Research Paper

Measuring disparities in food access and its implications for nutrient-related diseases

an empirical study in metropolitan Atlanta

Yilun Zha, Georgia Institute of Technology, United States

Abstract

A considerable body of literature has investigated the role of food insecurity in resulting in a wide range of subsequent health. Using Yelp businesses and USDA SNAP retailer data, this study proposes a multidimensional and spatio-temporally dynamic approach to measuring food access from the perspective of Abundance, Diversity, and Healthiness. An inconsistent result has been found by comparing the space-based and place-based measurement, which suggests a non-dichotomous interpretation of urban foodscape. Furthermore, OLS and Spatial regression models are built to investigate the associations between food access and nutrient-related diseases. Results suggest that living in areas with great diversity of food choices have a strongly protective effect on prevalence rates of high blood pressure, diabetes, and coronary heart diseases, while there is no statistically significant tie between abundance of food and the occurrence of diseases. Findings emphasize the need for a deepened understanding of the components and manifestation of the concept and operationalization of food access in regional food systems planning.

Keywords

Food Insecurity; Food Access; Space-based and Place-based; Nutrient-related disease; Spatial regression

1.Introduction

1.1 Overview

The United Nations Declaration of Human Rights explicitly states that food is a basic human right (The United Nations 1948). Correspondingly, food insecurity, defined as 'the limited or uncertain availability of nutritionally adequate and safe foods or limited ability to acquire acceptable foods in socially acceptable ways', has been identified as a significant public health issues both globally and within the United States (Murrell and Jones 2020).

There is a common belief that food insecurity has explicit ties to economic conditions, which can be clearly manifested by the fact that in 2018, an estimated 821 million people in the world, more than 2/3 of which in the global north, are still undernourished (The United Nations 2020). However, what is a little bit surprising is that the issue of food insecurity has been proliferating during the last two decades in counties often considered with a relative richness of food supply, such as United States.

Unfortunately, the overwhelming and enduring confirmation of the severeness of food insecurity does not guarantee a comprehensive understanding of the prevalence of food insecurity among families living in urban areas of developed countries, let alone an effective solution to ameliorate the plague. Within the limited scope of the food insecurity research, disparities in food access are consistently identified as a potential vertex of a causal ecological model which, working with other factors both upstream and

downstream collectively, leads to unhealthy health behavior as a final output (Sadler, Gilliland and Arku 2016).

However, such convergence in causal reasoning has very limited ability to disentangle the true impact food access exerts to population's ultimate food choice and nutrition intake. Reasons, which will be discussed in detail in the next section, are multi-faceted, includes but not limited to the conflict in the definition, high cost of measurement, availability of sensitive data, disciplinary bulwark, etc.

With the overarching aim to bridge the gap between food insecurity, food access, and nutrient-related implications, this study, using publicly available food providers and potential food consumer's data, aims to (1). constructing a probabilistic model to estimate the exposure of food choices among demographic groups ; (2). evaluating the food environment and food access based on space (egocentric neighborhood) and place (enumeration neighborhood); (3). applying statistical learning techniques to validate the associations between food access and the prevalence of nutrient-related diseases.

1.2 Previous Studies

There is a great abundance of literature in food insecurity and food access, which in general, although at some risk of oversimplification, can be traced back to two intellectual groups, one of which is made up of public health experts and nutrient epidemiologists, whose primary, if not all, focus is on the surveillance of insufficient or unhealthy nutrition intake and preventive treatment of food-related diseases, and the other is planners and geographers who specifically aim to decode the spatial aspect of food access and seek for spatial solutions to diminish the disparity. This paper will conduct a highly selected rather than a exhaustive review on topics that either lays the foundation of the research methods of this study, or still be left ambiguous and needs further examination.

There are a huge number of key health outcomes that are consistently believed to be associated with high food insecurity, such as risk factors for children's health and educational outcomes (Howard 2011), adverse consequences for overall health of children (Kirkpatrick, McIntyre and Potestio 2010), maternal depression and health status (Garg et al. 2015), mental health conditions (Heflin and Ziliak 2008), subsequent weight gain (Jones and Frongillo 2007), poor sleep outcomes (Na et al., n.d.), and suicide ideation (McIntyre et al. 2013). In particular, the adverse outcome that affects most population by food insecurity as a risk factor includes diabetes and cardiovascular diseases (South, Palakshappa and Brown 2019, CA 2018).

Among factors that are believed to be associated with food insecurity, geographic access to food is one of the most important dimensions in the US since it affects the cost of food that low-income consumers face and the decisions they make about which foods to purchase (Rose 2010). the definition of 'access to healthy food' as well as the variables that predict access to healthful food vary widely among different literature. USDA identifies food deserts across United States by factors including distance, income, urbanization, and vehicle access (USDA 2019). Attentions have been put on the neighborhood food environment and how it influences the access dimension of food insecurity, and, in turn, the consumption behavior and health outcomes of low-income consumers. Evidence shows that African Americans residing in census tracts with one or more supermarkets were more likely to meet their fruit and vegetable recommendations than those that did not (Morland, Wing and Diez Roux 2002). It is also found in North Carolina that proximity of supermarkets was shown to be positively associated with diet quality among pregnant women (Laraia et al. 2004).

2.Data and Methods

2.1 study area and research scope

The study area (figure1) centers on metropolitan Atlanta in Northwest Georgia, which is made up of 10 counties: Cherokee, Clayton, Cobb, DeKalb, Douglas, Fayette, Fulton, Gwinnett, Henry, and Rockdale. According to statistics, Metro Atlanta added nearly 64,000 people in 2019, bringing the 10-county region's total population to 4.7 million (Atlanta Regional Commission 2020). It is worth noting that the geographical scope of this study is smaller than the *metropolitan statistical area* (MSA) used by the Census Bureau, the latter of which consists of 29 counties and accommodates more than 6 million population (U.S. Census Bureau 2018a). The primary reason lies in the fact that counties in the periphery of the MSA, although defined as a part of the 'metropolitan', shows more rural characteristics de facto in terms of local food environment and available data points in those areas are in extreme sparsity, which will make the prediction unreliable.

The basic analytical units used in this study is census tract. Food access will be measured at the scope of census tract, which is consistent with most available lattice data. The regression model will use census tract as the geographical unit since the morbidity data is reported by tracts.

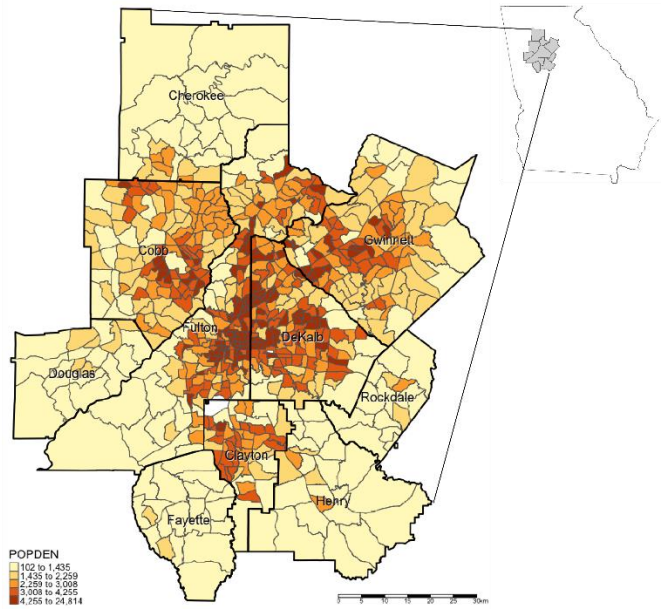

Figure 1. the study area and population density (person/sqmile) by census tracts

2.2 data source and descriptive statistics

There are primarily three data sources used in this study. Food provider's data is retrieved from a combination of Yelp Fusion API and US Department of Agriculture SNAP Retailer. The overarching assumption is that the two main sources of food exposure in the metropolitan area are from either restaurants or food retailers of all kinds. Yelp.com is an American company that operates an online platform for publishing crowd-sourced reviews about businesses, including restaurants of a wide variety of categories. Business-related data is collected and shared through Yelp Fusion API. SNAP retailer data is an open-source dataset published by USDA consisting of food retailers of 12 categories including convenience store, grocery store, supermarket, farmer's market, and so on.

Through combining the two datasets mentioned above, this study aims to exhaust food exposure within the research area as much as possible. After data cleaning and calibration, there are in total 10,444 point

food providers identified in the 10-county area: 5,635 groceries and 4,809 restaurants. Table 1 lists top 10 categories with most counts. It is not hard to see that most restaurant types listed among the top are typically not identified as 'healthy' food.

Restaurant Type	Count	Grocery Type	Count
Coffee	485	Convenience store	3066
American	440	Combination grocery	1040
Sandwiches	376	Small grocery store	526
Burgers	364	Super store	316
Chicken wings	349	Supermarket	299
Hotdogs	348	Medium grocery store	276
Breakfast brunch	321	Large grocery store	59
Mexican	307	Farmers' market	53
Bakeries	285		
Seafood	284		
TOTAL	9818		5635

Table 2. Top 10 categories of food providers with most counts

Notes: 1. Yelp and USDA-SNAP has different classification system: Categories of Yelp restaurants are based on flavor while SNAP retailers are categorized into business types; 2.the total count of restaurant types exceeds restaurant counts due to multiple labels. For instance, McDonald can be both labelled as 'burgers' and 'coffee'.

Figure 2 shows the geographical distribution of food providers of selected types in metro Atlanta, which gives an intuitive impression of the disparities in food exposure. Grocery stores tend to have a concentrated distribution in densely populated areas and transportation hubs and favor higher-status neighborhoods. In contrast, restaurants of usual (affordable, sometimes unhealthy) types spread more evenly across the metro.

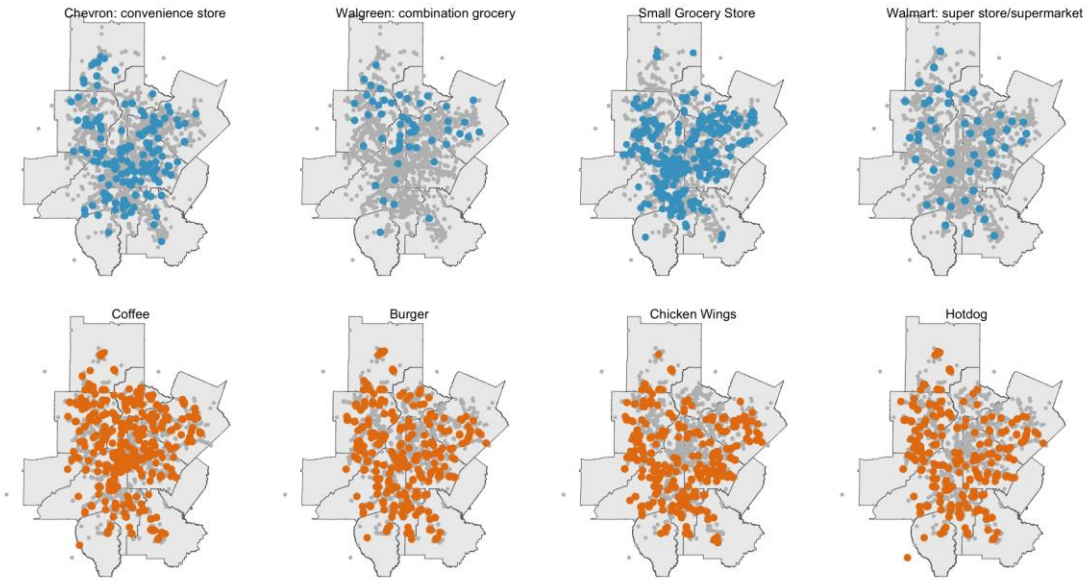

Figure 2. geographical distribution of food providers of selected types

The second data source is from the Census Bureau 2015-2019 American Community Survey 5-year estimates for socioeconomic status (SES) characteristics at the level of census tract (n = 735). Indicators include median household income (MHI), percent of car ownership (P.CAROWN), percent of population 60+ years old (P.OVERSIXTY), percent of uninsured population (P.UNINSURED), all of which either have a potential impact on food access, or serve as adjustment factors of disease prevalence.

Center of Disease Control and Prevention (CDC) initiated a project in 2016 called "500 Cities" predicting census tract-level area chronic disease risk factors, health outcomes, and clinical preventive service use for the largest 500 cities in the United States (CDC 2019). This study will use the 2019 estimates of crude prevalence of diseases that have been frequently and consistently identified as 'nutrient-related', including high blood pressure (HBP), coronary heart disease (CHD), and diabetes. Unfortunately, the prevalence data is not available for all census tracts but only for those lie in the municipal boundary of Atlanta, Alpharetta, Sandy Springs, and Johns Creek, which limits the sample size of the prediction model to n = 186.

The healthiness of each food providers is predicted by the *modified retail food environment index* (mRFEI), invented and published by the CDC's Division of Nutrition, Physical Activity and Obesity in 2011 (CDC 2011). The mRFEI is calculated at the census tract level by dividing total number of healthy food retailers by total number of food retailers. The value of mFREI is from 0 to 100. 100 indicates that all food providers in a census tract are healthy food providers. Detailed data manipulation will be introduced in the following section. Georgia reports an 8.0 average mRFEI score, compared to national average mRFEI = 10.0. The 10-county area has an 8.5, which is slightly higher than state average but still below national mean.

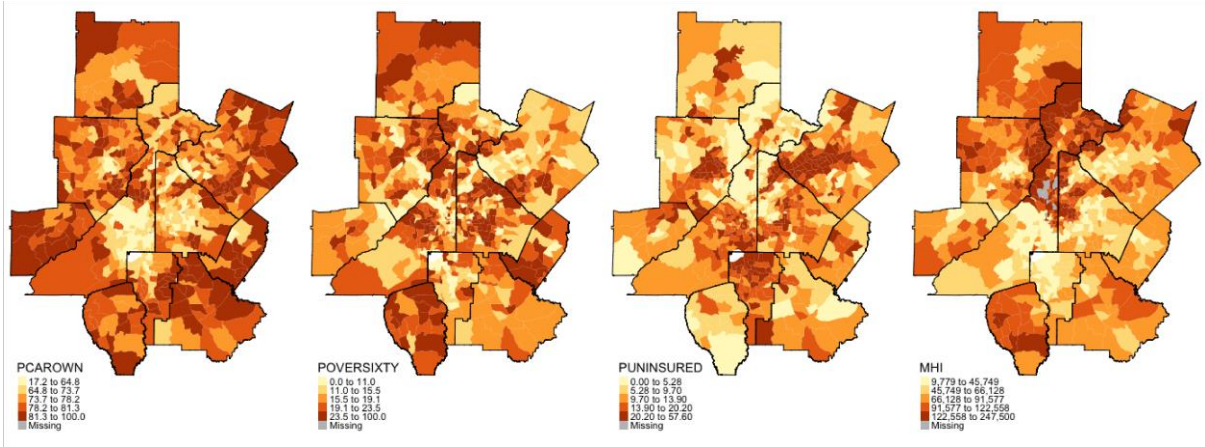

Figure 3. selected socioeconomic status characteristics by census tracts

	Unit	n	Mean	St. Dev.	Min	Q1	Q3	Max
Census: ACS 5-year est.								
MHI	USD	724	82,670	42,423	9,779	50,050	107,018	247,500
P.CAROWN	%	733	74.3	11.0	17.2	70.1	81.1	100
P.OVERSIXTY	%	734	18.1	8.0	0	12.8	22.5	100
P.UNINSURED	%	734	14.1	9.2	0	7.0	19.1	57.6
TOTAL.POP	person	735	6,043	3,117	148	4,013	7,486	25,063
POP.DEN	p/sqm	735	3,148	2,569	101.7	1,672	3,879	24,814

CDC: 500 Cities								
HBP	%	186	30.3	10.4	9.4	21.8	39.9	51.6
CHD	%	186	5.0	2.2	1.0	3.2	6.3	11.0
DIABETES	%	186	10.1	5.4	1.9	6.1	14.7	23.0
CDC: mRFEI								
mRFEI	%	735	8.5	43.8	0	5.0	12.0	51.0

Table 2. Descriptive statistics of selected variables

2.3 probabilistic model of relative food exposure

In essence, the food choice behavior of each individual is stochastic and highly unpredictable (Rodriguez, Hendrickson and Rasmussen 2018). Factors including geographic proximity, purchasing power, and cultural acceptance have either direct or indirect influences on the final decision of food choice. The most accurate approach to measuring actual food exposure is surveying, while it is tremendously time-expensive and requires a large sample size for statistical power.

As an alternative, this study proposes a discrete probabilistic model to simulate the food provider selection process of an aggregated population as a proxy of relative food exposure. Relative exposure, as a contrast to the actual one, takes a route of modelling the hypothetical incentive or disincentive to the exposure studied (Sherman et al. 2005). For instance, if the modelled probability of visiting A is 2 times higher than visiting B, we assume that relative exposure to A is 2 times more than B, although the actual situation may or may not coincide with the hypothetical share.

The fundamental assumption of this probabilistic model is that if there are in total n food providers, each denoted as Y_j, j = 1, 2, …, n, within a distance threshold d to an aggregated population X_i, the food-provider selection outcome follows a Multinomial distribution (n=1) and the probability mass function can be written as

$$\mathbb{P}(Y_j|X_i) = \prod_{k=1}^{n} p_{ik}^{\mathbf{1}(k=j)} \tag{1}$$

where the parameter p_{ik} is the probability that X_i visits Y_k. In other words, this study proposes a method to estimate the parameters of the probability distribution for each population mass. The estimation follows three main steps:

(1). For each population mass, every food provider within the catchment area will be initialized with equal probability of visiting: $p_{ik} = \frac{1}{n}$ regardless of any extraneous factors. Note that a variable catchment area will be used since for units in the outskirts of the metro area, there may not be sufficient number of food providers compared to more populous units. Figure 4 shows the concept of variable catchment area. For units located in the data-sparse region, the search radius is set to be the shortest distance within which there exists at least 10 food providers.

(2). Based on a gravity model, we apply two distance-decay functions $G(d_{ik})$ to each p_{ik} where d_{ik} is the distance between X_i and Y_k. The rationale behind different distance-decay functions is that there is a substantial difference between the distance tolerance and disincentive for population with and without

privately-owned automobiles. Hence, we define a Sigmoid friction function G_1 for automobiles and a Gaussian friction function G_2 for non-automobiles listed below (see Figure 5):

$$G_1(i,j) = \begin{cases} 1 - \frac{1}{1+e^{(5-d_{ij})}} & (d_{ij} \leq 10) \\ 0 & (d_{ij} > 10) \end{cases} \tag{2}$$

$$G_2(i,j) = \begin{cases} e^{-0.5d_{ij}} & (d_{ij} \leq 10) \\ 0 & (d_{ij} > 10) \end{cases} \tag{3}$$

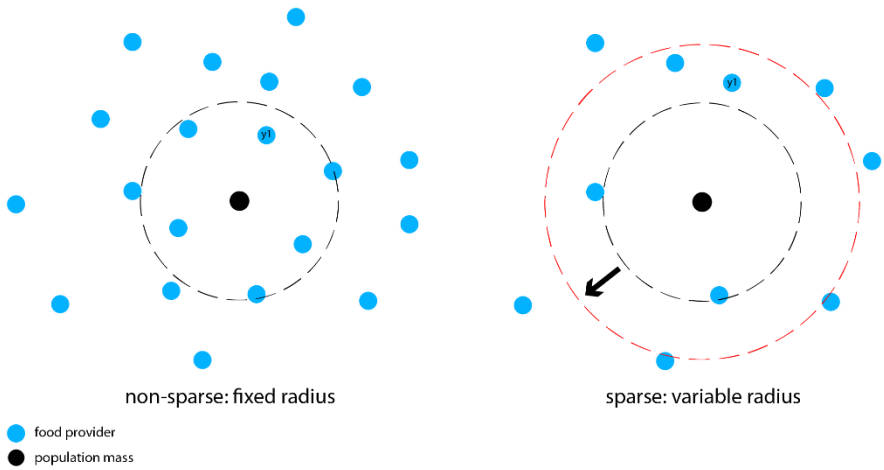

Figure 4. variable catchment area diagram

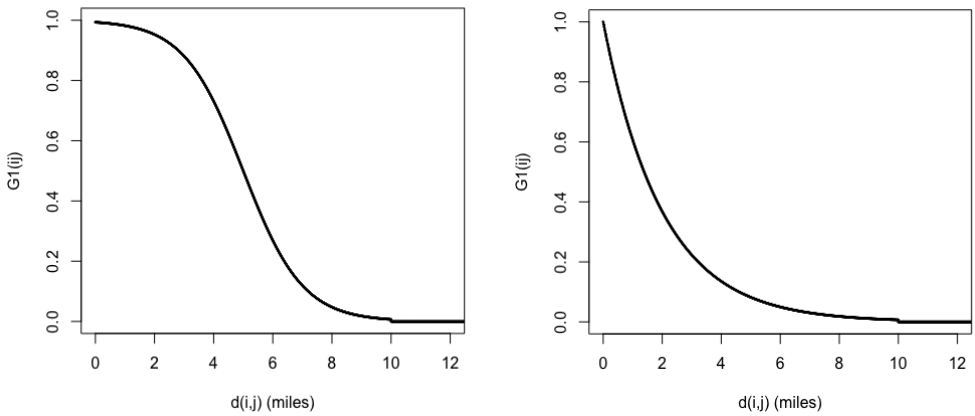

Figure 5. distance-decay functions (left: Gaussian; right: exponential)

Notes: The selection of distance-decay function is based on the rationale that short-distance travel via automobile is less sensitive than that via other methods of transportation like walk or bike. Correspondingly, the sigmoid function drops slowly in the beginning while Gaussian penalizes more on short distance.

(3). To put a penalty on the difference of purchasing power, we define a price friction function P(i,j) using the quantile of median household income $Q(MHI_i) \in [0, 1]$. $Q(MHI_i) = 1$ means the unit i has the highest MHI among all units in the metro area. The assumption is that the difference in purchasing power generates differentiated frictions to food providers of different price levels. Our data classifies the price of restaurants

into four categories: $, $$, $$$, $$$$. The high-level idea is that the price friction for affordable food providers is smaller than expensive ones.

Finally, we update the unstandardized probability $p_{ij}^0 = p_{ij} * G(i, j) * P(i, j)$. It can be further standardized by

$$p_{ij}^* = \frac{p'_{ij}}{\sum_{j=1}^{n} p'_{ij}} \tag{5}$$

as the final estimate of the adjusted probability that X_i will visit Y_j. Measurement of food access will be based on the unstandardized and standardized probability.

2.4 measurement of food access

As states in 2.3, the measurement of food access is based on the relative probability of a population mass X_i visiting food provider Y_j. It is noted from previous literature review that the definition of food access can cover a wide range of measurements. This study argues for a decomposition of food access into three main dimensions: *Abundance*, *Diversity*, and *Healthiness*. Abundance measures the total wealthiness of food choices, which is usually manifested by the concept of "food desert" (Rose 2010). Diversity is the representation of the range of available food choice types. Places provided with food choices with good quantity but poor diversity are usually referred to as "food swamp" (Amin, Badruddoza and McCluskey 2020). Finally, healthiness evaluates the attractiveness of healthy food providers as a proportion of that of total food providers.

Quantitatively, the abundance of place i, denoted as A_i, is computed by the raw sum of probability p_{ij}. Here we use unstandardized probability instead of standardized one since it is comparable between places. Diversity D_i is assessed by the Shannon Entropy of food provider types (Shannon 1997). Higher the entropy, more diversified the composition of food types is. Here we need to compute the standardized probability of each provider type k, then perform the standardization:

$$p_{ik}^* = \frac{\sum_{j \in k} p'_{ij}}{\sum_{k \in \mathcal{K}} \sum_{j \in k} p'_{ij}} \tag{7}$$

And the entropy (diversity) of each place i can be calculated by

$$D_i = \sum_{k \in \mathcal{K}} -p_{ik}^* * \log(p_{ik}^*) \tag{8}$$

Healthiness data of each provider is not available; thus, this study proposes an estimate using the CDC mRFEI index. We define healthiness-adjusted probability as

$$p_{ij}^{**} = mREFI_k * p'_{ij}, \tag{9}$$

where Y_j is in unit k.

Moreover, this study adopts a spatiotemporally dynamic approach of exposure, which takes the 'activity space' of population into consideration. Defining the context of exposure using census residential areas has been criticized from different perspectives including that of the 'local trap' (Cummins, 2007). Census

data aggregates a number of persons into a population mass based on their residential address while it is very likely that the static residential address does not represent their actual space-time patterns, the most significant one of which is commuting. Employed population spends a large share of their daytime around the environment close to where they work, so is the food exposure. Thus, this study defines place-based access, which is entirely based on a static administrative unit (census tract), and space-based access, which examines the work-live space-time pattern of a population mass and split the total amount of exposure by residential address and employment address.

Here is an example. If census tract Xi has 1000 population, 500 of which works in tract Xj. Then, the place-based access will use Xi as the sole centroid while space-based access splits total exposure into $\frac{9}{14}$ share of live space and $\frac{5}{14}$ workspace. Then, the actual access is the weighted sum of population whose activity spaces are different. To clear this up, Table 3 shows the measure matrix where x-axis is the dimensions of access, and y-axis is the dimensions of space-time pattern.

2.5 validating associations by statistical learning algorithms

This study assesses the associations between food access and nutrient-oriented diseases by spatial regression. Due to the page limits, only the high-level concepts and ideas will be discussed.

Spatial regression, spatial lag/error model in particular, is often used to deal with the spatial autocorrelation issue of lattice spatial data. The ordinary least square (OLS) model assumes that the residual is independent and normally distributed with a constant mean of zero and constant variance (Dai 2010). However, socioeconomic attributes often have a spatial spillover effect which results in spatial autocorrelation (Anaya-Izquierdo and Alexander 2020).

3. Measuring food access in metropolitan Atlanta

Using the assessment method defined in the previous chapter, this study computes the abundance, diversity, and healthiness index in three scenarios: 1. with car ownership (distance-decay function G1); 2. without car ownership (distance-decay function G2); 3. car ownership weighted by PCAROWN, by two approaches: place-based, and space-based. It is important to note that either index only makes sense comparatively while their absolute value does not have any meaning. Thus, this study uses Quantile classification with labels from "very low" to ’very high’.

3.1 place-based food access

Figure 6 presents the place-based distribution of indices, which shows apparent, while not consistent, geographic disparities in the food access across metro Atlanta. Regarding the three dimensions of food access, the Diversity index clearly shows a north-south divide. It is consistent with the geographical disparities in other socioeconomic status characteristics, such as median household income and poverty rates. In contrast, Abundance index shows higher value around urban/town centers and highway corridors, which is quite understandable in terms of a commercial entity’s location selection procedure. In addition, the inner peripheral ring of the metro in general has higher Healthiness value than inner cities. Tracts in the outer ring are among the lowest rankings in both Abundance and Healthiness, partly due to the edge effect and the sparsity of food providers.

When comparing the different scenarios regarding car ownership and corresponding distance-decay function, it can be found that all three food access indices show lower values for w/o ownership than that

for w/ ownership. To adjust for the influence of private transportation ownership, this study continues to produce the car-ownership-adjusted indices denoted as *aAi, aEi, aHi*, which gives a better estimate of the overall access for an aggregated population mass.

In order to assess the global spatial autocorrelation effect and identify local hotspot, this study uses GeoDa software to compute the Global Moran's I index and Local Spatial Autocorrelation index (LISA) (Anselin and Kho 2006, Anselin 1995). The result shows that the global Moran's I for each index are: 0.599, 0.703, and 0.618, all of which indicate moderate to strong spatial spillover effect. It also provides a justification of the use of spatial regression in substitution for OLS regression in the later chapter.

With a strong global spatial spillover feature, the geographic distribution of food access also shows statistically significant local hotspots, which can be understood as 'high-access area' and 'low-access area' at the regional level. Figure 7 shows different hot spots and cold spots identified by LISA (α = 0.05). The result reinforces previous observations that an obvious hotspot of diversity in northern metro and low-value area of Abundance and Healthiness in the outskirt. There is no real-life evidence showing that there exists a healthy-food cluster in southern Cherokee county and Clayton County, the validity of which we need to test in the prediction model.

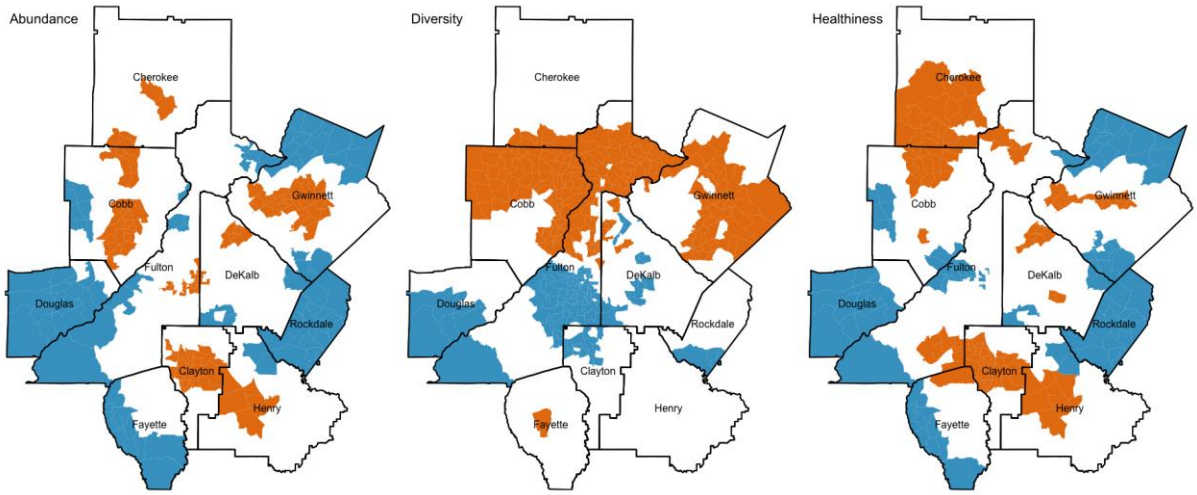

Figure 7. LISA hot spots (orange) and cold spots (blue)

4.2 space-based food access

Due to the intrinsic risk of solely using residential areas to define the context of exposure, this study proposes an alternative way to quantify food exposure considering the fact that the local scale is not the only meaningful unit of interest in health research; as a result context should not be exclusively defined as a local area (Purcell and Brown 2005). As a result, this study adopts a temporal structure of an activity space, defined by the frequency, regularity, and duration at which locations are visited (Perchoux et al. 2013).

Figure 8 shows a simple schematic activity space representation with nodes and links. A clearly defined spatio-temporal activity space is particularly relevant to quantify individual and collective spatial behavior in relation with the accessibility to all resources including food.

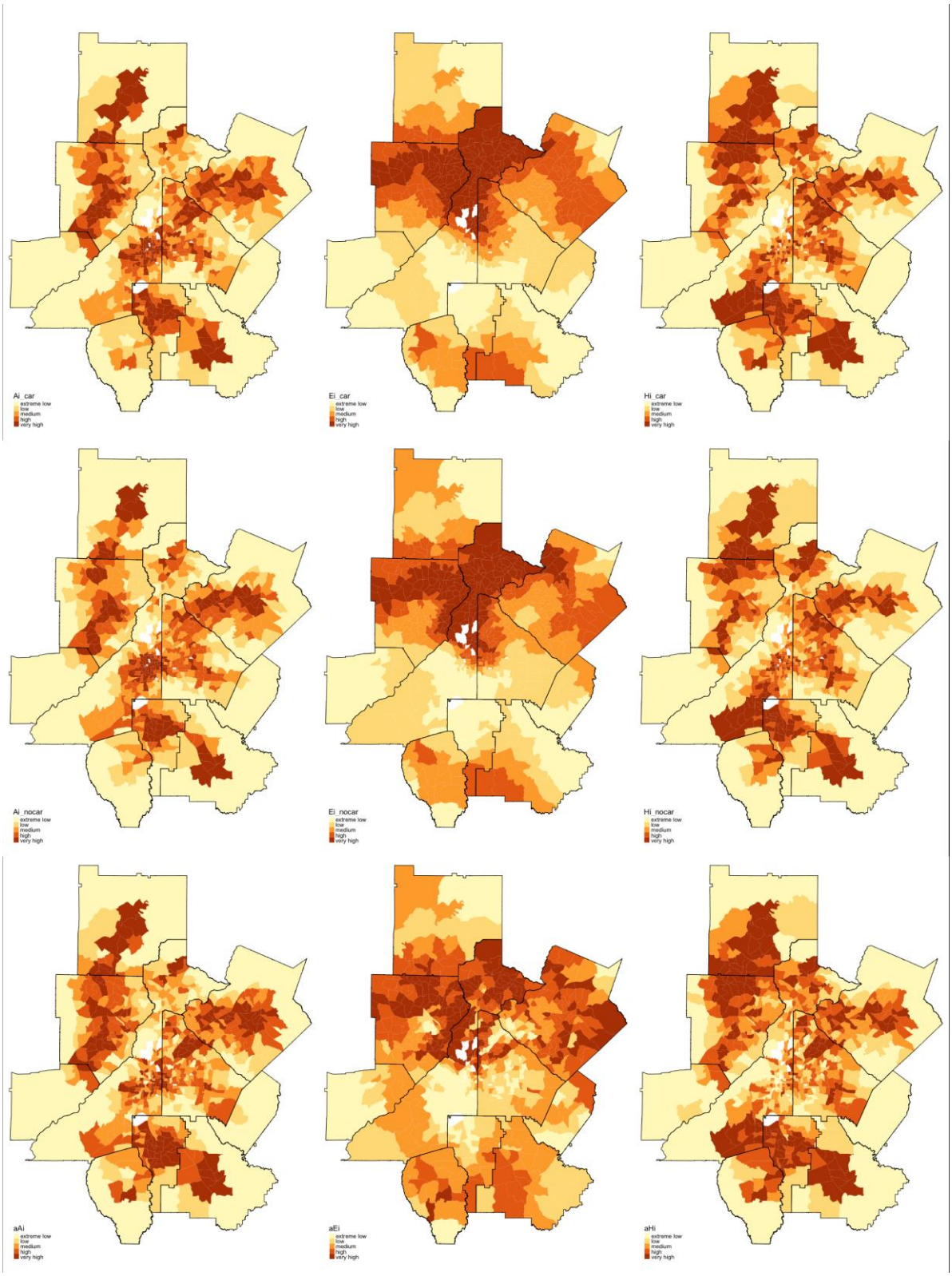

Figure 6. place-based food access index (from left to right: Abundance, Diversity, Healthiness)

4.2 space-based food access

Due to the intrinsic risk of solely using residential areas to define the context of exposure, this study proposes an alternative way to quantify food exposure considering the fact that the local scale is not the only meaningful unit of interest in health research; as a result, context should not be exclusively defined as a local area (Purcell and Brown 2005). As a result, this study adopts a temporal structure of an activity space, defined by the frequency, regularity, and duration at which locations are visited (Perchoux et al. 2013).

Figure 8 shows a simple schematic activity space representation with nodes and links. A clearly defined spatio-temporal activity space is particularly relevant to quantify individual and collective spatial behavior in relation with the accessibility to all resources including food.

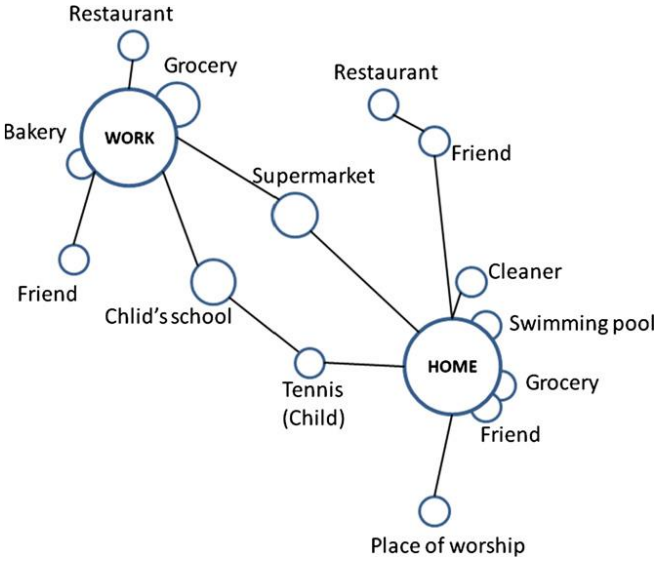

Figure 8. Schematic activity space representation with nodes and links

The simplest space-time path is work-home. To quantify this structure, this study uses LEHD Origin-Destination Employment Statistics (LODES) dataset published by Center for Economic Studies, Census Bureau to determine the destinations of workspace and the relative shares (U.S. Census Bureau 2018b). There are in total 405,105 tract-to-tract job flows within the 10-county metro area and in total 1,790,588 employment counts that have different residential address and work address, which shows that more than one third of the total population do not work in the census tract where they live. Figure 9 displays the direction and magnitude of job flow in metro Atlanta. There are a large amount of long-distance commuting flows from southern metro Atlanta where many disadvantaged neighborhoods locates, to the northern region, which is often considered as the job center.

Based on previous computation, space-based access can be obtained from adjusting for the multiplicity of exposure of a census-tract population mass. If the total population of census tract i is N with place-based index value I_i, and p_j is the share of population that work in tract j, the space-based access index is

$$I' = aI_i + (1 - a) * \sum_j p_j I_j \tag{12}$$

where a is the share of time spent in residential place. In this study, a is set to be 9/14 based on the assumption is that there are in total 14 times food choices (lunch + dinner x 7days) for one week and approximately 9 of them are made in the residential place.

Figure 10 shows the outcome of space-based food access index and LISA hot spots and cold spots. An interesting finding is that the global Moran's I drops significantly to 0.29, 0.25, and 0.21 correspondingly, which indicates that the spatial spillover effect is weaker than that under the place-based measure since in the space-based measure population are not tied to one single geographic unit any more while scatted across the metro area according to their commuting pattern. The outcome of LISA also differs in the sense that some southern tracts which are identified as hotspots before show no significance due to its high and volatile population fluidity induced by commuting.

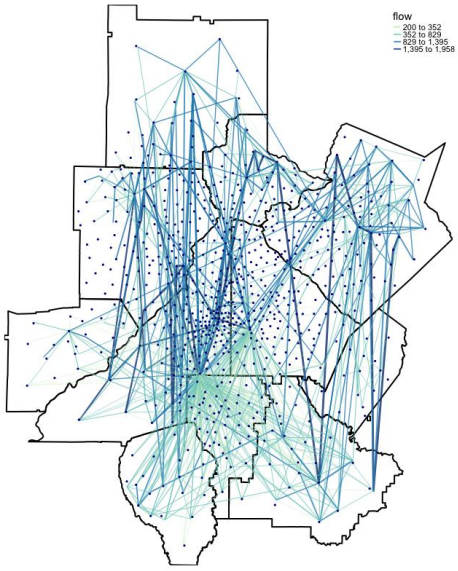

Figure 9. tract-to-tract job count flow chart (only showing count > 200)

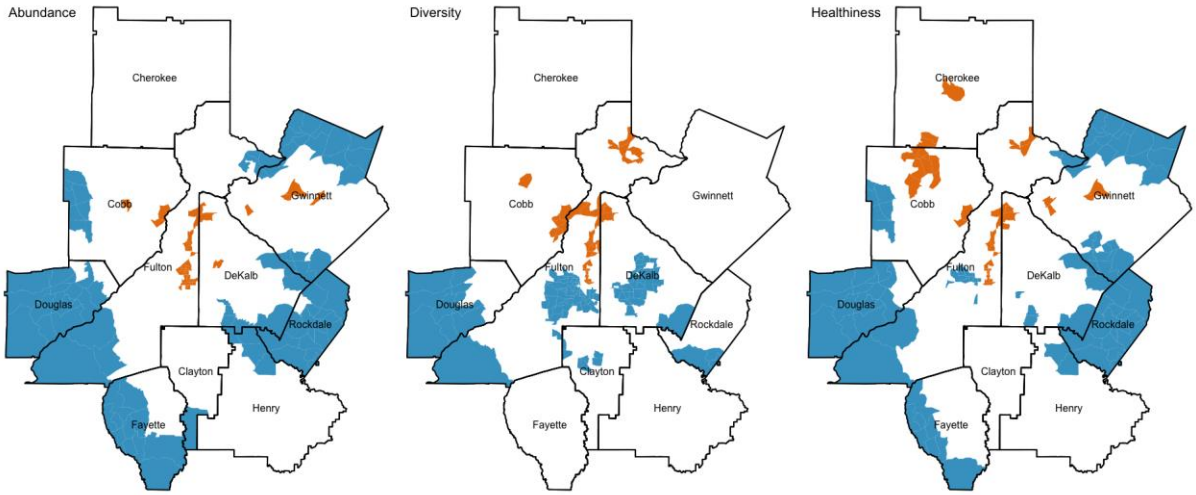

Figure 10. space-based food access LISA hot spots and cold spots

3.3 a multidimensional assessment of food access

After taking a full exploration of two different approaches of measurements and three dimensions of food access separately, it is natural to ask the question: what on earth does the whole picture of the food environment in metro Atlanta look like, and where are places that have good/bad food access?

There is no easy answer to either one of the questions. Clearly, disparities in food access does exist in metro Atlanta and take different forms by different dimensions. However, it is critical to note that disadvantaged residents in so-called food deserts did not always experience problems with food procurement PURELY because of where they live (Williams and Hubbard 2001). Recent literature has cautioned that existing studies may overstate the issue or even fabricate food deserts to support the rationale for subsequent studies (Cummins and Macintyre 2002). There are two intrinsic flaws embedded in the 'local trap': (1). It fails to recognize other factors that affect the food-selection behavior while belittles this issue to a physical determinism. (2). It fails to distinguish a poor access defined to a geographic unit from that to the population who live there.

Attempting to make a response to both flaws, this study argues that it is risky to produce a single picture of the urban foodscape, considering the extreme complexity it bears. As a result, it is also not reliable to make a dichotomous judgement of whether an area has good or bad access, especially when individual-level data is not available. Instead, this study substitutes existing terminology used to describe urban foodscape, such as 'food desert', 'food swamp', by a location in the three-dimensional A-D-H space. Figure 11 defines three possible locations in the value space where there can be much more positions according to the subjective definition and choices of the cutoff value: the sweet spot, where all dimensions are high; the sour spot with all low values, and the dilemma spot with high abundance but low diversity and healthiness.

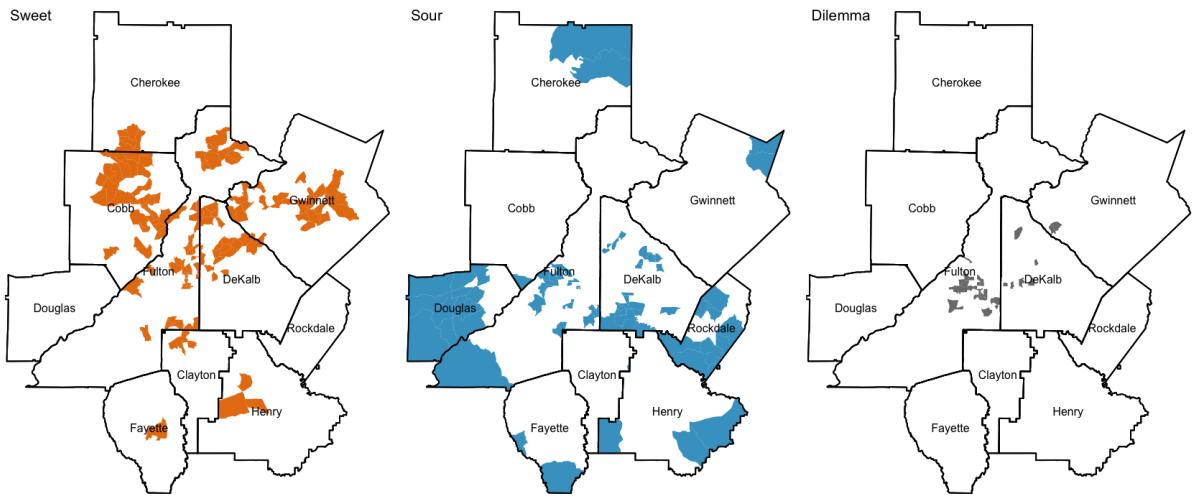

FIGURE 11. sweet spot, sour spot, and dilemma spot

5. Assessing the associations between food access and nutrient-related diseases

To test the validity and the predictability of food access indices proposed before, this study uses OLS regression and spatial regression to construct the prediction model. The predictive variables are census-tract level crude prevalence of nutrient-related diseases: High Blood Pressure (HBP), Diabetes, and coronary heart disease (CHD). Four control SES variables are included in the model to adjust for other relevant factors. Figure 12 shows the correlation coefficients of all explanatory variables. To avoid multicollinearity, PCAROWN is removed from the list since it has a correlation coefficient of 0.935 to D-index, plus the fact that the measurement has included the information of car ownership in each index.

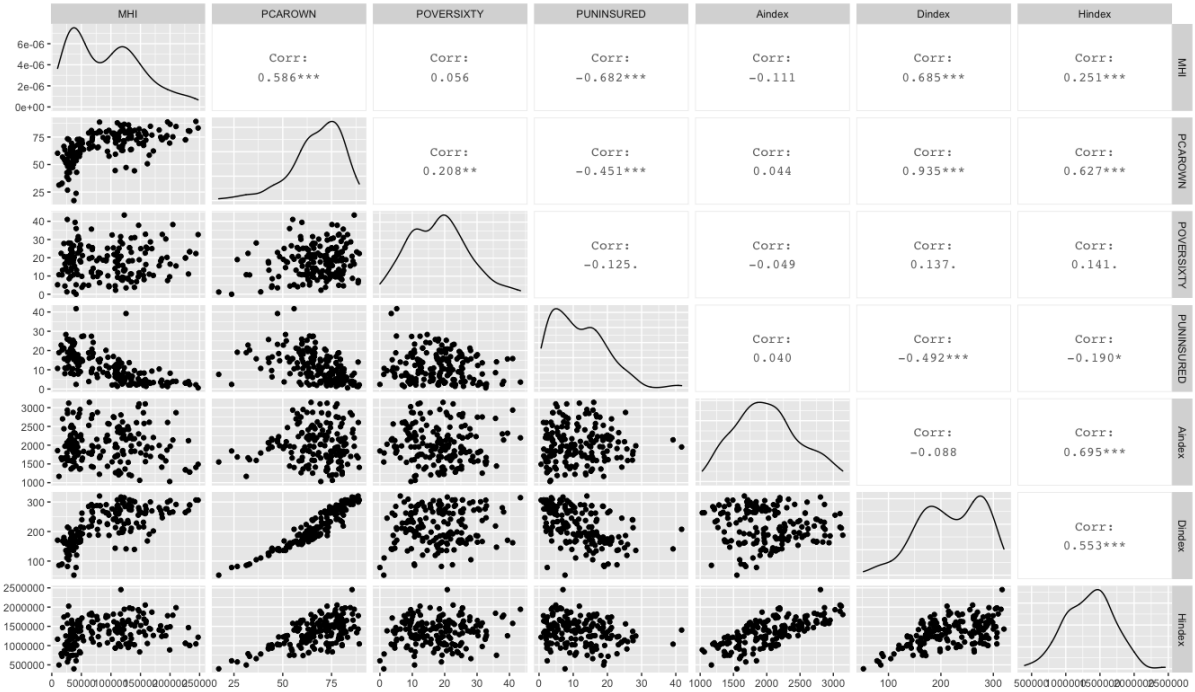

FIGURE 12. Correlation matrix and density estimation

Table 4 - 6 shows three sets of the results of the regression models with different predictive variables. In general, all three models achieve satisfying goodness of fit (R^2 > 0.80), which indicates a good prediction performance. The coefficients indicate a mixed result. In the OLS regression, all three indices are statistically significant. Abundance and Diversity index show a negative correlation with the crude prevalence, while the healthiness index show a very mild but positive correlation, adjusted for SES factors which are all significant as well.

The spatial lag model and error model both return significant spatial lagged variables: ρ and λ consistently. In comparison, the error model returns larger Lagrange Multipliers (not shown in Table 4) than the lag model. The coefficients clearly show that only the diversity index is significantly correlated with disease prevalence while abundance and healthiness index are not significant any more after adjusted for the spatial lag effect.

There is certainly more than one interpretation of the mixed results obtained from the models. From the perspective of pure statistics, it can be concluded that after adjusting for spatial autocorrelation effect and related SES characteristics, only the diversity of food providers has a significant protective effect on the crude prevalence of nutrient-related diseases. It is not surprising that the general abundance of food choice

does not predict diseases well since living in a food-swamped area has little to do with what residents end up choosing to consume. In contrast, what does surprise us is that the healthiness index has an insignificant positive correlation with disease prevalence, which is obviously against the consistent results that most literature achieved before.

	Dependent variable:		
	BPHIGH_CrudePrev	CHD_CrudePrev	DIABETES_CrudePrev
	(1)	(2)	(3)
Aindex	−0.004*	−0.001**	−0.002*
	(0.002)	(0.0005)	(0.001)
Dindex	−0.09***	−0.02***	−0.05***
	(0.02)	(0.004)	(0.01)
Hindex	0.0000**	0.0000*	0.0000*
	(0.0000)	(0.0000)	(0.0000)
MHI	−0.0000***	−0.0000**	−0.0000***
	(0.0000)	(0.0000)	(0.0000)
POVERSIXTY	0.66***	0.16***	0.29***
	(0.05)	(0.01)	(0.02)
PUNINSURED	0.31***	0.07***	0.17***
	(0.07)	(0.02)	(0.04)
Constant	36.76***	5.89***	14.93***
	(3.88)	(0.88)	(1.97)
R^2	0.75	0.72	0.76
Adjusted R^2	0.74	0.71	0.75
F Statistic (df = 6; 169)	84.01***	72.06***	86.81***
Note:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:		
	BPHIGH_CrudePrev	CHD_CrudePrev	DIABETES_CrudePrev
	(1)	(2)	(3)
$\rho(lag\ coef.)$	0.538***	0.444***	0.531***
	(0.058)	(0.0005)	(0.062)
Aindex	0.001	−0.001	−0.000
	(0.003)	(0.001)	(0.001)
Dindex	-6.164***	-1.333**	-3.151***
	(1.678)	(0.42)	(0.901)
Hindex	0.0000	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
MHI	−0.0000	−0.0000	−0.0000
	(0.0000)	(0.0000)	(0.0000)
POVERSIXTY	0.407***	0.114***	0.164***
	(0.036)	(0.010)	(0.018)
PUNINSURED	0.222***	0.045***	0.121***
	(0.049)	(0.013)	(0.025)
Constant	23.731***	5.204**	11.642***
	(6.77)	(1.619)	(3.412)
R^2	0.88	0.81	0.88
Adjusted R^2	0.84	0.75	0.85
Breusch-Pagan Test	41.36***	36.98***	65.08***
Note:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:		
	BPHIGH_CrudePrev	CHD_CrudePrev	DIABETES_CrudePrev
	(1)	(2)	(3)
$\lambda(error\ coef.)$	0.538***	0.444***	0.588***
	(0.058)	(0.0005)	(0.076)
Aindex	-0.003	−0.002	−0.002
	(0.006)	(0.001)	(0.003)
Dindex	−15.608***	-3.221**	-8.223***
	(2.288)	(0.506)	(1.136)
Hindex	0.0001	0.001*	0.000
	(0.0001)	(0.0000)	(0.0000)
MHI	−0.0000	−0.0000	−0.0000**
	(0.0000)	(0.0000)	(0.000)
POVERSIXTY	0.406***	0.112***	0.159***
	(0.038)	(0.010)	(0.020)
PUNINSURED	0.218***	0.046**	0.120***
	(0.055)	(0.014)	(0.028)
Constant	69.07***	12.457***	31.915***
	(7.16)	(1.606)	(3.566)
R^2	0.86	0.79	0.86
Adjusted R^2	0.82	0.73	0.82
Breusch-Pagan Test	35.78***	25.06***	67.98***
Note:			*p<0.1; **p<0.05; ***p<0.01

TABLE 4-6. Results of OLS, Spatial Lag, and Spatial Error regression models

6. Discussion and Conclusion

Food insecurity is a serious national issue now as geographic and non-geographic disparities have been deepened and reinforced across the urban region. In 2016, a national average of 12.3% households are food insecure (A et al. 2017). The number is surging as the economy experiences a catastrophic attack from COVID-19. In spite of a strong urgency, it still bears a huge risk of wasting resources if policy makers rush to a 'one size fits all' solution, which is often solely based on physical determinism, without sorting out some fundamental questions as to what constitutes food access, and what in essence affects population's food choice.

With an aim to bridge the gap and explore the hidden information not uncovered by existing food access research, this study proposes a multidimensional and spatiotemporally dynamic approaches to delineate the food environment in metropolitan Atlanta. The decomposition of food access into abundance, diversity, and healthiness echoes some key conclusions in previous literature: food insecurity is not merely 'not enough to eat', but a complicated, multidimensional barrier to a high-quality, various, and desirable diet affected by a large number of factors. Focusing more on the geographic access, this study examines a subset of factors including socioeconomic status, distance, food type, price tolerance, etc., while still leaves a lot untouched. For instance, cultural barrier can serve as a huge incentive / disincentive to food choice as individual's cultural background has a huge influence on the diet pattern.

The inconsistency between different dimensions justifies the fact that it is not legitimate to make a dichotomous judgement of food access based on geographic place, especially when there exists consistent evidence showing that the reason for food insecurity is much more complicated than 'nowhere to buy'.

Furthermore, mixed results are found in the regression models assessing the association between food access index and disease morbidity rate. There is overwhelming evidence that diversity of food choices is strongly correlated with disease prevalence while the quantity of choices is not. Based on results of the regression models, it can be asserted that the abundance index and diversity index show satisfying accuracy in representing the true picture and in making predictions. However, the reliability of the healthiness index is still questionable.

The limitation of this study is obvious and may pose a challenge to the reliability and legitimacy of the results obtained. First, however sophisticated the algorithm is designed, there is hardly any way that we can determine the difference from 'true' exposure to the relative exposure hypothesized by assumptions and mathematical models. Secondly, as a cross-sectional study, it has very limited capability in proving any causality by displaying a clear prospective temporality from food access to nutrient-related diseases. Finally, constrained by the limited availability of data, some dimensions of food access, for instance healthiness, are still difficult to be accurately uncovered despite its great significance. New sources of data, such as social media network, can be further explored.

Nonetheless, this study provides novel insights into the measurement of food access and the ability of food choice in explaining the prevalence of relevant diseases, which can be a useful information for community planners and health policy makers who seek to understand the overall situation of local foodscape and make plans to facilitate better food access for population living in food insecurity.

References

A, Coleman-Jensen, Rabbitt MP, Gregory CA and Singh A. 2017. Household food insecurity in the United States in 2016. *USDA-ERS Economic Research Report.*

Amin, M. D., S. Badruddoza and J. J. McCluskey. (2020). Predicting access to healthful food retailers with machine learning. Food Policy, 101985. issn: 0306-9192 (Print) 0306-9192 (Linking). https://doi.org/10.1016/j.foodpol.2020.101985. Available at <https://www. ncbi.nlm.nih.gov/pubmed/33082618%20https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC7564312/pdf/main.pdf>.

Anaya-Izquierdo, K., and N. Alexander. (2020). Spatial regression and spillover effects in cluster randomized trials with count outcomes. *Biometrics,* issn: 1541-0420 (Electronic) 0006341X (Linking). https://doi.org/10.1111/biom.13316. Available at <https://www.ncbi.nlm. nih.gov/pubmed/32557560>.

Anselin, Ibnu Syabri, Luc, and Youngihn Kho. (2006). An Introduction to Spatial Data Analysis. *Geographical Analysis.*

Anselin, Luc. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis.*
Atlanta Regional Commission. 2020. About the Atlanta Region. *ARC* (August). Available at <https://atlantaregional.org/atlanta-region/about-the-atlanta-region>.

CA,Tait. (2018).Theassociationbetweenfoodinsecurityandincidenttype2diabetesinCanada: apopulation-basedcohortstudy.*PLoSONE,*https://doi.org/10.1371/journal.pone.0195962.

CDC. (2011). Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI). Available at <https://www.cdc.gov/obesity/downloads/census-tract-level-statemaps-mrfei_TAG508.pdf>.

CDC. (2019). 500 Cities Project: Local data for better health, December. Available at <https: //www.cdc.gov/500cities/index.htm>.

Cummins, Steven, and Sally Macintyre. (2002). A Systematic Study of an Urban Foodscape: The Price and Availability of Food in Greater Glasgow. *Urban Studies* 39 (11):2115– 2130. Available at <https://doi.org/10.1080/0042098022000011399>.

Dai, D. (2010). Black residential segregation, disparities in spatial access to health care facilities, and late-stage breast cancer diagnosis in metropolitan Detroit. *Health Place* 16 (5):1038–52. issn: 1873-2054 (Electronic) 1353-8292 (Linking). https://doi.org/10.1016/j.healthplace. 2010.06.012. Available at <https://www.ncbi.nlm.nih.gov/pubmed/20630792>.

Garg, A., S. Toy, Y. Tripodis, J. Cook and N. Cordella. (2015). Influence of maternal depression on household food insecurity for low-income families. *Acad Pediatr* 15 (3):305–10. issn: 1876-2867 (Electronic) 1876-2859 (Linking). https://doi.org/10.1016/j.acap.2014.10.002.
Available at <https://www.ncbi.nlm.nih.gov/pubmed/25454368>.

Heflin, Colleen M., and James P. Ziliak. (2008). Food Insufficiency, Food Stamp Participation, and Mental Health*. *Social Science Quarterly* 89 (3):706–727. https://doi.org/https: //doi.org/10.1111/j.1540-

6237.2008.00556.x. eprint: https://onlinelibrary.wiley.com/doi/ pdf/10.1111/j.1540-6237.2008.00556.x. Available at <https://onlinelibrary.wiley.com/doi/ abs/10.1111/j.1540-6237.2008.00556.x>.

Howard, Larry L. (2011). Does Food Insecurity at Home Affect Non-Cognitive Performance at School? A Longitudinal Analysis of Elementary Student Classroom Behavior. *Economics of Education Review.*

Jones, S. J., and E. A. Frongillo.(2007). Food insecurity and subsequent weight gain in women. *Public Health Nutr* 10 (2):145–51. issn: 1368-9800 (Print) 1368-9800 (Linking). https: //doi.org/10.1017/S1368980007246737. Available at <https://www.ncbi.nlm.nih.gov/ pubmed/17261223>.

Kirkpatrick, S. I., L. McIntyre and M. L. Potestio. (2010). Child hunger and long-term adverse consequences for health. *Arch Pediatr Adolesc Med* 164 (8):754–62. issn: 1538-3628 (Electronic) 1072-4710 (Linking). https://doi.org/10.1001/archpediatrics.2010.117. Available at <https://www.ncbi.nlm.nih.gov/pubmed/20679167>.

Laraia, B. A., A. M. Siega-Riz, J. S. Kaufman and S. J. Jones. (2004). Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Prev Med* 39 (5):869–75. issn: 0091-7435 (Print) 0091-7435. https://doi.org/10.1016/j.ypmed.2004.03.018.

McIntyre, L., J. V. Williams, D. H. Lavorato and S. Patten. (2013). Depression and suicide ideation in late adolescence and early adulthood are an outcome of child hunger. *J Affect Disord* 150 (1):123–9. issn: 0165-0327. https://doi.org/10.1016/j.jad.2012.11.029.

Morland, K., S. Wing and A. Diez Roux. (2002). The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health* 92 (11):1761–7. issn: 0090-0036 (Print) 0090-0036. https://doi.org/10.2105/ajph.92.11. 1761.

Murrell, A., and R. Jones. (2020). Measuring Food Insecurity Using the Food Abundance Index: Implications for Economic, Health and Social Well-Being. *Int J Environ Res Public Health* 17 (7). issn: 1660-4601 (Electronic) 1660-4601 (Linking). https://doi.org/10.3390/ ijerph17072434. Available at <https://www.ncbi.nlm.nih.gov/pubmed/32260107%20https: //res.mdpi.com/d_attachment/ijerph/ijerph-17-02434/article_deploy/ijerph-17-02434v2.pdf>.

Na, M., S. Eagleton, L. Jomaa, K. Lawton and J. Savage. n.d. (2019) *Food Insecurity Is Associated with Suboptimal Sleep Quality Among Low-income Head Start Preschool-aged Children (P04-071-19).* Curr Dev Nutr. 2019 Jun 13;3(Suppl 1):nzz051.P04-071-19. doi: 10.1093/cdn/nzz051.P04-071-19. eCollection 2019 Jun. Book. isbn: 2475-2991 (Electronic).

Perchoux,C.,B.Chaix,S.CumminsandY.Kestens.(2013).Conceptualizationandmeasurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility. *Health Place* 21:86–93. issn: 1873-2054 (Electronic) 1353-8292 (Linking). https: //doi.org/10.1016/j.healthplace.2013.01.005. Available at <https://www.ncbi.nlm.nih.gov/ pubmed/23454664>.

Purcell, Mark, and J. Christopher Brown. (2005). Against the local trap: scale and the study of environment and development. *Progress in Development Studies* 5 (4):279–297. https://doi. org/10.1191/1464993405ps122oa. Available at <https://doi.org/10.1191/1464993405ps 122oa>.

Rodriguez, L. R., K. L. Hendrickson and E. B. Rasmussen. (2018). Development and quantification of measures for risky and delayed food and monetary outcome choices. *Behav Processes* 151:16–26. issn: 0376-6357. https://doi.org/10.1016/j.beproc.2018.02.020.

Rose, D. (2010). Access to healthy food: a key focus for research on domestic food insecurity. *J Nutr* 140 (6):1167–9. issn: 1541-6100 (Electronic) 0022-3166 (Linking). https://doi.org/ 10.3945/jn.109.113183. Available at <https://www.ncbi.nlm.nih.gov/pubmed/20410087% 20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2869501/pdf/nut1401167.pdf>.

Sadler, R. C., J. A. Gilliland and G. Arku. (2016). Theoretical issues in the 'food desert' debate and ways forward. *Geojournal* 81 (3):443–455. issn: 0343-2521. https://doi.org/10.1007/ s10708-015-9634-6. Available at <https://link.springer.com/content/pdf/10.1007/s10708015-9634-6.pdf>.

Shannon, C. E. (1997). The mathematical theory of communication. *MD Comput* 14 (4):306–17. issn: 0724-6811 (Print) 0724-6811 (Linking). Available at <https://www.ncbi. nlm.nih.gov/pubmed/9230594>.

Sherman, J. E., J. Spencer, J. S. Preisser, W. M. Gesler and T. A. Arcury. (2005). A suite of methodsforrepresentingactivityspaceinahealthcareaccessibilitystudy.*IntJHealthGeogr* 4:24. issn: 1476-072X (Electronic) 1476-072X (Linking). https://doi.org/10.1186/1476072X-4-24. Available at <https://www.ncbi.nlm.nih.gov/pubmed/16236174>.

South, A. M., D. Palakshappa and C. L. Brown. (2019). Relationship between food insecurity and high blood pressure in a national sample of children and adolescents. *Pediatr Nephrol* 34 (9):1583–1590. issn: 0931-041X (Print) 0931-041x. https://doi.org/10.1007/s00467019-04253-3.

The United Nations. (1948). *Universal Declaration of Human Rights.*

U.S. Census Bureau. (2018a). Age and Sex, 2013-2018 American Community Survey 5-year estimates. U.S. Census Bureau. Available at <https://data.census.gov/cedsci/table?t=Age% 20and%20Sex&tid=ACSST1Y2019.S0101&hidePreview=false>.

U.S. Census Bureau. (2018b). LEHD Origin-Destination Employment Statistics Data (20022018). U.S. Census Bureau. Available at <https://lehd.ces.census.gov/data/#lodes>.

Williams, Peter, and Phil Hubbard. 2001. Who is disadvantaged? Retail change and social exclusion. *The International Review of Retail, Distribution and Consumer Research* 11 (3):267–286. https://doi.org/10.1080/713770599. eprint: https://www.tandfonline.com/ doi/pdf/10.1080/713770599. Available at <https://www.tandfonline.com/doi/abs/10.1080/ 713770599>.