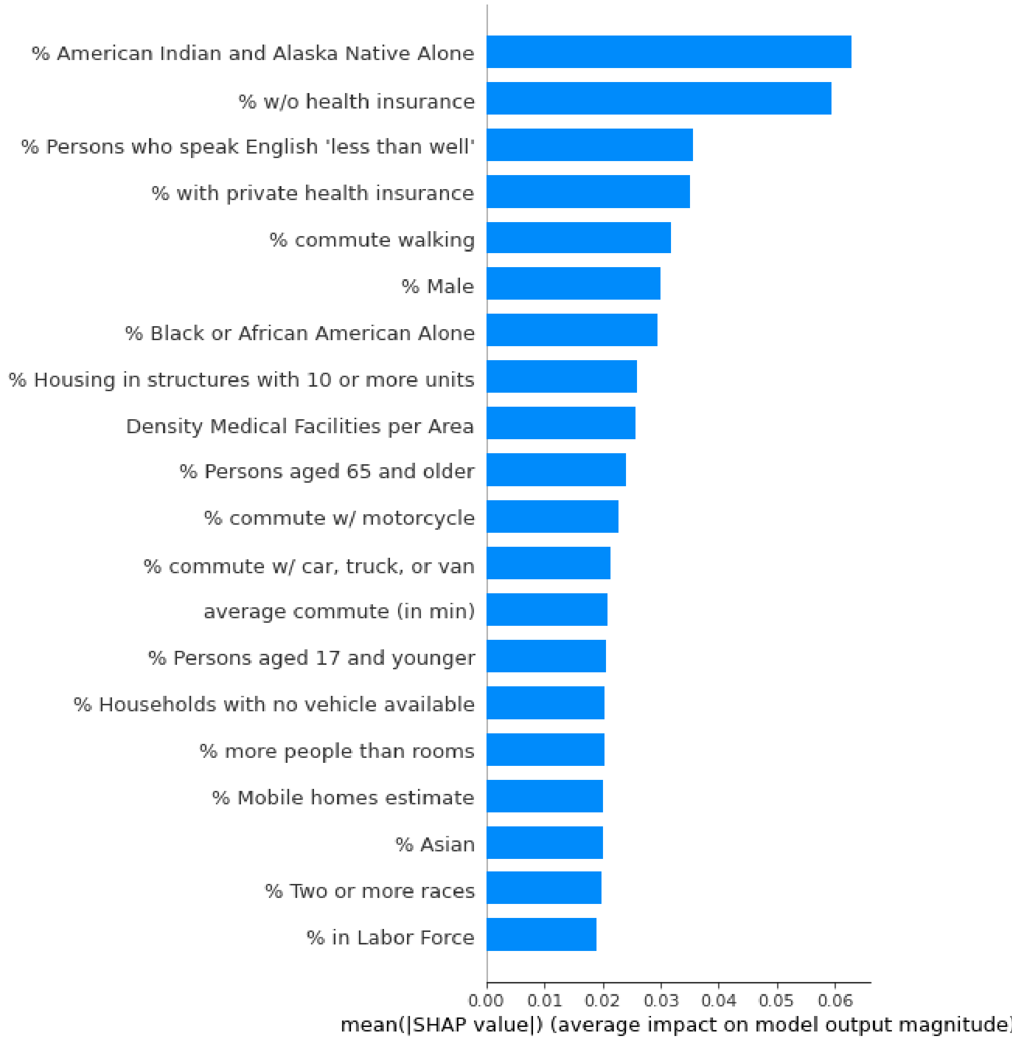
	feature	importance
26	% daytime population	1.67E-02
27	% Single-parent households	1.65E-02
28	% household income between 75 to 100 (k USD)	1.42E-02
29	% unemployed	1.38E-02
30	Gini Index	1.32E-02
31	% in armed forces	1.24E-02
32	% of Housing cost-burdened occupied housing units	1.23E-02
33	% adult population	1.19E-02
34	% disabled	1.12E-02
35	% commute w/ bicycle	1.01E-02
36	Density Medical Facilities per Population	9.77E-03
37	% Some other race, not Hispanic or Latino persons	9.30E-03
38	% commute other means	7.67E-03
39	Density Transportation Stops per Area	5.52E-03
40	Stops per Facility	3.01E-03
41	Flag - Theme Housing Type/ Transportation	2.36E-03
42	Flag - Theme Household Characteristics	6.21E-04
43	Flag - persons with no high school diploma	5.93E-04
44	Unemployment Rate	3.73E-04
45	% of persons below 150% poverty	2.46E-04
46	Flag - Households with no vehicle available	0.00E+00
47	Flag - Single-parent households	0.00E+00
48	Flag - Persons aged 17 and younger	0.00E+00
49	Flag - Housing cost-burdened occupied housing units	0.00E+00

To focus on the most impactful variables in the model, we limited the selection to the 20 most important features, ranking them according to their SHAP-scores. This approach streamlines the analysis, allowing us to concentrate on the key drivers of the target variable, ensuring a more efficient and interpretable examination of the problem at hand. Including additional features beyond the top 20 would likely have minimal impact on the overall analysis, reinforcing our decision to focus on the most influential variables. Additionally, the reduction simplifies the model for decision-makers and enables them to use our findings more effectively in their policy and planning efforts.

As shown in , all the features contribute almost equally to vaccination attainment. An exception is the “percentage of American Indian and Alaska Native Alone persons” and “percentage of persons without Health Insurance Coverage features,” which are significantly more meaningful.

Figure 1:

Most important 20 features driving vaccine access



4.1.2.1 Important Socioeconomic groups influencing vaccine access equity

In the following section, we provide some interpretation of the meaning and importance of each of the most influential 20 factors:

Percentage of American Indian and Alaska native alone population: Historically, American Indian and Alaska Native populations have faced disparities in access to healthcare services, including vaccinations. Inadequate healthcare infrastructure, cultural barriers, and higher poverty rates influence these disparities (Adakai, 2018).

Percentage of persons without health insurance coverage: Due to costs or other associated barriers, individuals without health insurance are less likely to access healthcare services, including vaccinations.

Percentage of persons (age 5+) who speak English "less than well": Language barriers can negatively impact access to healthcare, including vaccinations. In areas with higher percentages of individuals with limited English proficiency, developing language-appropriate resources and educational materials is crucial to ensure equitable access to vaccination services.

Percentage of persons with private health insurance coverage: Private health insurance generally improves access to healthcare services (Arnett et al., 2019), including vaccinations.

Percentage of workers 16 years and over that walk to work: Transportation access is a significant factor in determining an individual's ability to access vaccination services. Transportation barriers may hinder equitable vaccine access in areas with more workers walking to work.

Percentage of the male population: Gender can play a role in healthcare access, including vaccination, through various mechanisms influenced by cultural, social, and economic factors.

Percentage of Black or African Americans alone: Similar to American Indian and Alaska Native populations, Black or African American populations have also historically faced disparities in healthcare services, including vaccinations. These disparities can result from systemic racism, lack of access to quality healthcare, and cultural barriers (Byrd & Clayton, 1992).

Percentage of housing in structures with ten or more units estimate: In more urbanized areas with a higher percentage of multi-unit housing structures, access to healthcare services may be better due to the proximity of medical facilities. However, urban areas may also face overcrowding, which can impact the equitable distribution of vaccines.

Percentage of persons aged 65 and older estimate: Older adults are often more vulnerable to infectious diseases and may require prioritization in vaccination campaigns. Areas with higher percentages of older adults may need additional resources and targeted outreach to ensure equitable vaccine access for this population.

In summary, the socioeconomic features analysis highlights the importance of race, health insurance coverage, language proficiency, and gender in shaping equitable vaccine access. These factors emphasize the need to address disparities and barriers different populations face to ensure equal access to vaccination services.

4.1.2.2 Spatial Factors

The following section discusses the importance of the most meaningful spatial features that influence vaccine access.

The density of medical facilities per area: A higher density of medical facilities within a given area can lead to more accessible healthcare services, including vaccinations. However, ensuring adequate staffing and resources to meet the population's needs in areas with higher facility density is crucial. Overcrowding and long waiting times can result from insufficient capacity at medical facilities.

The average commute to work: Longer commute times can impact individuals' access to healthcare services like vaccinations. A long commute may mean less time and flexibility to schedule vaccination appointments, especially if working hours are inflexible. Policymakers could consider extending clinic hours, establishing mobile vaccination clinics, or partnering with employers to provide on-site vaccination services.

The density of medical facilities per population: Higher density of medical facilities indicates a better potential for healthcare access, including vaccinations. However, it is essential to consider factors such as transportation, cultural barriers, and socioeconomic factors that may influence the utilization of these facilities. For example, a densely populated urban area with numerous medical facilities may face challenges in achieving equitable vaccine distribution if public transportation is inadequate or if specific population subgroups face language barriers.

The density of transportation stops per area: Areas with higher transportation stops can facilitate more accessible healthcare services by providing convenient transportation options to reach vaccination sites. Public transportation is critical in ensuring equitable access to healthcare services, particularly for low-income populations or those with limited personal transportation options. Policymakers should consider expanding public transportation and ensuring it connects communities to vaccination sites effectively To improve vaccine distribution equity.

The number of transportation stops per medical facility: More transportation stops per medical facility can improve access to healthcare services, including vaccinations, by making it easier for individuals to reach these facilities using public transportation. This metric can help identify areas where transportation options are limited, which can inform targeted interventions to improve transportation access. Policymakers could consider increasing the number of transportation stops near medical facilities or providing shuttle services to bridge gaps in transportation access.

Transportation modes and preferences: Different transportation modes and preferences can influence an individual's ability to access vaccination services. Understanding how these preferences interact with each other and the local environment can provide valuable insights for policymakers to improve vaccine distribution equity.

Walk to work: In areas where walking is a more common mode of transportation, it is essential to consider the walkability of neighborhoods and the accessibility of vaccination sites. Ensuring vaccination sites are within walking distance of residential areas or creating safe walking routes can help increase access to vaccination services for those who primarily walk to work.

Motorcycle for transport: Motorcycle use as a primary mode of transportation may indicate limited access to public transportation or a preference for personal transportation options. Policymakers should consider road safety, parking availability, and access to vaccination sites to ensure that individuals using motorcycles can easily access vaccination services.

Car, truck, or van for transport: In areas where private vehicle use is more common, ensuring adequate parking near vaccination sites and providing clear directions can help improve access to vaccination services. Drive-through vaccination clinics can also be considered as they facilitate quick and convenient access to vaccines for individuals with personal transportation.

Carpooled for transport: Carpooling can indicate a reliance on shared transportation options, which may present challenges for individuals seeking vaccination services. Policymakers should consider coordinating with employers to provide on-site vaccination clinics or offering incentives for carpooling to vaccination sites to accommodate the needs of carpoolers.

Public transportation: In areas with higher public transportation usage, it is crucial to effectively ensure that public transit routes connect communities to vaccination sites. Expanding public transportation options or adjusting schedules to match vaccination clinic hours can help improve access to vaccination services for those who rely on public transportation.

4.1.3 Feature interactions

By exploring the socioeconomic and spatial dimensions of vaccine distribution equity, we can understand various communities' challenges and develop targeted interventions to address these challenges effectively.

As we continue to analyze the spatial features, we must recognize that an interaction between socioeconomic and spatial factors may exist. For instance, the density of medical facilities, transportation infrastructure, or commute times may affect rich and poor people differently. Since the XGBoost model is tree-based, it can identify those interactions independently.

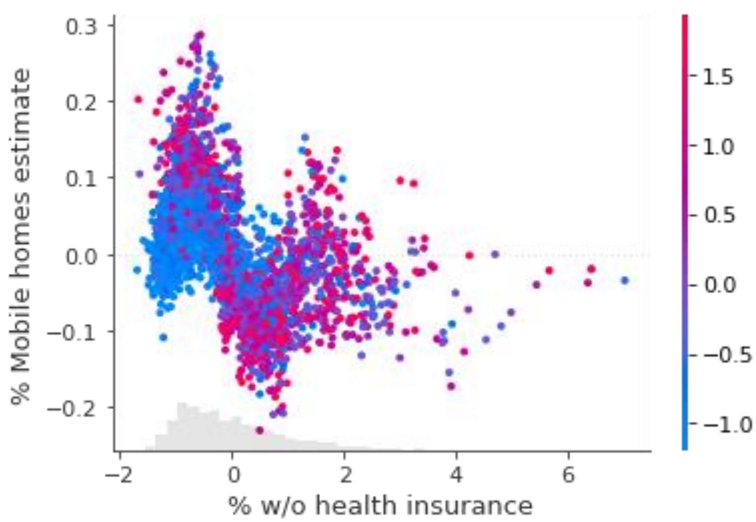
After determining the relevant features, we used SHAP interactions to identify the most important interaction effects on the model's predictions between pairs of features. The interaction effect between two features is the additional contribution that arises when both features are present together beyond their individual effects. SHAP interaction values help identify which pairs of features significantly impact the model's predictions when they interact. Besides, we calculated the Pearson correlation coefficient for the top interacting feature pairs to evaluate the strength and direction of the linear relationship between those features and Spearman's rank correlation and Kendall's tau coefficients to capture the non-linear relationships.

4.1.3.1 Interaction terms

We discuss the top ten interactions in the following section, following the approach discussed in section 3.4.

Interaction 1 (Figure 2): Percentage No. Health Insurance Coverage and Percentage of mobile homes estimate, coefficients: Pearson (0.36), Spearman (0.44), Kendall's tau (0.30).

Figure 2:
Percentage No. Health Insurance Coverage and Percentage of mobile homes estimate



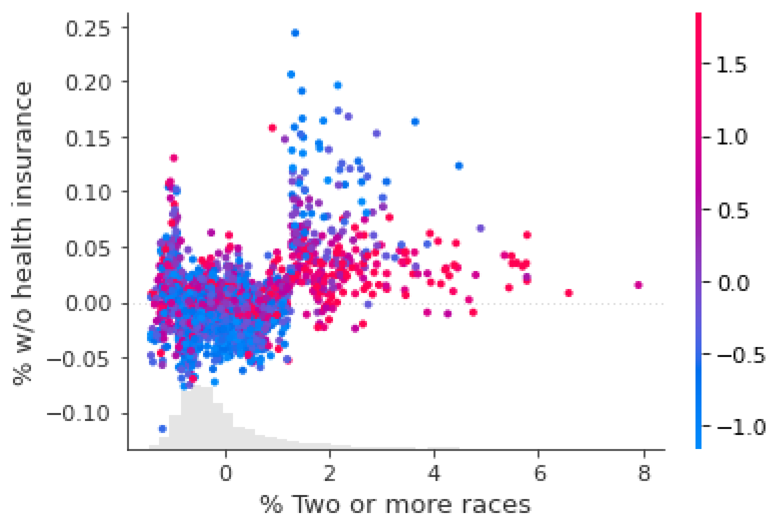
Note. The graph shows a scatterplot between the two variables "% w/o health insurance" (x-axis) and "percentage of mobile homes" (left y-axis). The values on both axes represent the normalized feature values, thus negative values or values exceeding 100% can appear. The color indicates the impact (based on the SHAP-value) on the target variable (vaccination attainment index), and the legend can be found on the right-hand side. A blue color indicates that it impacts the target negatively, i. e., better vaccination attainment, while a red color indicates a positive impact on the vaccination attainment index ("red is bad"). The grey histogram on the bottom shows the distribution of the x-values. If, like here, colors cannot be clearly separated, it indicates an interaction between those two variables. Tree-based models like XGBoost consider those interactions. The SHAP-module automatically selects the most interesting interaction for a given feature.

This interaction shows a positive relationship between the percentage of individuals without health insurance coverage and the percentage of mobile homes in an area. It indicates that areas with more mobile homes tend to have more people without health insurance coverage. Mobile home communities may face unique challenges in accessing healthcare services due to their transient nature, limited resources, or lack of infrastructure. The higher Spearman and Kendall's tau coefficients suggest that the relationship is not entirely linear and is more accurately captured by rank correlations.

Interaction 2 (Figure 3): Percentage of persons with Two or More Races and percentage of persons without Health Insurance Coverage, coefficients: Pearson (0.31), Spearman (0.20), Kendall's tau (0.14).

Figure 3:

Interaction Percentage of persons with Two or More Races and percentage of persons without Health Insurance Coverage

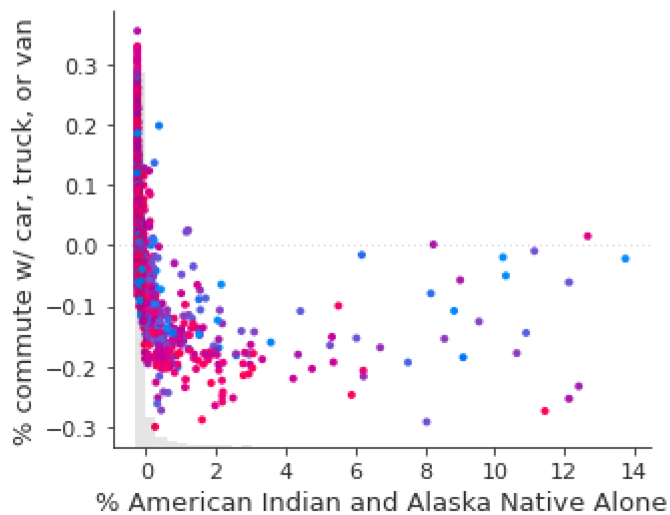


This interaction reveals a positive relationship between the percentage of persons of two or more races and those without health insurance coverage. This finding suggests that mixed-race populations may experience higher healthcare access challenges or face unique barriers in obtaining health insurance coverage. The lower Spearman and Kendall's tau coefficients indicate that the relationship may not be as strong as the Pearson coefficient suggests.

Interaction 3 (Figure 4): Percentage of American Indian and Alaska Native Alone and percentage of Workers 16 Years and Over, using a Car, Truck, or Van to go to work, coefficients: Pearson (-0.08), Spearman (-0.19), Kendall's tau (-0.13).

Figure 4:

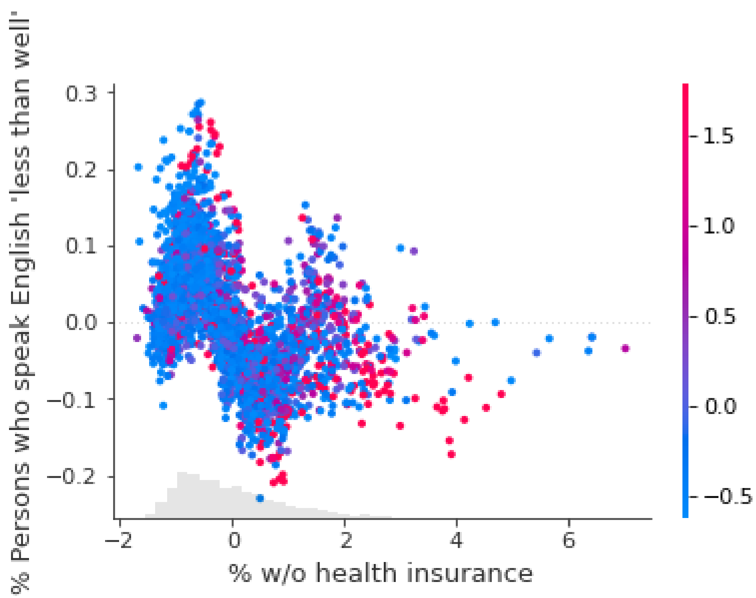
Interaction Percentage of American Indian and Alaska Native Alone and percentage of Workers 16 Years and Over, using a Car, Truck, or Van to go to work



This interaction shows a weak negative relationship between the percentage of American Indian and Alaska Native Alone individuals and workers using a car, truck, or van for their daily commute. It could imply that American Indian and Alaska Native communities may rely more on other modes of transportation, such as walking or public transit, or face transportation challenges that impact their ability to access healthcare services, including vaccination. The weak negative correlation suggests this relationship is weaker than other interactions.

Interaction 4 (Figure 5): Percentage of persons without Health Insurance Coverage and Percentage of persons (age 5+) who speak English “less than well” estimate, coefficients: Pearson (0.33), Spearman (0.26), Kendall’s tau (0.18).

Figure 5:
Interaction Percentage of persons without Health Insurance Coverage and Percentage of persons (age 5+) who speak English “less than well”

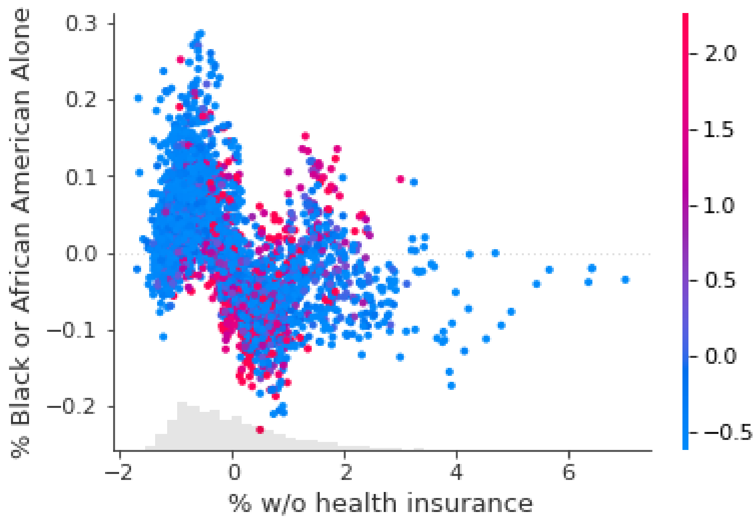


This interaction indicates a positive relationship between the percentage of persons without health insurance coverage and those who speak English “less than well.” It suggests that language barriers can significantly impact healthcare access and contribute to a higher rate of people without health insurance coverage. Limited English proficiency may make it difficult for individuals to navigate healthcare systems, understand health insurance options, and access vaccination services. The lower Spearman and Kendall’s tau coefficients indicate that the relationship may not be linear.

Interaction 5 (Figure 6): Percentage of persons without Health Insurance Coverage and percentage of Black or African American Alone, coefficients: Pearson (0.15), Spearman (0.17), Kendall's tau (0.11).

Figure 6:

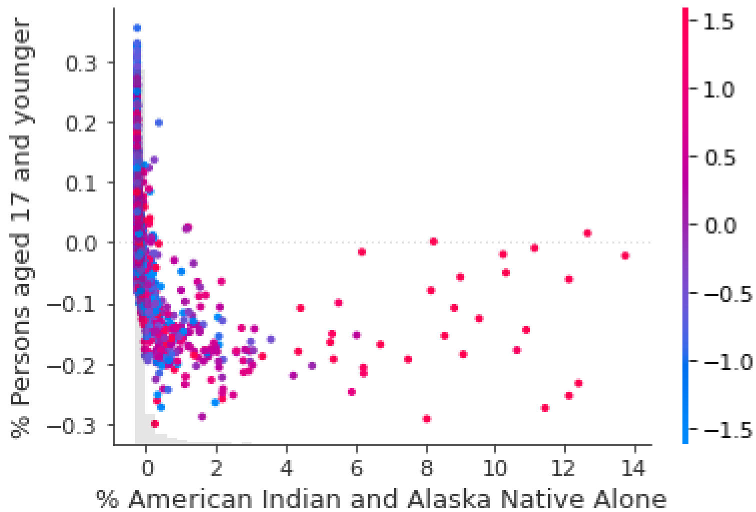
Interaction Percentage of persons without Health Insurance Coverage and percentage of Black or African American Alone



This interaction shows a weak positive relationship between the percentage of persons without health insurance coverage and the percentage of Black or African American Alone individuals. This finding implies that Black or African American communities may face unique challenges in accessing healthcare services, including obtaining health insurance coverage. The weak correlation coefficients suggest that the relationship is not as strong or direct as other interactions.

Interaction 6 (Figure 7): Percentage of American Indians and Alaska Native Alone and the Percentage of persons aged 17 and younger estimate, coefficients: Pearson (0.28), Spearman (0.10), Kendall's tau (0.07).

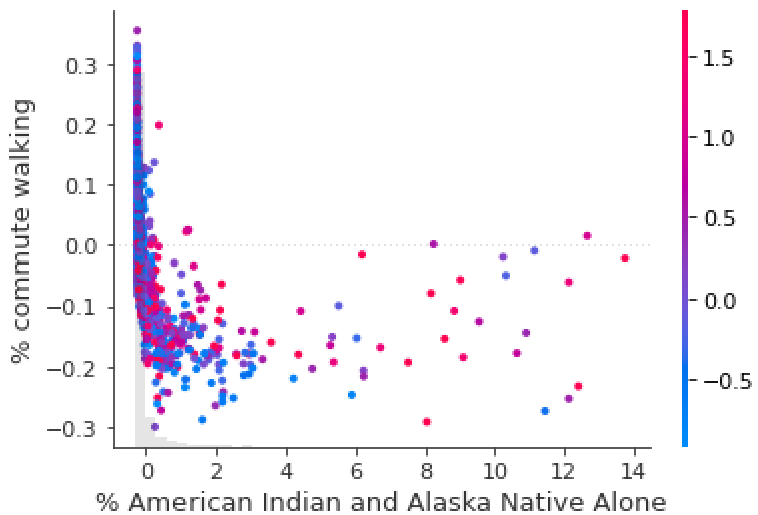
Figure 7:
Interaction Percentage of American Indians and Alaska Native Alone and the Percentage of persons aged 17 and younger



This interaction reveals a positive relationship between the percentage of American Indian and Alaska Native Alone individuals and the percentage of persons aged 17 and younger. The positive correlation indicates that areas with a higher percentage of American Indian and Alaska Native Alone individuals may have a higher proportion of young people. However, the lower Spearman and Kendall's tau coefficients suggest that the relationship is not as consistent as the Pearson coefficient implies.

Interaction 7 (Figure 8): Percentage of American Indian and Alaska Native Alone and percentage of Workers 16 Years and Over who walk to work, coefficients: Pearson (0.14), Spearman (0.18), Kendall's tau (0.12).

Figure 8:
Interaction Percentage of American Indian and Alaska Native Alone and percentage of Workers 16 Years and over who walk to work

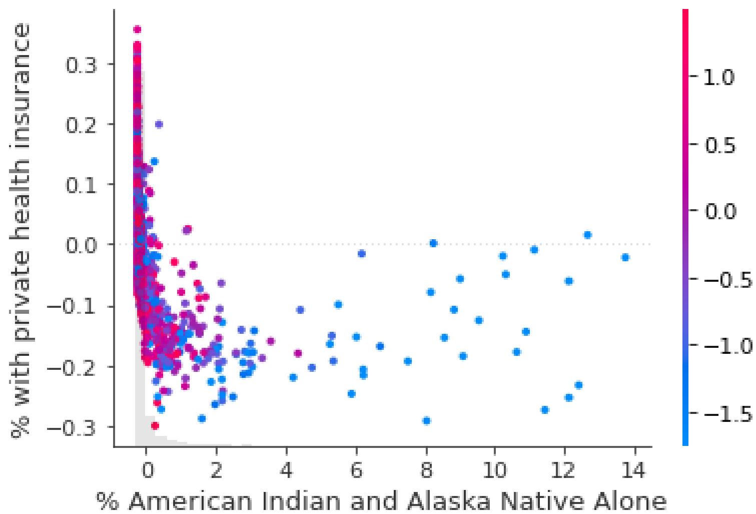


This interaction demonstrates a weak positive relationship between the percentage of American Indian and Alaska Native Alone individuals and the percentage of workers who walk to work. It suggests that these communities may be more likely to use walking as a mode of transportation. The weak correlation coefficients imply that this relationship is not as strong or direct as other interactions.

Interaction 8 (Figure 9): Percentage of American Indians and Alaska Native Alone and percentage of people with Private Health Insurance, coefficients: Pearson (0.14), Spearman (0.18), Kendall's tau (0.12).

Figure 9:

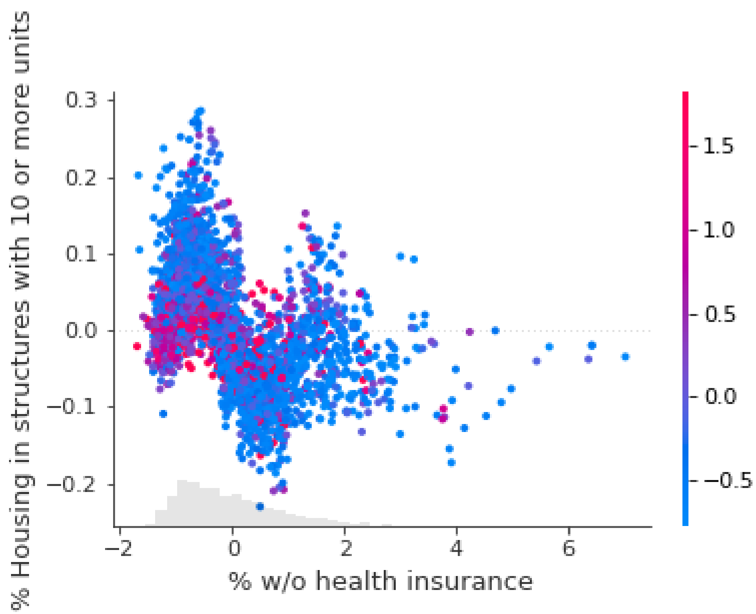
Interaction percentage of American Indians and Alaska Natives alone and percentage of people with private health insurance



This interaction shows a negative relationship between the percentage of American Indian and Alaska Native Alone individuals and the percentage of people with private health insurance. This finding suggests that American Indian and Alaska Native communities may face challenges obtaining private health insurance, which could impact their healthcare access, including vaccination services. However, the low Spearman and Kendall's tau coefficients indicate that the relationship may not be as consistent as the Pearson coefficient suggests.

Interaction 9 (Figure 10): Percentage of persons without Health Insurance Coverage and Percentage of housing in structures with ten or more units estimate, coefficients: Pearson (-0.17), Spearman (-0.27), Kendall's tau (-0.19).

Figure 10:
Interaction Percentage of persons without Health Insurance Coverage and Percentage of housing in structures with ten or more units

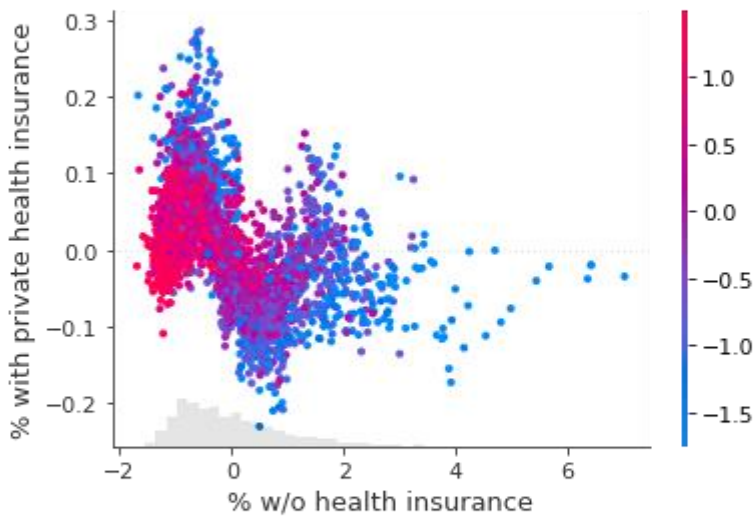


This interaction indicates a weak negative relationship between the percentage of persons without health insurance coverage and the percentage of housing in structures with ten or more units. It suggests that areas with larger housing structures may have fewer people without health insurance coverage. This relationship could be attributed to differences in socioeconomic factors, such as income levels, affecting housing choices and health insurance coverage. The higher Spearman and Kendall's tau coefficients imply that rank correlations might better capture the relationship.

Interaction 10 (Figure 11): Percentage of persons without Health Insurance Coverage and percentage of persons with Private Health Insurance, coefficients: Pearson (-0.59), Spearman (-0.58), Kendall's tau (-0.41).

Figure 11:

Interaction Percentage of persons without Health Insurance Coverage and percentage of persons with Private Health Insurance



This interaction shows a strong negative relationship between the percentage of persons without health insurance coverage and those with private health insurance. This finding is expected, as an increase in private health insurance coverage would lead to fewer individuals without health insurance coverage. The strong negative correlation coefficients across Pearson, Spearman, and Kendall's tau indicate a consistent and monotonic relationship.

4.1.3.2 Interaction themes

The top 10 interactions identified using SHAP values highlight the complex relationships between socioeconomic factors and their impact on vaccination rates. Key themes include health insurance coverage and the unique challenges faced by American Indian and Alaska Native Alone populations.

Health insurance coverage: Interactions 1, 2, 4, 5, 9, and 10 involve the percentage of persons without health insurance coverage. These interactions highlight healthcare access and affordability disparities, which can significantly impact vaccination rates. Individuals without health insurance might face financial barriers to accessing vaccination services or be more hesitant to seek healthcare due to cost concerns.

Interactions 1, 9, and 10 involve the relationship between health insurance coverage and mobile homes, housing structures, and private health insurance, respectively. These interactions emphasize the disparities in healthcare access and affordability. Communities with a higher proportion of mobile homes or housing structures with ten or more units might face unique challenges in accessing healthcare services, including vaccinations.

Interactions 2, 4, and 5 reveal that mixed-race, non-English speaking populations and Black or African American populations may face unique challenges in accessing healthcare services. These interactions emphasize the importance of addressing racial, ethnic, and linguistic disparities to improve vaccination rates.

American Indian and Alaska native alone population: Interactions 3, 6, 7, and 8 involve the percentage of American Indian and Alaska Native Alone persons. These interactions highlight how this particular population may face unique challenges in accessing healthcare services, including vaccination.

Interactions 3 and 7 suggest that transportation options may be limited in communities with higher percentages of American Indian and Alaska Native Alone persons, impacting their access to vaccination services.

Interaction 6 reveals that American Indian and Alaska Native Alone communities may have higher proportions of younger individuals with unique vaccination needs and healthcare access challenges.

Interaction 8 highlights a negative correlation between the percentage of American Indian and Alaska Native Alone persons and the percentage of people with private health insurance, suggesting that these communities may face disparities in healthcare access and affordability, influencing vaccination rates.

In summary, our understanding of these complex interactions and our analysis of both spatial and socioeconomic features has revealed the importance of factors such as race, health insurance coverage, language proficiency, gender, transportation infrastructure, commute times, and the density of medical facilities in shaping equitable access to vaccine distribution. The interplay between these socioeconomic and spatial dimensions highlights the complex nature of vaccine distribution equity and the need to address disparities and barriers different population groups face. As we move forward, these identified features serve as valuable predictors for modeling the vaccination attainment index on the tract level.

By incorporating these insights and understanding the relationships between these factors, we developed a more accurate and robust predictive model that can help policymakers and public health officials identify areas needing targeted interventions and work towards achieving better equity in vaccine distribution.

4.2 Model Validation on county and tract levels

After selecting the relevant features using SHAP, the model was trained at the county level using these selected features to predict the vaccination attainment index at the tract level. This approach aimed to identify the most critical factors contributing to the vaccination attainment index and to develop a more accurate and robust predictive model. Our cross-validation technic on the model demonstrated that we have a stable model based on MSE and MAE metrics across the folds, i. e., validation runs (see 3.5), as shown in Figure 12 and Figure 13 below:

Figure 12:
Mean Absolute Error in the county cross-validation

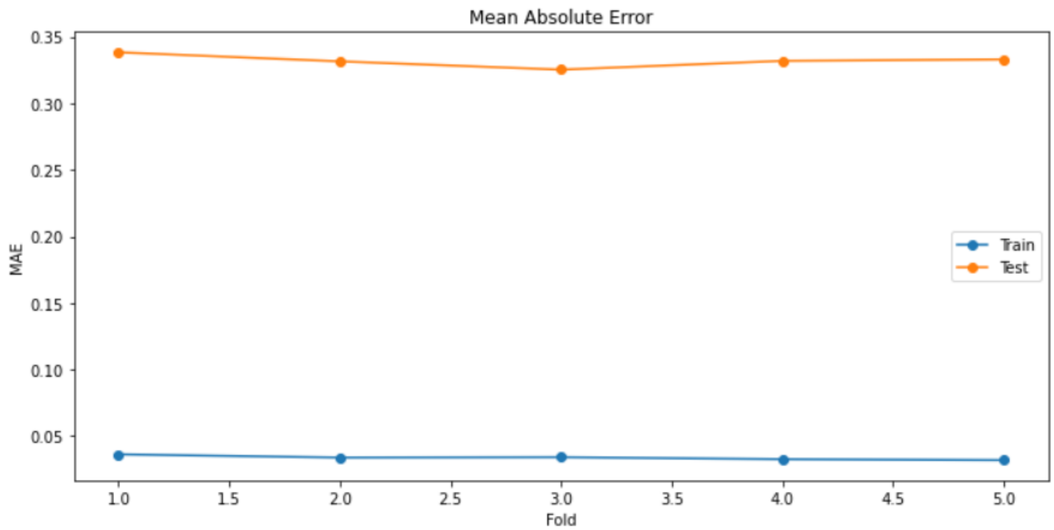
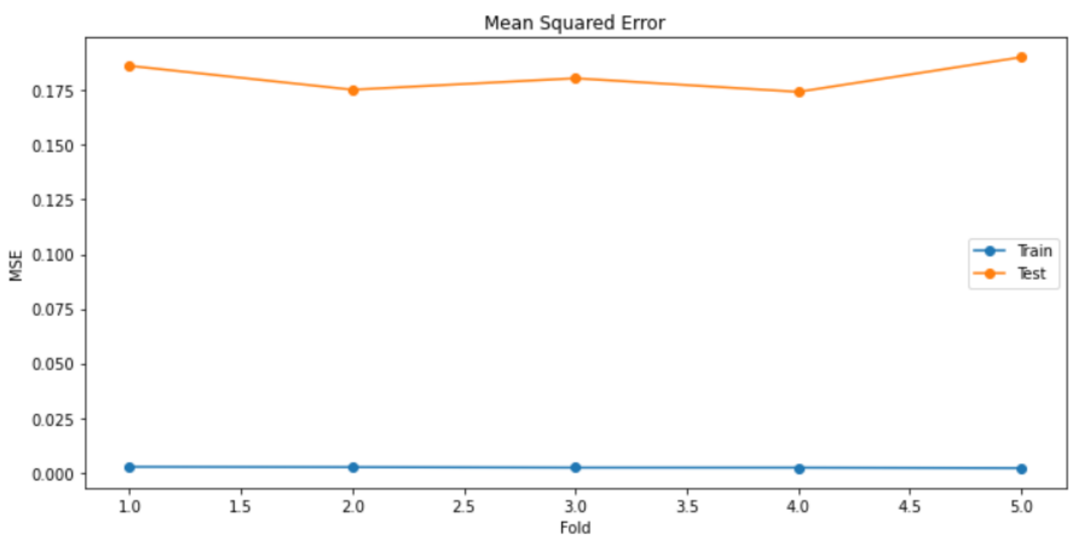


Figure 13:
Mean Squared Error in the county cross-validation



Once the model was trained, it was applied to predict the vaccination attainment index for each tract. The predicted index values were aggregated using a weighted average based on the population of each tract to evaluate the model's performance at the county level. This approach ensures that tracts with larger populations contribute more to the aggregated county-level index, thus providing a more representative assessment of the vaccination attainment index within each county.

The predicted county-level vaccination attainment index was then compared to the actual index to assess the model's accuracy and performance. Three evaluation metrics were used to quantify the model's performance: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Table 6:
Overview of model valuation metrics

Metric	Value	Interpretation
MAE	0.37	This metric represents the average absolute difference between the predicted and actual index values. The lower the MAE, the better the model's performance. An MAE of 0.37 indicates a moderate level of accuracy in the model's predictions considering the range of 1.61 (min: -1.59, max: 0.02) of the vaccination attainment index.
MSE	0.21	MSE of 0.21 indicates that the average squared difference between the predicted and actual index values is relatively low, suggesting that the model has a moderate level of accuracy in predicting the vaccination attainment index at the county level. However, it is essential to note that the MSE is sensitive to outliers, and more significant errors will significantly impact this metric.
RMSE	0.46	This metric is the square root of the MSE and represents the standard deviation of the residuals (prediction errors). The RMSE is more interpretable than the MSE, as it has the same unit as the dependent variable. An RMSE of 0.46 indicates a moderate level of accuracy in the model's predictions considering the range of 1.61 (min: -1.59, max: 0.02) of the vaccination attainment index.

Note. Compiled by the authors. Copyright 2023 by the authors.

These evaluation metrics reveal that the model reasonably approximates the vaccination attainment index at the tract level, although there is room for improvement. The selected features and their

SHAP importance values played a crucial role in the model's performance. By refining the model and incorporating more relevant features or optimizing the model's parameters, it may be possible to improve the accuracy of the predictions further and better understand the factors driving the vaccination attainment index.

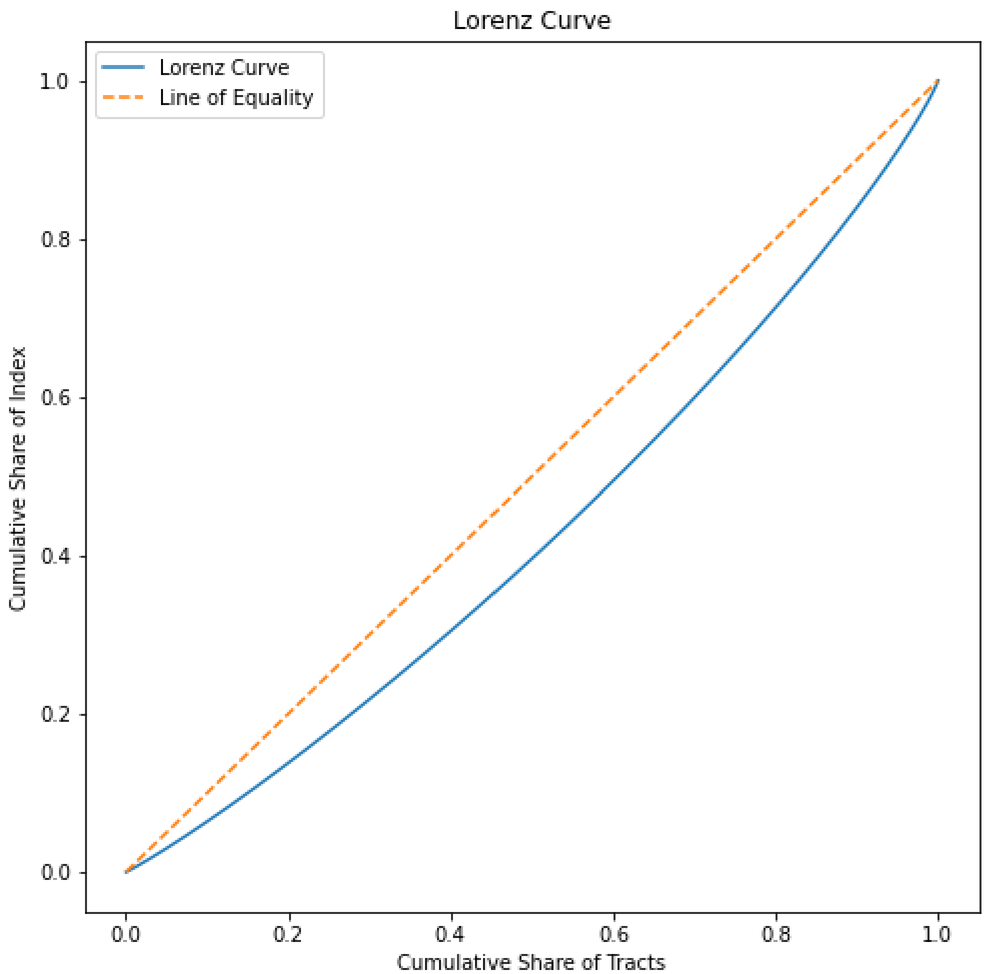
4.3 Equity indices and their performance

In our study, the Gini index for vaccination at the county level is 0.27, and at the tract level, it is 0.15. The higher Gini index at the county level indicates a greater degree of inequality in vaccination distribution compared to the tract level. Our results show larger disparities between counties than within counties (across tracts).

The lower Gini index at the tract level, as seen on the Lorenz curve in Figure 14, may indicate that the variation of vaccination attainment is smaller within individual counties. However, the disparities become more pronounced in the entire population across counties. A Theil index of 0.13 on the county level confirmed this, while on the tract level, we obtained 0.04.

Figure 14:

The Gini index on all United States tracts



Note. The Lorenz curve (blue curve) in this graph represents the distribution of our vaccination attainment index in all the tracts. The 45-degree line (orange line), also known as the line of equality, depicts perfect equality where each percentage of the tracts attains the same level of vaccination attainment index. The farther the Lorenz curve bows away from this line, the greater the inequality.

By analyzing the Gini Index and the Theil Index values alongside real-world examples from the United States, we can understand how the previously discussed factors might influence the disparities in vaccination distribution:

Health insurance coverage: In states like Texas, which has one of the highest uninsured rates in the nation, limited access to healthcare might hinder vaccination rates. The higher Gini Index value at the county level indicates that the disparity in vaccination rates between different counties may be linked to the variation in health insurance coverage in each county.

American Indian and Alaska native alone population: Regions with higher concentrations of American Indian and Alaska Native Alone populations, such as parts of Arizona and the Dakotas, face unique healthcare access and affordability challenges. The difference in Gini Index values between county and tract levels implies that these challenges might be more pronounced at the broader geographic scale (county) than within the smaller, localized areas (tract).

Transportation options: In rural communities, transportation options may be limited, making it difficult for residents to access vaccination services. For example, rural counties in states like Montana or Wyoming might experience more significant disparities in vaccination rates compared to more urbanized counties. The higher Gini Index value at the county level suggests this factor might significantly contribute to the disparities between different counties.

The disparities in vaccination distribution across counties and tracts in the United States reflect the complex interactions between socioeconomic and spatial factors. By examining real-world examples and the Gini Index values for the vaccination attainment index, we can better understand the factors driving these disparities and devise strategies to address them.

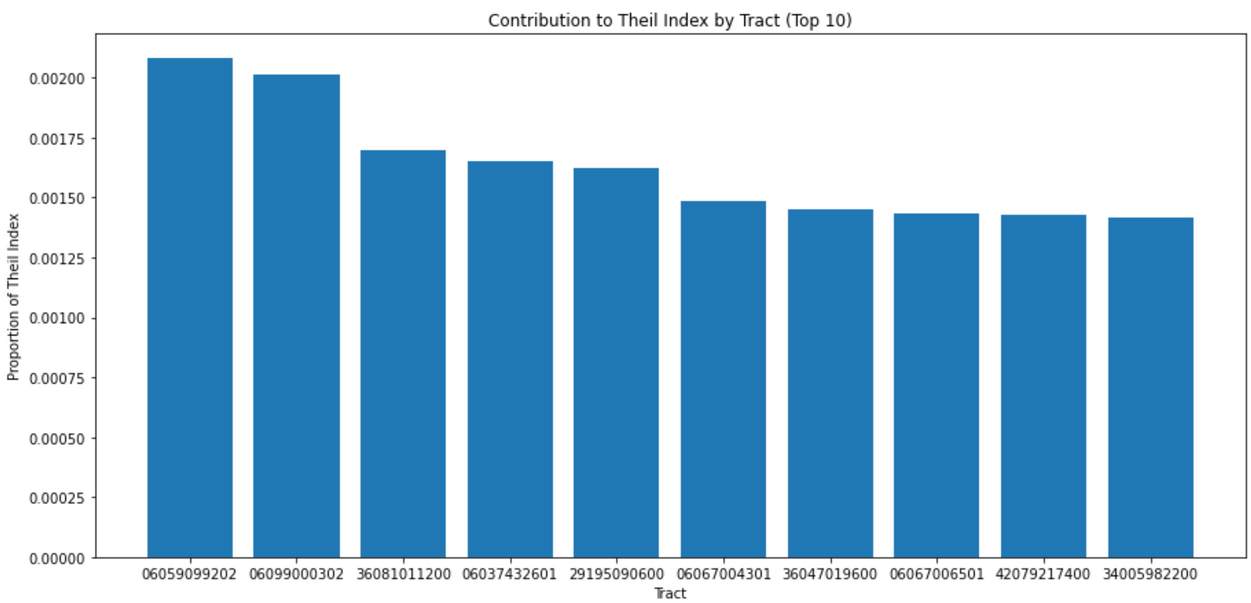
Comparing the Gini and Theil Index values for the vaccination data, both indices show a similar trend: greater inequality exists between counties than within tracts. However, it is crucial to consider that the Gini Index reacts stronger to changes in the center of the data than at the tails, whereas the Theil Index is sensitive to changes across the entire distribution (Gastwirth, 2017).

In the context of vaccination distribution, the Gini Index may better capture disparities caused by health insurance coverage and racial/ ethnic disparities, which are more prevalent in the middle of the distribution. On the other hand, the Theil Index could better capture disparities resulting from extreme values, such as isolated rural communities with poor access to healthcare services.

4.4 Scenario analysis on all the tracts of the United States

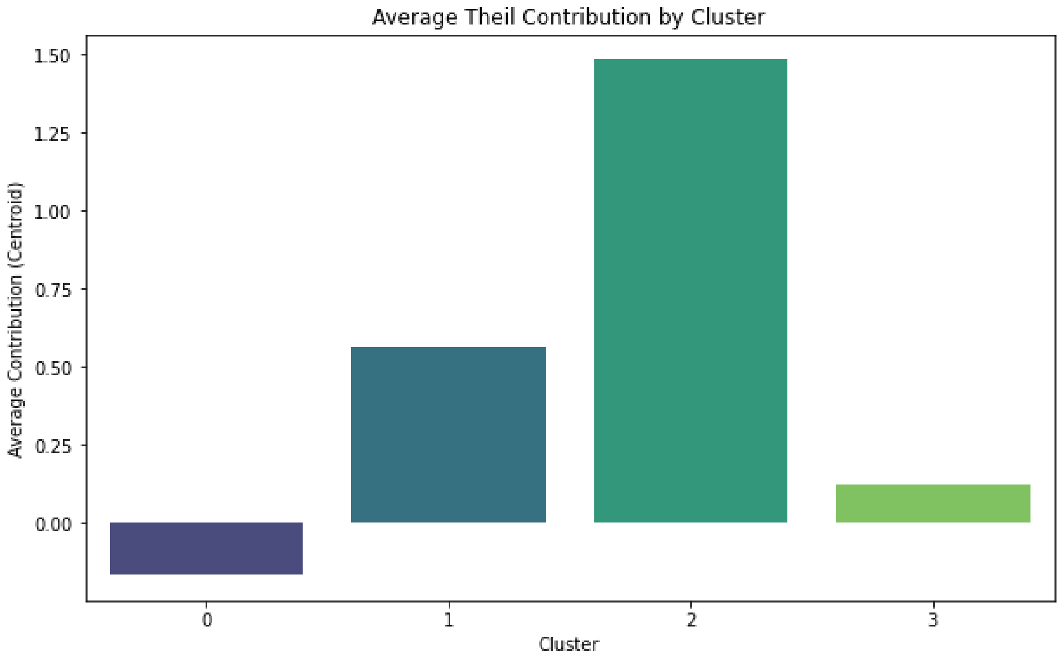
When looking at all the tracts in the United States, we can cluster them according to their contribution to the Theil index. Looking at Figure 15, we can see the contribution of the top 10 tracts to inequity.

Figure 15:
Top ten tracts contributing to inequality measured through the Theil index



However, to better understand the behavior of all more than 70’000 tracts, we clustered them based on the Theil contributions in four clusters using the K-means and Elbow method discussed in the methodology section. As shown in Figure 16, clusters 1 (22689 tracts) and 2 (7668 tracts) contribute most to inequity, which is not the case for cluster 0 (38567 tracts), which contributes negatively (reduces the impact on inequity) to better equity. Cluster 3 (1251 tracts) has a moderately low impact on inequity. This can be interpreted by comparing the height of the bar, representing the average proportion of the overall Theil index contributed by that tract. A higher bar indicates that the tract contributes more to the inequity, while a lower bar implies a smaller contribution.

Figure 16:
The average contribution of all United States tracts by cluster

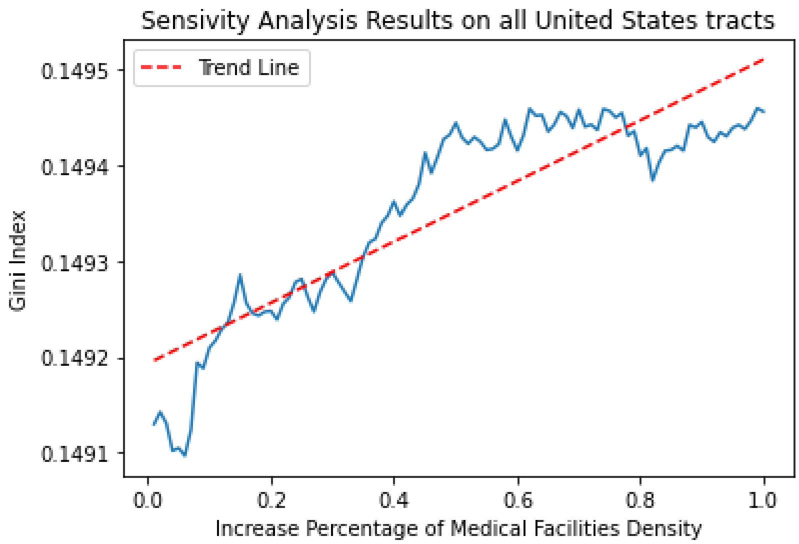


4.4.1 Adjusting the density of medical facilities in all the tracts uniformly

We conducted a scenario analysis that involved increasing the density of medical facilities equally across all tracts in the United States and calculating the Gini Index for the vaccination attainment index after each increment. We observed an increasing trend in the Gini Index, indicating that inequality in the vaccination attainment index rises as medical facility density increases (see Figure 17).

Figure 17:

The Gini index in the scenario with an equal increase of density in all the United States tracts



This outcome may seem counterintuitive, but it can be explained by the increased medical facility density applied equally across all tracts, regardless of their specific needs. This approach emphasizes equality rather than equity. While equality aims to provide the same resources to everyone, equity focuses on distributing resources based on the unique needs of each community.

The increasing Gini Index could be a result of the following factors:

Unaddressed disparities: By increasing medical facility density equally, the scenario fails to address pre-existing disparities among tracts, such as differences in socioeconomic conditions, access to transportation, or health insurance coverage. The needs of particular communities may be overlooked, leading to widening inequality in the vaccination attainment index.

Resource inefficiency: In regions with sufficient medical facilities, further increasing the density might not significantly improve the vaccination attainment index. On the other hand, regions with higher needs might still face a shortage of medical facilities, leading to an unequal distribution of vaccination resources.

Potential over-provision: Applying the same increase across all tracts might result in over-provision of medical facilities in certain areas. These additional resources might not substantially improve the vaccination attainment index, especially if other factors such as education, income, or cultural beliefs are more significant barriers to vaccination.

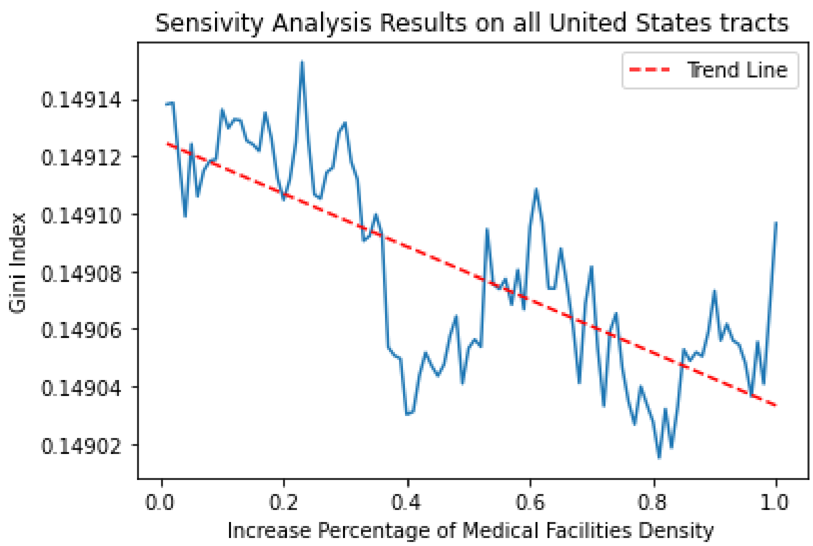
4.4.2 Adjusting proportionally density of medical facilities in tracts on population density

It would be more effective to tailor the increase in medical facility density based on the specific needs and challenges each tract faces to address these issues and achieve better equity in vaccination coverage. In the second scenario, we increased the density of medical facilities proportionally to the population density across all tracts in the United States.

By calculating the Gini Index and plotting the results, we observed a decreasing trend in the Gini Index, indicating that inequality in the vaccination attainment index is reducing as medical facility density increases in this manner (see Figure 18).

Figure 18:

The Gini index of the scenario with a proportional increase based on population density in the United States



This outcome can be explained by the fact that the increase in medical facility density was more targeted, considering each tract's population density. This approach is more aligned with the concept of equity, focusing on distributing resources based on the unique needs of each community.

The decreasing Gini Index in this scenario can be attributed to the following factors:

Addressing disparities: By increasing medical facility density proportionally to the population density, the scenario is better suited to address pre-existing disparities among tracts, such as differences in socioeconomic conditions, access to transportation, or health insurance coverage. This approach helps better to meet the needs of communities with higher population densities, reducing inequality in the vaccination attainment index.

Resource efficiency: By allocating medical facilities in proportion to population density, the resources are being used more efficiently. Regions with higher population density receive more medical facilities, which can lead to a more significant improvement in the vaccination attainment index. Conversely, regions with lower population density receive fewer additional facilities, avoiding potential over-provision and resource wastage.

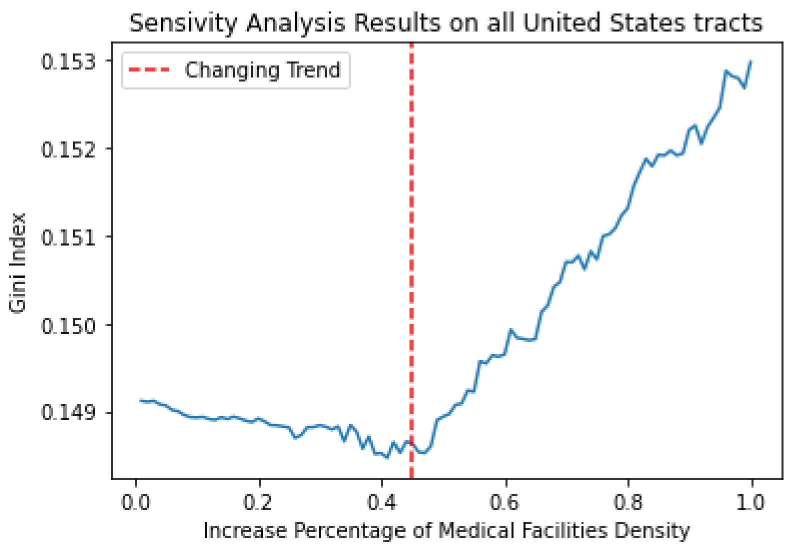
Comparing the two scenarios, it is evident that the second scenario is more effective in reducing inequality in the vaccination attainment index. The second scenario accounts for population density, an essential factor in addressing the unique needs of each community and promoting better equity in vaccination coverage. By considering the findings from our previous analysis, such as the impact of health insurance coverage, racial and ethnic disparities, and transportation options, we can develop even more targeted interventions to further reduce inequality in the vaccination attainment index.

4.4.3 Adjusting proportionally density of medical facilities in tracts on average time to commute

In the third scenario, we increased the density of medical facilities proportionally to the average time to commute across all tracts in the United States. Upon calculating the Gini Index and plotting the results, we observed that the Gini Index decreased for the increasing percentage up to 45% and then went up drastically (see Figure 19). It indicates that the relationship between the allocation of medical facilities based on the average commute time and vaccination attainment index inequality is more complex than in the previous scenarios.

Figure 19:

The Gini index in the scenario with a proportional increase based on the average time to commute in the United States



The initial decrease in the Gini Index up to the 45% increase suggests that allocating medical facilities based on average commute times can help reduce inequality in the vaccination attainment index to some extent. This approach addresses the accessibility issue for communities with longer commute times, potentially improving vaccination rates in areas where transportation is a significant barrier.

However, the drastic increase in the Gini Index after the 45% increase point implies that allocating medical facilities solely based on average commute time may not be the most efficient strategy for reducing inequality in the vaccination attainment index. Allocating resources based on commute times alone may lead to overemphasizing areas with longer commutes, neglecting other factors contributing to disparities in vaccination rates, such as population density, health insurance coverage, and socioeconomic conditions.

4.4.4 Conclusion

Compared to simply proving all tracts with the same number of medical facilities, a more balanced approach makes it clear that considering multiple factors contributing to vaccination disparities will likely yield better results in reducing inequality. Allocating medical facilities based on population density produced a more consistent decrease in the Gini Index, suggesting a more effective resource allocation strategy. In conclusion, resource allocation should consider multiple factors, including population density, health insurance coverage, and transportation options, to maximize the effectiveness of interventions and ensure that resources are utilized efficiently.

4.5 Cambridge, MA, case study

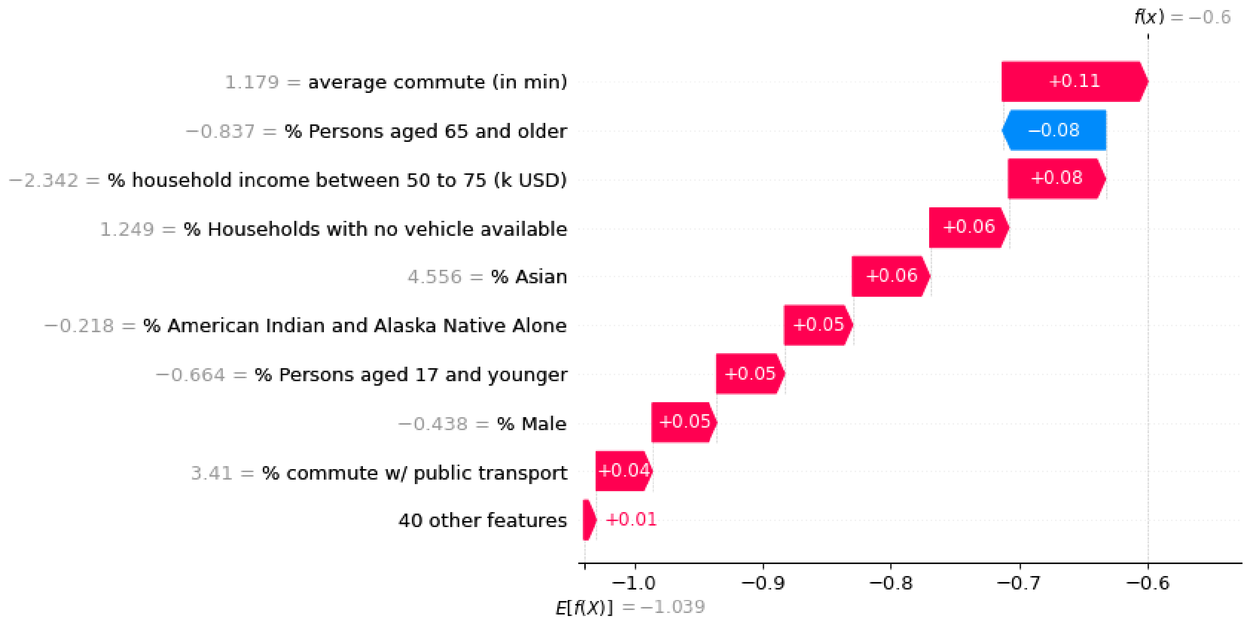
After running our scenarios on all the tracts of the United States, we focused on a smaller scope to see the impact in a specific area. In our case, we selected Cambridge, MA, as a geographic area to measure the impact of adjusting the density of medical facilities on the tracts within the city and how it would improve the equity of access.

4.5.1 Relevant Features

Before trying to improve equity of vaccine access in our case study on Cambridge, MA, we looked at the relevant features of Middlesex County (county of Cambridge tracts) compared to all the counties in the United States we defined in the previous sections. In Middlesex County, the SHAP waterfall plot (Figure 20) highlights several factors influencing the vaccination attainment index, along with their associated contributions:

Figure 20:

SHAP waterfall plot of the relevant factors in Middlesex County



Note. The SHAP-waterfall plot shows the impact of certain features on the target (vaccination attainment index). A red bar indicates a positive influence. E. g., the average commute to work duration increases the vaccine attainment index by 0.12. A blue bar indicates a negative effect. Since tree-based models, like XGBoost, capture non-linearities and interactions, the effects can differentiate between samples (see 4.5.2 for an example). Those plots can be made for the explanatory model as well as for predictions made.

- **The average commute to work [+0.11]:** Longer commutes may be associated with a higher vaccination attainment index, possibly because individuals with longer commutes may have less flexible schedules, making it more challenging to schedule vaccination appointments, thus slowing down the vaccination process.
- **Percentage of persons aged 65 and older estimate [-0.08]**: A higher percentage of elderly individuals may be associated with a lower vaccination attainment index. Older populations are often prioritized for vaccinations due to their higher vulnerability to severe COVID-19 outcomes, leading to faster vaccination rates for this group.
- **Percentage of households income from $50,000 to $74,999 [+0.08]**: A higher percentage of households within this income bracket might be related to a higher vaccination attainment index,

suggesting that people in this income range may face challenges in accessing vaccination services or may have different attitudes toward vaccination compared to other income groups.

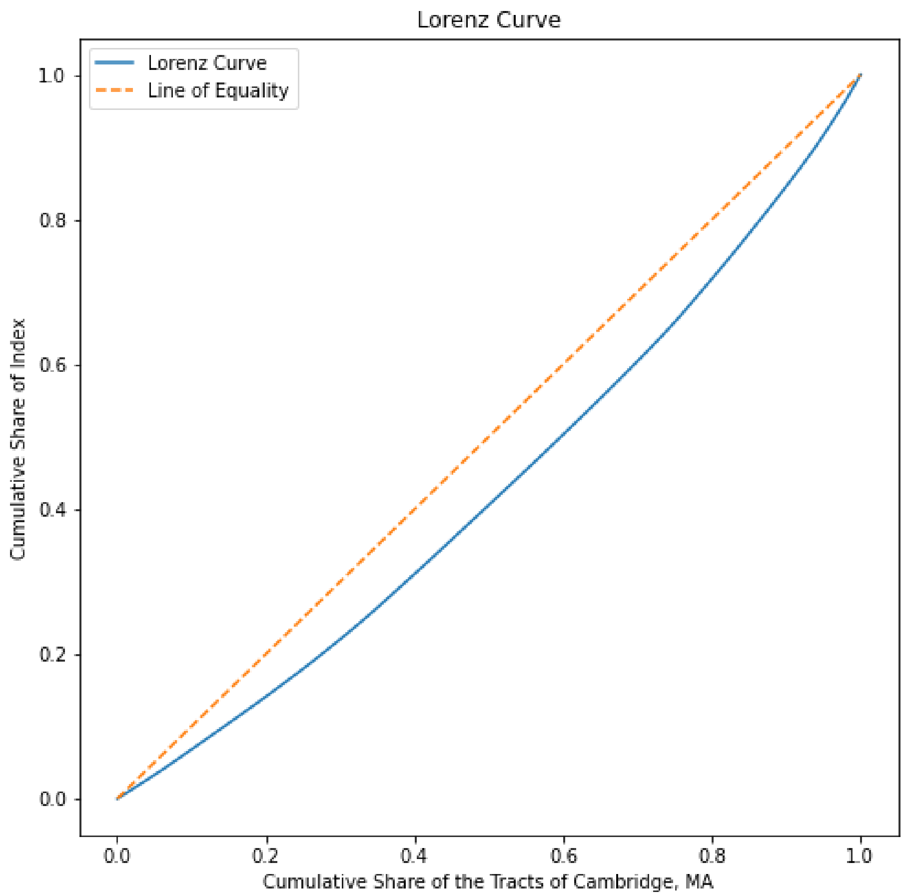
- **Percentage of households with no vehicle available estimate [+0.06]:** A higher percentage of households with no vehicle available is positively associated with the vaccination attainment index, suggesting poorer vaccination performance in areas where more households lack access to a vehicle. A reason may lie in difficulties in reaching vaccination sites for those without personal transportation, as public transportation options may be limited or inconvenient for some individuals.
- **Percentage of American Indian and Alaska native alone [+0.06]**, and **Percentage of Asian Alone persons [+0.06]**: A higher percentage of these racial/ ethnic groups might be associated with a higher vaccination attainment index, which could be due to language barriers, limited access to vaccination sites, or cultural factors that affect vaccination uptake in these communities.
- **Percentage of males [+0.05]**: A higher percentage of males might be linked to a higher vaccination attainment index, which could be related to factors such as men's health-seeking behaviors, job types, or other socioeconomic factors that make it more difficult for men to access vaccinations or prioritize getting vaccinated.

While some factors may be similar to those seen in other United States counties, some factors, such as health insurance coverage and housing structure, are not as prominent in Middlesex County. The importance of each factor may vary depending on the local context, suggesting that the factors driving vaccination rates may vary between regions.

4.5.2 Equity metrics

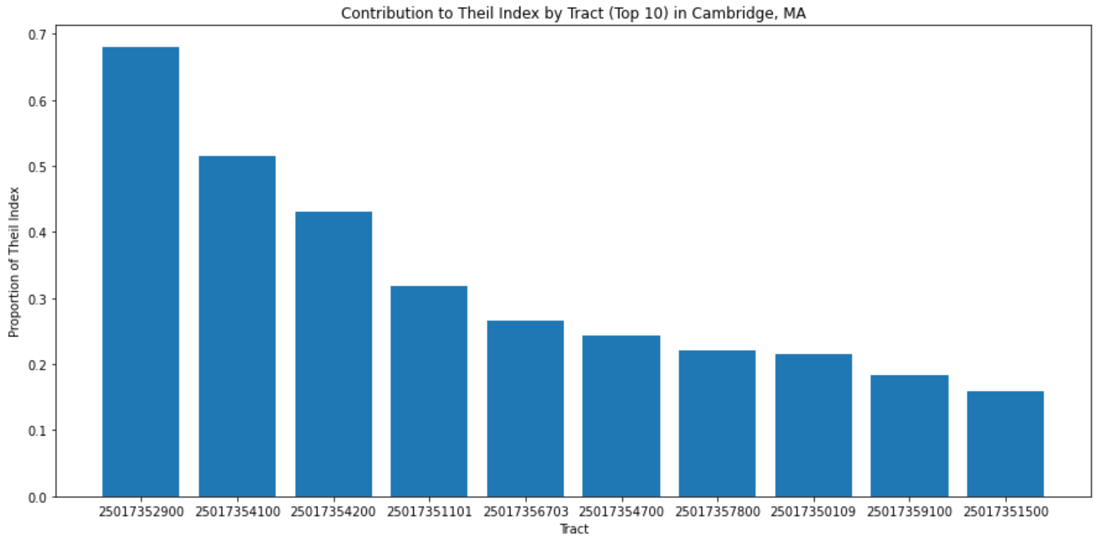
Cambridge, MA, had a Gini Index of 0.14 at the tract level, suggesting a relatively low inequity in vaccination attainment within the city, although some disparities still exist (see Figure 21). However, as shown on the Lorenz curve below, it has the same shape as the nationwide curve as the national tract-level Gini Index (0.15). Nevertheless, there is still room for improvement in addressing disparities.

Figure 21:
Lorenz Curve of Census Tracts in Cambridge, MA



We confirmed this observation in Cambridge, MA, by obtaining the Theil index, resulting in a value of 0.03. It shows a similar better position of the city of Cambridge compared to the nationwide index of 0.04 on all the tracts of the United States. Moreover, we can see in Figure 22 that despite the first four tracts, all the tracts contribute equally to the inequity based on the proportion of the total Theil index contributed by each tract.

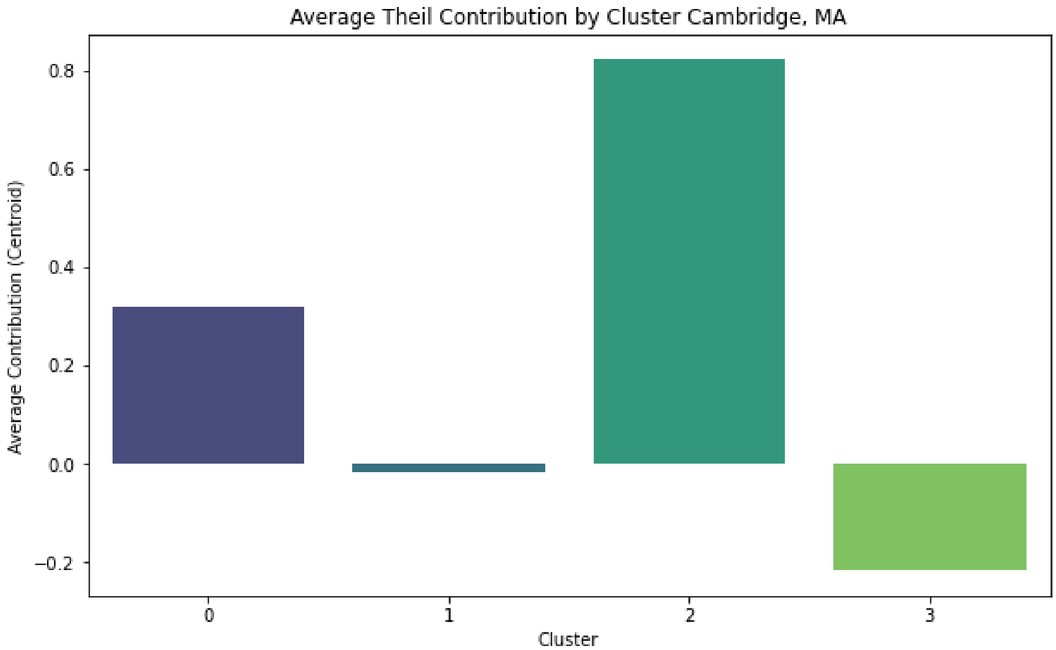
Figure 22:
Contribution to the Theil Index in Cambridge, MA



Based on the Theil Index contributions, we clustered based on the Theil contributions the tracts of Cambridge, MA, in four clusters using the K-means and Elbow method discussed in the methodology section. Furthermore, as shown in Figure 23, clusters 0 and 2 contribute most to inequity, which is not the case for cluster 3, which contributes negatively (reduces the impact on inequity) to better equity. This can be interpreted by comparing the height of the bar, representing the average proportion of the overall Theil index contributed by that tract. A higher bar indicates that the tract contributes more to the inequity, while a lower bar implies a smaller contribution.

Figure 23:

Clusters in Cambridge, MA, based on Theil's Contributions



These differences in the equity of vaccine distributions by tract are related to the fact that factors influencing the attainment of vaccination differ from one tract to another. If we take two tracts from Cambridge, MA, we can see that the top relevant features impacting our vaccination attainment index differ. We can see that in the following two graphs (Figure 24 and Figure 25) based on SHAP values. Moreover, even if we have standard features, the contributions for each one are not the same, as the entities' interactions and dynamics are different. Notably, according to the SHAP analysis for tract 25017359400, the density of medical facilities is important, while it is not for tract 25017359303.

Figure 24:
Relevant Features for the Tract 25017359400

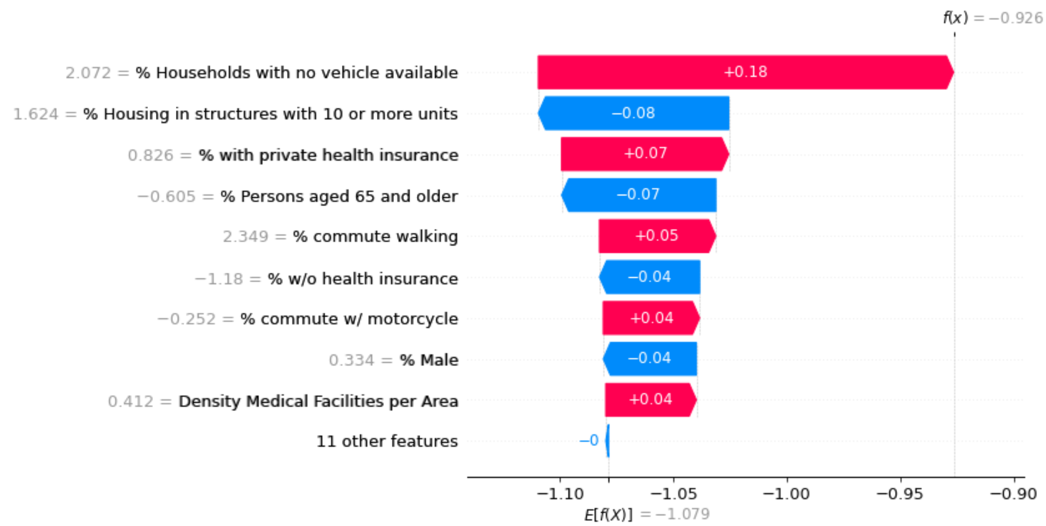
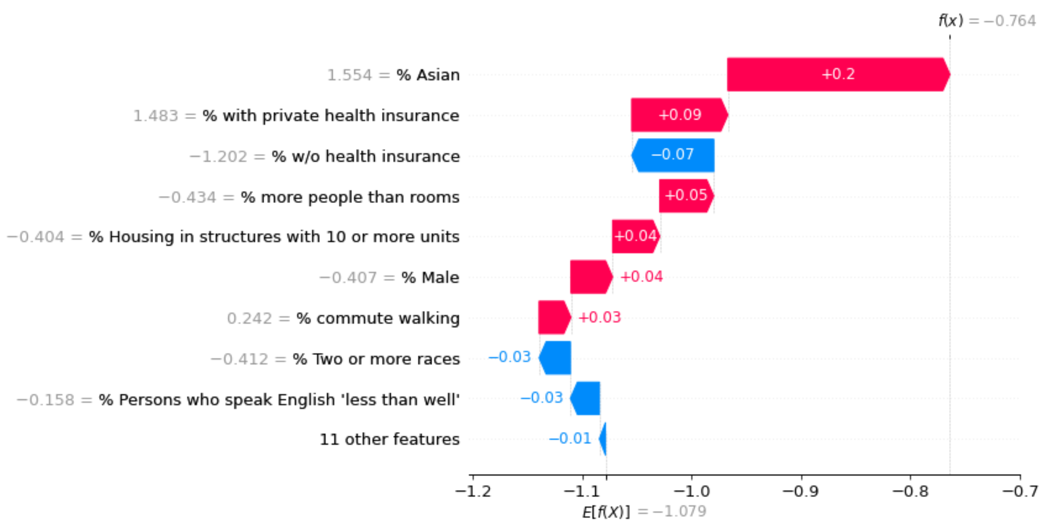
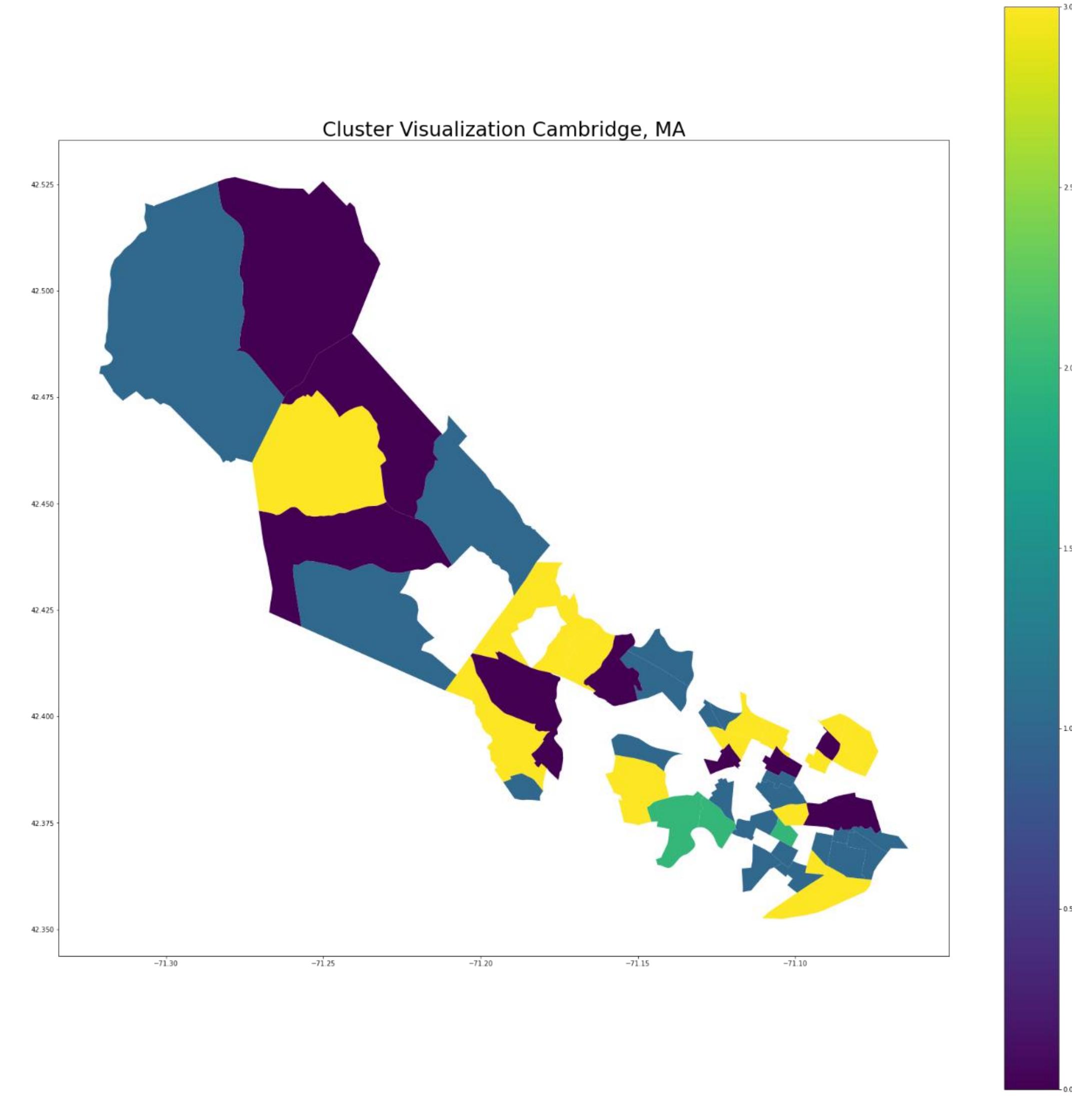


Figure 25:
Relevant Features for Tract 25017359303



The map below (Figure 26) shows where the clusters are concentrated. The tracts in light blue and green are the tracts that need specific attention regarding access to vaccination. As they impact the inequity of access, those tracts require a specific and targeted strategy of vaccination centers distribution to improve the vaccination rate and reduce the inequity between tracts.

Figure 26:
Clusters in Cambridge, MA, according to their Theil contribution



Note. This map shows the repartition of the clustered tracts in Cambridge based on their Theil contribution to inequity identified previously (see Figure 20). The yellow and green tracts negatively contribute to inequity (reducing inequity). The dark blue indicates the tracts contributing the most to inequity. The light blue represents those tracts that do not contribute to inequity.

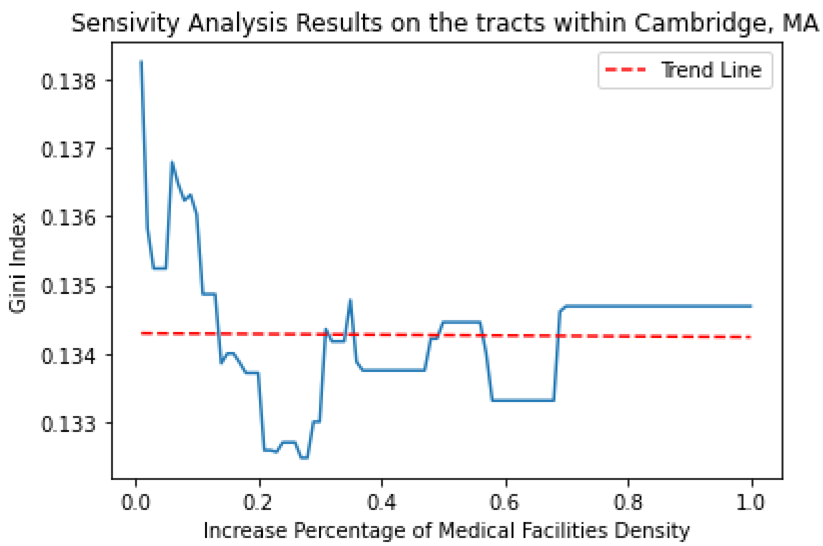
4.5.3 Scenario analysis on Cambridge, MA

After analyzing the as-is situation in Cambridge, MA, we sought to improve equity in vaccine access. We start with a uniform increase in the density of facilities (4.5.3.1), continue with a proportional adjustment of the density in relation to the population density (4.5.3.2), and conclude by adjusting the density of medical facilities proportionally to the average time to commute (4.5.3.3).

4.5.3.1 Adjusting the density of medical facilities in tracts within Cambridge, MA, uniformly

We uniformly increased medical facilities' density throughout all tracts in Cambridge, MA, with the Gini Index for the vaccination attainment index being calculated after each increment. Figure 27 shows a downward trend in the Gini Index at the start but flattened at the end, suggesting that it does not improve inequality drastically.

Figure 27:
The Gini index in the scenario with an equal increase of density in Cambridge, MA



The lack of improvement may occur due to various factors, as equality of resource availability is different than allocating the resources based on the needs of the tracts. As in all the United States, the potential reasons could be:

- Demand for medical services, including immunization, may vary around Cambridge, MA. If medical facilities are increased equitably, tracts with higher needs or demands may still have a shortfall, whereas tracts with lower needs may have too many. This mismatch can worsen immunization inequities, raising the Gini index.

- Variations in population density: Increasing medical facility density uniformly may not serve highly populated zones. Thus, vaccination rates may not rise as projected, increasing inequality.
- Income, education, and transportation may affect vaccination rates. Inequalities in immunization rates may persist or even increase if medical facilities are uniformly dense.
- Increased medical facilities may have heterogeneous consequences. Some areas may benefit immediately from the improved availability of medical care due to enhanced access to information and resources. Vaccination inequalities may increase.

4.5.3.2 Adjusting proportionally density of medical facilities in tracts on population density

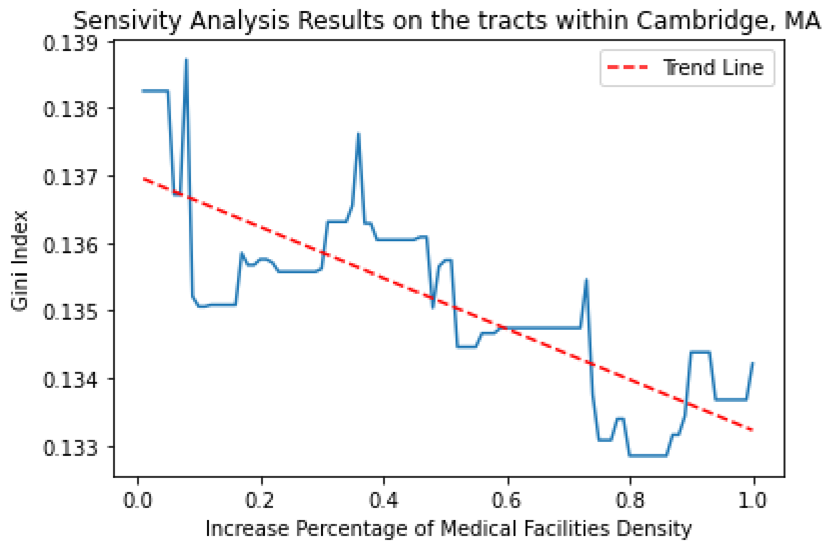
When we increased the density of medical facilities proportionally to the population density in each tract of Cambridge, MA, we observed a decreasing trend in the Gini index. This result suggests that allocating resources based on the needs of each tract is more effective in reducing inequality in vaccination rates than uniformly increasing the density of medical facilities across all tracts.

The Gini index increased in the previous scenario, where medical facilities were increased uniformly across all tracts. This outcome indicated that a uniform increase in resources did not address each tract's unique needs and characteristics, leading to a rise in inequality.

In this scenario, the decreasing trend in the Gini index demonstrates that considering population density when increasing medical facilities helps better match resources with the demand for healthcare services in each tract. By allocating resources in this manner, it better addresses disparities in vaccination rates and contributes to a more equitable distribution of healthcare services (see Figure 28).

Figure 28:

The Gini index of the scenario with a proportional increase based on population density in Cambridge, MA



4.5.3.3 Adjusting the density of medical facilities proportionally to the average time to commute

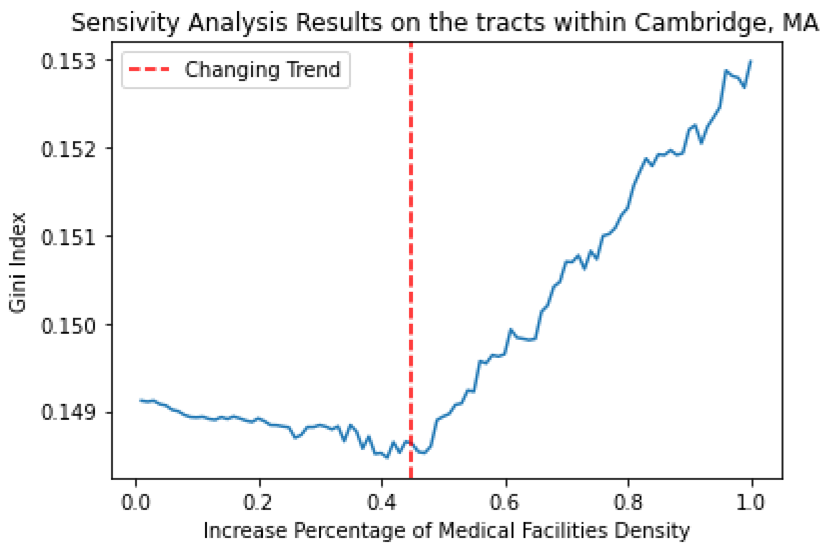
Previously, we increased the density of medical facilities proportionally to the average commute time on all the Cambridge, MA, tracts and calculated the Gini index. We found that the Gini index decreased initially (up to a 45% increase) and then increased drastically afterward, reproducing the same behavior as seen in the scenario for the entire United States.

This result suggests that allocating resources based on the average time to commute can help reduce inequality in vaccine access. When medical facilities are distributed according to commute times, those areas with longer commute times get additional resources, making it more convenient for residents to access healthcare services, including vaccinations. This approach can lead to an improvement in equity for areas that might have been underserved before.

However, the drastic increase in the Gini index after the 45% increase (as shown in Figure 29) suggests that the allocation strategy based solely on commute times may not be the most effective approach to reducing inequality in vaccination rates beyond a certain point. One explanation could be that commute times alone do not capture all the factors contributing to disparities in healthcare access, such as population density, socioeconomic status, or racial and ethnic composition.

Figure 29:

The Gini index in the scenario with a proportional increase based on the average time to commute in Cambridge, MA



5 Discussion

The results showed that several socioeconomic factors, such as transportation options and health insurance coverage, could affect county and tract vaccination attainment. We also examined the effect of adjusting the density of medical facilities on the equity metrics for the United States and Cambridge, MA. Consequently, we remarked that depending on the characteristics of each geographical entity, the impact of allocating more medical facilities differs from county to county or census tract to census tract. In this section, we discuss our results and compare them to previous research studies conducted in the field, yielding insights to policymakers. We conclude this chapter by summarizing the limitations we see in our work.

5.1 Contribution

When we started our study, we initially hypothesized that once we subtract all the effects of spatial factors from vaccination attainment, equity can be achieved by having an equal outcome for the socioeconomic features selected. Measures like the Gini or Theil indices can only measure deviation from equality and thus, as Whitehead et al. (2019) asserted, may only lead to equal outcomes. Therefore, it was necessary to reduce the problem to a point where we could safely state that equality also means equity. We, however, found that complex interactions between spatial and socioeconomic variables made it difficult to subtract these effects from the outcome. We apply the WHO's (2022) definition of equity, stating that the absence of avoidable or remediable differences between groups of people, whether these groups are socially, economically, demographically, or geographically defined, constitutes equity.

However, complex interactions between spatial and socioeconomic variables are present in the outcome. Thus, we could not easily separate the effects of those factors. Nevertheless, the vaccination attainment index and SHAP allowed us to identify the drivers of the outcome and their main interactions on the individual tract level. Based on the interactions and the importance of the features, we can see that the American Indian and Alaska Native Alone population, health insurance coverage, and transportation options are the main drivers for vaccination attainment and its equity. Our findings on transportation means confirm the results of Field (2000) and Jin et al. (2022) regarding means of transportation and their importance to equity in healthcare. Regarding the American Indian and Alaska Native Alone populations, they still have difficulties accessing health care, as mentioned in the study conducted in Arizona in 2018 (Adakai, 2018). Health insurance coverage is also crucial for healthcare access in general and in the vaccination context in particular, as it relates to the economic situation of individuals and the affordability of the insurance to be covered by the health system, which impacts vaccination equity.

We, therefore, assumed that carefully applying the indices to the outcome allows us to make conclusions about the equity of the outcome proxied through equality. To achieve equity, we sought to enable the population to experience the same ease and convenience of access based on their socioeconomic status in each tract. By looking at the (granular) tract level, we hoped to account for differences in the population. For our analysis, we used the Theil Index and the Gini Index because the Theil Index helped us identify those tracts contributing most to inequity, as recommended by Jin et al. (2022).

Dastgoshade et al. (2022) also stated that it may not be possible to incorporate all potentially relevant factors in all contexts. We sought to overcome this limitation by identifying those socioeconomic factors that best describe the population in an area. However, we must acknowledge that in our study, we found that different regions have different drivers of vaccination attainment and, thus, may react differently when changing variables to improve the equity of vaccine access.

Following our approach, proxying equity through equality of the outcome (see 2.2.4.3 and 3.6) and the granularity of our study, we arrived at a measure for equity through the Gini and the Theil indices on the vaccination attainment index. However, we did not only want to measure the equity of vaccine access. We also wanted to propose how to improve it. We studied how changing factors affect the outcome. When medical facilities are distributed according to commute times, those areas with longer commute times get additional resources, making it more convenient for residents to access healthcare services, including vaccinations. This approach can lead to an improvement in equity for areas that might have been underserved before. It reinforces the *equity needs approach* that most socially disadvantaged populations should receive more services to provide these groups with favorable conditions they would not otherwise have (Nicholls, 2001).

In conclusion, while allocating resources based on commute times can positively affect equity in healthcare access, other factors, such as population density, seem to provide a more effective allocation strategy for reducing disparities in vaccination rates. It is crucial to consider multiple factors when planning resource allocation to maximize the use of resources and ensure equitable access to healthcare services, including vaccinations: transportation means, health insurance coverage, and being an American Indian and Alaska Native. This idea agrees with the vertical spatial equity definition of Ashik et al. (2020) by addressing the needs of the affected population. However, applying a horizontal spatial equity approach (see 4.4.1 and 4.5.3.1), distributing facilities equally to residents regardless of a population's need, disagreed with other equity concepts, particularly the WHO's.

When comparing the scenario, which increased the density of medical facilities proportionally to population density, the decreasing trend in the Gini index observed in the previous scenario suggests that population density might be a better indicator for allocating resources and reducing inequality in vaccination rates (see 4.4.2 and 4.5.3.2). It also aligns with our hypothesis that for the short term, the only good factor to be influenced by policymakers is the density of vaccination centers, proxied through the density of medical facilities. Policymakers should focus on ensuring that the distribution of medical facilities matches the population's needs and consider expanding services in underserved areas.

5.2 Limitations

While our approach to defining and measuring the equity of vaccine access on a microscopic level is novel, we discuss some limitations in the following section.

First, we cannot model every individual, even with the census tract and county commuting data. Our suggested measures are calculated based on summary statistics representing census tracts as the smallest unit. The model likely does not provide accurate results if a unique situation can be found in any of the tracts. Due to a lack of more granular data (even vaccination attainment rates on tract level could help), we cannot assess its impact.

Second, with COVID-19, we got the chance to have country-wide data to evaluate the as-is situation using our approach. However, we are dealing with a single event. All conclusions we draw are based on this one event, and we cannot assess whether specific characteristics are unique to this disease. We also want to highlight that, with the data available, we did not account for any uncertainty. Furthermore, we only looked at data from the United States which may further limit the generalizability to other countries with different socioeconomic properties.

Third, our approach looked at the individual counties or census tracts, respectively. We, therefore, neglected the potential effects of neighboring counties and tracts. In particular, on the tract level, they may affect the population closer to the borders.

Fourth, both the Gini and Theil indices have limitations when measuring equity. The Gini index is less sensitive to extreme values and is a relative measure, making it unsuitable for comparing different vaccination indices. It can be ambiguous in interpretation and lacks decomposability, hindering subgroup analysis. The Theil index is less intuitive and also a relative measure. Although it is more sensitive to extreme values than the Gini index, outliers can still influence its results. Despite these limitations, both indices are valuable tools for assessing inequality and are often used alongside other measures for a comprehensive understanding of equity in a given population.

Fifth, we initially hypothesized that we should assume equality after subtracting the effects of spatial factors. However, since we had to shift our methodology to using an XGBoost model, we could not numerically isolate the effects of spatial variables. While we could argue that our design allowed us to assume equity is equality for the outcome, we failed to prove it numerically (see also 2.2.4.3 and 3.6).

Sixth, we also limited our study to ensuring equitable access to the vaccine network and neglected factors like the reputation of sites or individual preferences. We did this because it is the first and most crucial step to ensure equitable accessibility. Ensuring other aspects of vaccination is essential but, therefore, out of the scope of this capstone project.

Finally, only considering access: in the COVID data set, we calculated the vaccination attainment index where we assumed that differences are alone access related. We, therefore, disregarded the implications of factors like vaccination hesitance and skepticism. It may be essential because some underserved socioeconomic groups may have developed skepticism because of events happening in the past, and thus their underserving may not be related to access. For instance, the black population may have lost trust in the health system due to the 1932 Syphilis Study (Centers for Disease Control and Prevention, 2022), where syphilis was studied on the population without their consent or knowledge.

6 Conclusion

In this study, we aimed to understand the factors influencing vaccination rates and equity across the United States by conducting a scenario analysis on the tract level. By focusing on tract-level data, we were able to capture a finer granularity of information, allowing for the identification of localized disparities and trends that might be obscured when analyzing data at a broader scale, such as the county level. We also focused our analysis on Cambridge, MA, as a case study to provide more in-depth insights into the local dynamics and challenges a specific community faces. By examining the effects of various scenarios on Cambridge, MA, we can better understand how different interventions and resource allocations may impact vaccination rates and equity within the city and provide valuable information for policymakers and public health professionals to develop targeted strategies to improve vaccination uptake and reduce disparities.

Unsurprisingly, we found that population density is a primary driver of equitable access. However, future research should further investigate how neighboring areas interact with the access. Additionally, we found that the main features influencing vaccination attainment are transportation means, health insurance coverage, and being an American Indian and Alaska Native. Considering this, those effects can be addressed by policymakers in the long term.

Exploring other factors influencing vaccination rates, such as vaccine hesitancy, skepticism, and individual preferences, could provide a more comprehensive understanding of vaccination equity. Evaluating the impact of targeted interventions to address these factors could also contribute to developing more effective strategies for improving vaccination equity.

Moreover, future research could benefit from examining alternative equity measures and employing different methodologies to better isolate spatial variables' effects. It would enable researchers to accurately assess the relationship between equality and equity, potentially informing more effective policies to achieve equitable healthcare access. While we confirmed our hypothesis that equal treatment of local communities does not contribute to an equitable outcome, addressing their needs could improve it. We, therefore, propose that future research focuses more on using the results from SHAP and tailor the addressing of needs based on the most influential factors identified: SHAP helps to understand complex interactions and provides entry points for further study (1) on how socioeconomic groups affect the outcome and on (2) the spatial factors interacting with them.

Furthermore, assessing the impact of various resource allocation strategies on vaccination equity, considering both horizontal and vertical spatial equity, would help policymakers identify the most effective approaches to ensure equitable access to healthcare services, including vaccinations. This could improve overall public health outcomes and reduce health disparities.

Finally, with our defined set of potential metrics, it is of utmost importance to approach policymakers now with our findings, e. g., members of the COVID-19 Vaccine Equity Initiative, and suggest they consider them in their decision-making and risk management processes. A successful implementation can maximize immunization coverage by incorporating equity considerations and better preparing for the next pandemic.

References

Adakai, M. (2018). Health Disparities Among American Indians/Alaska Natives—Arizona, 2017. *MMWR. Morbidity and Mortality Weekly Report*, *67*. https://doi.org/10.15585/mmwr.mm6747a4

American Community Survey. (2017). https://www.socialexplorer.com/data/ACS2021_5yr/

Arnett, D. K., Blumenthal, R. S., Albert, M. A., Buroker, A. B., Goldberger, Z. D., Hahn, E. J., Himmelfarb, C. D., Khera, A., Lloyd-Jones, D., McEvoy, J. W., Michos, E. D., Miedema, M. D., Muñoz, D., Smith, S. C., Virani, S. S., Williams, K. A., Yeboah, J., & Ziaeian, B. (2019). 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*, *140*(11), e563–e595. https://doi.org/10.1161/CIR.0000000000000677

Ashik, F. R., Mim, S. A., & Neema, M. N. (2020). Towards vertical spatial equity of urban facilities: An integration of spatial and aspatial accessibility. *Journal of Urban Management*, *9*(1), 77–92. https://doi.org/10.1016/j.jum.2019.11.004

Askari, H., & Mirakhor, A. (2020). *Conceptions of Justice from Islam to the Present*. Springer International Publishing. https://doi.org/10.1007/978-3-030-16084-5

Bardosh, K., Figueiredo, A. de, Gur-Arie, R., Jamrozik, E., Doidge, J., Lemmens, T., Keshavjee, S., Graham, J. E., & Baral, S. (2022). The unintended consequences of COVID-19 vaccine policy: Why mandates, passports and restrictions may cause more harm than good. *BMJ Global Health*, *7*(5), e008684. https://doi.org/10.1136/bmjgh-2022-008684

Basu, S., Andrews, J., Kishore, S., Panjabi, R., & Stuckler, D. (2012). Comparative Performance of Private and Public Healthcare Systems in Low- and Middle-Income Countries: A Systematic Review. *PLoS Medicine*, *9*(6), e1001244. https://doi.org/10.1371/journal.pmed.1001244

Behbahani, H., Nazari, S., Jafari Kang, M., & Litman, T. (2019). A conceptual framework to formulate transportation network design problem considering social equity criteria. *Transportation Research Part A: Policy and Practice*, *125*, 171–183. https://doi.org/10.1016/j.tra.2018.04.005

Boone, C. G., Buckley, G. L., Grove, J. M., & Sister, C. (2009). Parks and People: An Environmental Justice Inquiry in Baltimore, Maryland. *Annals of the Association of American Geographers*, *99*(4), 767–787. https://doi.org/10.1080/00045600903102949

Brown, K. (2021). *5 Global Issues to Watch in 2022*. Unfoundation.Org. https://unfoundation.org/blog/post/5-global-issues-to-watch-in-2022/

Bureau of Transportation Statistics. (2023). *National Transit Map Stops*. https://services.arcgis.com/xOi1kZaI0eWDREZv/arcgis/rest/services/NTM_Stops/FeatureServer/0

Byrd, W. M., & Clayton, L. A. (1992). An American health dilemma: A history of blacks in the health system. *Journal of the National Medical Association*, *84*(2), 189–200.

Centers for Disease Control and Prevention. (2020a). Agency for Toxic Substances and Disease Registry/Geospatial Research, Analysis, and Services Program. *CDC/ATSDR Social Vulnerability Index*.

Centers for Disease Control and Prevention. (2020b). *Paving the Road to Health Equity—OMHHE - CDC*. https://www.cdc.gov/minorityhealth/publications/health_equity/index.html

Centers for Disease Control and Prevention. (2021). *Who Sets the Immunization Schedule?* Centers for Disease Control and Prevention. https://www.cdc.gov/vaccines/parents/schedules/sets-schedule.html

Centers for Disease Control and Prevention. (2022, December 20). *Tuskegee Study—Timeline*. https://www.cdc.gov/tuskegee/timeline.htm

Centers for Disease Control and Prevention. (2023). *COVID-19 Vaccinations in the United States, County*. https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh

Chang, H.-S., & Liao, C.-H. (2011). Exploring an integrated method for measuring the relative spatial equity in public facilities in the context of urban parks. *Cities*, *28*(5), 361–371. https://doi.org/10.1016/j.cities.2011.04.002

Chen, J. T., & Krieger, N. (2021). Revealing the Unequal Burden of COVID-19 by Income, Race/Ethnicity, and Household Crowding: US County Versus Zip Code Analyses. *Journal of Public Health Management and Practice*, *27*(Supplement 1), S43. https://doi.org/10.1097/PHH.0000000000001263

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cho, C.-J. (1998). An equity-efficiency trade-off model for the optimum location of medical care facilities. *Socio-Economic Planning Sciences*, *32*(2), 99–112. https://doi.org/10.1016/S0038-0121(97)00007-4

Chok, N. S. (2010, September 24). *Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data* [University of Pittsburgh ETD]. University of Pittsburgh. http://d-scholarship.pitt.edu/8056/

Cook, K. S., & Hegtvedt, K. A. (1983). Distributive Justice, Equity, and Equality. *Annual Review of Sociology*, *9*, 217–241.

Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. *Accounting, Auditing and Finance*, *1*(1), 5–8. https://doi.org/10.23977/accaf.2020.010102

Dadashpoor, H., Rostami, F., & Alizadeh, B. (2016). Is inequality in the distribution of urban facilities inequitable? Exploring a method for identifying spatial inequity in an Iranian city. *Cities*, *52*, 159–172. https://doi.org/10.1016/j.cities.2015.12.007

Dastgoshade, S., Shafiee, M., Klibi, W., & Shishebori, D. (2022). Social equity-based distribution networks design for the COVID-19 vaccine. *International Journal of Production Economics*, 108684. https://doi.org/10.1016/j.ijpe.2022.108684

Dear, M. J. (1974). A Paradigm for Public Facility Location Theory. *Antipode*, *6*(1), 46–50. https://doi.org/10.1111/j.1467-8330.1974.tb00583.x

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., & Schiavina, M. (2021). Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, *125*, 103312. https://doi.org/10.1016/j.jue.2020.103312

Dijkstra, L., & Poelman, H. (2014). *A harmonised definition of cities and rural areas: The new degree of urbanisation*.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dukhovnov, D., & Barbieri, M. (2021). County-level socio-economic disparities in COVID-19 mortality in the USA. *International Journal of Epidemiology*, *51*(2), 418–428. https://doi.org/10.1093/ije/dyab267

Enayati, S., & Özaltın, O. Y. (2020). Optimal influenza vaccine distribution with equity. *European Journal of Operational Research*, *283*(2), 714–725. https://doi.org/10.1016/j.ejor.2019.11.025

Fahlevi, M. (2019). ISLAMIC ECONOMY AND POLITICS IN THE VIEW OF MUHAMMAD BAQIR SADR. *Economics and Management*, *13*(2), 6.

Field, K. (2000). Measuring the need for primary health care: An index of relative disadvantage. *Applied Geography*, *20*(4), 305–332. https://doi.org/10.1016/S0143-6228(00)00015-1

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Gastwirth, J. L. (2017). Is the Gini Index of Inequality Overly Sensitive to Changes in the Middle of the Income Distribution? *Statistics and Public Policy*, *4*(1), 1–11. https://doi.org/10.1080/2330443X.2017.1360813

Goentzel, J., Russell, T., Carretti, H. R., & Hashimoto, Y. (2022). Vaccine network design to maximize immunization coverage. *Journal of Humanitarian Logistics and Supply Chain Management*. https://doi.org/10.1108/JHLSCM-10-2021-0101

Greenberger, M. (2018). Better Prepare Than React: Reordering Public Health Priorities 100 Years After the Spanish Flu Epidemic. *American Journal of Public Health*, *108*(11), 1465–1468. https://doi.org/10.2105/AJPH.2018.304682

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed). McGraw-Hill Irwin.

Jin, T., Cheng, L., Wang, K., Cao, J., Huang, H., & Witlox, F. (2022). Examining equity in accessibility to multi-tier healthcare services across different income households using estimated travel time. *Transport Policy*, *121*, 1–13. https://doi.org/10.1016/j.tranpol.2022.03.014

Kelobonye, K., McCarney, G., Xia, J. (Cecilia), Swapan, M. S. H., Mao, F., & Zhou, H. (2019). Relative accessibility analysis for key land uses: A spatial equity perspective. *Journal of Transport Geography*, *75*, 82–93. https://doi.org/10.1016/j.jtrangeo.2019.01.015

Kilinc, M. S., Stamm, J. H., & Milburn, A. B. (2018). Using spatial data analytics to identify associations between home healthcare services' accessibility and socioeconomic factors. *IISE Transactions on Healthcare Systems Engineering*, *8*(4), 250–267. https://doi.org/10.1080/24725579.2018.1502222

Killeen, R. M. (2007). Vaccines—One of the Greatest Medical Advances of Modern Times. *Canadian Pharmacists Journal / Revue Des Pharmaciens Du Canada*, *140*(2_suppl), S2–S2. https://doi.org/10.3821/1913-701X-140-Sp2.S2

Lazarus, J. V., Wyka, K., White, T. M., Picchio, C. A., Rabin, K., Ratzan, S. C., Parsons Leigh, J., Hu, J., & El-Mohandes, A. (2022). Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-31441-x

Lewis, E. O., MacKenzie, D., & Kaminsky, J. (2021). Exploring equity: How equity norms have been applied implicitly and explicitly in transportation research and practice. *Transportation Research Interdisciplinary Perspectives*, *9*, 100332. https://doi.org/10.1016/j.trip.2021.100332

Liao, C.-H., Hsueh-Sheng, C., & Tsou, K.-W. (2009). Explore the spatial equity of urban public facility allocation based on sustainable development viewpoint. In *REAL CORP 2009: Cities 3.0—Smart, sustainable, integrative: Proceedings of 14th International Conference on Urban Planning, Regional Development and Information Society ; Beiträge zur 14. Internatinalen Konferenz zu Stadtplanung, Regionalentwicklung und Informationsgesellschaft ; [strategies, concepts and technologies for planning the urban future ; 22—25 April 2009, Centre de Disseny de Sitges, Catalonia, Spain ; Tagungsband]* (pp. 137–145). CORP, Competence Center of Urban and Regional Planning.

Lucy, W. (1981). Equity and Planning For Local Services. *Journal of the American Planning Association*, *47*(4), 447–457. https://doi.org/10.1080/01944368108976526

Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*.

Mashrur, R., Meher, N. N., Mashrur, R., & Meher, N. N. (2015). A GIS Based Integrated Approach to Measure the Spatial Equity of Community Facilities of Bangladesh. *AIMS Geosciences*, *1*(1), Article geosci-01-00021. https://doi.org/10.3934/geosci.2015.1.21

Mehrabi, N., Huang, Y., & Morstatter, F. (2020). *Statistical Equity: A Fairness Classification Objective* (arXiv:2005.07293). arXiv. http://arxiv.org/abs/2005.07293

Mohammadi, M., Dehghan, M., Pirayesh, A., & Dolgui, A. (2022). Bi-objective optimization of a stochastic resilient vaccine distribution network in the context of the COVID-19 pandemic. *Omega*, *113*, 102725. https://doi.org/10.1016/j.omega.2022.102725

Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics: Conference Series*, *1361*(1), 012015. https://doi.org/10.1088/1742-6596/1361/1/012015

National Institute of Allergy and Infectious Diseases. (2019). *Smallpox Vaccine Supply & Strength*. https://www.niaid.nih.gov/diseases-conditions/smallpox-vaccine

Nicholls, S. (2001). Measuring the accessibility and equity of public parks: A case study using GIS. *Managing Leisure*, *6*(4), 201–219. https://doi.org/10.1080/13606710110084651

Nicholls, S., & Shafer, C. S. (2001). Measuring accessibility and equity in a local park system: The utility of geospatial technologies to park and recreation professionals. *Journal of Park and Recreation Administration*, *19*, 102–124.

Norton, B. G. (1989). Intergenerational equity and environmental decisions: A model using Rawls' veil of ignorance. *Ecological Economics*, *1*(2), 137–159. https://doi.org/10.1016/0921-8009(89)90002-5

O'Donnell, C., & Erman, M. (2021, December 9). U.S. campaign to vaccinate young children off to sluggish start despite abundant supply. *Reuters*. https://www.reuters.com/business/healthcare-pharmaceuticals/us-campaign-vaccinate-young-children-off-sluggish-start-despite-abundant-supply-2021-12-09/

Organisation for Economic Co-operation and Development. (2020). *The territorial impact of COVID-19: Managing the crisis across levels of government*. OECD. https://www.oecd.org/coronavirus/policy-responses/the-territorial-impact-of-covid-19-managing-the-crisis-across-levels-of-government-d3e314e1/

Pereira, R. H. M., Schwanen, T., & Banister, D. (2017). Distributive justice and equity in transportation. *Transport Reviews*, *37*(2), 170–191. https://doi.org/10.1080/01441647.2016.1257660

Reda, A. (2014). Weber and Baqir as-Sadr: The Paradox of Economic Development in Islamic Societies. *The American Journal of Economics and Sociology*, *73*(1), 151–177. https://doi.org/10.1111/ajes.12069

Richter, F. (2021, December 9). *Infographic: The Countries Best and Worst Prepared for a Pandemic* [Statistics]. Statista Infographics. https://www.statista.com/chart/20629/ability-to-respond-to-an-epidemic-or-pandemic/

Rydland, H. T., Friedman, J., Stringhini, S., Link, B. G., & Eikemo, T. A. (2022). The radically unequal distribution of Covid-19 vaccinations: A predictable yet avoidable symptom of the fundamental causes of inequality. *Humanities and Social Sciences Communications*, *9*(1), Article 1. https://doi.org/10.1057/s41599-022-01073-z

Shapley, L. S. (1953). A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Simmons-Duffin, S., & Huang, P. (2021). Here's How The U.S. Can Jump-Start Its Sluggish COVID-19 Vaccine Rollout. *NPR*. https://www.npr.org/sections/health-shots/2021/01/08/954662748/heres-how-the-u-s-can-jump-start-its-sluggish-covid-19-vaccine-rollout

Smoyer-Tomic, K. E., Hewko, J. N., & Hodgson, M. J. (2004). Spatial accessibility and equity of playgrounds in Edmonton, Canada. *The Canadian Geographer / Le Géographe Canadien*, *48*(3), 287–302. https://doi.org/10.1111/j.0008-3658.2004.00061.x

Talen, E., & Anselin, L. (1998). Assessing Spatial Equity: An Evaluation of Measures of Accessibility to Public Playgrounds. *Environment and Planning A: Economy and Space*, *30*(4), 595–613. https://doi.org/10.1068/a300595

Tian, M., Yuan, L., Guo, R., Wu, Y., & Liu, X. (2022). Evaluations of Spatial Accessibility and Equity of Multi-Tiered Medical System: A Case Study of Shenzhen, China. *International Journal of Environmental Research and Public Health*, *19*(5), 3017. https://doi.org/10.3390/ijerph19053017

Tsou, K.-W., Hung, Y.-T., & Chang, Y.-L. (2005). An accessibility-based integrated measure of relative spatial equity in urban public facilities. *Cities*, *22*(6), 424–435. https://doi.org/10.1016/j.cities.2005.07.004

Ugochukwu, O. C., Chukwudi, E. I., Nkechi, A. A., & Ogochukwu, I. M. (2020). Covid-19 Pandemic in Nigeria: The Essential Roles of Biomedical Engineers & Technologists. *Journal of Bioscience & Biomedical Engineering*, *1*(3). https://doi.org/10.47485/2693-2504.1024

United Nations. (1948). *Universal Declaration of Human Rights*. United Nations; United Nations. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations. (2022). *Health*. United Nations; United Nations. https://www.un.org/en/global-issues/health

Wang, F., & Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health & Place*, *11*(2), 131–146. https://doi.org/10.1016/j.healthplace.2004.02.003

Whitehead, J., L. Pearson, A., Lawrenson, R., & Atatoa-Carr, P. (2019). How can the spatial equity of health services be defined and measured? A systematic review of spatial equity definitions and methods. *Journal of Health Services Research & Policy*, *24*(4), 270–278. https://doi.org/10.1177/1355819619837292

Wilensky, G. R. (2021). 2020 Revealed How Poorly the US Was Prepared for COVID-19—And Future Pandemics. *JAMA*, *325*(11), 1029–1030. https://doi.org/10.1001/jama.2021.1046

World Bank, & World Health Organization. (2017). *World Bank and WHO: Half the world lacks access to essential health services, 100 million still pushed into extreme poverty because of health expenses*. https://www.who.int/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses

World Health Organization. (2022). *Health equity*. https://www.who.int/westernpacific/health-topics/equity

Yuan, Y., Xu, J., & Wang, Z. (2017). Spatial Equity Measure on Urban Ecological Space Layout Based on Accessibility of Socially Vulnerable Groups—A Case Study of Changting, China. *Sustainability*, *9*(9), Article 9. https://doi.org/10.3390/su9091552

Zhao, G., Cao, X., & Ma, C. (2020). Accounting for horizontal inequity in the delivery of health care: A framework for measurement and decomposition. *International Review of Economics & Finance*, *66*