ARTICLE IN PRESS

Applied Geography xxx (xxxx) xxx

Contents lists available at ScienceDirect

Applied Geography

journal homepage: http://www.elsevier.com/locate/apgeog

Using geospatial networking tools to optimize source locations as applied to the study of food availability: A study in Guilford County, North Carolina

Timothy Mulrooney [a,*], Richard Foster [a], Manoj Jha [b], Leila Hashemi Beni [b], Lyubov Kurkalova [b], Chyi Lyi Liang [b], Haoran Miao [b], Greg Monty [b]

[a] Department of Environmental, Earth and Geospatial Sciences, North Carolina Central University, Durham, NC, USA

[b] North Carolina Agricultural and Technical State University, Greensboro, NC, USA

A B S T R A C T

In the study of local-level food security, terms such as food variety, availability, accessibility and utilization represent quantitative metrics to describe one's relationship to the tangible and intangible food environment. Food availability entails how close one is located to the nearest food location. These locations could be healthy and fresh food as applied explicitly to the study food deserts, generally considered to be low-income areas that are far from healthy and fresh food. In the Geographic Information Systems (GIS) network model where travel times and distances are either calculated along a line network such as a series of roads or via more traditional techniques such as Manhattan or Euclidean distance, healthy and fresh food locations are defined as destinations. The places people are traveling from are referred to as sources. However, modeling source locations can be increasingly complex. In just measuring food availability between all residential parcels to the closest healthy food destination in Guilford County, North Carolina, it requires more than 177,000 route calculations, one for each of the residential parcels in Guilford County, North Carolina. Research (Zandbergen and Hart 2009; Fischer 2004; Sahar et al. 2019; Winn 2014) has highlighted the challenges in efficiently locating many addresses and calculating so many routes. In order to simplify the number of network calculations, this research explores ways to model, agglomerate or simplify source locations to decrease the sheer number of calculations while not degrading results when compared to calculations using all original 177,000 source locations. Studies in the field of food security have modeled source locations as census tract centroids, block group centroids, as well as random points and even fishnets or grids. In this paper, we explore the use of different techniques to simulate source locations in the study of food availability in Guilford County, North Carolina. These results are compared to calculations using all residential source locations in the county as a baseline. While all eleven techniques, which include random, stratified and systematic, as well as combinations of them, showed some level of agreement with baseline measurements, sources simulated as block centroids, population-weighted block group centroids and even a randomized-strata technique were strongest using t-tests of two means and equivalence tests for dissimilarity for both drive-distance and drive-time.

1. Introduction

It is difficult to quantitively explain the nexus at the spatial distribution of healthy food, spending patterns, health outcomes and explanatory factors behind them. The body of knowledge transcending studies in food deserts and the food environment entails a combination of qualitative methods, quantitative calculations and mixed methods, which collect and analyze both qualitative and quantitative data within the same study. Contemporary research has been trending in this direction (Brindle-Fitzpatrick, 2015; Chrisinger, 2016; Shannon, 2015) since they also take into account individual perceptions and descriptions of the food environment that can not be encapsulated using data agglomerated within enumerations units such as block groups, census tracts, zip codes and counties. Quantitative methods explore numerical relationships between food availability, food access, food utilization and explain it for a variety of metrics such as race/ethnicity, poverty status and access to transportation, among other things. These quantitative metrics are immune to the effects of opinion and perception, and help provide an impartial look at one's relationship to the food environment.

A Geographic Information System (GIS) serves as a powerful tool to examine quantitative spatial relationships that exist between and among the various agents within the food environment. These agents include source locations (where people are traveling from) and destinations (where people are traveling to) in order to procure all forms of food, both healthy and unhealthy. As applied to this study, the Food and Agriculture Organization of the United Nations (FAO) definition of food availability (FAO 2006) will be used, which measures how close one is located to the nearest food location. Food availability, sometimes referred to as proximity (Thornton et al., 2011), differs slightly from food access, which takes into account both availability and a transportation component (FAO 2006). While healthy food may be readily available, it may not be accessible if one does not have the means of

* Corresponding author.
E-mail address: tmulroon@nccu.edu (T. Mulrooney).

https://doi.org/10.1016/j.apgeog.2021.102415
Received 19 February 2020; Received in revised form 20 January 2021; Accepted 6 February 2021

transportation to get to it. As with both of these metrics, both availability and access can be qualified given the various modes of individualized and non-individualized transportation available to some populations based purely on geography alone (i.e. proximity to bus routes or sidewalks) or rurality where a car is the only reasonable mode of transportation.

Ways to measure this availability can be in absolute units such as minutes or miles which are self-explanatory. More recently, relative unitless metrics (Zenk et al., 2014; Clary et al., 2015; Mason et al., 2013) have been used as alternatives to absolute measures since a 10-min drive time to the nearest fresh food source in a downtown urban area means something much different than a 10-min drive to the nearest fresh food in a rural area. However, relative metrics have their challenges and potential for misuse much like their absolute measure counterparts. For example, the binary categorization of a food store as just 'UNHEALTHY' or 'HEALTHY' deemphasizes the relative importance of one individual store. A farmers' market and a large supermarket both have the same value when applied in a ratio although the large supermarket is utilized by much more of the population than the farmers' market. Furthermore, proportions which represent the percentage of healthy food outlets against a denominator can be misleading if the denominator includes only a limited classification of food outlets (healthy food + fast food outlets + convenience stores) versus a more comprehensive listing of all outlets where food can be purchased (Thornton et al., 2020) which may include bakeries, hardware stores, clothing outlets, gyms and even laundromats (Lucan et al., 2018). The use of weights to measure an outlet's contribution to the overall quantitative assessment of the food environment (Thornton et al., 2012) or novel relative metrics (Wilkins et al., 2019) using various cohorts of the food environment (fast food, total outlets, restaurants, etc.) are being devised to make stronger associations with health outcomes. Nonetheless, the use of either relative or absolute metrics require additional interpretation beyond the actual numbers yielded as a result of their calculations.

A GIS is used to measure availability between these sources and destinations. This availability can be measured as Euclidean Distance (straight-line), Manhattan Distance (distance between two points at right angles), drive distance or even drive time between a source and destination. While Euclidean and Manhattan Distance are relatively easy and fast to compute within a GIS, the calculation of drive distance and drive time, which is more practical, requires more input parameters, is dependent upon high-quality data and is more resource-intensive. In other cases, the number and types of destinations within a source can be found within an area of interest (AOI) or multiple areas of interest. These AOIs are typically referred to as buffers. Methods have been used to measure how availability and access, whether via distance, drive times or grouped within buffers of varying lengths correlates to socio-economics and ultimately health outcomes. Studies (Pearce et al., 2008, Winkler 2006; Wang et al., 2006) at different scales have found varying degrees of success finding associations between these factors, ranging from strong associations between socio-economic status (SES) and healthy food availability (Giang et al., 2008; Lewis et al., 2005; Morland & Filomena, 2007, Powell et al., 2007) to no relationship between the two (Zenk et al., 2005; Morland et al., 2002). However, regardless of method used, most of the historical literature within the United States has cited that underrepresented communities (poor, minority, rural or combinations thereof) are typically further and less accessible to healthy food, cost more and have less healthy options than their counterparts.

Despite advances in GIS technologies, it is difficult to model all potential drive-time scenarios between a source and destination using desktop computing solutions in a timely manner. In the study area of Guilford County where the population exceeds 500,00 people, there exists 220,702 individual parcels of which 177,080 are residential parcels. Using Dijkstra's Shortest Path First (SPF) algorithm (1959) and a road network with more than 98,000 vertices, a best-case scenario for calculating just one path requires a minimum of 98,000 calculations and a worst-case scenario of more than 9 billion ($98{,}000^2$) calculations for each of the 177,000 paths to be calculated (Mehlhorn 2008). This means there are literally trillions of possible calculations between these residential parcels and potential destinations. While applications using Python, Stata (Huber & Rust, 2016) and R programming solutions make this process faster than Esri's Network Analyst calculations using a GUI (Graphical User Interface), they are less intuitive for the average GIS user and still require billions of calculations.

Research (Fischer, 2004; Sahar et al., 2019; Winn, 2014; Zandbergen & Hart, 2009) has highlighted the challenges in efficiently locating many addresses and calculating so many routes. Because of this, geographers and GIS professionals have explored methods to simplify the number of potential source locations which would in turn decrease the number of calculations, while ensuring this decrease in source locations does not degrade results when compared to the entirety of source locations. This simplification or sampling is an attempt to circumvent issues of scale and in this case size by reducing the number of potential source locations while not compromising results. Within many different applications, prior work has approximated possible source locations as both population and geometry-based centroids (Berke and Shi 2005), random points (Mulrooney, Beratan, et al., 2017), grid-based points (Economic Research Service 2019) and random points within a stratum (Hilson et al., 2015) as highlighted in Table 1. This paper explores a variety of sampling methods for potential source locations in the quantitative measure of food availability using Networking GIS tools and how these methods compare to measurements for all residential locations in Guilford County, North Carolina.

1.1. Literature review

The mapping and delineation of food-unavailable areas within the digital environment has been made exponentially easier with GIS technologies. While originally used as an aesthetic tool to map study areas (Wrigley et al., 2002) or display underlying explanatory variables (Guy et al., 2004), GIS has since been used to measure distances, quantitatively express availability and render this availability with statistical significance using a variety of analytical, geostatistical and cartographic techniques. Among the first to do this within the realm of the food environment were Donkin et al. (1999), Lovett et al. (2002) and Pearce (2006).

Within the GIS data environment, ways to express quantitative dimensions of the food environment vary from study to study. Prior research has expressed these measures as absolute linear units such as kilometers or miles (Jago et al., 2007), travel time in minutes (VerPloeg et al., 2009; Jiao et al., 2012) and densities such as the number of food options per square mile by census tract (Block et al., 2004), as well as derived metrics such as the cost to operate a car (Hallett & McDermott, 2011). More recently, the aforementioned relative metrics have been implemented as complements to absolute metrics. The Retail Food Environmental Index (RFEI) and the Expanded RFEI (Cooksey-Stowers et al., 2017, Lucan et al., 2015) are examples of widely-accepted relative metrics, while others (Mulrooney, McGinn, et al., 2017; Rose et al., 2009) have derived their own metrics and subsequent interpretations to define spatial extents of food deserts and food swamps using the RFEI, Expanded RFEI and Modified RFEI (mRFEI) developed by the Centers for Disease Control (2009) and Food Balance Metric (Gallagher, 2006) as guidelines.

Nonetheless, food availability can be mapped at different scales or within different enumeration units such as those seen in Fig. 1 which highlight travel time to the nearest healthy food location using two different sized enumeration units, the census block group and the census tract. In the United States, counties are further divided into census tracts, census block groups and census blocks. While counties have been used as the basis for food environment research (California Center for Public Health Advocacy, 2007; Sisiopiku & Barbour, 2014), it is not at a scale that facilitates local-level interventions necessary to implement

Table 1
Different sampling techniques used in this study.

Number	Method	Explanation	Total # Points	Prior Work
1	Guilford Residential Parcels	The centroids of all Guilford County Residential parcels were extracted as centroids and used as source locations in Network Analyst model. Calculations were grouped into block groups and compared to other modeling techniques grouped at the BG level.	177,080	
2	Block Group Centroid	The centroids of all block groups in Guilford County were extracted as centroids and used as source locations.	292	Berke and Shi 2005
3	Pop. Weighted BG Centroid	The weighted centroids of all block groups, weighted by population at the block level, in Guilford County were extracted as centroids and used as source locations.	291	Berke and Shi 2005
4	Block Centroid	The centroids of all census blocks in Guilford County and were extracted as centroids. These blocks were grouped into the appropriate block group.	8183	Berke and Shi 2005
5	Populated Block Centroid	The centroids of all census blocks in Guilford County with a population 50 or more were extracted as centroids. These blocks were grouped into the appropriate block group.	2745	Berke and Shi 2005
6	1/2 Mile Fishnet	Source locations automatically placed at 1/2 mile increments and grouped into appropriate BG after Network Analyst calculation.	2631	Economic Research Service 2019
7	1/4 Mile Fishnet	Source locations automatically placed at 1/4 mile increments and grouped into appropriate BG after Network Analyst calculation.	10,520	Economic Research Service 2019
8	Random Points (1,000)	Random points randomly distributed throughout study area and grouped into appropriate BG after Network Analyst calculation.	5000	Mulrooney, Beratan, et al. (2017)
9	Random Points (5,000)	Random points randomly distributed throughout study area and grouped into appropriate BG after Network Analyst calculation.	1000	Mulrooney, Beratan, et al. (2017)
10	Random Points (4 per BG)	Area is split into strata (BGs) and each strata contains 4 points, randomly placed throughout BG.	1168	Hilson et al., 2015
11	Stratified Random - Area (BG)	Area is split into strata (BGs) and each strata contains proportional number of points based on the area of the BG, randomly placed throughout BG.	5000	Hilson et al., 2015
12	Stratified Random - Pop (BG)	Area is split into strata (BGs) and each strata contains proportional to the population of the BG, randomly placed throughout BG.	5000	Hilson et al., 2015

practical policy. In North Carolina, there are exactly 100 counties. Within them, there exist 2174 census tracts and 6107 census block groups. Contemporary neighborhood-level research has used these units as a basis for research because they 1) are able to articulate sub-county/neighborhoods patterns and 2) explanatory data via the census, American Community Survey (ACS) or spending patterns are also available at that scale, which in turn make for easy comparison, correlation and associations.

Food availability studies at the tract level by Baker et al. (2006), Block et al. (2004) and Zenk and Powell (2008) measure metrics such as clustering to supermarkets, density to fast food and density of convenience stores within for particular study areas, respectively, while the mRFEI (Centers for Disease Control and Prevention, 2011) measures relative food availability (ratio of healthy to unhealthy food) at the tract level across the entire United States. The gold standard for measuring food access in the United States, the United States Department of Agriculture Food Access Atlas (Economic Research Service 2019) specifically measures food deserts (low income and limited access) as well as the individual components that make up this metric combined with ancillary measures such as income, vehicle access and high-density housing at the tract level.

In this work, evaluating and displaying food availability at the household level becomes one issue, the scale at which to render results becomes another. Household-level results such as those highlighted in Fig. 2, while useful, require display at an extremely high scale. This becomes problematic when exploring and comparing counties, tracts and even block groups, as the 292 block groups in Guilford County average 680 households per block group. However, block group level analysis has been performed by Jiao et al. (2012), Sharkey and Horel (2008), Mulrooney, McGinn, et al. (2017) and Chen and Clark (2013). Other research such as Van Hoesen (2014) has used towns as enumeration units for a study in Vermont. Towns in Vermont are non-overlapping polygonal units that are smaller than census tracts, but larger than block groups. Some other states such as New Jersey, Pennsylvania, Michigan and Ohio are divided into sub-county units such as townships, cities, towns and boroughs at which high-quality information is also collected.

Using vector-based tools, network analysis and geoprocessing algorithms can compute availability between a destination location from a particular source. In some cases, these destination locations may be retail food and fast food (Wang et al., 2007), any location selling fruits and vegetables (Winkler 2006), fast food and convenience stores (Zenk & Powell, 2008) or a combination of supermarkets, grocery stores, convenience stores and fast good (Liu et al., 2007). As previously mentioned, this availability can be measured in different ways. While Wang et al. (2007) and Winkler (2006) measured this availability using

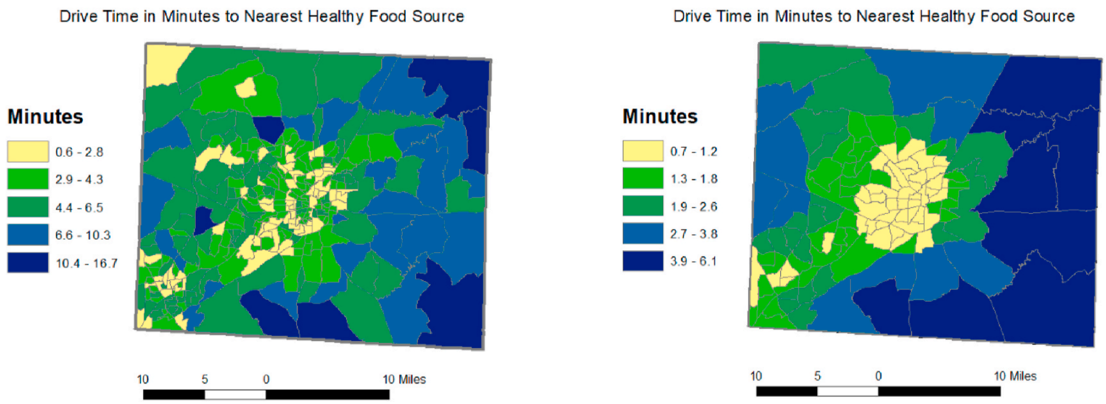

Fig. 1. Average drive-time to healthy food sources in Guilford County for 292 Block Groups (left) and 119 Tracts (right).

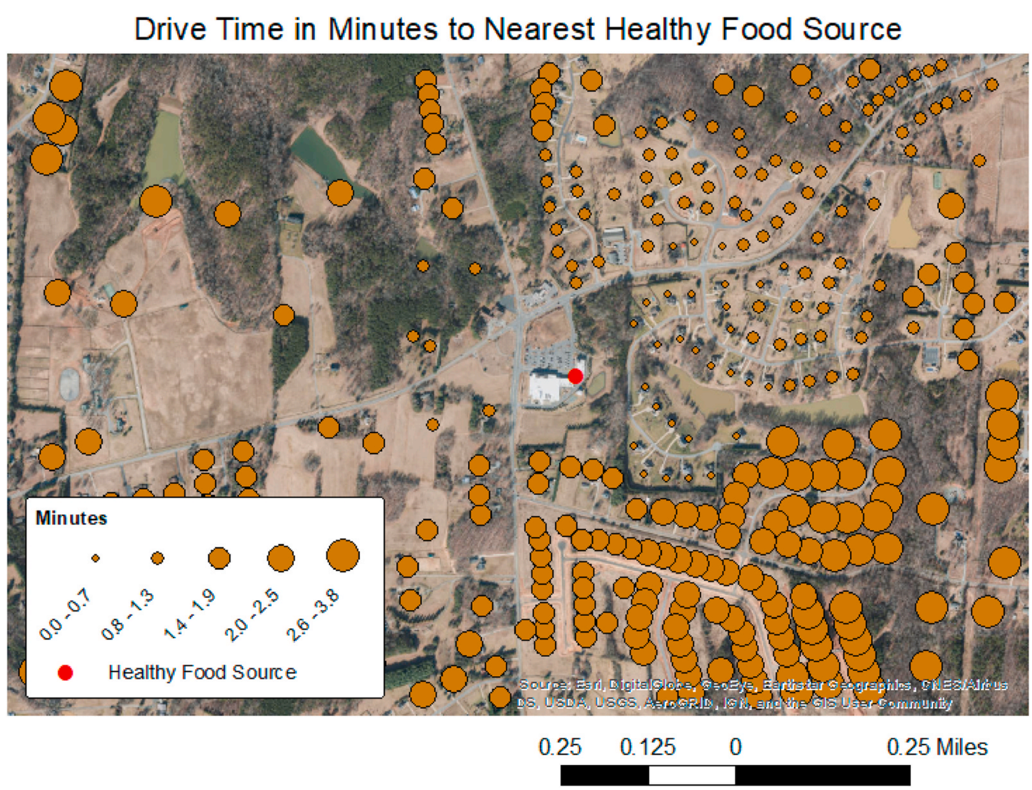

Fig. 2. Example of household-level drive-time analysis.

Euclidean distance, Zenk and Powell (2008) measured walking distance while Liu et al. (2007) and Chen and Clark (2013) used driving distance. Pearce et al. (2008) measured this availability as a function of drive time.

While network-based (drive-time and drive-distance) measurements at the block group level serve as the foundation for this research, this paper focuses specifically on the representation of sources, whether it be schools (Austin et al., 2005), census blocks (Smoyer-Tomic et al., 2008), centroids of block groups (Gallagher, 2006), the actual addresses of households (Van Hoesen 2014; Bodor et al., 2008, Burdette et al., 2004) or even random points (Mulrooney, Beratan, et al., 2017) using Network Analyst tools as applied in the study of food availability. Using Network Analyst, one is able to determine the closest destination (fresh food) using a travel-time and distance impedance along a network from a source. The United States Department of Agriculture (USDA) Food Access Atlas uses the center of ½ km grids, or a fishnet, as their source in creating a continuous raster surface of access in concert with socio-economic data (access to transportation, poverty, etc.) rasterized at this scale from existing census data grouped at various scales (United States Department of Agriculture, 2020). In the case of schools and block groups where the number of sources is relatively small (in the dozens to perhaps hundreds), the calculations of these networks is relatively fast and comparable to non-network based calculations of Euclidean and Manhattan distance using tradition GIS geoprocessing tools. However, as the number of source locations increase, these calculations become more complex and longer to process.

In field-based research revolving around quality assurance/quality control (QA/QC), geographical sampling presents a number of issues in trying to avoid non-probabilistic sampling techniques. For example, inaccessible sample points pose logistical challenges for those physically visiting these points that do not exist in the digital environment. In this case where sample points are simulated in a GIS environment in order to reduce the sheer number of calculations without degrading results, various methods exist to accomplish this. Where little is known about the data, a random distribution of points may be preferable. In other cases where an equal distribution of points is required, a systematic approach such as a fishnet such as those used by the USDA Food Access Atlas can be used. Still in other cases where more data is known, a stratified sampling technique based on strata (block groups) may be used (Wood, 1955) to allow certain strata to be weighted more based on some variable (population or area). Various techniques which use or combine systematic, random and stratified techniques for various sample sizes are highlighted in Table 1 and four of these methods are shown in Fig. 3 as well.

1.2. Study area

As part of a larger research project revolving around food insecurity in three different counties across North Carolina, we conducted this pilot study in Guilford County (Fig. 4). Located in the Piedmont Region of North Carolina, the city of Greensboro dominates the county and serves as the economic and cultural center of this county, which has an area of about 1700 km^2 (658 mi^2) and a population of approximately 502,000 people. Besides Greensboro (population 294,000), Guilford County is also composed of other cities, municipalities and newly incorporated municipalities (NIMs) to include High Point (pop. 112,00), Stokesdale (pop. 5400 and incorporated in 1989), Summerfield (pop. 11,300 and incorporated 1996), Jamestown (pop. 4500), Sedalia (pop. 660 and incorporated in 1997), Oak Ridge (7000 and incorporated in 1998). Despite this, rural regions, using the definition provided by the United States Census (2010), are abundant in the southeastern, northeastern and northwestern portions of the county and represent 16.4% of the population and 55.3% of the land area within the study area.

2. Methodology

In order to test the usefulness of sampling algorithms to model travel time to the nearest healthy food locations, data related to healthy food locations (destinations), residential locations (sources) and a network connecting them (roads) were extracted from existing GIS data sources. The use of the terms source and destination align with their usage in technical applications (Environmental Systems Research Institute 2010) as used in Network Analyst calculations created in support of this research and reinforced by GIS literature (Abousaeidi, Fauzi, & Muhamad, , Raja et al., 2008) which represent travel-time and travel-distance between two locations. Healthy food locations were derived from their NAICS (North American Industry Classification Standard) code, a multinational (United States, Canada and Mexico) standard which classifies business establishments by their primary economic activity. All businesses are provided as point locations by InfoUSA for the year 2018.

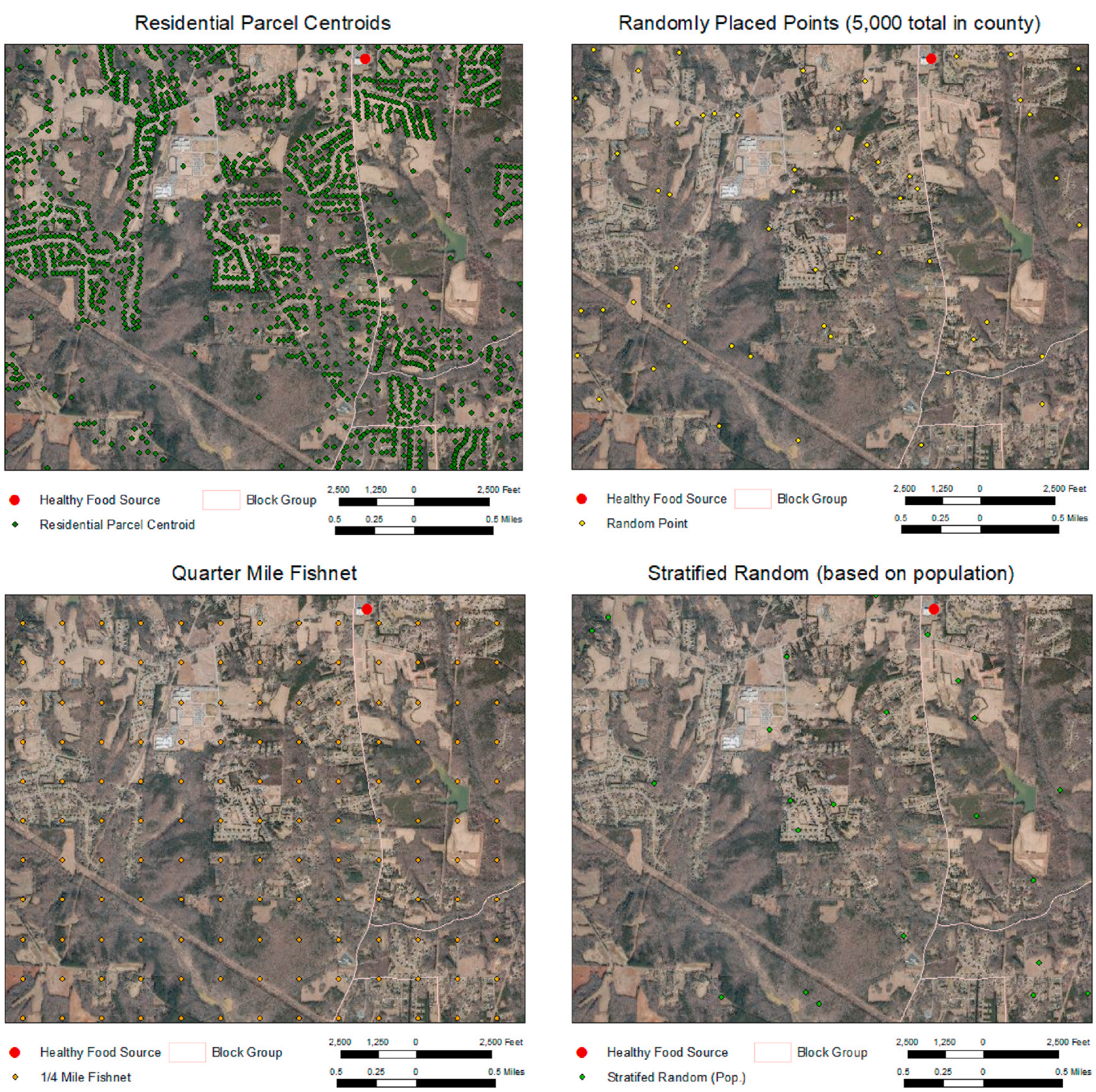

Fig. 3. Examples of different techniques to model source locations in the GIS. They include using all residential parcels (upper left), randomly placed points (upper right), stratified random where number of points are proportional to block group's population (lower right) and ¼ mile fishnet (lower left).

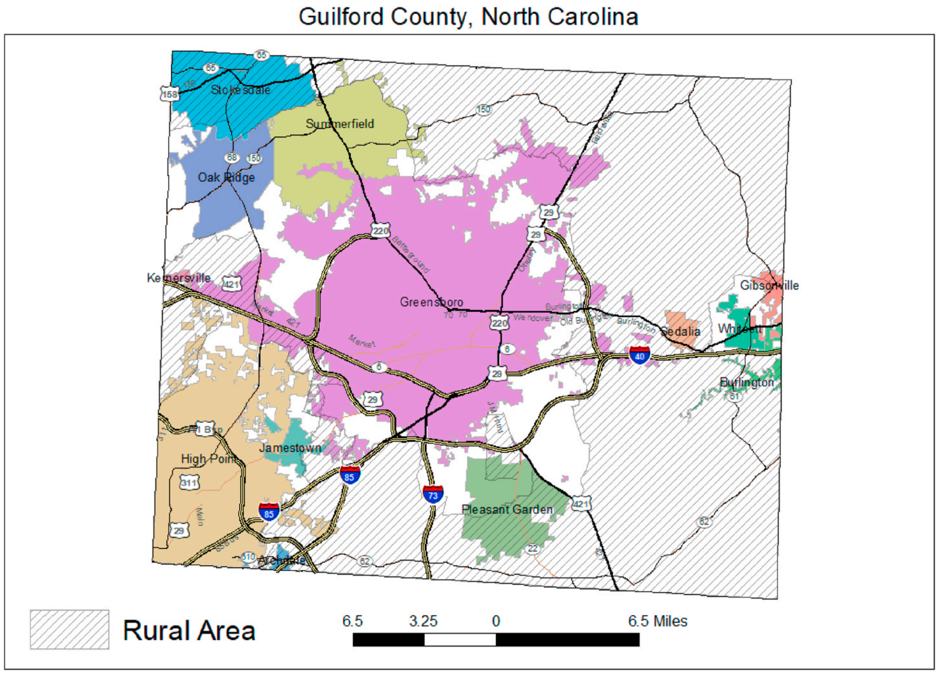

Fig. 4. The study area of Guilford County, North Carolina.

NAICS codes beginning with 44,511 represent supermarkets and other groceries, with further delineations (44511001 through 4451106) to represent increasing larger stores based on the number of employees. In addition, InfoUSA provides attributes about the store's number of employees. Only large (more than10 employees) supermarkets were extracted from this database, regardless of NAICS code since most people would not patronize small stores classified as groceries to satisfy their healthy food needs. In addition, all other large stores that sell fresh food not denoted by this NAICS code (such as = WalMart Supercenter and Target) were merged into this database, as well as farmers' markets. Finally, a 10-mile buffer was applied around Guilford County to include stores outside of the county that may be frequented from Guilford County source locations. In all, a total of 118 healthy food sellers, to be used as destinations, were exported into the final version of this GIS database.

In GIS, a network can be modeled in a way so that cumulative impedances represented quantitatively through its attributes (such as travel time or stream flow) can be calculated above and beyond real-world distance measures (such as meters and miles) that are inherent in GIS. Network calculations have advantages over the aforementioned Euclidean and Manhattan calculations in that they better represent real-world conditions and a human understanding of travel. All travelers care more about travel time or travel distance as opposed to the straight-line distance (Euclidean) or minimum distance when traveling at right angles (Manhattan Distance). While the latter are easy to calculate in a GIS, the former require more data, the correct type of data and the correct attributes in order to make these more complex calculations. The North Carolina Department of Transportation (NCDOT) provides high-quality road data through both services (NCDOT 2020a) and downloadable sites (NCDOT 2020b) that can be used in these calculations. This road network consists of 1,072,300 individual road segments with more than 90 attributes, including road length and speed limit, so high-precision travel time calculations can be derived. These road network data for the entire state of North Carolina were integrated into this GIS database for use in these calculations.

Finally, source locations were extracted or created using a variety of different formats as highlighted in Table 1. In one calculation (block group centroids, Method 2), the centroid of each of the 292 block groups was extracted and used as a source in the Esri Network Analyst calculation to find the nearest facility (in terms of travel-time) between the 292 sources and 118 destinations. In another scenario (Method 6), source locations were created at ½ mile intervals using the *Create Fishnet* function in Esri software. The specific networking tool in which routes between sources and destinations were computed was Esri's *Nearest Facility* tool, which is able to determine the closest facility (healthy food source) using a travel-time impedance along a network from another set of points (modeled sources).

For each of the scenarios described in Table 1, a Network Analyst computation using Dijkstra's Shortest Path First (SPF) algorithm (1959) was run to determine the drive-time and drive-distance between each of the sources and destinations. The results of the calculation are a set of routes, represented as a line, connecting the source and the nearest destination. As highlighted in Table 1, given that the number of sources used in each modeling scheme ranges from the hundreds (block group centroid = 292 sources) to hundreds of thousands (Guilford County Residential Parcels = 177,080 sources), the number of actual calculations varies greatly amongst these techniques which in turn cannot be supported by typical desktop computing solutions. In running the Esri Network Analyst calculation, all desktop computers used by the research team ran out of memory when calculating 177,080 possible routes for all Guilford County residential locations. As a result, the sources were broken down in the manageable subsets of sources and the Network Analyst calculation was run successfully. This process, even automated in the Python programming environment, is time-consuming and this conundrum with in the Esri platform serves as the impetus for this research project.

Using GIS techniques, each route was grouped into its appropriate block group. From the appropriate Federal Information Processing Standard (FIPS) code. For those sources not related to a FIPS code such as the fishnet and random points (Fig. 3), a *Spatial Join* function was run, grouping and averaging each location and calculated travel time/distance with a block group that can be mapped. The result is a choropleth map, highlighting the 292 block groups in the study area and their calculated average drive time using the techniques described above using the same destinations, but different sources as described above.

Given the base enumeration and display unit is the block group, it must be noted that in some cases, there is an equal distribution of source points within each block group while in other cases there may be none.

In Methods 2 (Block Group Centroid) and 10 (4 sources in each BG), each block group contains the same number of source points. In Method 6 where points are systematically placed ½ mile apart, 22 block groups did not have a source point located within it while one block group, the largest in terms of area but the 57th largest in terms of population, contains 127 source points. In Method 9 where 5000 points were randomly distributed throughout the study area, this same block group contains 227 source points while 19 block groups did not contain a source point. Still yet, Methods 11 and 12 create a proportionate distribution of points within each block group based on area and population, respectively. As a result, while the aforementioned block group has the greatest number of source points (236) based on relative area as highlighted in Method 11, it contains only 24 source points based on its population in relation to the rest of the other block groups (Method 12).

For each of the techniques described in Table 1, a choropleth map was created such as those seen in Fig. 5. These maps are based on tabular data representing each of the 292 block groups and the calculated drive time to the nearest healthy food source for each scenario as shown in Tables 1 and 2. For each of the techniques highlighted in Fig. 5, the patterns appear generally the same since they show the same information (drive time to the nearest healthy food facility), albeit calculated differently using the techniques described above. Noticeable is the block group in the northwest part of Guilford County where the block group occurs in the first quantile of data one calculation (source is estimated as the center of the block group) while it occurs in the second quantile for the map to the right. In this case, the healthy food source is located close (just over 1 mile) to the block group centroid which is used as a source. In the second calculation to the right, there are 39 simulated sources located throughout the block group which obfuscate values closer to the destination in this part of the county.

As alluded to before, while there are 292 block groups in Guilford County, the number of block groups with actual sources and calculated values within them that are mapped may be slightly less depending upon the sampling method used. For example, using the ½ mile fishnet technique where source points are evenly spaced ½ mile apart, there were cases where a source point did not fall within a block group, resulting in a Null calculation. This occurs in urban areas where block groups are much smaller. As a result, the sample size for this calculation is 270 since only 270 block groups contain a simulated source point using this technique. Nonetheless, in cases where the block group is used as a strata unit, all 292 block groups will have a calculated value within it. In cases of stratified random techniques based on area and population, block groups may have such small area and/or population that their allocation of source points is rounded down to 0, hence resulting in a number slightly less than 292. In the case of population-weighted centroids, only 291 block groups had a population and one centroid was not computed.

2.1. T-test of two means

Using the Guilford Residential Parcel calculation (Method 1) as a baseline, an independent *t*-test of two means was run between the dataset as a result of the Guilford Residential Parcel calculation and each of the eleven other scenarios to determine if there was a difference between the different cohorts of accuracies.

Average travel time of Method 1 = average travel time other method

$$H_0 : \mu_1 = \mu_2$$

Average travel time of Method 1 ≠ average travel time for other method

$$H_a : \mu_1 \neq \mu_2$$

Using the derived average drive-times for each block group ($\widehat{Y}_1$ and $\widehat{Y}_2$) and the sample sizes for each cohort (N_1 and N_2), this test helps determine the criteria in order to reject the Null hypothesis (drive-times or distances from Guilford Residential Parcel calculation are equal) and accept the alternate hypothesis (drive-times or distances from Guilford Residential Parcel calculation are not equal to each other).

2.2. Equivalence testing

One of the challenges in working with the *t*-test of two means is the interpretation of results. When running these tests, insufficient evidence to reject the null hypothesis as shown above may not necessarily imply similarity and the probability of making a Type II error still exists. This non-rejection of the null hypothesis as a means to assume equivalence can be thought of as convoluted and speculative. As a result, researchers have explored testing mechanisms to directly measure equivalence (Berger & Hsu, 1996; Cribbie et al., 2004; Seaman & Serlin, 1998) between different groups. These tests use statistical tools to determine if the means between two groups are small enough to be considered inconsequential.

In equivalence testing, the null and alternate hypotheses are essentially switched compared to difference-based t-tests. In equivalence testing, the null hypothesis dictates the difference of the means fall outside of an equivalence interval (θ), which may not necessarily be symmetric about a mean. The alternate hypothesis states the difference of the means falls within this confidence interval.

$$H_{01} : \mu_1 - \mu_2 \leq \theta_1 \, H_{a1} : \mu_1 - \mu_2 > \theta_1$$

$$H_{02} : \mu_1 - \mu_2 \geq \theta_2 H_{a2} : \mu_1 - \mu_2 < \theta_2$$

In order to further support the results from the t-tests, paired equivalence tests were run between Method 1 (sources simulated for all residential parcels in Guilford County) and each of the other methods

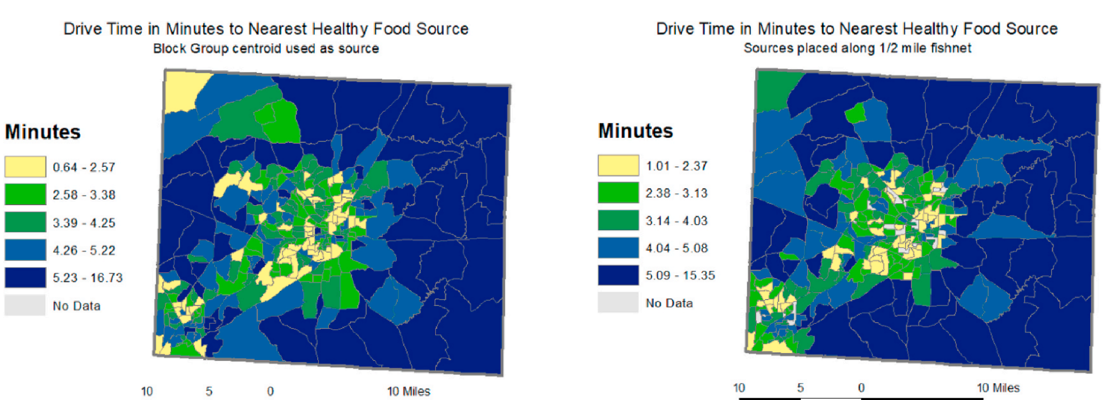

Fig. 5. Results of Nearest Facility analysis in which routes between sources and destinations are agglomerated at the block group level when block group centroids are used as sources (left) and ½ fishnet is implemented (right) using a quantile mapping classification.

Table 2
Results of paired two-tail *t*-test when comparing Method 1 with each of the other Methods.

				Drive-Distance (Miles)			Drive-Time (Minutes)		
Method	Method Number	N	n (BGs)	$\bar{x}$	t	ρ	$\bar{x}$	T	ρ
Guilford Residential Parcels	1	177,080	292	1.701	–	–	3.998	–	–
Block Group Centroid	2	292	292	1.801	−0.844	*0.399*	4.229	−1.299	*0.194*
Pop. Weighted BG Centroid	3	291	291	1.781	−0.662	*0.508*	4.154	−0.891	*0.373*
Block Centroid	4	8183	292	1.703	−0.017	*0.987*	4.016	0.132	*0.895*
Populated Block Centroids	5	2745	291	1.730	−0.265	*0.791*	4.202	−1.294	*0.196*
1/2 Mile Fishnet	6	2631	270	1.800	−0.858	*0.391*	3.958	0.246	*0.806*
1/4 Mile Fishnet	7	10,520	290	1.789	−0.790	*0.430*	4.006	−0.046	*0.964*
Random Points (1000)	8	1000	196	2.050	−2.686	0.007	4.359	−1.909	*0.057*
Random Points (5000)	9	5000	273	1.810	−0.944	*0.346*	4.016	−0.102	*0.919*
Random Points (4 per BG)	10	1168	292	1.751	−0.447	*0.655*	3.917	0.495	*0.621*
Stratified Random - Area (BG)	11	5000	290	1.789	−0.790	*0.430*	3.989	0.059	*0.953*
Stratified Random - Pop (BG)	12	5000	291	1.748	−0.418	*0.676*	3.893	0.655	*0.513*

Null Hypothesis (Means Equal to Each Other) Acceptable at β (.05)

highlighted in Table 1. Steeped in behavioral research, the paired-sample test for equivalence developed by Wellek (2003) uses a standardized equivalence interval, population mean difference score and population standard deviation of the differences to derive a t-statistic that can be compared to a critical value. The Two One-Sided Test for Equivalence (TOST) purported by Seaman and Serlin (1998) implements two separate one-sided tests using raw mean differences $(\mu_1 - \mu_2)$ instead of mean difference scores.

3. Results

Descriptive statistics are highlighted in Table 2 for each of the methods used to approximate source locations. A general observation shows the average drive-distance for when grouped at the block group level was lowest when using all Guilford County residential parcels compared to the other methods. However, average drive-time ranked 5th out of the 11 different methods run. This table also highlights results from the *t*-test and resulting ρ score when comparing results from all Guilford County Residential Parcels (Method 1) versus the other eleven methods used. This test essentially compares the average drive-distance and drive-time when grouped at the block level when all Guilford County Residential Parcels are used as sources when independently compared to each of other methods employed in this study (Methods 2 through 12).

In 10 of the 11 t-tests for drive-distance and in all 11 of the t-tests for drive-time, the corresponding p-values were computed to be above the acceptable limits for accepting the alternative hypothesis at that significance level (α = 0.05). The only exception was Method 7 (1000 random points used as sources). As a result, given this sample size and confidence, we can confidently state that there is no difference between drive-time between when grouped at the block group level for sources simulated for all Guilford County Residents versus the other methods used. When simulating drive-distance, all methods except 1 articulated there were no differences using the *t*-test.

The paired *t*-test for equivalence and TOST were run in R, an open source environment used primarily for statistical computing and the results of the test for dissimilarity (the null hypothesis), are shown in Table 4. Indicators of similarity, or rejection of dissimilarity, appeared across all eleven methods at some level. However, their strength varied between the type of test (TOST vs. Paired T-Test) and measurement (drive-time vs. drive-distance) as shown in Table 3. In three cases, where sources were approximated at the population weighted block group centroid (Method 3), block centroid (Method 4) and random points (4 per block group) in Method 10, dissimilarity was rejected across all four variations of the measurement. Combined with the fact that combined p-values between the two methods were so high in the test of two means (indicating extremely little chance the set of values are statistically different), it is safe to assume these methods are excellent ways to model source locations as opposed to running all possible calculations. In addition, other methods such as the populated block centroids (Method 5), ½ mile fishnet (Method 6), ¼ mile fishnet (Method 7), 1000 random points (Method 8) and 5000 random points (Method 9) satisfied three out of the four criteria for the test of dissimilarity. Except for Method 8 which failed the two-tail *t*-test for drive-time, these methods would also be candidates for source modeling routines. Lastly, the block group centroid (Method 2), Stratified Random - Area (Method 11) and Stratified Random - Population (Method 12) only satisfied two of the dissimilarity criteria.

Table 3
Results for the Test for Dissimilarity (Null Hypothesis). Both the Two One-Sided Test for Equivalence (TOST) and Paired T-Test for Equivalence were run between Method 1 and each of the other 11 methods.

				Drive-Distance (Miles)			Drive-Time (Minutes)		
Method	Method Number	N	n (BGs)	$\bar{x}$	TOST	Paired T-Test	$\bar{x}$	TOST	Paired T-Test
Guilford Residential Parcels	1	177,080	292	1.701	–	–	3.998	–	–
Block Group Centroid	2	292	292	1.801	Rejected	Not Rejected	4.229	Rejected	Not Rejected
Pop. Weighted BG Centroid	3	291	291	1.781	Rejected	Rejected	4.154	Rejected	Rejected
Block Centroid	4	8183	292	1.703	Rejected	Rejected	4.016	Rejected	Rejected
Populated Block Centroids	5	2745	291	1.730	Rejected	Rejected	4.202	Rejected	Not Rejected
1/2 Mile Fishnet	6	2631	270	1.800	Rejected	Rejected	3.958	Rejected	Not Rejected
1/4 Mile Fishnet	7	10,520	290	1.789	Rejected	Not Rejected	4.006	Rejected	Rejected
Random Points (1000)	8	1000	196	2.050	Rejected	Not Rejected	4.359	Rejected	Rejected
Random Points (5000)	9	5000	273	1.810	Rejected	Not Rejected	4.016	Rejected	Rejected
Random Points (4 per BG)	10	1168	292	1.751	Rejected	Rejected	3.917	Rejected	Rejected
Stratified Random - Area (BG)	11	5000	290	1.789	Not Rejected	Not Rejected	3.989	Rejected	Rejected
Stratified Random - Pop (BG)	12	5000	291	1.748	Rejected	Rejected	3.893	Not Rejected	Not Rejected

Table 4
Assessment of each of the techniques used in this research and justification for each technique.

Method	Method Number	n	Strength of Technique	Justification
Guilford Residential Parcels	1	177,080	–	Simulates geographic reality, but is time and resource intensive on most desktop computers.
Block Group Centroid	2	292	Fairly Strong	Satisfied t-tests for differences, but only satisfied 2 of the 4 equivalence tests for dissimiliarity.
Pop. Weighted BG Centroid	3	291	Strongest	Satisfied t-tests for differences and satisfied all 4 equivalence tests for dissimiliarity.
Block Centroid	4	8183	Strongest	Satisfied t-tests for differences and satisfied all 4 equivalence tests for dissimiliarity.
Populated Block Centroids	5	2745	Strong	Satisfied t-tests for differences, but only satisfied 3 of the 4 equivalence tests for dissimiliarity.
1/2 Mile Fishnet	6	2631	Strong	Satisfied t-tests for differences, but only satisfied 3 of the 4 equivalence tests for dissimiliarity.
1/4 Mile Fishnet	7	10,520	Strong	Satisfied t-tests for differences, but only satisfied 3 of the 4 equivalence tests for dissimiliarity.
Random Points (1000)	8	5000	Weak	Did not satisfy *t*-test for differences for drive-distance and barely passed for drive-time, but satisfied for 3 of the 4 equivalence tests for dissimiliarty.
Random Points (5000)	9	1000	Strong	Satisfied t-tests for differences, but only satisfied 3 of the 4 equivalence tests for dissimiliarity.
Random Points (4 per BG)	10	1168	Strongest	Satisfied t-tests for differences and satisfied all 4 equivalence tests for dissimiliarity.
Stratified Random - Area (BG)	11	5000	Fairly Strong	Satisfied t-tests for differences, but only satisfied 2 of the 4 equivalence tests for dissimiliarity.
Stratified Random - Pop (BG)	12	5000	Fairly Strong	Satisfied t-tests for differences, but only satisfied 2 of the 4 equivalence tests for dissimiliarity.

4. Discussion

The decision to patronize certain healthy food establishments is a function of many quantitative, qualitative and in-situ factors. This research attempted to model travel scenarios to the nearest healthy food source based solely on geographic proximity. Using the techniques described above, different ways to approximate source locations resulted in varying degrees of success, with population weighted block group centroids (Method 3), block centroids (Method 4) and random points (4 per block group) in Method 10 being the strongest of these methods.

In the course of research, opportunities for future research were elucidated above and beyond those addressed in this paper. These opportunities revolve around the concept of the quantitative assessment of one's relationship to the food environment which would complement the research highlighted here. While Guilford County residential parcels, as well as other modeling scenarios were used as source locations, it is only an attempt to model people's actual consumer behaviors. People may shop for food on the way home from work or as part of other errands. As well, availability is a principal determinant for a primary grocery shopping location for just under half (48%) of US residents (Food Marketing Institute, 2016). Another study (Zenk et al., 2013) found that more than half (53.9%) of residents in Detroit often shopped within 2 miles of their residence. Those who bypass the closest store cite reasons such as lower prices, lower prices on wanted items, better selection and better quality of fresh foods as reasons for doing so (Food Marketing Institute, 2016). Lower income residents may not have the means to be as choosy, as almost half of all Americans have cited that it is 'sometimes' or 'often' true that they would've purchased healthy food options instead of unhealthy ones for economic reasons (International Food Information Council Foundation, 2018). These lower income residents thus are subjected to the grocery store and their options, or lack thereof, that geography dictates, compared to their higher-income counterparts, as well as the inability to procure healthy food options at these stores. This is difficult to encapsulate within this research.

The second quantifiable factor taken from this study is food away from home (FAFH). FAFH can be thought of as food obtained or consumed at fast food establishment (which accounts for the largest percentage of FAFH), full-service restaurants and other (catered affairs, food trucks and vending machines) establishments. Quick Service Restaurants (QSRs) have driven the growth of FAFH in the last two decades as restaurants such as Chipotle Mexican Grill and Panera Bread provide facets of both fast food (counter service) and full-service restaurants (perceived ambiance and food quality). FAFH accounts for one-half of Americans' food budgets and Americans' share of energy intake from FAFH is 34%, double that from 1978 (United States Department of Agriculture, 2018). Nonetheless, it is difficult to model travel patterns attached to the consumption of FAFH, which may or may not be healthy. While spending patterns are available at the block group level, controlling for it within the confines of this study falls outside of the scope of this research.

As applied explicitly to GIS applications related to the quality of spatial food environment data, work has proliferated as research in the spatial analysis and representation of the food environment has increased and a need has arisen to answer questions about the validity of data on which decisions are made. (Forysth et al. (2010) understood these challenges, which include the reliability and validity of data (proper addresses and classifications of stores) as well detail and completeness (enough information is stored that can be useful in food environment analysis). Wilkins et al. (2017) further expounded on these dimensions to include the quality of geocoding processes, the definition of food outlet constructs (what is the definition of healthy, use of proprietary codes, etc.) and ways to measure access and via a reportable standard called Geo-FERN (Food Environment Reporting). Comprehensive studies (Liese et al., 2010; Auchincloss et al., 2012) have explored the quality of large spatial databases purchased from independent sources among and between disparate datasets and providers which serve as the basis for retail businesses. Larger-scale studies (Rummo et al., 2015; Han et al., 2012; Hosler & Dharssi, 2010; Mendez et al., 2016) were performed for Durham, Chicago, Albany and Pittsburgh respectively. All cited some degree of difference between different CAB databases such as InfoUSA, Dunn and Bradstreet, TDLinx, as well as field-based and automated methods, noting that caution must be taken when using CAB databases. Powell et al. (2011) research reinforced the idea of uncertainty absorption within this narrow focus (validity of GIS data in measuring the food environment), highlighting the reconciliation that must be made between the sheer number of data sources provided by CAB databases, the time needed for field verification and the need for high-quality data. While exploring these differences falls outside of the scope of this research, they must be noted.

Results of drive-time calculations agglomerated within enumeration units such as census block groups can potentially be misleading because of the Modifiable Areal Unit Problem (MAUP) given the varying sizes and shapes of enumerations units to which data are grouped. Research (Wieczorek et al., 2011) showed the sizes and shapes of enumeration units do in fact affect the patterns of phenomenon based on the way in which they are grouped, whether intentionally or unintentionally. While this research's end-results focuses solely on census block groups, results may be different if data are grouped within census tracts or even zip codes. Using the techniques described previously, data were grouped into 119 census tracts and 24 zip codes. Patterns highlighted at the census block group mimicked those patterns at the coarser tract and zip code scale.

Lastly, while the locations of sources are dictated by explicit locations or centroids in some scenarios (Methods 1 through 4), others have some degree of randomness attached to them. In Methods 7 through 11, random points or parcel locations within the study area or strata are used to model source locations. Geostatistical tools such as Nearest Neighbor Analysis can truly dictate the degree of randomness between and amongst these points and python programming solutions combined with modeling tools can be used to run a Monte Carlo simulation, ensuring randomness for each of these methods over many different iterations of random source location placement. In Methods 5 and 6, the fishnet is placed at designated intervals starting from the bottom left/southwestern corner of the map extent, the minimum latitudes and longitudes of geographic boundaries for Guilford County. If these points were placed starting from the upper right/northeastern corner of the map, they may be in slightly different locations about the study area. Further research would be required to assess to what degree these differences may or may not affect outcomes.

5. Conclusions

While the concepts 'food desert' and 'food swamp' have many theoretical definitions, they have practical and applied applications. They exist in the real world and people have an innate understanding of them. Computationally, they represent a combination of availability, or lack thereof, to food sources (healthy and/or unhealthy) and usually a socio-economic component such as income or poverty. GIS tools are being integrated into the study of food deserts and food swamps to map the extent of these phenomena and further explore policy-driven solutions (Liese et al., 2014; Shannon, 2015; Story et al., 2008) in order to improve community health. In this study, we explicitly use GIS functionality to measure food availability, one of the fundamental pillars of food insecurity, as a function of drive-time and drive-distance between a source and destination in and around Guilford County, North Carolina, with an end goal of determining how source locations can be modeled or sampled from a larger database of more than 177,000 possible sources which represent all Guilford County residential locations. Traditional desktop computing solutions have found computing these 177,000 routes resource-intensive, especially within the Esri software environment. This study explores different sampling methods from this population and statistical analysis to ensure results when agglomerated at the block group level are consistent with results when the entire population of 177,000 sources are used and all 177,000 routes are calculated.

A complete and comprehensive evaluation of the food environment at a local scale requires qualitative techniques and quantitative assessment. A GIS has been used as a powerful tool to make these quantitative assessments which require spatial calculation such as proximity (drive-time or drive-distance) to the nearest resource(s) and density of resources within enumerations units such as census block groups, census tracts and zip codes. In addition, spatial clustering of healthy food outlets, spending information or health outcomes with statistical significance can be measured using tools such as Local Moran's I and Getis-Ord. Qualitative variables such as personal dietary preferences, one's perception of the food environment and one's understanding of relative distance are difficult to quantify. In addition, overly-difficult-to-quantify metrics such as Food Away From Home (FAFH), people selecting more distant supermarkets than the closest to their home and the purchasing of food in conjunction with other activities (work, recreation, etc.) are difficult to quantify and more difficult to map within the confines of this research. As a result, this research focuses solely on the availability, or proximity, of source locations to the nearest healthy food option, the basis for most food desert and food swamp research as well as the popular USDA Food Access Atlas.

In evaluating all sampling techniques, all 11 had some level of agreement when grouped at the block group level when compared to all Guilford residential parcels as highlighted in Table 4. However, three of them (Population-Weighted Block Group Centroids, Block Centroids and Random Points (4 per block group)) rejected all four equivalence tests for dissimilarity and returned p-values that satisfied research-grade ($\alpha = 0.05$) and less-stringent significance levels ($\alpha = 0.1$ or $\alpha = 0.2$), highlighting similarity between the Guilford County residential parcels and these three techniques. These were denoted as the strongest sampling techniques and decisions rating the other techniques as strong, fairly strong and weak were made based on these criteria. This supports work by Berke and Shi (2009), as well as Henry and Boscoe (2009) which highlighted how population-based measures such as the population centroid were both adequate and better than geometry-based methods. As a result, this research suggests population-weighted block group centroids (Method #3) best simulate true residential locations in agreement with this prior research, its alignment with population-based measures and a 600-fold decrease in the number of calculations required. While two other techniques (block centroids and four random points per block group) did satisfy all aforementioned criteria just as well, they also require more calculations (28 and 4 times more, respectively) than the population-weighted block groups.

Calculations using these source locations can be used as part of unit-based metrics used by the USDA Food Access Atlas which uses distance in miles or integrated into unitless metrics or ratios (Zenk et al., 2014; Clary et al., 2015; Mason et al., 2013; Mulrooney, McGinn, et al., 2017). However, care must be used when implementing unit-based metrics. In applying unit-based metrics, the size of study areas have ranged from the county scale (Zenk 2005), such as this one, to the multi-county (Murrell & Jones, 2020) and state (Mulangu & Clark, 2012) scale. While the study areas come in all shapes and sizes, care must be taken to ensure these unit-based metrics are applied correctly. The USDA Food Access Atlas ensures to explicitly differentiate between poor availability in urban areas versus poor availability in rural regions using a distance qualifier while Mulangu and Clark (2012) qualified their drive-time calculations between rural and urban areas.

The mapping of food availability can be done a number of different ways and the use of sources, from which distance and drive-time to destinations are computed, can be estimated as actual street addresses, blocks, block groups, random points and along grids, as is the case of the USDA Food Access Atlas. The research community accepts the results and accompanying maps at face value with little question as to the how values were computed and at what scale they were computed. Since managing and making drive-time calculations with all sources can be resource-intensive, this research is an attempt to determine how well various sampling techniques align with results from using calculations where all parcels are used as sources. In this paper, it was found that:

- Population-weighted centroids, block centroids and a stratified-random technique grouped at the block group scale aligned best with results from all Guilford County residential parcels.
- All three of these techniques satisfied t-tests of two differences as well as two separate tests of dissimilarity (Two One-Sided Test for

Equivalence and Paired T-Test for Equivalence) for both drive-distance and drive-time.

- Of these three, population-weighted centroids require less calculations because there are less source points (291) compared to block centroids (8,183) and the stratified-random (1,168) techniques.
- 1000 random points was the weakest of these methods because it did not satisfy a *t*-test of two differences and passed 3 of 4 tests for dissimilarity.

Nonetheless, this holistic and critical look at the source data and techniques used in the quantitative assessment of the food environment can serve as an impetus for larger work in the policy and subsequent remediation of food deserts and food swamps at a scale necessary to affect neighborhood-scale health outcomes and other related quality of life measures.

Author statement

Timothy Mulrooney: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Data curation, Resources, Richard Foster: Conceptualization, Software, Formal analysis, Visualization, Methodology, Writing – original draft, Writing – review & editing, Data curation, Manoj Jha: Project administration, Supervision, Funding acquisition, Data curation, Resources, Leila Hashemi Beni: Project administration, Supervision, Methodology, Lyubov Kurkalova: Conceptualization, Project administration, Supervision, Chyi Lyi Liang: Conceptualization, Project administration, Supervision, Writing – review & editing, Haoran Miao: Conceptualization, Supervision, Methodology, Visualization, Greg Monty: Project administration, Supervision

Acknowledgement

This material is supported by the National Science Foundation under Grant No. 1824949. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The project was also supported by the Agricultural and Food Research Initiative Competitive Program of the USDA National Institute of Food and Agriculture (NIFA), grant number 2021-67021-34152.

References

Abousaeidi, M., Fauzi, R., & Muhamad, R. Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. Saudi Journal of Biological Sciences, 23(5), 555-564.

Auchincloss, A., Moore, K., Moore, L., & Diez-Roux, A. (2012). Improving retrospective characterization of the food environment for a larger region in the United State during a historic time period. *Health & Place, 18*(6), 1341–1347.

Austin, S. B., Melly, S. J., Sanchez, B. N., Patel, A., Buka, S., & Gortmaker, S. L. (2005). Clustering of fast-food restaurants around schools: A novel application of spatial statistics to the study of food environments. *American Journal of Public Health, 95*, 1575–1581.

Baker, E. A., Schootman, M., Barnidge, E., et al. (2006). The role of race and poverty in access to foods that enable individuals to adhere to dietary guidelines. *Preventing Chronic Disease, 3*, A76.

Berger, R. L., & Hsu, J. C. (1996). Bio-equivalence trials, intersection-union test and equivalence confidence sets. *Statistical Science, 11*, 283–302.

Berke, E. M., & Shi, X. (2009). Computing travel time when the exact address is unknown: A comparison of point and polygon ZIP code approximation methods. *International Journal of Health Geographics, 8*, 23. https://doi.org/10.1186/1476-072X-8-23

Block, J. P., Scribner, R. A., & DeSalvo, K. B. (2004). Fast food, race/ethnicity, and income: A geographic analysis. *American Journal of Preventative Medicine, 27*, 211–217.

Bodor, J. N., Rose, D., Farley, T. A., Swalm, C., & Scott, S. K. (2008). Neighbourhood fruit and vegetable availability and consumption: The role of small food stores in an urban environment. *Public Health Nutrition, 11*, 413–420.

Brindle-Fitzpatrick, S. (2015). Food deserts or food swamps?: A mixed-methods study of local food environments in a Mexican city. *Social Science & Medicine, 142*, 202–213. http://doi:0.1016/j.socscimed.2015.08.010.

Burdette, H. L., & Whitaker, R. C. (2004). Neighborhood playgrounds, fast food restaurants, and crime: Relationships to overweight in low-income preschool children. *Preventative Medicine, 38*, 57–63.

California Center for Public Health Advocacy. (2007). *Searching for healthy food: The food landscape in California cities and counties*. Retrieved from http://www.publichealthadvocacy.org/RFEI/expanded percent20methods.pdf.

Centers for Disease Control and Prevention. (2011). *The state indicator report on fruits and vegetables*. U.S. Department of Health and Human Services. Retrieved from https://www.cdc.gov/obesity/downloads/census-tract-level-state-maps-mrfei_TAG508.pdf.

Chrisinger, B. (2016). A mixed-method assessment of a New supermarket in a food desert: Contributions to everyday life and health. *Journal of Urban Health, 93*(3), 425–437.

Clary, C. M., Ramos, Y., Shareck, M., & Kestens, Y. (2015). Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. *Prevantative Medicine (Baltimore), 71*, 83–87.

Cooksey-Stowers, K., Schwartz, M. B., & Brownell, K. D. (2017). Food swamps predict obesity rates better than food deserts in the United States. *International Journal of Environmental Research and Public Health, 14*, 1366.

Cribbie, R. A., Gruman, J. A., & Arpin-Cribbie, C. A. (2004). Recommendations for applying tests of equivalence. *Journal of Clinical Psychology, 60*, 1–10.

Dijkstra, E. W. (1959). A note of two problems in connexion with graphs. *Numberische Mathematik, 1*, 269–297.

Donkin, A. J., Dowler, E. A., Stevenson, S. J., & Turner, S. A. (1999). Mapping access to food at a local level. *British Food Journal, 101*(7), 554.

Environmental Systems Research Institute (Esri). (2010). Network Analyst tutorial. Retrieved from http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/network-analyst-tutorial.pdf.

Fischer, M. M. (2004). In D. A. Hensher, K. J. Button, K. E. Haynes, & P. R. Stopher (Eds.), *Handbook of transport geography and spatial Systems: Vol. 5. GIS and network analysis* (pp. 391–408). Emerald Group Publishing Limited. https://doi.org/10.1108/9781615832538-022.

Food Marketing Institute. (2016). *U.S. Grocery shopping trends, 2016*. Arlington, VA: Food Marketing Institute.

Forysth, A., Lytle, L., & Van Riper, D. (2010). Finding food: Issues and challenges in using geographic information Systems to measure food access. *Journal of Transport and Land Use, 3*(1), 43–65.

Gallagher, M. (2006). *Examining the impact of food deserts on public health in Chicago*. Mari Gallagher Research and Consulting Group. Retrieved from http://www.marigallagher.com/site_media/dynamic/project_files/1_ChicagoFoodDesertReport-Full_.pdf.

Giang, T., Karpyn, A., Laurison, H. B., Hillier, A., & Perry, R. D. (2008). Closing the grocery gap in underserved communities: The creation of the Pennsylvania fresh food financing initiative. *Journal of Public Health Management and Practice, 14*(3), 272–279.

Guy, C. M., Clarke, G., & Eyre, H. (2004). Food retail change and the growth of food deserts: A case study of cardiff. *International Journal of Retail & Distribution Management, 32*, 72–88.

Hallett, L., & McDermott, D. (2011). Quantifying the extent and cost of food deserts in Lawrence, Kansas, USA. *Applied Geography, 31*, 1210–1215.

Han, E., Powell, L., Zenk, S., Rimkus, L., Ohri-Vachaspati, P., & Chaloupka, F. (2012). Classification bias in commercial business lists for retail food stores in the U.S. *International Journal of Behavioral Nutrition and Physical Activity, 46*. https://doi.org/10.1186/1479-5868-9-46

Henry, K. A., & Boscoe, F. P. (2008). Estimating the accuracy of geographical imputation. *International Journal of Health Geographics, 7*, 3. https://doi:10.1186/1476-072X-7-3.

Hillson, R., et al. (2020). *Stratified sampling of neighborhood sections for population estimation: A case study of bo city, Sierra Leone* (Vol. 10). Gale Academic OneFile Accessed . (Accessed 17 June 2020).

Hosler, A., & Dharssi, A. (2010). Identifying retail food stores to evaluate the food environment. *American Journal of Preventive Medicine, 39*(1), 41–44.

Huber, S., & Rust, C. (2016). Calculate travel time and distance with OpenStreetMap data using the open source routing machine (OSRM). *Stata Journal, StataCorp LP, 16*(2), 416–423.

International Food Information Council Foundation. (2018). *Food and health Survey*. Retrieved from https://foodinsight.org/wp-content/uploads/2018/05/2018-FHS-Report-FINAL.pdf.

Jago, R., Baranowski, T., & Baranowski, J. C. (2007). Distance to food stores and adolescent male fruit and vegetable consumption: Mediation effects. *International Journal of Behavioral Nutrition and Physical Activity, 4*, 35.

Jiao, J., Moudon, A., Ulmer, J., Hurvitz, P., & Drewnowski, J. (2012). How to identify food deserts: Measuring physical and economic access to supermarkets in King County, Washington. *American Journal of Public Health, 102*(10), 32–39.

Lewis, L. B., Sloane, D. C., Nascimento, L. M., Diamant, A. L., Guinyard, J. J., Yancey, A. K., et al. (2005). African Americans' access to healthy food options in South Los Angeles restaurants. *Research and Practice, 95*(4), 668–673.

Liese, A., Clabianchi, N., Lamichhane, A., Barnes, T., Hibbert, J., Porter, D., Nichols, M., & Lawson, A. (2010). Validation of 3 food outlet databases: Completeness and geospatial accuracy in rural and urban food environments. *American Journal of Epidemiology, 172*(11), 1324–1333.

Liese, A. D., Hibbert, J. D., Ma, X., Bell, B. A., & Battersby, S. E. (2014). Where are the food deserts? An evaluation of policy-relevant measures of community food access in South Carolina. *Journal of Hunger & Environmental Nutrition, 9*(1), 16–32.

Liu, G. C., Wilson, J. S., Qi, R., & Ying, J. (2007). Green neighborhoods, food retail and childhood overweight: Differences by population density. *American Journal of Health Promotion, 21*, 317–325.

Lovett, A., Haynes, R., Sunnenberg, G., & Gale, S. (2002). Car travel time and accessibility by bus to general practitioner services: A study using patient registers and GIS. *Social Science & Medicine, 55*(1), 97–111.

Lucan, S. C. (2015). Concerning limitations of food-environment research: A narrative review and commentary framed around obesity and diet-related diseases in youth. *Journal of the Academy of Nutrition and Dietetics, 115*, 205–212.

Lucan, S. C., Maroko, A. R., Seitchik, J. L., Yoon, D. H., Sperry, L. E., & Schechter, C. B. (2018). Unexpected neighborhood sources of food and drink: Implications for research and community health. *American Journal of Preventative Medicine, 55*(2), e29–38.

Mason, K. E., Bentley, R. J., & Kavanagh, A. M (2013). Fruit and vegetable purchasing and the relative density of healthy and unhealthy food sources: Evidence from an Australian multilevel city. *Journal of Epidemiology & Community Health, 3*, 231–236.

Melhorn, K., & Sanders, P. (2008). *Chapter 10. Shortest paths. Algorithms and data structure: The basic toolbox*. Springer.

Mendez, D., Kim, K., Hardaway, C., & Fabio, A. (2016). Neighborhood racial and socioeconomic disparities in the food and alcohol environment: Are there difference by commercial data sources? *Journal of Racial and Ethnic Health Disparities, 3*, 108–116.

Morland, L., & Filomena, S. (2007). Disparities in the availability of fruits and vegetables between racially segregated urban neighbourhoods. *Public Health Nutrition, 10*, 1481–1489.

Morland, K., Wing, S., Diez Roux, A. V., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine, 22*(1), 23–29.

Mulangu, F., & Clark, J. (2012). Identifying and measuring food deserts in rural Ohio. *Journal of Extension, 50*(3), 3FEA6. Retrieved fromhttp://www.joe.org/joe/2012june/a6.php.

Mulrooney, T., Beratan, K., McGinn, B., & Branch, B. (2017a). A comparison of raster-based travel time surfaces against vector-based network calculations as applied in the study of rural food insecurity. *Applied Geography, 78*, 12–21.

Mulrooney, T., McGinn, C., Branch, B., Madumere, C., & Ifediora, B. (2017b). *A* New raster-based metric to measure relative food availability in rural areas: A case study in southeastern North Carolina. *Southeastern Geographer, 57*(2), 151–178.

Murrell, A., & Jones, R. (2020). Measuring food insecurity using abundance index: Implications for economic, health and social well-being. *International Journal of Environmental Research and Public Health, 17*(7), 2434. bb.

North Carolina Department of Transportation. (2020a). *NCDOT maps and apps*. Retrieved from http://ncdot.maps.arcgis.com/home/index.html.

North Carolina Department of Transportation. (2020b). *GIS Data Layers*. Retrieved from https://connect.ncdot.gov/resources/gis/Pages/GIS-Data-Layers.aspx.

Pearce, J., Hiscock, R., Blakely, T., et al. (2008). The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption. *Journal of Epidemiology & Community Health, 62*, 198–201.

Powell, L. M., Han, E., Zenk, S. N., et al. (2011). Field validation of secondary commercial data sources on the retail food outlet environment in the United States. *Health & Place, 5*, 1122–1131.

Raja, S., Born, B., & Russell, J. K. (2008). A planners guide to community and regional food planning: Transforming food environments, facilitating healthy eating. *Planning advisory service report No. 554*. Chicago: American Planning Association.

Rummo, P., Albrecht, S., & Gordon-Larsen, P. (2015). Field validation of food outlet databases: The latino food environment in North Carolina, USA. *Public Health Nutrition, 18*(6), 977–982.

Sahar, L., Foster, S. L., Sherman, R. L., Henry, K. A., Goldberg, D. W., Stinchcomb, D. G., & Bauer, J. E. (2019). GIScience and cancer: State of the art and trends for cancer surveillance and epidemiology. *Cancer, 125*(15), 2544–2560.

Seaman, M. A., & Serlin, R. C. (1998). Equivalence confidence intervals for two-group comparisons of means. *Psychological Methods, 3*, 403–411.

Shannon, J. (2015). Rethinking food deserts using mixed-methods GIS. *Cityscape, 17*(1), 85–96.

Sharkey, R., & Horel, S. (2008). Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. *Journal of Nutrition, 30*(5), 620–627.

Sisiopiku, V. P., & Barbour, N. (2014). Use of GIS spatial analysis to identify food deserts in the state of Alabama. *Athens Journal of Health, 1*(2), 91–103.

Smoyer-Tomic, K. E., Spence, J. C., Raine, K. D., et al. (2008). The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets. *Health & Place, 14*, 740–754.

Story, M., Kaphingst, K. M., Robinson-O'Brien, R., & Glanz, K. (2008). Creating healthy food and eating environments: Policy and environmental approaches. *Annual Review of Public Health, 29*, 253–272.

Thornton, L. E., et al. (2011). Using geographic information Systems (GIS) to assess the role of the built environment in influencing obesity: A glossary. *International Journal of Behavioral Nutrition and Physical Activity, 8*, 71.

Thornton, L. E., & Kavanagh, A. M. (2012). Association between fast food purchasing and the local food environment. *Nutrition & Diabetes, 2*, e53.

Thornton, L. E., Lamb, K. E., & White, S. R. (2020). The use and misuse of ratio and proportion exposure measures in food environment research. *International Journal of Behavioral Nutrition and Physical Activity, 17*, 118. https://doi.org/10.1186/s12966-020-01019-1

United States Census Bureau. (2010). *2010 census urban and rural classification and urban area criteria*. Retrieved from https://www.census.gov/geo/reference/ua/urban-rural-2010.html.

United States Department of Agriculture. (2018). America's eating habits: Food away From Home. Washington D.C *Economic Research Service*. Retrieved fromhttps://www.ers.usda.gov/webdocs/publications/90228/eib-196.pdf.

United States Department of Agriculture. (2020). *Food access Atlas documentation*. Retrieved from https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/.

Van Hoesen, J. G., Bunkley, B., & Currier, C. (2013). A GIS-based methodology toward refining the concept of rural food deserts: A case study from rutland county, Vermont. *Journal of Agriculture, Food Systems, and Community Development, 3*(2), 61–76.

Ver Ploeg, M., Breneman, V., Farrigan, T., Hamrick, K., Hopkins, D., et al. (2009). *Access to affordable and nutritious food–measuring and understanding food deserts and their consequences: Report to congress*. Washington, D.C.: United States Department of Agriculture, Economic Research Service.

Wang, M. C., Kim, S., Gonzalez, A. A., MacLeod, K. E., & Winkleby, M. (2007). Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *Journal of Epidemiology & Community Health, 61*, 491–498.

Wellek, S. (2003). *Testing statistical hypotheses of equivalence*. New York: Chapman &Hall/CRC.

Wilkins, E., et al. (2019). Methods of measuring associations between the Retail Food Environment and weight status: Importance of classifications and metrics. *SSM - Population Health, 8*, 100404.

Wilkins, E., Morris, A., Radley, D., & Griffith, C. (2017). Using Geographic Information Systems to measure retail food environments: Discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). *Health & Place, 44*, 100–117.

Winkler, E., Turrell, G., & Patterson, C. (2006). *Does living in a disadvantaged area mean fewer opportunities to purchase fresh fruit and vegetables in the area? Findings from the*.

Winn, M. T. (2014). *A road network shortest path analysis: Applying time-varying travel-time costs for emergency response vehicle routing*. Davis County, Utah: A Thesis Presented to the Department of Humanities and Social Sciences in Candidacy for the Degree of Master of Science.

Wood, W. F. (1955). Use of Stratified random samples in land-use study. *Annals of the Association of American Geographers, 45*, 350–367.

Wrigley, N. (2002). 'Food deserts' in British cities: Policy context and research priorities. *Urban Studies, 39*(11), 2029–2040.

Zandbergen, P. A., & Hart, T. C. (2009). Geocoding accuracy considerations in determining residency restrictions for sex offenders. *Criminal Justice Policy Review, 20*(1), 62–90. https://doi.org/10.1177/0887403408323690

Zenk, S. N., & Powell, L. M. (2008). US secondary schools and food outlets. *Health & Place, 14*, 336–346.

Zenk, S. N., Schulz, A. J., Israel, B. A., James, S. A., Bao, S., & Wilson, M. L. (2005). Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in Metropolitan Detroit. *American Journal of Public Health, 95*(4), 660–667.

Zenk, S., Schulz, A., Israel, B., Mentz, G., Miranda, P., Opperman, A., & Odoms-Young, A. (2014). Food shopping behaviors and exposure to discrimination. *Public Health Nutrition, 17*, 1167–1176.