

Puente-Beccar, Manuela

**Article — Published Version**

## Health Preferences and Sorting in the City

Health Economics

*Suggested Citation:* Puente-Beccar, Manuela (2026) : Health Preferences and Sorting in the City, Health Economics, ISSN 1099-1050, Vol. 35, Iss. 7, pp. 1033-1055, <https://doi.org/10.1002/hec.70098>

This Version is available at:

<https://hdl.handle.net/10419/342910>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

## RESEARCH ARTICLE OPEN ACCESS

# Health Preferences and Sorting in the City

Manuela Puente-Beccar 

Max Planck Institute for Behavioral Economics, Bonn, Germany

**Correspondence:** Manuela Puente-Beccar ([puente@coll.mpg.de](mailto:puente@coll.mpg.de))**Received:** 11 July 2025 | **Revised:** 17 February 2026 | **Accepted:** 3 March 2026**Keywords:** health preferences | location choice | sorting

## ABSTRACT

There are large health inequalities between neighborhoods in many cities of the world. This paper studies individuals' sorting based on health amenities and exposes an important connection between health preferences and the housing market. I estimate a neighborhood choice model using geolocated data from a health survey in New York City and neighborhood characteristics from different data sources. I find that individuals with higher preferences for being healthy are more likely to choose neighborhoods close to health amenities such as parks, and this sorting explains a large part of observed geographical inequalities in health behaviors. Moreover, the model predicts that a higher demand for health amenities capitalizes into house prices and displaces poor individuals to neighborhoods further from these amenities. This implies that health awareness campaigns might actually worsen the inequality in access to public health amenities.

**JEL Classification:** D01, I12, I14, R21

## 1 | Introduction

In many cities of the developed world, individuals in one neighborhood can expect to live 10 years longer than those in other neighborhoods a few miles away (Gourevitch 2019; O'Mahony 2019), a life expectancy gap comparable to that between the United States and Laos or Senegal.<sup>1</sup> Similar levels of inequality are observed in health behaviors, which are more directly related to individuals' preferences. These preferences, in turn, may influence where individuals choose to live.

The first aim of this paper is to study the role of the sorting channel in the observed differences in health between neighborhoods, a channel that has been acknowledged but understudied at such fine geographical units. The proposed mechanism is that individuals with stronger health preferences place a higher value on health amenities. This leads to a sorting pattern where healthier individuals are more likely to live near health amenities and near others with similar preferences. Additionally, this higher valuation would be capitalized into

house prices, providing a direct connection between health preferences and the housing market.

This connection brings interesting and unstudied dynamics. It would imply that changes in health preferences will change the equilibrium of the housing market through both the capitalization of health amenities into house prices and the relocation of individuals. The second aim of this paper is, therefore, to test the potential effect of health awareness campaigns. Specifically, it explores whether health campaigns can raise rents and housing prices in areas with more health-related amenities, possibly pushing low-income residents further from these amenities.

To answer these two research questions I propose a model of neighborhood choice where individuals have heterogeneous preferences for health amenities and other neighborhood characteristics. In the model, individuals have different incomes, which constrain their choice of housing. Additionally, they differ in their level of education, age, race, gender, family composition, and health preferences. Housing prices are an

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Health Economics* published by John Wiley & Sons Ltd.

equilibrium outcome depending on the overall demand and supply of housing in each neighborhood.

To identify the sorting mechanism behind neighborhood choice, given the observed location of individuals, I exploit exogenous variation in access to health amenities provided by parks since their location was determined many years ago, especially in developed, densely populated cities. Parks provide space for physical activities, clean air (Jo et al. 2019) and can positively impact mental health (Wood et al. 2017). Furthermore, to reduce concerns of reverse causality, where parks affect behavior, I proxy health preferences with an index based on health behavior variables that do not directly relate to parks, such as having unsafe sex, while excluding behaviors that could be affected by proximity to parks, such as exercising.

Estimating the model, I find that independent of other individual characteristics, such as income, individuals with high health preferences are more likely to choose neighborhoods closer to parks, and this result is robust to different specifications of the model and the main variables. The sorting model explains more than 60% of the variation in health preferences by neighborhoods, highlighting that when studying neighborhood effects on health, the location of individuals cannot be taken as given.

To answer the second research question, I use the estimated model to predict changes in housing prices and neighborhood composition under a counterfactual scenario in which everyone's health preference index increases (numerically set to the observed increase in a 10-year period). The effects are an increase of around 2% in prices of neighborhoods close to parks, due to the increased demand for health amenities, and a reduction of 2%–3% in prices of neighborhoods that are distant from parks. Given the importance of prices for the location decision of low-income individuals, the model predicts that these price increases displace poor individuals to cheaper neighborhoods far from health amenities. This simulation shows clearly that health policies can have effects on the housing market and the configuration of neighborhoods. Additionally, observed housing prices after this 10-year period confirm a higher valuation of parks, providing evidence that these predictions are verified in the real world.

The main data for this analysis is an individual-level health survey in New York City of more than 40,000 individuals located in 128 neighborhoods, from a pool of cross-sections for the period 2009–2013. This dataset allows me to observe individuals' detailed health behavior, their income category, and their geolocation, a very unique combination of data types. Using New York City as a case study has several other advantages. The first one is data availability, which allows me to match the health survey with publicly available geographical data on park zones and administrative data on housing and neighborhood characteristics, including house prices and crime rates. Another advantage is that due to geographical constraints, the structure of the city remains constant, which allows me to interpret the results of a static model.

This paper contributes mainly to the understanding of the connection between health preferences and the housing market, a phenomenon that is only recently being studied (Bishop et al. 2024). Additionally, it contributes to the literature in urban economics studying health amenities as determinants of location choice. So far the focus has been on temperature and air quality at coarse geographical levels (e.g., Chay and Greenstone 2005; Bayer et al. 2009; Albouy et al. 2016; Heblich et al. 2021; Mathes 2025). A couple of papers looking at sorting and health behaviors are Plantinga and Bernell (2007), which studies obesity considering counties with differences in urban sprawl, and Darden (2021), which studies smoking prevalence in urban versus rural areas. This literature has not yet studied health amenities and sorting at the neighborhood level and so is not able to explain health inequalities between neighborhoods.

This paper also contributes to the literature on neighborhood (or place) effects on health (e.g., Ellen et al. 2001; Bilger and Carrieri 2013; Alexander and Currie 2017; Deryugina and Molitor 2020, 2021; Percoco 2021; Finkelstein et al. 2021) and in particular on health behavior (e.g., Kling et al. 2007; Eid et al. 2008; Zhao and Kaestner 2010; Ludwig et al. 2013; Dang 2015; Ou 2019; Allcott et al. 2019). Most of these papers overcome the problem of individuals' sorting with (quasi)experimental designs but do not study it. My contribution is to specifically study the role of sorting and to show that it is the main mechanism behind the neighborhood-health relation.

The rest of the paper is organized as follows. Section 2 describes the data. A reduced form analysis is presented in Section 3. Section 4 sets out the structural model of location choice. Section 5 shows the estimation results, while the counterfactual evaluation is presented in Section 6. Section 7 presents a discussion of the identification assumptions and the results, and Section 8 concludes.

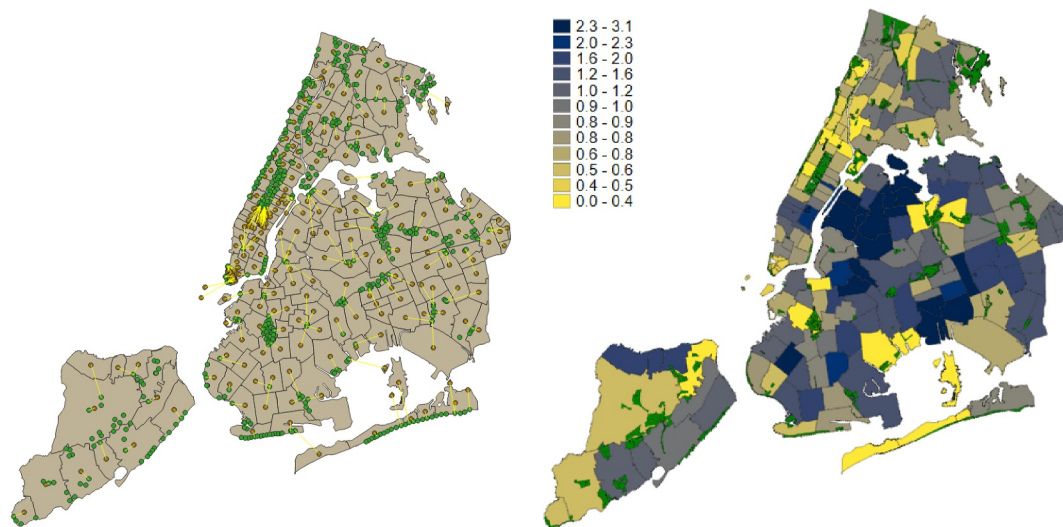
## 2 | Data

To understand individuals' location decisions as determinants of health in neighborhoods, we need data on both neighborhood and individual characteristics. I describe below the data used for the main analysis.

### 2.1 | Data Sources and Preparation

#### 2.1.1 | Data on Health and Demographics

Individual level data comes from the Community Health Survey (CHS) of the New York City Department of Health and Mental Hygiene (DOHMH 2013). I got restricted access to the geolocation of individuals at the ZIP code level.<sup>2</sup> For confidentiality reasons, the data is a pooled cross-section for the period 2009–2013. This data gives information on health behavior and demographics, including income categories and age groups, for more than 40 thousand individuals in New York City (NYC) located in 128 ZIP codes.



**FIGURE 1** | Distance to the closest park. The left panel shows the calculation of distance to the closest park by taking the linear distance from the centroids of ZIP codes to the centroids of park zones. The right panel shows the resulting attribution of distance to each neighborhood (ZIP code) in Km. *Source:* Own elaboration from geographic data on ZIP codes and park zones from NYC Open Data. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 2.1.2 | Data on Housing Prices

I use all sales in the city, publicly available at *NYC Open Data*, to calculate average house prices for every neighborhood. I cleaned the house price data by excluding sales of non-residential units, sales of empty (or almost) lots, and sales registered with a symbolic price (usually from family transfers). To be able to compare housing expenditures with yearly income, I calculate annual rents with the price-to-rent ratio for each borough in NYC. I considered sales in 2009, 2011, and 2013<sup>3</sup> to get representative prices faced by the individuals in the health survey. The cleaned file contains approximately 135 thousand sales.

### 2.1.3 | Data on Housing and Neighborhood Characteristics

Primary Land Use Tax Lot Output (PLUTO) files, also publicly available at *NYC Open Data*, are used to calculate housing characteristics at the neighborhood level, namely, average number of floors (a proxy for population density), percentage of detached houses, average age of the building, house size, and area destined for commercial or industrial use. The cleaned file for 2009<sup>4</sup> has approximately 850 thousand observations.

To calculate crime rates by neighborhood, I use data from the *NYPD crime statistics* on all registered instances of violent crimes by police precinct. I estimated the yearly frequency of violent crimes<sup>5</sup> per 1000 residents by neighborhood, with a mapping between precincts and ZIP codes. There are approximately 350 thousand occurrences of these crimes in the period 2009–2013.

### 2.1.4 | Geographic Data

*NYC Open Data* provides shape files for the ZIP codes of NYC and the park zones. To calculate distance to the Central

Business District (CBD), I used the Google Maps API from the centroids of each ZIP code to Times Square<sup>6</sup> in Manhattan, using transit mode.

Distance to the closest park was calculated by the minimum aerial distance from the centroids of each ZIP code to the centroids of park zones using QGIS. This can be seen in the left panel of Figure 1, where green dots represent the centroids of park zones and brown dots represent the centroids of ZIP codes. The right panel of Figure 1 shows the resulting distance by neighborhood.

## 2.2 | Parks

A couple of clarifications are useful on the definition of parks and distance. I consider “parks” all the green areas that the NYC Department of Parks and Recreation considers park zones, except for park areas smaller than 2 ha (5 acres - to exclude playgrounds and to avoid endogeneity stemming from conversion of some areas into green areas). Park zones, therefore, include beaches and exclude green areas such as cemeteries or natural reserves. The analysis includes more than 250 parks, and the majority of these were established between 1700 and 1950,<sup>7</sup> which implies that their location is not the result of neighbors’ decisions. Regarding distance, I choose to calculate linear distance to parks and transit distance to the CBD because it is more likely that New Yorkers go to the CBD by car or public transportation but that they go walking, running or cycling to the closest park in their neighborhood.

Finally, it is worth mentioning that parks are amenities for many different reasons. One reason is their appeal for exercise by providing space for physical activities. They also offer clean air, moderate temperatures and quiet, all characteristics that can be beneficial for health. Independently, one might value parks due to their recreation opportunities, especially for children, the opportunities for community engagement, and due to their

aesthetics. These last reasons, however, especially controlling for income and family composition, should be independent from health and health behaviors.

## 2.3 | Descriptive Statistics

Table 1 shows descriptive statistics of individuals' demographics, health behavior variables, and neighborhood characteristics. The variables in the first block of the left panel are included as controls. Individuals may differ in their preferences for neighborhoods (and parks) depending on their income, education, age, and race. Being a parent can have a stronger effect on the valuation of nearby parks, while having health insurance can be correlated with health behavior. The second block summarizes the health behaviors present in the data, which I use to construct a health preference index, as explained in the next subsection.

An individual is poor if her household income is below 100% of the Federal Poverty Line, and a "Potential parent" is an

individual in the age Group 25–44 who is also married (or cohabiting).

Unfortunately, some of the demographics in the individual-level data are coded into categories. The data gives information on four age groups (omitted category 18–24) and five income groups: from less than 100% of the Federal Poverty Line (FPL) to more than 600% of the FPL. Income is an important variable for the location choice of an individual, so I include an approximation of it in the analysis. To take into account that not all individuals pay the same rent in each neighborhood, I assign rents (divided into 5 quantiles) to each individual corresponding to their income category. Details on this assignment, the approximation of income, and a discussion of the assumptions made can be found in Appendix B.

The right panel of Table 1 shows descriptive statistics of the estimated neighborhood variables that are used in the analysis that follows. Most of them can be considered exogenous or fixed in the short term. The exceptions are crime rate –which I include because it is a very important determinant of neighborhood choice– and percentage white, percentage black/

**TABLE 1** | Descriptive statistics.

|                      | Individual |     |     |       |                            | Neighborhood |       |       |        |
|----------------------|------------|-----|-----|-------|----------------------------|--------------|-------|-------|--------|
|                      | Mean       | Min | Max | N     |                            | Mean         | SD    | Min   | Max    |
| Income (1000 USD)    | 51.79      | 9.4 | 136 | 43659 | Annual rent (1000 USD)     | 11.54        | 6.95  | 4.02  | 45.76  |
| Poor (< 100% FPL)    | 0.22       | 0.0 | 1.0 | 43659 | Distance CBD (min)         | 47.89        | 20.34 | 12.00 | 108.68 |
| College              | 0.33       | 0.0 | 1.0 | 43659 | Distance park              | 1.11         | 0.69  | 0.04  | 3.11   |
| Female               | 0.54       | 0.0 | 1.0 | 43659 | Detached houses            | 0.26         | 0.24  | 0.00  | 0.85   |
| Potential parent     | 0.21       | 0.0 | 1.0 | 43659 | Number of floors           | 3.09         | 1.90  | 1.66  | 10.85  |
| Married              | 0.49       | 0.0 | 1.0 | 43659 | Building age               | 78.23        | 14.16 | 33.89 | 102.59 |
| Age 25–44            | 0.40       | 0.0 | 1.0 | 43659 | House size ( $m^2$ )       | 112.06       | 17.87 | 70.19 | 159.50 |
| Age 45–64            | 0.32       | 0.0 | 1.0 | 43659 | Commercial area ( $Km^2$ ) | 1.34         | 2.03  | 0.13  | 16.42  |
| Age 65+              | 0.15       | 0.0 | 1.0 | 43659 | Industrial area ( $Km^2$ ) | 0.07         | 0.15  | 0.00  | 1.21   |
| Insured              | 0.81       | 0.0 | 1.0 | 43659 | Crime rate (per 1000 res)  | 19.17        | 16.09 | 0.83  | 83.56  |
| White                | 0.36       | 0.0 | 1.0 | 43659 | Percentage white           | 0.37         | 0.28  | 0.00  | 0.89   |
| Black/Hispanic       | 0.49       | 0.0 | 1.0 | 43659 | Percentage black/hisp      | 0.48         | 0.31  | 0.05  | 0.97   |
| Ever smoker          | 0.36       | 0.0 | 1.0 | 43659 |                            |              |       |       |        |
| Current smoker       | 0.15       | 0.0 | 1.0 | 43659 |                            |              |       |       |        |
| Unhealthy drinker    | 0.09       | 0.0 | 1.0 | 42971 |                            |              |       |       |        |
| Unsafe sex           | 0.23       | 0.0 | 1.0 | 24678 |                            |              |       |       |        |
| Flu vaccine          | 0.37       | 0.0 | 1.0 | 43421 |                            |              |       |       |        |
| Colonoscopy          | 0.60       | 0.0 | 1.0 | 27829 |                            |              |       |       |        |
| > = 1 sweetened bev. | 0.29       | 0.0 | 1.0 | 43278 |                            |              |       |       |        |
| < = 1 fruit/veg      | 0.35       | 0.0 | 1.0 | 42727 |                            |              |       |       |        |
| > = 4 fruit/veg      | 0.21       | 0.0 | 1.0 | 42727 |                            |              |       |       |        |
| Exercise             | 0.75       | 0.0 | 1.0 | 43623 |                            |              |       |       |        |

*Note:* Individual characteristics come from the CHS, as well as percentages of whites and blacks/Hispanics by neighborhood, using survey weights. (nominal) Income is approximated as is explained in the Appendix. Poor = 1 if household income is below 100% of the Federal Poverty Line and "potential parent" is an individual in the age Group 25–44 who is also married (or cohabiting). Definitions of health behavior variables can be found in Appendix Table A1. (nominal) Annual rent is estimated from administrative data on all sales in the city; distance to the CBD is calculated with the Google Maps API; distance to the closest park is estimated with GIS software; crime rates are estimated from NYPD crime statistics while all other neighborhood characteristics come from PLUTO files. There are 128 observations for all neighborhood variables.

**TABLE 2** | Correlations among health behaviors. [Colour table can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

|                   | Ever smoker | Current smoker | Unhlt drink | Unsafe sex | Soda  | ≤ 1 Fruit/veg | Flu vaccine | Colo noscopy | Exercise | ≥ 4 Fruit/veg |
|-------------------|-------------|----------------|-------------|------------|-------|---------------|-------------|--------------|----------|---------------|
| Ever smoker*      | 1           |                |             |            |       |               |             |              |          |               |
| Current smoker    | 0.45        | 1              |             |            |       |               |             |              |          |               |
| Unhealthy drinker | 0.13        | 0.11           | 1           |            |       |               |             |              |          |               |
| Unsafe sex*       | 0.05        | 0.07           | 0.11        | 1          |       |               |             |              |          |               |
| Soda              | 0.04        | 0.11           | −0.03       | 0.02       | 1     |               |             |              |          |               |
| ≤ 1 fruit/veg     | 0.03        | 0.12           | −0.01       | 0.002      | 0.13  | 1             |             |              |          |               |
| Flu vaccine*      | 0.03        | −0.07          | −0.01       | −0.06      | −0.04 | −0.05         | 1           |              |          |               |
| Colonoscopy*      | 0.00        | −0.12          | −0.01       | −0.04      | −0.06 | −0.06         | 0.20        | 1            |          |               |
| Exercise          | −0.02       | −0.06          | 0.04        | 0.05       | −0.09 | −0.18         | 0.02        | 0.05         | 1        |               |
| ≥ 4 fruit/veg     | −0.01       | −0.10          | 0.01        | −0.01      | −0.13 | −0.37         | 0.04        | 0.05         | 0.14     | 1             |

*Note:* Correlations between the health behavior variables present in the survey data. The first six are bad health behaviors, while the last 4 are good health behaviors. Behaviors with an asterisk (\*) can be considered exogenous from parks. Green numbers show a positive correlation between behaviors, while yellow numbers show negative correlations. Correlations below 0.05 in absolute value, in gray, are omitted from this analysis.

*Source:* own elaboration using data from CHS 2009–2013.

Hispanic, which are determined in equilibrium with the estimation of the model and allow for racial homophily.

The numbers presented in the table are averages of averages, for example, all neighborhoods have, on average, 3-floor buildings; the “lowest” neighborhood has an average of 1.7 floors, while the “tallest” neighborhood has an average of 11 floors. The geographic distribution of these variables can be found in the Appendix Figure A1.

## 2.4 | Characterizing Health Preferences

I take observable health behaviors as a proxy for unobservable health preferences.<sup>8</sup> Table 2 shows a correlation analysis between the different health behavior variables present in the survey (Appendix Table A1 presents the definition of each variable). Looking at correlation coefficients that are greater or equal to 0.05<sup>9</sup>, we can see that, in general, bad (good) health behaviors are positively correlated with each other and negatively correlated with good (bad) health behaviors.

The correlations presented show consistency in health behavior. It is not the case that individuals compensate some bad behaviors with other good behaviors; instead, these behaviors are consistent with a preference for being healthy. In other words, this indicates that behaviors such as colonoscopies or vaccines, which are more likely in the older population, are not a response to poor health but reflect a general preference for keeping good health. This consistency in behaviors allows us to create an index with these variables to proxy individuals' health preferences.

A valid concern is the possible reverse causality between neighborhood characteristics and health behavior. We could think that if we randomly allocate an individual close to a park, she starts

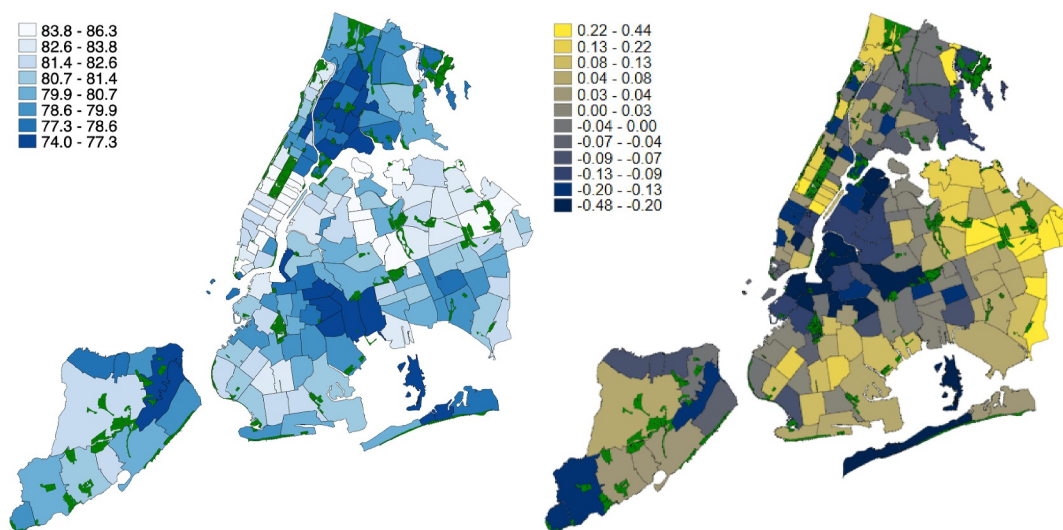
exercising, which then causes her to eat healthier and to smoke and drink less, while if the random allocation had been distant from parks, this change in behavior would not have happened. To minimize this concern, I generate two health preference indices. The first index uses all the variables in Table 2 while the second index uses the more exogenous health behavior variables marked with an asterisk. To generate the indices I add (subtract) 1 for each good (bad) health behavior and then standardize, so the index has mean 0 and standard deviation 1.<sup>10</sup>

This second health preference index is less affected by the endogeneity problem, since moving next to a park should not change whether an individual is an ever smoker (i.e., a pre-determined variable) or if she has unsafe sex.<sup>11</sup> Even if the second index is generated with less than half of the variables used for the first index, the correlation between the two is very high (0.73), so in the rest of the paper, I use the second, more exogenous health preference index.

In order for this index to be a good proxy for health preferences, it should be correlated with health outcomes. Figure 2 shows a map of life expectancy on the first panel, with data from the U.S. Small-area Life Expectancy Estimates Project, and a map of the health preference index on the right panel. This figure shows that the geographic inequalities of (the approximated) health preferences are very similar to the geographic inequalities in life expectancy, showing healthy areas mainly in the center of Manhattan and the north of Queens,<sup>12</sup> and an overall pairwise correlation of 0.37 (significant at the 1% level).

## 3 | Reduced Form Analysis

In this section, I present a correlational analysis of health, distance to parks, and housing prices that offers a straightforward view of the relations between these variables.



**FIGURE 2** | Life expectancy and health preferences. Life Expectancy 2010–2015. *Source:* National vital statistics system. Average of the Health preference index by ZIP code. Higher values indicate higher preferences for being healthy. *Source:* own elaboration using data from CHS, 2009–2013. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 3.1 | Health and Distance to Parks

If individuals with high health preferences would choose to live closer to health amenities such as parks, we should observe that in the data. Table 3 shows linear regressions at both neighborhood and individual level to see the relation between health preferences and distance to parks, controlling for many individual characteristics that can be correlated to health preferences. We see that, indeed, health preferences decline with distance to parks. This negative correlation is maintained if we control for individual's demographic characteristics and distance to the city center.

The signs of the correlations with all the variables included in the regression are those that we would expect: preferences for being healthy are higher for older (omitted category 18–24), highly educated, married, insured individuals, females, and potential parents, while the value of the health preference index is lower if an individual is poor or is white, black or Hispanic (compared to the omitted races, mainly composed by Asian).

If we look at the standardized coefficient (which allows to compare the strength of each explanatory variable) in column 5 we can see that the correlation with distance to parks is higher than the correlation between the health preference index and being poor.

### 3.2 | House Prices and Distance to Parks

If parks are considered amenities, we should observe higher prices in neighborhoods closer to parks; indeed, other papers in other contexts find a positive marginal willingness to pay to live close to parks (e.g., Poudyal et al. 2009; Wen et al. 2015; Du and Zhang 2020). Table 4 shows a negative correlation between house prices in a neighborhood and the distance to the closest park, which is maintained after controlling for other neighborhood and housing characteristics.

In Table 4, we also see that house prices are lower if we move away from the CBD and if there is a higher crime rate, while house prices are higher in neighborhoods with detached houses, bigger houses, and taller buildings.<sup>13</sup> The third column shows that the effect of distance to parks is 37% of the effect of distance to the CBD, which is thought of as one of the main determinants of house prices.

### 3.3 | Need for Structural Estimation

This reduced-form analysis has two important findings. The first one is a positive correlation between health preferences and living close to parks. Considering that this correlation is positive when we proxy preferences with health behavior variables that are unlikely to be affected by proximity to parks and that this correlation is maintained if we control for other relevant variables for health preferences, such as age, race, income, and education, this supports the hypothesis that individuals that care about their health choose to locate close to parks. The second important finding is that, as Table 4 shows, this preference for parks has an effect on the housing market by increasing housing prices close to parks.

The analysis of this section is useful to understand the basic relations between the variables of interest. However, a structural model that allows us to understand the interactions between health, location choice, and equilibrium prices is a better alternative. A model of location choice with endogenous prices allows to answer the following questions:

- What is the role of health preferences in individuals' location choice?
- How much of the geographical inequalities in health preferences can be explained by sorting?
- How do people relocate if there are changes in people's health preferences?

**TABLE 3** | Health preferences and distance to parks.

|                             | Neighborhood level   |                      | Individual level     |                      |                   |
|-----------------------------|----------------------|----------------------|----------------------|----------------------|-------------------|
|                             | (1)<br>Coef./std.err | (2)<br>Coef./std.err | (3)<br>Coef./std.err | (4)<br>Coef./std.err | (5)<br>Beta coef. |
| Distance park ( <i>Km</i> ) | −0.02<br>(0.02)      | −0.03*<br>(0.02)     | −0.03*<br>(0.02)     | −0.03**<br>(0.01)    | −0.02             |
| Distance CBD (min)          |                      | −0.00<br>(0.00)      |                      | −0.00*<br>(0.00)     | −0.01             |
| Poor                        |                      | 0.17<br>(0.17)       |                      | −0.03*<br>(0.02)     | 0.01              |
| College                     |                      | 0.43***<br>(0.14)    |                      | 0.15***<br>(0.02)    | 0.07              |
| Female                      |                      | 0.18<br>(0.54)       |                      | 0.17***<br>(0.02)    | 0.09              |
| Potential parent            |                      | −0.43<br>(0.49)      |                      | 0.09***<br>(0.03)    | 0.04              |
| Married                     |                      | 0.75**<br>(0.33)     |                      | 0.30***<br>(0.02)    | 0.15              |
| Age 25–44                   |                      | −0.52<br>(0.35)      |                      | −0.32***<br>(0.03)   | −0.16             |
| Age 45–64                   |                      | 0.53<br>(0.43)       |                      | 0.36***<br>(0.03)    | 0.17              |
| Age 65+                     |                      | 1.18***<br>(0.39)    |                      | 0.84***<br>(0.03)    | 0.30              |
| Insured                     |                      | −0.03<br>(0.20)      |                      | 0.24***<br>(0.02)    | 0.09              |
| White                       |                      | −0.35***<br>(0.10)   |                      | −0.27***<br>(0.03)   | −0.13             |
| Black/Hispanic              |                      | 0.02<br>(0.11)       |                      | −0.06**<br>(0.02)    | −0.03             |
| Constant                    | 0.02<br>(0.03)       | −0.47<br>(0.36)      | 0.04**<br>(0.02)     | −0.44***<br>(0.04)   |                   |
| Obs                         | 128                  | 128                  | 43659                | 43659                |                   |
| $R^2$                       | 0.01                 | 0.57                 | 0.00                 | 0.22                 |                   |

Note: The first two columns show the coefficients from a linear regression where the dependent variable is the mean of the health preference index by neighborhood. Columns 3, 4 and 5 use instead the individual health preference index. Column 5 shows the coefficients of column 4 standardized. Standard errors in parentheses, clustered by neighborhood.

Source: own elaboration using data from CHS 2009–2013 and distances calculated as explained in Section 2.

\* $p < 0.10$ .

\*\* $p < 0.05$ .

\*\*\* $p < 0.01$ .

— How does this affect geographical inequalities in health preferences and the equilibrium in the housing market?

An important characteristic of New York City is that location is not determined by inheritance of a house,<sup>14</sup> implying that most of its inhabitants actually choose where to live. First of all, the home ownership rate is around 30%. From the NYC-HANES survey 2014, we know that 50% of individuals moved in the

past five years while 90% moved after age 18. Other survey data<sup>15</sup> shows that 80% of the individuals who moved did so to a different neighborhood. This evidence allows me to interpret the observed location of individuals as their choice.

In the next section, I present a structural model that allows to study the location choice of individuals with heterogeneous demand for amenities and endogenous housing prices.

**TABLE 4** | House prices and distance to parks.

|                                            | Annual rent (1000 USD) |                        |                   |
|--------------------------------------------|------------------------|------------------------|-------------------|
|                                            | (1)<br>Coef./std.error | (2)<br>Coef./std.error | (3)<br>Beta coef. |
| Distance park ( <i>Km</i> )                | −0.57<br>(0.64)        | −1.09**<br>(0.54)      | −0.15             |
| Distance CBD (min)                         |                        | −0.10**<br>(0.04)      | −0.41             |
| Detached houses                            |                        | 5.74**<br>(2.47)       | 0.27              |
| Number of floors                           |                        | 1.13***<br>(0.35)      | 0.43              |
| Building age                               |                        | 0.01<br>(0.04)         | 0.03              |
| House size ( <i>m</i> <sup>2</sup> )       |                        | 0.07**<br>(0.03)       | 0.25              |
| Commercial area ( <i>Km</i> <sup>2</sup> ) |                        | 0.29<br>(0.23)         | 0.12              |
| Industrial area ( <i>Km</i> <sup>2</sup> ) |                        | 6.44**<br>(2.78)       | 0.19              |
| Crime rate (per 1000 res)                  |                        | −0.07***<br>(0.02)     | −0.22             |
| Constant                                   | 10.48***<br>(0.66)     | 2.38<br>(5.84)         |                   |
| Obs.                                       | 128                    | 128                    |                   |
| <i>R</i> <sup>2</sup>                      | 0.03                   | 0.46                   |                   |

Note: Coefficients from a linear regression at the neighborhood level. Column 3 shows the coefficients of column 2 standardized. Standard errors in parentheses. Source: own elaboration using data from different sources as explained in Section 2.

\**p* < 0.10.

\*\**p* < 0.05.

\*\*\**p* < 0.01.

#### 4 | Location Choice Model

To model the choice of neighborhood, I use a discrete choice model initially developed by McFadden (1973) and improved in the Industrial Organization literature (Berry et al. 1995; Nevo 2000). I follow Bayer et al. (2005) –which I will refer to as BMR– who adapted this model for the urban setting by providing equilibrium proofs for endogenously generated neighborhood characteristics, present, for example, in case of racial homophily.

The indirect utility of agent *i* in neighborhood *j* is given by:

$$V_{ij} = \beta^{DI} \log(I_i - P_j) + \mathbf{Z}_j' \beta_i^Z + \bar{\mathbf{R}}_j' \beta_i^R + \xi_j + \epsilon_{ij} \quad (1)$$

where *I<sub>i</sub>* is the individual's income, *P<sub>j</sub>* is the price of housing at neighborhood *j*, *Z<sub>j</sub>* is a vector of relevant characteristics of neighborhood *j* and  $\bar{R}_j$  is the average of individuals' race in neighborhood *j*, a variable determined in equilibrium. The coefficient  $\beta^{DI}$  captures the effect of Disposable Income, that is, the income individuals can dispose of after paying for housing. I

include the logarithm of income minus price to allow for wealth effects, that is, to account for the fact that individuals spend a considerable amount of their income on housing, implying that rents and income must be considered together. An alternative specification is considered at the end of this section with results in Section 5.3.

The error term is divided into two components,  $\xi_j$ , an unobserved (by the econometrician) element of neighborhood characteristics relevant for location choice, and  $\epsilon_{ij}$ , an individual specific taste shock. The vector of observed neighborhood characteristics, *Z<sub>j</sub>* contains the distance to the Central Business District (CBD), the distance to the closest park and other housing and neighborhood characteristics, such as the average building age in the neighborhood or the commercial area. The CBD is a proxy for occupation location and for other amenities such as theaters.

The model allows each individual to respond differently to *Z<sub>j</sub>* and  $\bar{R}_j$ :

$$\beta_{i,k}^r = \beta_{0,k}^r + \beta_{1,k}^r \mathbf{y}_i, \quad r = Z, R \quad (2)$$

where  $k$  indicates each component of  $\mathbf{Z}_j$  and  $\bar{R}_j$ , and  $\mathbf{y}_i$  is a vector of individual characteristics, such as education, race, and health preferences. This specification relaxes the Independence of Irrelevant Alternatives assumption since different individuals will have a different ordering of preferred alternatives (Nevo 2000).

#### 4.1 | Estimation of Parameters

It is convenient to rewrite this model grouping terms that vary only by alternative (neighborhood) and those that vary by individual and alternative.

$$V_{ij} = \lambda_{ij} + \delta_j + \epsilon_{ij}$$

$$\lambda_{ij} = \beta^{DI} \log(I_i - P_j) + \mathbf{Z}'_i \beta^Z_1 \mathbf{y}_i + \bar{\mathbf{R}}'_j \beta^R_1 \mathbf{y}_i \quad (3)$$

$$\delta_j = \mathbf{Z}'_j \beta^Z_0 + \bar{\mathbf{R}}'_j \beta^R_0 + \xi_j \quad (4)$$

Individuals will choose the neighborhood that gives them the highest utility, that is, individual  $i$  will choose neighborhood  $j$  only if

$$V_{ij} \geq V_{ik} \iff (\lambda_{ij} + \delta_j) - (\lambda_{ik} + \delta_k) \geq \epsilon_{ik} - \epsilon_{ij} \quad \forall k \neq j$$

Assuming that the individual specific shocks follow the extreme value distribution, the probability that individual  $i$  chooses neighborhood  $j$  is then given by:

$$p_{ij} = \frac{\exp(\delta_j + \lambda_{ij})}{\sum_l \exp(\delta_l + \lambda_{il})} \quad (5)$$

If we take  $\delta_j$  as neighborhood fixed effects, we get a Conditional Logit with alternative varying regressors and an alternative specific constant. McFadden (1973) showed that this model can be estimated through maximum likelihood by selecting the parameters that increase the probability that each individual makes the choice observed in the data.<sup>16</sup>

The coefficients  $\delta_j$  can be estimated by the contraction mapping of Equation (6), for each combination of parameters in  $\lambda_{ij}$ , where  $t$  denotes the iteration. Take, for example, an initial case in which the observed share of individuals living in neighborhood  $j$  ( $S_j$ ) is higher than the share calculated by the model ( $\sum_i \hat{p}_{ij}^t$ ), then the logarithm of a number greater than 1 will increase the value of  $\delta$  for the next iteration. That higher  $\delta^{t+1}$  increases the value of  $\hat{p}_{ij}^{t+1}$ , bringing the summation closer to  $S_j$ . This procedure is iterated until an equilibrium is reached. In this way, the  $\delta_j$  coefficients will be those that equate the observed shares of each alternative,  $S_j$  (the supply of housing in each neighborhood) to the shares estimated by the model (the demand for housing in each neighborhood).

$$\delta_j^{t+1} = \delta_j^t + \log\left(\frac{S_j}{\sum_i \hat{p}_{ij}^t}\right) \quad (6)$$

The  $\beta_0$ 's of  $\delta_j$  can be estimated in a second stage, through a linear regression of Equation (4), by plugging in the left side the  $\delta_j$  estimated with Equation (6) (through convergence of the contraction mapping), and the observed neighborhood characteristics in the right side. The residuals of this regression are the unobserved neighborhood characteristics.

The last step is to take into account that the data used is survey data, which implies that each observation has a different weight, affecting the relative importance of each  $p_{ij}$  (Bruch and Mare 2012).

#### 4.2 | Identification of Parameters

This interaction coefficient between distance to parks and an individual's health preference index will be an unbiased estimator of the preferences that individuals with high health preferences have to live close to parks only if distance to the closest park is uncorrelated with other unobservable characteristics captured in  $\xi_j$ . This exogeneity assumption would be invalid in case of endogenous health amenities if, for example, healthy stores would open in the vicinity of parks. The coefficient would nevertheless capture the preferences of intrinsically healthy individuals to live closer to healthy neighborhoods. A discussion of the exogeneity assumption of distance to parks and of the interpretation of results in the presence of endogenous health amenities and peer effects in health is presented in Section 7.

There are two other assumptions regarding the endogenous variables of the model  $P_j$  and  $\bar{R}_j$ , since components in  $\xi_j$  probably influence prices and can influence a neighborhood's racial composition. The potential endogeneity between racial composition and  $\xi_j$  is of minor concern as long as we assume that a neighborhood's racial composition is independent of distance to parks and individual's health preferences, and this is not a very strong assumption.

Regarding prices, the assumption that allows the identification of parameters without instruments is the functional form specification of income and house prices. The advantage of this specification, apart from being more appropriate by allowing wealth effects, is that  $\delta_j$  (which comprises  $\xi_j$ ) acts as a neighborhood fixed effect and addresses the endogeneity problem between prices and the unobserved characteristics in  $\xi_j$ , by estimating the parameters on within-neighborhood variation. There is no endogeneity concern in the second stage of the estimation either, since prices are not an explanatory variable of that regression.

An alternative specification, more common in the literature, is to let utility depend linearly on income minus price (instead of log-linearly). In this case income becomes irrelevant since it does not vary by neighborhood, and prices are captured by  $\delta_j$ , entering the second stage of the estimation where we need to address the endogeneity present between prices and the error term of this second stage  $\xi_j$ . This alternative specification with a

suitable instrument for prices is estimated as robustness and presented in Section 5.3.

## 5 | Results

This section shows the parameter estimates of the location choice model, calculations of marginal willingness to pay, and an analysis of model performance. Definitions and sources of all the variables used were presented in Section 2.

### 5.1 | Parameter Estimates

Table 5 shows important interaction coefficients between neighborhood and individual characteristics: the  $\beta_1$  (Tables A2 and A3 show the full set of coefficients). The signs of the coefficients can be directly interpretable, and in general, validate the findings of Section 3.

The main coefficient of interest is that of the interaction between distance to parks and the health preference index. This coefficient is negative and significant, which implies that people who care about their health are less likely to choose neighborhoods that are more distant from parks. In other words, after controlling for other reasons why people might value parks (proxied by age, having children, income, or education), an individual's health preferences have predictive power in the choice of neighborhoods close to parks.

To understand how the Marginal Willingness to Pay (MWTP) for living close to parks compares to the MWTP for other neighborhood characteristics, I calculate the MWTP for a change in one standard deviation of each characteristic for a typical New Yorker (non-poor, educated, and in working age) and a typical New Yorker who is also healthy (health preference index above average). Results are presented in Table A4 in the Appendix. These calculations show that an individual with high health preferences would be willing to increase her annual rent by 11 thousand dollars to live one standard deviation (20 min)

closer to the CBD, a similar amount to a typical individual, but she would be willing to increase her rent by almost 5 thousand dollars to live one standard deviation closer to a park, *three times as much* as a typical New Yorker. Considering that both income and the actual rents paid by individuals are approximations, these values are useful mainly to compare the relative effect of each characteristic and should not be taken as precise quantitative estimates. A more detailed discussion is presented in Section 7.

### 5.2 | Model Performance

I use the model to assign each individual to a neighborhood according to the estimated probabilities (the exact procedure is explained in Appendix C). As a measure of fit, I look at the percentage of the true variance in the health preference index that the model is able to predict, similar to an  $R^2$  measure. The model is able to capture 61% of the variation, so it is able to replicate a lot of the geographical variation in health behaviors observed in the data. The left panel of Figure A2 shows the predicted geographic distribution of health preferences, which looks very similar to the actual distribution of preferences in the right panel of Figure 2.

### 5.3 | Robustness Checks

I perform several robustness checks that confirm the findings presented so far. Appendix table A5 shows robustness to the definition of income. The first column shows the coefficients of the interaction with the health preference index for a model that allows for heterogeneous preferences by disposable income, that is, estimating  $\beta_i^{DI}$  instead of  $\beta^{DI}$ . All coefficients are similar to those of the main specification.

The second column shows the estimated coefficients of a model that includes income linearly instead of log-linearly. This implies that income drops from the estimation because it does not vary with neighborhood choice. Annual rent, which then varies only

TABLE 5 | Selected interaction coefficients.

|                | Poor       | College   | Parent     | Age65 m   | White    | Health index |
|----------------|------------|-----------|------------|-----------|----------|--------------|
| Distance CBD   | −0.0044**  | −0.008*** | −0.0029    | −0.0012   | 0.0022   | −0.0014      |
| Distance park  | −0.0365    | −0.0328   | −0.0206    | 0.0193    | −0.034   | −0.0436**    |
| Detached house | −0.194*    | 0.2984*   | −0.0741    | 0.0626    | −0.2269* | 0.1081       |
| Number floors  | −0.0674*** | 0.087***  | −0.0173    | −0.0061   | −0.0083  | 0.033***     |
| Building age   | 0.0003     | 0.0043**  | −0.0013    | 0.0000    | −0.0016  | −0.0013      |
| House size     | −0.0034*** | 0.0022    | 0.0003     | 0.0008    | −0.0012  | 0.0018**     |
| Comm area      | −0.0074    | −0.0121   | 0.0113     | 0.0084    | −0.0103  | −0.0072      |
| Indus area     | −0.0192    | −0.1207   | −0.1546    | −0.2663** | 0.1429*  | −0.0032      |
| Crime rate     | 0.0001     | −0.0019   | −0.0025*** | −0.0002   | −0.0008  | −0.0007      |

Note: Coefficients from the interactions between neighborhood and individual characteristics. Omitted columns show interaction coefficients with *female*, *married*, *age25–44*, *age45–64*, *insured* and *black/hisp*, the full set of coefficients and standard errors is in Appendix Table A2. Std errors and *p*-values calculated with clustering at the neighborhood level (since there are *j* alternatives for each individual).

\**p* < 0.10.

\*\**p* < 0.05.

\*\*\**p* < 0.01.

by alternative, becomes part of  $\delta_j$  and enters the second stage of the estimation (Equation 4 in Section 4). The likely correlation between rents and the unobserved characteristics  $\xi_j$  requires the instrumentation of these prices. Appropriate instruments for  $P_j$  are the exogenous characteristics of neighborhoods adjacent to  $j$ , an instrument very similar to that used by Bayer et al. (2007). The logic is the following: when deciding on a neighborhood, a valuable but scarce characteristic of houses in neighborhood  $k$ , adjacent to  $j$ , should increase prices of houses with that characteristic also in neighborhood  $j$ , but once an individual chooses neighborhood  $j$ , her utility is not affected by the characteristics of houses in  $k$ , that is, the characteristics of adjacent neighborhoods affect prices in  $j$  but not utility in  $j$ . Table A7 in the Appendix shows the second stage of the estimation using (only) exogenous characteristics of adjacent neighborhoods (building age, number of floors, etc.) as instruments for prices, while Table A6 shows instrument relevance with an F-statistic of 19.6. The results are very similar to those of the main specification.

To check robustness with respect to the main variables of the analysis, I estimate the model using the health preference index calculated with all the health behavior variables instead of the more exogenous health preference index of the main specification. The interaction coefficient between distance to the closest park and the health preference index gives a negative and significant coefficient. I also consider an alternative index formed only with the variables *ever smoker* and *unsafe sex*, omitting *flu vaccine* and *colonoscopy*, to take into account the possibility the latter two are capturing individuals' frailty (and therefore bad health) or quality of health insurance. Approximating health preferences with only these two variables is somewhat far-fetched<sup>17</sup> (the correlation between this index and the index with all health behaviors is 0.48), but the sign of the interaction coefficient between this index and distance to parks remains negative. These results are presented in the first two columns of Table A8.

The other columns of this table show robustness to the variable measuring distance to parks. Column 3 shows that using park area in each neighborhood instead of distance to the closest park gives a positive (even if non-significant) coefficient (0.014), suggesting that individuals with high health preferences are more likely to choose neighborhoods with (larger) parks in them. In Column 5, instead, the model is estimated with the logarithm of distance to parks to allow for nonlinear effects.

The last column of Appendix Table A8 shows the results of the estimation with a weighted version of distance to parks to control for park quality. It is possible that some parks are less *amoenus* than others because they are dangerous. Some parks in NYC have relatively high levels of crime and therefore lose their amenity value.<sup>18</sup> To take this into account, I weigh distance to parks by multiplying it by the logarithm of the number of crimes per  $Km^2$  in that park (which makes unsafe parks more distant than they actually are). The correlation between the raw and the weighted distance is 0.83, while the correlation between the raw and weighted distances with the survey question on exercise (at the neighborhood level) goes from  $-0.15$  to  $-0.17$ , which shows that correcting by crime levels at parks helps to capture their amenity value. As Table A8 shows, this coefficient is negative and highly significant.

Finally, to check robustness to the sample, I estimate the model excluding each borough at the time. Excluding Manhattan, for example, rules out the possibility that the effects are driven by Central Park. By excluding Staten Island, I also reduce the concerns that the results are driven by recently established parks. The resulting coefficients are numerically similar, and the qualitative results are maintained. Appendix Table A9 presents these coefficients for the most relevant variables. This check shows that the statistical results hold even with subsamples of the data, but the underlying assumption is a reduction of the choice set, which cannot be justified on empirical grounds.

## 6 | Improvements in Individuals' Health Preferences

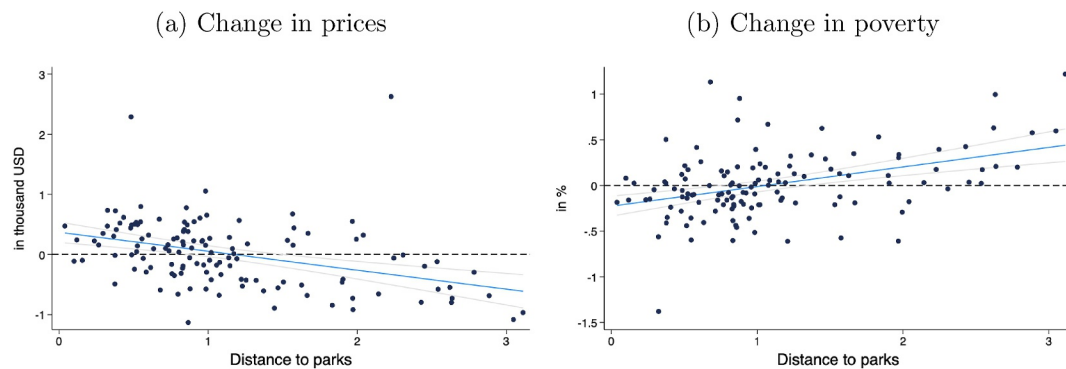
In this section, I use the model to understand the consequences in terms of city configurations and equilibrium prices in response to an improvement in residents' health preferences.

Changes in individuals' characteristics will change the probabilities with which each agent chooses their neighborhood. These changes in demand for neighborhoods, keeping supply fixed, must be accompanied by new housing prices to find a new equilibrium. Furthermore, since people have preferences over the racial composition of neighborhoods, we must make sure that the model is in equilibrium regarding both prices and racial composition. The steps to calculate the new equilibrium are presented in Appendix C.

A widely used policy to improve health is information campaigns. For example, from 2002 to 2013, the New York City government used information campaigns to address tobacco use, obesity, and HIV infection (Fairchild et al. 2015). To run this counterfactual analysis, I increase everyone's health preference index by 0.3 standard deviations, which is the observed improvement in the period 2009–2019.<sup>19</sup>

Figure A2 shows the geographical distribution of health preferences predicted by the model and the changes in the geographic distribution of health preferences with the counterfactual evaluation. Considering that the counterfactual increases everyone's health preferences, we could expect no changes in the distribution of these preferences by neighborhood; instead, the simulations show that some neighborhoods become healthier while others become less healthy, in particular, some originally unhealthy neighborhoods (low average health preference index) become even more unhealthy after the implementation of the information campaign. There are, however, many variables that affect individuals' location decisions in different directions, so there is no clear pattern of relocation. The average observed changes in the health preferences of neighborhoods are small, with a maximum effect of  $\pm 0.01$  standard deviations.

Where we do see stronger effects is in equilibrium prices. The left panel of Figure 3 shows the changes in prices with the counterfactual evaluation, with a clear negative (and significant) relation with distance to the closest park. We can see in Figure A3, which shows the geographic distribution of these



**FIGURE 3** | Counterfactual - effects on Prices and Poverty prevalence. The left panel shows the difference between equilibrium prices from the counterfactual evaluation and benchmark prices in the y-axis (in thousands of USD), and distance to the closest park in the x-axis. The right panel shows the difference between the poverty prevalence predicted in the counterfactual evaluation and benchmark poverty. Both plots include the 95% confidence interval. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

changes, that the simulation predicts reductions in prices of up to USD 1100 in annual rent in the Southwest part of Queens, since this is an area that had relatively large prices but is the most distant from parks. For the North of Queens and a big part of Manhattan and the Bronx the simulation predicts price increases, being areas closer to parks. On average, prices of neighborhoods within 1 Km of a park show an increase in the annual rent of USD 188, a 1.8% increase with respect to original prices in those neighborhoods. Instead, neighborhoods that are 1 Km distant from parks show a reduction in annual rents of USD 213, a 2.5% reduction, while those that are more than 2 Km distant from parks show a decrease of USD 286, a 3.3% reduction with respect to the original prices.<sup>20</sup>

Since poor individuals respond very strongly to prices, these changes are also reflected in the new equilibrium of poverty prevalence in each neighborhood. The right panel of Figure 3 shows a (significant) increase in poverty prevalence in neighborhoods further from parks. As can be seen in Figure A3, the neighborhoods that become cheaper attract poorer individuals who are pushed away from neighborhoods that become more expensive.

This counterfactual simulation shows that a policy that can be considered successful from a public health point of view can generate unexpected negative consequences through interactions with the housing market. The rise in demand for neighborhoods closer to health amenities due to the health campaign causes an increase in prices in these neighborhoods, while neighborhoods that are distant from health amenities experience a reduction in prices. Poor individuals, therefore, relocate to neighborhoods further from health amenities, a negative result in terms of equity of access to health amenities.

Promoting better health behaviors is a first-order policy concern, therefore, understanding the connection between health preferences and the housing market can help us in defining policies for modern cities where the population is increasingly aware of the benefits of a healthy lifestyle.

## 6.1 | Model Predictions on Real Data

To check whether the simulation findings translate into real world findings, I calculate housing prices for 2019 (using the same data source and data cleaning procedures explained in Section 2). Table 6 shows regressions of annual rents on neighborhood characteristics in two panels, excluding or including borough fixed effects. Column 1 of each panel uses observed prices for the period 2009–2013 (repeating the results presented in Table 4). Column 2 of each panel uses instead the prices from the simulation exercise, while column 3 uses observed prices in 2019.

These regressions show that the correlation between house prices and distance to parks is stronger for both, the simulation prices, and the observed prices in 2019, confirming that parks are more valued in both of these scenarios. Since health preferences were higher in 2019 than in 2009, by the 0.3 standard deviations used in the simulation, this table provides evidence that simulation findings translate into real-world findings.

As an example, the simulation shows an increase in prices in Harlem and Washington Heights, both in the north of Manhattan, which are two of the most gentrified neighborhoods in the United States in the 2000s (Szekely 2018), providing anecdotal evidence of these effects on real data. These increases in prices due to increased preferences for carrying a healthy lifestyle, could then be one of the causes behind the gentrification of some neighborhoods.

## 7 | Discussion

In this section, I briefly discuss identification issues, the implications of the data that I use, and the interpretation of my results.

### 7.1 | Identification of Preferences

The interaction coefficient between distance to the park and the health preference index can be interpreted as evidence of higher

**TABLE 6** | Equilibrium house prices under counterfactual scenarios.

|                                            | Annual rent (1000 USD) |                    |                   |                    |                   |                   |
|--------------------------------------------|------------------------|--------------------|-------------------|--------------------|-------------------|-------------------|
|                                            | (Without borough FEs)  |                    |                   | (With borough FEs) |                   |                   |
|                                            | 09–13                  | Counterfactual     | 2019              | 09–13              | Counterfactual    | 2019              |
| Distance CBD (min)                         | −0.10**<br>(0.04)      | −0.12***<br>(0.04) | −0.23*<br>(0.13)  | −0.07<br>(0.05)    | −0.08*<br>(0.05)  | −0.21<br>(0.17)   |
| Distance park ( <i>Km</i> )                | −1.09**<br>(0.54)      | −1.38**<br>(0.53)  | −2.54<br>(1.74)   | −0.94*<br>(0.56)   | −1.23**<br>(0.54) | −3.80**<br>(1.85) |
| Detached houses                            | 5.74**<br>(2.47)       | 7.05***<br>(2.40)  | 11.40<br>(7.93)   | 1.26<br>(3.33)     | 2.35<br>(3.24)    | 3.60<br>(11.04)   |
| Number of floors                           | 1.13***<br>(0.35)      | 1.37***<br>(0.34)  | 1.78<br>(1.12)    | 0.62<br>(0.43)     | 0.89**<br>(0.42)  | 2.73*<br>(1.42)   |
| Building age                               | 0.01<br>(0.04)         | 0.00<br>(0.03)     | 0.14<br>(0.11)    | 0.02<br>(0.04)     | 0.01<br>(0.04)    | 0.18<br>(0.13)    |
| House size ( <i>m</i> <sup>2</sup> )       | 0.07**<br>(0.03)       | 0.08***<br>(0.03)  | 0.12<br>(0.09)    | 0.10***<br>(0.03)  | 0.11***<br>(0.03) | 0.12<br>(0.10)    |
| Commercial area ( <i>Km</i> <sup>2</sup> ) | 0.29<br>(0.23)         | 0.21<br>(0.22)     | 0.67<br>(0.73)    | 0.32<br>(0.22)     | 0.24<br>(0.21)    | 0.46<br>(0.73)    |
| Industrial area ( <i>Km</i> <sup>2</sup> ) | 6.44**<br>(2.78)       | 6.41**<br>(2.70)   | 4.86<br>(8.94)    | 6.82**<br>(2.73)   | 6.81**<br>(2.66)  | 1.79<br>(9.05)    |
| Crime rate (per 1000 res)                  | −0.07***<br>(0.02)     | −0.06***<br>(0.02) | −0.19**<br>(0.08) | −0.04*<br>(0.03)   | −0.04*<br>(0.02)  | −0.12<br>(0.08)   |
| Constant                                   | 2.38<br>(5.84)         | 2.12<br>(5.68)     | 2.98<br>(18.77)   | −3.37<br>(6.04)    | −3.74<br>(5.88)   | −6.65<br>(20.04)  |
| Borough FE                                 | No                     | No                 | No                | Yes                | Yes               | Yes               |
| Obs.                                       | 128                    | 128                | 128               | 128                | 128               | 128               |
| <i>R</i> <sup>2</sup>                      | 0.46                   | 0.52               | 0.31              | 0.52               | 0.57              | 0.35              |

*Note:* Coefficients from a linear regression of equilibrium prices on neighborhood characteristics, with and without borough fixed effects. The first and fourth columns show the regression coefficients of the data used for the calibration of the model (observed rents for the period 2009–2013). The second and fifth columns show regression coefficients of the equilibrium rent under the counterfactual. The third and sixth columns use observed rents in 2019. Standard errors in parenthesis.

\**p* < 0.10.

\*\**p* < 0.05.

\*\*\**p* < 0.01.

preferences to live close to parks for individuals who care about their health. The assumption behind this interpretation is that  $\xi_j$  does not include variables correlated with distance to parks.

The following, however, is also possible. A group of people who care about their health locate close to parks because they have a preference to stay near green areas. Given the demand for healthy products or services by these people, healthy stores or gyms open in the vicinity of parks, attracting other healthy individuals who did not have particular preferences for parks. As long as these healthy products or services do not affect the behaviors used for the construction of the health preference index (and in the absence of peer effects in these behaviors), the coefficient of interest would still be unbiased, but in this case, it would capture higher preferences to live close to health amenities, not exclusively parks.

The coefficient of interest will be biased if these endogenous amenities affect the behaviors used for the construction of the

index. For example, if clinics open in the vicinity of parks, they can increase the number of colonoscopies. Nevertheless, the only sensible reason for clinics to open in the vicinity of parks is the existence in these areas of people with higher demand for their services, that is people with higher health preferences, which supports the sorting mechanism. In this case, I do not capture preferences for living close to parks, but I capture a combined effect that must be initiated by preferences for living close to parks. The same is true if we consider the possibility of peer effects in flu vaccines, colonoscopies, and unsafe sex.

The existence of peer effects in the “exogenous” behaviors and the endogenous generation of health services that may affect these behaviors is plausible. This implies that the identification strategy that I propose has drawbacks. Unfortunately, finding a more suitable instrument in this setting is extremely difficult due to the endogenous nature of these variables and the limitations on data availability on individuals’ health, demographics, and location, given their sensitivity and data protection regulations.

## 7.2 | Data Limitations

Prices and income are extremely important for the choice of neighborhood. Unfortunately, it is very difficult to have access to detailed data on health that also provides detailed information on income and geographical location. The data I use provides information on five income categories, and I constructed representative house prices for these five income groups. These approximations are useful to obtain an estimate of the effect of income and house prices on location choice and give us ballpark figures of individuals' marginal willingness to pay.

Another limitation is the relative coarseness of ZIP codes as the geographical unit of observation. This data limitation implies that many neighborhoods are composed of thousands of housing units, with possibly different living conditions within them that cannot be distinguished and with some margin of error in the calculation of distance to the closest park. The robustness of the results to the use of park area instead of distance to parks is, however, reassuring.

All these limitations imply that the exact quantitative results should be taken with a grain of salt. Nevertheless, the robustness of the main results, plus the intuitive direction of the relations between variables, indicate that despite these limitations, the qualitative conclusions of this analysis are valuable.

## 7.3 | Implications of Rent Regulations

New York City has around 1 million housing units under rent regulation, representing almost 30% of all housing (Reiss 2022) (only 1% of housing units are under the more strict “rent control” policy). Rent-regulated units are hard to inherit (since they require the new tenant to have lived in that unit for at least 2 years), and their rent is not so different from market values, especially taking into account that landlords compensate lower rents with lower maintenance (Gyourko and Linneman 1990). Landlords of rent-regulated units cannot increase rent above official levels agreed upon every year. During the years of this study (2009–2013), rents were allowed to increase between 2% and 4% per year.

Under the counterfactual scenarios, the largest rent increases are around 11% in 10 years, which corresponds to yearly increases of 1.1%, which are below the rent increase caps. In this sense, the policies on rent regulation should not affect the results or implications of my analysis.

## 8 | Conclusion

This paper addresses one acknowledged but untested channel in the neighborhood-health relation. I propose and estimate a model of location choice that takes into account health amenities, proxied by parks, and health preferences, proxied by health behavior. I find that individuals who care about their health are more likely to choose neighborhoods close to health amenities, providing empirical evidence of the sorting

mechanism behind the observed differences in health behaviors between neighborhoods of the same city.

I use the model for a counterfactual evaluation that shows unintended consequences in house prices and the displacement of poor individuals after a well-intended information campaign on health. This evaluation brings to light an important connection between individuals' health preferences and house prices and implies that the increased attention to a healthy lifestyle can be a contributing factor to gentrification. Consequently, any policy that tries to affect health preferences must take into account potential equilibrium effects in the housing market.

This is the first paper studying sorting based on health preferences and health amenities in order to understand the observed large inequalities in health behaviors between neighborhoods and the consequences of this sorting on the housing market. Improvements in data availability, such as panel data on individuals' health, income, and geographical location at the neighborhood level, which allow incorporating forward-looking behaviors, or data at finer geographical levels, which allow for alternative ways to deal with the endogeneity of location choice, health, and amenities, will allow future work to generate further insights into the mechanics and consequences of location choice based on health preferences and health amenities.

---

### Acknowledgments

I thank Jerome Adda, Gianmarco Ottaviano, and Matthias Sutter for suggestions that substantially improved this research. Open Access funding enabled and organized by Projekt DEAL.

### Funding

The author has nothing to report.

### Ethics Statement

From the ethics committee of Bocconi University, No SA000127.

### Conflicts of Interest

The author declares no conflicts of interest.

### Data Availability Statement

The data that support the findings of this study are available from NYC Department of Health and Mental Hygiene. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of NYC Department of Health and Mental Hygiene.

### Endnotes

<sup>1</sup> World Health Organization 2019.

<sup>2</sup> In the data made available by the DOHMH, respondents residing in a ZIP Code with a population less than 30,000 were combined to an adjacent ZIP code within their United Hospital Fund (UHF). A discussion of taking neighborhoods as ZIP codes is presented in section 7.

<sup>3</sup> House prices in NYC reached a peak in 2007 and touched bottom in the first quarter of 2009. In the second quarter of 2013, prices were back to the 2007 level.

- <sup>4</sup> The average characteristics of 2009 are virtually equal to those of 2013. This is a confirmation that these characteristics are fixed in the short term.
- <sup>5</sup> Violent crime includes murder, rape and felony assault, to which I add dangerous drugs, and dangerous weapons, given their relevance both in prevalence and in perceived neighborhood quality. I do not include robbery (officially classified as violent crime) since property crimes are mechanically positively correlated with rents, which introduces noise into the correlations.
- <sup>6</sup> Using Wall Street gives practically the same results.
- <sup>7</sup> The website <https://www.nycgovparks.org/about/history/timeline> shows the history of NYC parks. Only seven large parks were created after 1950.
- <sup>8</sup> Behaviors are the result of preferences, information, beliefs, and other constraints. I cannot separate these factors from the available data, but this is not fundamental for the purposes of my analysis.
- <sup>9</sup> Given the large sample size (> 40K obs.) almost all correlations are significant at the 1% level.
- <sup>10</sup> Principal Component Analysis (pca) gives highly correlated indices. The disadvantage of pca is that it cannot be calculated when some observations are missing, which greatly restricts the sample size, given many missing observations for colonoscopies and unsafe sex (see Table 1).
- <sup>11</sup> See Section 7 for a discussion on the implications of this assumption and the effects of peers and endogenous amenities.
- <sup>12</sup> A map of New York City boroughs is presented in Appendix A.
- <sup>13</sup> This regression also shows higher prices for neighborhoods with large industrial areas. This coefficient is very dependent on one neighborhood, Long Island City, a redeveloped industrial area that has become increasingly more expensive.
- <sup>14</sup> Rent-controlled apartments are usually inherited, but they represent less than 1% of housing units. Apartments under rent regulation, which are quite common, are hard to inherit. A discussion on the effects of rent regulation is presented in Section 7.
- <sup>15</sup> <https://www.brickunderground.com/live/why-new-yorkers-move-so-often>.
- <sup>16</sup> The Conditional Logit parameters in this model are estimated with Stata's in-built *cmlogit* function, which uses the Newton-Raphson optimization algorithm and results are robust to different starting values. The same parameters are obtained with Matlab's *fminunc* function, based on the Quasi-Newton method.
- <sup>17</sup> Consider that relatively old adults could have smoked when they were young, when the negative effects of smoking on health were not as evident as today.
- <sup>18</sup> In 10% of parks, there are three or more crime events per year (considering murder, rape, and felony assault; dangerous drugs and dangerous weapons are not reported). The median number of crimes per park is zero.
- <sup>19</sup> For this calculation, I used the publicly available yearly datasets from the Community Health Survey of DOHMH (2019). The improvement in health behavior is virtually the same in both high- and low-educated individuals, plus Stead et al. (2019) find neutral or inconsistent evidence of differential effects of information campaigns by subpopulations such as ethnicity or socioeconomic group.
- <sup>20</sup> A discussion of the effect of rent regulation on market equilibrium prices is presented in Section 7.

## References

Albouy, D., W. Graf, R. Kellogg, and H. Wolff. 2016. "Climate Amenities, Climate Change and American Quality of Life." *Journal of the*

*Association of Environmental and Resource Economists* 3, no. 1: 205–246. <https://doi.org/10.1086/684573>.

Alexander, D., and J. Currie. 2017. "Is it Who You Are or Where You Live? Residential Segregation and Racial Gaps in Childhood Asthma." *Journal of Health Economics* 55: 186–200. <https://doi.org/10.1016/j.jhealeco.2017.07.003>.

Allcott, H., R. Diamond, J. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *Quarterly Journal of Economics* 134, no. 4: 1793–1844. <https://doi.org/10.1093/qje/qjz015>.

Bayer, P., F. Ferreira, and R. McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115, no. 4: 588–638. <https://doi.org/10.1086/522381>.

Bayer, P., N. Keohane, and C. Timmins. 2009. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management* 58, no. 1: 1–14. <https://doi.org/10.1016/j.jeem.2008.08.004>.

Bayer, P., R. McMillan, and K. Rueben. 2005. *An Equilibrium Model of Sorting in an Urban Housing Market*. Working Paper 10865. NBER.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63, no. 4: 841–890. <https://doi.org/10.2307/2171802>.

Bilger, M., and V. Carrieri. 2013. "Health in the Cities: When the Neighborhood Matters More than Income." *Journal of Health Economics* 32, no. 1: 1–11. <https://doi.org/10.1016/j.jhealeco.2012.09.010>.

Bishop, K., N. V. Kuminoff, S. Mathes, and A. Murphy. 2024. "The Marginal Cost of Mortality Risk Reduction: Evidence From Housing Markets." *Journal of Urban Economics* 139: 103627. <https://doi.org/10.1016/j.jue.2023.103627>.

Bruch, E. E., and R. D. Mare. 2012. "Methodological Issues in the Analysis of Residential Preferences, Residential Mobility, and Neighborhood Change." *Sociological Methodology* 42, no. 1: 103–154. <https://doi.org/10.1177/0081175012444105>.

Chay, K. Y., and M. Greenstone. 2005. "Does Air Quality Matter? Evidence From the Housing Market." *Journal of Political Economy* 113, no. 2: 376–424. <https://doi.org/10.1086/427462>.

Dang, R. 2015. Spillover Effects of Local Human Capital Stock on Adult Obesity-Evidence From German Neighborhoods: SSRN 2683509.

Darden, M. E. 2021. "Cities and Smoking." *Journal of Urban Economics* 122: 103319. <https://doi.org/10.1016/j.jue.2021.103319>.

Deryugina, T., and D. Molitor. 2020. "Does when You Die Depend on Where You Live? Evidence From Hurricane Katrina." *American Economic Review* 110, no. 11: 3602–3633. <https://doi.org/10.1257/aer.2018.1026>.

Deryugina, T., and D. Molitor. 2021. "The Causal Effects of Place on Health and Longevity." *Journal of Economic Perspectives* 35, no. 4: 147–170. <https://doi.org/10.1257/jep.35.4.147>.

DOHMH. 2009–2013. *Community Health Survey*. Department of Health and Mental Hygiene. Restricted access to individuals' ZIP Code

DOHMH. 2009–2019. *Community Health Survey*. Department of Health and Mental Hygiene. <https://www1.nyc.gov/site/doh/data/data-sets/community-health-survey-public-use-data.page>.

Du, M., and X. Zhang. 2020. "Urban Greening: A New Paradox of Economic or Social Sustainability?" *Land Use Policy* 92: 104487. <https://doi.org/10.1016/j.landusepol.2020.104487>.

Eid, J., H. G. Overman, D. Puga, and M. A. Turner. 2008. "Fat City: Questioning the Relationship Between Urban Sprawl and Obesity." *Journal of Urban Economics* 63, no. 2: 385–404. <https://doi.org/10.1016/j.jue.2007.12.002>.

- Ellen, I. G., T. Mijanovich, and K.-N. Dillman. 2001. "Neighborhood Effects on Health: Exploring the Links and Assessing the Evidence." *Journal of Urban Affairs* 23, no. 3–4: 391–408. <https://doi.org/10.1111/0735-2166.00096>.
- Fairchild, A. L., R. Bayer, and J. Colgrove. 2015. "Risky Business: New York City's Experience With Fear-Based Public Health Campaigns." *Health Affairs* 34, no. 5: 844–851. <https://doi.org/10.1377/hlthaff.2014.1236>.
- Finkelstein, A., M. Gentzkow, and H. Williams. 2021. "Place-Based Drivers of Mortality: Evidence From Migration." *American Economic Review* 111, no. 8: 2697–2735. <https://doi.org/10.1257/aer.20190825>.
- Gourevitch, M. 2019. "Large Life Expectancy Gaps in US Cities Linked to Racial and Ethnic Segregation by Neighborhood." Published on June 5, 2019. <https://nyulangone.org/news/large-life-expectancy-gaps-us-cities-linked-racial-ethnic-segregation-neighborhood>.
- Gyourko, J., and P. Linneman. 1990. "Rent Controls and Rental Housing Quality: A Note on the Effects of New York City's Old Controls." *Journal of Urban Economics* 27, no. 3: 398–409. [https://doi.org/10.1016/0094-1190\(90\)90009-c](https://doi.org/10.1016/0094-1190(90)90009-c).
- Heblich, S., A. Trew, and Y. Zylberberg. 2021. "East-Side Story: Historical Pollution and Persistent Neighborhood Sorting." *Journal of Political Economy* 129, no. 5: 1508–1552. <https://doi.org/10.1086/713101>.
- IPUMS-USA. 2019. University of Minnesota. [www.ipums.org](http://www.ipums.org).
- Jo, H.-K., J.-Y. Kim, and H.-M. Park. 2019. "Carbon Reduction and Planning Strategies for Urban Parks in Seoul." *Urban Forestry and Urban Greening* 41: 48–54. <https://doi.org/10.1016/j.ufug.2019.03.009>.
- Kling, J. R., J. B. Liebman, and L. F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75, no. 1: 83–119. <https://doi.org/10.1111/j.1468-0262.2007.00733.x>.
- Ludwig, J., G. J. Duncan, L. A. Gennetian, et al. 2013. "Long-Term Neighborhood Effects on Low-Income Families: Evidence From Moving to Opportunity." *American Economic Review* 103, no. 3: 226–231. <https://doi.org/10.1257/aer.103.3.226>.
- Mathes, S. M. 2025. "The Dynamics of Residential Sorting and Health: Implications of Climate Change in the United States." *International Economic Review* 66, no. 2: 711–756. <https://doi.org/10.1111/iere.12739>.
- McFadden, D. 1973. Conditional Logit Analysis of Qualitative Choice Behavior.
- Nevo, A. 2000. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics and Management Strategy* 9, no. 4: 513–548. <https://doi.org/10.1162/105864000567954>.
- O'Mahony, D. 2019. "10-Year Life Expectancy Gap Between London Streets Revealed by Postcode Inequality Study." Published on March 25, 2019. <https://www.standard.co.uk/news/london/10year-life-expectancy-gap-between-london-streets-revealed-by-postcode-inequality-study-a4099701.html>.
- Ou, S. 2019. "Are Some Neighborhoods Bad for Your Waistline? A Test of Neighborhood Exposure Effects on BMI." *Journal of Health Economics* 63: 52–63. <https://doi.org/10.1016/j.jhealeco.2018.10.007>.
- Percoco, M. 2021. "Spatial Health Inequality and Regional Disparities Historical Evidence From Malaria in Italy." *Region* 8, no. 1: 53–73. <https://doi.org/10.18335/region.v8i1.325>.
- Pew. 2016. "Household Expenditures and Income." Published on March 30, 2016. <https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2016/03/household-expenditures-and-income::text=However>.
- Plantinga, A. J., and S. Bernell. 2007. "Can Urban Planning Reduce Obesity? The Role of Self-Selection in Explaining the Link Between Weight and Urban Sprawl." *Review of Agricultural Economics* 29, no. 3: 557–563. <https://doi.org/10.1111/j.1467-9353.2007.00370.x>.
- Poudyal, N. C., D. G. Hodges, and C. D. Merrett. 2009. "A Hedonic Analysis of the Demand for and Benefits of Urban Recreation Parks." *Land Use Policy* 26, no. 4: 975–983. <https://doi.org/10.1016/j.landusepol.2008.11.008>.
- Reiss, D. 2022. NYC: *Rents, Guidelines and Trends*. NYC Rent Guidelines Board.
- Stead, M., K. Angus, T. Langle, et al. 2019. *Mass Media to Communicate Public Health Messages in Six Health Topic Areas: A Systematic Review and Other Reviews of the Evidence*. Public Health Research.
- Szekely, B. 2018. "Downtown LA's 90014 Heads the List of Fastest-Gentrifying Zips Since the Turn of the Millennium." Published on February 26, 2018. <https://www.rentcafe.com/blog/rental-market/real-estate-news/top-20-gentrified-zip-codes/>.
- Wen, H., Y. Zhang, and L. Zhang. 2015. "Assessing Amenity Effects of Urban Landscapes on Housing Price in Hangzhou, China." *Urban Forestry and Urban Greening* 14, no. 4: 1017–1026. <https://doi.org/10.1016/j.ufug.2015.09.013>.
- Wood, L., P. Hooper, S. Foster, and F. Bull. 2017. "Public Green Spaces and Positive Mental Health—Investigating the Relationship Between Access, Quantity and Types of Parks and Mental Wellbeing." *Health & Place* 48: 63–71. <https://doi.org/10.1016/j.healthplace.2017.09.002>.
- Zhao, Z., and R. Kaestner. 2010. "Effects of Urban Sprawl on Obesity." *Journal of Health Economics* 29, no. 6: 779–787. <https://doi.org/10.1016/j.jhealeco.2010.07.006>.



## Appendix B: Calculations for Income and House Prices

The individual-level data from *CHS* has information on five income groups: from less than 100% of the Federal Poverty Line (FPL) to more than 600% of the FPL. The 2010 poverty line guidelines of the Department of Health and Human Services give an FPL of USD 14,500 for two-person households (the mean household size in NYC is 2.4). I consider the threshold for two-person households because the assumption that housing is paid by two household members is less strong than the assumption that it is paid by one of them. With household income data from ACS 5-year estimates for 2009–2013 (downloaded from IPUMS-USA 2019), I calculate the median income for each income category in the survey.

The table below shows the approximations I used to estimate a yearly income in USD from the income categories.

| Category | Definition        | Range         | Approximate income |
|----------|-------------------|---------------|--------------------|
| 1        | < 100% FPL        | min–14,499    | 9430               |
| 2        | 100% – < 200% FPL | 14,500–28,999 | 21,380             |
| 3        | 200% – < 400% FPL | 29,000–57,999 | 42,161             |
| 4        | 400% – < 600% FPL | 58,000–86,999 | 71,037             |
| 5        | > 600% FPL        | 87,000–max    | 136,000            |

In the health survey data, 22% of individuals declare a household income below 100% FPL (using survey weights), while using ACS data, only 12% should belong to this category. This is probably due to underestimation of income in the health survey but I attain to it in the analysis.

To be able to compare housing expenditures with yearly income, I calculate annual rents with the price-to-rent ratio for each borough in NYC (from [smartasset.com](https://www.smartasset.com)—data from 2015 using the Wayback Machine from the internet Archive): Bronx 28, Brooklyn 43, Manhattan 50, Queens 30 and Staten Island 36. Considering that income is capped at USD 136,000, I also cap rents to ensure a reasonable disposable income after rent at USD 100,000 (affecting 1% of estimated rents).

To assign the price that each individual paid for housing, I calculated five quantiles of rents and assigned quantile 10 to individuals in the first income category, quantile 30 to individuals in the second income category, quantile 50 for the third income category, quantile 70 for the fourth and quantile 90 for the last income category.

This assignment is more realistic than assigning the median rent to all individuals in one neighborhood, and it ensures that individuals have positive disposable income. The fact that individuals with lower income live in houses with lower rents is intuitive but is also confirmed in the data (Pew 2016). Furthermore, this assignment implies income sorting only within neighborhoods and not at the neighborhood level, so it does not affect the estimation of the model.

## Appendix C: Methodological Aspects

### C.1 Marginal Willingness to Pay

The Marginal Willingness to Pay (MWTP) for characteristic  $x$  is the marginal rate of substitution between  $x$  and income. From the indirect utility function in Equation (1), I obtain:

$$MWTP_i(x) = \frac{\partial V / \partial x}{\partial V / \partial I} = \frac{\beta_i^x}{\beta_i^I} (I_i - P_j)$$

The total effect on utility from a given characteristic is the sum between the  $\beta_0$  presented in Table A3 and the  $\beta_1$  presented in Table 5 (and in Table A2) interacted with individuals' characteristics. These effects, therefore, are individual-specific.

## C.2 Assignment of Individuals to Neighborhoods

I use the model to assign each individual to a neighborhood according to the estimated probabilities. To do this assignment, I first calculate the cumulative probability that each individual lives in each (ordered) neighborhood. This cumulative probability goes from 0 to 1. Then, I draw random variables from a Uniform distribution in [0, 1] and assign the individual to the corresponding neighborhood in the cumulative probability. For example, suppose an individual can choose between 4 neighborhoods, and the following are the probabilities given by the model:

$$[0.1 \quad 0.5 \quad 0.2 \quad 0.2]$$

Then, the cumulative probability vector would be

$$[0.1 \quad 0.6 \quad 0.8 \quad 1]$$

A variable drawn from a Uniform [0, 1] would have exactly a 10% probability of falling in the interval [0; 0.1], implying that the chosen neighborhood is the first one, 50% probability of falling in the interval [0.1; 0.6] so that the chosen neighborhood is the second one and 20% probability of falling in the intervals [0.6; 0.8] or [0.8; 1], implying the choice of the third or the fourth neighborhood respectively.

With this assignment I calculate the population distribution (in terms of health preferences and poverty) and the new prices that allow for an equilibrium of this new distribution. Then, to avoid results to be driven by the randomness of the draw, I repeat this procedure 1000 times and take the average.

Notice that assigning individuals to neighborhoods based only on the highest probability would lead to some attractive neighborhoods having too many people and some empty unattractive neighborhoods.

## C.3 Calculation of New Equilibrium Prices

The steps to calculate the new equilibrium after a change in individuals' utility, as those resulting from an increase in health preferences are:

1. Take the sociodemographic composition as given
2. Calculate the prices that clear the market with the following contraction mapping:

$$\hat{P}_j^{t+1} = \hat{P}_j^t + \ln \left( \hat{S}_j(\hat{P}_j^t) / S_j \right)$$

where  $\hat{S}_j(\hat{P}_j^t) = \sum_i \hat{p}_{ij}(\hat{P}_j^t)$ , with  $\hat{p}_{ij}$  the estimated probability that individual  $i$  chooses neighborhood  $j$  and  $t$  the iteration

3. With the new equilibrium prices and the corresponding choice probabilities, calculate the new sociodemographic composition:

$$\hat{Y}_j = \sum_i y_i \cdot \hat{p}_{ij}$$

4. Repeat steps 1 to 3 until  $\hat{Y}_j^{t+1} = \hat{Y}_j^t$

Understanding the second step of this iterative procedure is important to understand the mechanics of the model. Notice first that the observed shares must be maintained since the supply of housing is fixed in the short term. Therefore, if the observed share (supply) is higher than the share predicted by the model (demand), prices must go down. Lower prices will attract individuals until an equilibrium is reached. Analogously, if the supply of housing is lower than the demand predicted by the model, prices must increase to reach an equilibrium.

**TABLE A1** | Definition of health behavior variables.

| Variable             | Definition                                                                                                                                  |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Ever smoker          | Smoked at least 100 cigarettes                                                                                                              |
| Current smoker       | Smokes cigarettes now if ever smoker                                                                                                        |
| Unhealthy drinker    | More than one drink on average (per day) more than 5 times a month<br>Or more than 3 drinks on average (per day) if more than twice a month |
| Unsafe sex           | No condom if not with a stable partner (among sexually active in last 12 months)                                                            |
| Flu vaccine          | Flu vaccine (shot or spray) in the last 12 months                                                                                           |
| Colonoscopy          | Colonoscopy performed in the last 10 years for individuals of 45+ years                                                                     |
| > = 1 sweetened bevr | One or more soda or other sugar-sweetened beverages consumed per day                                                                        |
| < = 1 fruit/veg      | One or less servings of fruit and/or vegetables per day                                                                                     |
| > = 4 fruit/veg      | Four or more servings of fruit and/or vegetables per day                                                                                    |
| Exercise             | Physical activity (including walking as exercise) in the last month                                                                         |

Source: Codebook of the CHS and own definition of variables.

## Appendix D: Appendix Tables and Figures D

**TABLE A2** | Interaction coefficients, full set.

|                    | Poor                | Educated            | Female              | Parent              | Married             | age2544             | age4564             | age65 m             | White               | Black/hisp          | Insured             | Healthy             |
|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Distance CBD       | −0.0044<br>(0.0019) | −0.008<br>(0.0023)  | 0.0009<br>(0.0005)  | −0.0029<br>(0.002)  | 0.0029<br>(0.0014)  | 0.0017<br>(0.0022)  | 0.0012<br>(0.0017)  | −0.0012<br>(0.0018) | 0.0022<br>(0.0018)  | −0.0001<br>(0.002)  | 0.0052<br>(0.002)   | −0.0014<br>(0.0012) |
| Distance park      | −0.0365<br>(0.0257) | −0.0328<br>(0.03)   | 0.0017<br>(0.0055)  | −0.0206<br>(0.0234) | 0.0426<br>(0.0188)  | 0.0175<br>(0.0281)  | 0.0082<br>(0.0256)  | 0.0193<br>(0.0248)  | −0.034<br>(0.0231)  | −0.0609<br>(0.0261) | −0.016<br>(0.0264)  | −0.0436<br>(0.0191) |
| Detached house     | −0.194<br>(0.1157)  | 0.2984<br>(0.1541)  | −0.0604<br>(0.0299) | −0.0741<br>(0.1101) | 0.1584<br>(0.0792)  | −0.0358<br>(0.1388) | −0.0148<br>(0.1177) | 0.0626<br>(0.1155)  | −0.2269<br>(0.1342) | −0.0291<br>(0.1294) | −0.0329<br>(0.1014) | 0.1081<br>(0.0875)  |
| Number floors      | −0.0674<br>(0.0199) | 0.087<br>(0.0218)   | 0.0098<br>(0.0038)  | −0.0173<br>(0.0183) | −0.0285<br>(0.0151) | 0.0058<br>(0.0198)  | −0.0064<br>(0.0163) | −0.0061<br>(0.0169) | −0.0083<br>(0.0127) | 0.0204<br>(0.0183)  | 0.063<br>(0.0204)   | 0.033<br>(0.0111)   |
| Building age       | 0.0003<br>(0.0013)  | 0.0043<br>(0.0018)  | −0.0003<br>(0.0003) | −0.0013<br>(0.0016) | 0.0002<br>(0.001)   | 0.0029<br>(0.0017)  | 0.0008<br>(0.0015)  | 0.0000<br>(0.0015)  | −0.0016<br>(0.0016) | −0.0014<br>(0.0019) | 0.0006<br>(0.0016)  | −0.0013<br>(0.001)  |
| House size         | −0.0034<br>(0.0013) | 0.0022<br>(0.0015)  | 0.001<br>(0.0003)   | 0.0003<br>(0.0012)  | −0.0007<br>(0.0009) | 0.0001<br>(0.0015)  | 0.0009<br>(0.0013)  | 0.0008<br>(0.0013)  | −0.0012<br>(0.0012) | −0.0004<br>(0.0012) | 0.0014<br>(0.0012)  | 0.0018<br>(0.0008)  |
| Comm area          | −0.0074<br>(0.0102) | −0.0121<br>(0.0087) | −0.0078<br>(0.0029) | 0.0113<br>(0.0097)  | −0.0162<br>(0.0081) | 0.0222<br>(0.0119)  | 0.0181<br>(0.0076)  | 0.0084<br>(0.0073)  | −0.0103<br>(0.0084) | −0.0094<br>(0.0098) | 0.0027<br>(0.0074)  | −0.0072<br>(0.0049) |
| Indus area         | −0.0192<br>(0.1197) | −0.1207<br>(0.1924) | 0.0248<br>(0.0256)  | −0.1546<br>(0.1405) | 0.2778<br>(0.1508)  | 0.0222<br>(0.1703)  | −0.2349<br>(0.1592) | −0.2663<br>(0.1336) | 0.1429<br>(0.0805)  | 0.1407<br>(0.0803)  | 0.1231<br>(0.0811)  | −0.0032<br>(0.0698) |
| Crime rate         | 0.0001<br>(0.0009)  | −0.0019<br>(0.0014) | 0.0000<br>(0.0003)  | −0.0025<br>(0.0012) | −0.0008<br>(0.0009) | −0.0005<br>(0.0012) | −0.0008<br>(0.0012) | −0.0002<br>(0.0011) | −0.0008<br>(0.0014) | −0.0002<br>(0.0016) | 0.0003<br>(0.0015)  | −0.0007<br>(0.0008) |
| Percentage white   | −0.1371<br>(0.1081) | 0.5614<br>(0.1584)  | −0.0431<br>(0.0398) | 0.1614<br>(0.1317)  | −0.1195<br>(0.109)  | −0.2725<br>(0.1602) | −0.3419<br>(0.1465) | −0.349<br>(0.1307)  | 3.1307<br>(0.1168)  | 1.9908<br>(0.1669)  | 0.3842<br>(0.0915)  | −0.1928<br>(0.0844) |
| Percentage bl/hisp | −0.2925<br>(0.1058) | 0.3559<br>(0.1465)  | −0.0842<br>(0.0391) | 0.2072<br>(0.1189)  | −0.3679<br>(0.0996) | −0.0157<br>(0.144)  | 0.0225<br>(0.1368)  | −0.1<br>(0.1249)    | 1.7742<br>(0.1371)  | 3.514<br>(0.1661)   | 0.4063<br>(0.0974)  | −0.0592<br>(0.0727) |
| $\beta^{DI}$       | 0.7714<br>(0.2260)  |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |

Note: All interaction coefficients ( $\beta_i$ ) from the main specification of the model, plus  $\beta^{DI}$ . Standard errors in parentheses, clustered by neighborhood.

**TABLE A3** | Delta decomposition.

|                            | Coef.     | Std.error |
|----------------------------|-----------|-----------|
| Distance CBD (min)         | −0.004    | (0.00)    |
| Distance park (Km)         | 0.007     | (0.05)    |
| Detached houses            | 0.140     | (0.26)    |
| Number of floors           | −0.134*** | (0.03)    |
| Building age               | 0.005**   | (0.00)    |
| House size                 | −0.002    | (0.00)    |
| Commercial area ( $Km^2$ ) | 0.083***  | (0.02)    |
| Industrial area ( $Km^2$ ) | 0.038     | (0.25)    |
| Crime rate (per 1000 res)  | −0.011*** | (0.00)    |
| Percentage white           | −0.200    | (0.31)    |
| Percentage black/hisp      | −0.299    | (0.28)    |
| Obs.                       | 128       |           |
| $R^2$                      | 0.85      |           |

Note: Coefficients from a linear regression on the alternative specific constants on neighborhood characteristics.

\* $p < 0.10$ .

\*\* $p < 0.05$ .

\*\*\* $p < 0.01$ .

**TABLE A4** | mean MWTP for one SD change (in thousand USD).

|                       | Std. dev. | Typical | Healthy |
|-----------------------|-----------|---------|---------|
| Distance CBD          | 20.3      | −10.6   | −11.8   |
| Distance park         | 0.7       | −1.4    | −4.7    |
| Detached houses       | 0.2       | 5.9     | 8.6     |
| Number of floors      | 1.9       | 3.7     | 11.4    |
| Building age          | 14.2      | 15      | 13      |
| House size            | 17.9      | 7.1     | 11.6    |
| Commercial area       | 2.0       | 14.2    | 12.7    |
| Industrial area       | 0.1       | −0.7    | −1.4    |
| Crime rate            | 22.8      | −30     | −31.9   |
| Percentage white      | 0.3       | 46.4    | 42.3    |
| Percentage black/hisp | 0.3       | −1.2    | −2.5    |

Note: Mean Marginal Willingness to Pay for a change of one standard deviation in each neighborhood characteristic at the time for a *Typical* individual (non-poor, working age and educated) and a typical *Healthy* individual (upper half of the health preference index distribution). Calculation using survey weights.

**TABLE A5** | Robustness to income: Coefficients of interactions with “health index”, with variations for income.

|                    | $\beta_i^{DI}$      | Linear price        |
|--------------------|---------------------|---------------------|
| Distance CBD       | −0.0011<br>(0.0012) | −0.0014<br>(0.0012) |
| Distance park      | −0.0399<br>(0.0190) | −0.0439<br>(0.0191) |
| Detached house     | 0.0942<br>(0.0875)  | 0.1066<br>(0.0874)  |
| Number floors      | 0.0300<br>(0.0105)  | 0.0328<br>(0.0110)  |
| Building age       | −0.0014<br>(0.0010) | −0.0013<br>(0.0010) |
| House size         | 0.0017<br>(0.0008)  | 0.0018<br>(0.0008)  |
| Comm area          | −0.0072<br>(0.0046) | −0.0070<br>(0.0049) |
| Indus area         | −0.0187<br>(0.0674) | −0.0016<br>(0.0698) |
| Crime rate         | −0.0007<br>(0.0008) | −0.0007<br>(0.0008) |
| Percentage white   | −0.1971<br>(0.0844) | −0.1958<br>(0.0842) |
| Percentage bl/hisp | −0.0277<br>(0.0740) | −0.0613<br>(0.0727) |

Note: The first column allows for a heterogeneous response to disposable income ( $\beta_i^{DI}$  instead of  $\beta^{DI}$ ). The second column shows the resulting coefficients with a linear (instead of log-linear) specification for income. Standard errors in parentheses, with clustering at the neighborhood level.

**TABLE A6** | Robustness to income: Instrument relevance.

|                       | Annual rent (1000 USD) |            |
|-----------------------|------------------------|------------|
|                       | Coef.                  | Std. error |
| Detached houses adj.  | 3.202                  | (2.60)     |
| Building age adj.     | 0.091**                | (0.04)     |
| Number of floors adj. | 1.911***               | (0.26)     |
| House size adj.       | 0.083**                | (0.04)     |
| Constant              | −13.551**              | (5.99)     |
| Obs.                  | 127                    |            |
| F-stat                | 19.61                  |            |

Note: First stage of the instrumental variable estimation of the Delta coefficients, from the linear (instead of log-linear) specification on income and price (presented in the last column of Table A5). The second stage is presented in Table A7 and explained in Section 5.3. Explanatory variables are exogenous characteristics of adjacent neighborhoods.

\* $p < 0.10$ .

\*\* $p < 0.05$ .

\*\*\* $p < 0.01$ .

**TABLE A7** | Robustness to income: Delta coefficients with a linear specification of income.

|                            | Delta     |            |
|----------------------------|-----------|------------|
|                            | Coef.     | Std. error |
| Annual rent (1000 USD)     | −0.039    | (0.03)     |
| Distance CBD (min)         | −0.004    | (0.00)     |
| Distance park              | −0.012    | (0.02)     |
| Detached houses            | 0.169     | (0.26)     |
| Number of floors           | −0.146*** | (0.05)     |
| Building age               | 0.008***  | (0.00)     |
| House size                 | −0.001    | (0.00)     |
| Commercial area ( $Km^2$ ) | 0.093***  | (0.02)     |
| Industrial area ( $Km^2$ ) | 0.120     | (0.30)     |
| Crime rate (per 1000 res)  | −0.011*** | (0.00)     |
| Percentage white           | −0.182    | (0.30)     |
| Percentage black/hisp      | −0.265    | (0.32)     |
| Obs.                       | 127       |            |

Note: Delta coefficients of the linear (instead of log-linear) specification on income and price (presented in the last column of Table A5). This specification requires instrumentation of prices at the second stage of the estimation, as explained in Section 5.3.

\* $p < 0.10$ .

\*\* $p < 0.05$ .

\*\*\* $p < 0.01$ .

**TABLE A8** | Robustness to variable specification: Coefficients of interactions with “health index”, with alternatives for main variables.

|                    | Hlt idx1            | Hlt idx3            | Park area           | Log dPark           | Crime weight        |
|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Distance CBD       | −0.0012<br>(0.0012) | 0.0020<br>(0.0014)  | −0.0011<br>(0.0013) | −0.0014<br>(0.0012) | −0.0019<br>(0.0012) |
| Distance park      | −0.0341<br>(0.0207) | −0.0128<br>(0.0224) |                     | −0.0390<br>(0.0181) | −0.0270<br>(0.0076) |
| Park area          |                     |                     | 0.0144<br>(0.0129)  |                     |                     |
| Detached house     | 0.1304<br>(0.0874)  | −0.0397<br>(0.1031) | 0.0922<br>(0.0926)  | 0.1036<br>(0.0880)  | 0.0765<br>(0.0833)  |
| Number floors      | 0.0368<br>(0.0106)  | 0.0152<br>(0.0123)  | 0.0365<br>(0.0117)  | 0.0326<br>(0.0112)  | 0.0295<br>(0.0107)  |
| Building age       | 0.0003<br>(0.0011)  | 0.0011<br>(0.0013)  | −0.0006<br>(0.0011) | −0.0011<br>(0.0010) | −0.0017<br>(0.0011) |
| House size         | 0.0008<br>(0.0008)  | 0.0012<br>(0.0010)  | 0.0017<br>(0.0008)  | 0.0019<br>(0.0008)  | 0.0019<br>(0.0008)  |
| Comm area          | −0.0042<br>(0.0046) | 0.0008<br>(0.0056)  | −0.0103<br>(0.0055) | −0.0070<br>(0.0051) | −0.0076<br>(0.0051) |
| Indus area         | 0.0038<br>(0.0811)  | 0.0328<br>(0.0739)  | −0.0519<br>(0.0643) | −0.0315<br>(0.0650) | 0.0130<br>(0.0632)  |
| Crime rate         | −0.0013<br>(0.0008) | −0.0004<br>(0.0009) | −0.0000<br>(0.0009) | −0.0006<br>(0.0009) | −0.0008<br>(0.0009) |
| Percentage white   | −0.1053<br>(0.0790) | −0.2537<br>(0.0889) | −0.1822<br>(0.0928) | −0.1794<br>(0.0869) | −0.1949<br>(0.0844) |
| Percentage bl/hisp | −0.0774<br>(0.0764) | −0.1041<br>(0.0869) | −0.0576<br>(0.0774) | −0.0547<br>(0.0741) | −0.0415<br>(0.0739) |

*Note:* Estimation coefficients by replacing the main variables of the analysis with an alternative. In column 1, the more exogenous health preference index is replaced with another health index that includes all behaviors, while in column 2, it is replaced by an index containing only the variables *ever smoker* and *unsafe sex*. In column 3, distance to the park is replaced with park area by neighborhood, while in column 4, it is replaced with its logarithm. In column 5, distance to the park is weighted by a measure of criminality in parks. Standard errors in parenthesis, with clustering at the neighborhood level.

**TABLE A9** | Robustness to the sample: Coefficients of interactions with “health index”, excluding one borough at the time.

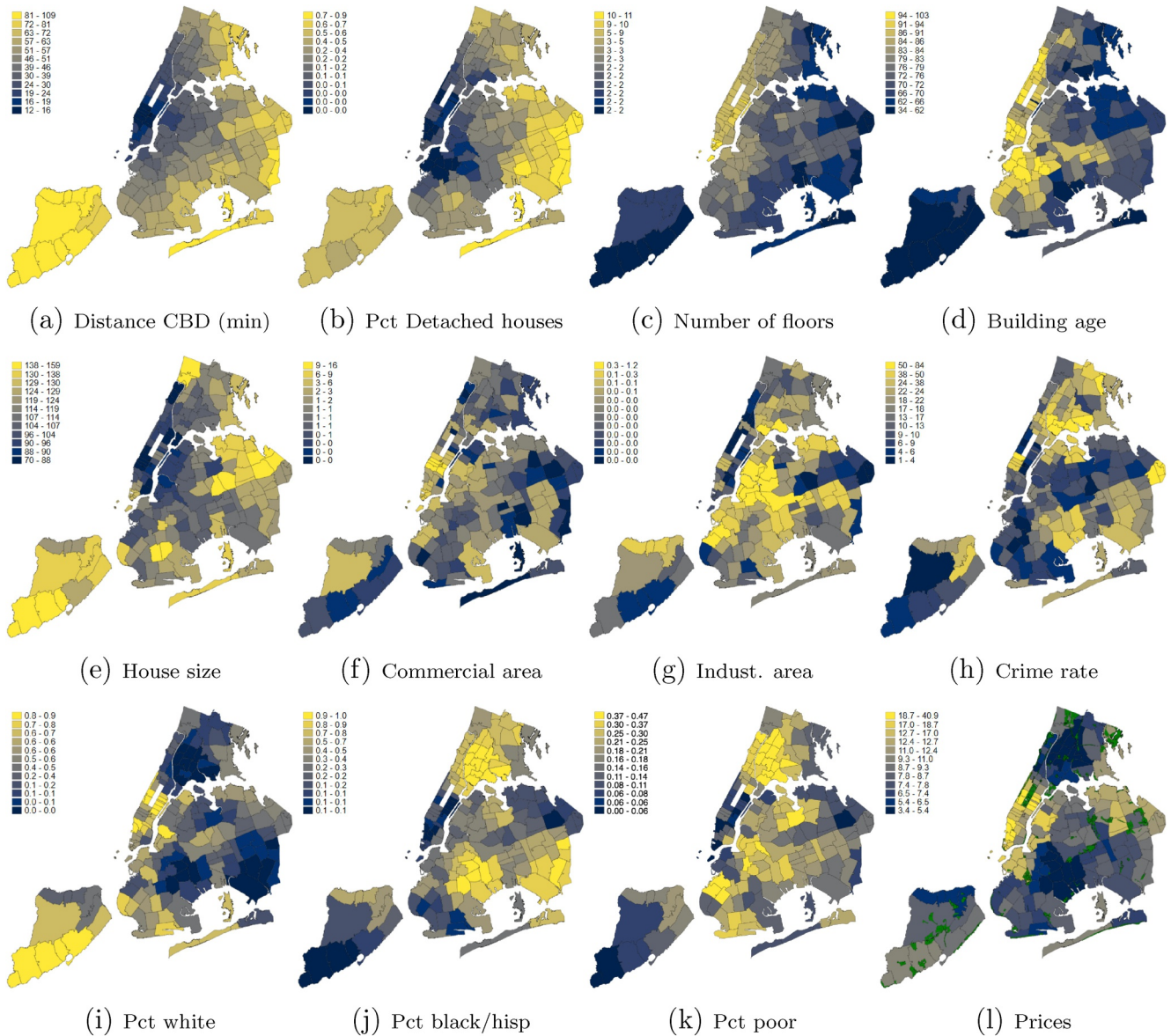
|                | ¬ Bronx             | ¬ Brooklyn          | ¬ Manhattan         | ¬ Queens            | ¬ st Island         |
|----------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Distance CBD   | −0.0012<br>(0.0014) | −0.0034<br>(0.0013) | −0.0020<br>(0.0015) | −0.0024<br>(0.0017) | −0.0010<br>(0.0015) |
| Distance park  | −0.0413<br>(0.0211) | −0.0776<br>(0.0176) | −0.0431<br>(0.0221) | −0.0136<br>(0.0274) | −0.0470<br>(0.0201) |
| Detached house | 0.1155<br>(0.0940)  | 0.2756<br>(0.0993)  | 0.0597<br>(0.0890)  | 0.1091<br>(0.1392)  | 0.1049<br>(0.0877)  |
| Number floors  | 0.0343<br>(0.0113)  | 0.0350<br>(0.0110)  | −0.0419<br>(0.0510) | 0.0219<br>(0.0115)  | 0.0362<br>(0.0123)  |
| Building age   | −0.0014<br>(0.0012) | −0.0028<br>(0.0011) | −0.0008<br>(0.0012) | −0.0012<br>(0.0012) | −0.0015<br>(0.0012) |
| House size     | 0.0016<br>(0.0009)  | 0.0009<br>(0.0008)  | 0.0030<br>(0.0011)  | 0.0017<br>(0.0009)  | 0.0020<br>(0.0008)  |
| Comm area      | −0.0065<br>(0.0051) | −0.0084<br>(0.0045) | 0.0104<br>(0.0120)  | −0.0106<br>(0.0053) | −0.0079<br>(0.0053) |
| Indus area     | −0.0124<br>(0.0762) | 0.0836<br>(0.0600)  | −0.0407<br>(0.0817) | −0.0696<br>(0.1139) | 0.0115<br>(0.0722)  |

(Continues)

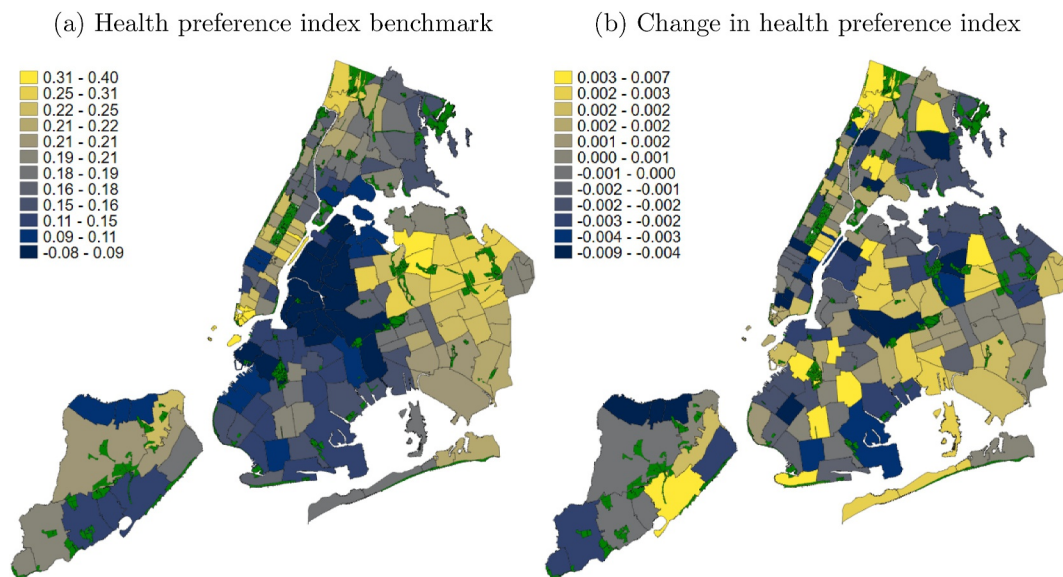
**TABLE A9** | (Continued)

|             | ↖ Bronx             | ↖ Brooklyn          | ↖ Manhattan         | ↖ Queens            | ↖ st Island         |
|-------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Crime rate  | -0.0013<br>(0.0012) | -0.0011<br>(0.0008) | -0.0002<br>(0.0010) | 0.0001<br>(0.0010)  | -0.0004<br>(0.0008) |
| Pct white   | -0.1718<br>(0.0881) | -0.1684<br>(0.0941) | -0.2176<br>(0.0951) | -0.2902<br>(0.1505) | -0.1759<br>(0.0913) |
| Pct bl/hisp | -0.0344<br>(0.0815) | -0.0020<br>(0.0689) | -0.0449<br>(0.0833) | -0.2231<br>(0.1531) | -0.0623<br>(0.0732) |

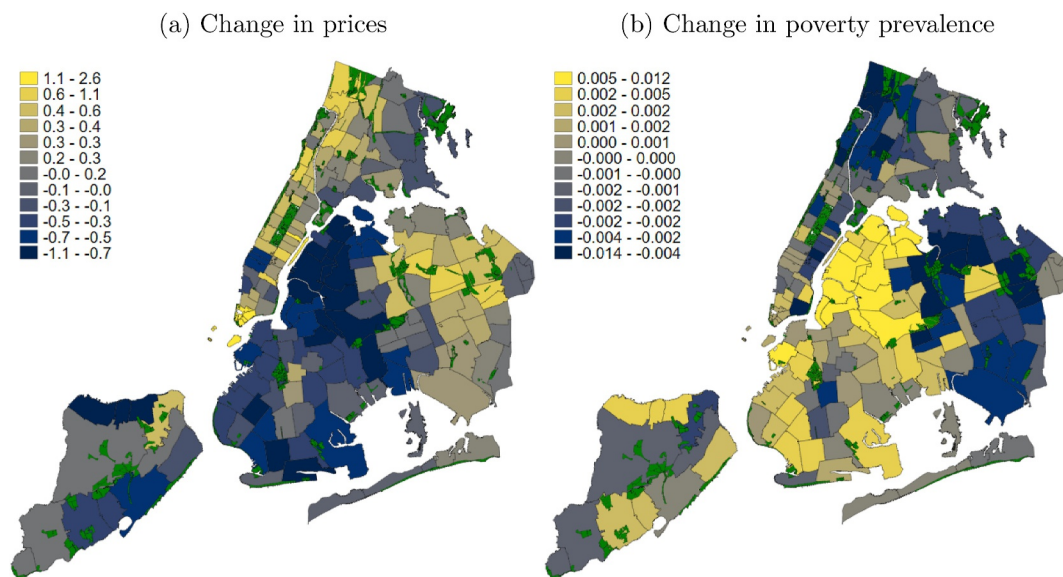
*Note:* The table shows  $\beta_j$  coefficients of the interaction with “health index” excluding the column-borough from the estimation. Standard errors in parentheses, with clustering at the neighborhood level.



**FIGURE A1** | Neighborhood characteristics. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE A2** | Counterfactual - geographical distribution of Health preferences. The left plot shows the distribution of health preferences by neighborhood predicted by the model, used as the benchmark. This is calculated by averaging the health preference index of all individuals assigned to each neighborhood according to the model parameters. The right plot shows the difference between the average health preference index by neighborhood after relocation due to the counterfactual evaluation (before the improvement of 0.3 standard deviations) and the benchmark distribution of health preferences. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE A3** | Counterfactual - Geographical distribution of Prices and Poverty prevalence. The left plot shows the difference between equilibrium prices from the counterfactual evaluation and benchmark. that is observed, prices (in thousands of USD). The right panel shows the difference between the poverty prevalence predicted in the counterfactual evaluation and benchmark poverty (predicted by the model to fit observed data). [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]