Detailed Methodology

INTRODUCTION

This study analyzed data from the 2005 California Health Interview Survey, the 2005 InfoUSA Business File, and the 2000 US Census. Geographic Information System (GIS) software was used to examine the association of the retail food environment with obesity and diabetes in California, with consideration for the effect of community income on that relationship. This document describes the data sources used, refinements to those data sources, the mapping of the retail food outlets, and the construction of a local Retail Food Environment Index (RFEI), which describes the relative density of different types of retail food outlets around individuals' homes.

DATA SOURCES

Individual health outcomes and demographics. Individual-level data were drawn from the 2005 California Health Interview Survey (CHIS 2005). CHIS is a biennial, random-digit-dial (RDD) telephone survey that collects information from a representative sample of California's non-institutionalized population. Survey topics include individual health behaviors, health outcomes, and socio-demographic characteristics. CHIS 2005 completed interviews with more than 43,000 adults drawn from every county in the state, in English, Spanish, Chinese (both Mandarin and Cantonese), Vietnamese, and Korean. CHIS 2005 includes geocoded home addresses for approximately 90 percent of the adult sample. The CHIS sample represents the geographic diversity of California, and the available multi-language interviews accommodate the state's rich ethnic diversity. CHIS is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. The survey has been conducted every two years since 2001. For more information about CHIS, please visit www.chis.ucla.edu. The following health outcomes were analyzed in this study:

- **Obesity.** Obesity was defined as having a Body Mass Index (kg/m^2) of 30 or greater. Body mass index (BMI) was calculated based on CHIS 2005 respondents' self-reported height and weight values. Obesity, rather than overweight, was studied because it is an established risk factor for chronic illness and mortality. Although BMI calculated from self-reported height and weight is highly correlated with measured BMI, self-reporting may underestimate the actual prevalence of obesity due to social desirability bias.

- **Diabetes.** Estimates of diabetes were based on responses to the question, "Other than during pregnancy, has a doctor ever told you that you have diabetes or sugar diabetes?" Diabetes rates may be underestimated or underreported in underserved populations that do not have adequate access to healthcare and in whom the disease has therefore not been diagnosed.

Food retailers. The InfoUSA Business File was purchased from ESRI (Redlands, CA), a private vendor, in the spring of 2005. It included more than 88,000 food retailers in California. Since other commonly used food retail data sources, such as phonebooks, are known to be flawed, a data-cleaning process was employed to minimize error in the ESRI dataset. For a description of the quality-control process, see "Detailed Methodology" for the brief, *Searching for Healthy Food: The Food Landscape in California Cities and Counties* (http://www.publichealthadvocacy.org/RFEI/expanded%20methods.pdf). From this dataset, the following four types of food retailers were selected for inclusion in this study:

- **Fast-food restaurants.** 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 outlets are characterized by meal service (vs. snacks, dessert, coffee) and lower price (less than $7/meal). We began with businesses with a North American Industry Classification System (NAICS) code for restaurants (72211002, 72211011, 72211012, 72211013, 72211016, 72211020, 72221101, 72221103, 72221104, and 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.

- **Convenience stores.** Convenience stores were defined as businesses with NAICS code 44512001 that do not sell gasoline or other fuel. This list includes primarily 7-Elevens and other chains. In order to include smaller chains and family-owned convenience stores, we included businesses with NAICS codes for supermarkets and grocery stores that had two or fewer employees (44511001, 44511002, 44511003, 44511004, and 44511005).

▪ **Grocery stores.** Supermarkets and grocery stores (referred to collectively as grocery stores in this study) were identified based on a modification of the Food Marketing Institute (FMI) definition of a supermarket. FMI defines supermarkets and grocery stores as businesses that earn annual revenues of $2 million or more each year; however, in this study, we defined supermarkets and grocery stores as those that earn annual revenues of at least $1 million. We made this modification to include smaller markets that sometimes play an important role in communities. 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, and Ranch 99) or stores with the word "supermarket" in the business name were included. NAICS codes included 44511001, 44511002, 44511003, 44511004, and 44511005.

▪ **Produce vendors.** Produce vendors were defined as produce stores and farmers' markets. Produce stores included all businesses with NAICS codes 44523001 and 44523003. Farmers' markets included all certified farmers' markets listed on the website of the California Federation of Certified Farmers' Markets (www.cafarmersmarkets.com). We adjusted the number of farmers' markets to include only markets in unique places. For example, the Davis Farmers' Market is held both Wednesdays and Saturdays; we included only a single location record for this market. Actual physical locations (which were provided in downloadable files from the website) were used instead of mailing addresses.

Community Income. Community income data were taken from the 2000 US Census. Lower-income communities were defined as census tracts in which at least 30 percent of households have incomes below 200 percent of the federal poverty level (FPL).[1] At the time of the 2000 Census, 200 percent FPL was $21,738 for a family of two and $34,058 for a family of four.[2,3] Using these definitions, approximately 46 percent of our sample lived in lower-income communities.[4]

DATA ANALYSIS
This study used Geographic Information System (GIS) software to examine the distribution of retail food outlets in California communities relative to the geocoded home addresses of CHIS respondents. Census tracts were used as proxies for neighborhoods in order to identify lower- and higher-income communities.

Construction of the Retail Food Environment Index (RFEI).
The local food environment of individual CHIS respondents was characterized by the relative availability of fast-food restaurants and convenience stores compared to grocery stores and produce vendors around respondents' home addresses. The size of the local food environment (i.e. the buffer, or radius, used to draw a circle around an individual's home) varied by the type of community in which the respondent lived. Different community types were identified using an urbanicity variable provided by Claritas Inc., a marketing research firm that assigns zip codes to one of four categories based on population density and location relative to a Census-designated population center.[5] The four categories include: urban, smaller city, suburban, and rural/small town. Smaller city (small or "second" cities that are population centers) and suburban (adjacent to major metropolitan areas) zip codes both have moderate population densities, while urban density is high and rural/small town density is low. Our urbanicity-specific buffers (0.5 mile in urban areas, 1 mile in smaller cities and suburban areas, and 5 miles in rural areas) were selected based on sample size, food retail density, transportation expectancies, and conventional trade area size for food retailers.[6]

The number of each type of food retailer within the urbanicity-specific buffer around the home address of each CHIS respondent was used to calculate the Retail Food Environment Index (RFEI). The numerator of the RFEI comprised the sum of the number of convenience stores and fast-food restaurants , while the denominator comprised the sum of the number of grocery stores and produce vendors. A higher RFEI indicates that a respondent lives near a larger number of fast-food restaurants and convenience stores relative to the number of grocery stores and produce vendors. For example, an individual with an RFEI of 2.0 has twice as many fast-food restaurants and convenience stores as grocery stores and produce vendors nearby.

Twenty-eight percent of California adults do not have any grocery stores or produce vendors within the specified buffers around their homes; therefore, the RFEI could not be calculated for these individuals. This population includes two subgroups: 18 percent who only have fast-food restaurants and convenience stores near home (but no grocery stores or produce vendors), and 10 percent who have none of the food outlets examined in this study near their homes. Individuals with no grocery stores near home are demographically heterogeneous, including affluent individuals who may have chosen to live farther away from commercial areas as well as lower-income individuals living in places with limited food options. Therefore, it is difficult to draw conclusions about the association between the retail food environment and health outcomes for these groups.

Construction of the RFEI assumes that fast-food restaurants and convenience stores are less likely to stock fresh fruits and vegetables and other healthy foods while grocery stores and produce vendors provide more healthy food options. The RFEI was used to assess the relative availability of different types of food retailers, as the absolute numbers of fast-food restaurants or grocery stores taken alone may describe local food environments less fully. While other studies have targeted "food deserts" (areas where there is no food access at all), this investigation was focused primarily on the implications of the balance of retailers in one's local food environment.

The exclusion of convenience stores associated with gas stations from the analyses leads our study to undercount the total number of convenience stores around respondents' homes. This will tend to bias the numerator of the RFEI downward. The use of an annual sales cutoff of $1 million to define grocery stores rather than the industry standard of $2 million leads our count of grocery stores to include small neighborhood stores as well as larger supermarkets. This will tend to bias the denominator of the RFEI upward. These two points, taken together, lead to a downward bias in our RFEI estimate for California adults.

Statistical analyses. We used descriptive and inferential statistics to analyze the relationships between the local food environment, community income, and individual health outcomes. Frequency tables were used to determine the prevalence of obesity and diabetes, and chi-square tests were used to detect significant differences. The average RFEI was also calculated for lower-income and higher-income communities, and two-tailed *t*-tests were used to detect significant differences. To examine the association of the RFEI with the prevalence of obesity and diabetes while adjusting for other factors, weighted logistic regression models were conducted using SAS and SUDAAN to account for the survey design of CHIS 2005.

To examine mean RFEIs at the county level, individual's RFEIs were averaged for residents of California counties with populations over 250,000. These counties represent over 90% of the state's population. These average RFEIs characterize the food environment for a typical adult in each of these counties. Corresponding data on obesity prevalence and age-adjusted diabetes prevalence for each of these counties was obtained from CHIS 2005.

All findings in the policy brief that are based on comparisons between groups or express probability are statistically significant ($p < 0.05$) except where otherwise noted. The cross-sectional nature of these data precludes conclusions regarding causality.

Notes
1. 200 percent FPL was used instead of 100 percent to account for the higher cost of living in California.
2. http://www.census.gov/prod/2003pubs/c2kbr-19.pdf
3. Analysts have used cutoffs of 20, 30, and 40 percent to determine whether or not a given neighborhood is low-income. See Paul Jargowsky, *Stunning Progress, Hidden Problems: The Dramatic Decline of Concentrated Poverty in the 1990s* (Washington, D.C.: The Brookings Institution, 2003); Tom Kingsley et al., *Concentrated Poverty: A Change in Course* (Washington, D.C.: The Urban Institute, 2003); Alemayehu Bishaw, *Areas With Concentrated Poverty: 1999* (Washington, D.C.: US Census, 2005; available at http://www.census.gov/prod/2005pubs/censr-16.pdf).
4. 2000 Census and 2005 California Health Interview Survey.
5. Claritas assigns zip codes to urbanization categories based on the analysis of population density grids of 1990 geoboundaries, 2000 redistricting updates, and 2001 population estimates. The following four classes were identified: 1) Urban areas have population density scores mostly between 85 and 99. They include both the downtowns of major cities and surrounding neighborhoods. Households within this classification live within the classic high-density neighborhoods found in the heart of America's largest cities. While almost always anchored by the downtown central business district, these areas often extend beyond city limits and into surrounding jurisdictions to encompass most of America's earliest suburban expansions. 2) Smaller cities are less densely populated than urban areas, with population density scores typically between 40 and 85, and are the population centers of their surrounding communities. This category also includes thousands of satellite cities—higher density suburbs encircling major metropolitan centers. 3) Suburbs have population density scores between 40 and 90. Unlike smaller cities, they are not the population center of their surrounding community, but rather a continuation of the density decline moving out from the city center. 4) Rural areas, collapsed into a single urbanization category, have population density scores under 40. This category includes exurbs, towns, farming communities and other sparsely populated portions of the state.
6. Two miles is the conventional trade area for most large grocery stores (>50,000 ft^2); e.g. suburban residents travel an average of 1.1 miles to go grocery shopping (this may not necessarily be the distance between their home and the closest store, as some are willing to travel farther if they have access to convenient transport options). See Susan L. Handy and Kelly J. Clifton. Local shopping as a strategy for reducing automobile travel. *Transportation.* 2001;28:317–346.