Identifying food deserts in the rural South: a comparison of food access measures

Angela D. Liese
Bethany A. Bell
Sarah E. Battersby
James D. Hibbert

and

Xiaoguang Ma

University of South Carolina

This report contains two scientific publications which we plan to submit to their respective target journals in January 2013. All of the aims of the grant proposal have been addressed in these manuscripts.
Who lives in a food desert? A comparison of policy-relevant measures of community food access

Angela D. Liese\textsuperscript{a}, James D. Hibbert\textsuperscript{a}, Xiaoguang Ma\textsuperscript{a}, Bethany A. Bell\textsuperscript{b}, Sarah E. Battersby\textsuperscript{c}

\textsuperscript{a} Department of Epidemiology and Biostatistics and Center for Research in Nutrition and Health Disparities, Arnold School of Public Health, University of South Carolina, 921 Assembly Street, Columbia, SC 29208, USA, AL: liese@sc.edu; JH: hibbert@mailbox.sc.edu; XM: ma26@email.sc.edu

\textsuperscript{b} College of Education, University of South Carolina, 820 South Main Street, Columbia, SC 29208, USA, BB: bellb@mailbox.sc.edu

\textsuperscript{c} Department of Geography, University of South Carolina, 709 Bull Street, Columbia, SC 29208, USA, SB: battersby@sc.edu

Address for corresponding author:

Angela D. Liese, PhD, MPH
Center for Research in Nutrition and Health Disparities, Arnold School of Public Health, University of South Carolina, 921 Assembly Street, Columbia, SC 29208
Phone: (803) 777-9414; Fax: (803) 777-2504; Email: liese@sc.edu
ABSTRACT

Background: Improving spatial access to healthy food retailers has emerged as a novel focus of public policy in the United States. Various agencies have developed specific measures of community food access but these indicators have not been compared systematically in terms of the areas identified or the populations affected.

Methods: We replicated the USDA Economic Research Service’s (ERS) food deserts criteria, the Centers for Disease Control and Prevention’s (CDC) tracts with healthier food retailers, and The Reinvestment Fund (TRF) definition of low access areas using a 2009 food environment validation study conducted in eight South Carolina counties, 2000 US Census data and geographic information systems.

Results: According to the USDA ERS criteria, only 10% (n=15) of 150 Census tracts in the study area were designated as food deserts, compared to 28.7% (n=43) designated as non-healthier retail tracts by CDC, and 29.3% (n=44) limited supermarket access areas by TRF. Despite these differences, the geographic overlap was quite high (USDA ERS vs. CDC: 71% of tracts; CDC vs. TRF: 77%; USDA ERS and TRF: 65%). The population estimated to be residing in areas with poor access to healthy food choices ranged from 50,085 to more than 201,300 residents.

Conclusions: Marked differences were observed in the underlying methodologies in the estimates of prevalence of areas with low healthy food access, their geographic locations, and the affected populations. The findings suggested a need for clear communication of conceptual differences between the measures of food access.

Keywords: spatial access, food environment, food desert, policy
BACKGROUND

Improving spatial access to healthy food retailers has emerged as a novel approach in public policy in the United States (US), complementing long-standing policies on improving economic food access.\textsuperscript{1-6} Various US agencies have developed measures of community food access for purposes ranging from surveillance to policy implementation.\textsuperscript{1,7-11} The US Department of Agriculture’s Economic Research Service (USDA ERS) has issued a definition for food deserts, which is embedded in the USDA Food Environment Atlas and the Food Desert Locator\textsuperscript{7,8} and linked to the Healthy Food Financing Initiative.\textsuperscript{1} The Centers for Disease Control and Prevention (CDC) have reported on tracts with healthier food retailers in the context of their \textit{State Indicator Report on Fruits and Vegetables},\textsuperscript{8} and The Reinvestment Fund (TRF) has defined limited supermarket access areas.\textsuperscript{10,12,13}

According to the USDA ERS report, 6,500 food desert tracts (i.e. about 10\% of a total of 65,334) in the continental US in 2006 were identified as low income tracts in which residents had low access to healthier food retailers.\textsuperscript{14} In comparison, the CDC reported that in 2007, 72\% of Census tracts were considered healthier food retail tracts, because they contained at least one supermarket or large grocery store, which suggests that 28\% were not considered healthier food retail tracts.\textsuperscript{8} A brief comparison of the published methodologies of these measures of community food access reveals that they rely on very different, secondary data sources listing retail food outlets, and are based on unique sets of criteria and methods. To the best of our knowledge, these measures of community food access have not been compared systematically in terms of the areas identified or the size of the populations that are designated as living in areas with poor food retail access.
One approach to a systematic evaluation of these community food access measures would be to use a single, high-quality data source on food retail outlets and then replicate the measures according to each agency’s criteria and compare the results directly to one another. We had the opportunity to conduct such an evaluation, based on data from a previous eight-county food environment validation study which had verified the exact geospatial location of each food outlet in the area and included ascertainment of outlet attributes such as the outlet type. Thus, the purpose of this project was to replicate the USDA ERS’s food desert measure, CDC’s healthier food retail tracts measure, and the Reinvestment Fund (TRF) limited supermarket access areas measure, using a single, validated data source of the food environment and then to compare the three measures with respect to the areas identified as having poor food access and the populations affected.

METHODS

Study Area and Food Environment Data

We utilized the data of a field census (i.e. on the ground verification) of retail food outlets conducted in 2009. The food environment data included geospatial information and store type attributes on all retail food outlets located in seven rural (Chester, Lancaster, Fairfield, Kershaw, Calhoun, Clarendon and Orangeburg) and one urban county (Richland) in South Carolina (Figure 1). Of the 2,208 food outlets situated within the boundaries of the 150 Census tracts in our study area, the locations of 108 food retail outlets, including supermarkets, supercenters, warehouse clubs, large grocery stores and green grocers, were used for the replication of the three community food access measures.
Additionally, a 10-mile buffer area (Euclidian distance) was created around the study area, using two sources of readily available and existing but non-verified data in order to account for stores that could lie just outside the boundary. The data sources for this buffer area included the South Carolina Department of Environmental Health and Control Licensed Food Services Facilities Database and InfoUSA, corresponding to the time period of the field census. An additional 92 grocery stores were contained within this buffer area, resulting in an overarching total for the study area plus buffer of 200.

Geographic Information Systems (GIS) Analyses

We replicated the three measures of community food access using the food environment data described above using GIS-based methods and ArcGIS 10 (Environmental Systems Research Institute (ESRI) 2011, Redlands, CA) software. For network distance calculations, street centerlines from Streetmap Premium (ESRI, 2011) based on commercial street centerline data from NAVTEQ and Tom Tom were used within the Network Analyst extension of ArcGIS.

USDA ERS Food Deserts. This measure designates areas as food deserts at the Census tract level based on income and distance, identifying tracts that are low income in which residents have low access to a supermarket (including supercenters, warehouse club, large grocery stores of which a total of 194 were located within in our study area). A tract is considered a food desert if 1) the tract meets the US Treasury Department’s New Markets Tax Credit (NMTC) program eligibility criteria and 2) 33% of tract population (or a minimum of 500 people) live outside the threshold distance to the nearest supermarket (1 mile in urban areas or 10 miles in rural areas). For a tract to be eligible for the NMTC it must have 1) a poverty rate of at least 20% or 2) a median family income less than 80% of the statewide median family income.
(for tracts not in metropolitan areas), or a median family income less than 80% of the metropolitan area median family income or less than 80% of the greater state median family income (for tracts in metropolitan areas).

Population and economic data were derived from the 1km x 1km gridded population estimates published by the Socioeconomic Data and Applications Center at Columbia University (SEDAC). Because income is a primary determining factor for the identification of food deserts, only tracts meeting the low income criteria were used in the GIS model. The polygonal 1km x 1km SEDAC population grids were converted to point data using a centroid approach retaining the SEDAC population estimates of all people living within each grid cell. Euclidean distance from each SEDAC grid cell centroid to the nearest food outlet was calculated in miles. Distance results in conjunction with income, urbanicity, and population counts were used in ArcGIS to identify a tract as a food desert. Urbanicity was determined by the intersection of tract centroids with Metropolitan Statistical Areas (MSAs). A tract was considered “urban” if its centroid fell within an MSA, otherwise the tract was considered to be “rural.” SEDAC population data points located in low income tracts that exceeded a threshold distance of 1 mile (urban) or 10 miles (rural) were summed within their corresponding tract boundary to obtain a total population of low access individuals. Tracts containing greater than 500 individuals classified as having low access or greater than 33% of the total population classified as having low access were designated food deserts.

**CDC Non-Healthier Retail Tracts.** We focused on the logical counterpart to the CDC’s healthier retail tract measure, those Census tracts which do not contain healthier food retailers. This measure designates a Census tract as a non-healthier retail tract based on the lack of a supermarket (including supercenters, warehouse club), large grocery store (defined as having 50
or more employees) or a green grocer within the Census tract or a half-mile outside of the tract. A total of 200 eligible food outlets were present in our area. Census tracts in the study area were buffered externally by a half mile. Counts of food outlets were determined using a spatial join between the Census tract buffers and food outlets.

**TRF Limited supermarket access areas.** This measure designates areas as low access based on income and distance at the Census block group level. Low access block groups are then clustered into limited supermarket access areas. First, block groups are classified by income into low-moderate income and high-income classes based on median family income from Census Summary File 3 (US Census Bureau, 2000) data. A block group is considered low-moderate income if its median family income is below the median for the metropolitan statistical area (MSA) in which it is located. For those block groups outside of an MSA, low-moderate income was determined by the median family income of the county in which the block group is located. Then, the mean network distance (shortest street path) from the population-weighted centroids of all non-low-moderate income block groups to the nearest supermarket was calculated to determine the benchmark distance. A block group is considered low access if 1) it is classified as a low-moderate income block group and 2) the network distance of the population-weighted centroid to the nearest supermarket exceeds the benchmark distance. Lastly, low access block groups are clustered into limited supermarket access areas based on an adjacency using first and second order neighbors; low access block groups must have at least two adjacent block groups that are low access, similarly these 1st order neighbors must have two adjacent block groups that are also low access. In order to accommodate this process, a 10-mile buffer was used around the study area to include potential low access block group neighbors at the edge of the study area. Low access areas were aggregated to Census tracts through an areal aggregation process.
whereby if more than 50% of the land area of a tract was composed of LAAs, then the entire tract was considered low access. The store types included in this method were limited to supermarkets, supercenters, warehouse clubs and large grocery stores (a total of 194 within our area).

**Urban versus Non-Urban Areas.** Rural-Urban Commuting Areas (RUCAs) have been developed to characterize Census tracts in the US with respect to their rural and urban status. We converted the ten-tiered RUCA codes developed by the USDA ERS into a dichotomy using only the primary and secondary RUCA codes, thereby differentiating urban core from all other areas non-urban, including sub-urban areas, large rural towns, and small towns and isolated rural areas.

**Statistical Analyses**

Statistical analyses included calculation of the percent of Census tracts designated as meeting a given criteria. Ninety-five percent confidence intervals were calculated by approximating the binomial distribution with a normal distribution. Analyses were conducted using SAS software (Version 9.3, SAS Institute, Cary, NC).

**RESULTS**

Areas identified as having poor access to healthier food retailers according to each of the three measures of community food access are shown in Figure 2. Panels a, b, and c present the entire eight-county study area. Panels d, e, and f focus on the largely urban area around the city of Columbia. Marked differences were observed in which Census tracts were designated as
having poor food access, with the fewest number of Census tracts being identified by the USDA ERS measure, followed by the CDC measure and lastly the TRF measure.

The quantitative comparison between the three measures of community food access (Table 1), reveals that according to the USDA ERS only 15 (i.e. 10%) of the 150 Census tracts in the study area were designated as food deserts, compared to 43 (28.7%) tracts considered non-healthier retail tracts according to the CDC’s measure, and 44 (29.3%) considered limited supermarket access areas by TRF. The population estimated to be residing in areas with poor access to healthier foods ranged from 50,085 to more than 201,300. Large differences in the size of the areas were also observed, ranging from 193 square miles according to USDA ERS to 4,170 square miles according to TRF (see also Figure 1). The prevalence of areas with poor food access was somewhat more similar in urban areas (17.7% vs. 11.4% vs. 2.5% of tracts designated as having poor food access for USDA ERS vs. CDC vs. TRF) than in non-urban areas (1.4% vs. 47.9% vs. 59.2%).

Between 65 and 76% of Census tracts received an identical designation as being an area of poor food access or not between any two community food access measures (Table 2). The consistency was highest between CDC and TRF (77%), followed by USDA ERS and CDC at 71%, and 65% for USDA ERS and TRF. The consistency was markedly higher in urban areas (ranging from 85% to 91%) than in non-urban areas (ranging from 42% to 61%).

Figure 3 visualizes these analyses on a map, indicating which areas were consistently identified as having poor food access by none, one, two, or all three measures of community food access. Of the 150 Census tracts, n=69 (46.0%) were not considered to have poor food access by any of the three measures, n=65 (43.3%) were designated as low access by one, n=11 (7.3%) by two measures, and only n=5 (3.3%) by all three measures. This distribution differed
between urban (n=79 tracts) and non-urban (n=71) tracts in that there were 36 (45.6%), 33 (41.8%), 5 (6.3%), and 5 (6.3%) tracts for urban, and 33 (46.5%), 32 (45.1%), 6 (8.5%), and 0 (0.0%) tracts for rural identified as limited supermarket access areas by no, one, two, and all three measures, respectively.

**DISCUSSION**

Research on access to healthy food, the retail food environment, and food choices increasingly seemed to suggest that lack of access to supermarkets contributes to poor diet quality.\(^20\text{-}24\) Several subsequent reviews of the food environment literature synthesized the research basis for potential policy approaches to modifying the food environment\(^25,26\) and described the presence of areas characterized by poor spatial access to healthier food retailers.\(^11,27\text{-}29\) However, more recently, several larger studies not showing any evidence of associations with dietary intake or obesity have been published.\(^30\text{-}33\) None the less, in response to earlier findings, several US government agencies recently issued policies aimed at improving spatial access to healthier food retail outlets.\(^1,8,9\)

We replicated three distinct measures of community food access that are relevant to US food access policy based on a single, verified dataset of the food environment and a standardized GIS protocol in a contiguous study area. This approach allowed us to focus on the differences in the methods and implications of the three measures that were solely due to their definitions, un-confounded by any inconsistencies in the underlying GIS methodologies or data. This study demonstrates a number of substantial and important differences. The USDA ERS food desert measure\(^11\) identified the fewest number of Census tracts. Despite the focus of its name on food access, this measure actually identified low income tracts which have low access. Of the 150 tracts, 73 were considered low income and of those only 15 had low food access, i.e. met the
criteria for food deserts. This emphasis reflects the USDA ERS’ commitment to socio-economically disadvantaged populations.\textsuperscript{11,34} The food desert measure inherently distinguished access in urban and rural settings.

In contrast, the CDC’s non-healthier retail tract measure was first presented as an environmental indicator in the context of a national report of state-level fruit and vegetable consumption.\textsuperscript{8} No distinction by income levels was made, nor was consideration given to potential access differences due to the generally larger areas comprised by rural Census tracts than urban tracts. Thus, the CDC’s approach was clearly more population-wide than the USDA ERS approach. The TRF measure,\textsuperscript{10,12,13} on the other hand, seems conceptually somewhat similar to the USDA ERS measure in its consideration of income factors related to access, such as household income and car ownership, yet additionally adds the dimension of adjacency/spatial clustering in its definition. In summary, a key implication for future work on these measures of community food access is that the conceptual differences between them need to be communicated very clearly. The term “food desert” has become a catchphrase, frequently used by the media and the public in ways that may not correspond to the intent of the respective developers.\textsuperscript{35-37}

There are a number of GIS-related methodological issues related to the three measures of community food access that are worth considering. Distances were important components of the computational algorithm for both the USDA ERS food desert and the TRF low access area measure. Whereas the USDA ERS used Euclidian (straight-line) distances in its computations, the TRF limited supermarket access areas were based on network distances. Had the USDA ERS algorithm been applied using network distances instead of Euclidian distances, the number of Census tracts considered food deserts would have been markedly higher, i.e. 31 (21%) instead of
15 (10%). Furthermore, the consideration of an exterior buffer was of differing importance to the three measures. Inclusion of a buffer zone did not change any of the USDA food desert designations, but one additional Census tract was designated a non-healthier retail tract according to the CDC measures. However, the buffer area was critically important to the TRF limited supermarket access area measure which builds on a complex adjacency system. Without the buffer zone, 57 Census tracts (38%) would have been considered limited supermarket access areas, instead of the 43 tracts (29%) after inclusion of the buffer. Last but not least, our replication of the three measures used Census tract boundaries corresponding to the year 2000. Had we used the geographic boundaries of Census tracts in 2010, which correspond to a total of 169 tracts in the study area, the number of food deserts, non-healthier retail tracts and limited supermarket access areas would have been higher for the USDA food desert (19% instead of 10%), but virtually unchanged for the other measures (CDC 28%, TRF 25%).

In summary, the replication of the three measure of community food access revealed a number of important GIS-related methodological differences. Furthermore, replication of both the USDA ERS and the TRF measure required more advanced GIS methods, while the CDC’s measure was clearly the easiest to replicate.

A number of limitations and strengths are worth considering. Even though the publicly available documentation on the USDA ERS food desert and the CDC healthier retail tract measure are extremely detailed, we additionally reviewed and vetted our implementation approach with representatives of each of these agencies. Unfortunately, the TRF methodology, while described, was not as detailed as needed and we were unable to have any questions answered by TRF representatives. Thus, while we are certain that our approach to the TRF methodology captured the essence of their measure, we do not claim an identical replication as
there may have been details of which we were unaware. Furthermore, we recognize that our boundary buffer area was not verified, which may have resulted in a small amount of error given known inaccuracies in secondary data sources. However, given that we used multiple data sources to create this buffer, including the data source known to be of the highest quality in South Carolina, the amount of any error will have been very small. Lastly, we do not claim that our study area findings are generalizable to the entire US, however, they are very consistent with data published by two of the agencies. Specifically, the USDA ERS food desert atlas shows that for our study area 10% of Census tracts were considered food deserts in 2006 and the CDC’s data for the study area indicates that in 2007 28% of the study area’s tracts are considered non-healthier retail tracts (L. Moore, personal communication).

Over the past years, the US government has invested several millions of dollars to improve retail food environments, particularly in underserved communities, through initiatives such as Communities Putting Prevention to Work and the Healthy Food Financing Initiative. Public health practitioners and food policy councils are increasingly conducting local assessments of food retail environments in their neighborhoods, communities or states. Tailored to practitioners, the CDC recently released an action guide and toolkit along with the Children’s Food Environment State Indicator Report.

CONCLUSIONS

The results of the present report offer a direct and systematic comparison between three different measures of community food access and highlight both similarities and differences in the conceptual frameworks and methodologies between these measures. For local health practitioners, the USDA ERS food desert measure is likely the easiest to obtain given the
availability of the USDA Food Desert locator. However, it is important to recognize that this measure specifically targets low income areas which have low access to healthier food retailers. The food desert measure is likely to identify the smallest number of Census tracts as our study would suggest. Conceptually, the CDC’s healthier food retailer measure may be the most intuitive, in the sense that it focuses on the availability of a supermarket or large grocery store within the boundaries of a Census tracts (and ½ mile thereof). This measure identified a significantly larger number of Census tracts as having poor food access than the USDA ERS food desert. Lastly, the TRF limited supermarket access area measure is the most complicated to replicate. As the development of spatial food access policies in the US evolves, the various agencies involved in these efforts may want to consider working toward a harmonization of their various definitions. At a minimum, it is important that any communication regarding specific food access measure is very clear about fundamental assumptions in their definitions that may not be obvious from the respective names.

**List of Abbreviations**

U.S. Department of Agriculture, USDA; Economic Research Service, ERS; Center for Disease Control and Prevention, CDC; The Reinvestment Fund, TRF; United States, US; Geographic Information System, GIS.

**Competing Interests**

None

**Authors’ Contributions**
AL developed the idea for this manuscript, acquired and interpreted the data, and drafted the manuscript. JH participated in acquisition of data, geocoded the data and conducted GIS-based data management. XM conducted statistical analyses. BB provided statistical expertise. SB provided geographic expertise. All authors reviewed and edited the manuscript, and approved the final version of the manuscript.

Acknowledgement

Funding for this project was received from a grant by the RIDGE Center for Targeted Studies at the Southern Rural Development Center at Mississippi State University. The food environment data were funded by NIH 1R21CA132133. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the RIDGE Center for Targeted Studies or the National Cancer Institute or the National Institutes of Health.

REFERENCES


Figure 1. South Carolina Study Area
Figure 2. Geographic variation of three measures of community food access, for entire South Carolina area (panels a-c) and urban areas only (panels d-f).
Figure 3. Geographic consistency between measures of community food access
Table 1. Characteristics of areas designated as having poor food access according to three measures of community food access

<table>
<thead>
<tr>
<th></th>
<th>USDA ERS Food Desert Tracts</th>
<th>CDC Non-Healthier Retail Tracts</th>
<th>TRF Limited Supermarket Access Areas (Tract-level)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All (n=150)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Tracts/Block Groups, N</td>
<td>15</td>
<td>43</td>
<td>44</td>
</tr>
<tr>
<td>Percent (95% CI), %</td>
<td>10.0 (5.2, 14.8)</td>
<td>28.7 (21.4, 35.9)</td>
<td>29.3 (22.1, 36.6)</td>
</tr>
<tr>
<td>Affected Population, N</td>
<td>50,085</td>
<td>148,933</td>
<td>201,301</td>
</tr>
<tr>
<td>Affected Area, SQ MI</td>
<td>193</td>
<td>2,974</td>
<td>4,170</td>
</tr>
<tr>
<td><strong>Non-Urban (n=71)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Tracts, N</td>
<td>1</td>
<td>34</td>
<td>42</td>
</tr>
<tr>
<td>Percentage, %</td>
<td>1.4 (0.0, 4.1)</td>
<td>47.9 (36.3, 59.5)</td>
<td>59.2 (47.7, 70.6)</td>
</tr>
<tr>
<td>Affected Population, N</td>
<td>2,217</td>
<td>120,892</td>
<td>195,871</td>
</tr>
<tr>
<td>Affected Area, SQ MI</td>
<td>152</td>
<td>2,936</td>
<td>4,161</td>
</tr>
<tr>
<td><strong>Urban (n=79)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Tracts, N</td>
<td>14</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Percentage, %</td>
<td>17.7 (9.3, 26.1)</td>
<td>11.4 (4.4, 18.4)</td>
<td>2.5 (0.0, 6.0)</td>
</tr>
<tr>
<td>Affected Population, N</td>
<td>47,868</td>
<td>28,041</td>
<td>5,430</td>
</tr>
<tr>
<td>Affected Area, SQ MI</td>
<td>41</td>
<td>38</td>
<td>9</td>
</tr>
</tbody>
</table>
Table 2. Geographic agreement between community food access measures

<table>
<thead>
<tr>
<th>Overall Agreement (%)</th>
<th>USDA ERS vs. CDC</th>
<th>USDA ERS vs. TRF</th>
<th>CDC vs. TRF</th>
<th>USDA ERS vs. CDC vs. TRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>70.7</td>
<td>64.7</td>
<td>76.7</td>
<td>49.3</td>
</tr>
<tr>
<td>Non-Urban</td>
<td>53.5</td>
<td>42.3</td>
<td>60.6</td>
<td>46.5</td>
</tr>
<tr>
<td>Urban</td>
<td>86.1</td>
<td>84.8</td>
<td>91.1</td>
<td>51.9</td>
</tr>
</tbody>
</table>
Variation in low food access areas due to data source inaccuracies

Xiaoguang Ma¹, Sarah E. Battersby², Bethany A. Bell³, James D. Hibbert⁴, Timothy L. Barnes¹, Angela D. Liese¹,⁴*

1. Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC
2. Department of Geography, University of South Carolina, Columbia, SC
3. College of Education, University of South Carolina, Columbia, SC
4. Center for Research in Nutrition and Health Disparities, Arnold School of Public Health, University of South Carolina, Columbia, SC

*Corresponding Author:
Angela D. Liese
Department of Epidemiology and Biostatistics
Center for Research in Nutrition and Health Disparities
Arnold School of Public Health, University of South Carolina
921 Assembly Street, Columbia, SC, 29208
Email: liese@sc.edu
Phone: (803) 777-9414
Fax: (803) 777-2524
Abstract

US agencies have developed measures of community food access based on secondary food outlet data sources, however very little is known regarding the influence of inaccuracies in these data sources on identifying these measures. This study aims to replicate the food access measures using different food outlet data sources and examine the validity of secondary data sources by comparison to the spatially and temporally validated field census. With respect to areas designated by the U.S. Department of Agriculture Economic Research Service (USDA ERS) as food deserts, relying on the secondary data sources identified fewer food deserts than actually existed in the eight county study area. However, there was little over-ascertainment compared to the reference data. Although the secondary data sources identified nearly the same total number of tracts identified as non-healthier retail tracts by Centers for Disease Control and Prevention (CDC) as the reference data, they did not consistently identify the same tracts, due largely to under-ascertainment and some over-ascertainment of areas with poor food access. Secondary data sources showed little under- and over-ascertainment on the Reinvestment Fund (TRF) measures. Inaccuracies in secondary data sources used by agencies influence the resulting designations as areas with poor food access.

Count of word: 198 (abstract); 4,022 (text)
INTRODUCTION

Neighborhood characteristics have been shown to be associated with community food access measures, which in turn can influence healthy dietary behaviors.\textsuperscript{1-3} A number of studies have shown that low access to healthier food outlets, specifically supermarkets, can contribute to a poor dietary intake.\textsuperscript{4-6} Additionally, access to unhealthy food outlets such as convenience stores and fast food restaurants also contributes to poor diet quality.\textsuperscript{7,8}

Improving access to healthy and affordable food is an explicit goal of several federal and state policy initiatives in the United States (US). These include the Healthy Food Financing Initiative (HFFI) which is a partnership of the Department of Agriculture (USDA), Department of the Treasury (Treasury), and Department of Health and Human Services (DHHS).\textsuperscript{9} Furthermore, there are several food environment initiatives by the Centers for Disease Control and Prevention (CDC)\textsuperscript{11,12} and a variety of state efforts, such as the Pennsylvania Fresh Food Financing Initiative (FFFI).\textsuperscript{10} In order to identify areas eligible for these federal support initiatives, several agencies have developed different measures of community food access, including the food desert (FD) by USDA Economic Research Service (ERS),\textsuperscript{13} non-healthier food retail tract (NHT) by CDC,\textsuperscript{14,15} and limited supermarket access area (LSA) by the Reinvestment Fund (TRF).\textsuperscript{16}

However, each of these measures of community food access relies on different sources of secondary data to locate retail food outlets. For instance, a database of stores authorized to receive Supplemental Nutrition Assistance Program (SNAP) benefits and data from Trade Dimensions TDLinx (a Nielsen company) in 2006 were used in the definition of the USDA ERS FD.\textsuperscript{17} CDC used the Dun & Bradstreet (D&B) data (Short Hills, New Jersey) for locations of supermarkets in 2007 to define NHT\textsuperscript{14}, and TRF relied solely on TDLinx data to define LSA.\textsuperscript{16}
Previous research by our group and others have demonstrated that secondary data sources such as Dun & Bradstreet and InfoUSA (Omaha, Nebraska) contain substantial amounts of error including undercounts of outlets, overcounts, geospatial inaccuracies, and incorrect assignments of store types. Errors in these secondary data sources might introduce bias into studies focusing on individual behaviors and also affect policy-level food environment indicators such as FD, NHT and LSA. To date, very little is known regarding the magnitude and the direction of the influences of inaccuracies in secondary data sources on identifying community food access measures.

This study aimed to designate low food access areas as defined by three agency-developed community food access measures (FD, NHT, LSA) using different food outlet data sources (D&B, InfoUSA and data from a ground-truthed field census). We also evaluated the accuracy and validity of secondary data sources on identifying community food access measures by comparison to the spatially and temporally validated field census.
MATERIALS AND METHODS

Study Area

The study area included eight contiguous counties (Calhoun, Chester, Clarendon, Fairfield, Kershaw, Lancaster, Orangeburg, and Richland) in the Midlands of South Carolina (Figure 1). The area covers approximately 5,575 square miles and a population of more than 620,000 which accounts for about 15% of the total population in South Carolina. Geographically, in the 2000 U.S. Census this area included 150 Census tracts, containing 489 Census block groups.

Data Sources

Field Census on Food Outlets (Reference Data)

A field census of retail food outlets obtained via direct observation and geographic positioning systems (GPS) verification of all food outlets was conducted from September 2008 to July 2009. The type of food outlet was assigned using an algorithm described previously. This verified dataset was considered the “gold-standard” or reference for the replication of the three measures of community food access in the study area.

Secondary Data on Food Outlets

Two commercially available secondary data sources, D&B and InfoUSA, were used to create the three measures of community food access (FD, NHRT and LSA) in the study area designated by Census tracts and block groups. Both datasets were obtained in the context of the study described above, i.e. prior to the start of the field census. Both data sources list businesses according to the North American Industry Classification System (NAICS), and
included geo-coordinates of each outlet and other outlet information such as employee member, and sales volume. The NAICS codes were used to assign each listed outlet to an outlet type as described previously. For the purposes of this analysis, only the NAICS codes 445110, 452910, 452990, and 453998 corresponding to supermarket and grocery stores (includes stores retailing a general line of food, supercenters, and warehouse clubs) and code 445230 corresponding to green grocers were relevant.

US Census Data

Population and demographic data were readily available to use from the U.S. Census 2000. Household income was obtained from Census 2000 Summary File 3. The data were downloaded from the Socioeconomic Data and Applications Center (SEDAC) hosted at Columbia University. The details of how these data were used in the definitions of the various community food access measures are described below.

GIS Computation of Community Food Access Measures

We subsequently replicated all three community food access measures (USDA ERS FD, CDC NHRT, TRF LSA) based on each of the three data sources, i.e. the field census, D&B and InfoUSA. This allowed us to focus exclusively on the impact of the inaccuracies in the data sources on the three measures of community food access. Figure 2 illustrates the dataflow diagram and data sources used for identifying the three community food access measures. The algorithms were implemented using Esri’s ArcGIS (version 10.0) software and related extensions. We refer to the areas designated by each of the three measures (i.e. FD, NHT, LSA) as low access areas.
Before the replication of each food access measure, a 10-mile exterior buffer corridor was created around the study area in ArcGIS using two sources of readily available and existing but not ground-truthed data (InfoUSA and the Licensed Food Services Facilities Database from the Department of Health and Environmental Control in South Carolina which were used previously\textsuperscript{18}) in order to account for stores that could lie just outside the boundaries of our study area.

\textit{USDA ERS Food Desert}

According to the USDA ERS, a FD is defined as a low-income Census tract where a substantial number or share of residents has low access to a supermarket or large grocery store. The food desert definition is informed by the USDA ERS report \textit{Access to Affordable and Nutritious Food - Measuring and Understanding Food Deserts and Their Consequences}.\textsuperscript{13} A tract is considered as low-income if it has a poverty rate of 20 percent or higher or a median family income at or below 80 percent of the area's median family income. A tract is considered as low-access if at least 500 people and/or at least 33 percent of the census tract's population reside more than 1 mile (for urban tracts) or 10 miles (for rural tracts) from a supermarket or large grocery store.\textsuperscript{17}

At first, we identified the low income Census tracts based on the low income criteria discussed above. Then, polygonal 1km x 1km SEDAC population grids were used to evaluate distance to supermarkets or grocery stores. To examine the distance, we converted the SEDAC grids to point data using a centroid approach retaining the SEDAC population estimates of all people living within each grid cell.\textsuperscript{25} Distance from each SEDAC grid cell centroid to the nearest food outlet was calculated in miles using Euclidean (straight-line distance) and network (shortest
street distance) approaches. For network distance, street centerlines from Streetmap Premium (ESRI, 2011) based on commercial street centerline data from NAVTEQ and Tom Tom were used. Distances were calculated using the Network Analyst (ESRI, 2011) extension for ArcGIS. Low access was evaluated differently according to USDA guidelines for urban and rural areas. Urbanicity was determined by the intersection of tract centroids with Metropolitan Statistical Areas (MSAs). A tract was considered “urban” if its centroid fell within an MSA, otherwise the tract was considered to be “rural.” SEDAC population data points located in low income tracts that exceeded a threshold distance of 1 mile (urban) or 10 miles (rural) were summed within their corresponding tract boundary to obtain a total population of low access individuals.

**CDC Non-Healthier Tract**

In CDC’s 2009 *State Indicator Report on Fruit and Vegetables*, the percentage of a state’s Census tracts supporting healthy food choices was used as an indicator to quantify access to fruits and vegetables in the neighborhood. This measure defines a Census tract as healthy or non-healthy based on availability of healthier food retailers (e.g. supermarkets, large grocery stores, warehouse clubs and fruit and vegetable markets) located within the tract or within a 0.5 mile buffer surrounding the tract boundaries. In order to make it comparable to other food access measures, we used the logical counterpart to the healthier tract in this study, the NHT. Counts of food outlets were determined using a spatial join between the tract buffers and food outlets. Tract buffers, and subsequently, tracts that contained no healthier retailers were classified as NHT.

**TRF Limited Supermarket Access Area**
TRF defines LSA areas as the areas in which residents must travel significantly farther to the nearest full-service grocery store than residents of areas showing similar population density and car-ownership characteristics as well as median household incomes greater than 120% of the area median.\textsuperscript{16} Block group was used as the unit of analysis. Population weighted centroids were calculated for each block group and the network (shortest street path) distance in miles from each centroid to its nearest food outlet was calculated using the Network Analyst extension in ArcGIS 10 (ESRI, 2011). Streetmap Premium (ESRI, 2011) street centerlines were used with the Network Analyst extension in ArcGIS 10. Median family income from Census Summary File 3 data (US Census Bureau, 2000) was then used to determine a benchmark distance. To determine the benchmark distance, block groups were first classified into low-moderate income and high-income categories. A block group is considered low-moderate income if its median family income is below the median family income for the metropolitan statistical area (MSA) in which it’s located. For those block groups outside of an MSA, low-moderate income was determined by the median family income of the county in which the block group is located. The benchmark distance is the mean network distance of all high income block groups. A block group is considered low access if 1) it is classified as a low-moderate income block group \textit{and} 2) the road distance of the population-weighted centroid to the nearest supermarket exceeds the benchmark distance.

\textbf{Statistical Analysis}

Number and percentage of low food access tracts (for FD and NHT) or block groups (for LSA) were identified using the list of food outlets from the field census, D&B, and InfoUSA databases, respectively. Validity statistics were calculated using 2x2 tables including count of
agreement on low food access areas (+agree), count of agreement on non-low food access areas (-agree), count of disagreement (disagree), percentage of concordance, sensitivity, specificity, positive predicted value (PPV), and negative predicted value (PPV). The concordance of food access measure designations between each secondary data source (D&B and InfoUSA) and field census data was calculated as the fraction of the tracts or block groups with agreement on low and on non-low food access areas in all tracts or block groups. Moreover, the total agreements were drawn into maps in the ArcGIS as graphic overlaps. We calculated sensitivity as low food access areas identified by the secondary data source and the field census data”/ “low food access areas identified by field census data”). Under-ascertainment was estimated by 100 minus sensitivity. Specificity was calculated as “non-low food access areas identified by the secondary data source and the field census data”/ “non-low food access areas identified by field census data”). PPV was calculated as “low food access areas identified by the secondary data source and the field census data”/ “low food access areas identified by the secondary database”). Over-ascertainment was estimated by 100 minus PPV. NPV was calculated as “non-low food access areas identified by the secondary data source and the field census data”/ “non-low food access areas identified by the secondary database”). We calculated 95% confidence intervals (CI) for each of these proportions by approximating the binomial distribution with a normal distribution. Statistical analyses were conducted using STATA (version 11.0, College Station, TX).

RESULTS

The number of food outlets by categories and number and percentage of low food access areas were summarized for each community food access measure by different data sources in Table 1. The field census included less supermarkets and fruit and vegetable markets but more
supercenters than listed in secondary data sources. It also included outlets in the categories warehouse club and large grocery store which not distinguishable in the secondary data sources because of lack of specific NAICS codes. Compared to the field census, the secondary data sources identified fewer tracts with low access, especially for USDA ERS FD (D&B: 11 tracts, InfoUSA: 13 tracts vs. field census: 15 tracts out of the 150 tracts in the study area). Using InfoUSA data, 99 (20.3%) of block groups were defined as TRF LSA, which was less than those using field census data (113, 23.1%) and D&B data (115, 23.5%). For the CDC NHT, using either secondary data source to identify areas with low food access yielded reasonably consistent results in terms of the number and percent of tracts.

Validity statistics including agreement, disagreement, concordance, sensitivity, specificity, PPV and NPV are shown in Table 2 for the ability of the secondary data sources to identify correctly areas designated as low access compared to the field census (gold-standard). USDA ERS FD and TRF LSA showed excellent overall geographic concordance between each of the two secondary data sources and the field census data (from 95.9% to 98.7%). The concordance of the CDC NHT was good but somewhat lower for both D&B and InfoUSA data (89.3% for D&B and 88.7% for InfoUSA). For both the USDA ERS FD and CDC NHT, both secondary data sources were less likely to identify areas with food access areas than the data from the field census (sensitivity was 66.7% for FD using D&B and approximately 81% for NHT using both D&B and InfoUSA). The PPV values were about 81% for NHT for both D&B and InfoUSA, which meant there was a small amount of over-ascertainment compared to field census data. However, for FD, very little over-ascertainment was identified by secondary data sources. When identifying LSA, secondary data sources both have little under- and over-ascertainment of areas with low food access compared to field census data. In general, under-
ascertainment of areas with poor food access seemed to be a somewhat more pronounced problem, especially for the USDA ERS food desert measure, rather than over-ascertainment.

DISCUSSION

According to the results in this study, the consistency and validity of secondary food outlet data sources needs to be improved when they were used to identify low food access areas in communities. With respect to areas designated by the USDA ERS as FD, both secondary data sources identified fewer FDs, but showed an excellent agreement (due to the large number of non-FD) and little over-ascertainment compared to our field census. Although the secondary data sources identified nearly the same total number of tracts identified as NHT by the CDC’s algorithm as the field census, they did not consistently identify the same tracts, due to both under- and over-ascertainment of areas with low food access. Only LSA presented consistent results on validity across different data sources. Under-ascertainment of areas with poor food access seemed to be a more pronounced problem.

The vast majority of epidemiological studies measured food access based on readily available secondary databases such as D&B and InfoUSA. Likewise, government agencies have also utilized such secondary databases to develop community food access measures. However, the secondary data used in the development of these policy-relevant measures of community food access have not been verified or ground-truthed, which may introduce bias into the studies relying on these data. Based on a verified field census, our previous study evaluated the validity of three secondary databases (D&B, InfoUSA, and the Licensed Food Services Facilities Database from the South Carolina Department of Health and Environmental Control) in the same eight-county study region in South Carolina and found that secondary
databases were more likely to undercount food outlets and have geospatial inaccuracies compared to our field census data.\textsuperscript{18} Building on this validation study, we observed an increase on both undercount and overcount when one additionally considered errors in the assignment of food outlet type based on NAICS codes.\textsuperscript{21}

In present study, we found that both D&B and InfoUSA tended to identify fewer low food access areas and showed more under-ascertainment than our field census database, especially for USDA ERS FD. Although there were substantial differences in the concepts behind and interpretations of the various measures of community food access and the areas designated as having low food access, we used the same GIS algorithm and Census data when identifying each community food access measure between secondary databases and field census data. Therefore, the discrepant results should result from the difference on the count and geographic accuracy of supermarkets and large grocery stores (and fruit and vegetable markets for NHT). The under-ascertainment and undercount of low access areas is likely attributable to the undercount of food outlets (D&B 24\% and InfoUSA 29\%) and geographic inaccuracies of secondary databases.\textsuperscript{18}

Even though D&B and InfoUSA located almost the same number of NHTs as the field census data, these NHTs were not identical. Less than 90\% of concordance was obtained between these two secondary datasets and the field census data. Both under-ascertainment (sensitivity was approximately 81\%) and over-ascertainment (PPV was approximately 80\%) were found for the secondary databases. According to the results in this study, using D&B without validation would introduce errors in both ways when identifying NHTs. Of the three community food access measures examined in this study, NHT was the only measure based solely on food outlet data and not Census data. Thus, the accuracy of NHT designations was only
dependent on the accuracy and validity of the food outlet database. All the errors in the database would be transferred into the designation of NHT.

In the present study, TRF’s LSA seemed to be consistent among secondary and field census databases in terms of number of LSAs, and validity statistics of secondary databases compared to field census data. However, identification of LSAs differed from the other two measures of food access in multiple ways. For example, unlike FD and NHT which are measured at the tract level, LSA is calculated at the Census block group level. Moreover, except food outlet data, LSA designation needs information on population density, car ownership, and median income whereas FD only includes median income and NHT only depends on food outlet data. In addition, LSA calculations are based on the benchmark distance between population-weighted centroid to the nearest supermarket and large grocery store, whereas FD based on the distance between the geographic center of the 1-km square grid and the nearest healthy food outlet, and NHT did not even based on the distance. With an algorithm with more parameters beyond food outlet data, the impact of the quality of a food outlet database might be balanced out. This could be the reason for relatively good accuracy for secondary food outlet databases when identifying LSA areas.

According to the results in this study, InfoUSA showed less under-ascertainment on FD and more under-ascertainment on LSA, but showed less over-ascertainment on both FD and LSA compared to D&B. In our previous study, InfoUSA was less likely to overcount supermarket and grocery stores than D&B, which was confirmed in our study by identifying community food access measures. However, the relatively low sensitivity (66.7%) of D&B on FD versus InfoUSA (86.7%) could not be fully explained by inaccurate count of food outlets in the databases. In our recent study considering errors in both count and type of food outlet, we
performed the stratification validity analysis by income and poverty and found that D&B was more likely to undercount supermarkets and grocery stores in low-income and poor areas. According to USDA ERS’s definition, all FDs were low-income areas. Therefore, the discrepancy between D&B and InfoUSA might be because they were used to identify low access measures in low-income areas.

According to agency standard data and publication, 23 and 31 tracts are designated as FD and NHT in the study area, respectively. For USDA ERS’s FD, the tracts identified by the agency standard data are more than those identified by any three data sources in this study, however for CDC’s NHT, the number is much less by agency standard than by data sources in this study. USDA ERS used TDLinx and SNAP data in 2006 to identify the FDs. The different food outlet database might be the reasons for different number of FD. However, we did not have access to the original secondary data utilized by USDA ERS, SNAP and TDLinx. Our results suggest that it may be worthwhile to conduct a formal evaluation of the accuracy of these specific secondary data sources used by the USDA ERS food desert measure. In this study, the findings related to D&B are directly applicable and informative for the CDC’s NHT, because D&B data were used in the original CDC publication. However, we identified many more NHTs even using the same secondary database (D&B). It might be because we use the data in a more recent year. In this study, a 10-mile buffer area was added around the study area; however actually, the number of NHT was not changed after removing the buffer area. We have tried to replicate the algorithm to designate low food access areas based on the publications from the various agencies; however, there are some proprietary aspects of the algorithms, so that even though the agencies publish the general method, the datasets used as inputs (including the date of
each dataset), and possibly some other GIS analysis settings cannot be known and then replicated.

There were several limitations in this study. First, in order to estimate the food access for residents living at the edge of the study area, we added a 10-mile buffer around the study area using the combination of InfoUSA and SC DHEC data. In this way, not all food outlets in the “gold standard” were spatially verified by field census. However, the buffer area was small compared to the whole study area. Moreover, we used the combination of two secondary databases to increase the accuracy. In our previous validation study, we found that combinations of secondary databases increased the validity. Second, the field census was conducted in a 10-month range. There were possibilities that the status of food outlets were updated in that time. Third, some food outlets may be listed in the secondary databases but under an NAICS code that we did not request (e.g., code 446191 – food/health supplement stores). However, the number of such food outlets should be very small. In addition, we replicated the algorithms of identifying each community food access measure based on the descriptions from the documentations published by the agencies. There might be differences on GIS programs with government agencies. Lastly, it is important to notice that the secondary databases in this study were developed mainly for commercial purpose rather than scientific research.

This study is the first study to replicate the three agency standard food access measures using different food outlet data sources, and to examine the accuracy of secondary food outlet databases on identifying agencies’ community food access measures. In addition, the gold standard database, field census data, was the largest validation effort of its kind to date in terms of both geographic area and the number of food outlets. Moreover, we included three community food access measures which are most widely used to identify low food access areas by US
government agencies. D&B and InfoUSA are the most commonly used secondary food outlet databases used in epidemiological studies.

**Conclusion**

Secondary data sources did not indicate good consistency and validity when identifying USDA ERS’s FD and CDC’s NHT; however, they might be acceptable for identifying TRF’s LSA. Our results suggested that it may be worthwhile to conduct a formal evaluation of the accuracy of the secondary data sources before they are used by the agencies to identify areas with low food access.

**List of Abbreviations**

Healthy Food Financing Initiative, HFFI; U.S. Department of Agriculture, USDA; Department of the Treasury, Treasury; Department of Health and Human Services, DHHS; Pennsylvania Fresh Food Financing Initiative, FFFI; Centers for Disease Control and Prevention, CDC; food desert, FD; Economic Research Service, ERS; non-healthier food retail tract, NHT; limited supermarket access area, LSA; Reinvestment Fund, TRF; Supplemental Nutrition Assistance Program, SNAP; Dun & Bradstreet, D&B; geographic positioning systems, GPS; North American Industry Classification System, NAICS; Socioeconomic Data and Applications Center, SEDAC; Metropolitan Statistical Area, MSA; positive predicted value, PPV; negative predicted value, PPV.

**Competing Interests**

None
Authors’ Contributions

XM conducted statistical analyses and drafted the manuscript; SB provided geographic expertise; BB provided statistical expertise; JH participated in acquisition of data, geocoded the data and conducted GIS-based data management; TB participated in collecting the filed census data; AL wrote the funding application, developed the idea for this manuscript, acquired and interpreted the data. All authors reviewed and edited the manuscript, and approved the final version of the manuscript.

Acknowledgement

Funding for this project was received from a grant by the RIDGE Center for Targeted Studies at the Southern Rural Development Center at Mississippi State University. The food environment data were funded by NIH 1R21CA132133. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the RIDGE Center for Targeted Studies or the National Cancer Institute or the National Institutes of Health.

REFERENCES


Figure 1. South Carolina Study Area
Figure 2. Dataflow diagram for identifying three community food access measures
Table 1. Number of food outlets, and low food access measures by different data sources

<table>
<thead>
<tr>
<th>Food Outlet in the Study Area, N</th>
<th>Field Census</th>
<th>D&amp;B</th>
<th>InfoUSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supermarket</td>
<td>81</td>
<td>89</td>
<td>82</td>
</tr>
<tr>
<td>Supercenter</td>
<td>13</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Warehouse Clubs</td>
<td>1</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Large Grocery Store</td>
<td>7</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Fruit and Vegetable Market</td>
<td>6</td>
<td>11</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food Outlet in the Study Area with 10-Mile Buffer, N</th>
<th>Field Census</th>
<th>D&amp;B</th>
<th>InfoUSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supermarket</td>
<td>167</td>
<td>174</td>
<td>167</td>
</tr>
<tr>
<td>Supercenter</td>
<td>19</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Warehouse Clubs</td>
<td>1</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Large Grocery Store</td>
<td>7</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Fruit and Vegetable Market</td>
<td>6</td>
<td>11</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identified Low Food Access Areas, N (%)</th>
<th>Field Census</th>
<th>D&amp;B</th>
<th>InfoUSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD*</td>
<td>15 (10.0)</td>
<td>11 (7.3)</td>
<td>13 (9.3)</td>
</tr>
<tr>
<td>NHT*</td>
<td>43 (28.7)</td>
<td>43 (28.7)</td>
<td>42 (28.0)</td>
</tr>
<tr>
<td>LSA†</td>
<td>113 (23.1)</td>
<td>115 (23.5)</td>
<td>99 (20.3)</td>
</tr>
</tbody>
</table>

--: No NAICS codes for those categories in D&B and InfoUSA.
*: The food access measure was in Census tract level and there are 150 Census tracts in the study area.
†: The food access measure was in Census block group level and there are 489 Census block groups in the study area.
Table 2. Validity statistics of D&B and InfoUSA compared to field census data for each community food access measure

<table>
<thead>
<tr>
<th></th>
<th>+Agree (count)</th>
<th>-Agree (count)</th>
<th>Disagree (count)</th>
<th>Concordance (%)</th>
<th>Sensitivity (95% CI %)</th>
<th>Specificity (95% CI %)</th>
<th>PPV (95% CI %)</th>
<th>NPV (95% CI %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD (Field Census: N=15/150)*</td>
<td>10</td>
<td>134</td>
<td>6</td>
<td>96.0</td>
<td>66.7 (41.7, 84.8)</td>
<td>99.3 (95.5, 99.9)</td>
<td>90.9 (62.3, 98.4)</td>
<td>96.4 (91.9, 98.5)</td>
</tr>
<tr>
<td>D&amp;B</td>
<td>13</td>
<td>135</td>
<td>2</td>
<td>98.7</td>
<td>86.7 (62.1, 96.3)</td>
<td>100</td>
<td>100</td>
<td>98.5 (94.8, 99.6)</td>
</tr>
<tr>
<td>InfoUSA</td>
<td>13</td>
<td>135</td>
<td>2</td>
<td>98.7</td>
<td>86.7 (62.1, 96.3)</td>
<td>100</td>
<td>100</td>
<td>98.5 (94.8, 99.6)</td>
</tr>
<tr>
<td>NHT (Field Census: N=43/150)*</td>
<td>35</td>
<td>99</td>
<td>16</td>
<td>89.3</td>
<td>81.4 (67.4, 90.3)</td>
<td>92.5 (85.9, 96.2)</td>
<td>81.4 (67.4, 90.3)</td>
<td>92.5 (85.9, 96.2)</td>
</tr>
<tr>
<td>D&amp;B</td>
<td>34</td>
<td>99</td>
<td>17</td>
<td>88.7</td>
<td>81.0 (66.7, 90.0)</td>
<td>91.7 (84.9, 95.6)</td>
<td>79.1 (64.8, 88.6)</td>
<td>92.5 (85.9, 96.2)</td>
</tr>
<tr>
<td>InfoUSA</td>
<td>34</td>
<td>99</td>
<td>17</td>
<td>88.7</td>
<td>81.0 (66.7, 90.0)</td>
<td>91.7 (84.9, 95.6)</td>
<td>79.1 (64.8, 88.6)</td>
<td>92.5 (85.9, 96.2)</td>
</tr>
<tr>
<td>LSA (Field Census: N=113/489)†</td>
<td>104</td>
<td>365</td>
<td>20</td>
<td>95.9</td>
<td>92.0 (85.6, 95.8)</td>
<td>97.1 (94.8, 98.4)</td>
<td>90.4 (83.7, 94.6)</td>
<td>97.6 (95.5, 98.7)</td>
</tr>
<tr>
<td>D&amp;B</td>
<td>97</td>
<td>374</td>
<td>18</td>
<td>96.3</td>
<td>85.8 (78.2, 91.1)</td>
<td>99.5 (98.1, 99.9)</td>
<td>98.0 (92.9, 99.4)</td>
<td>95.9 (93.4, 97.5)</td>
</tr>
<tr>
<td>InfoUSA</td>
<td>97</td>
<td>374</td>
<td>18</td>
<td>96.3</td>
<td>85.8 (78.2, 91.1)</td>
<td>99.5 (98.1, 99.9)</td>
<td>98.0 (92.9, 99.4)</td>
<td>95.9 (93.4, 97.5)</td>
</tr>
</tbody>
</table>

*a. “+Agree” means the agreement on low food access areas.

b. “-Agree” means the agreement on non-low food access areas.

c. Concordance = (“+Agree” + “-Agree”)/ (“+Agree” + “-Agree” + “Disagree”)

*: The food access measure was at Census tract level and there are 150 Census tracts in the study area.

†: The food access measure was at Census block group level and there are 489 Census block groups in the study area.