City University of New York (CUNY)

CUNY Academic Works

Publications and Research

CUNY Graduate School of Public Health &
Health Policy

2020

Google-truthing to assess hot spots of food retail change: A repeat cross-sectional Street View of food environments in the Bronx, New York

Nevin Cohen
CUNY School of Public Health

Michael Chrobok
University of Toronto

Olivia Caruso
University of Toronto

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/sph_pubs/272

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

Google-truthing to assess hot spots of food retail change: A repeat cross-sectional Street View of food environments in the Bronx, New York

Nevin Cohen^{a,*}, Michael Chrobok^b, Olivia Caruso^c

^a CUNY Graduate School of Public Health and Health Policy, Department of Health Policy & Management, 55 W 125th Street, Room 605, New York City, New York, 10027, United States

^b Department of Geography & Planning, University of Toronto, 100 St. George Street, Room 5047, Toronto, Ontario, M5S 3G3, Canada

^c Health Studies, University of Toronto, 15 King's College Circle, Room H012, Toronto, Ontario, M5S 3H7, Canada

ARTICLE INFO

Keywords:

Food environment
Food access
Food retail
Ground-truthing
Hot spot analysis
Urban planning

ABSTRACT

Google Street View (GSV) images can be used to “ground-truth” current and historical food retail data from approximately 2007 - when GSV was launched in a few US cities - to the present, facilitating analyses of food environments over time. A review of GSV images of all food retailers listed in a government database of licensed establishments in the Bronx, New York enabled records to be verified, businesses classified, and retail change quantified. The data revealed several trends likely to affect food access and health: increasing overall numbers of food retailers; the growth of dollar stores; and numerous openings, closings, and ownership changes across all food retail segments. Hot spot analysis identified statistically significant clusters of new dollar stores and bodegas, purveyors of less healthy processed foods, in lower-income neighborhoods in the South Bronx that face elevated rates of diet-related diseases. This article demonstrates the benefits and limitations of using GSV to conduct “virtual” food environment research.

1. Introduction

Over the past few decades, governments have implemented varied supermarket incentive programs, from low-cost financing to zoning bonuses, to increase access to healthy food and reduce diet-related chronic diseases in underserved communities (Chrisinger, 2015; Rosenberg and Cohen, 2018). These policies have been based primarily on studies of the distribution of food retail establishments that identify disparities in their density or proximity to vulnerable populations (United States Department of Agriculture USDA, Economic Research Service, 2009; Walker et al., 2010).

Analyses of changes to food environments over time have been less common, but a growing body of research using repeat cross-sectional or longitudinal methods has examined temporal trends in the prevalence of “healthy” or “unhealthy” food retailers, and changes in retail distribution by socioeconomic status or race and ethnicity of surrounding communities (Berger et al., 2019; Filomena et al., 2013; James et al., 2017; Kolak et al., 2018; Larsen and Gilliland, 2008; Luan et al., 2015; Maguire et al., 2015; Mundorf et al., 2015; Ohri-Vachaspati et al., 2019; Richardson et al., 2014; Rummo et al., 2017).

The results of this research have been inconsistent (Lucan, 2015).

Some studies have documented overall increased supermarket density over time (Filomena et al., 2013; Maguire et al., 2015; Richardson et al., 2014), while others have shown declining numbers of supermarkets and greater spatial dispersion, linked to grocery retail consolidation and the suburbanization of supermarkets and populations (Bedore, 2013; Larsen and Gilliland, 2008; Semple and Giguere, 2018).

Attempts to measure associations between socioeconomic status and changes to food environments over time have been similarly inconsistent. Several studies have noted increasing numbers of fast food outlets and convenience stores in low-income neighborhoods (Berger et al., 2019; Maguire et al., 2015; Richardson et al., 2014; Rummo et al., 2017). Others have documented the more rapid growth of fast food in wealthier neighborhoods (James et al., 2017) and decreasing numbers of convenience stores in low and medium-income communities (Ohri-Vachaspati et al., 2019). Research on supermarkets have identified positive (Kolak et al., 2018; Ohri-Vachaspati et al., 2019), negative (Richardson et al., 2014), and no relationships (Larsen and Gilliland, 2008; Maguire et al., 2015) between access and socioeconomic status over time.

One of the challenges of researching food environment change has been the dearth of accurate, reliable data on food retailers. Researchers

* Corresponding author.

E-mail addresses: nevin.cohen@sph.cuny.edu (N. Cohen), michael.chrobok@mail.utoronto.ca (M. Chrobok), olivia.caruso@mail.utoronto.ca (O. Caruso).

<https://doi.org/10.1016/j.healthplace.2020.102291>

Received 9 May 2019; Received in revised form 3 January 2020; Accepted 21 January 2020

Available online 10 February 2020

1353-8292/© 2020 Elsevier Ltd. All rights reserved.

Table 1

Market segments and classification criteria of food retailers.

Market Segment	Description of Classification Criteria
Bakeries, Meat or Fish Markets, & Produce Markets	Includes stores that primarily sell baked goods; meat, fish, or seafood; or fruits and/or vegetables, respectively. Operating banners and signage of stores in this segment often explicitly reference the types of goods being sold (i.e., 'bakery', 'butcher', 'produce', or 'fish market').
Bodegas, Small Corner Grocery Stores, & Delis	Includes stores with a relatively small footprint that sell a limited range of grocery items, deli sandwiches, beer and soda, cigarettes, candy, and/or lottery tickets. Operating banners and signage of stores in this segment often explicitly reference the types of goods being sold (i.e., 'beer and soda', 'cigarettes', and/or 'deli').
Dollar/Discount Stores	Includes stores that sell discounted goods or products for \$1. Operating banners of stores in this segment often include '99 cents' or 'dollar'.
Major Convenience Chain Stores & Gas Station Snack Shops	Includes stores that sell snacks and lottery tickets, with fewer grocery items being sold than at a bodega, small corner grocery store, or deli. Stores in this segment belong to a major regional or national chain, and are often (but not exclusively) located at a gas station.
Mass Merchandise, Department, Club Format, & Wholesale Stores	Includes stores that primarily sell non-food items but have grocery departments or sell wholesale goods to the public.
Other Specialty Food Stores	Includes stores that specialize in a specific set of products that are not bread, meat, fish, or produce (i.e., cheese, chocolate, pasta, and spice shops; gift basket retailers; and health food, vitamin, and supplement stores).
Pharmacies	Includes stores that sell primarily pharmaceutical goods (i.e., medication), but also food, as indicated by an A&M food retail license. Operating banners of stores in this segment often explicitly reference the types of goods being sold (i.e., 'pharmacy' or 'drug').
Supermarkets	Includes stores that sell a full range of grocery items. Stores in this segment have larger footprints than bodegas, small corner grocery stores, and delis.

and policymakers often use business data from secondary sources, either proprietary commercial business lists or government license and inspection databases (Fleischhacker et al., 2013). Although these data are easily accessible, and in the case of government sources are free, ground-truthing indicates that both public and private datasets are frequently inaccurate, incomplete, or out of date (Lake et al., 2010; Liese et al., 2013; Wong et al., 2017). One study found that a commercial list of food retailers undercounted food stores by 16% and an Internet-based list missed 34% of existing retailers (Paquet et al., 2008). Liese et al. (2010) observed that 37%–39% of the food retailers in a region in South Carolina were not recorded in two common business databases sold by Dun & Bradstreet and InfoUSA. A study in the Bronx, New York ground-truthed food retail data obtained from Infogroup (formerly InfoUSA) and found that for strict matches (i.e., businesses with the same or consistent names), the business list only identified existing grocers 25% of the time, and identified only a third of specialty food stores, with a positive predictive value of just 45.5% even after data cleaning (Lucan et al., 2013).

Measurement errors and biases occur in commercial databases due to systematic undercounting of specific types of stores. More than 10% of businesses opt out of being listed in commercial directories, and these are often independent stores like ethnic markets that are important sources of groceries in low-income and immigrant communities (Hosler and Dharssi, 2010). Smaller grocers that open and close, or change owners, more frequently than national chains are undercounted as well

Fig. 1. Google Street View images of 1384 Nelson Ave., Bronx, NY, showing a parking lot circa 2008 (top) and a full-service grocer in 2017 (bottom). This address record was coded as a “New Store” in the “Supermarket” segment.

Table 2

Changes in Bronx food retail by market segment, 2008 to 2017.

Market Segment	Number of Stores, 2008	Number of Stores, 2017	Percent Change, 2008–2017
Bakeries, Meat or Fish Markets, & Produce Markets	258	262	1.55%
Bodegas, Small Corner Grocery Stores, & Delis	1,498	1,660	10.81%
Dollar/Discount Stores	119	171	43.70%
Major Convenience Chain Stores & Gas Station Snack Shops	58	65	12.07%
Mass Merchandise, Department, Club Format, & Wholesale Stores	7	10	28.57%
Other Specialty Food Stores	30	26	–13.33%
Pharmacies	74	80	8.11%
Supermarkets	161	190	18.01%
^a Not Classified	26	22	–15.38%
TOTAL	2,231	2,486	11.43%

^a Denotes stores that could not easily be classified due to an absence of clear GSV imagery or because these retailers fell into multiple categories.

(Caspi and Friebur, 2016; Powell et al., 2011). Government records are often more comprehensive than private databases, and generally include smaller stores missed by the commercial lists, but are not necessarily complete or accurate. One study found that the food retailers identified in eight government agency lists covering a single community varied significantly from list to list (Hosler and Dharssi, 2010). In addition to gaps in store coverage, private and public databases frequently have inaccurate information due to self-reported addresses, data entry errors, and insufficient error-checking. Errors may include incorrect street addresses and geocoordinates, misspellings, and missing data fields. In one study, the geocoordinates of food retailers correctly identified only about half of the stores within 100 m of their actual locations (Liese et al., 2010). Data cleaning can address some, but not all of these errors, and requires significant effort.

Food retail studies based on inaccurate retail data may produce

Table 3
Changes in food retail segments, 2008 to 2017, by number of stores and type of change.

Food Retail Segment	2008	Segment Stability		Segment Loss		Segment Gain		2017
	Number of Food Retailers	Continuously Operating (No Change)	Replacement in Same Segment	Closed, No Food Retail Replacement	Replaced by Food Retailer in Different Segment (13)	Opened, No Previous Food Retailer	Replaced Food Retailer in Different Segment	Number of Food Retailers
Bakeries, Meat or Fish Markets, & Produce Markets	258	117	59	(69)		73	13	262
Bodegas, Small Corner Grocery Stores, & Delis	1,498	385	866	(222)	(25)	390	19	1,660
Dollar/Discount Stores	119	45	40	(25)	(9)	76	10	171
Major Convenience Chain Stores & Gas Station Snack Shops	58	36	20	(2)	0	7	2	65
Mass Merchandise, Department, Club Format, & Wholesale Stores	7	5	0	(2)	0	5	0	10
Other Specialty Food Stores	30	17	0	(13)	0	9	0	26
Pharmacies	74	59	2	(12)	(1)	15	4	80
Supermarkets	161	88	62	(9)	(2)	33	7	190
Not Classified	26	4	2	(10)	(10)	11	5	22
TOTAL	2,231	756	1,051	(364)	(60)	608	60	2,486

Fig. 2. Bodegas, corner stores, and delis opened between 2008 and 2017. Each dot represents the geocoded location of one new store.

misleading results (Kelly et al., 2011). Estimating the number of supermarkets in an area by sorting retailers by business names can be inaccurate because store names may include descriptors that misidentify business type. Bodegas and convenience stores often contain variations of the term “supermarket” in their operating banners, risking misclassification as full-service grocery retailers. Some researchers classify supermarkets by store size, yet store footprint data can also be unreliable, particularly if self-reported or if based on property data for buildings that have unusual configurations or house multiple tenants. Size cutoffs for supermarkets based on industry averages can underestimate their prevalence in expensive real estate markets in which full-service supermarkets may occupy the footprint of a typical convenience store, or overestimate their prevalence in communities where low rents enable

Fig. 3. Statistically significant cluster of new bodegas, corner stores, and delis opened between 2008 and 2017, and percentage of households below the poverty level by census Public Use Microdata Area (PUMA).

Table 4
Diet-related disease prevalence in part of the South Bronx (PUMA 3710/Community Districts 1 & 2) compared to Bronx County and New York City, in percent of adults with each condition. Source: NYC DOHMH, Community Health Survey, 2015–2016.

	South Bronx neighborhoods of Hunts Point, Longwood, Mott Haven & Melrose (PUMA 3710/Bronx Community Districts 1 & 2)	Bronx County	New York City
Obesity	42%	32%	24%
Diabetes	20%	16%	11%
Hypertension	38%	36%	28%

Fig. 4. Dollar/discount stores opened between 2008 and 2017, with a statistically significant hot spot. Each dot represents the geocoded location of one new store.

convenience stores to be larger. These factors may lead to misclassifications that are biased. For example, a study of the classification of food stores by commercial business lists compared to ground-truthed retail data found that convenience stores were more likely to be misclassified as full-service grocery stores in communities of color (Han et al., 2012).

Ground-truthing, physically canvassing the streets within a geographic area, often reveals that a significant portion of existing businesses differ from those listed in commercial or government databases. For example, Caspi and Frieur (2016) ground-truthed food retailers by driving along each street in their study area to log the geocoordinates of each establishment, photograph the storefront, and conduct a windshield survey of store name, address, and hours of service. They found that of the 136 stores identified, 34% were not in ESRI's Business Analyst database, composed of Dun & Bradstreet data, and that 45% of the stores on the ESRI list either did not exist or were not actually food retailers (Caspi and Frieur, 2016).

To reduce the time and cost of physical surveys, researchers have begun to verify retailers using virtual representations of streetscapes, like the images in databases such as Google Street View (GSV), which has compiled photographs of streets in major cities since approximately 2007 (Chau, 2007). This technique is particularly useful for large areas that would be expensive or time-consuming to survey in person (Pliakas et al., 2017). GSV has been used to measure various environmental features such as transportation infrastructure, land uses, vegetation, commercial businesses, sidewalk activity, and street conditions. Studies have found consistency between in-person observations of streetscape

features and assessments made by viewing GSV images, establishing the validity of GSV audits (Hwang and Sampson, 2014; Kelly et al., 2013; Rundle et al., 2011; Vargo et al., 2012). In particular, the method has been found to be reliable for identifying food retailers (Clarke et al., 2010; Rossen et al., 2012; Wilson et al., 2012).

This paper illustrates that GSV can also be used to ground-truth historical streetscape data and to measure changes over time. Prior to 2014, GSV images recorded at different times were woven together, limiting GSV's use in documenting the continuity or dynamics of environmental conditions in a particular place (Curtis et al., 2013; Gibbs, 2014). Since 2014, GSV has included a feature called "Time Machine" that enables viewers to see historical, time-stamped images for specific months and years between 2007 and the present. Researchers have begun to use these historical images to track changes in the physical environment (Cândido et al., 2018; Schootman et al., 2016). However, GSV has not commonly been used to analyze how health-related aspects of the built environment have changed over time (Rzotkiewicz et al., 2018). As we illustrate in the following case study, historical GSV images facilitate verification of the number and types of food establishments in an area over time, and can show whether and in what ways the food environment in a community has changed.

2. Methods

2.1. Case selection

To illustrate GSV's usefulness in repeat cross-sectional food

Fig. 5. Supermarkets opened between 2008 and 2017. Each dot represents the geocoded location of one new store.

environment research, we analyzed government food retail data in the Bronx, one of New York City's five boroughs, from 2008 and 2017. We used GSV images to visually verify food retailers and identify changes to stores (including openings, closings, and shifts in brand, type, or ownership) that occurred between these years.

We focused on food retail change in the Bronx for three reasons: (1) the borough has significantly higher rates of obesity and diet-related chronic diseases than the rest of the city (NYC Department of Health and Mental Hygiene, 2015); (2) much of the borough is eligible for incentives under New York City's Food Retail Expansion to Support Health (FRESH) program, which offers tax abatements and zoning bonuses to spur supermarket construction, renovation, and expansion (NYC Department of Health and Mental Hygiene, 2014); and (3) economic growth and rezoning in the borough has attracted new development and more affluent residents (Austensen et al., 2016), raising fears that gentrification may have affected food retailers.

We chose 2008 and 2017 as our study periods because we had ready access to a food retail dataset from 2008, and 2007 GSV images were available by 2008. At the start of our research in early 2018, the most up-to-date government dataset and GSV images were from 2017.

2.2. Data cleaning and joining

We obtained publicly-available food establishment data for the years 2008 and 2017 from the New York State Department of Agriculture and Markets (A&M), which licenses food retailers in New York State, and filtered the records to create lists of stores in Bronx County for each year. The A&M database required substantial data cleaning due to misspellings, typographical errors, inconsistent abbreviations, and address aliases. Data cleaning involved four steps: (1) formatting addresses for spelling and street nomenclature consistency; (2) deduplication of

address entries; (3) conversion of street aliases to standardized US postal addresses; and (4) verification of inconsistent or unclear addresses using Google Maps. We performed this work in Microsoft Excel.

We used Trifacta, a data wrangling application, to join the cleaned Bronx addresses from the 2008 and 2017 lists using street address as the primary join key and license number as a secondary key. The joined dataset contained the names and addresses of all licensed Bronx food retail establishments in 2008 and 2017, with links between those establishments listed at the same address in both years.

2.3. Classification scheme development

The authors collectively developed a classification scheme (i.e., rubric) based on a preliminary review of the joined dataset and our knowledge of food retail in New York City. We used the rubric to tentatively classify 300 address entries (approximately 10% of the joined dataset), discussed and modified the classification scheme, and developed strategies to incorporate GSV image data into the validation and coding of food retailers. This involved comparing descriptors in the A&M "entity" and "doing-business-as" names (e.g., "deli"; "fish store"; "supermarket") with the visual identity of the storefront and street-level references to the types of food sold (i.e., in overhead or window signs and displays). One co-author used the consensus rubric and classification methods (Table 1) to classify all retailers in the cleaned and joined dataset according to the procedures described below.

2.4. Data validation and categorization using observation of GSV imagery

To validate and categorize food retail establishments, we retrieved GSV images from 2008 to 2017 for each address in the cleaned and joined dataset. We noted missing images and consulted and compared

Fig. 6. Food retail closures across all market segments between 2008 and 2017. Each dot represents the geocoded location of a closed store.

close-in-time data (i.e., 2007 and 2009 photos in lieu of a 2008 image).

It was beyond the scope of our study to manually scan all street segments in Bronx County to determine if food retailers were present at addresses not listed in the joined dataset (i.e., absent from the A&M database in both study years). However, for some street addresses that had just one associated entry (i.e., 2008 or 2017 only), we checked the location on GSV to confirm whether food retailers were actually present in both study years. If so, we added these missing records to our joined dataset.

We then reviewed each GSV image to validate the store's banner (the name on the store's sign). Visually identifying the banner name was important because the "entity" (i.e., corporate) and "doing-business-as" names in the A&M database are, at best, approximations of the banner name, and often were completely different terms that would make it impossible to tally by company or brand.

Next, we used the GSV images to code the market segment of each food retailer by applying the classification rubric in Table 1. For images that did not clearly indicate the store type, we searched online to find store web pages or consumer reviews that described the type of establishment, and also considered the store size reported in the A&M database. We noted establishments that were clearly not food businesses.

2.5. Measuring food retail dynamics

Building on Filomena et al. (2013), for each street address in the joined A&M database, we identified any of the following changes that occurred between 2008 and 2017: (1) different owner or banner names; (2) changed retail segments; and (3) business closures, openings, and associated land use changes. We categorized each record as one of the following: "continuously operating store"; "replaced store - same

segment"; "replaced store - different segment"; "new store"; or "closed store."

Continuously operating stores were those with unchanged banners and market segments between 2008 and 2017. Replaced stores were new retailers in either the same or different food retail segment in 2017 that replaced a store that had been operating in 2008. New stores were operating in 2017 at locations that were not food stores in 2008. Closed stores were open in 2008 but had no food retailer present at that location in 2017. Stores with slightly different banner names in 2008 and 2017 (e.g., Bob's Quick Deli vs. Bob's Express Deli) were classified as continuously operating if GSV indicated no other differences except the name variation. Descriptive statistics of these changes are presented in the Results section.

For establishments that changed, we noted details from the GSV images that explained the nature of the change, including former or current land uses, façade changes, name changes, new construction, demolition, or building replacement. Examples include: "was a parking lot in 2008, now a supermarket" (see Fig. 1); "building was under development in 2008"; "store has closed and building sits vacant"; "building demolished for a new mixed-use development"; and "now a beauty salon."

2.6. Hot spot analysis

We conducted hot spot analyses in ArcGIS Online to test for statistically significant spatial clustering of food retail establishment openings, closings, and changes among retailers (i.e., name and license changes) that might indicate turbulence in the food retail environment or the concentration of particular types of retailers in specific neighborhoods. ArcGIS applies the Getis-Ord G_i^* test of spatial clustering to

Fig. 7. Food retailers in all market segments with name or license changes between 2008 and 2017. Each dot represents the geocoded location of one food retailer.

identify statistically significant clusters of a high or low value of phenomena of interest. We analyzed the spatial arrangement of retailers within a fishnet grid, normalized by ESRI-estimated 2019 population, to measure clusters of stores within the entire borough. All of the results presented in the Results section are significant at the $p < .01$ level.

3. Results

3.1. Discrepancies between GSV and A&M data

Google-truthing identified discrepancies between listings in the A&M database and images of food retailers in GSV. GSV showed 11 more retailers than the A&M database in 2008 and 68 fewer than A&M in 2017. The differences resulted from: (1) businesses being incorrectly identified as food establishments in the A&M database (e.g., a long-standing flat-tire repair shop listed as a deli); (2) closed businesses having active licenses; (3) duplicate entries; (4) separately-licensed food establishments existing within larger stores (e.g., a licensed sushi vendor within a supermarket); and (5) food establishments appearing in GSV but missing from the A&M database (74 in 2008 and 54 in 2017).

3.2. Food retail dynamics, 2008–2017

The Google-truthed data showed that the total number of Bronx food retail establishments grew from 2,231 in 2008 to 2,486 in 2017, an increase of 11.43%. The growth in food retailers was about twice the 5.7% growth in the borough's population, which increased from 1.39 million in 2008 to 1.47 million in 2017 (New York University Furman Center, 2019).

As Table 2 indicates, GSV revealed important differences within

specific food retail segments. Bodegas, small corner grocery stores, and delis remained the most common food retailers in the Bronx, with a net increase of 162 stores between 2008 and 2017. Yet, because of increases in other food retail types, their share of food retail remained stable at about 67%. Dollar and discount stores comprised the fastest-growing food retail segment in the Bronx, increasing 43.70% between 2008 and 2017, from 119 to 171. Dollar Tree had just one Bronx location in 2008 but 20 by 2017, and Family Dollar stores grew from 15 to 24. Despite the increasing grocery market share captured by dollar stores and other non-traditional food retailers, such as pharmacies, the number of supermarkets increased by 18.01% in the Bronx, from 161 in 2008 to 190 in 2017, and account for approximately two-thirds of all grocery expenditures in the borough (Food Trade News, 2018). The number of independent butchers, bakers, fish markets, and produce stands, often thought of as anachronisms, increased slightly from 258 in 2008 to 262 in 2017.

In addition to food retail growth, there were significant changes in store ownership and banners over the study period, as Table 3 illustrates. For example, though there was a net increase of 162 bodegas, small corner grocery stores, and delis between 2008 and 2017, 222 retailers in this segment closed and were not replaced; 866 were replaced by another store in the same category; 25 were lost to a different market segment; 19 were converted from another segment; and 390 opened on property that had not previously been used for food retail. This dynamism might reflect innovation and competition, changing rents, or business instability. The effects on consumers might be positive or negative: new owners and brands may be accompanied by store upgrades and new products that meet consumer demand, or they could be disorienting and dissatisfying for some, such as longtime residents who may view retail change as a harbinger of gentrification.

Fig. 8. Hot spot of food retailers with name or license changes between 2008 and 2017.

3.3. Hot spots of food retail openings

Fig. 2 presents the spatial distribution of new bodegas, corner stores, and delis that opened in the Bronx between 2008 and 2017, including openings on sites with no previous food retailer (not replacements of food retailers in the same or different market segment).

While the new stores are located throughout the borough, hot spot testing reveals a statistically significant cluster of bodegas, corner stores, and delis in the South Bronx. As Fig. 3 and Table 4 show, the South Bronx has a higher percentage of households below the poverty level than the rest of the borough; higher rates of poverty (29%) and unemployment (12%) compared to New York City as a whole (20% and 9%, respectively); and a higher diet-related disease prevalence than the borough and city.

Dollar and discount stores, like bodegas, market their products to low-income households. New dollar stores opened throughout the Bronx between 2008 and 2017, as Fig. 4 shows, but were also clustered in the South Bronx.

New supermarkets opened between 2008 and 2017 throughout the borough. These openings were not clustered in any census tracts, as Fig. 5 indicates.

3.4. Food retail closings

Food retailers closed throughout the borough between 2008 and 2017, as Fig. 6 illustrates. There were no statistically significant clusters of closed retailers, and there were too few closed supermarkets to test for spatial clustering of this market segment.

3.5. Hot spots of food retailers in transition

We tested for spatial clustering of food retailers in all retail segments that made a business transition between 2008 and 2017, either a change to the business' name or its license, suggesting a change in ownership, supplier, business model, or a possible effort to re-brand the business or attract different customers. These changes occurred throughout the borough, as Fig. 7 shows. There was a statistically significant cluster of retailers that changed their name or license in the South Bronx, as Fig. 8 illustrates.

4. Discussion and concluding remarks

Our Bronx study categorized food retailers in business in 2017 and approximately 10 years prior by applying a simple rubric to manual views of GSV images. We used this data to identify overall trends in food retail in the borough, document the growth or decline of specific types of food stores, and analyze spatial clustering of store openings, closings, and name or license changes. The results illustrate that GSV can be used to verify and clean retail databases, confirm business names and addresses, classify market segments, and identify missing or redundant data.

The study also showed how researchers can use historical images on GSV to verify the food retailers that existed in a geographic area, and to compare older and current images to understand the nature of retail change over time. GSV enables measuring such changes store by store, including detecting changes not recorded in government data, like store façade renovations, changes in brands, expansions, and openings, and closures that may not be reflected in a single cross-sectional dataset.

Moreover, while business databases may categorize establishments by NAICS code, these are usually too general to distinguish between store types such as fish markets and butchers, or ethnic grocers and national chains. These details, which can be detected through visual assessment of storefront images, can be critical to understanding the variety of food sold in a community, how accessible it is to different populations, and how these factors change over time.

GSV also can save time and the cost of in-person ground-truthing because it enables researchers to conduct historical cross-sectional studies or longitudinal research from the convenience of a computer screen. It took between 140 and 190 person-hours, or between 15 and 20 addresses per person-hour, to verify Bronx store names and addresses, classify stores by market segment, capture images and record the data, and triangulate with other data sources, such as retailers' websites. This effort is not insignificant, but not having to travel to field research sites to observe store façades reduces the time it takes to do this research by approximately 70 percent, according to one New York City study that compared using GSV to conducting conventional street audits (Bader et al., 2017). For researchers who wish to compare food environments in multiple cities, GSV may be the only feasible option. As text and image recognition software improves, it can potentially be used to automate this process, reducing costs.

Our study identified hot spots of retail change in the South Bronx showing clusters of new bodegas and dollar stores, and stores with changes to business names and licenses, but much more extensive analyses are possible with GSV images. Researchers can test for correlations among changes in food retail and real estate values, development trends, neighborhood socio-demographics, and other variables. It is also possible to assess the context of change by using GSV images to evaluate adjacent and nearby streets, or entire commercial corridors, to understand how changes in food retail are linked to the changing characteristics of other commercial sectors, transportation infrastructure, residential development, and other aspects of the built environment. We only examined data from two years, but our method allows researchers to examine finer-grain changes in retail that occur over shorter intervals.

GSV can also help researchers evaluate policies that aim to influence the food retail sector, such as supermarket incentive programs. For example, our study found that by 2017, 102 of the Bronx's 190 supermarkets were either completely new, having not existed as food retail in 2008 ($n = 33$), were replacements of pre-existing supermarkets ($n = 62$), or were replacements of other types of food stores ($n = 7$). By comparison, the Food Retail Expansion to Support Health (FRESH) incentive program has contributed to just 6 completed supermarket construction or renovation projects in the Bronx over the same time period, raising the question of the magnitude of the policy's impact on what has otherwise been a dynamic and growing part of the food retail sector.

The potential of using Google Street View to ground-truth food outlet data and to examine changes to retail environments over time is contingent upon both the *availability* and *quality* of published images. While our study had the advantage of being set in one of Google's earliest and most comprehensively photographed cities (Chau, 2007), images circa 2008 in the Bronx were less comprehensive than those taken in subsequent years. In verifying and coding Bronx address records, 8% of addresses did not have an associated 2007–09 image. By 2016–18, the percentage of addresses without a photograph dropped to approximately 1%. Image availability and quality varies based on road layouts that present travel choices for GSV data collectors (e.g., split-lane streets with express and collector routes or above-ground passages and tunnels), detours due to construction, shopping centers that are photographed only from the street, and obstructions such as delivery trucks that limit visibility of storefronts. Older images have lower resolution than those taken by newer cameras, and weather or time-of-day conditions (e.g., sun glare) can reduce image clarity. Seeking alternative vantage points or reference images, as we did here – for instance, by observing a street segment from the opposite direction or a nearby cross-street, and consulting curbside signage to identify

shops in malls – can mitigate some but not all of these limitations. Nonetheless, GSV is a useful tool for food environment research in locations with a record of repeat and extensive coverage, which currently include larger cities in the Global North (Weiner, 2019).

Beyond these technical issues, any visual assessment, whether real or virtual, requires careful verification and the use of multiple reviewers to increase reliability. Our small study relied on the assessment of a few individuals, and although guided by a rubric and the triangulation of different data points for greater reliability, the classification nonetheless involved interpretations of the GSV images that were not analyzed for inter-reviewer reliability. Furthermore, two-dimensional images on the computer cannot completely substitute for the ability to inspect food retailers in real life. GSV may not be appropriate for analyses that require images taken at exactly the same time, as the dates of images vary throughout the streetscape (Curtis et al., 2013). Furthermore, GSV data are owned by Google, which can limit future access.

Despite these limitations, GSV and other image databases can be valuable tools to study temporal change in a community's food retail sector, and historical images are indispensable for researchers who do not have reliable, detailed, historical business datasets. Data about the dynamics of food retailers, particularly information on whether, to what extent, and how a community's food establishments change over time, and where those changes occur, can improve policy effectiveness by facilitating targeted food retail incentives and reducing unnecessary subsidies. Detecting physical upgrades in stores and changes in brands may also provide an early warning to policymakers of "food gentrification", the process by which new retail investments and marketing efforts cater to more affluent shoppers, potentially decreasing affordability and access to longstanding residents (Cohen, 2018). Without accurate historical data showing changes in food retail, public health practitioners and policymakers risk "fighting the last war" by designing interventions based on outdated assumptions about food retail businesses. By developing new tools and methods to mine GSV and other sources of big data, public health researchers will be better able to document, analyze, and respond to food retail dynamics.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Declaration of competing interest

None.

References

- Austensen, M., Ellen, I.G., Herrine, L., Karfunkel, B., Jush, G.K., Yager, J., 2016. State of New York City's housing and neighborhoods in 2015. Retrieved from. http://furmancenter.org/files/sotc/NYUFurmanCenter_SOCin2015_9JUNE2016.pdf.
- Bader, M.D., Mooney, S.J., Bennett, B., Rundle, A.G., 2017. The promise, practicalities, and perils of virtually auditing neighborhoods using Google Street View. *Ann. Am. Acad. Polit. Soc. Sci.* 669 (1), 18–40. <https://doi.org/10.1177/0002716216681488>.
- Bedore, M., 2013. Geographies of capital formation and rescaling: a historical-geographical approach to the food desert problem. *Can. Geogr.* 57 (2), 133–153. <https://doi.org/10.1111/j.1541-0064.2012.00454.x>.
- Berger, N., Kaufman, T.K., Bader, M.D.M., Rundle, A.G., Mooney, S.J., Neckerman, K.M., Lovasi, G.S., 2019. Disparities in trajectories of changes in the unhealthy food environment in New York City: a latent class growth analysis, 1990–2010. *Soc. Sci. Med.* 234, 112362. <https://doi.org/10.1016/j.socscimed.2019.112362>.
- Cândido, R.L., Steinmetz-Wood, M., Morency, P., Kestens, Y., 2018. Reassessing urban health interventions: back to the future with Google Street View time machine. *Am. J. Prev. Med.* 55 (5), 662–669. <https://doi.org/10.1016/j.amepre.2018.04.047>.
- Caspi, C., Friebur, R., 2016. Modified ground-truthing: an accurate and cost-effective food environment validation method for town and rural areas. *Int. J. Behav. Nutr. Phys. Activ.* 13 (1), 37. <https://doi.org/10.1186/s12966-016-0360-3>.
- Chau, S., 2007. Introducing... Street View!. Retrieved from. <https://maps.googleblog.com/2007/05/introducing-street-view.html>.
- Chrisinger, B.W., 2015. If you build it, will they come, and what will they eat? Investigating supermarket development in food deserts. Retrieved from. <http://repository.upenn.edu/dissertations/AAI3709431>.

- Clarke, P., Ailshire, J., Melendez, R., Bader, M., Morenoff, J., 2010. Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. *Health Place* 16 (6), 1224–1229. <https://doi.org/10.1016/j.healthplace.2010.08.007>.
- Cohen, N.C., 2018, Mar. 27. Feeding or starving gentrification: the role of food policy. Retrieved from. <http://www.cunurbanfoodpolicy.org/news/2018/3/27/feeding-or-starving-gentrification-the-role-of-food-policy>.
- Curtis, J.W., Curtis, A., Mapes, J., Szell, A.B., Cinderich, A., 2013. Using Google Street View for systematic observation of the built environment: analysis of spatio-temporal instability of imagery dates. *Int. J. Health Geogr.* 12 (1), 53. <https://doi.org/10.1186/1476-072x-12-53>.
- Filomena, S., Scanlin, K., Morland, K.B., 2013. Brooklyn, New York foodscape 2007–2011: a five-year analysis of stability in food retail environments. *Int. J. Behav. Nutr. Phys. Activ.* 10 (1), 46. <https://doi.org/10.1186/1479-5868-10-46>.
- Fleischhacker, S.E., Evenson, K.R., Sharkey, J., Pitts, S.B., Rodriguez, D.A., 2013. Validity of secondary retail food outlet data: a systematic review. *Am. J. Prev. Med.* 45 (4), 462–473. <https://doi.org/10.1016/j.amepre.2013.06.009>.
- Food Trade News, 2018. 2018 market study. June 2018, Retrieved from. <https://best-met.com/issue/2018-market-study-2/>.
- Gibbs, S., 2014, Apr. 23. Google Introduces 'Time Machine' Feature in Street View. *The Guardian*. Retrieved from. <https://www.theguardian.com/technology/2014/apr/23/google-introduces-time-machine-feature-in-street-view>.
- Han, E., Powell, L.M., Zenk, S.N., Rinkus, L., Ohri-Vachaspati, P., Chaloupka, F.J., 2012. Classification bias in commercial business lists for retail food stores in the U.S. *Int. J. Behav. Nutr. Phys. Activ.* 9 (1), 46. <https://doi.org/10.1186/1479-5868-9-46>.
- Hosler, A.S., Dharssi, A., 2010. Identifying retail food stores to evaluate the food environment. *Am. J. Prev. Med.* 39 (1), 41–44. <https://doi.org/10.1016/j.amepre.2010.03.006>.
- Hwang, J., Sampson, R.J., 2014. Divergent pathways of gentrification: racial inequality and the social order of renewal in Chicago neighborhoods. *Am. Socio. Rev.* 79 (4), 726–751. <https://doi.org/10.1177/0003122414535774>.
- James, P., Seward, M.W., O'Malley, A.J., Subramanian, S., Block, J.P., 2017. Changes in the food environment over time: examining 40 years of data in the Framingham heart study. *Int. J. Behav. Nutr. Phys. Activ.* 14, 84. <https://doi.org/10.1186/s12966-017-0537-4>.
- Kelly, B., Flood, V.M., Yeatman, H., 2011. Measuring local food environments: an overview of available methods and measures. *Health Place* 17 (6), 1284–1293. <https://doi.org/10.1016/j.healthplace.2011.08.014>.
- Kelly, C.M., Wilson, J.S., Baker, E.A., Miller, D.K., Schootman, M., 2013. Using Google Street View to audit the built environment: inter-rater reliability results. *Ann. Behav. Med.* 45 (Suppl. 1), 108–112. <https://doi.org/10.1007/s12160-012-9419-9>.
- Kolak, M., Bradley, M., Block, D.R., Pool, L., Garg, G., Toman, C.K., et al., 2018. Urban foodscape trends: disparities in healthy food access in Chicago, 2007–2014. *Health Place* 52, 231–239. <https://doi.org/10.1016/j.healthplace.2018.06.003>.
- Lake, A.A., Burgoine, T., Greenhalgh, F., Stamp, E., Tyrrell, R., 2010. The foodscape: classification and field validation of secondary data sources. *Health Place* 16 (4), 666–673. <https://doi.org/10.1016/j.healthplace.2010.02.004>.
- Larsen, K., Gilliland, J., 2008. Mapping the evolution of “food deserts” in a Canadian city: supermarket accessibility in London, Ontario, 1961–2005. *Int. J. Health Geogr.* 7, 16. <https://doi.org/10.1186/1476-072x-7-16>.
- Liese, A.D., Barnes, T.L., Lamichhane, A.P., Hibbert, J.D., Colabianchi, N., Lawson, A.B., 2013. Characterizing the food retail environment: impact of count, type, and geospatial error in 2 secondary data sources. *J. Nutr. Educ. Behav.* 45 (5), 435–442. <https://doi.org/10.1016/j.jneb.2013.01.021>.
- Liese, A.D., Colabianchi, N., Lamichhane, A.P., Barnes, T.L., Hibbert, J.D., Porter, D.E., et al., 2010. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. *Am. J. Epidemiol.* 172 (11), 1324–1333. <https://doi.org/10.1093/aje/kwq292>.
- Luan, H., Law, J., Quick, M., 2015. Identifying food deserts and swamps based on relative healthy food access: a spatio-temporal Bayesian approach. *Int. J. Health Geogr.* 14, 37. <https://doi.org/10.1186/s12942-015-0030-8>.
- Lucan, S.C., 2015. Concerning limitations of food-environment research: a narrative review and commentary framed around obesity and diet-related diseases in youth. *J. Acad. Nutr. Diet.* 115 (2), 205–212. <https://doi.org/10.1016/j.jand.2014.08.019>.
- Lucan, S.C., Maroko, A.R., Bumol, J., Torrens, L., Varona, M., Berke, E.M., 2013. Business list vs ground observation for measuring a food environment: saving time or waste of time (or worse)? *J. Acad. Nutr. Diet.* 113 (10), 1332–1339. <https://doi.org/10.1016/j.jand.2013.05.011>.
- Maguire, E.R., Burgoine, T., Monsivais, P., 2015. Area deprivation and the food environment over time: a repeated cross-sectional study on takeaway outlet density and supermarket presence in Norfolk, UK, 1990–2008. *Health Place* 33 (Complete), 142–147. <https://doi.org/10.1016/j.healthplace.2015.02.012>.
- Mundorf, A.R., Willits-Smith, A., Rose, D., 2015. 10 years later: changes in food access and disparities in New Orleans since Hurricane Katrina. *J. Urban Health* 92 (4), 605–610. <https://doi.org/10.1007/s11524-015-9969-9>.
- New York City Department of Health and Mental Hygiene, 2014. Shop Healthy NYC year 1 evaluation report: West Farms and Fordham, Bronx, 2014. Retrieved from. <http://www.nyc.gov/html/doh/downloads/pdf/pan/shop-sealthy-report.pdf>.
- New York City Department of Health and Mental Hygiene, 2015, Mar. 26. Health outcomes in the Bronx improving. Retrieved from. <https://www1.nyc.gov/site/doh/about/press/pr2015/pr014-15.page>.
- New York University Furman Center, 2019. CoreData.nyc. Population by borough. United states census, American community survey and NYU furman center. Retrieved from. <http://coredata.nyc/>.
- Ohri-Vachaspati, P., DeWeese, R.S., Acciai, F., DeLia, D., Tulloch, D., Tong, D., et al., 2019. Healthy food access in low-income high-minority communities: a longitudinal assessment – 2009–2017. *Int. J. Environ. Res. Publ. Health* 16 (13), 2354. <https://doi.org/10.3390/ijerph16132354>.
- Paquet, C., Daniel, M., Kestens, Y., Léger, K., Gauvin, L., 2008. Field validation of listings of food stores and commercial physical activity establishments from secondary data. *Int. J. Behav. Nutr. Phys. Activ.* 5 (1), 58. <https://doi.org/10.1186/1479-5868-5-58>.
- Pliakas, T., Hawkesworth, S., Silverwood, R.J., Nanchahal, K., Grundy, C., Armstrong, B., et al., 2017. Optimising measurement of health-related characteristics of the built environment: comparing data collected by foot-based street audits, virtual street audits and routine secondary data sources. *Health Place* 43, 75–84. <https://doi.org/10.1016/j.healthplace.2016.10.001>.
- Powell, L.M., Han, E., Zenk, S.N., Khan, T., Quinn, C.M., Gibbs, K.P., et al., 2011. Field validation of secondary commercial data sources on the retail food outlet environment in the U.S. *Health Place* 17 (5), 1122–1131. <https://doi.org/10.1016/j.healthplace.2011.05.010>.
- Richardson, A.S., Meyer, K.A., Howard, A.G., Boone-Heinonen, J., Popkin, B.M., Evenson, K.R., et al., 2014. Neighborhood socioeconomic status and food environment: a 20-year longitudinal latent class analysis among CARDIA participants. *Health Place* 30, 145–153. <https://doi.org/10.1016/j.healthplace.2014.08.011>.
- Rosenberg, N.A., Cohen, N., 2018. Let them eat kale: the misplaced narrative of food access. *Fordham Urban Law J.* 45 (4), 1091–1120. <https://ir.lawnet.fordham.edu/ulj/vol45/iss4/5>.
- Rossen, L.M., Pollack, K.M., Curriero, F.C., 2012. Verification of retail food outlet location data from a local health department using ground-truthing and remote-sensing technology: assessing differences by neighborhood characteristics. *Health Place* 18 (5), 956–962. <https://doi.org/10.1016/j.healthplace.2012.06.012>.
- Rummo, P.E., Guilkey, D.K., Ng, S.W., Popkin, B.M., Evenson, K.R., Gordon-Larsen, P., 2017. Beyond supermarkets: food outlet location selection in four U.S. cities over time. *Am. J. Prev. Med.* 52 (3), 300–310. <https://doi.org/10.1016/j.amepre.2016.08.042>.
- Rundle, A.G., Bader, M., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using Google Street View to audit neighborhood environments. *Am. J. Prev. Med.* 40 (1), 94–100. <https://doi.org/10.1016/j.amepre.2010.09.034>.
- Rzotkiewicz, A., Pearson, A.L., Dougherty, B.V., Shortridge, A., Wilson, N., 2018. Systematic review of the use of Google Street View in health research: major themes, strengths, weaknesses and possibilities for future research. *Health Place* 52, 240–246. <https://doi.org/10.1016/j.healthplace.2018.07.001>.
- Schootman, M., Nelson, E., Werner, K., Shacham, E., Elliott, M., Ratnapradipa, K., et al., 2016. Emerging technologies to measure neighborhood conditions in public health: implications for interventions and next steps. *Int. J. Health Geogr.* 15 (1), 20. <https://doi.org/10.1186/s12942-016-0050-z>.
- Simple, H., Giguere, A., 2018. The evolution of food deserts in a small midwestern city: the case of Ypsilanti, Michigan: 1970 to 2010. *J. Plann. Educ. Res.* 38 (3), 359–370. <https://doi.org/10.1177/0739456x17702222>.
- United States Department of Agriculture (Usda), Economic Research Service, 2009. Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. USDA, Washington, DC.
- Vargo, J., Stone, B., Glanz, K., 2012. Google walkability: a new tool for local planning and public health research? *J. Phys. Activ. Health* 9 (5), 689–697. <https://doi.org/10.1123/jpah.9.5.689>.
- Walker, R.E., Keane, C.R., Burke, J.G., 2010. Disparities and access to healthy food in the United States: a review of food deserts literature. *Health Place* 16 (5), 876–884. <https://doi.org/10.1016/j.healthplace.2010.04.013>.
- Weiner, C., 2019. He's trying to fill in the gaps on Google Street View – Starting with Zimbabwe. Retrieved from. <https://www.npr.org/2019/09/22/760572640/hes-trying-to-fill-in-the-gaps-on-google-street-view-starting-with-zimbabwe>.
- Wilson, J.S., Kelly, C.M., Schootman, M., Baker, E.A., Banerjee, A., Glennin, M., Miller, D. K., 2012. Assessing the built environment using omnidirectional imagery. *Am. J. Prev. Med.* 42 (2), 193–199. <https://doi.org/10.1016/j.amepre.2011.09.029>.
- Wong, M.S., Peyton, J.M., Shields, T.M., Curriero, F.C., Gudzone, K.A., 2017. Comparing the accuracy of food outlet datasets in an urban environment. *Geospatial Health* 12 (1), 546. <https://doi.org/10.4081/gh.2017.546>.