



HHS Public Access

Author manuscript

Geospat Health. Author manuscript; available in PMC 2019 March 11.

Published in final edited form as:

Geospat Health. ; 12(1): 546. doi:10.4081/gh.2017.546.

Comparing the accuracy of food outlet datasets in an urban environment

Michelle S. Wong¹, Jennifer M. Peyton², Timothy M. Shields³, Frank C. Curriero³, and Kimberly A. Gudzone^{2,4}

¹Department of Health Policy and Management, Johns Hopkins School of Public Health

²Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University School of Medicine

³Department of Epidemiology, Johns Hopkins School of Public Health

⁴Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract

Studies that investigate the relationship between the retail food environment and health outcomes often use geospatial datasets. Prior studies have identified challenges of using the most common data sources. Retail food environment datasets created through academic-government partnership present an alternative, but their validity (retail existence, type, location) has not been assessed yet. In our study, we used ground-truth data to compare the validity of two datasets, a 2015 commercial dataset (InfoUSA) and data collected from 2012 to 2014 through the Maryland Food Systems Mapping Project (MFSMP), an academic-government partnership, on the retail food environment in two low-income, inner city neighbourhoods in Baltimore City. We compared sensitivity and positive predictive value (PPV) of the commercial and academic-government partnership data to ground-truth data for two broad categories of unhealthy food retailers: small food retailers and quick-service restaurants. Ground-truth data was collected in 2015 and analysed in 2016. Compared to the ground-truth data, MFSMP and InfoUSA generally had similar sensitivity that was greater than 85%. MFSMP had higher PPV compared to InfoUSA for both small food retailers (MFSMP: 56.3% vs InfoUSA: 40.7%) and quick-service restaurants (MFSMP: 58.6% vs InfoUSA: 36.4%). We conclude that data from academic-government partnerships like MFSMP might be an attractive alternative option and improvement to relying only on commercial data. Other research institutes or cities might consider efforts to create and maintain such an environmental dataset. Even if these datasets cannot be updated on an annual basis, they are likely more accurate than commercial data.

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Correspondence: Michelle S. Wong, Department of Health Policy and Management, Johns Hopkins School of Public Health, 624 N. Broadway, Baltimore, MD 21205, USA., Tel: +1.714.421.3122. mwong23@jhu.edu.

Contributions: MSW and KAG conceptualised the study. JMP conducted the ground truthing. MSW and KAG created the maps and conducted the statistical analysis. MSW drafted the manuscript. All authors reviewed and edited drafts of the paper and provided content for the final draft of the paper.

Conflict of interest: the authors declare no potential conflict of interest.

Keywords

Retail food environment; Urban food environment; Validity; Geographic Information Systems

Introduction

In recent years, there has been significant interest in the role of the neighbourhood retail food environment on diet and weight (Feng *et al.*, 2010). Retail stores that provide fresh fruits and vegetables, such as supermarkets, can promote a healthy diet, while food offered by other retailers, such as fast food restaurants, may lead to excess calorie consumption. Studies that investigate the relationship between the retail food environment and health outcomes often use geospatial information with geographic information systems (GIS) software and databases. Yet, the validity of the results depends upon whether or not retail store location and type are accurate within the data.

Two of the most common sources used to document the retail food environment are data collected through direct observation (*i.e.* *ground-truthed* data) and commercially available secondary data. Each has strengths and weaknesses (Fleischhacker *et al.*, 2013). Ground-truthed data are usually considered the *gold standard*, but collecting such data is labour- and cost-intensive, particularly for studies that cover wide geographic areas, and cannot be used for retrospective analyses (Powell *et al.*, 2011). Commercial datasets, such as those obtained from InfoUSA, Inc. (InfoUSA, 2016) or Dun & Bradstreet, Inc. (Dun&Bradstreet, 2016) compile information from a variety of sources, including business listings and Yellow Page listings. Commercial data come in an easy-to-use format, cover large geographic regions and can provide historical information. However, there are concerns about data accuracy. Because these data sources are developed for marketing purposes, they may not capture small, independently-owned businesses. Previous studies report sensitivities (*i.e.* probability that the dataset correctly identifies stores) ranging from 0.20 (slight) to 0.99 (almost perfect) and positive predicted values (*i.e.* probability that the stores listed in the dataset actually exist) ranging from 0.39 (fair) to 1.00 (perfect in principle), and suggest that InfoUSA's data may be slightly better than those of Dun & Bradstreet (Fleischhacker *et al.*, 2013).

As an alternative to these data sources, several research institutions have formed collaborations with local government agencies to create robust data sources of the retail food environment. For example, the Center for a Livable Future (CLF) at the Johns Hopkins School of Public Health has developed and maintains the Maryland Food Systems Map Project (MFSMP) (Center for a Livable Future, 2015). The MFSMP is a publically available, web-based mapping tool that has GIS-supported data on Maryland's food system, including information about retail food outlets. MFSMP is specifically maintained for research purposes and includes food store type and location as well as information that may be of interest to researchers but are not available in commercial datasets. For example, information about whether retailers accept Supplemental Nutrition Assistance Program (SNAP) benefits is available in the MFSMP but not in commercial datasets. Data for the MFSMP come from a variety of sources including the Baltimore City Health Department (BCHD) and food environment assessments conducted by the CLF. However, maintaining the MFSMP dataset

is labour-intensive, which results in CLF only updating this dataset every few years. Another example is Ohio State University's Mapping the Food Environment Project. (<http://foodmapping.osu.edu/>) This project is working to create a publically-available GIS data hub to facilitate research on the food environment. Data come from both primary and secondary data sources, such as the United States Department of Agriculture (USDA)'s food environmental atlas. However, the accuracy of these data compared to that of commercial sources is unknown.

Our study aimed to assess the validity of two datasets compared to 2015 ground-truthed data, a 2015 commercial dataset (InfoUSA) and data from the MFSMP collected from 2012 through 2014, on the retail food environment in two low-income, inner city neighbourhoods in Baltimore City, MD, USA. Despite MFSMP data being older than the commercial data, we hypothesised that MFSMP data would be more accurate than the commercial dataset, as we suspect that it will better capture small, independently owned food retailers that are common in inner city environments.

Materials and Methods

Study region

This assessment was part of a larger study assessing the built and social environments surrounding two public housing developments in Baltimore City. One development is located in East Baltimore, in the Oldtown/Middle East neighbourhood, and the other is located in West Baltimore, in the Sandtown-Winchester neighbourhood. Both neighbourhoods are located within food deserts (Buczynski *et al.*, 2015) where most residents are black and belong to the low-income bracket. The median household incomes are \$14,000 and \$24,000 for Oldtown/Middle East and Sandtown-Winchester, respectively (Baltimore Neighborhood Indicators Alliance - Jacob France Institute, 2015) Our analysis area included retailers within a 0.75-mile radius of the centroid of each public housing development.

Datasets

We compared three different data sources: ground-truthed, commercial and academic-government partnership.

Ground-truthed data

We conducted a ground-truth assessment in September 2015 of all food retail stores within a 1-mile radius of the centroid of each public housing development. We chose a zone that was larger than our 0.75-miles analysis area to decrease variability that might occur from *edge effects* (*i.e.* our ground-truth efforts might miss some stores at or just beyond the boundary). If we had only covered the 0.75-mile analysis area, we might conclude that the comparison datasets – which were limited to retailers within a 0.75 mile radius – included false positives (*i.e.* comparison datasets included retailers that did not actually exist), when in fact, these retailers do exist, but we missed them in our ground-truth assessment because they were located near the edge of the 0.75-buffer.

A trained observer systematically canvassed the area by car to collect data visible on the exterior of all food establishments. The observer did not enter into the stores but listed the names of outlets, addresses (or intersection if the address was unable to be determined) and classified the food retailer type. Categories included grocery stores, corner stores, convenience stores, chain fast food restaurants, and takeout restaurants. As done in previous validation studies, we excluded from our analysis stores that primarily sold liquor or were primarily bars (n=37) (Powell *et al.*, 2011). The observer determined this aspect based on establishment name (*e.g.* liquor or bar). Any missing data (*e.g.*, missing address or clarification regarding food retailer classification) from the initial ground-truth assessment was addressed during a follow-up assessment in April 2016 by two trained observers, using the same methodology as the initial one.

Commercially available data

These data on food retailers came from InfoUSA covering the year 2015 (InfoUSA, 2016). This provider contains information about US businesses including address, size of business, and North American Industry Classification System (NAICS) designations. InfoUSA obtains data about retailers from various sources, including Yellow Page directories and corporate websites. We included businesses with the following NAICS codes in our analysis: convenience stores (445120), grocery stores (445110), limited-service restaurants (722513) and full-service restaurants (722511).

Academic-government partnership data

The MFSMP was developed and is maintained by the Johns Hopkins Bloomberg School of Public Health Center for a Livable Future (2015). Data on the locations of food retailers in Baltimore City were originally derived from the BCHD's food permit list from August 2011. CLF regularly updates this list based upon store closings and changes. In the summer of 2012, they conducted a food store survey in Baltimore City. Two trained observers went into all stores on the BCHD's food permit list to verify the store name, existence and location. They also assessed each store's health food availability index (HFAI), a score based on the store's availability of healthy foods, and determined whether stores accepted SNAP and the special supplemental nutrition program for women, infants, and children (WIC) benefits. Observers also noted stores that had been closed or renamed and added new stores that were not on the BCHD's food permit list. CLF used this information to update their food retailer database, which classifies stores based on BCHD listing, industry standards, and CLF's own research. Our analysis included the following MFSMP categories: chain fast food outlets, carry-outs, corner stores, convenience stores, behind-glass-stores (*i.e.* subset of corner stores commonly found in lower-income communities characterised by having a Plexiglas barrier that separates customers from retail items and the store worker/owner) and small grocery stores. We used the most recently available data for each of these categories. The data for the chain fast food outlets, carryouts, and remaining types of store were updated in 2013, 2014, and 2012, respectively.

Study variables

We examined two broad categories of outlets that typically offer what is summarily called unhealthy food: *small food retailers* and *quick-service restaurants*. Detailed definitions for

both are available in Table 1. Stores that sell packaged snack foods and beverages that are high in calories and poor in nutrition (*e.g.*, soda, chips) were considered *small food retailers* in this study. For the ground-truthed data, we classified corner stores and convenience stores as *small food retailers*. The commercial data do not differentiate between smaller, independently-owned grocery stores, many of which are considered corner stores in Baltimore City, from larger chain supermarkets. We identified these smaller stores if they had a NAICS code designation of grocery store and had less than four employees and classified these select small grocery stores also as *small food retailers*. For the academic-government partnership data, we classified corner stores, convenience stores, behind-glass-stores and small grocery stores as small food retailers. We defined quick-service restaurants as outlets that only sell calorie-dense foods prepared on the premises, which patrons typically consume as *take-out*. For the ground-truthed data, we classified chain fast food and take-out restaurants as *quick-service restaurants*. For the commercial data, limited-service restaurants were closest to what we defined as quick-service restaurants. However, a substantial portion of these outlets was classified as full-service restaurants. A standardised definition distinguishing between limited and full-service restaurants does not exist. Thus, in our analysis, quick-service restaurants included both businesses with either a primary or secondary NAICS designation of limited-service restaurants (NAICS code: 722513) or a primary NAICS designation of full-service restaurant (NAICS code: 722511) and had the following key words in the business name: carryout/carry-out, chicken, trout, Chinese/China and pizza/pizzeria. We selected these key words because they are frequently included in restaurant names for take-out type restaurants (*e.g.*, *Fried Chicken take-out*). For the academic-government partnership data, we classified carry-out and fast food chain restaurants as quick-service restaurants.

Mapping and statistical analysis

We used ArcGIS, version 10.3 (ESRI, Redlands, CA, USA) to map store locations. For the ground-truthed data, we used the business address to geocode (*i.e.* link street address to an electronic street map) outlet location. All locations were successfully geocoded (69% automatically). Remaining addresses were manually geocoded with 14% that needed to be geocoded as the nearest intersection. For InfoUSA, we used latitude and longitude information to plot the store location. MFSMP data were already geocoded and in shapefile form from CLF. We used ArcGIS to create maps for both food retailer categories. All stores from the InfoUSA and MFSMP comparison datasets with the classifications of interest within a 0.75-mile radius of the centroid of each public housing development were selected for further analysis.

We identified store matches based on i) stores having the same name and address, and ii) stores with different names but located at the same address and being the same category of store.

To compare the accuracy of MFSMP's and InfoUSA's datasets to the ground-truthed data, we calculated sensitivity (*i.e.* the proportion of food outlets listed in both the ground-truth dataset and the comparison datasets out of all relevant businesses from the comparison datasets) and positive predictive value (PPV) (*i.e.* the proportion of food outlets listed in

both the ground-truth dataset and the comparison dataset out of all relevant businesses in the ground-truth dataset) for both retail food categories. We did not calculate negative predictive value or specificity because we did not identify *true negative* stores (*i.e.* stores that do not belong to either retail category of interest) in the ground-truth assessment. Because there might be potential misclassification of store type, we also assessed sensitivity and PPV for a combined list of both quick-service restaurants and small food retailers. Statistical analyses were conducted using Stata, version 14/IC (StataCorp, College Station, TX, USA)

Results

Figure 1 shows the locations of the two low-income, inner city neighbourhoods in Baltimore City communities, selected as study areas, and the surrounding retail environment. While, visually, there appears to be some overlap between all three data sources, there are also areas where only two data sources overlapped. These included the areas south of the West Baltimore community and west of the East Baltimore community: there was some overlap between MFSMP and the ground-truth assessment with reference to quick-service restaurants, but not including InfoUSA.

Ground-truth assessment identified most small food retailers and quick-service restaurants within our study areas (Figure 2), followed by MFSMP data and InfoUSA in that order. Table 1 provides counts of small food retailers and quick-service restaurants by data source specific-categories. Most of the small food retailers were identified as corner stores in the ground-truthed and MFSMP datasets, and as small grocery stores in InfoUSA. Most of the quick-service restaurants were identified as take out restaurants in the ground-truthed data, as carry-outs in the MFSMP, and were more evenly split between limited service restaurants and full-service restaurants in InfoUSA.

Figure 3 presents validation statistics for both data sources by food retailer category. Compared to the ground-truthed data, MFSMP and InfoUSA had a sensitivity of 91.6 and 84.6% for small food retailers, respectively. Sensitivity for quick-service restaurants was similar for both data sources. MFSMP had a higher PPV than InfoUSA for both retailer categories. When we combined both food retailer types, sensitivity and PPV improved for InfoUSA, but were similar for MFSMP.

Discussion

This study is the first to compare the accuracy of data from an academic-government partnership and a commercial source to the *gold standard* of ground-truth data in low-income, inner city communities. Both data sources had high degrees of sensitivity (>80%). However, the academic-government partnership data (MFSMP) had higher PPV than the commercial data source (InfoUSA) in these communities. If a store is listed in one of these secondary data sources, it is likely to actually exist. However, it is important to note that both datasets likely only capture a fraction of all small food retailers and quick-service restaurants that exist in low-income, inner city communities.

Other studies have calculated similar sensitivity estimates for commercial data as we obtained in our study (Paquet *et al.*, 2008; Liese *et al.*, 2010; Han *et al.*, 2012; Rossen *et al.*,

2012) but our PPVs were lower than those estimated in studies conducted in urban areas (Paquet *et al.*, 2008; Liese *et al.*, 2010; Han *et al.*, 2012; Rossen *et al.*, 2012; Lucan *et al.*, 2013). One study has found that neighbourhood characteristics, such as the proportion of blacks in the neighbourhoods (Han *et al.*, 2012) are associated with a lower accuracy in secondary data sources. Our communities were predominantly black, so this may explain our lower PPVs. However, not all studies have found these differences in data accuracy (Bader *et al.*, 2010). These differing results may suggest that these factors vary by geography. Thus, when possible, local efforts, such as those of the MFSMP, are likely to be a more accurate option for measuring the retail food environment.

Prior GIS validation studies have also assessed the validity of data from government agencies. City or state health departments maintain food retail listings for licensing and inspection purposes. These data sources have generally been found to be more accurate than commercial data sources and are less likely to systematically omit small, independent businesses (Fleischhacker *et al.*, 2013). This is consistent with our findings that the MFSMP data are more accurate than InfoUSA as BCHD data was one of the data sources for the MFSMP. While we did not specifically compare MFSMP and BCHD data, we believe that the MFSMP data likely address some aspects of weakness of government data. Inaccuracies in health department data may arise due to lists being out of date (*e.g.*, business closure) or the store type may be listed incorrectly (Lyseen and Hansen, 2014). CLF periodically updates the BCHD food retail listings based upon their own assessments. Additionally, government data are not created for research purposes and may require substantial investments of time to reformat the data so they can be used, while MFSMP data are publically available and can be downloaded in an easily useable format.

Although the MFSMP data was older – up to 3 years old – this did not seem to substantially affect accuracy of stores listed. This might be due to how we identified matches: stores that had a different name but were of the same type and in the same location were considered a match. While there were some store closures, we found that new stores within the same categories were likely open in the same location. We believe that data that can identify the type of store in a particular location is sufficient for studying the food environment.

One challenge of using InfoUSA was selecting the appropriate NAICS codes. Previous research has used a variety of NAICS codes, including convenience stores, fast food restaurants and pizza restaurants, general merchandise stores (Fleischhacker, *et al.*, 2013). We noticed that some of the ground-truthed locations actually did exist in the InfoUSA dataset but were classified by other NAICS codes that we did not consider, including all those referred to as *other general merchandise stores*, and *full-service restaurants*. We were cautious in using the codes we considered to maintain the trade-off between PPV and sensitivity. Inclusion of more codes can improve the PPV, but may also decrease sensitivity. Identifying quick-service restaurants in InfoUSA was particularly challenging. Previous studies have noted similar issues with identifying quick-service restaurants, especially chain fast food outlets, in commercially available data (Sturm, 2008; Powell *et al.*, 2011). The most commonly used methods for identifying these food outlets was through the primary NAICS code for limited-service restaurants. However, we found that using only this code under-counted the number of quick-service restaurants because many chain fast food

restaurants had a primary NAICS code for full-service restaurant and a secondary one for limited-service restaurants. Additionally, some take-out restaurants only had the NAICS code designation of full-service restaurant. In fact, the most common designation for ground-truthed retailers that existed in the InfoUSA dataset was *full-service restaurant* but this was not considered in our analysis.

Using NAICS codes to identify restaurants, especially quick-service restaurants, may require additional modifications, such as expanding the NAICS code search to include secondary NAICS codes or manually identifying the names of large fast food chain restaurants (*e.g.*, McDonalds) and key terms for independently-owned, take-out restaurants (*e.g.*, chicken). Prior studies have used similar approaches given the limitations of the NAICS codes (Fleischhacker *et al.*, 2013; Wilkins *et al.*, 2017). More research may be required to establish the best algorithm to identify quick-service restaurants. This algorithm may have regional variations. For example, our study areas had very few chain fast food outlets, so challenges in identifying these types of outlets were not of particular concern for this study. However, this may be of concern in areas where fast food restaurants are more common. In contrast, our study areas had numerous independently-owned takeout restaurants, which we attempted to identify through keyword searches among full-service restaurants. Similarly, using only the convenience store NAICS code undercounted the number of small food retailers as this code was primarily limited to chain convenience stores. A modified search that combined NAICS codes and number of employees helped to identify independently-owned food stores.

There are several limitations to our study. While our ground-truth assessment was conducted by an investigator familiar with the community, it is possible that we may have overlooked some food outlets. Our ground-truth assessments only categorised stores based on their external appearance, which might result in misclassification of some stores. Some of the stores we identified in the ground-truth assessment might exist in the MFSMP or InfoUSA datasets under a different code or category that we did not consider; however, we used a variety of MFSMP categories and NAICS codes to reduce this possibility. We likely undercounted the number of quick-service restaurants in InfoUSA, as some has NAICS code designations of *full-service restaurant*. We used key words to add in some full-service restaurants, but could have missed some. We only focused on food outlets that are typically considered *unhealthy*, but it may also be important to also understand how these datasets perform in accurately identifying healthy food sources such as supermarkets. Unfortunately, we could not assess this, as few or none of these outlets exist in our study region. Our findings, conducted in low-income, inner city communities, cannot be extended to other areas, such as suburban or rural neighbourhoods.

Conclusions

Data from academic-government partnerships like MFSMP might be an attractive alternative to relying only on commercial data to identify small food retailers and quick-service restaurants. While the MFSMP is derived from the local health department, CLF has invested resources to validate the data, improve its quality, and transform it into an easy-to-use format for research. We found these data to include less misclassification and ambiguity

in identifying appropriate food outlets compared to InfoUSA. Given the potential strengths of the academic-partnership data compared to commercial data, other research institutes or cities might consider replicating CLF's efforts to create and maintain this type of environmental dataset, although we acknowledge that this would require an investment of time and money. Even if such data could not be updated on an annual basis, our results suggest that the information provided to researchers would still provide more accurate information than the most up-to-date commercial data available for low-income, inner city communities.

Acknowledgements:

the authors would like to thank James Harding from the Center for a Livable Future for assistance with MFSMP data. We would also like to acknowledge our funding sources: MSW was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (T32DK062707). KAG was supported by a career development award from the National Heart, Lung, and Blood Institute (K23HL116601).

References

- Bader MD, Ailshire JA, Morenoff JD, House JS, 2010 Measurement of the local food environment: a comparison of existing data sources. *Am J Epidemiol* 171:609–17. [PubMed: 20123688]
- Baltimore Neighborhood Indicators Alliance - Jacob France Institute, 2015 Vital Signs 14, Baltimore, MD, USA Available from: http://www.bnijfi.org/vital_signs
- Buczynski A, Freishtat H, Buzogany S, 2015, Mapping Baltimore City's food environment: 2015 report Baltimore, MD, USA Available from: <http://mdfoodsystemmap.org/wpcontent/uploads/2015/06/Baltimore-Food-Environment-Report-2015-1.pdf>
- Center for a Livable Future, 2015 Center for a Livable Future, Baltimore, MD, USA Available from: <http://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-a-livable-future/>
- Dun&Bradstreet, 2016 Dun&Bradstreet, Short Hills, NJ, USA Available at <http://www.dnb.com/>
- Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS, 2010 The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 16:175–90. [PubMed: 19880341]
- Fleischhacker SE, Evenson KR, Sharkey J, Pitts SB, Rodriguez DA, 2013 Validity of secondary retail food outlet data: a systematic review. *Am J Prev Med* 45:462–73. [PubMed: 24050423]
- Han E, Powell LM, Zenk SN, Rimkus L, Ohri-Vachaspati P, Chaloupka FJ, 2012 Classification bias in commercial business lists for retail food stores in the U.S. *Int J Behav Nutr Phys Act* 9:46. [PubMed: 22512874]
- InfoUSA, 2016 InfoUSA, Papillion, NE, USA Available from: <https://www.infousa.com/>
- Liese AD, Colabianchi N, Lamichhane AP, Barnes TL, Hibbert JD, Porter DE, Nichols MD, Lawson AB, 2010 Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. *Am J Epidemiol* 172:1324–33. [PubMed: 20961970]
- Lucan SC, Maroko AR, Bumol J, Torrens L, Varona M, Berke EM, 2013 Business list vs ground observation for measuring a food environment: saving time or waste of time (or worse)? *J Acad Nutr Diet* 113:1332–9. [PubMed: 23871107]
- Lyseen AK, Hansen HS, 2014 Spatial and semantic validation of secondary food source data. *ISPRS Intl J Geo-Inf* 3:236–53.
- Paquet C, Daniel M, Kestens Y, Leger K, Gauvin L, 2008 Field validation of listings of food stores and commercial physical activity establishments from secondary data. *Int J Behav Nutr Phys Act* 5:58. [PubMed: 19000319]
- Powell LM, Han E, Zenk SN, Khan T, Quinn CM, Gibbs KP, Pugach O, Barker DC, Resnick EA, Myllyluoma J, 2011 Field validation of secondary commercial data sources on the retail food outlet environment in the US. *Health Place* 17:1122–31. [PubMed: 21741875]

- Rossen LM, Pollack KM, Curriero FC, 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* 18:956–62. [PubMed: 22818588]
- Sturm R, 2008 Disparities in the food environment surrounding US middle and high schools. *Public Health* 122:681–90. [PubMed: 18207475]
- Wilkins EL, Morris MA, Radley D, Griffiths C, 2017 Using Geographic Information Systems to measure retail food environments: discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). *Health Place* 44:110–7. [PubMed: 28236788]



Figure 1. Comparison of small food retailers and quick-service restaurants from three data sources.

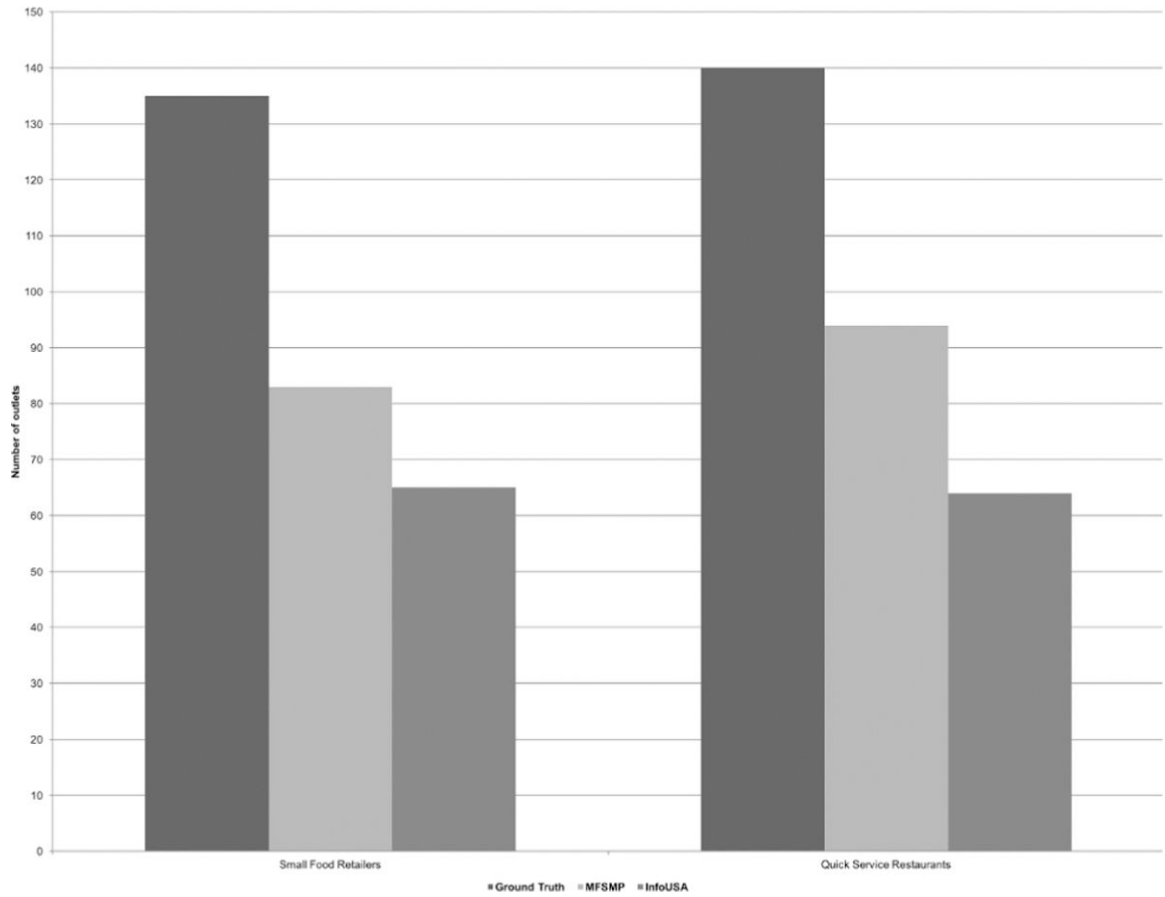


Figure 2. Counts of food outlets identified within 0.75-mile radius of housing developments by data sources.

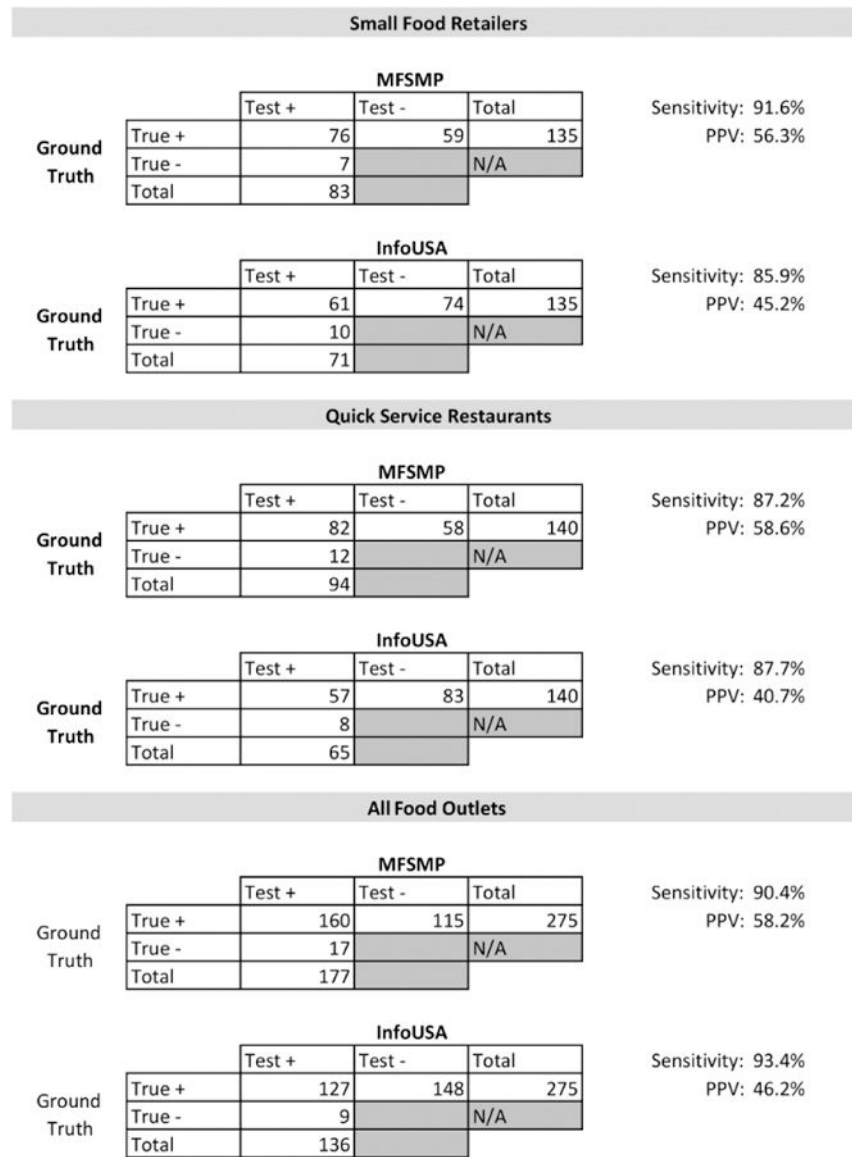


Figure 3. Validation statistics for small food retailers, quick-service restaurants, and all food outlets.

Table 1.

Definition of food outlets by data sources and counts by outlet type.

Data set	Small food retailers Definition			Quick-service restaurants		
	Outlet type	Definition	Number of outlets	Outlet type	Definition	Number of outlets
Ground truth	Corner stores	Small, locally owned businesses that primarily sell prepared foods and beverages; traditionally located at corners. Some of these stores may also have a deli counter that sells prepared foods, such as sandwiches; however, these typically occupy a small portion of the store. Corner stores that sell both food and liquor were excluded from this category	116	Fast food	Franchise restaurants that sell inexpensive food that is prepared quickly and has minimal table service	17
	Convenience stores	Similar to corner stores, but typically business is part of a larger chain	19	Take-out restaurants	Typically independently owned restaurants with minimal/no table service that serve inexpensive food that is intended to be consumed away from the restaurant's premise	123
MFSMP	Corner stores	Small-format grocery stores that are typically independently owned and operated; annual sales <\$2 million	46	Fast food chain local restaurant	National, regional, and chain name restaurants that offer take-out-ready food prepared quickly and with little to no table service and a consistent menu across locations	6
	Small grocery stores		1			
InfoUSA	Convenience stores	Stores that sell food products, but where food is not the main business (the majority of sales are made up from gas, cigarettes, pharmacy items, home good, etc.); includes chain and gas station convenience stores, drug stores or pharmacies, and discount/dollar stores; typically chain operated	9	Carry-outs	Independently owned restaurants serving <i>to-go</i> prepared foods, typically high in calories, salt, and fat	88
	Behind-glass stores	Similar to corner stores, but all of the goods for sale and the clerk are physically behind Plexiglas; typically found in areas with high crime	27			
InfoUSA	Convenience stores	Convenience stores as identified by NAICS code 445120	21	Limited-service restaurants	Primary or secondary NAICS code 722513	31
	Small grocery stores	Grocery stores as identified by NAICS code 445110 and has less than 4 employees	50	Full-service restaurants	Primary NAICS code 722511 with the following words in the restaurant name: carryout/carry-out, chicken, trout, pizza, Chinese/China	34