



Searching for Healthy Food

The Food Landscape in California Cities and Counties

JANUARY 2007

DETAILED METHODOLOGY

INTRODUCTION

This study used commercial data sources and Geographic Information System (GIS) software to examine the distribution of retail food outlets in California communities. This document describes the data sources used, refinements to those data sources, the mapping of the retail outlets, and the construction of an index describing the relative abundance of different types of retail food outlets.

DATA SOURCES

Food Retailers. A dataset was purchased from ESRI (Redlands, CA), a private vendor, in Spring, 2005. It included more than 88,000 food retailers in California. From this dataset, the following four types of food retailers were selected for inclusion in this study.

- **Fast-food restaurants** were defined following the National Restaurant Association's distinction between "table service" and "quick service (fast-food)" restaurants. In addition to counter service, fast-food restaurants are characterized by meal service (vs. snacks, dessert, coffee) and low price (less than \$7/meal). We began with businesses with a North American Industry Classification System (NAICS) code for restaurants (72211002, 72211011, 72211012, 7221013, 72221101, 72221103, 72211016, 72211020, 72221104, 72221105). From these businesses, we selected restaurants with five or more locations with the same name and that provided counter service meals. Major fast-food chains were included (e.g., McDonald's, Taco Bell, Carl's Jr.), as were smaller, regional or locally owned chains. In all, 14,823 fast-food restaurants are included in this study.
- **Convenience stores** were defined as businesses with NAICS code 44512001 that do not sell gasoline or other fuel. This list includes primarily 7-Eleven's and other chains (n=2,752). In order to include smaller chains and family-owned convenience stores, we included businesses with NAICS codes for supermarkets and grocery stores (44511001, 44511002, 44511003, 44511004, 44511005) that had two or fewer employees (n=3,914). In all, there are 6,659 convenience stores in this study.
- **Supermarkets** were identified based on a modification of the Food Marketing Institute definition of a supermarket as a business that sells a variety of food and that earns annual revenues of \$2 million or more each year. To include smaller markets that sometimes play an important role in urban communities, the selection criterion was defined as businesses with annual revenue greater than or equal to \$1 million *or* that were members of a chain (either a national chain such as Safeway, Albertsons, Trader Joe's, or a regional chain such as La Superior, Nugget, Henry's, 99 Ranch) *or* with the word "supermarket" in the business name (n=3,902). NAICS codes included 44511001, 44511002, 44511003, 44511004, 44511005. In all, 3,853 supermarkets are included in this study.

- **Produce vendors** were defined as the sum total of **produce stores** and **farmers markets** for a region. Produce stores included all businesses with NAICS codes 44523001, 44523003. Farmers markets included all certified farmers markets in California. We adjusted the number of farmers markets to include only markets in unique places. For example, the Davis Farmer’s Market is held both Wednesdays and Saturdays; we included only a single location record for this market. In addition, Web sites such as www.cafarmersmarkets.com were used to obtain the locations of additional farmers markets. This information was then geocoded in ArcGIS 9. Actual physical locations were used instead of mailing addresses (which were provided in downloadable files from the web site). In all, 847 produce stores and 445 farmers markets are included in this study.

GEOSPATIAL AND POPULATION DATA

For our analysis of geographic boundaries, we utilized a number of geospatial data from a variety of publishers. Principally, we have used geodata published by the California Spatial Information Library (<http://casil.ucdavis.edu/>); these data include county boundaries and congressional districts.

City designation took one of three forms: geocoding (as provided by California Department of Health Services), Census Designated Place, or proximity to Geographic Names Information System populated place (<http://geonames.usgs.gov/index.html>).

Census boundaries were obtained from the California Office of Statewide Health Planning & Development (<http://www.oshpd.cahwnet.gov/>) and value-added census tabular data were obtained from the California Governor’s Office of Emergency Services (<http://www.oes.ca.gov/>). Calculation of the RFEI was restricted to counties and cities determined by census and population projections to have a population of at least 250,000, as reported by the California Department of Finance for January 1, 2005 (<http://www.dof.ca.gov/HTML/DEMOGRAP/ReportsPapers/Estimates/E1/documents/E-1table.xls>).

All data were transformed to conform to the California Teale projection (Albers Equal Area NAD27), if not already transformed.

DATA ANALYSIS

Data sources were mapped using geographic information software ESRI ArcGIS v9.0 (ESRI; Redlands, CA). Tables were constructed of counts of the number of retail food outlets of each type within census tracts. Data for census tracts were aggregated to construct tables summarizing the number of each type of retail food outlet by geographic region. Geographic regions included the state as a whole, and cities and counties with populations of greater than 250,000.

CONSTRUCTION OF THE RETAIL FOOD ENVIRONMENT INDEX (RFEI)

The number of supermarkets, produce vendors (produce stores and farmers markets), convenience stores and fast-food restaurants was determined for each census district in the state. Census districts were then summed for various geographic regions to determine the total count of each type of retail food outlets in cities and counties. The Retail Food Environment Index (RFEI) was then constructed as follows: the numerator was comprised of the sum of the number of convenience stores and fast-food restaurants, while the denominator was comprised of the sum of the number of supermarkets and produce vendors. The index was determined for the various geographic areas, using geospatial data as described above.

QUALITY CONTROL

In addition to the use of Access queries, Microsoft Excel was used to perform random accuracy checks. The results were recalculated manually for several random counties and other geographic units by sorting and summing the figures for each unit, then comparing the results obtained in the worksheet to those in the final tables.

ADDITIONAL POINTS

- The exclusion of convenience stores associated with gas stations from the analysis leads our study to undercount the total number of convenience stores in geographic areas. This will tend to bias the numerator of the RFEI downward.
- The use of an annual sales cutoff of \$1 million to define supermarkets rather than the industry standard of \$2 million leads our count of supermarkets to include small neighborhood stores as well as larger supermarkets. This will tend to bias the denominator of the RFEI upward.
- The two points above together lead to a downward bias in our RFEI estimate for any given geographic area.



This paper provides a comprehensive overview of the research conducted for the study, *Searching for Healthy Food: The Food Landscape in California Cities and Counties*, available at www.publichealthadvocacy.org/searchingforhealthyfood.html.

Support for this project was provided by a grant from the California Vitamin Cases Consumer Settlement Fund.

For additional information, contact CCPHA at PO Box 2309, Davis, CA 95617; 530 297-6000;

www.publichealthadvocacy.org.

© January 2007 CCPHA