Neighborhood Racial and Socioeconomic Disparities in the Food and Alcohol Environment: Are There Differences by Commercial Data Sources?

Dara D. Mendez[1] · Kevin H. Kim[2] · Cecily R. Hardaway[3,4] · Anthony Fabio[1]

Abstract

Objectives This study examined neighborhood racial and socioeconomic disparities and the density of food and alcohol establishments. We also examined whether these disparities differed by data source.

Methods This study included commercial data for 2003 and 2009 from InfoUSA and Dun and Bradstreet (D&B) in 416 census tracts in Allegheny County, PA. Food and alcohol establishment densities were calculated by using area and population data from the 2000 US census. Differences between InfoUSA and D&B of food and alcohol densities across neighborhood racial and socioeconomic characteristics were tested using correlations and two-way mixed analysis of variance (ANOVA).

Results There were differences by data source in the association between neighborhood racial and socioeconomic characteristics and food/alcohol establishment density. There was a positive correlation between grocery store/supermarket density and percentage black, poverty, and percentage without a car among D&B data but not in InfoUSA. Alcohol outlet density (AOD) increased as neighborhood poverty increased for both data sources, but the mean difference in AOD between InfoUSA and D&B was highest among neighborhoods with 25–50 % poverty (Cohen's d −0.49, $p < 0.001$) compared to neighborhoods with lower or higher poverty (2003 data). Mean grocery store density increased as percentage poverty increased, but only among D&B (2009 data).

Conclusions Differences in commercial data in the location and numeration of food and alcohol establishments are associated with neighborhood racial and socioeconomic characteristics and may introduce biases concerning neighborhood food and alcohol environments, racial and socioeconomic disparities, and health.

Keywords Food and alcohol · Neighborhood · Disparity

✉ Dara D. Mendez
  ddm11@pitt.edu

1 Graduate School of Public Health, Department of Epidemiology, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261, USA

2 School of Education, Department of Psychology in Education, University of Pittsburgh, Pittsburgh, PA, USA

3 School of Medicine, Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

4 Social Science Research Institute, Duke University, Durham, NC, USA

Background

The influence of food and alcohol environments on health, well-being, and health behaviors has gained considerable attention in the public health discourse [1, 2]. The role of differential exposure or access to neighborhood environments in driving racial and socioeconomic health disparities has also been of concern [3–5]. In the context of food and alcohol environments, healthy food options and alcohol outlets are unequally distributed geographically (e.g., by race, class, and rurality) [6–11]. Inequities in neighborhood food and alcohol environments influence individual behaviors, are related to overall health and well-being, are barriers to healthy living environments, and are cited as important environmental justice issues [8, 11].

Many studies have explored whether neighborhood characteristics, including racial composition and poverty levels, are associated with availability of food (e.g., restaurants, grocery stores, and fast food options) or liquor/alcohol outlets [1,

6–8]. In general, these studies have found that neighborhoods that have predominately black or lower-income residents have fewer outlets that offer healthful food options and more liquor/alcohol outlets than neighborhoods with predominantly white or higher-income residents [1, 7–9, 11–14]. The researchers have also found that racial disparities in access to healthful food options and alcohol/liquor store densities persist even when neighborhood income levels are similar [9, 11, 12, 15].

Many studies use secondary, commercial data, such as InfoUSA and Dun and Bradstreet (D&B), to locate and measure food establishments [8, 16, 17] and secondary, Census business data to measure alcohol establishments [11, 18]. These datasets can be affordable and accessible, particularly in areas where local or publicly available data are limited or when primary data collection is not affordable or feasible. However, studies have pointed to the important differences between the sources of data, including variations in data collection procedures and reporting [17, 19, 20]. Prior studies have also tested the validity of these secondary data sources by comparing them with field surveys or similar methods [16, 17, 21, 22], with some studies suggesting that there may be differences between secondary and survey-based data of food establishments across neighborhood socioeconomic categorizations [6, 16, 23]. Still, little is known about the degree to which locations of food and alcohol establishments generated by secondary, commercial data sources such as InfoUSA and D&B may differ and how they may differ across various neighborhood racial and socioeconomic characteristics.

The purpose of this study was to (1) describe the associations between neighborhood racial and socioeconomic characteristics (i.e., racial composition, percentage poverty, percentage without a car, and city/non-city divisions) and density of food and alcohol establishments in the two different data sources; and (2) determine if these relationships differ by data source for key categories of food and alcohol establishments. We hypothesize that there will be an association between food/alcohol outlet density and neighborhood racial and socioeconomic characteristics, and that this relationship will differ between D&B and InfoUSA datasets. This will be the first study of our knowledge to assess the differences between two the widely used commercial datasets, D&B and InfoUSA, in the density of food and alcohol establishments and if these differences vary by various neighborhood racial and socioeconomic characteristics.

This study was not designed to assess the validity of these data sources against a "gold standard" or data collected through field surveys. However, this study will be useful in assessing and interpreting systematic biases in secondary commercial data of the food and alcohol environment and for understanding how the variations in data sources may be related to other neighborhood factors. Additionally, this study takes into account historic data in understanding changes in environments at two different time points. Finally, many prior studies focus on food outlets, but this study incorporates both food and alcohol outlets, including on (e.g., bars) and off premise (e.g., liquor stores) alcohol establishments, as well as other types of establishments such as pharmacies that may be limited food sources. The variations and limitations in these data may directly influence conclusions drawn about the associations between neighborhood resources such as grocery stores or food and alcohol establishments as well as health outcomes. Understanding these potential differences and biases is important for future research, interventions, and policies based on the results of analyses using these data sources.

Methods

Data Sources

This study includes data from InfoUSA/InfoGroup and D&B of food and alcohol establishments in Allegheny County, PA. InfoUSA and D&B are sales and marketing companies that collect data about businesses in the USA and Canada. The present study includes data collected by these two companies in 2009 and 2003. Although we expect there will be similar establishments during these two time periods, these data are cross sectional. We developed a methodology and process of combining these two commercial data sources to provide the best coverage of the various food and alcohol establishments in the study area [24]. The algorithm we developed as discussed [24] takes into location information, type of establishment, and other qualitative information about the establishment to cross-reference establishments across data sources. We repeated the same process for 2009 and 2003 data. There were a total of 7078 unique establishments identified in 2009 and 8705 unique establishments identified from data in 2003.

We also used data from the US census to calculate population and demographic variables and to calculate the density measures. The 2000 census was used since demographic and population trends between 2000 and 2010 were relatively stable for the region for the specific variables of interest for this study. Also, our key indicators for food and alcohol establishment density were based on 2003 and 2009 data before the 2010 census data were released. The unit of analysis for this study was the census tract, which was used as a proxy for neighborhood in this study. Census tracts are geographic units defined by the US census and include an average of 4000 residents (between 1000 and 8000) and are more homogenous in their social and demographic characteristics compared to other area-based units such as zip codes [25].

Measures

Food and alcohol establishments were categorized based on Standard Industrial Classification (SIC) codes used in the

previous work [7, 14, 16, 26]. The key establishments of interest included on and off premise alcohol outlets (SIC codes beginning; 5813, 5181, 5182, and 5921), grocery stores/supermarkets (SIC codes beginning; 5411), and restaurants (SIC codes beginning; 5812). We calculated census tract density measures by summing the number of establishments within each census tract by category based on per capita (per 10, 000 residents) and per area (square mile) from data generated from the census. Each commercial data source provides several SIC codes to categorize the type of establishment, including primary, secondary, and tertiary SIC codes to delineate the main type of industry. We created the final measures based on whether the establishment was categorized based on all SIC codes (i.e., primary, secondary, and tertiary).

We created measures of neighborhood racial and socioeconomic characteristics based on some previous work examining neighborhood socioeconomic deprivation and access to food and liquor sources [6, 7, 9]. The key US census-based measures were the percentage of households below the federal poverty level, percentage of households without a car, and percentage of black residents within the census tract. Since the majority of residents within the study area were either black or white with a small representation of populations from other racial and ethnic groups, the census tract variable percentage of black residents was used for this analysis. We also created a variable, city or non-city, based on whether a census tract was within the city limits.

Analysis

We calculated the Spearman correlation to examine the relationship between density measures and neighborhood characteristics (e.g., racial composition) and compared across the two datasets to determine if the association differed by data source. We conducted a formal statistical difference test to determine if the correlation differs across the data sources using Fisher's r-to-z transformation. The correlations between the density measures and neighborhood characteristics are presented for each data source followed by a difference test indicating whether there was a statistically significant difference in the correlations between the two data sources.

We then calculated the overall mean density and for each level of the neighborhood characteristic for each data source. Two-way mixed analysis of variance (ANOVA) was performed on density measures between the InfoUSA and D&B data by neighborhood characteristic. P values for the interaction effect of commercial data source (InfoUSA and D&B) by neighborhood characteristics are reported in the tables. An effect size, Cohen's d, stratified by neighborhood characteristics, is also reported for each comparison (Cohen's d value of 0.20 is small; 0.50 is medium; and 0.80 is large). Given the high correlations between the neighborhood racial and socioeconomic characteristics (ranged from 0.68 to 0.83, results not

shown), we did not conduct multivariate analyses with several neighborhood demographic measures but focused on the bivariate relationship between the neighborhood demographic measures and density measures. We used SAS v9.3 (Cary, NC) to conduct all analyses.

Results

Differences in the Correlation between Food/Alcohol Densities and Neighborhood Demographic Characteristics

For the 2003 data, there were significantly higher correlations between the density measures and the neighborhood racial and socioeconomic characteristics among D&B data compared to InfoUSA except (1) percentage without a car and restaurant density, (2) percentage without a car and wholesale food density, and (3) percentage below poverty and gas station density (Table 1). There was a significant difference in the correlations between alcohol density and percentage without a car when comparing InfoUSA and D&B ($r=0.38$ and 0.47, respectively; difference test $p<0.05$). There were significant, positive correlations between grocery store/supermarket density and percentage black, poverty, and percentage without a car ($r= 0.24$, 0.25, and 0.49, respectively), but only among D&B data; and these correlations were significantly different than the correlations for InfoUSA data. This may be due in part to considerable number of key supermarkets/grocery stores that were missing from the InfoUSA dataset (results not shown). There was a significantly smaller correlation between restaurant density and percentage black for InfoUSA versus D&B ($r=0.04$ and 0.20, respectively; difference $p=0.001$) and a significantly larger correlation between restaurant density and percentage without a car for InfoUSA versus D&B ($r= 0.68$ and 0.45, respectively; difference $p<0.001$). There was also a significant difference between InfoUSA and D&B in the size of the correlations between (1) wholesale density and percentage without a car, (2) general/convenience store and all three neighborhood demographic characteristics, (3) gas station density and percentage below poverty, (4) pharmacies and percentage below poverty, and finally (5) pharmacies and percentage without a car.

Similar to 2003, there were several cases in 2009 where the correlations between InfoUSA and D&B were significantly different (Table 1). There were significantly higher correlations between grocery store density and all neighborhood characteristics for D&B compared to InfoUSA, and the correlation between general/convenience store density and percentage without a car was significantly higher among D&B data versus InfoUSA. The correlation between pharmacies and percentage without a car was much lower in the 2009 data

Table 1 Spearman correlations and difference tests between various neighborhood factor and establishment densities (per square mile), Allegheny County, InfoUSA and D&B 2003 and 2009

Density measures	Database	US census measures		
		Percentage black	Percentage below poverty	Percentage without a car
2003				
Alcohol outlets	InfoUSA	0.11	0.29**	0.38*
	D&B	0.18**	0.37**	0.47**
	Difference test	0.146	0.068	0.025
Grocery stores/supermarkets	InfoUSA	0.00	0.01	0.07
	D&B	0.24**	0.35**	0.49**
	Difference test	<0.001	<0.001	<0.001
Restaurants	InfoUSA	0.04	0.21**	0.68**
	D&B	0.20**	0.29**	0.45**
	Difference test	0.001	0.083	<0.001
Wholesale foods	InfoUSA	−0.06	0.02	0.18**
	D&B	−0.03	0.01	0.04
	Difference test	0.080	0.670	0.004
General and convenience stores	InfoUSA	−0.03	0.00	0.01
	D & B	0.12*	0.19**	0.30**
	Difference test	0.002	<0.001	<0.001
Gas stations	InfoUSA	−0.10	0.10	−0.05
	D&B	−0.05	−0.01	0.02
	Difference test	0.300	0.033	0.160
Pharmacies	InfoUSA	−0.02	0.04	0.14*
	D&B	0.09	0.19**	0.30**
	Difference test	0.160	0.003	0.001
2009				
Alcohol outlets	InfoUSA	0.05	0.24**	0.32**
	D&B	0.06	0.26**	0.34**
	Difference test	0.84	0.67	0.65
Grocery stores/supermarkets	InfoUSA	0.008	0.07	0.12*
	D & B	0.17**	0.24**	0.34**
	Difference test	0.001	<0.001	<0.001
Restaurants	InfoUSA	0.02	0.19**	0.30**
	D&B	0.11*	0.23**	0.36**
	Difference test	0.07	0.40	0.17
Wholesale foods	InfoUSA	−0.05	0.02	−0.0002
	D&B	−0.05	−0.01	−0.05
	Difference test	1	0.54	0.31
General and convenience stores	InfoUSA	0.03	0.08	0.13
	D&B	0.08	0.13*	0.24**
	Difference test	0.31	0.30	*0.02*
Gas stations	InfoUSA	−0.07	−0.05	−0.04
	D&B	−0.09	−0.05	−0.02
	Difference test	0.68	1	0.68
Pharmacies	InfoUSA	−0.07	−0.03	0.05
	D&B	−0.12*	−0.05	−0.03
	Difference test	0.31	0.68	0.1

*$p<0.05$; **$p<0.01$

compared to the 2003 data although the correlations were not statistically significant in the 2009 data.

Differences in the Relationship between Establishment Density and Neighborhood Demographic Characteristics

We examined the relationship between neighborhood demographic characteristics and mean densities for key establishments of interest: alcohol outlets, grocery stores, and restaurants; and if the relationship differed across the two datasets: D&B and InfoUSA (Table 2). Out of the 416 census tracts in the study region, the majority of tracts had a low percentage of black residents, low poverty levels, and low percentages of residents without a car, and the majority was within the city proper. The overall mean and median alcohol and food outlet densities were higher for D&B data compared to InfoUSA data for 2003 and 2009. In 2003, the mean densities for each level of the various neighborhood characteristics differed across the data sources. For example, the Cohen's d ranged from −0.28 to −0.49 in comparing mean alcohol outlet density for each level of percentage black when comparing the two data sources. And in all cases, these differences across the data sources were statistically significant. In other words, the relationship between alcohol outlet density and neighborhood characteristics differed by data source. We also found that the mean alcohol, grocery store, and restaurant density were highest in the city compared to outside of the city.

There were similar trends for 2009 where the mean densities tended to be higher for D&B compared to InfoUSA for each level of neighborhood characteristics (Table 3). The association between the three density measures and neighborhood characteristics only differed across the data sources for the following combinations: (1) alcohol outlet density and percentage without a car, (2) alcohol outlet density and city location, and (3) grocery store/supermarket and all neighborhood characteristics. There was no statistically significant difference between the data sources in the relationship between restaurant density and neighborhood characteristics. Generally, there was an overall higher mean restaurant density for the highest poverty census tracts. Additional analyses showed an interaction between neighborhood poverty and neighborhood percentage black where the highest poverty tracts with the lowest percentage of black residents (i.e., 0–25 % black; predominately white) had a mean restaurant density of 80.9 (SD 92.9) and 84.7 (SD 104.6) for InfoUSA and D&B, respectively. Additionally, the highest poverty tracts with the highest percentage of black residents (i.e., 50–75 % black) had a mean restaurant density of 1.8 (SD 4.0) and 1.9 (SD 2.8) for InfoUSA and D&B, respectively (results not shown).

Discussion

This study examined the key differences in two commonly-used commercial data sources and then considered how these differences may influence the relationship between neighborhood racial and socioeconomic characteristics and the density of food and alcohol outlets. Given the increase in the use of commercial data for research purposes, understanding these differences is important [17, 19]. We found that there are systematic biases in these secondary data sources by neighborhood racial and socioeconomic characteristics. Although the associations with the neighborhood characteristics did not differ between the datasets for every kind of food and alcohol establishment, there were key differences in the associations between grocery store/supermarket densities and neighborhood characteristics across time. For example, there was a positive and significant correlation between grocery store/supermarket density and percentage black, poverty, and percentage without a car, but only among D&B data. This association was opposite of what was expected, but only held for the one data source, not for both. The mean densities were also significantly different across the two datasets and across racial and socioeconomic characteristics for all types of food and alcohol establishments in 2003 but only for the association between grocery stores/supermarkets and all racial/socioeconomic characteristics and alcohol outlets and percentage without a car and city/non-city boundaries for 2009. These findings coincide with a recent study that found some differences between two UK-based secondary data sources of food establishments across socioeconomic characteristics but did not examine neighborhood socioeconomic differences for the various types of establishments [23]. However, prior US-based studies found differences between on-the-ground survey data and secondary data for specific types of food establishments and across various types of neighborhoods [16, 17], but the intent of these studies was not to specifically assess the relationship between neighborhood characteristics and food establishments or differences between the two secondary sources outside of the on-the-ground survey data.

The intent of this study was to examine the differences in the relationship between neighborhood characteristics and densities of food and alcohol across data sources; however, we also found similarities across the data sources. We found that a higher percentage of neighborhood poverty and percentage without a car were positively correlated with alcohol outlet density and restaurant density in 2003 and 2009. We also found that grocery store density was positively correlated with percentage without a car across the two data sources for 2009 only, which was opposite of what was expected. Finally, pharmacy density was positively correlated with percentage without a car for the 2003 data. In examining similarities in mean densities and Cohen's d across data sources, there is a slightly different picture that cannot be captured solely with

Table 2 Mean (SD) and Cohen's d for selected establishment densities (per square mile) by various neighborhood characteristics, Allegheny county 2003

	N	Alcohol outlets InfoUSA Mean (SD)	D&B	Cohen's d	Grocery store/supermarkets InfoUSA Mean (SD)	D&B	Cohen's d	Restaurants InfoUSA Mean (SD)	D&B	Cohen's d
Median (Range)		0.83 (0, 54.6)	2.03 (0, 143.2)		0 (0, 12.33)	2.1 (0, 103.7)		3.2 (0, 251.1)	6.6 (0, 349.8)	
Mean (SD)		3.3 (6.5)	6.2 (12.9)	−0.28	0.3 (1.1)	6.1 (11.5)	−0.71	10.2 (23.7)	16 (32.3)	−0.20
Number of census tracts[a]										
Percentage black										
0–<25 %	336	3.1 (6.4)	5.6 (11.0)	−0.28	0.25 (1.0)	5.1 (9.9)	−0.69	9.9 (22.3)	14.7 (28.7)	−0.19
25–<50 %	25	7.0 (9.9)	14.5 (29.9)	−0.34	0.86 (2.3)	11.1 (17.8)	−0.81	20.9 (45.9)	34.0 (70.5)	−0.22
50–<75 %	22	3.1 (5.1)	5.3 (9.8)	−0.28	0.36 (1.1)	8.2 (13.1)	−0.84	7.3 (13.2)	15.8 (20.1)	−0.50
75–100 %	33	2.6 (5.0)	6.2 (9.1)	−0.49	0.11 (0.6)	11.2 (16.4)	−0.96	6.9 (16.3)	15.6 (25.4)	−0.41
p value[b]		0.0282	0.0094	*0.025*	0.0525	0.0021	*0.0015*	0.1030	0.0395	*0.0006*
Percentage below poverty										
0–<25 %	356	2.8 (5.5)	5.1 (9.7)	−0.29	0.21 (0.9)	4.8 (8.1)	−0.8	8.3 (16.3)	12.9 (21.1)	−0.24
25–<50 %	45	6.4 (8.7)	14.1 (24.4)	−0.42	0.84 (2.0)	14.8 (19.6)	−1.0	20.4 (40.4)	34.7 (58.7)	−0.28
50–100 %	15	4.4 (14.2)	8.2 (20.0)	−0.22	0.36 (1.4)	11.6 (26.1)	−0.61	23.9 (64.3)	33.5 (83.9)	−0.13
p value[b]		0.0016	<0.001	*<0.001*	0.0019	<0.001	*<0.0001*	0.0004	<0.001	*<0.0001*
Percentage without a car										
0–<25 %	308	2.0 (4.0)	3.6 (6.4)	−0.3	0.18 (0.9)	3.5 (5.9)	−0.79	6.2 (12.3)	9.9 (15.5)	−0.26
25–<50 %	81	8.1 (10.5)	15.5 (22.7)	−0.42	0.62 (1.8)	14.7 (19.4)	−1.0	25.5 (44.2)	38.4 (60.2)	−0.24
50–100 %	27	3.6 (7.5)	7.6 (14.8)	−0.34	0.37 (1.2)	10.3 (14.3)	−0.98	9.2 (15.5)	17.7 (25.4)	−0.40
p value[b]		<0.0001	<0.0001	*0.0002*	0.008	<0.0001	*<0.0001*	<0.0001	<0.0001	*<0.0001*
Non-City	267	2.2 (4.4)	3.8 (6.7)	−0.28	0.15 (0.55)	3.5 (6.4)	−0.74	6.0 (9.7)	9.6 (13.3)	−0.31
City	140	5.3 (9.0)	10.8 (19.3)	−0.37	0.54 (1.8)	11.2 (16.5)	−0.91	18.4 (37.2)	28.5 (50.2)	−0.23
p value[b]		<0.0001	<0.0001	*<0.0001*	0.0008	<0.0001	*<0.0001*	<0.0001	<0.0001	*<0.0001*

[a] Total of 416 census tracts in the study area

[b] The p value is indicated for the association between census variables and density variables for each data source (e.g., InfoUSA) (i.e., t test) and the p value for the interaction effect of whether the association between census variables and the density variables differs by data source (i.e., Cohen's d column, p value)

Table 3 Mean (SD) and Cohen's d for selected establishment densities (per square mile) by various neighborhood characteristics, Allegheny county 2009

		Alcohol outlets			Grocery store/supermarkets			Restaurants		
		InfoUSA	D&B	Cohen's d	InfoUSA	D&B	Cohen's d	InfoUSA	D&B	Cohen's d
Median (Range)		0.59 (0, 54.0)	1.27 (0, 73.8)		0 (0, 13.6)	1.2 (0, 76.4)		3.3 (0, 240.2)	4.4 (0, 267.4)	
Mean (SD)		3.1 (6.4)	4.3 (9.1)	−0.15	0.4 (1.4)	3.7 (7.6)	−0.60	11.4 (26.2)	12.3 (26.1)	−0.03
Number of census tracts[a]										
Percentage Black										
0–<25 %	336	3.2 (6.7)	4.3 (9.4)	−0.13	0.33 (1.4)	3.2 (6.9)	−0.58	11.1 (25.1)	12.2 (25.1)	−0.04
25–<50 %	25	4.9 (7.5)	6.1 (8.4)	−0.15	0.90 (1.9)	7.8 (13.0)	−0.74	22.3 (48.3)	20.4 (47.3)	−0.04
50–<75 %	22	2.2 (4.4)	3.8 (10.1)	−0.21	0.68 (1.5)	4.1 (8.6)	−0.55	8.9 (19.1)	10.0 (18.3)	−0.06
75–100 %	33	1.8 (4.1)	2.5 (3.9)	−0.17	0 (0)	5.5 (7.3)	−1.1	7.2 (14.5)	8.8 (16.7)	−0.10
p value[b]		0.28	0.5	0.94	0.06	0.01	0.007	0.15	0.36	0.13
Percentage below poverty										
0–<25 %	356	2.8 (6.0)	3.8 (8.4)	−0.14	0.31 (1.3)	2.9 (5.4)	−0.66	9.2 (19.5)	10.1 (18.3)	−0.05
25–<50 %	45	5.3 (8.1)	7.2 (11.2)	−0.19	0.63 (1.5)	8.6 (12.2)	−0.92	22.6 (43.4)	24.1 (43.9)	−0.03
50–100 %	15	4.8 (9.6)	6.6 (14.0)	−0.15	0.73 (2.8)	7.4 (19.5)	−0.50	28.1 (63.0)	29.4 (69.1)	−0.02
p value[b]		0.03	0.03	0.44	0.19	<0.001	<0.001	0.0002	<0.001	0.75
Percentage without a car										
0–<25 %	308	2.1 (4.2)	2.8 (5.7)	−0.14	0.22 (1.0)	2.1 (3.8)	−0.68	6.8 (13.7)	8.0 (14.2)	−0.09
25–<50 %	81	6.9 (11.0)	10.0 (15.7)	−0.23	0.91 (2.4)	9.0 (13.5)	−0.83	29.6 (48.5)	29.5 (48.1)	0.002
50–100≡	27	3.2 (5.8)	3.8 (6.3)	−0.10	0.27 (1.0)	5.3 (8.1)	−0.87	8.8 (15.4)	9.7 (14.5)	−0.06
p value[b]		<0.0001	<0.0001	0.0002	0.003	<0.0001	<0.001	<0.0001	<0.0001	0.26
Non-City	267	2.1 (3.8)	2.7 (5.7)	−0.12	0.24 (0.85)	2.1 (3.3)	−0.77	6.2 (9.7)	7.1 (10.9)	−0.09
City	140	5.1 (9.5)	7.3 (13.0)	−0.19	0.59 (2.0)	6.8 (11.6)	−0.75	21.6 (41.3)	22.7 (40.4)	−0.03
p value[b]		<0.0001	<0.0001	0.003	0.01	<0.0001	<0.001	<0.0001	<0.0001	0.84

[a] Total of 416 census tracts in the study area

[b] The p value is indicated for the association between census variables and density variables for each data source (e.g., InfoUSA) and the p value for the interaction effect of whether the association between the census variables and the density variables differs by data source (i.e., Cohen's d column, p value).

correlations. We found that the mean densities of alcohol outlets, grocery stores/supermarkets, and restaurants were highest among neighborhoods with 25–50 % black residents compared to all other neighborhoods (i.e., 0–25 % black, 50 %+). We see a similar pattern for percentage poverty and percentage without a car where the mean density for alcohol outlets and supermarkets/grocery stores is highest for neighborhoods with 25–50 % poverty or without a car. However, we find that the mean density of restaurants increases as the percentage of residents below poverty increases. Prior studies have found that higher neighborhood socioeconomic disadvantage was associated with more fast-food restaurants, less grocery stores, and more alcohol outlets [7, 8, 11, 14]. Many of these studies used commercial data such as InfoUSA and D&B but typically used one data source and did not specifically examine if the associations differed between commercial data sources.

Many studies also use commercial data to examine the food and alcohol environment in relation to health outcomes. For example, numerous studies have applied various methods to capture the food environment to understand its effects on diet and obesity [27–31]. A systematic review indicated that five studies found an association between neighborhood food environment and obesity while two studies did not, and most of these studies relied on commercial data to enumerate and measure the food environment [31]. The Centers for Disease Control and Prevention (CDC) currently provide state-level environmental and policy indicators and recommendations for food establishments based on commercial data from InfoUSA [32] and formerly based on data from D&B [17].

In studies of alcohol outlet density and related outcomes, data suggests that as alcohol becomes more available, the risk of alcohol-related disease increases. A recent systematic review by the Community Preventative Services Task Force concludes that there is strong evidence that privatization of alcohol sales leads to increases in excessive alcohol consumption [33]. Ecologic studies at the neighborhood level suggest that availability of alcohol is related to violence, injury traffic crashes, drunk driving offenses, cirrhosis mortality, assaultive violence, sexually transmitted diseases, and liquor law violations [34]. It is believed that less restrictive alcohol policy leads to increased availability through six mechanisms including alcohol outlet density.

There are several strengths to this study. The novelty of this study is the examination of the relationship between neighborhood characteristics and food/alcohol outlet density and if this relationship differed by data source. We conducted this analysis using commercial data readily used in prior studies. We also include key food and alcohol establishments that are of interest in public health research and present the results based on a variety of establishments and sociodemographic characteristics. In an effort to understand the larger resource environment, this study also includes pharmacies, convenience stores, and types of alcohol establishments that have not been assessed in prior work. In determining the key categorizations of establishments for this study, we used primary, secondary, and tertiary SIC codes where previous studies only used primary SIC codes but were comparing back to on-the-ground survey data of establishments.

There are also some limitations that warrant further exploration. Secondary commercial and business data sources are limited in terms of their ability to capture smaller, independent stores particularly in urban areas. However, in cases where primary data collection or neighborhood surveys are not feasible, these datasets provide a source of information that may be important for understanding resource environments. Additionally, as demonstrated as a strength of the present study, the use of more than one commercial data source is important for improving accuracy and mitigating bias in cases where other methods or approaches are not feasible [17]. Classifications of commercial businesses may vary by data source [19], and our ancillary analyses found some differences in how InfoUSA and D&B categorized key food and alcohol establishments. One study to date examined variations in classifications of food establishments by data source and found there were data source differences in classifications of convenience stores by the census tract percentage of black residents [19]. Although it is not possible to determine which data source is more accurate without first hand field observations [16, 17], utilization of multiple sources may limit error and lead to better informed decisions by researchers and public health workers. Unfortunately, every group may not have access to multiple databases due to cost. The present study also did not include the influences of adjacent census tracts, grouped census tracts, or alternative geographic units. Future studies could consider how spatial autocorrelation and groupings of regions or neighborhoods may influence these associations by data source.

Despite these limitations, this study highlights a need for understanding systematic biases in key commercial data sources used to measure and analyze neighborhood food and alcohol environments. Researchers using commercial data to measure the food/alcohol environment should interpret their results with caution and consider potential differences by other neighborhood characteristics. The differences we found across data sources should be considered in future studies and may have an influence on conclusions drawn about associations with health outcomes in prior research. Our findings suggest combining data sources when possible, particularly when using historical data, when on-the-ground survey methods are not feasible and in larger geographic regions. The results of this study provide insight into the variability in two widely used data sources to capture the food and alcohol environment and systematic differences across neighborhood racial and socioeconomic characteristics.

References

1. Larson NI, Story MT, Nelson MC. Neighborhood environments: Disparities in access to healthy foods in the U.S. Am J Prev Med. 2009;36(1):74–81.

2. Galster GC. The mechanism(s) of neighborhood effects: Theory, evidence, and policy implications. In: van Ham M et al., editors. Neighborhood effects research: new perspectives. London: Springer; 2012.

3. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep. 2001;116(5):404–16.

4. Osypuk TL, Acevedo-Garcia D. Beyond individual neighborhoods: a geography of opportunity perspective for understanding racial/ethnic health disparities. Health Place. 2010;16(6):1113–23.

5. Landrine H, Corral I. Separate and unequal: Residential segregation and black health disparities. Ethn Dis. 2009;19(2):179–84.

6. Sharkey JR, Horel S. Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. J Nutr. 2008;138(3):620–7.

7. Powell LM, Chaloupka FJ, Bao Y. The availability of fast-food and full-service restaurants in the United States: Associations with neighborhood characteristics. Am J Prev Med. 2007;33(4 Suppl): S240–5.

8. Hilmers A, Hilmers DC, Dave J. Neighborhood disparities in access to healthy foods and their effects on environmental justice. Am J Public Health. 2012;102(9):1644–54.

9. LaVeist TA, Wallace Jr JM. Health risk and inequitable distribution of liquor stores in African American neighborhood. Soc Sci Med. 2000;51(4):613–7.

10. Nielsen AL, Hill TD, French MT, Hernandez MN. Racial/Ethnic Composition, Social Disorganization, and Offsite Alcohol Availability in San Diego County, California. Soc Sci Res. 2010;39(1):165–75.

11. Romley JA, Cohen D, Ringel J, Sturm R. Alcohol and environmental justice: the density of liquor stores and bars in urban neighborhoods in the United States. J Stud Alcohol Drugs. 2007;68(1):48–55.

12. Zenk SN, Schulz A, Israel B, et al. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. Am J Public Health. 2005;95(4):660–7.

13. Pearce J, Blakely T, Witten K, et al. Neighborhood deprivation and access to fast-food retailing: a national study. Am J Prev Med. 2007;32(5):375–82.

14. Franco M, Diez-Roux AV, Hickson DA, et al. Neighborhood characteristics and availability of healthy foods in Baltimore. Am J Prev Med. 2008;35(6):561–7.

15. Morland K, Wing S, Diez-Roux AV, et al. Neighborhood characteristics associated with the location of food stores and food service places. Am J Prev Med. 2002;22(1):23–9.

16. 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):1122–31.

17. 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):1324–33.

18. Matthews SA, McCarthy JD, Rafail PS. Using ZIP code business patterns data to measure alcohol outlet density. Addict Behav. 2011;36(7):777–80.

19. Han E, Powell LM, Zenk SN, et al. Classification bias in commercial business lists for retail food stores in the U.S. Int J Behav Nutr Phys Act. 2012;9:46.

20. Ohri-Vachaspati P, Martinez D, Yedidia M, et al. Improving data accuracy of commercial food outlet databases. Am J Health Promot. 2011;26(2):116–22.

21. Seliske L, Pickett W, Bates R, et al. Field validation of food service listings: a comparison of commercial and online geographic information system databases. Int J Environ Res Public Health. 2012;9(8):2601–7.

22. Gustafson AA, Lewis S, Wilson C, et al. Validation of food store environment secondary data source and the role of neighborhood deprivation in Appalachia, Kentucky. BMC Public Health. 2012;12:688.

23. Burgoine T, Harrison F. Comparing the accuracy of two secondary food environment data sources in the UK across socio-economic and urban/rural divides. Int J Health Geogr. 2013;12:2.

24. Mendez DD, Duell J, Reiser S, et al. A methodology for combining multiple commercial data sources to improve measurement of the food and alcohol environment: Applications of geographical information systems. Geospat Health. 2014;9(1):71–96.

25. CensusBureau, Glossary of basic geographic and related terms—Census 2000 (http://www.census.gov/geo/www/tiger/glossary.html#glossary), 2001, US Department of Commerce. Washington, DC.

26. Treno AJ, Johnson FW, Remer LG, et al. The impact of outlet densities on alcohol-related crashes: a spatial panel approach. Accid Anal Prev. 2007;39(5):894–901.

27. Feng J, Glass TA, Curriero FC, et al. The built environment and obesity: a systematic review of the epidemiologic evidence. Health Place. 2010;16(2):175–90.

28. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, et al. Fast food restaurants and food stores: longitudinal associations with diet in young to middle-aged adults: the CARDIA study. Arch Intern Med. 2011;171(13):1162–70.

29. Fuller D, Cummins S, Matthews SA. Does transportation mode modify associations between distance to food store, fruit and vegetable consumption, and BMI in low-income neighborhoods? Am J Clin Nutr. 2013;97(1):167–72.

30. 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):723–30.

31. Holsten JE. Obesity and the community food environment: a systematic review. Public Health Nutr. 2009;12(3):397–405.

32. C.D.C., State Indicator Report on Fruits and Vegetables2013, Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services (http://www.cdc.gov/nutrition/downloads/State-Indicator-Report-Fruits-Vegetables-2013.pdf; Accessed July 11, 2013).

33. Hahn RA, Middleton JC, Elder R, et al. Effects of alcohol retail privatization on excessive alcohol consumption and related harms: a community guide systematic review. Am J Prev Med. 2012;42(4):418–27.

34. Theall KP, Scribner R, Cohen D, et al. The neighborhood alcohol environment and alcohol-related morbidity. Alcohol Alcohol. 2009;44(5):491–9.