NIH Public Access
Author Manuscript
J Transp Land Use. Author manuscript; available in PMC 2011 August 9.

Published in final edited form as:
J Transp Land Use. 2010 April 1; 3(1): 43–65. doi:10.5198/jtlu.v3i1.105.

Finding food:

Issues and challenges in using Geographic Information Systems to measure food access

Ann Forsyth[a]
Cornell University

Leslie Lytle
University of Washington

David Van Riper
University of Minnesota

Abstract

A significant amount of travel is undertaken to find food. This paper examines challenges in measuring access to food using Geographic Information Systems (GIS), important in studies of both travel and eating behavior. It compares different sources of data available including fieldwork, land use and parcel data, licensing information, commercial listings, taxation data, and online street-level photographs. It proposes methods to classify different kinds of food sales places in a way that says something about their potential for delivering healthy food options. In assessing the relationship between food access and travel behavior, analysts must clearly conceptualize key variables, document measurement processes, and be clear about the strengths and weaknesses of data.

Keywords

Food; Geographic Information Systems; GIS; Networks; Density; Access

1 Introduction

Access to local food is a topic of interest to the residents, travel researchers, and those concerned about the nutritional health of populations. Geographic information systems (GIS) have the potential to provide an important set of tools for better understanding issues of access and availability. The purposes of this paper are to reflect on current research on the food environment, to describe and evaluate available data, and to examine new research questions and opportunities for using GIS to study the food environment. To do this, we: 1) Present potential data sources for assessing the local food environment with GIS; 2) Describe food sales place classification methods; and 3) Discuss some of the challenges of using these methods to assess the food environment. These objectives are met using examples from a study of adolescents conducted in the Minneapolis-Saint Paul, Minnesota metropolitan area. While these examples come from a specific case of research on the food environment, they highlight a number of general issues faced by researchers in many locations. In addition, this paper specifically emphasizes distinguishing between different kinds or qualities of food environments, e.g. locations with healthy food options versus those without—an issue that has been of great interest in the field of nutrition. It concludes

[a] forsyth@cornell.edu .

with suggestions on how to manage gaps in existing GIS data, better document how GIS data are used in studies, and improve knowledge about the geography of food shopping so GIS analyses can be better targeted.

Research on the local food environment is complicated by the fact that the fields of transportation and nutrition use different theoretical approaches to examine people's behavior in relation to the food environment and have focused on somewhat different questions. In nutrition, ecological models emphasizing the built, social, and cultural environments as shaping food options are currently at the forefront of efforts to understand the nutritional health of populations, reflecting some disenchantment with models focusing solely on individual behavior and choices (Davison and Birch 2001; Story *et al.* 2008; Swinburn *et al.* 1999). From this perspective, the local food environment includes both the community nutrition environment (e.g. store and restaurant availability and location) and the consumer nutrition environment (the types, quality, and prices of foods in those places) (Black and Macinko 2008; Glanz *et al.* 2007; Lake and Townshend 2006; Maziak *et al.* 2008; Papas *et al.* 2007). Nutrition research examining access to healthy foods has often focused on food availability within homes, schools, and worksites. Where the larger physical environment has been a focus, the emphasis has been on how characteristics of neighborhoods may influence access and availability of foods in stores and restaurants or how local agriculture may affect a community's access to healthy options.[1] The presence of "food deserts" (areas devoid of food stores such as supermarkets) in low-income areas is another concern; studies in the U.S. have documented the existence of food deserts in low-income areas, but studies from other locations typically do not find the same pattern.[2] Other work examines how proximity to healthy or unhealthy foods affects the consumption patterns or weight status of local populations and often compares this with alternative explanations for food choices such as individual preference, economic situation, and social context.[3] Finally, in research examining active living and physical activity at the community level, the issue of ability to walk to restaurants or supermarkets nearby has been studied.[4]

In transportation, theories are often based on economic utility maximization, where individuals maximize their utility typically in terms of time and money—although individual attitudes, preferences, and social networks play roles (Axhausen 2005; Handy 2005; Timmermans and Zhang 2009). Alternatively, social learning theories and theories of planned behavior emphasize experiences of action and beliefs about outcomes of behavior (Forsyth *et al.* 2009; Gardner and Abraham 2008). In particular, trips or activities related to food may reflect habitual or indulgent behavior rather than specific cost-benefit calculations (Aarts *et al.* 1997; Gardner and Abraham 2008; Matthies *et al.* 2002). Research on access to food from a transportation perspective has often been part of a more general examination of shopping or nonwork travel activity (Bhat 1998; Handy 1996; Handy and Clifton 2001; Timmermans and Zhang 2009), or of trips to retail and service jobs (Grengs 2004). In these theories, the built environment presents opportunities for different levels of access or of trip generation related to food and other shopping; food quality or nutrition has been a minor concern to date (Cairns 1998; Cervero and Radisch 1996).[5]

---

[1]Representative studies include Block *et al.* (2004); Burns and Inglis (2007); Cummins *et al.* (2005); Glanz *et al.* (2007); Kaufman (1999); Moore and Diez Roux (2006); Morland *et al.* (2002a); Powell *et al.* (2007); Reidpath *et al.* (2002); Zenk (2006); Zenk *et al.* (2005)

[2]Representative studies on food deserts include Smoyer-Tomic *et al.* (2006, p. 319); Clarke *et al.* (2002); Cummins and Macintyre (1999); Morland *et al.* (2002a,b); Pearce *et al.* (2007a,b); White (2007); Zenk *et al.* (2005)

[3]Representative studies on proximity to healthy and unhealthy food and weight status or consumption include Burdette and Whitaker (2004); Giskes *et al.* (2007b); Jago *et al.* (2007); Laraia *et al.* (2004); Morland *et al.* (2006, 2002b); Simmons *et al.* (2005); Turrell *et al.* (2004).

[4]Studies on physical activity including food access measures include Forsyth *et al.* (2008); Jago *et al.* (2007); Lee and Moudon (2006).

The transportation literature, in summary, is concerned with the behavioral decision to take a trip to find food using a particular mode of transport, while the nutrition literature is concerned about access to healthy, affordable foods as it affects eating behavior. Both sets of theories about behavior, however, are extremely broad and can benefit from better measures documenting where food is located (environments). In addition, there is a need for more research on the specifics of where people obtain food —so that researchers can know, for example, how big the "local" food environment is. Such behavioral research is, however, typically expensive, time-consuming, and intrusive. Absent this behavioral information, data on the locations of food sources are useful in creating usable approximations of the local food environment. How to use current GIS data sources in the most effective way is a focus of this paper.

Research on the local food environment has typically focused on food sold in restaurants and stores. However, households may obtain food in a number of less traditional ways—for example, by growing it themselves or by purchasing it from farmers. Figure 1 shows how transportation and land use issues intersect with preferences and actual food purchasing and consumption. Areas marked with an asterisk are those measurable using GIS and are, incidentally, areas in which land use and transportation planners and policy makers have some influence, however limited.

As shown in Figure 1, the connections that link the individual to their food purchasing and consumption behaviors include individual, social, and economic factors. Many of these factors are related to availability and accessibility at a neighborhood and community level. GIS provides tools for analyzing availability and accessibility, but its use is more complicated than it may first appear. The next section describes some of the issues to be considered when using GIS data to understand the relationship between land use, transportation, and people's access and availability to healthful foods in a geographical area.

In this we use some examples from the TREC-IDEA study (Lytle 2009a). IDEA is a longitudinal cohort study examining the etiology of unhealthy weight gain in adolescents. The IDEA cohort includes dyads of 349 youths aged 10 to 16 and a parent in a seven-county area in the Minneapolis-Saint Paul metropolitan area. The first round of data collection occurred in 2006–2007. Data were collected at the individual level (including physical biomarkers, behavioral information on eating and activity patterns, and psychosocial factors such as attitudes towards eating, activity, and weight) and at the family and home level (including data on parental attitudes and behaviors and an assessment of the home environment). Data were also collected on the school environment and an assessment of each participant's school and neighborhood environments was made through GIS. Undertaking this research involved analyzing past practice in measuring the food environment and developing a set of measurement strategies. This paper reflects that assessment. Actual measures used in the study are available online (Forsyth 2007).

[5]Available data reflect those more general concerns. For example, in the National Household Transportation Survey, trips to buy food may be coded under several categories including "buy goods," "shopping/errands," and "buy gas." There is no specific food shopping category. NHTS meal eating codes are more specific: "get/eat meal," "coffee/ice-cream/snacks," and "meals," but food consumption could also be coded under "attend meeting," "visit public place," "visit friends/relatives," "attend funeral/wedding," "go out/hang out," "social event," "go to school" and so on (U.S. Department of Transportation 2004, Appendix I). Of the literature specifically addressing transportation and food, much relates to issues apart from personal consumption such as freight movement or, more recently, the tradeoff between agricultural production for biofuels vs. food. The relatively small amount of work on personal travel to food has examined a range of topics including trip generation rates of different kinds of stores, the role of food in vacation travel, access to shopping for low-income people, and whether presence of local stores affects mode choices and shopping patterns (Al-Zahrani and Hasan 2008; Cairns 1998; Clifton 2004; Handy and Clifton 2001; Smoyer-Tomic *et al.* 2006). Some of this work distinguishes between convenience shopping (frequent, local) and comparison shopping (less frequent, people can travel further); other studies point to other social and psychological motivations for shopping in particular locations (see review in Handy and Clifton 2001).

2 Potential data sources for assessing the local food environment

2.1 Data sources

The first question in studying food access is how to locate food sources. In public health, four main sources of data have been used to examine access to food sales places: fieldwork, parcel, licensing, and business data. Two other sources that are emerging or have been used in some contexts are online photography and taxation data. In transportation, the focus has been on shopping trips (identified through surveys) and trip generation of types of stores (using a variety of data including fieldwork) although some studies have used commercial databases of businesses (Bhat and Steed 2002; Cairns 1998; Hensen 1988). Larson *et al.* (2009) provide a comprehensive review of studies in nutrition and related fields with a brief listing of how food sources are defined (e.g. SIC codes and Dun and Bradstreet data, fieldwork). As has been noted before, many studies in the transportation area have not been clear about their data sources (Forsyth *et al.* 2006).

Below we examine the potential approaches in more detail citing example studies. We only include studies that locate individual stores—not ones using aggregate data (e.g. the United States Economic Census).

Fieldwork—Field observation by trained data collectors has been used to examine where food is sold, documenting the existence of a store or a restaurant or the type, quality, and prices of food available in the establishment. Fieldwork may be used to establish the criterion validity of measures believed to be less precise. For example, some field work is used as a way to “ground-truth” or to document that the food store identified through a database such as Dun and Bradstreet actually exists and is the type of food store originally identified in the database. Sometimes fieldwork is used to extend another database (for example, to add a neighborhood to an existing database) or to assess the availability of healthful foods in stores (Clarke *et al*. 2002; Sharkey and Horel 2008). For cost reasons, such fieldwork usually considers only a few key food items (e.g. Block and Kouba 2007; Galvez *et al*. 2008; Horowitz *et al*. 2004; Jetter and Cassady 2006). In transportation, such local field inventories have typically been part of larger shopping studies (Handy and Clifton 2001).

Land use and parcel data—Collected for urban planning and tax assessment purposes, these data are typically available in municipal GIS databases. These data sources contain information about land parcels and buildings, including uses, though the level of detail and categories of classification vary by location. The review of over two dozen nutrition studies by Larson *et al*. (2009) did not list any using land use data to locate food stores. Urban planners have used such data to document food store and restaurant location or to access other aspects of the environment such as opportunities for physical activity (e.g. Lee and Moudon 2006).

Health and agriculture department licensing data—Collected at the municipal, county, or state levels; often, food stores and restaurants are licensed by different departments, and each jurisdiction collects different information for different categories of stores. For example, some databases classify stores by items sold, others by store area, and others by cash register numbers (e.g. Laraia *et al*. 2004; Morland *et al*. 2006, 2002b; Zenk *et al*. 2005).

Commercially available business data—Some of the most accessible forms of data collected for business purposes are telephone directories, electronic and printed business directories, and company web sites (Pearce *et al*. 2007a,b; Smoyer-Tomic *et al*. 2006). In the

United States, two main data sources contain national information and significant numbers of data fields beyond name and address as well as standardized industrial classification codes: one is based on the yellow pages (InfoUSA, Business Analyst)6 and the other on self-reporting for credit purposes (Dun and Bradstreet)7 Trade Dimensions, another national source of data, only includes larger stores (Wang *et al*. 2006). These data sources are expensive—in the range of thousands and tens of thousands of dollars even at reduced academic prices (Moore and Diez Roux 2006). As will be shown later, they also contain many inconsistencies.

Taxation data—Governments collect sales and employment tax data, although food is not always a taxable item. Such data are typically heavily protected in terms of confidentiality and government agencies are rarely prepared to share such data. If they are, the address for taxation purposes may not match the store location (Krizek and Johnson 2006; Wang *et al*. 2006).

Online photographs—With the rise of Internet mapping, new databases have emerged. One genre is represented by sites such as Google Street View (launched in May 2007 as a feature of Google maps, maps.google.com) that shows building facades along streets in areas covered by the database. These databases are too new to have been used in research.

Other sources of data—A chamber of commerce may list members, or a council might inventory its supermarkets (Clarke et al. 2002). These sources are, in a sense, fieldwork by others. They are typically local, limiting larger scale or comparative work.

Overall, these sources vary substantially in several dimensions, as outlined in Table 1, and demonstrate a quality versus cost tradeoff. Given substantial resources, fieldwork could be used to collect high-quality data to answer a specific research question. However, the cost of such collection may be prohibitive and the degree of detail available not necessary to the research question. If such information can be found for a large enough area and with appropriate classifications, licensing data are very useful as they are quite comprehensive—only missing some smaller, unlicensed premises. However, licensing data are rarely available for large geographic areas. Similarly, land use data has different levels of detail in different places and the specific cost-benefit calculation will vary with location. Some places will have better licensing or land use data than others—making data use even trickier for comparative studies. In addition, commercial data are attractive due to the potential for consistency across large areas and national coverage; greater flexibility is the tradeoff for accuracy. Of course, these methods can be combined. For example, (Glanz *et al*. 2007) used GIS to locate stores in four neighborhoods and used fieldwork to assess food quality. But as geographical areas get larger, such combined measurement techniques become cumbersome and introduce new issues of consistency in checking and cleaning data.

---

[6]Business Analyst (InfoUSA/ESRI): ESRI (originally the Environmental Systems Research Institute) is the creator of the ArcGIS suite of GIS software. Business Analyst is integrated with their software and uses InfoUSA as the business database—a listing of 11 million U.S. businesses by “business name, industry description or SIC/North American Industry Classification System, sales, employees, and location” (http://www.esri.com/software/businessanalyst/index.html). It also includes the Directory of Major Malls, a list of 4,000 larger shopping centers. InfoUSA data are compiled from phone books, business directories, public filings, and U.S. Postal Service National Change of Address files, checked by phoning businesses. Using 2006 data, Krizek *et al.* (2007) counted 93,840 businesses in the Minneapolis-Saint Paul metropolitan area. In 2007, one year of data cost approximately $3,500 for the United States.

[7]Dun and Bradstreet is a business information provider. Companies apply for a free D-U-N-S number (Dun and Bradstreet Universal Numbering System). Data include exact address, a primary 4-digit SIC code, a primary 6-digit NAICS code, company names, business descriptions, number of employees, sales volume, and square footage of buildings. The educational rate for a metropolitan area can easily be over $10,000. Using 2005 data, Krizek *et al.* (2007) counted 123,668 businesses in the Minneapolis-Saint Paul metropolitan area.

3 Food Sales Place Classification Methods

A second issue faced by researchers looking at access to healthy food is how to use the data to find specific kinds of food sources that may be of interest because they represent healthy or unhealthy options. How to classify a food sales place is determined, in part, by the data chosen. Unfortunately, without fieldwork there is no simple way to find many specific types of places that sell foods that are of interest to researchers; for example, all stores selling fresh fruits and vegetables or all fast food restaurants.

Studies typically examine one or more specific types of stores selling food, including the following, though few studies look at all these categories:

- All stores that sell foods—these include businesses from pharmacies and discount department stores to gas stations, aiming to be quite comprehensive. (Block and Kouba 2007; Moore and Diez Roux 2006; Sharkey and Horel 2008).
- All grocery stores except convenience stores—these are the stores most likely to have a good selection of food at reasonable prices (Alwitt and Donley 1997), and these have been of interest to those interested in food access for those with low incomes (Clarke *et al*. 2002).
- Large supermarkets and super centers—a combination of inexpensive food and a wide selection (Moore *et al*. 2008; Morton and Blanchard 2007).
- Convenience stores including gas stations, which are locally accessible but may have fewer healthy options (Zenk and Powell 2008).
- Farmers markets that typically emphasize local food.
- Stores where alcohol is sold (as alcoholic beverages provide calories) (Pasch *et al*. 2009). Similarly, with restaurants one may be interested in:
- All restaurants—a measure of the availability of away-from-home meals (Morland *et al*. 2002b).
- Fast food restaurants—a measure of inexpensive and possibly less nutritious food (Zenk and Powell 2008).
- Certain types of fast food restaurants, e.g. hamburger and french fries (Wang *et al*. 2007).
- Non-fast food restaurants, including chains and independently owned restaurants.

While these categories may appear to be simple enough, they are not necessarily clearly defined in the data. Instead, with the exception of expensive fieldwork measures, researchers need to define usable approximations. How this is typically done is described below. Work focusing on health and nutrition has gone into the most detail. In transportation and planning, food stores are likely to be placed in larger categories such as local shopping or retail, and some detail is lost (Krizek and Johnson 2006).

Researcher-defined categories

The gold standard in this field is probably researcher-defined categories of food sources obtained through fieldwork. For example, the most accurate way to identify stores that sell fresh fruits and vegetables in an area is to send research staff into all stores and restaurants —and potentially to all other sales places (such as vending carts). However, this kind of work is extremely time-consuming. With the rise of online street-level photography for major sections of some cities, researchers may well be able to use these sources to identify some store types from the exterior of the building. However, field work is ideal to maximize validity, sensitivity, and specificity.

Names of chains

Some researchers use kinds of stores or national chains as proxies for healthy or unhealthy food. Land use and parcel data ofen are not helpful in identifying these kinds of property because this parcel information typically pertains to the landowner and many stores rent their properties. Food licensing data and commercial lists are better data sources. For example, Zenk *et al*. (2005), working with Michigan Department of Agriculture data, used store names to identify major chains. This approach can be used across data sources—one could compare names from licensing data in one place with names from the phone book in another; however, it may overlook independent restaurants or smaller grocery and convenience stores. Lists of chain names are available from commercial and online sources (Quick Serve Restaurant Magazine 2006; Supermarket News 2010; Wikipedia 2006). A complicating issue is that “fast food” or “limited service” restaurants form a large category including not only the prototypical “burger joints” but also such chains as Starbucks or Chipotle.

Land use categories

Developed for the purpose of planning regulation, these categories vary by municipality. Specific categories may be very broad, such as: agricultural, residential, commercial, mixed use, and institutional. While classifications of up to twenty or more different land uses are common for metropolitan-wide data sets, they do not necessarily indicate food stores; such places may be subsumed in a category such as “retail and other commercial.” In contrast, an individual municipality may have dozens of classifications that include many sub-types of such stores (Lee and Moudon 2006). Overall, the scale of the study matters and studies looking at broader geographical areas will typically need to use other data to identify businesses selling food.

Industrial classification codes

Other researchers have used industrial classification codes to classify types of stores and restaurants (Krizek 2003; Moore and Diez Roux 2006). In the U.S., these codes are used by the federal government for economic reporting and analysis. Standard Industrial Classification (SIC) codes were replaced by the North American Industrial Classification System (NAICS) codes in 1997; NAICS were updated again in 2002 and 2007.8 They can be quite detailed, but many data sources (e.g. Dun and Bradstreet) use self-reported codes which may be inconsistently applied.

Store size

Store size data may be available in commercial, licensing, or land use data sets and can be used to distinguish stores; bigger stores often provide a wider range of food options, including healthy options. Such data are not always complete, however, and different data sources use different categories (e.g. licensing may consider number of cash registers in one jurisdiction and number of employees in another) (Kaufman 1999; Krizek 2003).

---

[8]A full list of codes is available at http://www.census.gov/epcd/naics02/naicod02.htm and a comparison between NAICS and SIC codes is available at http://www.census.gov/epcd/naics02/N2SIC44.htm (U.S. Census Bureau 2002a,b).

Proprietary codes

Finally, some databases, and in particular Dun and Bradstreet, include proprietary codes such as “pizza shop.” These are potentially very useful for examining types of restaurants; however, these codes rely on self report and are highly inconsistent.

Overall, names of chains are easiest to use across data sources but are a partial list even of fast food restaurants. Industrial classification codes are potentially more comprehensive but are frequently misapplied as we describe below.

4 Challenges in Making Sense of Local Food Environment Data

Beyond data sources and classification, several additional challenges confront analysts in assessing and making sense of food environment data including: geographies, completeness and detail, reliability, data reduction, and consideration of how much we can expect from assessment of the physical environment (see also Table 1). In examining these challenges, we draw on the TREC-IDEA study described above. This metropolitan-wide study drew on local and national GIS databases and highlighted many of the issues likely to be faced by other researchers using GIS in food-environment research.

4.1 Geographies

How far do people travel to buy food? This behavioral issue is of great importance in defining the physical food environment. In the past, researchers had little choice but to use existing geographical units such as ZIP codes or census tracts (areas of approximately 5000 people) to identify an area they named “neighborhood.” Recent developments in GIS and data available for individual parcels and businesses now make it possible to construct measures of “neighborhood” or the local food environment that can be individualized to a specific home, worksite, school, or other community address (e.g. straight-line and network buffers around sites, as well as buffers around common travel routes such as the path from home to school). In addition, it is possible to use GPS units or receipts to track actual purchase locations, although this approach is in its infancy and raises issues of burden and accuracy (Brownson *et al.* 2009; Cummins 2007; Grengs *et al.* 2008). However, Giskes *et al.* (2007b) defined and examined neighborhoods using census tracts and identified the types and numbers of grocery stores in the area. They also sampled people living in the area to find out where they usually shop. While most shopped locally, nearly 15 percent of people shopped outside of their census tract area. Tabulations from the 2001 National Household Transportation Survey (NHTS) for the trip purposes of “meals,” “get/eat meal,” and “coffee/ice-cream/snacks” suggest that six percent of such trips were over a mile, if a block is 1/8 of a mile as defined in the 2008 NHTS (U.S. Department of Transportation 2004, 2008). However, many of these trips would have started at a point outside the home (e.g. school, worksite, or another interim destination) meaning the six percent is an underestimate of the number of such purchases occurring far from home. Other questions remain. It may be that the number or density of stores close to a household matters, the distance to the nearest facility, or the stores available on some normally traveled route (i.e. transit from school to home). This is an area where further research is needed to define the local food environment more precisely.

4.2 Detail and Completeness

The difficulty of creating a complete list of stores selling food from existing data stems from two issues: incomplete geographical coverage and incomplete data on types of stores. Geography is the more problematic. For example, store and restaurant licensing data are typically only available for some areas and, if available, are often collected using different definitions in different municipalities. Even licensing data may not cover all stores selling food, however, as in the case of stores selling candy bars.

Detecting subcategories of stores can be difficult. As an example, the NAICS code for food stores is 455. Major subcategories in NAICS 455 include grocery stores (4451); specialty food stores (4452); and beer, wine, and liquor stores (4453). Within these subcategories, grocery stores (4451) are further subdivided into supermarkets and other grocery stores (44511) and convenience stores (44512). Even this very comprehensive list of NAICS "food stores" does not include some types of stores that commonly sell significant amounts of food including warehouse clubs and super centers, department stores, and various kinds of pharmacies and drug stores—these do have NAICS codes, but not in the food store category. Thus care must be taken in using such codes.

Reliability and Validity—Reliability and validity are reflected in several dimensions, including whether the information is described well enough to be replicated across data collectors and over time, whether the location is correct and can be mapped, and whether the business classification is accurate (Lytle 2009b). Many existing data sources are checked only in an ad hoc manner.

Some address data are coded to a specific point on a street or parcel; some need to be geocoded to transform a written address into a specific point. Few data sources of either type have 100 percent of businesses matched to an exact street address. Table 2 shows an example of attempting to match Dun and Bradstreet data to a variety of stores selling foods in one metropolitan area. At a first pass, using automated means, we were unable to match between 12 and 22 percent of various categories of Dun and Bradstreet food sales places. It took one to two months of work by research assistants to get most addresses matched. The most common address issues were incorrect ZIP codes, missing directionals (such as Oak Street rather than Oak Street North), and addresses with extra numbers (e.g. 1000 instead of 100). Research assistants used online search engines or company websites to find the correct ZIP codes, directionals, or addresses, then re-geocoded the corrected addresses. Unfortunately, even this research was unable to locate every address in the Dun and Bradstreet database. This is a common problem. Data from the alternative source, Business Analyst, comes pre-geocoded but according to its own manual as of 2006, only 90 percent are at the address level, the rest are at the ZIP code level (Forsyth 2007).

There were several reasons for these problems:

- In some cases, the ZIP codes on the Dun and Bradstreet (DB) address did not match the ZIP code of the street network for the same address (e.g. 500 1st St 55117 for DB and 500 1st St 55116 for the street network). The street network layer being used was the street layer endorsed by the metropolitan area's regional planning agency, the Metropolitan Council, and produced by The Lawrence Group (TLG). Unfortunately, there is no easy way around this problem in that if the ZIP code were left, ArcGIS would match to all "500 1st St" addresses in the street network.
- In other cases, the address range on the street network did not contain the DB address. Every segment on the street network contained an address range (e.g. 103 to 199 for the odd side and 102 to 198 for the even side). Sometimes, the DB address would be 100 but the address range on the network did not contain 100; this yielded a low-match score that required checking.
- Custom geocoders were needed for addresses such as "1000 County Highway D" or "2000 State Highway 152." When ArcGIS creates an address locator, it drops

the words “County” or “State” and shortens “Highway” to “Hwy” and this must be managed. In addition, some addresses were incorrect (e.g. with typographical errors).

- Finally, some street addresses may have been incorrect in the street database, but in a large project, this is very time-consuming to check.

While these are specific problems on one metropolitan area, such problems are likely to be present in other locations. The large cost of data—for Dun and Bradstreet, in the tens of thousands of dollars per metropolitan area—makes comparative analyses of such issues difficult.

Importantly, and of general interest, few researchers have clearly indicated how they have dealt with such issues as poor address matching—did they do weeks of time-consuming address checking through field work or “ground-truthing” or were they content with 12 to 22 percent of addresses being unusable after they did an initial automated pass-through with standard tolerances? Two researchers using the same commercial data set, for example, may obtain different results due to different address matching strategies.

Another challenge is the discrepancy in classifying the same businesses using different data sources. In an exploratory study, we compared data from Business Analyst and Dun and Bradstreet in 50 randomly selected ZIP codes in the Minneapolis-Saint Paul metropolitan area (Table 3). ZIP codes were selected as roughly similar sized geographies; other smaller and more consistent units are better for analysis; these were also ZIP codes for which we had licensing data although the licensing data did not include NAICS codes. While both had roughly similar numbers of food and beverage stores (code 455) and food services and drinking places (code 722), even these two sources differed in the number of subtypes of stores identified in each code. For example, the Business Analyst data source identified 117 supermarkets and other grocery stores in the specified areas, while Dun and Bradstreet identified only 70, a 40 percent difference. Differences in full-service and limited-service restaurants were even more striking and were also evident in data for the entire metropolitan area.

These inconsistencies are due to lists being initially incomplete and also outdated (e.g. stores that have closed). While it is possible to combine data sources, identify potential duplicates, and phone or field-check each one, this is very time-consuming and undermines some of the point of using such data rather than fieldwork. Again, while these results reflect only one metropolitan area, they are likely to be repeated elsewhere.

There are even challenges in documenting the total numbers of business by data sources. In their historical work, Wang *et al*. (2006) compared California State Board of Equalization (SBOE) data with telephone directory listings for retail food stores (supermarkets, small grocery stores, and chain convenience stores) in four California cities in the period of 1979–1990. The researchers calculated the Spearman correlations for each type of store at the tract level in each of the 84 study tracts and found the correlation to be 0.50. Licensing data are presumably more accurate but are rarely available consistently across an entire metropolitan area. As a comparison, we examined the total numbers of food stores, restaurants, and bars in 50 ZIP codes in Minnesota based on Dun and Bradstreet, Business Analyst, and local licensing data using methods similar to Wang *et al*. (2006). Correlations between the two commercial sources were high (r= 0.96) but were lower between licensing data and the commercial sources (in both cases r=0.70). This probably overstates the level of agreement, as the numbers were likely composed of different stores as has been pointed out by researchers in Florida (Zhao *et al*. 1999).

While the IDEA study used Dun and Bradstreet data, this was an expensive choice. We now consider Business Analyst to be the preferred data source for business locations for several reasons. First, the data cost significantly less than the Dun and Bradstreet business data. Second, the geographic coverage of Business Analyst is better than Dun and Bradstreet; after subscribing to Business Analyst, researchers have access to all business locations in the United States, while Dun and Bradstreet charges different rates for different amounts of data. Third, since Business Analyst data are geocoded (though not all at the address level), researchers may focus on the data (mis)classification (e.g. NAICS codes) instead of worrying about both data (mis)classification and geocoding for Dun and Bradstreet. However, Business Analyst is only available for the current time period, so those wishing to find historical data need to use other sources such as InfoUSA (the ungeocoded data that is used in Business Analyst) or Dun and Bradstreet.

Other Data Issues—Several other issues make such research a challenge including how often data are updated and the costs of data cleaning and checking (see Table 1). In addition, researchers examining the relationship between residences and food sources in the local area also are faced with a large amount of data generated in terms of distances to various kinds of stores, different kinds of stores in buffers of differing sizes, and stores near worksites or schools as well as near residences. There is little, if any, guidance in the literature on how to systematically approach the task of reducing data (that is, dealing with having dozens of variables measured at multiple geographic scales) (Lytle 2009b; Oakes *et al*. 2009). There is a concern that if enough relationships are examined, something will emerge as significant just by chance. For example, research may examine buffers of 200, 400, 800, 1200, and 3000 meters around an individual's home, school, or worksite to see what food stores the individual is exposed to on their way from home to a destination.[9] Using these geographies and measures may result in hundreds of associations at least some of which will be statistically significant merely due to chance (given 90 or 95 percent confidence intervals). Some sophisticated statistical techniques may introduce assumptions that are not adequately examined (Oakes *et al*. 2009).

5 Moving Forward

GIS is a potentially powerful tool in measuring access to food (and many other things); however, data sources have a number of limitations that may not be obvious to analysts. They may provide a broad overview of food access but a misleading picture at a local level. This kind of problem has been noted in other areas as well, for example when using GIS to examine air quality (Ong *et al*. 2006). Typical researchers are not yet explicit enough about the limitations of their data and the work they perform to clean and manipulate it. Table 4 outlines some key challenges for future development of this area, based on the discussions in the previous section.

However, moving forward with GIS analyses requires some further behavioral research. Assessing the food environment around a person's residence, school, or worksite does not necessarily reflect the stores and restaurants from which they actually purchase food. While we know a great deal about travel mode choice and dietary intake, we know comparatively little about the actual locations where people shop, how many stores they typically frequent, how far they will travel to get the foods (and prices) that they want, and the extent of home delivery of food. Those few studies that have looked at shopping in some depth have found

---

[9]There is no current consensus on the best way to measure neighborhoods geographically, but typically both straight-line and network buffers may be used around home, school, or workplace. Major routes, such as home to school, is also a potential for buffering. Buffers are typically done at a range of distances from 200m to 3km. Technologies such as GPS would allow more precise determination of important locations, but at present have unresolved issues related to cost, privacy, and burden.

much nonlocal travel to shops even for low income people and people with stores nearby (Clifton 2004; Cummins *et al*. 2005; Giskes *et al*. 2007a,b; Handy and Clifon 2001; Lytle 2009b; Wrigley *et al*. 2003).

This behavioral information is important for measuring the physical environment—it can provide crucial information on what parts of the environment matter. In addition, food may be so widely available in homes and neighborhoods that the specific local environment is less important and other factors (e.g. price, taste) are more critical. Alternatively, the importance of the availability and accessibility of foods in the local environment may vary according to the resources or other constraints of the population living in that local environment; for those with abundant resources, traveling farther to get healthy foods may be less of an issue than for members of disadvantaged populations (Lytle 2009b).

Further, while the objective local nutrition environment is important, so is the perceived environment—people's mental maps of an area may be quite different from the measured environment. Giskes *et al*. (2007b) highlight the effect of people's perceptions of the availability and price of foods in their local area on their purchasing patterns. Still, planning occurs at a local level and so policy interventions occur (at least partly) in such places—making local analyses important for policy. Better theory is needed as to how the decision to shop in a particular place occurs; transportation and nutrition theories are currently comprehensive but not very specific about this issue.[10] Theory that can better link local, regional, and perceived environments, and can capture the role of environment relative to other factors (e.g. economics, taste, culture), will help move research forward.

Currently GIS data are subject to a number of quality limitations in terms of consistency across jurisdictions, original purpose of data collection, geographical units, completeness of store listings, frequency of updates, errors, and costs. In spite of the current limitations of GIS-based measures, having consistent and reliable GIS-based measures is essential in helping researchers and practitioners move forward in answering important questions about food and the environment, including the relationship between the perceived and researcher-observed environments. Further collaborations between transportation and nutrition researchers could do much to enhance knowledge about access to food.

Acknowledgments

Several GIS technicians developed the GIS steps for measuring the variables and have our thanks, particularly Ed D'Sousa. The study for which this paper was written was funded through a grant from the National Cancer Institutes as part of their Transdisciplinary Research in Energetics and Cancer Initiative, grant 1U54CA116849-01, "Examining the Obesity Epidemic Through Youth, Family & Young Adults," Principal Investigator: Robert Jeffery.

References

Aarts H, Verplanken B, van Knippenberg A. Habit and information use in travel mode choices. Acta Psychologica. 1997; 96(1–2):1–14.

Al-Zahrani A, Hasan T. Trip generation at fast food restaurants in Saudi Arabia. ITE Journal. 2008; 78(2):24.

Alwitt L, Donley T. Retail stores in poor urban neighborhoods. Journal of Consumer Affairs. 1997; 31(1):139–164.

[10]In terms of data, the future researchers could use food purchasing logs, computerized databases, global positioning systems, or food receipts to examine the stores and restaurants people actually visit rather than relying on recall during a survey or focusing only on local stores.

Axhausen, K. Social networks and travel: Some hypotheses. In: Donaghy, K.; Pop-pelreuter, S.; Rudinger, G., editors. Social Dimensions of Sustainable Transport: Transatlantic Perspectives. Vol. Transport and Society, chapter 7. Ashgate; London: 2005. p. 90-108.

Bhat CR. Analysis of travel mode and departure time choice for urban shopping trips. Transportation Research Part B. 1998; 32(6):361–371. doi: 10.1016/S0191-2615(98)00004-6.

Bhat CR, Steed JL. A continuous-time model of departure time choice for urban shopping trips. Transportation Research Part B. 2002; 36(3):207–224. doi: 10.1016/S0191-2615(00)00047-3.

Black J, Macinko J. Neighborhoods and obesity. Nutrition Reviews. 2008; 66(1):2–20. [PubMed: 18254880]

Block D, Kouba J. A comparison of the availability and affordability of a market basket in two communities in the Chicago area. Public Health Nutrition. 2007; 9(7):837–845. doi: 10.1017/PHN2005924. [PubMed: 17010248]

Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: A geographic analysis. American Journal of Preventive Medicine. 2004; 27(3):211–217. doi: 10.1016/j.amepre.2004.06.007. [PubMed: 15450633]

Brownson R, Hoehner C, Day K, Forsyth A. Measuring the built environment for physical activity: State of the science. American Journal of Preventive Medicine. 2009; 36(4, supplement):S99–S123.e12. doi: 10.1016/j.amepre.2009.01.005. [PubMed: 19285216]

Burdette H, Whitaker R. Neighborhood playgrounds, fast food restaurants, and crime: Relationships to overweight in low-income preschool children. Preventive Medicine. 2004; 38(1):57–63. doi: 10.1016/j.ypmed.2003.09.029. [PubMed: 14672642]

Burns C, Inglis A. Measuring food access in Melbourne: Access to healthy and fast foods by car, bus and foot in an urban municipality in Melbourne. Health & Place. 2007; 13(4):877–885. doi: 10.1016/j.healthplace.2007.02.005. [PubMed: 17470408]

Cairns S. Promises and problems: Using GIS to analyse shopping travel. Journal of Transport Geography. 1998; 6(4):273–284. doi: 10.1016/S0966-6923(98)00026-X.

Cervero R, Radisch C. Travel choices in pedestrian versus automobile oriented neighborhoods. Transport Policy. 1996; 3(3):127–141. doi: 10.1016/0967-070X(96)00016-9.

Clarke G, Eyre H, Guy C. Deriving indicators of access to food retail provision in British cities: Studies of Cardiff, Leeds and Bradford. Urban Studies. 2002; 39(11):2041–2060. doi: 10.1080/0042098022000011353.

Clifton KJ. Mobility strategies and food shopping for low-income families: A case study. Journal of Planning Education and Research. 2004; 23(4):402–413. doi: 10.1177/0739456X04264919.

Cummins S. Commentary: Investigating neighbourhood effects on health - avoiding the `local trap'. International Journal of Epidemiology. 2007 doi: 10.1093/ije/dym033.

Cummins S, Macintyre S. The location of food stores in urban areas: A case study in Glasgow. British Food Journal. 1999; 101:545–553. doi: 10.1108/00070709910279027.

Cummins S, Petticrew M, Higgins C, Findlay A, Sparks L. Large scale food retailing as an intervention for diet and health: Quasi-experimental evaluation of a natural experiment. Journal of Epidemiology and Community Health. 2005; 59(12):1035–1040. doi: 10.1136/jech.2004.029843. [PubMed: 16286490]

Davison KK, Birch LL. Childhood overweight: A contextual model and recommendations for future research. Obesity Reviews. 2001; 2(3):159–171. doi: 10.1046/j.1467-789x.2001.00036.x. [PubMed: 12120101]

Forsyth, A. Environment, food, and youth: GIS protocols. 2007. doi: 2007. URL http://www.designforhealth.net/resources/trec.html

Forsyth A, Hearst M, Oakes JM, Schmitz KH. Design and destinations: Factors influencing walking and total physical activity. Urban Studies. 2008; 45(9):1973–1996. doi: 10.1177/0042098008093386.

Forsyth A, Krizek K, Rodriguez D. Non-motorized travel: Research and contemporary planning initiatives. Progress in Planning. 2009; 71:170–184.

Forsyth A, Schmitz KH, Oakes M, Zimmerman J, Koepp J. Standards for environmental measurement using GIS: Toward a protocol for protocols. Journal of Physical Activity & Health. 2006; 3:S241–S257.

Galvez MP, Morland K, Raines C, Kobil J, Siskind J, Godbold J, Brenner B. Race and food store availability in an inner-city neighbourhood. Public Health Nutrition. 2008; 11(06):624–631. doi: 10.1017/S1368980007001097. [PubMed: 17935646]

Gardner B, Abraham C. Psychological correlates of car use: A meta-analysis. Transportation Research Part F. 2008; 11(4):300–311. doi: 10.1016/j.trf.2008.01.004.

Giskes K, Kamphuis CB, van Lenthe FJ, Kremers S, Droomers M, Brug J. A systematic review of associations between environmental factors, energy and fat intakes among adults: Is there evidence for environments that encourage obesogenic dietary intakes? Public Health Nutrition. 2007a; 10(10):1005–1017. doi: 10.1017/S1368980007665525. [PubMed: 17381942]

Giskes K, Lenthe FV, Brug J, Mackenbach J, Turrell G. Socioeconomic inequalities in food purchasing: The contribution of respondent-perceived and actual (objectively measured) price and availability of foods. Preventive Medicine. 2007b; 45(1):41–48. doi: 10.1016/j.ypmed. 2007.04.007. [PubMed: 17532036]

Glanz K, Sallis JF, Saelens BE, Frank LD. Nutrition environment measures survey in stores (NEMS-S): Development and evaluation. American Journal of Preventive Medicine. 2007; 32(4):282–289. doi: 10.1016/j.amepre.2006.12.019. [PubMed: 17383559]

Grengs J. Measuring change in small-scale transit accessibility with Geographic Information Systems: Buffalo and Rochester, New York. Transportation Research Record. 2004; 1887(1):10–17. doi: 10.3141/1887-02.

Grengs J, Wang X, Kostyniuk L. Using GPS data to understand driving behavior. Journal of Urban Technology. 2008; 15(2):33–53. doi: 10.1080/10630730802401942.

Handy SL. Understanding the link between urban form and nonwork travel behavior. Journal of Planning Education and Research. 1996; 15(3):183–198. doi: 10.1177/0739456X9601500303.

Handy, SL. Critical assessment of the literature on the relationships among transportation, land use, and physical activity. Prepared as a resource paper for TRB Special Report 282: Does the Built Environment Influence Physical Activity?. Examining the Evidence. 2005. URL http://onlinepubs.trb.org/onlinepubs/archive/downloads/sr282papers/sr282Handy.pdf

Handy SL, Clifton KJ. Local shopping as a strategy for reducing automobile travel. Transportation. 2001; 28(4):317–346. doi: 10.1023/A:1011850618753.

Hensen R. Trip generation rates for different types of grocery stores. ITE Journal. 1988; 58(10):21–22.

Horowitz C, Colson K, Hebert P, Lancaster K. Barriers to buying healthy foods for people with diabetes: Evidence of environmental disparities. American Journal of Public Health. 2004; 94(9): 1549–1554. doi: 10.2105/AJPH.94.9.1549. [PubMed: 15333313]

Jago R, Baranowski T, Baranowski JC, Cullen KW, Thompson D. Distance to food stores & adolescent male fruit and vegetable consumption: Mediation effects. International Journal of Behavioral Nutrition and Physical Activity. 2007; 4 doi: 10.1186/1479-5868-4-35.

Jetter K, Cassady D. The availability and cost of healthier food alternatives. American Journal of Preventive Medicine. 2006; 30(1):38–44. doi: 10.1016/j.amepre.2005.08.039. [PubMed: 16414422]

Kaufman P. Rural poor have less access to supermarkets, large grocery stores. Rural Development Perspectives. 1999; 13:19–26. URL http://www.ers.usda.gov/publications/rdp/rdp1098/.

Krizek K. Operationalizing neighborhood accessibility for land use-travel behavior research and regional modeling. Journal of Planning Education and Research. 2003; 22(3):270–287. doi: 10.1177/0739456X02250315.

Krizek KJ, El-Geneidy A, Horning J. Parcel level land use and determining motorized and non-motorized accessibility. PowerPoint presentation, Active Communities and Transportation Research Group. 2007

Krizek KJ, Johnson PJ. Proximity to trails and retail: Effects on urban cycling and walking. Journal of the American Planning Association. 2006; 72(1):33–42. doi: 10.1080/01944360608976722.

Lake A, Townshend T. Obesogenic environments: Exploring the built and food environments. Journal of the Royal Society For the Promotion of Health. 2006; 126(6):262–267. doi: 10.1177/1466424006070487. [PubMed: 17152319]

Laraia BA, Siega-Riz AM, Kaufman JS, Jones SJ. Proximity of supermarkets is positively associated with diet quality index for pregnancy. Preventive Medicine. 2004; 39(5):869–875. doi: 10.1016/j.ypmed.2004.03.018. [PubMed: 15475018]

Larson NI, Story MT, Nelson MC. Neighborhood environments: Disparities in access to healthy foods in the U.S. American Journal of Preventive Medicine. 2009; 36(1):74–81.e10. doi: 10.1016/j.amepre.2008.09.025. [PubMed: 18977112]

Lee C, Moudon A. Correlates of walking for transportation or recreation purposes. Journal of Physical Activity & Health. 2006; 3:77–98.

Lytle LA. Examining the etiology of child obesity: The IDEA study. American Journal of Community Psychology. 2009a; 44(3–4):338–349. [PubMed: 19838791]

Lytle LA. Measuring the food environment: State of the science. American Journal of Preventive Medicine. 2009b; 36(4, Supplement 1):S134–S144. doi: 10.1016/j.amepre.2009.01.018. Measurement of the Food and Physical Activity Environments - Enhancing Research Relevant to Policy on Diet, Physical Activity, and Weight. [PubMed: 19285204]

Matthies E, Kuhn S, Klockner CA. Travel mode choice of women: The result of limitation, ecological norm, or weak habit? Environment and Behavior. 2002; 34(2):163–177. doi: 10.1177/0013916502034002001.

Maziak W, Wark K, Stockton MB. Childhood obesity: Are we missing the big picture? Obesity Reviews. 2008; 9:35–42. doi: 10.1111/j.1467-789X.2007.00376.x. [PubMed: 18154601]

Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. American Journal of Public Health. 2006; 96(2):325–331. doi: 10.2105/AJPH.2004.058040. [PubMed: 16380567]

Moore LV, Diez Roux AV, Nettleton JA, Jacobs J, David R. Associations of the local food environment with diet quality-a comparison of assessments based on surveys and geographic information systems: The multi-ethnic study of atherosclerosis. American Journal of Epidemiology. 2008; 167(8):917–924. doi: 10.1093/aje/kwm394. [PubMed: 18304960]

Morland K, Roux AVD, Wing S. Supermarkets, other food stores, and obesity: The Atherosclerosis Risk in Communities study. American Journal of Preventive Medicine. 2006; 30(4):333–339. doi: 10.1016/j.amepre.2005.11.003. [PubMed: 16530621]

Morland K, Wing S, Roux AD. The contextual effect of the local food environment on residents' diets: The Atherosclerosis Risk in Communities study. American Journal of Public Health. 2002a; 92(11):1761–1767. doi: 10.2105/AJPH.92.11.1761. [PubMed: 12406805]

Morland K, Wing S, Roux AD, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. American Journal of Preventive Medicine. 2002b; 22(1):23–29. doi: 10.1016/S0749-3797(01)00403-2. [PubMed: 11777675]

Morton L, Blanchard T. Starved for access: Life in rural America's food deserts. Rural Realities. 2007; 1(4):1–10. URL http://www.ruralsociology.org/StaticContent/Publications/Ruralrealities/pubs/RuralRealities1-4.pdf.

Oakes JM, Måsse LC, Messer LC. Work group III: Methodologic issues in research on the food and physical activity environments: Addressing data complexity. American Journal of Preventive Medicine. 2009; 36(4, Supplement 1):S177–S181. doi: 10.1016/j.amepre.2009.01.015. Work group results from the workshop “Measurement of the Food and Physical Activity Environments - Enhancing Research Relevant to Policy on Diet, Physical Activity, and Weight”. [PubMed: 19285211]

Ong P, Graham M, Houston D. Policy and programmatic importance of spatial alignment of data sources. American Journal of Public Health. 2006; 96(3):499–504. doi: 10.2105/AJPH.2005.071373. [PubMed: 16449576]

Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC. The built environment and obesity. Epidemiological Reviews. 2007; 29(1):129–143. doi: 10.1093/epirev/mxm009.

Pasch KE, Hearst MO, Nelson MC, Forsyth A, Lytle LA. Alcohol outlets and youth alcohol use: Exposure in suburban areas. Health & Place. 2009; 15(2):642–646. doi: 10.1016/j.healthplace.2008.10.002. [PubMed: 19084464]

Pearce J, Blakely T, Witten K, Bartie P. Neighborhood deprivation and access to fast-food retailing: A national study. American Journal of Preventive Medicine. 2007a; 32(5):375–382. doi: 10.1016/j.amepre.2007.01.009. [PubMed: 17478262]

Pearce J, Witten K, Hiscock R, Blakely T. Are socially disadvantaged neigh-bourhoods deprived of health-related community resources? International Journal of Epidemiology. 2007b; 36:348–355. [PubMed: 17182634]

Powell LM, Slater S, Mirtcheva D, Bao Y, Chaloupka FJ. Food store availability and neighborhood characteristics in the United States. Preventive Medicine. 2007; 44(3):189–195. doi: 10.1016/j.ypmed.2006.08.008. [PubMed: 16997358]

Quick Serve Restaurant Magazine. National chain restaurants. Quick Serve Restaurant Magazine Online. 2006; 50 URL http://www.qsrmagazine.com/shells/chains.phtml?id=1338.

Reidpath DD, Burns C, Garrard J, Mahoney M, Townsend M. An ecological study of the relationship between social and environmental determinants of obesity. Health & Place. 2002; 8(2):141–145. doi: 10.1016/S1353-8292(01)00028-4. [PubMed: 11943585]

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. Journal of Nutrition. 2008; 138(3):620–627. http://jn.nutrition. org/cgi/reprint/138/3/620.pdf. [PubMed: 18287376]

Simmons D, McKenzie A, Eaton S, Cox N, Khan MA, Shaw J, Zimmet P. Choice and availability of takeaway and restaurant food is not related to the prevalence of adult obesity in rural communities in Australia. International Journal of Obesity Related Metabolic Disorders. 2005; 29(6):703–710. doi: 10.1038/sj.ijo.0802941.

Smoyer-Tomic K, Spense J, Amrheim C. Food deserts in the prairies? supermarket accessibility and local need in Edmonton, Canada. The Professional Geographer. 2006; 58(3):307–326. doi: 10.1111/j.1467-9272.2006.00570.x.

Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: Policy and environmental approaches. Annual Review of Public Health. 2008; 29:253–272. doi: 10.1146/annurev.publhealth.29.020907.090926.

Supermarket News. SN's Top 75 Retailers for 2010. 2010. http://supermarketnews.com/profiles/top75/2010/

Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: The development and application of a framework for identifying and prioritizing environmental interventions for obesity. Preventive Medicine. 1999; 29(6):563–570. doi: 10.1006/pmed.1999.0585. [PubMed: 10600438]

Timmermans HJ, Zhang J. Modeling household activity travel behavior: Examples of state of the art modeling approaches and research agenda. Transportation Research Part B. 2009; 43(2):187–190. doi: 10.1016/j.trb.2008.06.004. Modeling Household Activity Travel Behavior.

Turrell G, Blakely T, Patterson C, Oldenburg B. A multilevel analysis of socioeconomic (small area) differences in household food purchasing behaviour. Journal of Epidemiology and Community Health. 2004; 58(3):208–215. doi: 10.1136/jech.2003.011031. [PubMed: 14966233]

U.S. Census Bureau. NAICS codes and titles. 2002a. URL http://www.census.gov/epcd/naics02/naicod02.htm

U.S. Census Bureau. NAICS matched to 1987 SIC. 2002b. URL http://www.census.gov/epcd/naics02/N2SIC44.HTM

U.S. Department of Transportation. CD-ROM. Bureau of Transportation Statistics; Washington, D.C.: 2004. 2001 National Household Travel Survey: National data and data analysis tool.

U.S. Department of Transportation. National Household Travel Survey household screener interview. 2008. URL http://nhts.ornl.gov/2008/doc/NHTS_2008_Questionnaire. pdf

Wang M, Gonzalez A, Ritchie L, Winkleby M. The neighborhood food environment: Sources of historical data on retail food stores. International Journal of Behavioral Nutrition and Physical Activity. 2006; 3(1):15.. doi: 10.1186/1479-5868-3-15. URL http://www.ijbnpa.org/content/3/1/15. [PubMed: 16846518]

Wang M, Kim S, Gonzalez AA, MacLeod KE, Winkleby MA. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index.

Journal of Epidemiology and Community Health. 2007; 61(6):491–498. doi: 10.1136/jech.2006.051680. [PubMed: 17496257]

White M. Food access and obesity. Obesity Reviews. 2007; 8:99–107. doi: 10.1111/j.1467-789X.2007.00327.x. [PubMed: 17316311]

Wikipedia. List of fast food restaurants. 2006. URL http://en.wikipedia.org/wiki/List_of_ fast-food_restaurants

Wrigley N, Warm D, Margetts B. Deprivation, diet, and food-retail access: Findings from the Leeds 'food deserts' study. Environment and Planning A. 2003; 35(1):151–188. doi: 10.1068/a35150.

Zenk SN. Fruit and vegetable access differs by community racial composition and socioeconomic position in Detroit, Michigan. Ethnicity and Disease. 2006; 16(1):275. [PubMed: 16599383]

Zenk SN, Powell LM. U.S. secondary schools and food outlets. Health & Place. 2008; 14(2):336–346. doi: 10.1016/j.healthplace.2007.08.003. [PubMed: 17881277]

Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. American Journal of Public Health. 2005; 95(4):660–667. doi: 10.2105/AJPH.2004.042150. [PubMed: 15798127]

Zhao F, Gan A, Lee S-C. Employment data comparison emphasizes importance of multiple data sources. Florida Transportation Modeling. 1999; 12:2–4.

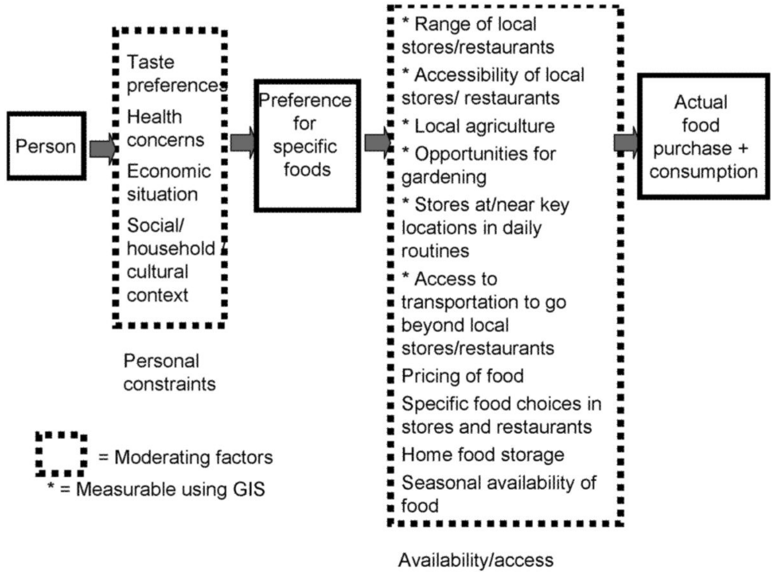

Figure 1.
Food preferences and consumption moderated by factors including land use and transportation options.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Differences in six key data sources for examining access to food

Data Purpose or Issue	Fieldwork	Land Use and Parcel Data	Business Licensing	Commercial Business Listing	Sales or Payroll Taxation Data	Online Street-level Images
Consistency across jurisdictions	Can be built into research design	Varies but generally limited, particularly for more detailed classifications	Varies but generally limited; municipal and county licensing is common	A strength of many of these data is their national range	Depends on the taxing authority but likely to be a strength⊠available at least at the state level	Consistent but coverage is only available in parts of urban areas international coverage
Original purpose of data collection	Data collection protocol developed for specific research needs	Property taxation and urban planning	Health regulation	Includes phone books, mailing, credit reporting	Taxation⊠typically state and national	Online searching; municipal databases and other clients
Geographies of data collection and representation	Can be tailored to suite research design	Land parcel	Business	Business	Typically available by address; office address may not be the store	Street facades
Accuracy of address and/or food type	Depends on researcher	Address accurate; type of store or restaurant may not be listed in sufficient detail	Type of store/restaurant is likely accurate	Typically 80–95% of businesses can be matched to an address; business listing may be obsolete or misclassified (see discussion in main text)	Likely high quality address data for office; may not have detail on food types	Cameras collect 360 degree view but researcher needs to locate address and assess food type
Completeness of listing of food sources	Depends on researcher	All parcels are listed but sources of food may not be obvious	The gold standard for completeness though regulatory categories may not match research categories; may miss informal businesses	Problems with addresses that may not be fully updated; self reporting of business codes	May be difficult to ascertain	Needs to be interpreted by research staff; stores without obvious facades would be missed
Reliability and validity	Depends on researcher	Information is updated but not checked	Information is updated but not checked	Information is updated but not checked	Likely very reliable	Unknown
Frequency of updates to database	Can be tailored	Varies by jurisdiction	Typically updated annually	Typically updated annually or quarterly; some data available in time series	Frequent	Updated but at different times for different places
Cos⊠ data, cleaning, coding	Expensive	Typically public data available for free or at cost	Typically public data; may require significant formatting	Can be very expensive though there are non-profit price breaks for most data it can still be many thousands of dollars for a metropolitan area	Requires permissions for detailed records; such permissions will likely be time	Base data are free; interpretation or coding of data could be quite time consuming

Data Purpose or Issue	Fieldwork	Land Use and Parcel Data	Business Licensing	Commercial Business Listing	Sales or Payroll Taxation Data	Online Street-level Images
					consuming to obtain	

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Address matching at first try and after extensive address cleaning and searching for Dun and Bradstreet data in Minneapolis-Saint Paul, Minnesota

Dataset	Number of Businesses	Unmatched first try	Unmatched after fixes
Eating and Drinking	4565	713 (16%)	97 (2%)
Food Stores	1806	286 (16%)	39 (2%)
Gas Stations	671	139 (21%)	21 (3%)
General Department Stores	394	87 (22%)	22 (6%)
Liquor Stores	443	91 (21%)	2 (0%)
Bakeries	33	4 (12%)	0

Table 3

Comparison of NAICS code distribution for 50 ZIP codes in Minneapolis-Saint Paul, MN

NAICS Code	Business Analyst	Dun and Bradstreet	Difference
455 Food and Beverage Stores	452	394	−13%
4451 Grocery Stores	242	218	−10%
44511 Supermarkets and Other Grocery (except Convenience) Stores	117	70	−40%
4452 Specialty Food Stores	119	82	−31%
44521 Meat Markets	17	18	6%
44522 Fish and Seafood Markets	1	2	100%
44523 Fruit and Vegetable Markets	12	2	−83%
44529 Other Specialty Food Stores	88	59	−33%
445291 Baked Goods Stores	0	21	–
445292 Confectionery and Nut Stores	12	13	8%
445299 All Other Specialty Food Stores	76	26	−66%
722 Food Services and Drinking Places	1004	950	−5%
7221 Full-Service Restaurants	34	432	1171%
7222 Limited-Service Eating Places	829	371	−55%
722212 Cafeterias	5	8	60%
722213 Snack and Nonalcoholic Beverage Bars	75	1	−99%
7224 Drinking Places (Alcoholic Beverages)	91	103	13%

Note: The NAICS codes are a nested hierarchy, so 722 is made up of 7221 + 7222 + 7223 (not reported here because it involves food services such as catering businesses) + 7224.

Table 4

Challenges for future development of GIS measures of the food environment.

Topic	Issue	Needed work
Geographies	Few studies of where people actually shop and those that exist show non-local shopping (Giskes *et al*. 2007a; Handy and Clifton 2001).	Studies of where different kinds of people actually shop, what they pass by, why. This will place the physically local food environment in context – when does it matter and for whom?
Detail and completeness	Data on food stores not available for every location; many kinds of stores sell food (e.g. department stores).	Two strategies are possible – (a) studying areas where data are better (e.g. where licensing is at the state level) and generalizing from those cases or (b) improving data more generally.
Reliability and validity	Numerous classification and address errors in business data in particular. Larger issue of measuring food quality.	At a minimum, researchers should report such issues as percentage of addresses matched and any cleaning and checking procedures used. Ultimately, data should be improved.
Other data issues	Large amount of data; potential for finding relationships due to chance	Develop better theory based on recent research.

Note: These topics are a shorter version of those in Table 1. For references, see text in Section 3.