A methodology for combining multiple commercial data sources to improve measurement of the food and alcohol environment: applications of geographical information systems

Dara D. Mendez[1], Jessica Duell[1], Sarah Reiser[1], Deborah Martin[1], Robert Gradeck[2], Anthony Fabio[1]

[1]University of Pittsburgh, Graduate School of Public Health, Department of Epidemiology, Pittsburgh, USA;
[2]University of Pittsburgh, Center for Social and Urban Research, Pittsburgh, USA

Abstract. Commercial data sources have been increasingly used to measure and locate community resources. We describe a methodology for combining and comparing the differences in commercial data of the food and alcohol environment. We used commercial data from two commercial databases (InfoUSA and Dun&Bradstreet) for 2003 and 2009 to obtain information on food and alcohol establishments and developed a matching process using computer algorithms and manual review by applying ArcGIS to geocode addresses, standard industrial classification and North American industry classification taxonomy for type of establishment and establishment name. We constructed population and area-based density measures (e.g. grocery stores) and assessed differences across data sources and used ArcGIS to map the densities. The matching process resulted in 8,705 and 7,078 unique establishments for 2003 and 2009, respectively. There were more establishments captured in the combined dataset than relying on one data source alone, and the additional establishments captured ranged from 1,255 to 2,752 in 2009. The correlations for the density measures between the two data sources was highest for alcohol outlets (r = 0.75 and 0.79 for per capita and area, respectively) and lowest for grocery stores/supermarkets (r = 0.32 for both). This process for applying geographical information systems to combine multiple commercial data sources and develop measures of the food and alcohol environment captured more establishments than relying on one data source alone. This replicable methodology was found to be useful for understanding the food and alcohol environment when local or public data are limited.

Keywords: food and alcohol establishments, neighbourhood, geocoding, geographical information systems, database, USA.

Introduction

Resources within community and neighbourhood contexts such as grocery stores or the placement of liquor stores may provide insight into important public health issues such as weight gain, obesity, violence and health behaviours including diet and alcohol use (Feng et al., 2010; Janevic et al., 2010; Ahern et al., 2013; Shimotsu et al., 2013). However, there are no consistent publicly-available sources of data on businesses such as food and alcohol establishments in the United States of America (USA). Public data that could be used for these purposes are often inaccessible or unavailable in a standardised format. As a result, there has been growing interest in commercial data for measuring and operationalising neighbourhood resource environments (Ohri-Vachaspati et al., 2011; Gustafson et al., 2012a, 2012b).

In applying business and commercial databases, researchers have cited limitations including inadequate numeration of establishments, differences in accuracy of establishment locations, differences in business classifications (e.g. fast food places *versus* regular restaurant or wholesale *versus* retail grocery), and variation in completeness of business information by neighbourhood demographics (e.g. racial composition, urban/rural location) (Liese et al., 2010; Ohri-Vachaspati et al., 2011; Han et al., 2012). However, these data sources may be an efficient and cost-effective method for enumerating establishments and characterising neighbourhood environments. Prior studies cite the benefit of using more than one business database when measuring the neighbourhood food environment and perhaps other related neighbourhood and community resources (Liese et al., 2010). However, we are not aware of any studies that discuss a methodology or process for combining multiple commercial data sources for applications in public health research.

Corresponding author:
Dara D. Mendez
University of Pittsburgh
Graduate School of Public Health, Department of Epidemiology
Crabtree Hall, A716, Pittsburgh, PA 15261, USA
Tel. +1 412-648-5664; Fax +1 412-642-7397
E-mail: ddm11@pitt.edu

The purpose of this study was to find and describe the process for combining commercial data to construct a geocoded, multisource database of the neighbourhood food and alcohol environment. We present here key issues to consider when using multiple commercial data sources for applications in public health research. We then assess the differences in density measures (e.g. density of grocery stores) between the individual data sources and the merged database.

Methods

Data sources

We used two commercial databases, InfoUSA (http://www.infousa.com/) and Dun&Bradstreet (D&B) (http://www.dnb.com/), based on data generated in 2009 and 2003 for Allegheny county, PA as a part a larger study investigating the relationship between neighbourhood food and alcohol environments and health. InfoUSA and D&B are sales and marketing companies that provide data of businesses in the USA and Canada. Both companies compile data through a combination of methods including telephone directories, public records and internet searchers, and the use of teleconference associates. Each data source contains information about the name, address and type of establishment. Types of establishments and associated classifications are based on standard industrial classification (SIC) and North American industry classification (NAICS) taxonomy. We also used the 2000 US Census population and demographic data including the number of residents per census tract and the area in square miles for each census tract in Allegheny county to calculate density measures (e.g. density of grocery stores within a geographic region).

Geocoding

We geocoded all establishments since longitude and latitude were not provided for all establishments. The reference database used to geocode establishments, which could influence the quality of our geocoding (Karimi et al., 2004), was based on local data provided by the Allegheny county GIS department. The local, reference database included a combination of street and rooftop, which produces higher match rates and positional accuracy (Roongpiboonsopit and Karimi, 2010). We used ArcGIS, version 10 (ESRI, Redlands, USA) to geocode establishments. Spelling sensitivity was set at 70, minimum candidate score at 10 and minimum match score at 60. The settings were chosen based on a review of business locations listed in the dataset; they provided the best matches with the fewest false positive matches. Ties were automatically assigned the first matched address, and we performed no manual matching to minimise subjectivity. Output included XY coordinates, which were spatially matched to census tracts using year 2000 government census shapefiles. For the 2003 data, we were able to successfully geocode 87.2% (n = 6,849; 1,008 were unmatched to address points in the geocoding reference file) of the establishments from and 83.1% (n = 3,316; 672 were unmatched) for InfoUSA. For data from 2009, we were able to successfully geocode 86.1% (n = 5,073; 817 were unmatched) of the D&B establishments and 84.9% (n = 3,679; 656 were unmatched) from InfoUSA. Based on prior work, our address match rates were optimal based on street and rooftop source files produced by the same agency (Roongpiboonsopit and Karimi, 2010).

The establishment matching process

To combine the datasets, it was necessary to first identify establishments that existed across InfoUSA and D&B. We developed an algorithm for use in SAS (SAS Institute Inc, Cary, USA) to automate the process. Our study team conducted an extensive process to match and cross-reference food and alcohol establishments across the InfoUSA and D&B datasets (see Fig. 1). We identified four items that consistently existed on the records in both databases: name or trade name of the business, address, phone number, and SIC code. Location match or an XY (longitude, latitude) match between the two sources was the first match test. However, a location match did not guarantee that the establishments were the same business. For example, relocation, obsolete sources data or a shopping centre with several establishments could result in an invalid match. Therefore, the programme conducted a name/trade name, phone number and SIC code comparison.

The computer name matching involved multiple levels. First, we applied a spelling distance algorithm to account for data entry errors. When establishment names did not match based on spelling distance, the programme matched them based on the number of words that were identical to account for different order of words in the establishment names in the two databases. Abbreviations used in the commercial database also prevented a programme match that was obvious when reviewed by study personnel. The study team also compiled a list of key words that when

Geocoded establishments in both data sources to generate X, Y (longitude, latitudes) coordinates

Determined if establishment matched on X, Y coordinates and the number of matches

XY match	Non XY match

One to many match	One to one match	XY match

Matching based exact establishment name, phone number and SIC code	Matching based on other location information (i.e. full address, street name and zip code)

If there is not an exact match based on this information, a high and low match should be determined based on phone number, establishment name, then SIC code.

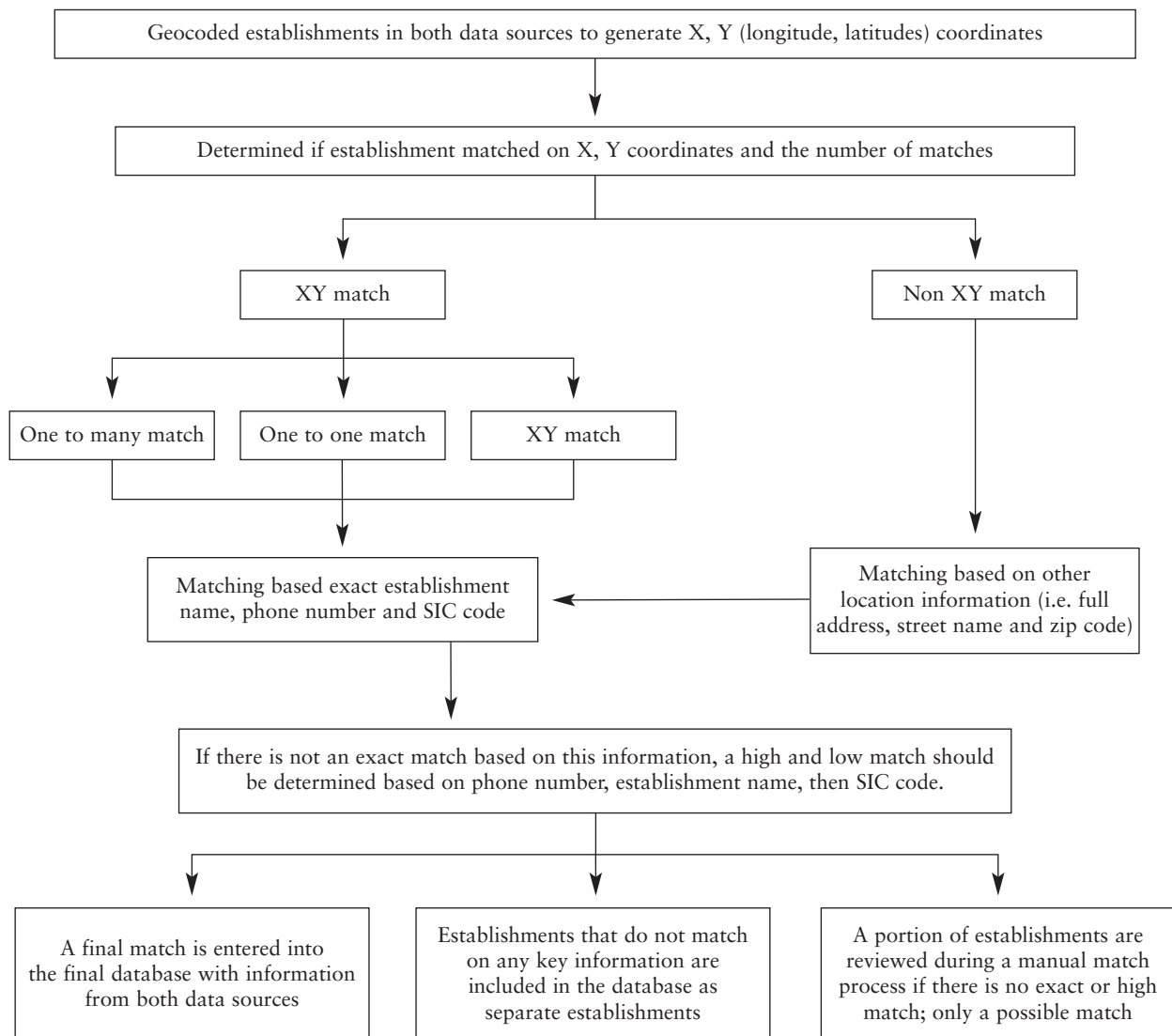| A final match is entered into the final database with information from both data sources	Establishments that do not match on any key information are included in the database as separate establishments	A portion of establishments are reviewed during a manual match process if there is no exact or high match; only a possible match

Fig. 1. Decision tree of key elements for building database and matching establishments.

found in both databases names, automatically matched the establishments to account for establishments that contained these keywords but otherwise did not match per the matching algorithm (e.g. state liquor stores had "LIQUOR" in the business name in both data sources). Any establishments that were not a definitive match based on XY coordinates, name, phone number and SIC through the computer programme were reviewed by 2-3 study personnel. A consensus on the matching status of each establishment was determined. For these establishments, additional variables were used during the review (e.g. CEO name, secondary and tertiary SIC and NAICS codes). Records that did not match across the two datasets based on XY coordinates were further examined for a match. Location match of these establishments was based on a street name, zip code, house number and the unit number if applicable. If the establishments matched (non-XY) on all characteristics and were on the same street, we reviewed them to ensure there were not two locations of the same chain (e.g. chain grocery store) on the same street. We also reviewed establishments that matched on two or more characteristics (e.g. phone number and name, same street and name, same street and phone number). In some cases, the address in one data source did not match the address in the other data source because of incorrect information (e.g. wrong/missing zip code, E(ast) instead of W(est), Old was missing). Finally, all matches were included in one database.

Measures and analysis

We operationalised the neighbourhood environment at the level of the census tract using US Census boundaries for the year 2000. Although other neighbourhood units were considered, prior work indicates census tracts as viable neighbourhood units of measuring the neighbourhood food and alcohol environment and may be best when calculating per capita and area density measures (Liese et al., 2010; Shier et al., 2012). First, various establishments were categorised according to SIC codes. The Supplement Table A includes a full listing of codes and how establishments were grouped. There were some differences in how establishments were categorised across datasets and both datasets included primary, secondary and sometimes tertiary SIC codes. Some examples are included in the Supplement, Table B. To account for differences in establishment classification, we created density measures based on primary SIC code and any SIC code (e.g. primary, secondary or tertiary). Establishments could be categorised in several different grouping, depending on the type of establishment. For example, restaurants serving alcohol could be categorised as an alcohol outlet and a restaurant.

We also created census-tract level density measures by summing the number of establishments within each census tract for each category and dividing based on per capita (per 10,000 residents) and per area (square mile) from population and area data generated from the US census. The key density measures presented in the results include on premise (i.e. bars and restaurants) and off premise alcohol outlets (i.e. state stores), grocery stores/supermarkets and restaurants. We determined the total number of establishments and mean density values for all census tracts in Allegheny county for the years 2003 and 2009. These values were calculated for InfoUSA and D&B data separately as well as for the combined InfoUSA and D&B dataset. We then conducted a comparison of mean density values and correlations of the same measures between InfoUSA and D&B.

Results

Matching and manual review results

For 2009, there were a total of 7,078 unique establishments represented in the combined dataset (results not shown). Out of the total 7,078 establishments, 3,071 were matches (i.e. a record of the establishments exists in both InfoUSA and D&B databases); 1,255 were from InfoUSA only, and 2,752 were from D&B

only. Among establishments with XY coordinates, there were a total of 64 combinations of establishments that were reviewed manually, and 42 were deemed matches. Among those without XY coordinates, 131 were deemed matches out of 316 that were manually reviewed. For 2003, there were a total of 8,705 unique establishments represented in the combined dataset (results not shown). There were 3,071 records across both sources; 946 from InfoUSA only and 4,738 from D&B only. Among establishments with XY coordinates, there were a total of 92 combinations of establishments that were reviewed manually, and 38 were deemed matches. Among those without XY coordinates, 204 were deemed matches out of 343 that were manually reviewed.

Descriptive results

We compared across datasets the total number of various types of establishments based on the primary and any SIC code (Tables 1a and 1b). Among establishments categorised by the primary SIC code, we found that combining the two data sources provided a greater number of establishment where establishments missing in one data source may have been included in the other data source. For example, in 2009, there were 540 alcohol outlets included in InfoUSA data, 703 in D&B, and combining the data sources produced 780 unique alcohol outlets. We also found that total number of establishments in D&B tended to be higher than the number of establishments in InfoUSA. We saw this trend across the various types of establishments. We also found that combining the datasets produced a higher number of establishments than relying on only one source. When we examined all SIC codes (i.e. primary, secondary, etc.), we find that D&B on average tended to have a higher number of establishments in each category compared to InfoUSA. In all categories, combining the two data sources resulted in a higher number of establishments compared to relying on one data source, similar to the patterns among the primary SIC code designation.

We also calculated the mean densities (per capita and area) for all census tracts in Allegheny county of establishments based on any SIC code, meaning establishments could be included in more than one category depending on the type of establishment (Tables 2a and 2b). Similar to the patterns seen in Tables 1a and 1b, the mean densities of the combined dataset were higher than for each single data source. For example, the mean density (population-based) of grocery stores per census tract in 2009 was 0.6 for InfoUSA, 12.6 for

Table 1a. Total number of establishments per census tract for each data source in Allegheny county, PA, USA in 2003 and 2009 based on the primary SIC code.

Year	2003			2009		
	InfoUSA	D&B	Combined	InfoUSA	D&B	Combined
Alcohol outlets	562	1,013	1,047	540	703	792
Grocery stores/supermarkets	0	1,170	1,065	0	739	656
Restaurants	2,152	3,350	3,951	2,439	2,583	3,337
Wholesale foods	132	220	286	117	225	286
General/convenience stores	64	398	417	147	337	384
Gas stations	198	334	391	212	259	335
Pharmacies	208	363	409	224	227	294

Table 1b. Total number of establishments per census tract for each data source in Allegheny county, PA, USA in 2003 and 2009 based on any SIC code.

Year	2003			2009		
	InfoUSA	D&B	Combined	InfoUSA	D&B	Combined
Alcohol outlets	669	1231	1,385	651	884	1,113
Grocery stores/supermarkets	66	1245	1,304	75	842	900
Restaurants	2,191	3,525	4,148	2,480	2,737	3,541
Wholesale foods	135	253	319	125	253	328
General/convenience stores	80	451	493	222	371	495
Gas stations	198	360	417	212	272	347
Pharmacies	208	383	429	225	250	317

D&B, and 16.7 for the combined dataset. The median grocery store density (population-based) in 2009 was 0 (range 0-37.6) for InfoUSA, 4.3 (range 0-902.3) for D&B and 4.5 (range 0-939.8) for the combined dataset (results not shown). The patterns of densities are similar for the per capita and area-based measures for 2009 and 2003. As can be seen in Tables 2a and 2b, the standard deviations for some of the density measures were quite high, indicating the range of establishments in areas that tended to be more commercial or were designated as shopping centres or malls.

We found a high correlation between InfoUSA and D&B for the density measures for alcohol outlets, grocery stores, restaurants, wholesale foods and gas stations (Table 3). The correlations ranged from 0.32 to 0.85 among the population-based density measures and ranged from 0.32 to 0.83 among the area-based density measures. The highest correlations were for both restaurant density measures, and the lowest correlations were for both grocery store measures. We additionally evaluated the association between land zoned for commercial use and various densities for the individual data sources and combined database

(results not shown). We did not find any major differences in the association between the density measures and land use for each of the data sources. ArcGIS was also used to develop density maps for Allegheny county. We included grocery store density maps for Allegheny county census tracts based on measures for InfoUSA only, D&B only and then the combined dataset (Fig. 2). The darker regions indicate a higher density. In the map of InfoUSA, there are many census tracts with a density of 0 while there is more of a range throughout the region in the D&B map. Although there were several tracts with a density of 0, the map with the combined InfoUSA and D&B data shows a wider range of densities throughout the region. We also saw a higher density in the combined data map, particularly within the centre of the map (within the city) compared to the maps of the separate datasets.

Discussion

The study presents a process and method of combining two secondary commercial data sources to capture the food and alcohol environment. Although this

Table 2a. Establishment density per 10,000 people and census tract for each data source in Allegheny county, PA, USA in 2003 and 2009.

Year	2003			2009		
	InfoUSA	D&B	Combined	InfoUSA	D&B	Combined
Alcohol outlets	7.0 (25.5)	14.0 (54.9)	15.7 (60.1)	9.9 (65)	12.6 (74.0)	16.7 (98.3)
Grocery stores/supermarkets	0.8 (6.0)	14.1 (70.3)	14.9 (75.8)	0.6 (2.4)	9.4 (47.8)	9.9 (49.6)
Restaurants	25.4 (91.6)	38.3 (125.5)	46.1 (156.3)	33.2 (160.1)	32.9 (137.5)	44.3 (198.5)
Wholesale foods	3.9 (44.6)	3.8 (44.2)	6.8 (69.9)	2.4 (25.3)	4.0 (37.2)	6.0 (57.2)
General/convenience stores	0.7 (2.5)	3.5 (4.7)	4.0 (5.4)	2.6 (13.5)	2.9 (4.4)	4.8 (14.5)
Gas stations	2.5 (15.5)	4.5 (26.8)	4.9 (26.9)	2.1 (9.3)	2.6 (9.5)	3.2 (9.7)
Pharmacies	1.6 (3.3)	3.1 (4.7)	3.5 (5.0)	1.8 (6.1)	1.8 (3.4)	2.5 (6.6)

Table 2b. Establishment density per square mile and census tract for each data source in Allegheny county, PA, USA in 2003 and 2009.

Year	2003			2009		
	InfoUSA	D&B	Combined	InfoUSA	D&B	Combined
Alcohol outlets	3.3 (6.5)	6.2 (12.9)	6.8 (13.8)	3.1 (6.4)	4.3 (9.1)	5.3 (10.6)
Grocery stores/supermarkets	0.3 (1.1)	6.1 (11.5)	6.4 (12.2)	0.4 (1.4)	3.7 (7.6)	3.9 (7.9)
Restaurants	10.2 (23.7)	16.0 (32.3)	18.9 (38.8)	11.4 (26.2)	12.3 (26.1)	15.9 (33.5)
Wholesale foods	0.4 (1.6)	0.8 (2.5)	1.0 (3.2)	0.4 (1.5)	0.9 (2.4)	1.1 (3.3)
General/convenience stores	0.3 (1.0)	2.0 (3.7)	2.2 (3.8)	0.9 (2.1)	1.5 (2.6)	2.0 (3.6)
Gas stations	0.6 (1.3)	1.3 (3.0)	1.5 (3.1)	0.8 (2.0)	1.0 (2.3)	1.3 (2.5)
Pharmacies	0.9 (2.1)	2.0 (4.2)	2.2 (4.6)	0.9 (2.1)	1.0 (2.6)	1.2 (2.9)

Table 3. Spearman correlations between InfoUSA and D&B for the various density measures 2009.

Density of establishments per 10,000 people	Correlation coefficient	P-value
Alcohol outlets	0.75	<0.001
Grocery stores/supermarkets	0.32	<0.001
Restaurants	0.85	<0.001
Wholesale foods	0.44	<0.001
General/convenience stores	0.49	<0.010
Gas stations	0.58	<0.001
Pharmacies	0.75	<0.010

Density of establishments per mile^2	Correlation coefficient	P-value
Alcohol outlets	0.79	<0.001
Grocery stores/supermarkets	0.32	<0.001
Restaurants	0.83	<0.001
Wholesale foods	0.42	<0.001
General/convenience stores	0.46	<0.001
Gas stations	0.59	<0.001
Pharmacies	0.76	<0.001

study is specific to InfoUSA and D&B data mainly due to their increased usage in public health literature (Liese et al., 2010; Gustafson et al., 2012a), this process can be applied among other secondary and local data sources. We found that there were some differences in the total number and density of establishments between InfoUSA and D&B. Combining the two data captured more establishments, but since we lacked comparable data from the same year (i.e. local historical data for 2003 and 2009), we could not determine whether we were under-counting or over-counting even when combining the data sources. Based on prior work, both sources typically undercount the true number of establishments (Liese et al., 2010). In our study, for some establishment categories, the range of combined establishments was approximately 40 to 900 compared to either single data source.

This study involves historical data (i.e. data from 2003), for which changes in the landscape of food, alcohol and other establishments could not be easily confirmed through other tools or methods such as online verification or "ground-truthing" (i.e. field sur-
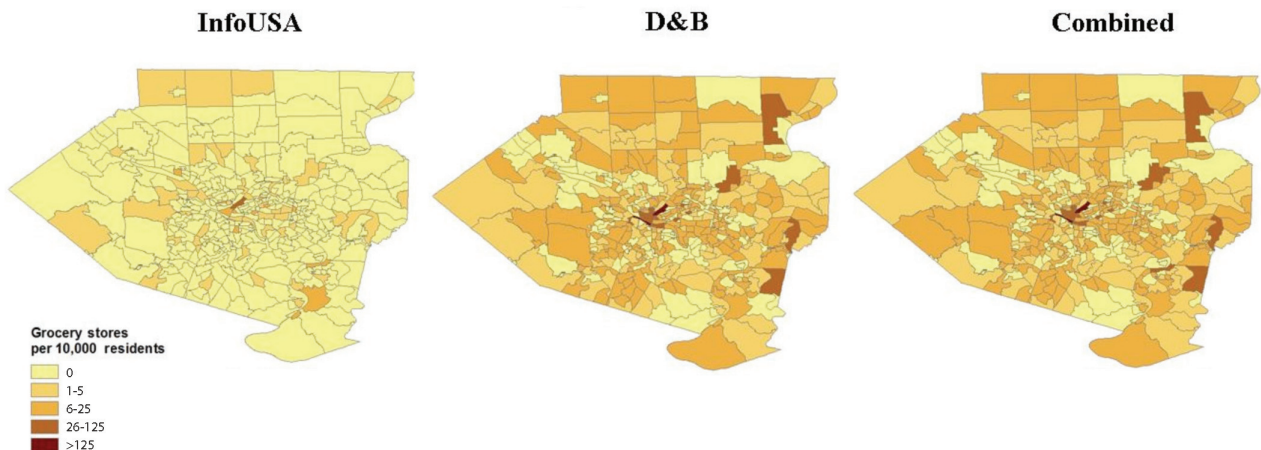
Fig. 2. Grocery store density maps (per capita) for the separate and combined commercial data sources for all census tracts in Allegheny county, PA, USA 2009. This figure includes grocery store density maps for each commercial data source, InfoUSA and D&B, and then the final map for both data sources combined. Density was calculated for each census tract as the number of grocery stores per 10,000 residents. The darker regions indicate a higher density. Data are based on census tracts in Allegheny county, PR, USA in 2009.

vey to verify a location). Additionally, data from tax records, which could possibly be used for the purposes of obtaining information about locations of food and alcohol establishments, are subject to confidentiality restrictions, and the use of local food inspection data to capture food establishments differs from location to location in terms of format, address quality and electronic availability. Combining secondary commercial data sources and local data sources when available is a viable option for research examining patterns of establishments that cannot be verified through neighbourhood audits.

Both data sources included SIC taxonomy in classifying establishments. However, the methodology used by each company differs, influencing how establishments are categorised. The processes used by these companies in identifying businesses and collecting information from businesses are proprietary, which limits an understanding of the data quality and composition. As a result, we decided to include measures and analyses that included classifications from both data sources. One study that examined various ways of reclassifying establishments did not compare classification differences across data sources, but found that classifications stemming exclusively from commercial data sources may be more likely to categorise establishments as "less healthy" options (e.g. as convenience stores or limited- service restaurants *versus* supermarkets or grocery stores) (Ohri-Vachaspati et al., 2011). Additionally, further work is needed to ensure consistent acquisition of information, includ-

ing how establishments are defined across time and geographic region for purposes other than marketing (Ohri-Vachaspati et al., 2011). In our study, we found some differences across the data sources in categorising restaurants and bars where InfoUSA primarily classified certain establishments as restaurants while D&B classified the same type establishment as a bar. In other cases, both classifications were provided for the establishment; one under the primary SIC and the other assigned to the secondary SIC. We identified similar patterns of differences in classification as well as missing data for various grocery stores and supermarkets. In addition to differences in classification, we found through name recognition and review that some establishments were not included in either data source. Our study team was able to fill in important gaps in information by using more than one data source, particularly when using historical data. Since our study relied upon secondary data and many other studies may use secondary data sources to understand such environments, we could not control for these inaccuracies.

In our study, we used a local geocoding reference file provided by the Allegheny county GIS department to geocode establishments. We were able to use local street and rooftop point data developed by our local county/city planning. Due to potential changes in streets and new developments in our area, local street network data provides the most up-to-date information for accurate mapping. Local street and rooftop point data may be of better quality (Roongpiboonsopit

and Karimi, 2010), especially given rigorous processes are in place to validate and update information. Geocoding software companies may not have more recent changes in an area and would, therefore, be unable to geocode accurately. Researchers should attempt to geocode addresses based on local street files when available rather than relying solely on the files provided by national software packages.

Finally, we examined potential visual differences in densities using maps of census tracts across the county. We were able to see some patterns across regions based on location within the county. Further work could examine whether these differences or patterns are systematic or based on other geographic characteristics. Future studies attempting to validate their secondary data through "ground-truthing" of a subset of communities in their area, considering other characteristics of the neighbourhood, such as the sociodemographics of the neighbourhood, may also provide additional insight into systematic differences across data sources.

Conclusion

The approach and rationale for combining commercial data to construct a database of the neighbourhood food and alcohol environment can be applied in other communities and geographic regions particularly for studies interested in understanding trends over time. For studies that cannot audit communities or conduct primary data collection of the resources in all regions of interests, these methods will be useful. By applying multiple data sources, we produced more establishments and coverage than relying only on one data source, potentially having an influence on how neighbourhoods are ultimately characterised. Finally, future studies should consider how differences across data sources may be associated with other neighbourhood characteristics and the relationship with important health outcomes.

Acknowledgements

References

Ahern J, Margerison-Zilko C, Hubbard A, Galea S, 2013. Alcohol outlets and binge drinking in urban neighborhoods: the implications of nonlinearity for intervention and policy. Am J Public Health 103, e81-e87.

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, 175-190.

Gustafson A, Hankins S, Jilcott S, 2012a. Measures of the consumer food store environment: a systematic review of the evidence 2000-2011. J Community Health, 37 897-911.

Gustafson A, Sharkey J, Samuel-Hodge CD, Jones-Smith JC, Cai J, Ammerman AS, 2012b. Food store environment modifies intervention effect on fruit and vegetable intake among low-income women in North Carolina. J Nutr Metab 2012, 932653.

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.

Janevic T, Borrell LN, Savitz DA, Herring AH, Rundle A, 2010. Neighbourhood food environment and gestational diabetes in New York City. Paediatr Perinat Epidemiol 24, 249-254.

Karimi HA, Durcik M, Rasdorf W, 2004. Evaluation of uncertainties associated with geocoding techniques. Comput-Aided Civ Inf 19, 170-185.

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, 1324-1333.

Ohri-Vachaspati P, Martinez D, Yedidia MJ, Petlick N, 2011. Improving data accuracy of commercial food outlet databases. Am J Health Promot 26, 116-122.

Roongpiboonsopit D, Karimi HA, 2010. Quality assessment of online street and rooftop geocoding services. Cartogr Geogr Inf Sci 37, 301-318.

Shier V, An R, Sturm R, 2012. Is there a robust relationship between neighbourhood food environment and childhood obesity in the USA? Public Health 126, 723-730.

Shimotsu ST, Jones-Webb RJ, Maclehose RF, Nelson TF, Forster JL, Lytle LA, 2013. Neighborhood socioeconomic characteristics, the retail environment, and alcohol consumption: a multilevel analysis. Drug Alcohol Depend 132, 449-456.

Supplement

Table A. SIC classifications and categorisations used for establishments.

Category	SIC	SIC EXT	SIC Title
On-premise alcohol outlet	**5813**		**Drinking places - Alcoholic beverages**
		581301	Bars
		581302	Discotheques
		581303	Cocktail lounges
		581304	Night clubs
		581305	Pubs
		581306	Cabarets
		581307	Comedy clubs
		581308	Night clubs information service
		581309	Karaoke clubs
		58130000	Drinking places
		58130100	Bars and lounges
		58130101	Bar (drinking places)
		58130102	Beer garden (drinking places)
		58130103	Cocktail lounge
		58130104	Saloon
		58130105	Tavern (drinking places)
		58130106	Wine bar
		58130200	Night clubs
		58130201	Cabaret
		58130202	Discotheque
Off-premise alcohol outlet	**5181**		**Wholesale beer and ale**
		518101	Beer & ale-wholesale
		51810000	Beer and ale
		51819900	Beer and ale, (NEC)
		51819901	Ale
		51819902	Beer and other fermented malt liquors
		51819903	Porter
	5182		**Wholesale wines and distilled alcoholic beverages**
		518201	Liquors-wholesale
		518202	Distillers representatives
		518203	Wines-wholesale
		518204	Wine brokers
		518205	Distillery equipment & supplies (WHLS)
		518206	Brewers equipment & supplies (WHLS)
		518207	Wines-sacramental (WHLS)
		51820000	Wine and distilled beverages
		51820100	Wine
		51820101	Brandy and brandy spirits
		51820102	Wine coolers, alcoholic
		51820200	Liquor
		51820201	Cocktails, alcoholic: premixed
		51820202	Neutral spirits
		51829900	Wine and distilled beverages, (NEC)
		51829901	Bottling wines and liquors

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
Off-premise alcohol outlet	**5921**		**Liquor stores**
		592101	Cocktail mixes
		592102	Liquors-retail
		592103	Wines-retail
		592104	Beer & ale-retail
		592105	Cordials
		592106	Daiquiri shops
		592107	Tasting rooms
		59210000	Liquor stores
		59210100	Wine and beer
		59210101	Beer (packaged)
		59210102	Wine
		59219900	Liquor stores, (NEC)
		59219901	Hard liquor
Wholesale foods, suppliers and distributors	**5141**		**Wholesale groceries - general line**
		514101	Food products (WHLS)
		514102	Food brokers
		514103	Foods-dehydrated (WHLS)
		514104	Mexican food products-wholesale
		514105	Grocers-wholesale
		514106	Delicatessens-wholesale
		514107	Bulk foods-wholesale
		514108	Delicatessen products-wholesale
		514109	Markets & grocers equip & supls (WHLS)
		514110	Brokers-grocery
		514111	Food agents
		514112	Food service-supplies (WHLS)
		514113	Kosher food products-wholesale
		514114	Macaroni products (WHLS)
		514115	Horseradish (WHLS)
		514116	Japanese food products-whls & mfrs
		51410000	Groceries, general line
		51419900	Groceries, general line, (NEC)
		51419901	Food brokers
	5142		**Wholesale packaged frozen foods**
		514201	Frozen foods-wholesale
		514202	Fruit juices (WHLS)
		514203	Meats-frozen (WHLS)
		51420100	Frozen vegetables and fruit products
		51420101	Fruit juices, frozen
		51420102	Fruit pies, frozen
		51420103	Fruits, frozen
		51420104	Vegetables, frozen
		51420200	Frozen fish, meat and poultry
		51420201	Fish, frozen: packaged
		51420202	Meat pies, frozen
		51420203	Meat, frozen: packaged
		51420204	Poultry pies, frozen
		51420205	Poultry, frozen: packaged

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		51429900	Packaged frozen goods, (NEC)
		51429901	Bakery products, frozen
		51429902	Dinners, frozen
		51429903	Soup, frozen
	5143		**Wholesale dairy products except dried or canned**
		514301	Cheese-wholesale
		514302	Butter (WHLS)
		514303	Dairy products-wholesale
		514304	Ice cream & frozen desserts-distributors
		514305	Ice cream mixes (WHLS)
		514306	Yogurt-wholesale
		514307	Milk buyers
		514308	Cheese-importing (WHLS)
		51430100	Milk
		51430101	Milk and cream, fluid
		51430102	Milk cooling stations
		51430103	Milk depot
		51439900	Dairy products, except dried or canned (NEC)
		51439901	Butter
		51439902	Cheese
		51439903	Dairy depot
		51439904	Frozen dairy desserts
		51439905	Ice cream and ices
		51439906	Yogurt
	5144		**Wholesale poultry and poultry products**
		514401	Poultry brokers
		514402	Poultry services (WHLS)
		514403	Eggs (WHLS)
		514404	Poultry-wholesale
		51440000	Poultry and poultry products
		51449900	Poultry and poultry products, (NEC)
		51449901	Eggs
		51449902	Eggs:cleaning, oil treating, packing, and grading
		51449903	Poultry products, (NEC)
		51449904	Poultry: live, dressed or frozen (unpackaged)
	5145		**Wholesale confectionary**
		514501	Candy & confectionery-wholesale
		514502	Candy & confectionery-brokers
		514503	Concessionaires equip & supplies (WHLS)
		514504	Potato chips (WHLS)
		514505	Popcorn & popcorn supplies-wholesale
		514506	Syrups-wholesale
		514507	Pretzels (WHLS)
		514508	Candy & confectionery-mfrs supls (WHLS)
		514509	Ice cream cone-distributors
		514510	Nuts-edible-wholesale & processing
		514511	Corn chips (WHLS)
		514512	Pistachios
		51450000	Confectionery

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		51450100	Fountain supplies
		51450101	Fruits, fountain
		51450102	Syrups, fountain
		51450103	Toppings, soda fountain
		51450200	Snack foods
		51450201	Corn chips
		51450202	Nuts, salted or roasted
		51450203	Popcorn and supplies
		51450204	Potato chips
		51450205	Pretzels
		51459900	Confectionery, (NEC)
		51459901	Candy
		51459902	Chewing gum
	5146		**Wholesale fish and seafoods**
		514601	Seafood-wholesale
		514602	Oysters-wholesale
		514603	Shrimp-wholesale
		514604	Fish & seafood-brokers
		514605	Fish buyers-fresh water
		51460000	Fish and seafoods
		51469900	Fish and seafoods, (NEC)
		51469901	Fish, cured
		51469902	Fish, fresh
		51469903	Fish, frozen, unpackaged
		51469904	Seafoods
	5147		**Wholesale meats and meat products**
		514701	Butchering
		514702	Meat jobbers
		514703	Meat brokers
		514704	Meat-wholesale
		514705	Sausages-wholesale
		514706	Pork dealers-wholesale
		514707	Frankfurters & rolls (WHLS)
		514708	Bacon & ham curers & suppliers
		514711	Meat-kosher-wholesale
		51470000	Meats and meat products
		51479900	Meats and meat products, (NEC)
		51479901	Lard
		51479902	Meat brokers
		51479903	Meats, cured or smoked
		51479904	Meats, fresh
	5148		**Wholesale fresh fruits and vegetables**
		514801	Fruits & vegetables-wholesale
		514802	Potatoes-wholesale
		514803	Produce-brokers
		514804	Fruits & vegetables-brokers
		514805	Fruits & vegetables-growers & shippers
		514806	Citrus buyers
		514807	Citrus fruits (WHLS)

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		514808	Bananas-wholesale
		514809	Citrus fruit products-wholesale
		514810	Fruit baskets-gift-wholesale
		514811	Apple brokers
		514812	Tomatoes-wholesale
		514813	Apple shippers
		514814	Cherries-maraschino (WHLS)
		514815	GARDENS-U-PICK (WHLS)
		51480000	Fresh fruits and vegetables
		51480100	Fruits
		51480101	Banana ripening
		51480102	Fruits, fresh
		51480200	Vegetables
		51480201	Potatoes, fresh
		51480202	Vegetables, fresh
	5149		**Wholesale groceries and related products not elsewhere classified**
		514901	Molasses (WHLS)
		514902	Bakers-wholesale
		514903	Baby foods (WHLS)
		514904	Chili & chili products (WHLS)
		514905	Coffee & tea-wholesale
		514906	Chocolate & cocoa (WHLS)
		514907	Commission merchants
		514908	Coffee roasting (WHLS)
		514909	Condiments & sauces (WHLS)
		514910	Pet foods-wholesale
		514911	Foods-foreign (WHLS)
		514912	Fruits-dried (WHLS)
		514913	Flour (WHLS)
		514914	Health food products-wholesale
		514915	Malts & hops (WHLS)
		514916	Honey (WHLS)
		514917	Yeast (WHLS)
		514918	Oils-vegetable (WHLS)
		514919	Olives (WHLS)
		514920	Pizza-wholesale
		514921	Rice-wholesale
		514922	Pizza equipment & supplies (WHLS)
		514923	Salads (WHLS)
		514924	Soft drinks (WHLS)
		514925	Sandwiches-wholesale
		514926	Tea-wholesale
		514927	Sugar-brokers & wholesalers
		514928	Tortillas (WHLS)
		514929	Clams (WHLS)
		514930	Pasta (WHLS)
		514931	Soybean products (WHLS)
		514932	Flavoring extracts (WHLS)

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		514933	Mayonnaise & salad dressings (WHLS)
		514934	Tamales (WHLS)
		514935	Ravioli (WHLS)
		514936	Corn products (WHLS)
		514937	Beverages (WHLS)
		514938	Spices-wholesale
		514939	Food service-distributors
		514940	Bagels-wholesale
		514941	Barbecue-wholesale (sauce)
		514942	Bird feed-wholesale
		514943	Canned goods-wholesale
		514944	Chinese foods-wholesale
		514945	Coffee break service & supplies-wholesale
		514946	Doughnuts-wholesale
		514947	Food specialties-wholesale
		514948	Maple sugar & syrup-wholesale
		514949	Natural foods-wholesale
		514950	Pickles & pickle products-wholesale
		514951	Pies-wholesale
		514952	Jams & preserves
		514953	Provisions-wholesale
		514954	Salt-wholesale
		514955	Wild rice-wholesale
		514956	Flavoring extracts-raw materialswholesale
		514957	Food facilities (WHLS)
		514958	Cocoa (WHLS)
		514959	Cookies & crackers-wholesale
		514960	Tea bag packers (WHLS)
		514961	Distributor-groceries
		514962	Cider-wholesale
		514963	Milk-powdered (WHLS)
		514964	Italian food products-wholesale
		514965	Malted milk (WHLS)
		514966	Homeopathic suppliers (WHLS)
		514967	Box lunches-wholesale
		514968	Oils-edible (WHLS)
		514969	Olive oil wholesale
		514970	Baby foods retail
		51490000	Groceries and related products, (NEC)
		51490100	Cooking oils and shortenings
		51490101	Cooking oils
		51490102	Margarine
		51490103	Oleomargarine
		51490104	Shortening, vegetable
		51490200	Pet foods
		51490201	Cat food
		51490202	Dog food
		51490300	Seasonings, sauces, and extracts
		51490301	Condiments

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		51490302	Flavourings and fragrances
		51490303	Fruit peel
		51490304	Hop extract
		51490305	Malt
		51490306	Malt extract
		51490307	Salad dressing
		51490308	Salt, edible
		51490309	Sauces
		51490310	Spices and seasonings
		51490400	Pasta and rice
		51490401	Macaroni
		51490402	Rice, polished
		51490403	Spaghetti
		51490500	Beverages, except coffee and tea
		51490501	Beverage concentrates
		51490502	Juices
		51490503	Mineral or spring water bottling
		51490504	Soft drinks
		51490505	Water, distilled
		51490600	Organic and diet food
		51490601	Diet foods
		51490602	Health foods
		51490603	Natural and organic foods
		51490604	Specialty food items
		51490700	Crackers, cookies, and bakery products
		51490701	Bakery products
		51490702	Cookies
		51490703	Crackers
		51490800	Dried or canned foods
		51490801	Canned goods: fruit, vegetables, seafood, meats, etc.
		51490802	Dairy products, dried or canned
		51490803	Fruits, dried
		51490804	Milk, canned or dried
		51490805	Soups, except frozen
		51490900	Coffee and tea
		51490901	Coffee, green or roasted
		51490902	Tea
		51490903	Tea bagging
		51491000	Sugar, honey, molasses, and syrups
		51491001	Honey
		51491002	Molasses, industrial
		51491003	Sugar, refined
		51491004	Syrups, except for fountain use
		51491100	Baking supplies
		51491101	Flour
		51491102	Starch
		51491103	Wet corn milling products
		51491104	Yeast
		51499900	Groceries and related products, (NEC)

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		51499901	Breakfast cereals
		51499902	Chocolate
		51499903	Cocoa
		51499904	Food gift baskets
		51499905	Pickles, preserves, jellies, and jams
		51499906	Pizza supplies
		51499907	Sandwiches
		51499908	Sausage casings
		51499909	Wine makers' equipment and supplies
		51499910	Sandwich supplies
		51499911	Breading mixes
General Convience Stores	**5331**		**Variety stores**
		533101	Variety stores
		533102	Trading posts
		533103	Soap mitts
		533104	Clothes posts
		533105	Hula supplies-retail
		53310000	Variety stores
	5399		**Miscellaneous general merchandise store**
		539901	General merchandise-retail
		539902	Country stores
		539903	Bead strings
		539904	Tote boxes pans & trays
		539905	Farmers co-op retail stores
		539906	Car seats-children
		539907	Sponges-retail
		53990000	Miscellaneous general merchandise stores
		53999900	Miscellaneous general merchandise stores, (NEC)
		53999901	Army-Navy goods stores
		53999902	Catalog showroom stores
		53999903	Country general stores
		53999904	Duty-free stores
		53999905	Surplus and salvage stores
		53999906	Warehouse club stores
	5411		**Grocery stores**
		541103	Convenience stores
		54110200	Convenience stores
		54110201	Convenience stores, chain
		54110202	Convenience stores, independent
Gas Stations	**5541**		**Gasoline service stations**
		554101	Service stations-gasoline & oil
		554102	Gas-diesel
		554103	Truck stops & plazas
		554104	Gas-leaded/lead-free
		554105	Kerosene
		554106	Service stations-marine
		554107	Oils-lubricating-retail
		554110	Alternative fuels
		554111	Diesel exhaust fluid

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		554112	Electric charging station
		55410000	Gasoline service stations
		55419900	Gasoline service stations, (NEC)
		55419901	Filling stations, gasoline
		55419902	Marine service station
		55419903	Truck stops
Grocery Stores/Supermarkets & other food markets	**5411**		**Grocery stores**
		541101	Food markets
		541102	Snack products
		541104	Food products-retail
		541105	Grocers-retail
		541106	Markets-kosher
		541107	Grocers-ethnic foods
		541108	Grocers-health foods
		541109	Grocers-take-out foods
		54110000	Grocery stores
		54110100	Supermarkets
		54110101	Supermarkets, chain
		54110102	Supermarkets, greater than 100,000 square feet (hypermarket)
		54110103	Supermarkets, independent
		54110104	Supermarkets, 55,000-65,000 square feet (superstore)
		54110105	Supermarkets, 66,000-99,000 square feet
		54119900	Grocery stores, (NEC)
		54119901	Cooperative food stores
		54119902	Delicatessen stores
		54119903	Frozen food and freezer plans, except meat
		54119904	Grocery stores, chain
		54119905	Grocery stores, independent
	5421		**Meat and seafood markets including freezer provisioners**
		542101	Seafood-retail
		542102	Food plans
		542103	Frozen foods-retail
		542104	Meat cutting service
		542105	Lobsters
		542106	Ham specialty stores
		542107	Meat-retail
		542108	Sausages
		542109	Shrimp-retail
		542110	Caviar
		542111	Meat markets-kosher
		542113	Hams
		542114	Smoked foods
		542115	Meat for freezers
		542116	Crab meat
		542117	Meats-barbecued
		542118	Smoking & curing service

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		542119	Freezers-provisioning
		54210000	Meat and fish markets
		54210100	Fish and seafood markets
		54210101	Fish markets
		54210102	Seafood markets
		54210200	Meat markets, including freezer provisioners
		54210201	Food and freezer plans, meat
		54210202	Freezer provisioners, meat
	5431		**Fruit and vegetable markets**
		543101	Fruits & vegetables & produce-retail
		543102	farm markets
		543103	Fruit packages
		543104	Juices-retail
		543105	Fruits-glace
		543106	Pick-your-own fruit & vegetables
		54310000	Fruit and vegetable markets
		54319900	Fruit and vegetable markets, (NEC)
		54319901	Fruit stands or markets
		54319902	Vegetable stands or markets
	5441		**Candy, nut and confectionery stores**
		544101	Candy & confectionery-retail
		544102	Nuts-edible
		544103	Popcorn & popcorn supplies
		544104	Candy making supplies
		544105	Wedding candy & confectioneries
		544106	Candy & cookie arrangements
		54410000	Candy, nut, and confectionery stores
		54419900	Candy, nut, and confectionery stores, (NEC)
		54419901	Candy
		54419902	Confectionery
		54419903	Confectionery produced for direct sale on the premises
		54419904	Nuts
		54419905	Popcorn, including caramel corn
	5451		**Dairy products stores**
		545101	Dairy products-retail
		545102	Yogurt
		545103	Cheese
		54510000	Dairy products stores
		54519900	Dairy products stores, (NEC)
		54519901	Butter
		54519902	Cheese
		54519903	Ice cream (packaged)
		54519904	Milk
	5461		**Retail bakeries**
		546101	Bagels
		546102	Bakers-retail
		546103	Bakers-cake & pie
		546104	Pies
		546105	Doughnuts

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		546106	Cake decorating
		546107	Cookies & crackers
		546108	Pretzels-retail
		546109	Cookie shops
		546110	Muffins
		546111	Matzos
		546112	Cake decorating instruction & supplies
		54610000	Retail bakeries
		54619900	Retail bakeries, (NEC)
		54619901	Bagels
		54619902	Bread
		54619903	Cakes
		54619904	Cookies
		54619905	Doughnuts
		54619906	Pastries
		54619907	Pies
		54619908	Pretzels
	5499		**Miscellaneous food stores**
		549901	Health & diet foods-retail
		549902	Spices
		549903	Water companies-bottled, bulk, etc
		549904	Vitamins
		549905	Weight control preparation
		549906	Salt
		549907	Poultry-retail
		549908	Food supplements
		549909	Foods-natural
		549910	Food specialties-retail
		549911	Game birds venison etc
		549912	Kosher foods
		549913	Herbs
		549914	Eggs-retail
		549915	Coffee & tea
		549916	Oriental food products
		549917	Chinese food products
		549918	Oriental goods
		549919	Japanese food products
		549920	Gourmet shops
		549921	Imported foods
		549922	Coffee roasting & handling equipment
		549923	Korean foods
		549924	Seeds-chinese preserved
		549925	Distilled water
		549926	Hawaiian foods
		549927	Mexican & latin american food products
		549928	Dietetic food products
		549929	Coconut products
		549930	British food products
		549931	Meat tenderizing preparations

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		549932	Soyfoods
		549933	Vegetable juices
		549934	Vitamins-raw materials
		549935	Organic foods & services
		549936	Nutrients
		549937	Vietnamese foods
		549938	Espresso & espresso bars
		549939	Ukrainian foods
		549940	Low carb foods & programs
		549941	Thai food
		549942	Beach service
		549943	Awareness products
		549944	Fitness supplements
		549945	Halal foods
		549946	Native american foods
		549947	Caribbean foods
		549999	Miscellaneous food stores
		54990000	Miscellaneous food stores
		54990100	Health and dietetic food stores
		54990101	Dietetic foods
		54990102	Health foods
		54990103	Vitamin food stores
		54990200	Beverage stores
		54990201	Coffee
		54990202	Juices, fruit or vegetable
		54990203	Soft drinks
		54990204	Tea
		54990205	Water: distilled mineral or spring
		54999900	Miscellaneous food stores, (NEC)
		54999901	Dried fruit
		54999902	Eggs and poultry
		54999903	Food gift baskets
		54999904	Gourmet food stores
		54999905	Spices and herbs
Restaurants (this includes all eating places)	**5812**		**Eating places**
		581201	Luaus
		581202	Greek food products
		581203	Ice cream parlors
		581204	Wedding cakes
		581205	Hamburger & hot dog stands
		581206	Foods-carry out
		581207	Foods-institutional
		581208	Restaurants
		581209	Delicatessens
		581210	Commissary contractors
		581211	Chicken dinners
		581212	Caterers
		581213	Cafeterias
		581214	Cafes

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		581215	Box lunches
		581216	Banquets-arranged
		581217	Appetizers & snacks etc
		581218	Soda fountain shops
		581219	Sandwiches
		581220	Restaurant management
		581221	Refreshment stands
		581222	Pizza
		581223	Banquet rooms
		581224	Barbecue
		581225	Beverages-non-alcoholic-retail
		581226	Theatres-dinner
		581227	Italian food products
		581228	Coffee shops
		581229	Deli-bakery
		581230	Restaurants-foods delivery
		581231	Food service-industrial
		581232	Chow mein & chop suey
		581233	Jewish foods
		581234	Reception centers
		581236	Tea rooms
		581237	Dessert preparations
		581238	Cafeterias-industrial-operators
		581239	Catering-kosher
		581240	Chuck wagon dinners
		581241	Milk bars
		581242	Mobile concessions
		581243	Luncheonette food supply
		581244	Oyster bars
		581245	Escargots
		581246	Chili parlors
		581248	Juice bars
		581249	Restaurants-reservations
		581250	Wedding rehearsal dinners-restaurants
		581251	Restaurants-family dining
		581252	Cajun products
		581253	Meal preparation services
		581254	Restaurants-cyber cafes
		581255	Hookah bars & lounges
		58120000	Eating places
		58120100	Ethnic food restaurants
		58120101	American restaurant
		58120102	Cajun restaurant
		58120103	Chinese restaurant
		58120104	French restaurant
		58120105	German restaurant
		58120106	Greek restaurant
		58120107	Indian/Pakistan restaurant
		58120108	Italian restaurant

WHLS = wholesale; NEC = not elsewhere classified.

continued

Category	SIC	SIC EXT	SIC Title
		58120109	Japanese restaurant
		58120110	Korean restaurant
		58120111	Lebanese restaurant
		58120112	Mexican restaurant
		58120113	Spanish restaurant
		58120114	Sushi bar
		58120115	Thai restaurant
		58120116	Vietnamese restaurant
		58120117	Pakistani restaurant
		58120200	Ice cream, soft drink and soda fountain stands
		58120201	Concessionaire
		58120202	Frozen yogurt stand
		58120203	Ice cream stands or dairy bars
		58120204	Snow cone stand
		58120205	Soda fountain
		58120206	Soft drink stand
		58120300	Fast food restaurants and stands
		58120301	Box lunch stand
		58120302	Carry-out only (except pizza) restaurant
		58120303	Chili stand
		58120304	Coffee shop
		58120305	Delicatessen (eating places)
		58120306	Drive-in restaurant
		58120307	Fast-food restaurant, chain
		58120308	Fast-food restaurant, independent
		58120309	Food bars
		58120310	Grills (eating places)
		58120311	Hamburger stand
		58120312	Hot dog stand
		58120313	Sandwiches and submarines shop
		58120314	Snack bar
		58120315	Snack shop
		58120400	Lunchrooms and cafeterias
		58120401	Automat (eating places)
		58120402	Cafeteria
		58120403	Luncheonette
		58120404	Lunchroom
		58120405	Restaurant, lunch counter
		58120500	Family restaurants
		58120501	Restaurant, family: chain
		58120502	Restaurant, family: independent
		58120600	Pizza restaurants
		58120601	Pizzeria, chain
		58120602	Pizzeria, independent
		58120700	Seafood restaurants
		58120701	Oyster bar
		58120702	Seafood shack
		58120800	Steak and barbecue restaurants
		58120801	Barbecue restaurant

WHLS = wholesale; NEC = not elsewhere classified.

Category	SIC	SIC EXT	SIC Title
		58120802	Steak restaurant
		58129900	Eating places, (NEC)
		58129901	Buffet (eating places)
		58129902	Cafe
		58129903	Caterers
		58129904	Chicken restaurant
		58129905	Commissary restaurant
		58129906	Contract food services
		58129907	Diner
		58129908	Dinner theater
		58129909	Health food restaurant
Pharmacies and Drug Stores	**5912**		**Drug Stores and Proprietary Stores**
		591201	Medicines-patent & proprietary
		591202	Health care products
		591203	First aid supplies
		591204	Elastic stockings
		591205	Pharmacies
		591206	Toilet articles
		591207	Pharmaceutical consultants
		591208	Razor sharpeners & stropper
		591209	Allergy resistant products
		591210	Convalescent supplies
		591211	Drugs-crude
		591212	Homeopathic remedies
		591213	Sun tan supplies
		591214	Compounding
		59120000	Drug stores and proprietary stores
		59129900	Drug stores and proprietary stores, (NEC)
		59129901	Drug stores
		59129902	Proprietary (non-prescription medicine) stores

WHLS = wholesale; NEC = not elsewhere classified.

Table B. Various establishments and their primary SIC classifications for InfoUSA and D&B.

InfoUSA Company Name	D&B Company Name	D&B Trade Name	InfoUSA Primary SIC code	D&B Primary SIC code	Final Category for Primary SIC code (based on InfoUSA)
Pop & Beer For Less	Mckees Rocks Pop & Beer For Le	Pop & Beer 4 Less	518101	5499	OffSiteAlc
Salute Distributing Inc	Salute Distributing Inc		518101	5499	OffSiteAlc
Ziegler's Uncommon Market	Zieglers Uncommon Market		514905	5499	Wholesale
Chips	Chips Inc	Value Added Foods	514101	5411	Wholesale
Schwan's Sales Enterprises	Schwan Food Company		514101	5411	Wholesale
Louis Greenwald Inc	Greenwald Louis Incorporated	Bells Market	514704	5411	Wholesale
Schwebel Baking Co	Schwebel Baking Co		514902	5461	Wholesale
Fortunes International Tees	Fortunes International Tees		514905	5499	Wholesale
Coffee Tree Roasters	Coffee Tree Roasters Corp	Warehouse	514905	5499	Wholesale
Asian Foods	Asian Foods Inc		514105	5411	Wholesale
Guy's Place Bakery	Meaj Inc	Guys Place Bakery	514902	5461	Wholesale
Mediterra Bakehouse	Mediterra Bakehouse		514902	5461	Wholesale
Old Vienna Baking Co	Feigs Bakery Inc	Old Vienna Baking Company	514902	5461	Wholesale
Herky's Food Products	Herkys Food Products Inc		514101	5421	Wholesale
Kramer Brothers Produce Co	Kramer Brothers Produce Co		514801	5431	Wholesale
Folino Brothers	Folino Brothers Produce	Folino Bros	514801	5421	Wholesale
Nickles Bakery Inc	Nickles Alfred Bakery Inc	Nickles Bakery	514902	5461	Wholesale
Bruegger's Bagels Bakery	Brueggers Bagels	Brueggers Bagels	514902	5461	Wholesale
O K Grocery	Giant Eagle Inc	Giant Eagle/O K Grocery	514105	5411	Wholesale
Tristate Petroleum Inc	Exxon Tristate Petroleum	Monroeville Exxon Shop	554101	5411	Gas
Buy 'n Fly	Mon Valley Petroleum Inc	Buy N Fly	554101	5411	Gas
Katie's Korner	Katies Korner		554101	5411	Gas
T J's Deli Mart	Kehm Oil Company	T Js Deli	554101	5411	Gas
Renaissance Deli Provision	Renaissance Deli Provision		581209	5411	Restaurant
Dairy Queen	Dairy Queen Grt Southern Shop	Dairy Queen	581203	5451	Restaurant
Rythm House Cafe	Rythm House Cafe		581206	5411	Restaurant
North Shore Delicatessen	North Shore Deli		581209	5411	Restaurant
Castellano's Deli	Castellanos Deli Inc		581208	5411	Restaurant
Billy's Troy Hill Bistro	Troy Hill Deli		581208	5411	Restaurant
Clairton's Fish & More	Clairtons Fish & More		581208	5421	Restaurant
Katerbean	1106 South Braddock Ave Inc	Katerbean	581228	5499	Restaurant
Rosebud Deli	Rosebud Deli Inc	Rosebud Deli Liberty Inc	581209	5411	Restaurant
Millie's Italian Deli	Millies Italian Deli		581209	5411	Restaurant
Au Bon Pain	A B P Corporation	Au Bon Pain	581208	5461	Restaurant
6th Ave Cafe & Deli	6th Ave Cafe & Deli Inc		581208	5411	Restaurant
Coffee A La Cart	Coffee A La Cart Inc	Cool Beans Coffee Co	581228	5499	Restaurant
Isaly Co	Isalys Store 170	Isaly	581209	5411	Restaurant
Deli Junction	Deli Junction LLC		581212	5411	Restaurant
Brown Bag Deli	Brown Bag Inc		581209	5411	Restaurant
Vento's Pizza	Vento Albert P	Ventos Dairy Delite	581222	5451	Restaurant
Grace Market	Grace Market		581209	5411	Restaurant
Valentina's Pizzeria & Deli	Valentinas Pizzeria And Deli		581209	5411	Restaurant

InfoUSA Company Name	D&B Company Name	D&B Trade Name	InfoUSA Primary SIC code	D&B Primary SIC code	Final Category for Primary SIC code (based on InfoUSA)
Cappo's Pizza & 5 Points Deli	Cappos Pizza & Five Point Deli	Cappos Pizza & 5 Point Deli	581208	5411	Restaurant
Bruster's Old Fashioned Ice	Brusters Old Fshiond Ice Crea		581203	5451	Restaurant
Dairy Queen	Dairy Queen Of Northern Pike	Dairy Queen	581203	5451	Restaurant
Carson Street Deli	Carson Street Deli		581212	5411	Restaurant
Panera Bread	Covelli Enterprises Inc	Panera Bread	581208	5461	Restaurant
Speedy Gourmet	Speedy Gourmet Inc		581209	5411	Restaurant
Panera Bread	Covelli Enterprises Inc	Panera Bread	581208	5461	Restaurant
Pittsburgh Presse Deli	Pittsburgh Presse Deli Holding		581209	5411	Restaurant
Sam's Market	Sams Market		581208	5411	Restaurant
La Prima Espresso Co	La Prima Expresso Co	La Prima Espresso	581208	5499	Restaurant
Dairy Queen	Dairy Queen Of South Side	Dairy Queen	581203	5451	Restaurant
Folino's Deli	Folinos Deli		581209	5411	Restaurant
Bruster's Old Fashioned Ice	Stacy Jos Ice Cream Inc	Brewsters	581203	5451	Restaurant
Cafe Venice Pizzeria & Deli	Cafe Venice Pizzeria & Deli		581208	5411	Restaurant
Victoria's Deli	Victorias Deli		581209	5411	Restaurant
Sal's Scotch Bottom Deli	Sals Scotch Bottom Deli LLC		581209	5411	Restaurant
Kazansky's Delicatessen	Bennett Brothers Inc	Kazanskys Delicatessen	581209	5411	Restaurant
61 C Cafe	61 C Cafe Inc		581228	5499	Restaurant
Moses Quality Market	Moses Quality Market		581209	5411	Restaurant
Panera Bread	Covelli Enterprises Inc	Panera Bread	581208	5461	Restaurant
Najanick & Ross Food & Snack	Najanick & Ross Food & Snack S		581209	5411	Restaurant
Scoops On Beverly	Scoops On Beverly		581203	5451	Restaurant
Panera Bread	Covelli Enterprises Inc		581208	5461	Restaurant
New Way Gourmet Deli & Rstrnt	New Way Gorment Food & Deli	New Way Gourmet	581208	5411	Restaurant
Webster Hall Deli Inc	Webster Hall Deli Inc		581209	5411	Restaurant
La Prima Espresso Co	La Prima Expresso Co	Cafe Zio	581208	5499	Restaurant
Salim's Middle Eastern Food	Salims Middle Eastern Food Co		581209	5411	Restaurant
Krispy Kreme Doughnuts	Metz & Associates Ltd	Century III Krispy Kreme 1101	581208	5461	Restaurant
Kribels Bakery	Kribels Bakery Inc	Kribel Bakery	581208	5461	Restaurant
Gloria Jean's Gourmet Coffees	Dan-Cam Inc	Gloria Jeans Gourmet Coffee	581228	5499	Restaurant
American Body Building-Pa W	American Body Building		581225	5499	Restaurant
Panera Bread	Covelli Enterprises Inc		581208	5461	Restaurant
Rita's Italian Ice	Ritas Italian Ices	Ritas Italian Ice	581203	5451	Restaurant
Pino's Mercato	Pinos Mercato Inc		581212	5411	Restaurant
Sandy's Deli	Sandys Delicatessen Inc		581209	5411	Restaurant
Sharon's Deli Mart	Sharon Deli Mart		581209	5411	Restaurant
Rita's Italian Ice	Ritas Italian Ice	Ritas Italian Ice	581203	5451	Restaurant
Martin's Fruit Basket	Martins Fruit Baskets	Martins Cafe-Deli	581209	5431	Restaurant
Liberty Dairy Store	Liberty Dairy Store		581209	5411	Restaurant
Alfred's Deli Plus	Alfreds Deli Plus		581209	5411	Restaurant
Nirvana Deli	Nirvana Deli		581208	5411	Restaurant
Campus Deli	Campus Deli		581208	5411	Restaurant
Panera Bread Cafe'	Covelli Enterprises Inc		581214	5461	Restaurant
Panera Bread	Covelli Enterprises Inc		581208	5461	Restaurant

InfoUSA Company Name	D&B Company Name	D&B Trade Name	InfoUSA Primary SIC code	D&B Primary SIC code	Final Category for Primary SIC code (based on InfoUSA)
Produce Plus	Produce Plus		581209	5431	Restaurant
Panera Bread	Covelli Enterprises Inc		581208	5461	Restaurant
Green Chef	Green Chef		581209	5411	Restaurant
Community Supermarket	Community Supermarket Inc		581209	5411	Restaurant
Panera Bread	Covelli Enterprises Inc	Panera Bread	581208	5461	Restaurant
Hot Tops Cake N Bake Pizza	Cal-A-Donut Inc		581208	5461	Restaurant
Smitys Market & Deli	Smitys Deli & Market	Smittys Market & Deli	581209	5411	Restaurant
Steamers Expresso Cafe Ltd	Steamers Espresso Cafe Ltd		581228	5499	Restaurant
Joe's Water Ice	Joes Water Ice		581203	5451	Restaurant
Great Valley Restaurant	Matco Foods Inc	Great Valley Foodland	581208	5411	Restaurant
Uncommon Market	Uncommon Market Inc		581212	5499	Restaurant
Bethel Bakery	Bethel Bakery Inc		581208	5461	Restaurant
Au Bon Pain Co	A B P Corporation	Au Bon Pain	581208	5461	Restaurant
Family Deli	Family Foods Inc	Family Deli	581209	5411	Restaurant
Panera Bread	Covelli Enterprises Inc		581208	5461	Restaurant
Marianne's Cool Treats	Mariannes Cool Treats		581203	5451	Restaurant
Gloria Jeans Gourmet Coffee	Gloria Jeans Gourmet Cof Beans	Gloria Jeans Gourmet Coffee	581228	5499	Restaurant
Frisch's Carrick Poultry	Carrick Poultry	Carrick Poultry Market	581212	5499	Restaurant
Main Street Deli	Main Street Deli		581209	5411	Restaurant
Castellano Deli Mart	Zeiglers Deli Mart		581209	5411	Restaurant
Jack's Pizza & Hoagies	Jacks Pizza & Hoagies		581208	5411	Restaurant
Dairy Queen	Dairy Qeen Brazier Natrona Hts	Dairy Queen	581203	5451	Restaurant
Starr Road Mushroom Market	Starr Road Mushroom Market		581209	5411	Restaurant
Glen's Frozen Custard	Glens Frozen Custard		581203	5451	Restaurant
Baskin-Robbins Ice Cream	George Robert Inc	Donut Connection	581203	5461	Restaurant
Panera Bread	Covelli Enterprises Inc		581208	5461	Restaurant
Boardwalk Custard	Boardwalk Custard		581203	5451	Restaurant
Wexford Post Office Deli	Wexford Po Deli & Catrg	Wexford Post Office Deli	581209	5411	Restaurant
Giant Eagle	Giant Eagle	Edgewood Giant Eagle	591205	5411	Pharmacy
Giant Eagle	Giant Eagle 6379 Inc	Giant Eagle	591205	5411	Pharmacy
Shop 'n Save Pharmacy	Fresh Foods Inc	Shop N Save Mount Lebanon	591205	5411	Pharmacy
Giant Eagle	Giant Eagle Inc	Giant Eagle	591205	5411	Pharmacy
Giant Eagle Pharmacy Dept	Giant Eagle Inc	Giant Eagle	591205	5411	Pharmacy