Potentials and Limitations of Large-scale Mobile Device Location Data for Food Access Analysis

Duanya Lyu [a], Luyu Liu [b, *], Catherine Campbell [c], Yuxuan Zhang [d], Xiang Yan [a]

[a] *Department of Civil & Coastal Engineering, University of Florida, Gainesville FL, United States*

[b] *Department of Geosciences, Auburn University, Auburn AL, United States*

[c] *Department of Family, Youth and Community Sciences, University of Florida, Gainesville FL, United States*

[d] *Department of Computer & Information Sciences & Engineering, University of Florida, Gainesville FL, United States*

*Corresponding author: Luyu Liu, email: luyuliu@auburn.edu

Potentials and Limitations of Large-scale Mobile Device Location Data for Food Access Analysis

Abstract:

Food insecurity is a major global challenge. Understanding household food access is crucial for developing strategies to combat food insecurity and obesity-related conditions like diabetes and cardiovascular diseases. Traditionally, researchers used surveys or specific location tracking data, which, while valuable, often had small sample sizes leading to limited generalizability. Large-scale mobile device GPS data offer detailed, high-resolution information on human mobility, allowing for a more comprehensive analysis of food access patterns. This paper assesses the potential and limitations of using mobile device location data, utilizing a Terabyte-level GPS dataset of 286 million records in Jacksonville, Florida. The results indicate that mobile GPS data can effectively capture households' food access activities and reveal richer spatiotemporal patterns. We validate our findings by comparing them with traditional approaches and conducting sensitivity analyses. The study highlights that results are sensitive to algorithm design and parameter settings, emphasizing the need for thorough validation. Our research underscores the importance of using GPS data for food access studies and informs policy discussions to improve food security.

Keywords: Food access; Food acquisition behavior; Food deserts; Mobile device location data; Human mobility data.

1. Background

Food insecurity, which means the lack of stable access to sufficient, safe, and nutritious food for a healthy, active life (Simelane and Worth 2020), is a major challenge faced by many households in the US. In 2022, the Department of Agriculture reported that 12.8% of U.S. households experienced food insecurity (Rabbitt *et al.* 2023). Meanwhile, prior studies have established strong associations between food insecurity and many critical health conditions, such as obesity, diabetes, and cardiovascular diseases (Bodor et al. 2010, Berkowitz et al. 2018). Food insecurity is a multi-dimensional issue. The Food and Nutrition Security (FNS) theory have outlined four key pillars of food security: availability, access, utilization, and stability, where the first three forming a consequential relationship, and stability adding a temporal dimension (Simelane and Worth 2020). As a crucial link in this framework, *food access* is widely regarded as an important concept for understanding and tackling food insecurity (Rabbitt *et al.* 2023).

Much of existing food access research focuses on the *supply* side, where measuring food *accessibility* is a critical topic. Researchers have used various datasets, such as point-of-interest data and satellite imagery, and developed both spatial and aspatial measurements (Larsen and Gilliland 2008, Nguyen, Hoang, *et al.* 2020). Recent years have witnessed an emerging number of studies on the *demand* side, shedding light into the household food acquisition behavior as well as the financial, mobility, and dietary considerations underlying the observed behavior (Simelane and Worth 2020). Commonly used approaches include gathering store data (e.g. transaction data or loyalty card information) or engaging individuals through surveys (Todd and Scharadin 2016) and recoding their activity-travel diaries (Wray *et al.* 2023). The studies offer detailed insights into household food access patterns as well as collecting rich purchase and attitudinal information. However, typically relying on *primary* data collection (Rabianski 2003), these studies tend to have small sample sizes, contain sampling bias, and face challenges with tracking behavioral differences across large areas and over time.

The recent availability of large-scale mobile device location GPS data offers a new approach to study human mobility (Kwan 2016). These passively collected *secondary* data can capture the movements of thousands of individuals over extended periods of weeks or months, with spatial and temporal resolutions of meters and seconds. Researchers have applied these data to study a variety of human mobility topics such as commuting patterns and evacuation behaviors (Kalter *et al.* 2021, Zhao *et al.* 2022, Horn *et al.* 2023). Large-scale GPS data holds promises for advancing our understanding on food access (Cetateanu and Jones 2016). However, the potentials and limitations of utilizing these data for food access analysis are not yet systematically studied (Chang *et al.* 2022, Horn *et al.* 2023, Xu *et al.* 2023).

To address this gap, this paper explores the use of a large-scale, individual-level, longitudinal secondary GPS dataset in food access analysis. We infer food acquisition activities from the data, analyze patterns and evaluate the findings. Specifically, we aim to: 1) assess the feasibility of using passively collected GPS data to replicate traditional food acquisition analyses and obtain common metrics; 2) identify the novel insights that the dataset can provide; 3) and evaluate the robustness of the findings generated from the dataset.

The paper is structured as follows. We begin with an overview of existing food access studies, examining their data sources and insights. Next, we introduce mobile location GPS data and its current applications in the food access literature. We then proceed to our case study, where we extract and analyze food-related trips from a large-scale mobile location GPS dataset of 1.5 months in Jacksonville, Florida. We compare our findings with those from existing studies with the traditional approaches.

Additionally, we conduct sensitivity analysis to evaluate the robustness of this mobile location GPS-based approach. Finally, we discuss the insights gained and the challenges encountered throughout this process.

2. Literature Review

2.1 Food Insecurity and Food Access

Figure 1 below shows the four pillars of food security established in the food and nutrition security theory: availability, access, utilization and stability (Simelane and Worth 2020). First, food *availability* is essential; then, stable quantity, quality of food may lead to food *access*; ultimately, stable food access, along with *utilization* that meets caloric and nutritional needs, contributes to achieving food and nutrition security. The four pillars are interconnected and food access is a crucial link.

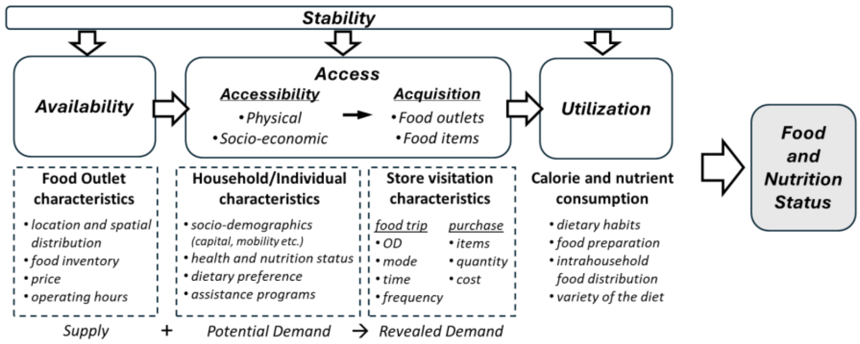

Figure 1 Food access in food and nutrition security theory framework. Generated with reference to *Simelane and Worth, 2020*

Food access research may focus on *potential* access (i.e. food *accessibility*) or *realized* access (i.e. food *acquisition*) (Khan 1992, Simelane and Worth 2020, Tadesse *et al.* 2020). Studies on *potential* food access primarily focus on the supply side. Researchers often model food outlet availability using datasets such as Point of Interest (POI) data (Larsen and Gilliland 2008) or satellite imagery (e.g., agricultural land use that informs food production and harvest patterns) (Nguyen, et al. 2020). Much work focuses on the development of *accessibility* measures, including measurements of “food deserts” (Berkowitz *et al.* 2018). By contrast, studies on *realized* food access examine the demand side by analyzing food *acquisition* behaviors. These analyses provide insights into how individuals perceive and interact with their food environment (Dubowitz et al. 2015). Specifically, studies on food acquisition behaviors often focus on two key aspects: trip characteristics and purchase behaviors (Hillier *et al.* 2017). Trip characteristics include where people shop, travel distances, travel modes, and trip duration. Studies of purchase behaviors examine the types of items bought and the amounts spent.

Next, we discuss how researchers study these behaviors, focusing on their data sources, research objectives, and the methodologies used or developed. We begin with reviewing traditional data sources and then the increasingly popular large-scale mobile device location data.

2.2 Traditional Approaches for Food Access Analysis

Traditional research methods used to understand household food access patterns include questionnaire surveys, interviews, and focus groups. Questionnaire surveys, such as the USDA National Household Food Acquisition and Purchase Survey (FoodAPS) or American Time Use Survey (ATUS), are widely used in food access studies to collect detailed data on where people shop, how much they spend, and the types of food they purchase (Coleman-Jensen et al. 2019). The structured format ensures consistent data collection, enabling comparisons across demographics and socioeconomic groups, and facilitating longitudinal studies on behavioral change (Anekwe and Zeballos 2019). Interviews and focus groups are effective for exploring participants' nuanced behaviors and perceptions of food quality, accessibility, and affordability. Moreover, they can reveal the cultural, social, and economic contexts that shape food acquisition behaviors, uncovering motivations and barriers influencing purchasing decisions. However, studies using these approaches tend to have limited sample sizes and undersample some groups or areas (Hillier et al. 2017). Another issue with traditional approaches is that the spatial and temporal information to be collected relies on manual recall, leading to potential inaccuracies and sparse data. This limitation often confines researchers to focus on a limited set of trip origins (e.g. work or home) or destinations (e.g. primary stores) in their analyses (Hillier *et al.* 2017).

Commented [XY1]: Not sure why you think surveys are not flexible and interviews/focus groups are. I do not think this is appropriate

2.3 Use of GPS Data for Food Access Analysis

Research in human mobility analysis has demonstrated GPS data's capability to track locations with high spatial and temporal resolution (Chen *et al.* 2016). Recently, food access researchers have utilized GPS data to explore food environment characteristics and their connections with space, time, and behaviors (Emish *et al.* 2023).

2.3.1 Primary GPS data

Researchers initially used GPS data to augment survey studies, where GPS tracking was integrated to gather more precise spatiotemporal movement information (Zenk *et al.* 2011). We refer to this type of data as *primary* GPS data. These studies typically involve distributing GPS tracking devices to participants for the recording of geo-tagged surveys (Elliston *et al.* 2020), geo-fenced visits (Wray *et al.* 2023), or, in more frequent tracking scenarios, movement trajectories (Zenk *et al.* 2011).

Including primary GPS data into surveys allows researchers to perform detailed spatiotemporal analyses and reconstruct individuals' food access activities. For instance, researchers can examine food exposure based on activity spaces, rather than home or workplace (Elliston *et al.* 2020); they can also calculate time-weighted exposure (Liu *et al.* 2020). More importantly, by using these new data, researchers can compare between *supply*, *potential* access, and *realized* access (see definitions in Section 2.1), leading to new behavioral insights. For example, Sadler and Gilliland (2015) found that proxy measurements consistently underestimate exposure to junk foods compared to objectively-derived GPS tracks ~~(Sadler and Gilliland 2015)~~. Elliston *et al.* (2020) noted a weak correlation between objective calculated and subjectively perceived outlet counts ~~(Elliston *et al.* 2020)~~. In addition, Liu *et al.* (2020) observed significant positive association between fast food meals consumed and time-weighted number of fast food outlets exposed but not with ratio exposed ~~(Liu et al. 2020)~~.

However, collecting primary GPS data is resource-intensive and so only small sample sizes are achieved, preventing these studies from engaging representative participant groups. Cetateanu and Jones (2016) and Siddiqui *et al.* (2024) reviewed the papers on GPS and food environment exposure and reported sample sizes ranging from 12 to 654 individuals. Moreover, the collected GPS data may suffer from Selective Daily Mobility Bias (SDMB). SDMB is a common issue in behavior research that relies on tracking movement, as individuals' mobility patterns are not random but shaped by their routines, preferences, and behaviors (Li et al. 2023). The awareness of carrying monitoring devices can also lead to increased consciousness of actions and potential behavioral changes over the study period (Zhang et al. 2021).

2.3.2 Secondary GPS Data

With the widespread use of location-enabled mobile devices, large-scale high-resolution individual-level GPS data are increasingly common (Chen et al., 2016). We refer to them as *secondary* GPS data since they are often collected by data vendors and social networking service platforms passively and not for specific purposes (Zhao *et al.* 2016). Secondary GPS data have become increasingly important in food access analysis, providing valuable behavioral insights.

Secondary GPS data are available to researchers in aggregated and disaggregated forms. *Aggregated* data are pre-processed datasets that capture human activities across geographic units such as census tracts and block groups. Compared to survey data, they offer much higher spatial resolution, larger sample sizes, and support more flexible spatiotemporal analyses. For instance, studies have observed links between visits to food retailers with diet-related diseases (Xie et al. 2023), grocery visits and regional socio-demographics (Smith et al. 2023), and alcohol outlet visits and domestic violence (Hu et al. 2021, Chang et al. 2022). *Disaggregated* data, consisting of raw GPS points at the individual level, offer detailed movement information (Zenk *et al.* 2018). Like primary GPS data, they facilitate detailed activity analyses, allowing for the comparison between supply, potential access, and realized access discussed earlier. However, they benefit from significantly larger population sizes and longer coverage periods. For example, Horn *et al.* (2023) analyzed data from over 240,000 smartphone users to correlate fast food outlet visits with diet-related diseases in Los Angeles County; García Bulle Bueno *et al.* continues Horn *et al.*'s study but focused on food visits and their influence on dietary choices; Xu *et al.* studied grocery shopping patterns over 50 weeks, highlighting spatial disparities in access among minority communities.

While secondary GPS data offer considerable potential, they also have drawbacks. Unlike primary GPS data, the data quality (accuracy, spatiotemporal coverage) may differ across individuals (Li et al. 2023). Also, algorithm design and parameter choices in the analysis of secondary GPS data can significantly impact the study results (Kwan 2016). The uncertainties in algorithm design is further exacerbated by the absence of contextual information about movement activities from the GPS data (Zhao et al. 2022). Moreover, despite larger sample sizes, studies have shown that secondary GPS data tended to under sample disadvantaged groups, which can bias study findings (Li et al. 2023).

Commented [XY2]: Ecological fallacy is not an inherent issue of GPS data. Researchers commit this mistake if they try to apply aggregate-level results to individuals.

2.4 Summary of Literature

Based on the literature review, we can identify several research opportunities. Firstly, there is a limited application of disaggregated secondary GPS data in food access studies compared to traditional methods, especially in the US South with high obesity rate and significant food access problem (Bodor *et al.* 2010). Second, existing applications of the new data primarily focus on specific type of food outlets, while regional food environments encompass a wide array of food outlets and can have different impact on community health and people's visiting patterns are different (Todd and Scharadin 2016, Balagtas *et al.* 2023). For instance, visits to fast food places tend to occur more frequently on weekdays, whereas supermarket more on weekends (East *et al.* 1994, García Bulle Bueno *et al.* 2024). While these distinctions have been explored using traditional approaches, they remain relatively underexplored with the new datasets (Todd and Scharadin 2016, Smith *et al.* 2023, Xu *et al.* 2023). Third, there is limited discussion in the literature regarding the sensitivity of the inferential processes in using secondary GPS data. It is crucial to address this sensitivity to ensure the reliability and accuracy of the findings from the data.

To address these gaps, we conduct a detailed case study in Jacksonville, Florida. We first apply a large-scale, individual-level, secondary GPS dataset to explore its potential in understanding the food access patterns within the study area. This investigation provides insights into the temporal and spatial patterns of food acquisition behavior across different types of food outlets in the study region. We have further conducted sensitivity analyses to test the robustness of study results in regards to algorithm design and parameter choices.

3. Case Study

3.1 Study Area

Our case study area is the City of Jacksonville, Florida, the largest municipality in the state. The city's demographics reveal a complex socioeconomic landscape. According to the American Community Survey (ACS) 2018-2022 five-year estimates (U.S. Census Bureau 2022), the population of 950,203 includes 53.1% White, 30.4% Black, and 11.3% Hispanic or Latino, and with a median age of 36.3 years, a median household income of \$64,138, and 14.8% living in poverty. *Figure 2* illustrates the spatial distribution of the sociodemographic characteristics by tracts. As shown in the figure, urban tracts generally exhibit higher population density, greater percentage of individuals aged 18-39, lower household vehicle ownership, and higher poverty rates. The northwest part of the study area demonstrates a lower percentage of White population.

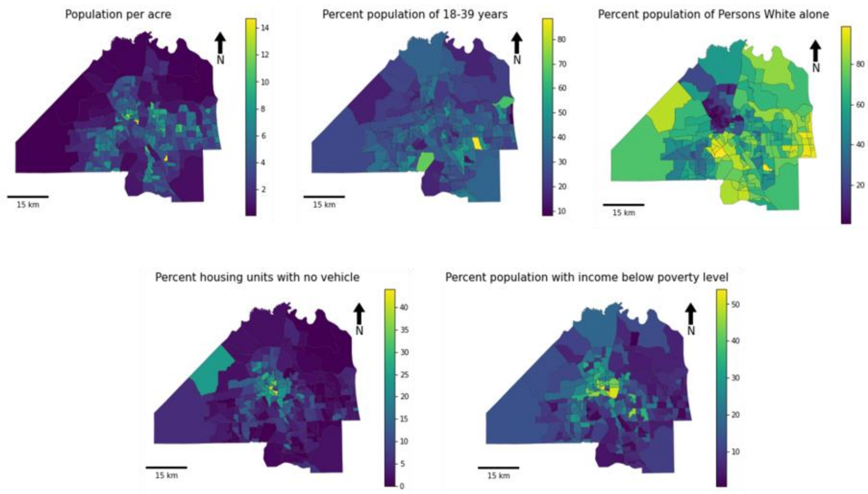

Figure 2 Socio-demographic characteristics of tracts in Jacksonville, Florida

3.2 Data

3.2.1 Mobile Device GPS data

This study uses a large-scale mobile device location dataset from Gravy Analytics, which aggregates data from over 150 million U.S. mobile devices through various apps (Gravy Analytics 2023a). According to the company, it complies with privacy laws, sourcing data only from users who opt in for device identifier and geolocation signal collection, with a 48- to 72-hour processing delay, where no location data is received or processed in real-time (Gravy Analytics 2023b).

The dataset is pre-processed for quality control, with accuracy indicated by a forensic identifier field, which measures GPS *positioning* errors caused by factors like spoofed locations, IP address-derived signals, and abnormal signal density (Xu *et al.* 2022, Gravy Analytics 2023c). For this study, we included only records classified as *High Accuracy* by Gravy Analytics, where GPS positioning errors do not exceed 35 meters. While lower accuracy data may be useful for broader neighborhood analyses or overall trend identification, it is not suitable for pinpointing specific customer visits to particular locations (Gravy Analytics 2023c). This pre-analysis filtering minimizes positioning errors, ensuring more reliable results. (However, inherent limitations in this GPS-based approach still affect accuracy and generalizability, as discussed later.)

The GPS dataset used in this study covers the area of Jacksonville, Florida, and it spans from September 1st, 2022, to October 15th, 2022, with a duration of 45 days. After pre-processing, we retained a total of 286.4 million disaggregated GPS records. The data fields used include device identifiers, latitude, longitude, geohash, and timestamp.

3.2.2 Food Outlets

We obtained a comprehensive database of food outlets in North Florida from the University of Florida GeoPlan Center (Alachua County 2022). This database covers various components of the local food system, including food production sites, retailers, and distribution sites. Our study focuses on the food retail category, which includes grocery stores, supermarkets, drug stores, corner stores, gas station stores[1] and restaurants. This study focuses solely on food-at-home (FAH) access, excluding restaurants and dine-out locations.

Trip purpose or purchasing information are usually absent from secondary GPS data. However, from a supply-side perspective, different food outlets offer varying types, prices, quantities, and varieties, which can significantly impact the health outcomes of individuals and communities (Ma *et al.* 2017). Also, from a demand-side perspective, individuals visit different food retailers for various purposes, not limited to food access. For instance, some may frequent big box stores for non-food items while visiting gas station stores for food. The USDA's FoodAPS indicates that SNAP households allocate 13% of their food spending to convenience stores, dollar stores, and pharmacies (Todd and Scharadin 2016).

To address these complexities, we build upon existing survey practices and literature (Todd and Scharadin 2016, Xu *et al.* 2023) to develop a two-step classification approach:

(1) *Food quality and physical size*, four types:

[1] The drug stores, corner stores, and gas station stores in the dataset have been verified to sell food. See additional details here: https://www.geoplan.ufl.edu/portfolio/foodshed/

- *Large Groceries or Supermarkets* that mainly sell food (e.g. Publix, Winn-Dixie)
- *Big Box Stores* that carry a full range of food products in addition to other products (e.g. Walmart Supercenter, Target)
- *Small Retailers* that sell healthy grocery items like milk, eggs (e.g. CVS, Dollar General)
- *Stores Selling Only Processed or Low-Quality Food* (e.g. Circle K, 7-Eleven)

(2) <u>*Purpose of visitation*</u>: We distinguish between

- *Stores Primarily Selling Food*, where individuals predominantly visit for food-related purposes (e.g., groceries, food marts).
- *Locations Visited for Various Purposes*, such as big box stores, gas stations, and *pharmacies*.

Commented [XY3]: Convert into narrative. Bullet points are discouraged for journal articles

We will first consider all POIs and categorized them into these four types, then narrow our analysis to those that primarily sell food. *Figure 3* shows the distribution of the stores under the two classifications and *Table 1* summarizes the numbers.

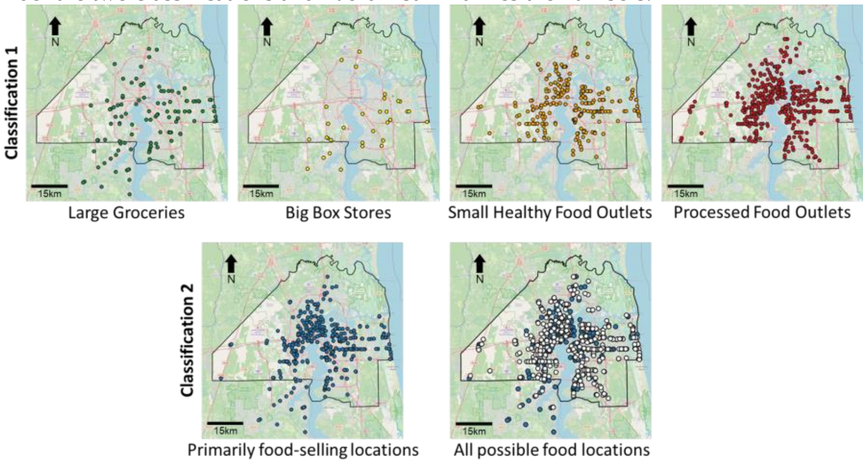

Figure 3 Distribution of different retailers in the study area

Table 1 Number of stores under different classifications

	Large Groceries	Big Box Stores	Small Healthy Outlets	Processed Food Outlets	Total
Primarily food-selling location	110	0	123	176	409
Other	6	31	153	323	513
Total	116	31	276	499	922

3.3 Method

Figure 4 shows the analytical framework of the case study. We begin by processing and merging the GPS, food outlets, and sociodemographic data. With the integrated data, we infer individual food access trips. From these activities, we then calculate food access metrics and analyze their patterns.

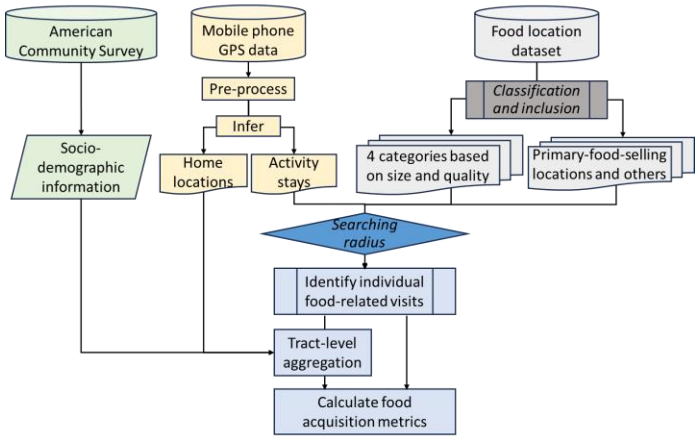

Figure 4 Analytical framework of the case study

3.3.1 Food Access Trip Inference

With the secondary GPS dataset, we first infer the home locations of the users and filter out the users without an inferred home. Then, we infer the stay points and trips of each user. With this information, we then merge the food store location data and the socio-demographic data, extract food-related visits and conduct further analysis.

(1) *Data preparation*. We first build a *geopandas* POINT object from the longitude, latitude, and timestamp of the pre-processed GPS data, along with the user identifiers.

(2) *Home inference*. Home is a crucial element in individuals' daily activity (Zenk *et al.* 2018) as well as food acquisition activity (Coleman-Jensen *et al.* 2019). It is also fundamental in aggregating the individuals and conducting tract-level analysis. Based on the pre-processed GPS data, we infer a home location for each user in the dataset. To do this, we adopted the proxy-home-location inference algorithm developed by Zhao et al. (Zhao *et al.* 2022). We first mesh the study area into 20-meter square grids and then count the number of GPS positions of each user during the period of 10:00 PM to 6:00 AM the next day. We assign the users' homes as the grids with the most data points. For the users whose home locations cannot be found through this, we then extract their weekend GPS data and consider the users' homes located in the grid where they spend the most amount of time during weekends.

(3) *Stop inference*. We conduct stop inference to extract activity stays from the GPS data. In this case study, we use the Python package *Trackintel* (Martin *et al.* 2023), a package for movement data processing. In *Trackintel* package and this study, an activity stay or stop is where the user made no large movement in a period. The package adopts a sliding window detection algorithm to detect point clusters and generate stops (the geometric center of the clusters). In this study, we set the staying thresholds to a space of 100-meter radius, and a duration of at least 5 minutes and at most 720 minutes.

(4) *Trip inference*. Apart from stop inference, we also conduct trip inference to extract the origins of the activities. In *Trackintel* package and this study, a trip is the chaining of the trip-legs that connect pairwise stops. The package adopts a

backward searching method to identify the temporally connected movements. It then groups them into one trip and extracts the trip starting point. In this study, we set the threshold for connecting point pairs to 60 minutes.

(5) *Identify food-related activities*. Among all the extracted stops, we then identified the ones within certain radii of the food outlets and considered those as food-related stops and their trips as food-related trips. We first pre-determine a set of radii and then discuss the implication of selecting those values (see *Sensitivity Analysis* section below). During this process, we exclude the stays points whose duration are longer than 2 hours. We adopted this threshold based on the American Time Use Survey (ATUS) result where the median grocery shopping activity duration is 30 minutes with a standard deviation of 30.6 minutes (Brown and Borisova 2007).

3.3.2 Descriptive and Spatiotemporal Analysis

Based on the inferred food-related stops and trips, we compute various metrics to analyze individuals' and populations' food acquisition patterns. Guided by literature, we calculate the following four metrics that have been widely adopted in survey studies (Leroy *et al.* 2015, Todd and Scharadin 2016):

(1) *Number of food retailer visits*. To describe the food acquisition frequency, we calculate the number of food retailer visits made by each individual during the study period.
(2) *Number of unique stores visited*. To capture the diversity and variability in food retailer visitation, we calculate the number of unique stores visited.
(3) *Home-to-store distance*. To analyze spatial characteristics, we calculate the distance between the user's *inferred home location* and the *visited* retailer store, as well as the distance to the *nearest* store. We discuss the differences between these values. Recognizing that network distance better reflects the accessibility and travel patterns in high-density urban areas, we used OpenStreetMap road network to calculate network distances.
(4) *Proportion of home-based visits*. To analyze the origins of the food acquisition activities, we calculate the proportion of home-based visits. We consider the trips with origins within 200 meters (circular buffer) of the user's home as home-based visits and divide the number of home-based trips by the total number of trips.

With the calculated metrics, we conduct spatial and temporal analysis and generate visualizations. In the temporal perspective, we select the number of visits as the metric for the visualization. We illustrate its time-of-day and the day-of-week pattern. For the time-of-day pattern, we plot the distribution of number of visits by hours. For the day-of-week pattern, we first compare the time-of-day pattern of weekdays and weekends and then plot the distribution of visits by day of the week. We also analyze the longitudinal pattern of the 1.5-month period. We calculate the total number of visits each day, plot the curves, and discuss the patterns with the major events in the calendar. In the spatial perspective, we select the home-to-store distance as the metric for the visualization. We aggregate the individual metrics by calculating the tract-level average and visualize them in maps. We also explore the differences in visit patterns across store types and their correlations with the sociodemographic characteristics.

3.4 Results

We present the empirical findings, and then discuss the potential of the dataset and the implications of its limitations.

3.4.1 Sampling Rate

Figure 5 shows the spatial distribution of the individual devices and the histogram of the tract-level sampling rate. We extracted 852,224 food-related stops generated by 93,854 individual devices. The average tract-level sampling rate is 10.4%. The map suggests a reasonable approximation of the population spatially, and the histogram shows a normal-like distribution centered around 8%. This indicates that despite some spatial variations, mobile device location data can significantly increase the sample size and spatial coverage of study populations.

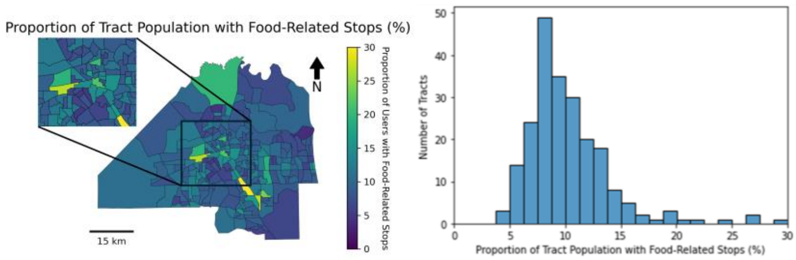

Figure 5 Spatial distribution and histogram of tract-level sampling rate

We notice significant variations in sampling rate from the histogram and the map. This aligns with findings from prior studies, suggesting the existence of geographic bias in GPS data (Li *et al.* 2023). In general, we observe lower sampling rates in the suburban and rural parts of the study area, which is not surprising as digital infrastructure (e.g., wireless network) tends to concentrate in urban areas. Interestingly, we also notice significant variability in sampling rates across tracts with lower socioeconomic status (i.e., areas with lower vehicle ownership and lower income levels). Unraveling the mechanisms behind this variability requires an in-depth investigation, which is beyond the scope of this study. Moreover, a further analysis that compares the demographics and socioeconomic characteristics of GPS device users to the general population in the study area would enhance our understanding of data representativeness and result generalizability; however, the absence of sociodemographic information from the GPS data prevents us from conducting such comparisons.

Commented [XY4]: “Fluctuations” are usually used to describe temporal trends.

3.4.2 Extracted Food-related Stops and Trips

Table 2 shows the number of food-related stops and trips extracted. As discussed in Section 3.3.1, we infer food-related stops to identify food acquisition visits, and trips to identify the origins of those visits. We pre-determined a set of searching radii for identifying food-related stops: 150m for Large Groceries, 200m for Big Box Stores, and 50m for Small Healthy Outlets and Processed Food Outlets, which are selected based on results from the sensitivity analysis (see the *Sensitivity Analysis* section). As shown in the table, the size of sample extracted from the GPS data is significantly larger than those of traditional surveys.

Table 2 Food-related stops and trips extracted

	Large Groceries	Big Box Stores	Small Healthy Outlets	Processed Food Outlets	Total
Stops	250,916	76,979	191,796	332,533	852,224
Trips	1,336	646	801	1,808	4,591

These results suggest that secondary GPS data can capture larger sample sizes, which allows us to calculate individual and sample average food access metrics (Section 3.4.3) and obtain more detailed spatiotemporal pattern (Section 3.4.4). As shown in Table 2, we were able to extract from the GPS data not only Jacksonville residents' visits at various types (i.e., stops) of food stores but also from where they traveled there (i.e., trips). Also, the timestamp in the extracted visits allows us to analyze the temporal patterns. Although, Table 2 also shows that inferred trips are significantly fewer than the inferred stops. As discussed in the *Data and Method* section, the drop is due to the reliance on continuous tracking points for inferring food acquisition trips. This condition can be very strict: it requires the devices to be on, tracking service to be activated, and coverage to be reliable. This is a challenge often encountered in secondary GPS data. In contrast, survey-based methods capture trip information in a single survey, but they can be susceptible to recall inaccuracies and low spatiotemporal resolution (Ver Ploeg *et al.* 2015). Primary GPS data, while offering high resolution, tends to have smaller samples due to cost constraints (Zenk *et al.* 2011). Therefore, it will be a difficult yet crucial task for future research to improve secondary GPS data quality and granularity and contextualize the dataset with trip purposes and demographic information, which will significantly enhance the dataset's power in studying food access patterns (Nguyen, Armoogum, *et al.* 2020). In the following subsections, we further explore the inferred stop and visit by calculating food access metrics from them.

3.4.3 Food Access Metrics

Here we present the area-wide metrics calculated from the extracted food acquisition activity and compare those in survey-based studies from prior literature. *Table 3* shows the four metrics for each type of store.

Table 3 Food acquisition metrics for each type of store

Metrics	Large Groceries	Big Box Stores	Small Healthy Outlets	Processed Food Outlets	All Food Locations	Nation-wide Survey Findings
Number of visits per individual (visits)	4.74	3.12	4.1	5.02	9.08	17.2 (Todd and Scharadin 2016)
Number of unique stores visited per individual (stores)	1.86	1.31	1.93	2.38	3.85	4.4 (retail banners, for grocery purpose, each month) (The Food Industry Association 2019)

Distance of visited store to home (km)	Euclidean	5.29	6.54	5.63	5.8	5.62	6.1 (Euclidean distance, to primary store) (Todd and Scharadin 2016)
	Network	7.43	8.69	7.21	7.47	7.61	
Proportion of home-based visits (%)		18.65	14.44	18.33	16.02	17.84	64 (grocery trips) (Ver Ploeg *et al.* 2015)

The descriptive analyses reveal the effectiveness of GPS data to estimate the same metrics that were traditionally extracted from surveys, such as visitation frequency, stores visited, and trip origin. However, we observe inconsistencies compared with prior studies. The GPS-based metrics exhibit significantly lower levels compared with their counterparts from survey-based studies, suggesting that the data may underestimate users' food access patterns. For instance, the results indicate an average of 9 visits over 1.5 months, translating to an average of 1.4 visits per week. In contrast, the nationwide USDA FoodAPS survey reports an average of 6.47 food acquisition events per week for its respondents (households, average size 2.42) (Todd and Scharadin 2016). A statewide study in Florida demonstrates a lower frequency, with 5.0% of respondents visiting grocery stores or other retail markets daily, 29.6% twice weekly, and 39.4% weekly (Hodges and Stevens 2013). The results from the statewide survey study, although lower, still exceeds the visitation frequency observed in our case study. In contrast to the American Time Use Survey, which indicates that around 64 percent of grocery shopping trips involve a direct journey from home to the store (Ver Ploeg *et al.* 2009), the GPS data results underestimate this proportion. Regarding the metric of home-to-store distance, the extracted values are more consistent with literature (3-4 miles, Euclidean distance) (Ver Ploeg *et al.* 2015).

The observed disparities in the secondary GPS dataset can be attributed to two factors. First, despite higher coverage at the population level, the temporal coverage of each individual's activities can still be low. Intermittent tracking, present in both primary and secondary GPS data, results in low temporal coverage and a decrease in the recorded number of visits. Secondary GPS data, in particular, encounters an additional challenge known as behavior bias (Li *et al.* 2023). This bias occurs when individuals activate location services selectively, typically during specific app usage, leading to inconsistencies in tracking. In the context of food access, this bias may arise when device users, who may be already familiar with nearby food retailers, refrain from activating location applications. This may contribute to the lower frequency and home-based visit percentage. Second, the inference procedure may also introduce errors, which is extensively discussed in the sensitivity analysis section later.

3.4.4 Spatiotemporal Patterns of Food Access

A major advantage of GPS data is its high resolution and ability to capture detailed and precise patterns in both spatial and temporal dimensions. For spatial analysis, we focus on the *home-to-store distance* metric, examining both the distance to the stores *visited* and the distance to the stores *nearest* to the individuals' homes. We first focus on each sample *individual*, plotting the distribution curves and the scatter density plots of the two distances in *Figure 6* and *Figure 7*. *Figure 7* is the iso-density graph of the visited v.s. nearest store distance points of each individual. Darker colors indicate higher density levels, meaning more individuals. We then aggregate the individual measurements to tract-level and visualize them in maps in *Figure 8*. *Figure 8* presents the tract-level average home-store distance values, where darker colors indicate larger distances.

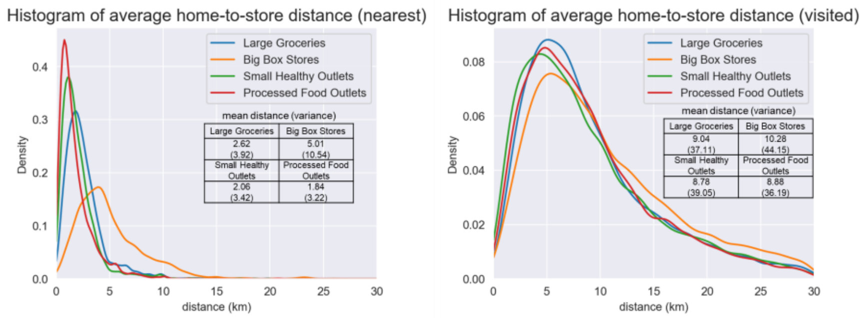

Figure 6 Density plots of home-to-store distances

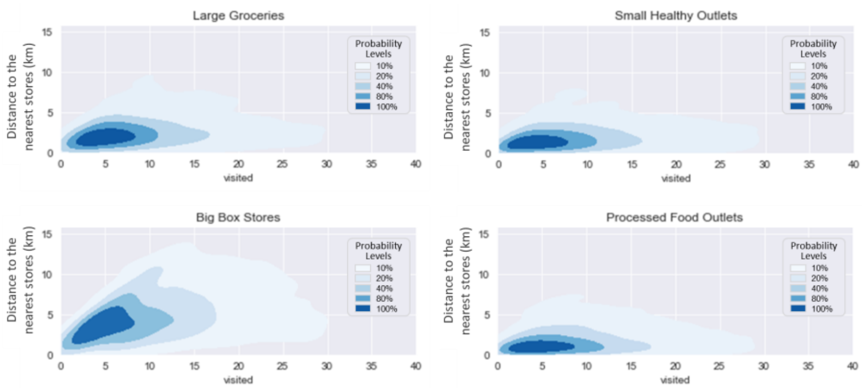

Figure 7 Density contours of home-to-store distances

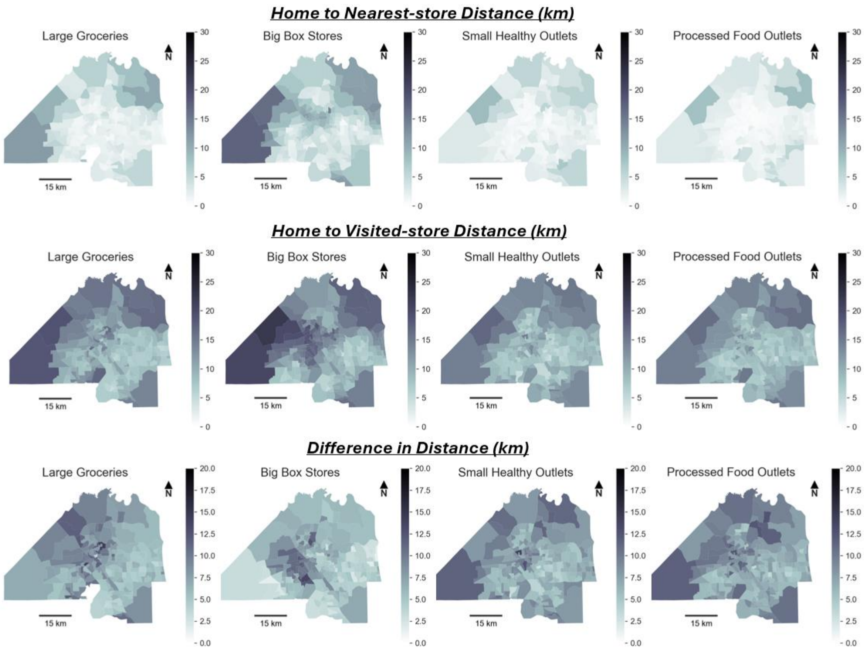

Figure 8 Spatial distribution of home-to-store distances

Firstly, the three figures illustrate notable trends that align with prior survey-based studies. For example, food-related visits in urban areas tend to cover shorter distances. The distances to the visited stores are generally larger than the distances to the nearest stores, suggesting that individuals may bypass the stores closest to their homes. These findings are also commonly observed in literature (Ver Ploeg *et al.* 2015). Moreover, we can observe detailed patterns that are often overlooked in survey-based studies due to limited sample sizes. The distribution curves and scatter density plots allow us to examine individual-level differences. As depicted in *Figure 6*, distances to the *nearest* big box stores surpass those to other store types, displaying not only greater magnitudes but also increased variability (in descending order: Big Box Stores, Large Groceries, Small Healthy Outlets, and Processed Food Outlets). On home-to-*visited*-store distances, both two individual-level figures (*Figure 8*) show that the four store types have similar patterns. This implies that a group of individuals may share similar patterns in visitation to various food retailers. This aligns with previous findings in travel behavior and food access that, while small samples may reveal distinct patterns for various stores based on individual characteristics (Liu *et al.* 2015), the results are usually similar when using aggregated data (Supernak 1967).

Our results also reveal the spatial heterogeneity in the visitation patterns, which are a major advantage of GPS data. *Figure 8* displays tract-level aggregated distances. Generally, the distances increase as one moves further away from the urban areas, but there are exceptions. Particularly, for big box stores, the urban core areas exhibit noticeably large distances. If we overlay these maps with socio-demographic characteristics (*Figure 2*), we can identify that these areas are marked by lower income, higher population density, and a higher percentage of non-white populations. Previous studies in Jacksonville also observed this urban food desert phenomenon (Lewis *et al.* 2018).

Another interesting phenomenon captured by *Figure 7* is the difference in slope between the four categories of food stores. The value of slope represents the people's preference to a closer store, with a slope of 1 meaning people will exclusively visit the closest store and a slope of 0 meaning people will not view distance as a factor when choosing stores to visit. Among the four types, the Big Box Stores have the steepest curve, followed by Large Groceries, while Small Healthy and Processed Food outlets have almost a flat curve. This also means that people tend to visit the closest big box stores and large groceries, while they do not view distance as a factor when choosing small healthy and processed food outlets. This can be primarily because the difference between stores at different locations can be less significant due to standardization guided by business strategies. Meanwhile, small healthy and processed food outlets are more heterogeneous and diverse in both quality and price, which drives people to visit specific locations.

Figure 9 illustrates the daily and weekly food outlets visitation patterns extracted from the dataset. We normalized the data with the total number of visits of the day/week periods to produce percentages. The time-of-day curves depicted in the top two figures show similar patterns between weekdays and weekends. However, the visits are more concentrated during the daytime on weekends and the evening peak is less pronounced on weekends and shifts towards noon. This is consistent with the literature, as people's primary locations and activities often differ on weekends (Zhao *et al.* 2022). This variation is more pronounced for big box stores compared to other types of stores, which is logical given the nature of weekend shopping activities (East *et al.* 1994).

From the day-of-week curves in the bottom figure, we can observe a peak on Friday. Other researchers have also observed a similar phenomenon (Cai 2006). Furthermore, we conducted longitudinal analyses. *Figure 10* shows the number of trips extracted each day during the study period. Note the significant drop around Labor Day (Sep 5, 2022).

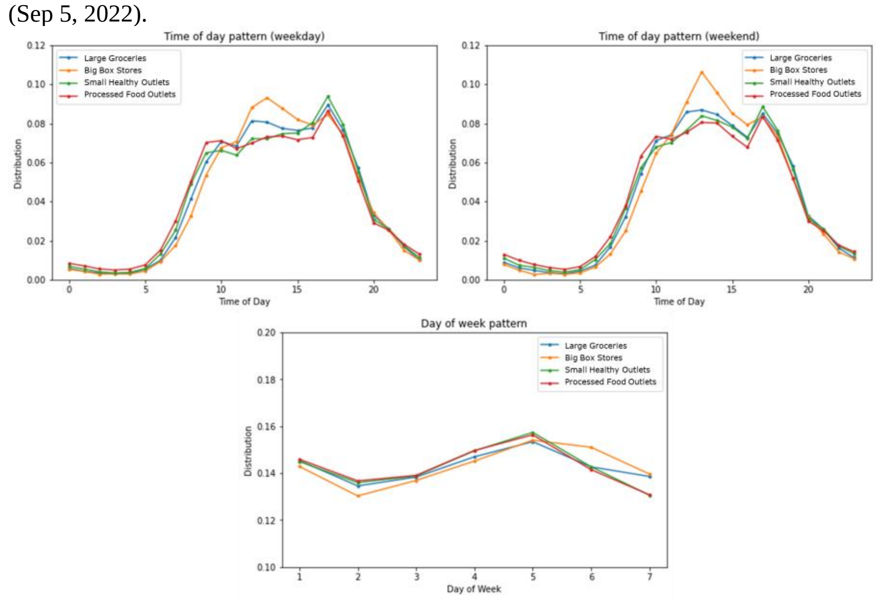

Figure 9 Temporal distribution of food trips for each type of food stores

The results of temporal analysis and longitudinal analysis are very coherent. Previous survey-based studies struggled to divide the reported trips into different periods on weekdays due to lower sample size. Longitudinal analysis requires continuous surveying over an extended period, which can be expensive and time-consuming. Moreover, the passive collection nature makes secondary GPS data an excellent source for these analyses. Theoretically, the data are generated continuously, providing 24/7 observations, and they offer more accurate timestamps compared to recall-based shopping logs. Therefore, we can conclude that secondary GPS have great potential as a complement to traditional methods, and they can offer more nuanced insights into the intricacies of the food access behavior.

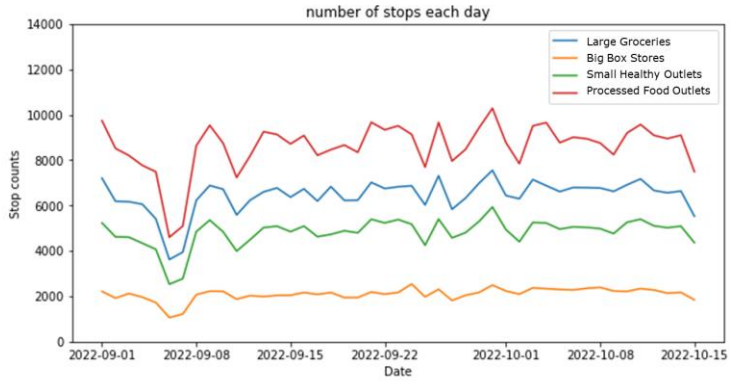

Figure 10 The change of the number of trips during the study period

4 Sensitivity Analysis

GPS data, especially secondary GPS data, lacks direct information about the activities performed, requiring inference procedure, which introduces uncertainty. Two possible sources of uncertainty are (1) the variability in activity purposes at the same location and (2) the parameters used in extracting the activities (Kwan 2016). It is important to evaluate how different design choices interact with each other and with the GPS data and produce varying results. In this study, we explore the first uncertainty by applying a new store classification that limits stores to primary food-selling locations (Section 4.1), and the second, by testing the radii for identifying food acquisition activity (Section 4.2).

4.1 Food Store Classification

The first aspect to be evaluated is regarding the store inclusion criteria, i.e., visit to which stores can be classified as food acquisition. In previous analysis, we considered all the possible food stores listed in the dataset, even the ones that also carry non-food items (such as big box retail stores). We now focus only on the *primarily* food-selling locations and recalculate the metrics in *Table 3*. *Table 4* and *Figure 11* show the results and comparison for both analyses. It is noteworthy that the type *Big Box Stores* are not included here because they all are not primarily food-selling locations.

Table 4 Food access metrics for each type, primary food-selling locations only

Metrics	Large Groceries	Small Healthy Outlets	Processed Food Outlets	All Food Locations
Number of visits per individual (visits)	5.13	3.47	6.04	5.13

Number of unique stores visited per individual (stores)	1.87	1.46	2.41	1.87
Distance of visited store to home, network distance (km)	8.15	7.82	7.03	7.52
Proportion of home-based visits (%)	17.95	21.62	18.9	17.95

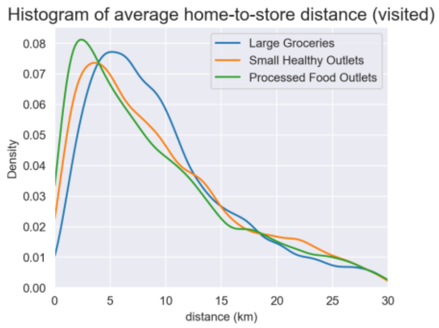

Figure 11 The extracted home-to-store distance, for primary food-selling locations only

We first examine all food retailers without classification. Comparing the last column (considering all store types) of *Table 4* to *Table 3*, we witness a decrease in both the total number of visits and unique stores visited. This is intuitive as we limited the stores considered. However, we also witness a decrease in home-to-store distance and an increase in the proportion of home-based visits. This can be explained by the activity space theory. According to prior literature (Gong *et al.* 2020), people exhibit different radii (home as the center) for different activities, with maintenance activity radius smaller and recreational activity space larger. By limiting the stores to primary food-selling locations, we could have reduced non-food-access-related visits.

We then distinguished different types of stores, and observed distinct patterns:

- The trips to Processed Food Outlets behave similarly to the pre-classification trend: fewer visits, fewer unique stores visited, smaller home-to-store distances, and a higher proportion of home-based visits.
- The trips to Healthy Outlets also behave similarly, except that the home-to-store distances increased instead of decreased.
- The trips to Large Groceries, however, behave differently. The home-to-store distances increased, similar to that of Small Healthy Outlets, but the proportion of home-based visits decreased. Also, the number of visits and the number of unique stores visited increased instead.

In our analysis, we calculated the metrics among the individuals who performed the activity (i.e. the population who visited the stores in the shorter list are different from those who visited the stores in the original list). So, these findings suggest that behavioral differences exist among individuals visiting primarily food-selling locations and those visiting complex food locations.

The increase in grocery visits per individual is worth noting. Visitors to grocery stores that primarily sell food items (e.g. Whole Foods) may visit more frequently and travel further compared to those to grocery stores that also carry non-food items (e.g. Winn-Dixie); visitors to gas station stores (a sub-type of Processed Food Outlets) may go more frequently than those visiting fast food and processed food stores (another sub-type of Processed Food Outlets). The individuals who tend to visit grocery stores that only sell food items may be the ones who already have relatively better access to those stores. They are able to travel longer distances, visit more frequently, and be more flexible regards the origins of the visits. On the other hand, there are individuals who frequently visit gas stations and dollar stores. They may have fewer options other than the unhealthy or innutritious food outlets.

These behavioral differences can inform designing food security or nutrition security interventions. There are multiple existing studies that support incorporating healthy food items into non-traditional outlets as an effective intervention approach from the supply-side perspective (Lucan *et al.* 2018, Chenarides *et al.* 2021). Researchers have argued that in areas with easier access to fast food restaurants and convenience stores but limited access to supermarkets, enhancing the variety of foods in existing stores may be more effective than opening new stores (Ver Ploeg *et al.* 2009). The strategy can also promote demand for healthy food and a healthy lifestyle. A national dollar store perception and utilization survey by the Center for Science in the Public Interest (CSPI) showed that dollar stores are important in food access in communities with limited resources and the residents strongly support offering healthy options at those locations (John *et al.* 2023), which is consistent with our findings from the case study section.

We also find that the high-quality grocery visitors in the study area tend to be more advantaged in terms of food access; however, there is a noticeable trend of reliance on gas station stores in terms of food access. *Table 1* shows that 65% of the processed food outlets in the study area are gas station stores and dollar stores that do not primarily sell food. After adding those less likely food locations, the number of visits per individual increased from 3.47 to 5.02. This can serve as a scenario simulation of adding healthy food options to those stores: both the physical accessibility and actual exposure to healthy food would increase after the intervention. This further highlights the strategy's potential as a successful approach to tackling food deserts moving forward.

4.2 Food-access Trips Identification

The second parameter to be assessed is regarding the radius of food store when identifying food access trips. *Figure 12* shows the satellite images of two food locations in Google Map. In the case of the supermarket on the left, a search radius too small may result in false-negative identifications, i.e., food access trips are misidentified as non-food access trips. Meanwhile, in the case of the small grocery store on the right, a radius too large might result in false-positive identifications, i.e., non-food access trips are misidentified as food access trips. Therefore, we test the radii of 50m, 100m, 150m, and 200m, and examine whether they would yield significantly different results.

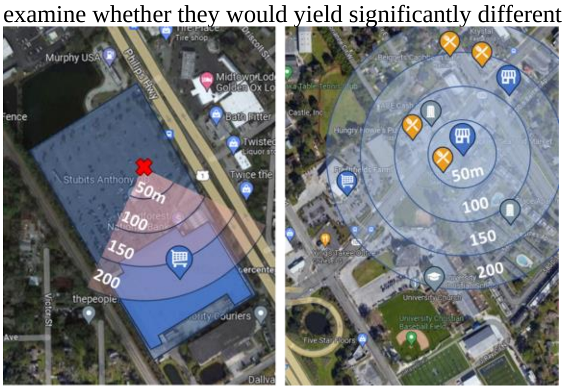

Figure 12 Identification radii in food-related trip extraction

In the previous analyses to extract food-related trips, we selected a food-related visit identification threshold of 150 meters for Large Groceries, 200 meters for Big Box Stores, and 50 meters for Small Healthy Outlets and Processed Food Outlets. However, this parameter can significantly influence the analysis. In light of this, our case study explores different values and examines the robustness of the results. *Figure 13* shows the food access metrics for different store types calculated with different thresholds.

Figure 13 The food access metrics calculated under different radii

The *Number of Visits per Individual* (the first figure) is the most fundamental metric among the four. We can observe steep increases for Small Healthy Outlets and Processed Food Outlets with larger radii, while the rise for Big Box Stores is relatively steady. These patterns align with expectations. Given the abundance of Processed Food Outlets and their concentration in densely populated urban areas (*Figure 3* and *Table 1*), they may be more sensitive to radii too large than big box stores in suburban or rural areas. However, as shown in the rest three figures, the other three metrics also change as the radius increases, and their changes are less intuitive.

This highlights the sensitivity of the GPS-based approach to trip identification parameters. We recommend different values for different types of stores instead of setting one value. For our case study area, Jacksonville, we suggest adopting 50m for Small Healthy Outlets and Processed Food Outlets, 200m for Big Box Stores, and 150m for Large Groceries. The values are based on *Figure 13*. Setting 50m for the two types of small food retailers is because their *Number of Visits per Individual* metrics increase significantly after that; the metric of big box stores remains steady increase. Setting 150m for groceries is because they are larger than small retailers but smaller than big box stores, and their home-to-store distance metric (third figure) switched to increase when set to 200m.

5 Discussion

This study systematically examines the potential and limitations of employing large-scale human mobility GPS data in the context of food access research. Using a Terabyte-level disaggregated GPS database with over 286.4 million GPS signal records, we have inferred food-related trips and stops and analyzed the food access patterns across different store types in Jacksonville, Florida. We offer several high-level discussions on this topic below.

5.1 Using Secondary GPS Data for Food Access Analysis

Overall, we find that the GPS data are promising for advancing food access research and can be used to inform policymaking. Compared to traditional approaches such as surveys, GPS data can lead to significantly higher sampling rate and broader spatial coverage of the study population. Moreover, our investigation reveals that secondary GPS data can generate the widely adopted metrics and replicate the analyses commonly done in traditional food-access studies. Notably, GPS data can facilitate a more nuanced understanding of the spatiotemporal patterns of food accessibility patterns than traditional approaches. A prime example is temporal analyses, which can be rather expensive for traditional approaches but are straightforward and cost-effective with GPS data.

However, our analysis also reveals several significant limitations of secondary GPS data in food access research, which affect the accuracy and representativeness of study findings. First, we noticed GPS data significantly undercount food-related activities in our case study. This could be due to *Selective Daily Mobility Bias* (SDMB), introduced by people's inconsistent activation of location-tracking services over time or their choice of not using navigation apps for grocery trips to stores they frequently visit. Moreover, we notice that different assumptions and parameter settings in the inference process can have major implications on the research results. The sensitivity analysis shows that the classification of food stores and the selection of identification radii have major implications on the four food-access metrics, which means that the robustness of study findings is subjective to algorithmic design and parameter choices (Kwan 2016). Furthermore, despite capturing larger sample sizes, the GPS data contain spatial biases. As shown in Section *3.4.1*, the sampling rate significantly varies across the area.

Finally, the absence of individual-level sociodemographic information from the GPS data limits its applicability to shed light on food access patterns across population groups. Existing studies have shown that while the GPS data are well-sampled across demographic categories (Squire 2019), some population groups are underrepresented in the data. For example, by using mobile device location data collected by SafeGraph Coston *et al.* (2021) showed that older and non-white individuals are under-sampled. Li *et al.* (2023) found that the Hispanic and low-income populations were underrepresented across the states in the U.S., while the advantaged groups, e.g., the high-income and highly educated people were overrepresented.

5.2 Policy Implications

First, our case study not only reaffirms some findings in the literature but also generates new insights that can inform future planning strategies to improve food access. For instance, in Section 4.1, when focusing on primary food-selling locations, we observed fewer visits to large grocery stores and more visits to smaller outlets (e.g., gas stations) that primarily sell processed food. This implies that adding healthy food options to processed food outlets could be a more effective strategy for increasing food access and exposure to healthy foods than building larger stores that sell healthy food.

Commented [XY5]: Please confirm this

Second, the limitations of GPS data discussed above suggest the need for caution when using such data to inform policy. The potential for over- or under-estimation of food-related visits could lead to biased results that mislead policy and planning efforts. For example, overestimating visits may create an inflated sense of accessibility, potentially diverting attention from necessary improvements in food access. The settings of the inference algorithm are the most crucial factors due to its rule-based nature. In this paper, we provide the recommended parameter values and settings for the future studies. It is also important to note that GPS data only captures food acquisition behavior, reflecting just one aspect of the larger issue of food security. It is essential to recognize that food purchasing, food consumption, or dietary outcomes are distinct and require separate considerations (Simelane and Worth 2020). The central problem and the pressing concerns of food insecurity/access is its association with negative nutrition- and diet-related health outcomes and its relationship with health disparities (Singleton *et al.* 2023).

Third, the continuous and passive nature of GPS data collection presents a unique opportunity for policymakers to monitor the long-term impacts of food security programs. For instance, by tracking changes in visitation patterns to grocery stores and other food outlets before and after the implementation of food assistance programs, policymakers could assess the effectiveness of these interventions. This would provide real-time feedback on the success of initiatives and allow for more agile adjustments to policy measures. This longitudinal approach would also remedy the drawbacks of the GPS data discussed above if pre- and post-program analysis follows the same inference rules.

Finally, our study underscores the importance of mixed-method research. While both big and small data—and the methods used to analyze them—have limitations, they can complement and enhance each other (Kwan 2016). As noted in the literature review, studies that integrate primary GPS data collection with surveys have uncovered novel insights that challenged previously held behavioral assumptions. We believe that applying a mixed-method approach to complement secondary GPS data with other data sources can provide a more holistic understanding of food access behaviors and improve the interpretation of key metrics, which is beneficial for future planning and policy making. For instance, conducting surveys on the same population from which the GPS data are collected would allow for the triangulation of results. Such data integration would facilitate statistical validation of the metrics extracted from both sources, addressing potential intrinsic or inference-induced inaccuracies in secondary GPS data and enhancing the generalizability of study findings.

5.3 Limitations and Future Research

This section notes several study limitations and offers several potential future research directions. First, the anonymity of secondary GPS data used here prevents us from distinguishing whether the observed differences in food access patterns across individuals are mainly shaped by their exposure to food environments or by their personal tastes (Jin *et al.* 2023). Populations groups tend to exhibit distinct perceptions and preferences towards various types of food outlets, but their behavior is also influenced by the food environment. Future research that isolates the independent effects of the two factors can lead to more targeted intervention strategies.

Another limitation is the temporal and spatial generalizability of the findings. Temporally, research on food sales has shown seasonality in food demand (Hu *et al.* 2021, Balagtas *et al.* 2023), which may influence mobility patterns. Therefore, generalizing our findings from the 45-day study period could introduce bias and limit the representativeness of the results. Spatially, study findings from Jacksonville may not be transferable to other contexts. A 2012 study noted disparities in food access among Health Zones within the city, with Urban Core residents facing a greater health burden (Healthy Jacksonville Children Obesity Prevention Coalition 2012). Additionally, Jacksonville's poverty rate (14.8%) exceeds both the national (12.5%) and state averages (12.9%) (U.S. Census Bureau 2022). These socio-economic factors should be considered when generalizing the results to other contexts.

Finally, this study focused solely on home locations, distinguishing between home-based and non-home-based activities. Expanding the analysis to include work and recreation locations could yield valuable insights (Ver Ploeg *et al.* 2015). In addition, improving food access has broader implications beyond food security. It addresses social equity, as marginalized communities face disproportionate barriers to accessing healthy food (Jin *et al.* 2023). It can also strengthen local economies by supporting community-based food businesses, promoting local sourcing, and reducing food transport—ultimately enhancing environmental sustainability (Lucan *et al.* 2018). Future research could explore sociodemographic factors like income, race, and age to better understand the spatial and temporal dynamics of food access and contribute to more comprehensive interventions (Li and Kim 2020, Zhao *et al.* 2022).

6 Conclusion

Mobile device location data presents a novel approach to studying food access. In this study, we systematically assess the potentials and limitations of human mobility GPS data compared with traditional approaches. We conducted a case study in Jacksonville, Florida, with a large-scale human mobility GPS dataset of 13 billion GPS records in the whole state and 286 million records in Jacksonville.

With the GPS dataset, we analyzed the distribution of four major metrics commonly employed in traditional food access studies and explored their spatial and temporal patterns. The results demonstrate the capability of GPS data to extract key insights regarding food access patterns that confirm findings from prior studies. On the other hand, our analysis also demonstrates that relying on GPS data would significantly underestimate the food-related activities and frequency. Our sensitivity analyses, focusing on the classifications of food-selling stores and identification radii of food-access trips, reveal some inherent challenges of extracting food-related trips from human mobility GPS data and algorithmic uncertainty in this process.

While affirming the potential of GPS data for food access analysis, the study also emphasizes the need for cautious use of such data. Future research should focus on improving data quality, refining activity inference algorithms, and incorporating diverse data sources and domain knowledge to gain a deeper understanding of people's food access patterns. Our research highlights the need for critical reflexivity, the detailed examination of the data and algorithms used, and the findings generated from them.

Acknowledgements

We thank Erik Finlay at the University of Florida GeoPlan Center for sharing the North Florida food retailer location database with us.

References

Alachua County, 2022. Alachua County Foodshed Map Sample.

Anekwe, T.D. and Zeballos, E., 2019. *Food-related time use: changes and demographic differences*.

Balagtas, J.V., Cooper, J., McLaughlin, P., and Qin, F., 2023. Consumer demand for food at home and food away from home: Understanding economic linkages during the pandemic. *Applied Economic Perspectives and Policy*, 45 (3), 1604–1617.

Berkowitz, S.A., Karter, A.J., Corbie-Smith, G., Seligman, H.K., Ackroyd, S.A., Barnard, L.S., Atlas, S.J., and Wexler, D.J., 2018. Food insecurity, food "deserts," and glycemic control in patients with diabetes: a longitudinal analysis. *Diabetes care*, 41 (6), 1188–1195.

Bodor, J.N., Rice, J.C., Farley, T.A., Swalm, C.M., and Rose, D., 2010. The association between obesity and urban food environments. *Journal of Urban Health*, 87, 771–781.

Brown, C. and Borisova, T., 2007. Understanding commuting and grocery shopping using the American Time Use Survey. *In*: *presentation at the International Association of Time Use Research XXIX Conference, Washington DC*. Citeseer.

Cai, J., 2006. Household production or leisure? A study of grocery and other shopping (Preliminary Draft).

Cetateanu, A. and Jones, A., 2016. How can GPS technology help us better understand exposure to the food environment? A systematic review. *SSM-population health*, 2, 196–205.

Chang, T., Hu, Y., Taylor, D., and Quigley, B.M., 2022. The role of alcohol outlet visits derived from mobile phone location data in enhancing domestic violence prediction at the neighborhood level. *Health & Place*, 73, 102736.

Chen, C., Ma, J., Susilo, Y., Liu, Y., and Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285–299.

Chenarides, L., Cho, C., Nayga Jr, R.M., and Thomsen, M.R., 2021. Dollar stores and food deserts. *Applied Geography*, 134, 102497.

Coalition, H.J.C.O.P., 2012. Duval County food summit fact sheet.

Coleman-Jensen, A., Rabbitt, M.P., Gregory, C.A., and Singh, A., 2019. Household food security in the United States in 2018. *USDA-ERS Economic Research Report*, (270).

Coston, A., Guha, N., Ouyang, D., Lu, L., Chouldechova, A., and Ho, D.E., 2021. Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy. *In*: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Presented at the FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, 173–184.

East, R., Lomax, W., Willson, G., and Harris, P., 1994. Decision making and habit in shopping times. *European Journal of Marketing*, 28 (4), 56–71.

Elliston, K.G., Schüz, B., Albion, T., and Ferguson, S.G., 2020. Comparison of Geographic Information System and Subjective Assessments of Momentary

Food Environments as Predictors of Food Intake: An Ecological Momentary Assessment Study. *JMIR mHealth and uHealth*, 8 (7), e15948.

Emish, M., Kelani, Z., Hassani, M., and Young, S.D., 2023. A Mobile Health Application Using Geolocation for Behavioral Activity Tracking. *Sensors*, 23 (18), 7917.

García Bulle Bueno, B., Horn, A.L., Bell, B.M., Bahrami, M., Bozkaya, B., Pentland, A., De La Haye, K., and Moro, E., 2024. Effect of mobile food environments on fast food visits. *Nature Communications*, 15 (1), 2291.

Gong, L., Jin, M., Liu, Q., Gong, Y., and Liu, Y., 2020. Identifying Urban Residents' Activity Space at Multiple Geographic Scales Using Mobile Phone Data. *ISPRS International Journal of Geo-Information*, 9, 241.

Gravy Analytics, 2023a. Frequently asked questions.

Gravy Analytics, 2023b. Consumer privacy.

Gravy Analytics, 2023c. How to Use Gravy Forensic Flags for Retail and Transportation Analysis.

Hillier, A., Smith, T.E., Whiteman, E.D., and Chrisinger, B.W., 2017. Discrete choice model of food store trips using National Household Food Acquisition and Purchase Survey (FoodAPS). *International Journal of Environmental Research and Public Health*, 14 (10), 1133.

Hodges, A.W. and Stevens, T.J., 2013. Local food systems in Florida: Consumer characteristics and economic impacts. *In*: *Proceedings of the Florida State Horticultural Society*. 338–345.

Horn, A.L., Bell, B.M., Bulle Bueno, B.G., Bahrami, M., Bozkaya, B., Cui, Y., Wilson, J.P., Pentland, A., Moro, E., and De La Haye, K., 2023. Population mobility data provides meaningful indicators of fast food intake and diet-related diseases in diverse populations. *npj Digital Medicine*, 6 (1), 208.

Hu, Y., Quigley, B.M., and Taylor, D., 2021. Human mobility data and machine learning reveal geographic differences in alcohol sales and alcohol outlet visits across U.S. states during COVID-19. *PLOS ONE*, 16 (12), e0255757.

Jin, A., Chen, X., Huang, X., Li, Z., Caspi, C.E., and Xu, R., 2023. Selective Daily Mobility Bias in the Community Food Environment: Case Study of Greater Hartford, Connecticut. *Nutrients*, 15 (2).

John, S., Sundermeir, S.M., and Gardner, K., 2023. Stretching the dollar: Community-informed opportunities to improve healthy food access through dollar stores.

Kalter, M.-J.O., Geurs, K.T., and Wismans, L., 2021. Post COVID-19 teleworking and car use intentions. Evidence from large scale GPS-tracking and survey data in the Netherlands. *Transportation Research Interdisciplinary Perspectives*, 12, 100498.

Khan, A.A., 1992. An integrated approach to measuring potential spatial access to health care services. *Socio-Economic Planning Sciences*, 26 (4), 275–287.

Kwan, M.-P., 2016. Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge. *Annals of the American Association of Geographers*, 106 (2), 274–282.

Larsen, K. and Gilliland, J., 2008. Mapping the evolution of'food deserts' in a Canadian city: Supermarket accessibility in London, Ontario, 1961–2005. *International journal of health geographics*, 7 (1), 1–16.

Leroy, J.L., Ruel, M., Frongillo, E.A., Harris, J., and Ballard, T.J., 2015. Measuring the food access dimension of food security: a critical review and mapping of indicators. *Food and nutrition bulletin*, 36 (2), 167–195.

Lewis, F.W., Mandal, A., Stevens, A., and Jones, C., 2018. Food deserts and food insecurity: a case study in Jacksonville, Florida. *Journal of Behavioral & Social Sciences*, 5 (1).

Li, J. and Kim, C., 2020. Exploring relationships of grocery shopping patterns and healthy food accessibility in residential neighborhoods and activity space. *Applied Geography*, 116, 102169.

Li, Z., Ning, H., Jing, F., and Lessani, M.N., 2023. Understanding the bias of mobile location data across spatial scales and over time: a comprehensive analysis of SafeGraph data in the United States. *Available at SSRN 4383333*.

Liu, B., Widener, M., Burgoine, T., and Hammond, D., 2020. Association between time-weighted activity space-based exposures to fast food outlets and fast food consumption among young adults in urban Canada. *International Journal of Behavioral Nutrition and Physical Activity*, 17 (1), 62.

Liu, J.L., Han, B., and Cohen, D.A., 2015. Peer reviewed: beyond neighborhood food environments: distance traveled to food establishments in 5 US Cities, 2009–2011. *Preventing chronic disease*, 12.

Lucan, S.C., Maroko, A.R., Seitchik, J.L., Yoon, D.H., Sperry, L.E., and Schechter, C.B., 2018. Unexpected neighborhood sources of food and drink: implications for research and community health. *American Journal of Preventive Medicine*, 55 (2), e29–e38.

Ma, X., Liese, A.D., Hibbert, J., Bell, B.A., Wilcox, S., and Sharpe, P.A., 2017. The association between food security and store-specific and overall food shopping behaviors. *Journal of the Academy of Nutrition and Dietetics*, 117 (12), 1931–1940.

Martin, H., Hong, Y., Wiedemann, N., Bucher, D., and Raubal, M., 2023. Trackintel: An open-source Python library for human mobility analysis. *Computers, Environment and Urban Systems*, 101, 101938.

Nguyen, M.H., Armoogum, J., Madre, J.-L., and Garcia, C., 2020. Reviewing trip purpose imputation in GPS-based travel surveys. *Journal of Traffic and Transportation Engineering (English Edition)*, 7 (4), 395–412.

Nguyen, T.T., Hoang, T.D., Pham, M.T., Vu, T.T., Nguyen, T.H., Huynh, Q.-T., and Jo, J., 2020. Monitoring agriculture areas with satellite images and deep learning. *Applied Soft Computing*, 95, 106565.

Rabbitt, M.P., Hales, L.J., Burke, M.P., and Coleman-Jensen, A., 2023. Household food security in the United States in 2022.

Rabianski, J.S., 2003. Primary and secondary data: Concepts, concerns, errors, and issues. *The Appraisal Journal*, 71 (1), 43–55.

Sadler, R.C. and Gilliland, J.A., 2015. Comparing children’s GPS tracks with geospatial proxies for exposure to junk food. *Spatial and spatio-temporal epidemiology*, 14, 55–61.

Siddiqui, N.Z., Wei, L., Mackenbach, J.D., Pinho, M.G.M., Helbich, M., Schoonmade, L.J., and Beulens, J.W.J., 2024. Global positioning system-based food environment exposures, diet-related, and cardiometabolic health outcomes: a systematic review and research agenda. *International Journal of Health Geographics*, 23 (1), 3.

Simelane, K.S. and Worth, S., 2020. Food and Nutrition Security Theory. *Food and Nutrition Bulletin*, 41 (3), 367–379.

Singleton, C.R., Wright, L.A., McDonald, M., Archer, I.G., Bell, C.N., McLoughlin, G.M., Houghtaling, B., Cooksey Stowers, K., and Anderson Steeves, E., 2023.

Structural racism and geographic access to food retailers in the United States: A scoping review. *Health & Place*, 83, 103089.

Smith, L.G., Ma, M.Y., Widener, M.J., and Farber, S., 2023. Geographies of grocery shopping in major Canadian cities: Evidence from large-scale mobile app data. *Environment and Planning B: Urban Analytics and City Science*, 50 (3), 723–739.

Squire, R.F., 2019. An Interactive Guide to Analyze Demographic Profiles from SafeGraph Patterns Data.

Supernak, J., 1967. Travel-time budget: a critique. *Transportation Science*, 1, 261–265.

Tadesse, G., Abate, G.T., and Zewdie, T., 2020. Biases in self-reported food insecurity measurement: A list experiment approach. *Food Policy*, 92, 101862.

The Food Industry Association, 2019. *U.S. grocery shopper trends*.

Todd, J.E. and Scharadin, B., 2016. *Where households get food in a typical week: Findings from USDA's FoodAPS*.

U.S. Census Bureau, 2022. American Community Survey 5-Year Estimates (2017–2021), Jacksonville city, Florida.

Ver Ploeg, M., Breneman, V., Farrigan, T., Hamrick, K., Hopkins, D., Kaufman, P., Lin, B.-H., Nord, M., Smith, T.A., Williams, R., and others, 2009. *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to congress*.

Ver Ploeg, M., Mancino, L., Todd, J.E., Clay, D.M., and Scharadin, B., 2015. *Where do Americans usually shop for food and how do they travel to get there? Initial findings from the National Household Food Acquisition and Purchase Survey*.

Wray, A., Martin, G., Doherty, S., and Gilliland, J., 2023. Analyzing differences between spatial exposure estimation methods: A case study of outdoor food and beverage advertising in London, Canada. *Health & Place*, 79, 102641.

Xie, H., Li, D., Wang, Y., and Kawai, Y., 2023. An early warning model of type 2 diabetes risk based on POI visit history and food access management. *PLOS ONE*, 18 (7), e0288231.

Xu, R., Huang, X., Zhang, K., Lyu, W., Ghosh, D., Li, Z., and Chen, X., 2023. Integrating human activity into food environments can better predict cardiometabolic diseases in the United States. *Nature Communications*, 14 (1), 7326.

Xu, Y., Zhao, X., Lovreglio, R., Kuligowski, E., Nilsson, D., Cova, T.J., and Yan, X., 2022. A highway vehicle routing dataset during the 2019 Kincade Fire evacuation. *Scientific data*, 9 (1), 608.

Zenk, S.N., Matthews, S.A., Kraft, A.N., and Jones, K.K., 2018. How many days of global positioning system (GPS) monitoring do you need to measure activity space environments in health research? *Health & place*, 51, 52–60.

Zenk, S.N., Schulz, A.J., Matthews, S.A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., Gibbs, K., Braunschweig, C., and Stokes, C., 2011. Activity space environment and dietary and physical activity behaviors: a pilot study. *Health & place*, 17 (5), 1150–1161.

Zhang, L., Misir, A., Boshuizen, H., and Ocké, M., 2021. A Systematic Review and Meta-Analysis of Validation Studies Performed on Dietary Record Apps. *Advances in Nutrition*, 12 (6), 2321–2332.

Zhao, K., Tarkoma, S., Liu, S., and Vo, H., 2016. Urban human mobility data mining: An overview. *In*: *2016 IEEE International Conference on Big Data (Big Data)*. 1911–1920.

Zhao, X., Xu, Y., Lovreglio, R., Kuligowski, E., Nilsson, D., Cova, T.J., Wu, A., and Yan, X., 2022. Estimating wildfire evacuation decision and departure timing using large-scale GPS data. *Transportation research part D: transport and environment*, 107, 103277.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.