Published in final edited form as:

J Acad Nutr Diet. 2013 October ; 113(10): 1332–1339. doi:10.1016/j.jand.2013.05.011.

Business list vs. ground observation for measuring a food environment: saving time or waste of time (or worse)?

Sean C. Lucan, MD, MPH, MS[a], **Andrew R. Maroko, PhD**[b], **Joel Bumol, MD**[a], **Luis Torrens, BA**[c], **Monica Varona, MEM**[d], and **Ethan M. Berke, MD, MPH**[e]

Sean C. Lucan: slucan@yahoo.com; Andrew R. Maroko: Andrew.Maroko@lehman.cuny.edu; Luis Torrens: edward.torrens@gmail.com; Monica Varona: mvarona1@gmail.com; Ethan M. Berke: ethan.berke@tdi.dartmouth.edu

[a]Department of Family and Social Medicine, Albert Einstein College of Medicine | Montefiore Medical Center, Bronx, NY

[b]Department of Health Sciences, Lehman College, City University of New York, Bronx, NY

[c]City University of New York School of Public Health at Hunter College, New York, NY

[d]Nicholas School of the Environment, Duke University, Durham, NC

[e]Prevention Research Center at Dartmouth, The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH (USA)

Abstract

In food-environment research, an alternative to resource-intensive direct observation on the ground has been the use of commercial business lists. We sought to determine how well a frequently-used commercial business list measures a dense urban food environment like the Bronx. On 155 Bronx street segments, investigators compared two different levels for “matches” between the business list and direct ground observation: *lenient* (by business type) and *strict* (by business name). For each level of matching, researchers calculated sensitivities and positive predictive values (PPVs) for the business list overall and by broad business categories: *General grocers* (e.g., supermarkets), *Specialty-food stores* (e.g., produce markets), *Restaurants*, and *Businesses not primarily selling food* (e.g., newsstands). Even after cleaning the business list (e.g., for cases of multiple listings at a single location), and allowing for inexactness in listed street addresses and spellings of business names, the overall performance of the business list was poor. For strict “matches”, the business list had an overall sensitivity of 39.3% and PPV of 45.5%. Sensitivities and PPVs by broad business categories were not meaningfully different from overall values, although sensitivity for *General grocers* and PPV for *Specialty-food stores* were particularly low: 26.2% and 32.0% respectively. For lenient “matches”, sensitivities and PPVs were somewhat higher but still poor: 52.4–60.0% and 60.0–75.0% respectively. The business list is inadequate to measures the actual food environment in the Bronx. If results represent

CORRESPONDING AUTHOR (ALSO REQUESTS FOR REPRINTS): Sean C. Lucan, MD, MPH, MS, Department of Family and Social Medicine, Albert Einstein College of Medicine, 1300 Morris Park Ave, Mazer Building, Room 410, Bronx, NY 10461 Tel: (718) 430-3667, Fax: (718) 430-8645, slucan@yahoo.com.

performance in other settings, findings from prior studies linking food environments to diet and diet-related health outcomes using such business lists are in question, and future studies of this type should avoid relying solely on such business lists.

Keywords

groundtruthing; commercial business list; food environment; sensitivity; positive predictive value; accuracy; validity; grocer stores; restaurants; food stores

INTRODUCTION

To measure food environments, an alternative to resource-intensive direct ground observation has been the use of commercial business lists like those maintained by Infogroup (formerly InfoUSA). Use of Infogroup's business lists to measure food environments is common, with several recent papers providing examples.[1–7] Such papers often provide little discussion of the business list's validity though, despite the fact that business-list data are primarily for business-to-business marketing, not food-environment research.

For food-environment research, validity of Infogroup's business lists may actually be quite modest. As compared to direct ground observation, studies report differing levels of agreement and highly variable sensitivities and positive predictive values (PPVs) depending on the specific food sources under consideration (e.g., supermarkets, convenience stores, restaurants) and the geographic locations of the studies.[8–11]

Prior Infogroup validation studies have been limited in both method and scope though. For instance, studies have considered only a limited sample of food sources: mostly select food stores and restaurants.[8–11] Neglected has been a range of potentially relevant additional retail, such as general merchandisers, gasoline service stations, and newsstands, which also offer food and beverages and which may also contribute meaningfully to an overall food environment.[12] Additionally, some Infogroup validation studies have been lenient in their definitions of "matches" between the business-list data and direct ground observation. For instance, "matches" based on general business category (e.g., any kind of fast-food restaurant at a location)[9] as opposed to precise business identity (e.g., a specific fast-food franchise at a location) would tend to overestimate the validity of the business list. Moreover, such matching could lead to important misclassification (e.g., a fast-food outlet like *Subway* may be different from a nutritional standpoint than a fast-food outlet like *Taco Bell*,[13] and counting *Subway* and *Taco Bell* as the same may be inappropriate for most purposes). Finally, no validation studies have occurred in New York City (NYC)—an urban setting with by far the most retail (food and non-food) in the U.S.[14]—even though foundational food-environment research using an Infogroup business list included hundreds of census tracts form NYC.[15]

The objective of this study was to rigorously evaluate the accuracy of Infogroup's business-list data over a wide range of food sources, and do so in the dense urban environment in NYC: the Bronx. Researchers sought to assess: (i) sensitivity (how often the business list identified food sources when they are actually present) and PPV (how often food sources are actually present when the business list said they are) relative to direct ground observation (primary data collection on Bronx streets) for a range of different types of food-related retail, (ii) different criteria for "matches" between the Infogroup data and direct ground observation, to understand nuances in how well the business list might perform for various types of food-environment research.

METHODS

This study did not involve human subjects and was deemed exempt by the Albert Einstein College of Medicine Institutional Review Board.

Business-list data

The business list for this study came from Infogroup (www.infogroup.com), downloaded April 2010 through a site license at Dartmouth's Tuck School of Business. Infogroup data include business name, business type, geographic location, and various company-relevant reports. Per the company's technical staff, data updates occur monthly, although randomly and without geographic basis. The company examines every address in the U.S. at least one to three times a year, with addresses in more densely-populated areas receiving the more-frequent updates.

Data cleaning

The Infogroup data commonly lists two or more discrete businesses at the same address. In some cases, such listings legitimately reflect two businesses operating at the same location (e.g., linked franchises like Dunkin Donuts and Baskin Robbins). More often, however, multiple listings result from having separate records for back offices and retail store fronts of single businesses (e.g., "Devo Food Corp" and "Fine Fare" at the same address for a single supermarket). We manually reconciled all addresses having multiple listings as not to unfairly disadvantage the business list. Specifically, we retained cases of dual businesses operating from a single storefront and purged records of back offices that would not contribute to the food environment or be observable to investigators on the ground.

Direct ground observation

Two teams of two investigators—one working July–August 2010, the other November 2010–March 2011—cumulatively observed both sides of 155 Bronx street segments (regions along streets from one intersection to the next), blinded to the Infogroup data. Both teams of investigators separately assessed a random sample of the same 30 street segments as a reliability check to make sure there was consistency in data collection and findings. Beyond this small area of overlap, both teams targeted separate random samples of street segments generated from a file containing all business lots in the Bronx (LotINFO, Space Track, Inc., New York, NY, 2008). Investigators recorded the names, addresses, and GPS coordinates of all storefront businesses offering any types of foods or beverages on sampled streets. When investigators could not determine if businesses offered foods or beverages from the sidewalk, investigators entered the businesses to check.

Categorizing businesses

Teams noted *broad business categories* for all businesses. These categories, which study investigators developed to facilitate comparisons between the business list and direct ground observation, were: *General grocers*, *Specialty-food stores*, *Restaurants*, and *Businesses not primarily selling food*) (see Table 1 for more details).

Determining "matches" between datasets

To be a "match" between the business list and direct ground observation, businesses had to be on the same street segment and have the same *broad business category*. Researchers then determined one of two levels of "matching" based on consistency in business name: (1) *strict matches*: businesses with the same or consistent name (e.g., "Franko Deli" and "Franco's Heroes and Sandwiches") vs. (2) *lenient matches*: businesses potentially having different names but thought to be of a consistent business type based on name (e.g., "Nacho Pizza" and "Original Tony's Pizza", both pizzerias; but not "Kim's Fruit Market" and "Triberia Fish Market", substantively different food outlets even though within the same *broad business category*). Substantial variations in listed names, notations, and spellings between business-list and direct-ground-observation datasets precluded using any automated approach to determining "matches" (see footnote to Table 2 for additional examples).

Statistical analysis

Using Stata 11 (Statacrop LP, College Station, TX), investigators calculated sensitivities, PPVs, and confidence intervals for the sample as a whole and for each of the four *broad business categories*, both by *strict* and *lenient* "matches".

RESULTS AND DISCUSSION

On 155 street segments across the Bronx (Figure 1), investigators observed 234 businesses offering any types of foods or beverages (there was complete agreement for the 30-segment reliability check between the two research teams). By comparison, the commercial business list identified only 202 food-related businesses (after 17 back-office listings were removed in the data-cleaning process).

By broad business category, direct ground observation showed 42 *General grocers*, 26 *Specialty-food stores*, 110 *Restaurants*, and 56 *Businesses not primarily selling food*. These values compared to 32, 25, 88, and 57 respectively for the business list.

Table 2 shows that both overall and by separate *broad business category*, sensitivities and PPVs were above 50% for *lenient* "matches", and generally less than 50% for *strict* "matches". For *strict* "matches", the sensitivity for *General grocers* was only 26.2% (i.e., the business list only identified existing grocers about one quarter of the time) and the PPV for *Specialty-food stores* was only 32.0% (i.e., only about a third of the time were specialty food stores actually on the street when the business list said they were). For *Businesses not primarily selling food* (a business category typically neglected in food-environment research), the business list had a *strict*-"match" sensitivity and PPV lower than for *Restaurants* (a business category typically measured in food-environment research). The business list might correctly identify existing food-related businesses just a third of the time under a worst-case *strict*-“match” scenario (i.e., lower confidence limit for overall sensitivity was 33.0%), and might mistakenly list non-existent food-related businesses as present more than a quarter of the time under a best-case *lenient*-“match” scenario (i.e., 1 - upper confidence limit for overall PPV of 73.7% = 26.3%).

No prior validation studies appear to have used criteria as rigorous as those established by our *strict* “matching”. *Strict* matching allowed for spelling errors and substantial imprecision in recorded names (see footnote to Table 2), but provided reasonable confidence that the two businesses being “matched” were indeed one and the same as opposed to just similar to some degree. *Strict* “matches” give the most stringent estimate of a business list’s accuracy, which— with upper confidence limits for sensitivities and PPVs generally around 50%— might be reasonably comparable to flipping a coin under the most optimistic scenarios.

Actually, a coin flip probably *overstates* the business list’s accuracy (even being optimistic) since investigators first cleaned the dataset to exclude back offices that would have increased “false positives” and thus would have reduced PPVs. Additionally, investigators required only that businesses in the commercial list be on the same street segment as the corresponding businesses in the direct-ground-observation data, rather than at the exact same address. With stricter criteria, address imprecision (e.g., 451 vs. 455 Morris Park Avenue) would have resulted in more missed businesses and thus lower sensitivities.

Data cleaning and address imprecision are relevant to both *strict* and *lenient*-“match” findings. Arguably, though, for the purposes of most food-environment research, the *lenient*-“match” findings could be most relevant. These findings allowed for certain record-keeping anomalies in the business list that may have created real but perhaps not meaningful differences. For instance, in cases where the business list might have retained the original business name when a shop changed hands (e.g., potentially “Nacho Pizza” vs. “Original Tony’s Pizza” in the current study), or listed the “doing business as” name as opposed to the retail store name (e.g., possibly “Tseng’s Ice Cream Shop” vs. “Baskin Robbins”), the errors in naming might not be so relevant to making important food-retail distinctions. Perhaps a pizzeria is a pizzeria and an ice- cream shop is an ice-cream shop, and the specific business names do not matter within business types.

Specific business names do matter *between* business types though (e.g., between pizzerias and ice-cream shops). At least one previous study did not consider business names to make such distinctions but rather relied on a less-stringent matching method based on general classification code only.[9] In this study, a “match” was when, for example, the business in each dataset was classified as a “fast-food restaurant”.[9] Most of the time, such a methodology would probably not be problematic: e.g., any fast-food burger franchise might contribute similarly to a food environment as any fast-food chicken outlet, even though they are clearly different types of fast- food businesses. However, if considering classification code only, theoretically one fast-food restaurant could be a burger franchise and another a fast-food chain that specializes in take-away salads. Such distinction would be relevant, and neglecting such distinction would be an important form of misclassification. Even our *lenient* “matches” would avoid such misclassification being based on business name to distinguish, for example, between burger joints and salad makers (e.g., *Checkers* vs. *Saladworks*). Notably, another study that based its matching on business name came to a similar conclusion that we did: that the Infogroup business list is “insufficient” to characterize the food environment.[11] Such conclusion contrasts to the more moderate appraisal made by the study using a general classification-based matching method: that the Infogroup business list must only “be used with caution”.[9]

The current study had several strengths. Investigators examined a full range of potentially relevant businesses selling foods and beverages[12]—including businesses not assessed in earlier studies like general merchandisers, newsstands, and clothing stores—and showed that the business list performed no better for these less-intuitive business types than for more-typically measured food-related businesses. The study used two teams of investigators to perform direct ground observations and verified consistency using a random sample of the same street segments for reliability checking. Notably, investigators made ground assessments blinded to data in the commercial business list. Further, investigators performed separate analyses using two different criteria for “matches” and conducted all match checking by hand after carefully hand cleaning the business-list data.

The study’s main limitation was a relatively small sample size, owing mostly to the labor-intensive method of hand cleaning and matching all results. Regardless, the sample covered the entire geographic area of the Bronx (Figure 1) and while confidence intervals might have been relatively wide, even upper limits were telling and did not meaningfully change core findings or implications. Another potential limitation was the time lag between acquisition of the business-list data and full completion of direct ground observation (almost 11 months). While it is possible in this time that some retailers opened for business or went out of business, it is unlikely that such occurrences accounted for the magnitude of discrepancy between the business list and direct ground observation the study found. For instance, in the reliability check for direct ground observation, teams visited the same 30 street segments separately at least six months apart and found no meaningful differences in the businesses thereon.

A final limitation is that the study did not assess for differential misclassification by type of food-related business or by whether a food-related business was present or not by neighborhood characteristics. Others researchers have shown differential misclassification with Infogroup data by neighborhood,[16] and differences in sensitivity,[10] positive predictive value,[10] or agreement with direct ground observation[9] by neighborhood characteristics. Some authors suggest that any business-list inaccuracy is only a problem if such differential misclassification is present.[17] This argument has merit when research finds *signal* in spite of *noise* in certain neighborhoods (e.g., food deserts in poor neighborhoods); but it does not address studies potentially finding no associations[4–6, 18] (e.g. the lack of food deserts in any neighborhoods[19]). An alternative argument is that differential misclassification matters little if overall performance is unacceptable as the current study suggests may be the case.

CONCLUSION

Early food-environment research appropriately made use of commercial business lists, like those maintained by Infogroup, to search for and establish foundational associations. For instance, in 2003 Moore and Diez-Roux obtained Infogroup (then InfoUSA) data because the business list was convenient, efficient, and the authors were “aware of no better source of data”.[15] Many years later, studies like ours make it clear that Infogroup data—while potentially quite excellent for intended marketing purposes—are not adequate for advancing food-environment research. The inadequacy might even be worse in rural settings[10] because the business-list data are updated less often in areas of low population density. Overall poor sensitivities and PPVs in a range of geographic areas raise concern about findings from prior studies linking Infogroup-determined food environments to diet or diet-related health outcomes. In cases where investigators failed to find robust associations,[4–6] were there actually no associations? Or was the dataset too insensitive to detect them? In cases of found associations, [1–3, 7] (sometimes in directions opposite of expected[5, 6]), did relationships actually exist? Or were they artifacts of false positivity?

Understandably, Infogroup data and other commercial business lists are attractive for research involving large geographic areas. But in order to be useable, such lists may require extensive cleaning and/or groundtruthing, which may negate any benefit in time, cost, or efficiency compared to direct ground observation. Alternative strategies like using Google Street View,[20] telephone and internet directories,[21–24] dining guides,[21] or government datasets (alone,[10, 21, 22, 24–32] combined with phone and web lists,[33] or even to supplement Infogroup data[17, 34]) have some advantages; but even under the best conditions these methods may not be sufficient in and of themselves. For instance, they do not provide information about what foods are available within the various food sources they identify, and studies show there is considerable variability in the types of foods offered by even a singe type of food source.[35–39] Until more sophisticated, nuanced, and accurate datasets are developed, primary data collection may be the only acceptable way forward when detailed understanding of a food environment is required. In the interim, based on our findings and the results of others,[9–11] unverified business-list data may no longer be acceptable as the sole measure of food sources for rigorous food-environment research.

References

1. Leung CW, Laraia BA, Kelly M, et al. The influence of neighborhood food stores on change in young girls’ body mass index. Am J Prev Med. 2011; 41(1):43–51. [PubMed: 21665062]
2. Babey SH, Wolstein J, Diamant AL. Food environments near home and school related to consumption of soda and fast food. Policy Brief UCLA Cent Health Policy Res. 2011; (PB2011-6):1–8.
3. Dubowitz T, Ghosh-Dastidar M, Eibner C, et al. The Women’s Health Initiative: the food environment, neighborhood socioeconomic status, BMI, and blood pressure. Obesity (Silver Spring). 2012; 20(4):862–871. [PubMed: 21660076]
4. Howard PH, Fitzpatrick M, Fulfrost B. Proximity of food retailers to schools and rates of overweight ninth grade students: an ecological study in California. BMC Public Health. 2011; 11(1):68. [PubMed: 21281492]
5. An R, Sturm R. School and residential neighborhood food environment and diet among california youth. Am J Prev Med. 2012; 42(2):129–135. [PubMed: 22261208]

6. Shier V, An R, Sturm R. Is there a robust relationship between neighbourhood food environment and childhood obesity in the USA? Public Health. 2012; 126(9):723–730. [PubMed: 22898435]

7. Sanchez BN, Sanchez-Vaznaugh EV, Uscilka A, Baek J, Zhang L. Differential associations between the food environment near schools and childhood overweight across race/ethnicity, gender, and grade. Am J Epidemiol. 2012; 17

8. Bader MD, Ailshire JA, Morenoff JD, House JS. Measurement of the local food environment: a comparison of existing data sources. Am J Epidemiol 1. 2010; 171(5):609–617.

9. Powell LM, Han E, Zenk SN, et al. Field validation of secondary commercial data sources on the retail food outlet environment in the U.S. Health & place. 2011; 17(5):1122–1131. [PubMed: 21741875]

10. Liese AD, Colabianchi N, Lamichhane AP, et al. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. Am J Epidemiol. 2010; 172(11):1324–1333. [PubMed: 20961970]

11. Gustafson AA, Lewis S, Wilson C, Pitts S. Validation of food store environment secondary data source and the role of neighborhood deprivation in Appalachia, Kentucky. BMC Public Health. 2012; 12(1):688. [PubMed: 22914100]

12. Farley TA, Baker ET, Futrell L, Rice JC. The ubiquity of energy-dense snack foods: a national multicity study. Am J Public Health. Feb; 2010 100(2):306–311. [PubMed: 20019297]

13. Kirkpatrick, S.; Reedy, J.; Harris, JL.; Ohri-Vachaspati, P. Fast food restaurant menu offerings score poorly in relation to dietary guidance. Amercian Publich Health Association Annual Meeting and Expo; San Fransisco, CA. 2012. https://apha.confex.com/apha/140am/webprogram/Paper264912.html

14. Center for Retail Research. [Accessed November 16, 2012.] World Shopping Capitals 2011 Survey. http://www.retailresearch.org/worldshoppingcapitals.php

15. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. Am J Public Health. 2006; 96(2):325–331. [PubMed: 16380567]

16. Han E, Powell LM, Zenk SN, Rimkus L, Ohri-Vachaspati P, Chaloupka FJ. Classification bias in commercial business lists for retail food stores in the U.S. Int J Behav Nutr Phys Act. 2012; 9:46. [PubMed: 22512874]

17. Lamichhane AP, Mayer-Davis EJ, Puett R, Bottai M, Porter DE, Liese AD. Associations of built food environment with dietary intake among youth with diabetes. J Nutr Educ Behav. Jan 9.2012

18. Richardson AS, Boone-Heinonen J, Popkin BM, Gordon-Larsen P. Neighborhood fast food restaurants and fast food consumption: a national study. BMC Public Health. 2011; 11:543. [PubMed: 21740571]

19. Kolata G. Studies Question the pairing of food deserts and obesity. The New York Times. Apr 17.2012 Health.

20. Rundle AG, Bader MD, Richards CA, Neckerman KM, Teitler JO. Using Google Street View to audit neighborhood environments. Am J Prev Med. 2011; 40(1):94–100. [PubMed: 21146773]

21. Timperio A, Ball K, Roberts R, Campbell K, Andrianopoulos N, Crawford D. Children's fruit and vegetable intake: associations with the neighbourhood food environment. Prev Med. 2008; 46(4): 331–335. [PubMed: 18164753]

22. Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. Am J Public Health. 2005; 95(4):660–667. [PubMed: 15798127]

23. Paquet C, Daniel M, Kestens Y, Leger K, Gauvin L. Field validation of listings of food stores and commercial physical activity establishments from secondary data. Int J Behav Nutr Phys Act. 2008; 5:58. [PubMed: 19000319]

24. Lake AA, Burgoine T, Greenhalgh F, Stamp E, Tyrrell R. The foodscape: classification and field validation of secondary data sources. Health Place. Jul; 2010 16(4):666–673. [PubMed: 20207577]

25. Toft U, Erbs-Maibing P, Glumer C. Identifying fast-food restaurants using a central register as a measure of the food environment. Scand J Public Health. 2011; 39(8):864–869. [PubMed: 21969329]

26. Pearce J, Hiscock R, Blakely T, Witten K. The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption. J Epidemiol Community Health. 2008; 62(3):198–201. [PubMed: 18272733]
27. Hosler AS, Dharssi A. Identifying retail food stores to evaluate the food environment. Am J Prev Med. 2010; 39(1):41–44. [PubMed: 20537845]
28. Ahern M, Brown C, Dukas S. A national study of the association between food environments and county-level health outcomes. J Rural Health. 2011; 27(4):367–379. [PubMed: 21967380]
29. Morland K, Wing S, Diez Roux A, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. Am J Prev Med. 2002; 22(1):23–29. [PubMed: 11777675]
30. Vernez Moudon A, Drewnowski A, Duncan GE, Hurvitz PM, Saelens BE, Scharnhorst E. Characterizing the food environment: pitfalls and future directions. Public Health Nutr. 2013:1–6.
31. Lake AA, Burgoine T, Stamp E, Grieve R. The foodscape: classification and field validation of secondary data sources across urban/rural and socio-economic classifications in England. Int J Behav Nutr Phys Act. 2012; 9:37. [PubMed: 22472206]
32. Cummins S, Macintyre S. Are secondary data sources on the neighbourhood food environment accurate? Case-study in Glasgow, UK. Prev Med. 2009; 49(6):527–528. [PubMed: 19850072]
33. Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: a geographic analysis. Am J Prev Med. 2004; 27(3):211–217. [PubMed: 15450633]
34. Lamichhane AP, Puett R, Porter DE, Bottai M, Mayer-Davis EJ, Liese AD. Associations of built food environment with body mass index and waist circumference among youth with diabetes. Int J Behav Nutr Phys Act 29. 2012; 9(1):81.
35. Gebauer H, Laska MN. Convenience stores surrounding urban schools: an assessment of healthy food availability, advertising, and product placement. J Urban Health. 2011; 88(4):616–622. [PubMed: 21491151]
36. Laska MN, Borradaile KE, Tester J, Foster GD, Gittelsohn J. Healthy food availability in small urban food stores: a comparison of four US cities. Public Health Nutr. 2010; 13(7):1031–1035. [PubMed: 19968901]
37. Farley TA, Rice J, Bodor JN, Cohen DA, Bluthenthal RN, Rose D. Measuring the food environment: shelf space of fruits, vegetables, and snack foods in stores. J Urban Health. Sep; 2009 86(5):672–682. [PubMed: 19603271]
38. Bodor JN, Rose D, Farley TA, Swalm C, Scott SK. Neighbourhood fruit and vegetable availability and consumption: the role of small food stores in an urban environment. Public Health Nutr. 2008; 11(4):413–420. [PubMed: 17617930]
39. Kersten E, Laraia B, Kelly M, Adler N, Yen IH. Small food stores and availability of nutritious foods: a comparison of database and in-store measures, northern california, 2009. Prev Chronic Dis. 2012; 9:E127. [PubMed: 22789445]

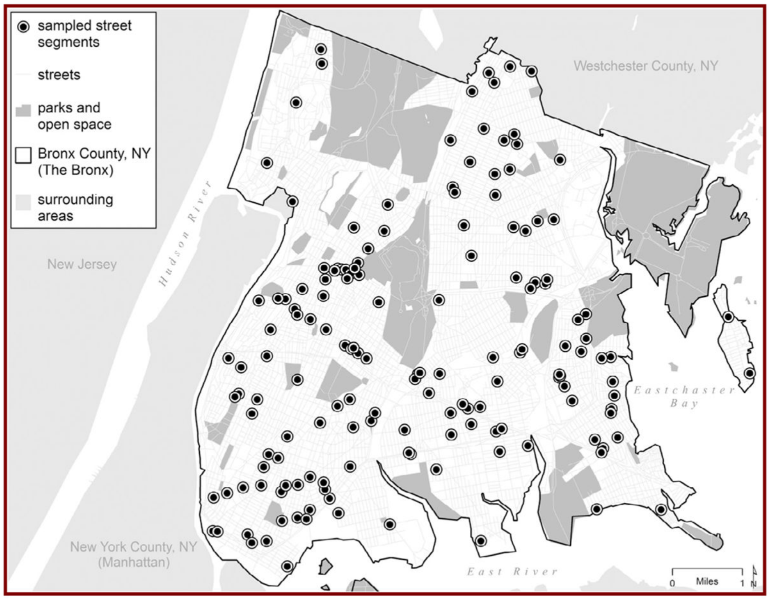

Figure 1. 155 street segments sampled from across the Bronx for direct ground observation

The count of sampled street segments may appear less than 155 due to the overlap of symbols at this scale. Street segments containing commercial business lots were close together in commercially-dense areas.

Table 1

Broad business categories for food-related retail, created to facilitate comparison between business-list data and direct ground observation[a]

Created broad business category[a]	Standardized types of retail[b]	Partial standardized definition[b]
General grocers (stores selling a wide variety of grocery items)	Grocery Stores	Supermarkets, food stores, and grocery stores, primarily engaged in the retail sale of all sorts of canned foods and dry goods … fresh fruits and vegetables … fresh and prepared meats, fish, and poultry …
Specialty-food stores (stores primarily selling one specific type of food or beverage)	Meat and Fish (Seafood) Markets	Establishments primarily engaged in the retail sale of fresh, frozen, or cured meats, fish, shellfish, and other seafood
	Fruit and Vegetable Markets	Establishments primarily engaged in the retail sale of fresh fruits and vegetables
	Candy, Nut, and Confectionery	Establishments primarily engaged in the retail sale of candy, nuts, popcorn, and other confections
	Retail Bakeries	Establishments primarily engaged in the retail sale of bakery products. The products may be purchased from others or made on the premises
	Miscellaneous Food Stores	Establishments primarily engaged in the retail sale of specialized foods, not elsewhere classified, such as eggs, poultry, health foods, spices, herbs, coffee, and tea
	Liquor Stores	Establishments primarily engaged in the retail sale of packaged alcoholic beverages, such as ale, beer, wine, and liquor, for consumption off the premises
Restaurants (outlets selling prepared food or drink for on-site or take-away consumption)	Eating Places	Establishments primarily engaged in the retail sale of prepared food and drinks for on-premise or immediate consumption
Businesses not primarily selling food[c] (businesses selling foods and/or beverages but not as their primary products)	Department stores	Retail stores generally carrying a general line of apparel, such as suits, coats, dresses; home furnishings, such as furniture, floor coverings, curtains, draperies, linens; major household appliances; and housewares, such as table and kitchen appliances, dishes, and utensils
	Variety Stores	Establishments primarily engaged in the retail sale of a variety of merchandise in the low and popular price ranges. These stores generally do not carry a complete line of merchandise and are not departmentalized
	Misc. General Merchandise Stores	Establishments primarily engaged in the retail sale of a general line of apparel, dry goods, hardware, housewares, groceries, and other lines in limited amounts
	Gasoline Service Stations	Gasoline service stations primarily engaged in selling gasoline and lubricating oils; also tires, batteries, and other auto parts
	Family Clothing Stores	Establishments primarily engaged in the retail sale of clothing, furnishings, and accessories for men, women, and children
	Drug Stores and Proprietary Stores	Establishments engaged in the retail sale of prescription drugs, proprietary drugs, and non-prescription medicines; may also carry a number of related lines: e.g., cosmetics, toiletries, tobacco, and novelty merchandise
	Tobacco Stores and Stands	Establishments primarily engaged in the retail sale of cigarettes, cigars, tobacco, and smokers' supplies
	News Dealers and Newsstands	Establishments primarily engaged in the retail sale of newspapers, magazines, and other periodicals

[a] Created by study investigators for the purposes of data collection and analysis in this study

[b] Types of retail and partial definitions based on Standardized Industrial Classifications available at www.osha.gov/pls/imis/sicsearch.html; complete definitions available at this site. For example, the complete definition of "Department stores" additionally includes: " … These and other merchandise lines are normally arranged in separate sections or departments with the accounting on a departmentalized basis. The departments and

functions are integrated under a single management. The stores usually provide their own charge accounts, deliver merchandise, and maintain open stocks. These stores normally have 50 employees or more. Establishments which sell a similar range of merchandise with less than 50 employees are classified in Industry 5399. Establishments which do not carry these general lines of merchandise are classified according to their primary activity." Such complicated definitions were impractical for use in collecting data through direct ground observation, prompting the creation of the simpler scheme shown in this table.

[c]As demonstrated in an multicity national study by Farley et al[12], food and beverage items are frequently available from a variety of non-intuitive retail outlets including those listed here in this table.

Table 2

Performance of the commercial business list relative to direct ground observation on 155 Bronx street segments, overall and by broad business category

		Broad Business Categories			
Businesses selling any food or beverage	**OVERALL (N = 234)**	**General grocers (N = 42)**	**Specialty-food stores (N = 26)**	**Restaurants (N = 110)**	**Businesses not primarily selling food (N = 56)**
By strict "matches"					
Sensitivity	39.3 (33.0, 45.9)	26.2 (13.9, 42.0)	30.8 (14.3, 51.8)	45.5 (35.9, 55.2)	41.1 (28.1, 55.0)
Positive predictive value	45.5 (38.5, 52.7)	34.4 (18.6, 53.2)	32.0 (14.9, 53.5)	56.8 (45.8, 67.3)	40.4 (27.6, 54.2)
By lenient "matches"					
Sensitivity	58.1 (51.5, 64.5)	52.4 (36.4, 68.0)	57.7 (36.9, 76.6)	60.0 (50.2, 69.2)	58.9 (45.0, 71.9)
Positive predictive value	67.3 (60.4, 73.7)	68.8 (50.0, 83.9)	60.0 (38.7, 78.9)	75.0 (64.6, 83.6)	57.9 (44.1, 70.9)

All values in table are percentages. Values in parentheses are 95% confidence intervals. N values in header are the numbers of businesses directly observed on the ground. *Strict match* = two businesses with the same or consistent name: could have difference in notation and/or spelling, but seemingly the same business in both datasets (examples: "Parrilla Latina Restaurant" vs. "Parilla Dominicano", "Franko Deli" vs. "Franco's Heroes and Sandwiches", "Jumbo Hamburger" vs. "Jimbo's Hamburgers"); *lenient match* = two businesses that may have different names in each dataset but thought to be of a consistent business type based on names (examples: "Nacho Pizza" vs. "Original Tony's Pizza", "C-Town Supermarket" vs. "Bravo Supermarket", "Tseng's Ice Cream Shop" vs. "Baskin Robbins").