IS THERE AN ASSOCIATION BETWEEN FRUITS AND VEGETABLES PURCHASED FOR USE AT HOME AND CHRONIC DISEASE PREVALENCE IN SAN DIEGO COUNTY?

A Thesis
Presented to the
Faculty of
San Diego State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Public Health
with a Concentration in
Epidemiology

by
Susan C. Farrish
Spring 2012
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Susan C. Farrish:

Is There an Association Between Fruits and Vegetables Purchased for Use at
Home and Chronic Disease Prevalence in San Diego County?

[Signatures]

Caroline A. Macera, Chair
Graduate School of Public Health

John E. Alcaraz
Graduate School of Public Health

Mark Kern
School of Exercise and Nutritional Sciences

April 9, 2012
Approval Date
Copyright © 2012
by
Susan C. Farrish
All Rights Reserved
ABSTRACT OF THE THESIS

Is There an Association Between Fruits and Vegetables Purchased for Use at Home and Chronic Disease Prevalence in San Diego County?

by

Susan C. Farrish

Master of Public Health with a Concentration in Epidemiology
San Diego State University, 2012

While the prevalence of infectious disease has decreased during the 20th century, the prevalence of chronic disease has increased. The four common risk factors for chronic diseases are lack of physical activity, poor nutrition, tobacco use, and excessive alcohol consumption. Public health policy has been geared towards decreasing the preponderance of these risk factors. Of interest is the attempt to improve the diet of the US populace. As part of this effort, the US Department of Agriculture has developed dietary guidelines, with the most recent being published in 2011. It advocates that half of the food on an individuals plate, by volume, should consist of fruits and vegetables. This cross-sectional, ecological study examines the proportion, by weight, of food purchased for home use that is composed of fruits and vegetables, to determine if it is associated with hypertension rates within the six geographic regions that make-up San Diego County. Socioeconomic status and race/ethnicity were included as covariates. The results indicate that, at a population level, the residents of San Diego County households purchase the same amount of fruits and vegetables regardless of which region they live in; 15.9% of the food purchased for use at home, by weight, is composed of fruits and vegetables. In a multiple linear regression, none of the factors or covariates were significantly related to hypertension rates. These results can be used by the county health department to develop education programs for the population, about the types of foods that should and should not be consumed. It is also the basis for further investigation examining alternative methods to collect food data for future studies.
**TABLE OF CONTENTS**

ABSTRACT ............................................................................................................................. iv  
LIST OF TABLES .................................................................................................................. vii  
LIST OF FIGURES ............................................................................................................... viii  
ACKNOWLEDGEMENTS ..................................................................................................... ix  
CHAPTER  
1 INTRODUCTION .........................................................................................................1  
   Background ..................................................................................................................1  
   Statement of the Problem..............................................................................................2  
   Purpose of the Study.....................................................................................................3  
   Goals and Hypotheses.................................................................................................4  
   Basic Assumptions.......................................................................................................4  
   Limitations of the Study..............................................................................................4  
   Definitions....................................................................................................................5  
2 LITERATURE REVIEW ..............................................................................................6  
   Methods for Collecting Nutrition Data.......................................................................6  
   Diet and Chronic Disease............................................................................................7  
   Bias in Food Survey Studies.......................................................................................8  
   Other Sources of Food Data.......................................................................................9  
   Summary......................................................................................................................10  
3 METHODS ..................................................................................................................11  
   Study Population.......................................................................................................11  
   Data Sources ............................................................................................................11  
   Data Collection .........................................................................................................13  
   Outcome......................................................................................................................14  
   Major Exposure Variable...........................................................................................14  
   Covariates....................................................................................................................14  
   Statistical Analysis.....................................................................................................14
LIST OF TABLES

Table 1. Selected Demographics of the County of San Diego, Health and Human Services Agency Geographic Service Regions .................................................................13

Table 2. Descriptive Statistics of San Diego County, 2009 (n=6) .................................................................16

Table 3. Weighted Multiple Linear Regression of Hypertension Prevalence Versus All Variables (n=6) ........................................................................................................17

Table 4. Weighted Simple Linear Regression of Hypertension Prevalence Versus Each Variable (n=6) ........................................................................................................17

Table 5. Weighted Multiple Linear Regression of Hypertension Prevalence Versus Fruit/Vegetable Proportion and Each Covariate (n=6) .................................................................18
LIST OF FIGURES

PAGE

Figure 1. My plate ......................................................................................................................................2
Figure 2. County of San Diego, Health and Human Services Agency geographic service regions. .................................................................12
ACKNOWLEDGEMENTS

I would like to thank my committee members, Drs. Macera, Alcaraz, and Kern for their guidance during the thesis process. They have made the process as painless as possible and it has helped me appreciate the research they are involved in.

I would also like to thank the epidemiologists at the County of San Diego’s Emergency Medical Services and Community Health Statistics Unit for their help and moral support. It has been motivating to see how a good team works together and to be allowed to be a part of that team.

And lastly, I would like to thank my husband and family. Without their time and support the process would have been much more difficult.
CHAPTER 1

INTRODUCTION

BACKGROUND

In 1990 five of the top ten causes of death in the US were infectious (National Center for Health Statistics [NCHS], 2009). In 2009, only two of the top ten causes of death in the US were infection related (Kochanek, Xu, Murphy, Minino, & Kung, 2011). Of the remaining eight causes of death, six are grouped in the category of chronic diseases; these include heart disease, cancer, chronic lung disease, stroke, diabetes mellitus, and kidney disease (Kochanek et al., 2011). These chronic diseases all have four modifiable risk factors in common, which include lack of physical activity, poor nutrition, tobacco use, and excessive alcohol consumption (Centers for Disease Control and Prevention [CDC], 2010). Many studies have been done to try and determine the optimal diet to minimize the risk of developing these chronic diseases and to promote better health.

The United States government has used this data, at the public health level, to craft campaigns to improve the health of the populace; examples include Healthy People 2020 and the National Prevention Strategy (National Prevention Council, 2011; United States Department of Health and Human Services, 2011). The data has also been used to help develop the dietary guidelines published by the US Department of Agriculture (USDA). In 1992, the first Food Pyramid was developed which recommended a balanced diet that includes whole grains, fruits, vegetables, low-fat dairy, and low-fat protein with the recommended number of servings based on an individual’s daily calorie needs. This pyramid recommended a minimum of five servings of fruits and vegetables per day. Many surveys and studies have used these guidelines as a benchmark to measure consumer intake of fruits and vegetables. The most current guideline was released in 2011 and is called “MyPlate.” (Figure 1) It recommends that fifty percent of a person’s food intake, by volume, should be from fruits and vegetables. It was developed as a visual depiction to better help consumers understand and remember the make-up, by proportion, of a healthy diet (USDA, 2011).
In San Diego County, the leading causes of death are similar to those at the national level. Of the top ten, six are due to chronic disease (County of San Diego, Health and Human Services Agency, Public Health Services, Epidemiology & Immunizations Branch, 2011). To combat this, the County public health officials have developed the 3-4-50 campaign that describes three behaviors – poor nutrition, lack of physical activity, and tobacco and substance abuse – that contribute to four diseases – cancer, vascular disease, diabetes type 2 (DM2) and lung diseases – that cause over 50% of all deaths in San Diego County. This campaign attempts to inform the public and to improve their health-related choices (County of San Diego, Health and Human Services Agency, Public Health Services, Community Health Statistics Unit, 2010a).

**STATEMENT OF THE PROBLEM**

The studies that have been used to help develop the USDA food guidelines, and other dietary guidelines, to improve the health of the population were accomplished by collecting
dietary information from study participants. The methods used to gather the diet information from the participants varies with the most common methods being the food frequency questionnaires (FFQ), 24-hour recall, and diet diaries. They collect information about an individual’s food and nutrient intake during a recent period (a week, month, year, etc.) along with information about diet-related lifestyle habits. The most dominant of these tools is the FFQ, which provides information about how often a particular food item is consumed. The patient is given a list of foods and asked to relate how often the item is consumed and may also be asked the portion size (Bountziouka & Panagiotakos, 2010). Studies have shown that there are biases when a subject is asked to self report their food intake. These biases are psychologically based. One in particular that has been studied is called social desirability bias. It is a bias that “reflect(s) the defensive tendency (of a respondent) to respond in a manner consistent with perceived social norms” (Hebert et al., 2008, p. 2265-2345). In an attempt to reduce this bias, this study will use data collected in a different manner, as food expenditures for a household, at an aggregate level, to determine if the same association exists between diet and the rate of one chronic disease, hypertension, since it is the most prevalent chronic disease in San Diego County. The study will also determine if the relationship is affected by the income level of the community, the education level of the community, or the primary racial/ethnic background of the community.

**PURPOSE OF THE STUDY**

The main purpose of this study is to determine if there is a difference in the food purchased by household, in different regions of San Diego County, and if this difference is associated with differences in hypertension prevalence. The study will also ascertain how race/ethnicity, education level, and income level of the community affect this relationship.

If the data give similar results as those studies that use FFQs and food diaries, then this may prove to be a new method of collecting data. This collection method could be used to examine further the relationship of diet/nutrition to chronic disease rates. If the results are not similar it could indicate that bias in reporting plays a bigger role than previously thought or, that this method of data collection is not an acceptable alternative.
GOALS AND HYPOTHESES

1. The proportion of the food purchased for home use, that is composed of fruits and vegetables, by weight, will account for the variability in hypertension among San Diego County Residents.
   • Those who purchase a larger proportion of fruits and vegetables for use at home will have a lower rate of hypertension.

2. Among the characteristics under investigation, race/ethnicity, socioeconomic status (income, educational attainment), and geography will be associated with the rate of hypertension for the six geographic service regions within San Diego County.
   • Those with a lower socioeconomic status and a minority race/ethnicity will have higher rates of hypertension.

BASIC ASSUMPTIONS

1. Results will be generalizable to residents in San Diego County.
2. Reported hypertension and socioeconomic status rates for each region are accurate.
3. Methods for assessing associations were consistent during the analysis.
4. Study data were entered completely and correctly.

LIMITATIONS OF THE STUDY

The main limitation of this study is that it is examining the weight of food purchased, in kilograms (kg), not the volume of food purchased. One hundred grams of raw spinach has a much different volume than 100g of cooked spinach. Conversely, if half of a person’s intake is fruits and vegetables, by volume, that may not be the same as 50% of the weight of the fruits and vegetables consumed.

The data used is the proportion of the weight of food for household use in the different categories of food. This does not directly correlate with the proportion of an individual’s diet that is consumed in each food category. If this study shows that further investigation is needed in this area, then studies will need to be done to determine the correlation between food purchased for a household and food consumed by an individual.

Another limitation of the study is that the data are not available at the household level. Although the data were collected at the household level, they were only available after being aggregated.
DEFINITIONS

Bureau of Labor Statistics, Consumer Expenditure Survey (BLS CEX): This survey is accomplished annually to determine consumer expenditures for goods in the US.

Cerebral Vascular Accident (CVA): More commonly known as a stroke.

Diabetes Mellitus type 2 (DM2): A disease associated with elevated levels of sugar in the blood.

Food Frequency Questionnaire (FFQ): This is a questionnaire used to gather information on the usual dietary intake of a study participant. It is often a list of foods (that the study is interested in) and the participant marks down how often they eat that item. Some FFQs also ask about quantity of foods consumed.

Hypertension (HTN): More commonly know as high blood pressure.

National Health and Nutrition Examination Survey (NHANES): This is a government survey that is accomplished periodically. It includes a physical exam and questions about a participant’s health and diet.

San Diego County: A county in the United States located in the southwestern corner of California and is composed of 18 incorporated cities and towns, including the City of San Diego

San Diego County’s Health and Human Services Agency Geographic Service Regions: Six geographic areas that collectively make-up San Diego County, that are used to deliver services to residents and collect demographic and health information.
CHAPTER 2
LITERATURE REVIEW

Research studies of morbidity and mortality due to chronic diseases have implicated diet, especially as it relates to being overweight or obese, as a major contributor (McCullough & Willett, 2006; Millen et al., 2001; Millen, Quatromoni, Nam et al., 2005; Millen, Quatromoni, Pencina et al., 2005). Within the United States, there is an obesity epidemic. In 2000 no state had an obesity prevalence of 30%; in 2010 there were 12 states with an obesity prevalence of 30% or more (CDC, 2011). The increase in obesity rates has been part of the impetus to do research to find a way to improve the diet of the US population. The research also continues to determine what type of diet will decrease an individual’s chance of developing a chronic disease.

A review of the research linking diet to increased morbidity and mortality that has been published in the last 10-15 years, reveals that many of the studies used data collected via food frequency questionnaires (FFQs) or 24-hr food diaries (Bountziouka & Panagiotakos, 2010). Both of these methods are prone to social desirability bias, which is defined as “the tendency of an individual to convey an image in keeping with the social norms and to avoid criticism in testing situations” (Hebert, Clemow, Pbert, Ockene, & Ockene, 1995). Studies have shown that data from some FFQs have more bias than a 24-hour diet recall; the variation can be a high or low estimate and can depend on the population demographics (Wirfalt, Jeffery, & Elmer, 1998). Also, women tend to show more bias than men resulting in an under-reporting of fat and calorie intake (Hebert et al., 2008).

METHODS FOR COLLECTING NUTRITION DATA

When a study intends to gather nutrition information there are several ways to gather the data. The researcher can ask the participant to complete a questionnaire about what they have consumed over a specified time frame, the participant can keep a diary, the participant can be interviewed, or the researcher can use a combination of these three methods. The information collected includes the type, and sometimes the quantity, of food consumed. The
researcher may ask about specific foods the participant consumed during a time period or may ask about what the participant typically consumes. Given that the data collected by the researchers is mostly subjective, the method of collection, as mentioned above, must be reliable and valid. A tool that is reliable measures data precisely, consistently; a tool that is valid measures accurately. There are methods available to evaluate the reliability and validity of a nutritional measurement tool (Bountziouka & Panagiotakos, 2010).

The most frequently used method of gathering nutrition information is the food frequency questionnaire (FFQ). It is a questionnaire that assesses how frequently a participant has consumed specific food items. Some FFQs are quite extensive attempting to gather detailed information on everything that was consumed. Others may be only be interested in specific nutrients so the list of foods may be much shorter. They are frequently used since they are easy to administer and inexpensive to process (Wirfalt et al., 1998).

Additional methods of collecting nutrition information are a food interview or a food diary. One of the largest studies in the US that uses these methods is the National Health and Nutrition Examination Survey (NHANES). This survey is conducted by the National Center for Health Statistics (NCHS), which is part of the Center for Disease Control and Prevention (CDC) (CDC/NCHS, 2011). The survey collects information about the participants food intake over the previous 24-hours, via an interview, at the initial encounter. A second 24-hour dietary recall is accomplished 3-10 days after the initial interview. For this recall the participant is instructed to keep track of their food intake for a 24-hour period, along with amount, and then relay that information to the interviewer (NHANES, 2012). When compared to the FFQ, both of these methods can be more burdensome for the participant (Bountziouka & Panagiotakos, 2010) but they are less prone to bias since it is usually a short-term evaluation (Hebert et al., 1995).

**Diet and Chronic Disease**

The Framingham Heart Study researchers have done much of the recent research that relates diet to increased risk for chronic diseases. Much of the research from this study was done on the relationship between diet and cardiovascular disease and other related diseases, such as carotid atherosclerosis. The studies have shown that men who eat a diet close to a heart healthy diet have a lower risk of developing heart disease (Millen, Quatromoni,
Pencina, et al., 2005); women who eat a diet consistent with the American Heart
Association’s Dietary Guidelines have a lower odds of developing carotid stenosis (Millen,
Quatromoni, Nam, et al., 2005).

Other studies from Framingham link diet and weight gain, which is a risk factor for
chronic diseases such as heart disease, stroke, cancer, diabetes, and arthritis (CDC, 2010).
The Framingham studies have shown that those individuals who adhere to the dietary
guidelines published by the USDA in 1995 have less weight gain over an eight-year period
(Quatromoni, Pencina, Cobain, Jacques, & D'Agostino, 2006).

Another study of note examined the Healthy Eating Index (HEI), which is a tool used
to numerically evaluate diet quality. It was developed by the USDA in 1995. Recently,
researchers changed the HEI by incorporating some of the more recent dietary
recommendations - developing the Alternate Healthy Eating Index (AHEI). The AHEI
showed stronger protective effects of a healthy diet compared to the HEI. Most notably, the
AHEI has been “shown to be twice as strong as the original HEI in predicting overall chronic
disease risk in US men and women, primarily driven by a marked inverse relation with
cardiovascular disease” (McCullough & Willett, 2006).

Food consumption patterns are a newer concept in chronic disease research. These
studies have shown that diets with high intake of fruits and vegetables, whole grains, and
low-fat items have a lower risk for some chronic diseases. This information will be helpful
in further understanding the disease discrepancies that have been shown to exist between
different ethnic groups. It will also be important in guiding the development of public health
food guidelines at all levels of government (Barkoukis, 2007).

**Bias in Food Survey Studies**

Of all the studies reviewed, many mentioned that self-recall could bias the collected
data but only one discussed bias similar to the previously discussed social desirability bias. It
stated that “the survey data used for this study was based on self-reports, and research has
shown that self-reported data, particularly pertaining to socially desirable behaviors, are
subject to biased recall” (Tsai et al., 2010).

Of the studies that used 24-hour dietary recall, the following limitations or biases
were discussed: misclassification of dietary components could underestimate the effects if
the dietary item of interest (Ajani, Ford, & Mokdad, 2004); given the short amount of time covered by a 24-hour dietary recall it “may not be adequate to assess individuals’ usual diet” (Chen, Cheskin, Shi, & Wang, 2011) “because day-to-day intake is highly variable for many persons” (Dubowitz et al., 2008); and lastly “a 24-hour dietary recall can be a restricted measure of overall nutrition and is subject to inaccuracies and recall bias” (Lane et al., 2008). As can be seen, the description of the limits in each of these studies are similar, usually focusing on inaccuracies in collecting dietary data.

The studies that used FFQs did not discuss the limits of this type of diet survey as frequently as the studies that used 24-hour recall diet diaries. In one study, a limitation was that the FFQ was truncated since it was only one component in a larger study (Chang et al., 2011). One other study discussed the limitations of the FFQ in more detail stating “that self-administered FFQs have limitations, particularly in respect to quantity” of intake (von Ruesten, Illner, Buijsse, Heidemann, & Boeing, 2010). Since the authors did not often mention the limits associated with FFQs this may indicate that researchers believe that FFQs are more reliable and valid at collecting dietary information from study participants.

**OTHER SOURCES OF FOOD DATA**

There are other types of information available about food consumption. Two studies were found that compared household food consumption to individual food consumption. Both of the studies found some correlation between the two. One found a relationship between household and individual food consumption present for major foods including cereal products, milk, cheese, meat and meat products, and fish (Becker, 2001). The other found a relationship between the household and individual consumption when looking at nutrients and additives within the foods (Lambe et al., 1998).

In a review of the literature, two sources of food expenditure data were found. One was the Consumer Expenditure Survey (CEX), which is conducted by the US Department of Agriculture. The other method is the collection of receipts from individuals/households for purchases of food items. The research that was done using the CEX data were evaluating the cost of food. One was examining the cost of a Calorie over time. It concluded that there are data showing that, relative to the general cost of food, the cost of fruits and vegetables has increased disproportionately over time and that this could be a minor contributor to the
increased trend in obesity (Christian & Rashad, 2009). No articles were found that were determining if there was an association between food purchased and chronic disease rates.

Two studies were found that used receipts for food purchased by a household. One compared the data they collected to equivalent data from the CEX from 2004-2005; the findings from the receipts were consistent with the CEX data (French, Wall, & Mitchell, 2010). The other study compared the proportion of the total dollars spent on different categories of food – protein, fruit, vegetables, grains, etc. They found differences in purchasing by ethnic group and concluded that this was a data collection method that should be further investigated (Cullen et al., 2007).

**SUMMARY**

The current literature, which examines the reasons for the increase in chronic disease rates, uses dietary recall diaries and FFQs. These have been shown to have bias towards under-reporting, especially when the study participants know that the data is being used to evaluate their health. It has also been reported that sometimes the participants report their diet as being more similar to the intervention they are involved in than it really is (Hebert et al., 2008).

This study is using data that was collected as part of an economic survey; the data used is the amount of money spent by a household on food for use at home. It is attempting to eliminate the social desirability bias from the equation by using objective dietary data. Studies have shown that the food purchased is equivalent to the food consumed (Becker, 2001) and that household purchases are equivalent to individual consumption level of foods (Lambe et al., 1998). Another study has compared the household food purchase data, gathered from household receipts, to those of CEX data from the same time frame, and the results were similar (French et al., 2010). This study is taking all of this into consideration.
CHAPTER 3

METHODS

This study is a cross-sectional ecological study to evaluate the nutritional intake across San Diego County in 2009. The Institutional Review Board at San Diego State University approved the study.

STUDY POPULATION

The study population is comprised of San Diego County residents. The residents are multi-racial; the racial/ethnic composition, per San Diego Association of Government’s (SANDAG) 2009 estimates, are 53.1% White, 29.6% Latino, 8.9% Asian, 5.1% African-American, 0.8% American-Indian/Alaska Native, 0.5% Native Hawaiian/Pacific Islander, and 2.1% identifying themselves as two or more races. Other characteristics that could be considered covariates for fruit and vegetable purchased by a household were obtained for each geographic service area of San Diego County. There are six geographic service regions within San Diego County that are used to deliver services to residents and collect demographic and health information (County of San Diego, Health and Human Services Agency, Public Health Services, Community Health Statistics Unit, 2010). (Figure 2 and Table 1)

DATA SOURCES

Consumer Expenditure (CEX) Database:

The CEX survey is a national level survey with data collected at a household level via quarterly interviews, for major item purchases. The food data come from a two week detailed diary kept for each household. The data for this study come from the Bureau of Labor Statistics Consumer Expenditure Surveys 2006 and 2007, adjusted for 2010 income levels; the adjustment was done by Esri (Redlands,CA) (Esri, 2010). Average annual consumer spending by household was aggregated by sub-regional area, as established by the San Diego Association of Governments (SANDAG); the epidemiologists at the County of San Diego’s Emergency Medical Services and Community Health Statistics Unit completed the data aggregation.

Food Pricing Database:

Data on average food costs were obtained from the USDA Quarterly Food-at-Home Price Database (QFAHPD) for 2009. This database uses data from Universal Product Codes (UPCs) and random-weight food items to estimate the household level food prices for 54 foods or food groups, per weight (100g). The data is collected in twenty-six metro and nine non-metro areas. The Metro California group includes San Diego and Sacramento Counties (Todd, Mancino, Leibtag, & Tripodo, 2010). For this study, the pricing data was presumed to be the same throughout the county.
Table 1. Selected Demographics of the County of San Diego, Health and Human Services Agency Geographic Service Regions

<table>
<thead>
<tr>
<th>HHSA Regions of San Diego County</th>
<th>Primary Race of Region</th>
<th>Percentage with Income Less than $50,000 Annually</th>
<th>Percentage with Education Level of HS Degree or Less</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Coastal</td>
<td>White (Non-Latino)</td>
<td>41.4%</td>
<td>26.8%</td>
</tr>
<tr>
<td>North Central</td>
<td>White (Non-Latino)</td>
<td>34.4%</td>
<td>16.8%</td>
</tr>
<tr>
<td>Central</td>
<td>Latino</td>
<td>59.3%</td>
<td>45.0%</td>
</tr>
<tr>
<td>South</td>
<td>Latino</td>
<td>56.5%</td>
<td>47.9%</td>
</tr>
<tr>
<td>East</td>
<td>White (Non-Latino)</td>
<td>42.9%</td>
<td>41.5%</td>
</tr>
<tr>
<td>North Inland</td>
<td>White (Non-Latino)</td>
<td>34.2%</td>
<td>38%</td>
</tr>
</tbody>
</table>


California Health Interview Survey:

Demographic and disease rate data were obtained from the California Health Interview Survey (CHIS) from 2009. CHIS is a random dial telephone survey administered to California residents every other year. The data was collected in a manner to ensure that it is a statistical representative of the California population. Data for San Diego County is available for the six geographic service regions (CHIS, 2011).

San Diego Association of Governments (SANDAG):

Population data were obtained from the San Diego Association of governments (SANDAG) from 2009. SANDAG is a public agency that creates and maintains demographic, economic, land use, transportation and criminal information about San Diego County (SANDAG, 2009).

**Data Collection**

The data for this study was combined into one data set per region from data obtained from the Consumer Expenditure Database, the US Department of Agriculture’s Food Pricing Database, the California Health Interview Survey, and the San Diego Association of Governments. The data set for each region included one value for the outcome variable, the major exposure variable, and each covariate, as described below.
**OUTCOME**

The outcome of interest is the prevalence of hypertension in 2009. The prevalence rates were available on-line through CHIS and are not age-adjusted. They are available for each of the six geographic regions within San Diego County.

**MAJOR EXPOSURE VARIABLE**

The major exposure variable is the proportion of food bought for home use, by weight, that is composed of fruits and vegetables. The data was available from the CEX database, in dollars, and was converted to weight using data from the USDA QFAHPD (Todd et al., 2010).

**COVARIATES**

The variables included as covariates for hypertension are income, education, and race/ethnicity. This data was obtained from the 2009 California Health Interview Survey (2011).

- **Income**: Income was assessed as the percentage of the population that had an annual household income less than $50,000.
- **Education**: Education status was assessed by the percentage of the population that had a high school diploma or less.
- **Race/ethnicity**: Race was assessed as the majority race in each geographic service region; the races included were white (non-Latino) and Latino.

**STATISTICAL ANALYSIS**

Data was analyzed using SAS version 9.2 (SAS Institutes, Cary, NC). Information provided by sub-regional area, on the proportion of food purchased for home use that was composed of fruits and vegetables, in percent weight, was compared by region. An analysis of variance was completed. The cases were weighted by population to ensure equal variances, as required by linear regression.

The means for fruit and vegetable proportions at the region level, from the analysis of variance, were combined with the population data on education, income, and race/ethnicity, to perform a multiple linear regression analysis. A linear regression analysis was completed for each variable, independently. Linear regression was also completed with fruit and vegetable proportion in addition to one other variable for a total of three more equations. For
all of these calculations, the variables were weighted by population to ensure equal variances. Univariate analyses were performed to provide descriptive statistics for the independent statistics.
CHAPTER 4

RESULTS

The fruit and vegetable data mean was evaluated by an analysis of variance for the 41 sub-regional areas within San Diego County. There was no significant variation noted between the mean values at the regional level (p=0.24). The sub-regional values ranged from 14.73% to 16.44%, with a mean of 15.9% and standard deviation of 0.37%. The mean and standard deviation for all variables plus population, at the regional level are in Table 2.

Linear regression failed to show a significant relationship between hypertension and all the variables in a full equation as in Table 3.

Table 2. Descriptive Statistics of San Diego County, 2009 (n=6)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension Prevalence</td>
<td>22.9%</td>
<td>30%</td>
<td>26.4%</td>
<td>2.9</td>
</tr>
<tr>
<td>Fruit / Vegetable Proportion</td>
<td>15.9%</td>
<td>16.2%</td>
<td>16.1%</td>
<td>0.1</td>
</tr>
<tr>
<td>Income less than $50K/yr</td>
<td>34.2%</td>
<td>59.3%</td>
<td>44.8%</td>
<td>10.8</td>
</tr>
<tr>
<td>Education of HS degree or less</td>
<td>16.8%</td>
<td>47.9%</td>
<td>36%</td>
<td>11.9</td>
</tr>
<tr>
<td>Population of regions</td>
<td>460,739</td>
<td>624,072</td>
<td>553,710</td>
<td>61,720</td>
</tr>
</tbody>
</table>

Further evaluation by linear regression with the variables independently and in combination with fruit/vegetable proportion failed to show any significant relationship, as in Table 4 and 5, respectively.
### Table 3. Weighted Multiple Linear Regression of Hypertension Prevalence Versus All Variables (n=6)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimated Coefficient</th>
<th>95% Confidence Interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit/Vegetable Proportion</td>
<td>-175.6</td>
<td>-1655.2, 1303.9</td>
<td>0.37</td>
</tr>
<tr>
<td>Primary Race/Ethnicity</td>
<td>65.9</td>
<td>-454.7, 586.5</td>
<td>0.36</td>
</tr>
<tr>
<td>Income below $50K per year</td>
<td>-1.7</td>
<td>-14.0, 10.7</td>
<td>0.33</td>
</tr>
<tr>
<td>Education of HS or less</td>
<td>-0.5</td>
<td>-6.9, 5.9</td>
<td>0.49</td>
</tr>
</tbody>
</table>

### Table 4. Weighted Simple Linear Regression of Hypertension Prevalence Versus Each Variable (n=6)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimated Coefficient</th>
<th>95% Confidence Interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit/Vegetable Proportion</td>
<td>0.8</td>
<td>-47.6, 49.3</td>
<td>0.96</td>
</tr>
<tr>
<td>Primary Race/Ethnicity</td>
<td>2.3</td>
<td>-5.3, 10.0</td>
<td>0.45</td>
</tr>
<tr>
<td>Income below $50K per year</td>
<td>0.1</td>
<td>-0.3, 0.5</td>
<td>0.53</td>
</tr>
<tr>
<td>Education of HS or less</td>
<td>0.2</td>
<td>-0.1, 0.4</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Table 5. Weighted Multiple Linear Regression of Hypertension Prevalence Versus Fruit/Vegetable Proportion and Each Covariate (n=6)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Estimated Coefficient</th>
<th>95% Confidence Interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F/V Proportion</td>
<td>-19.8</td>
<td>-98.9, 59.2</td>
<td>0.48</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>4.8</td>
<td>-8.7, 18.4</td>
<td>0.34</td>
</tr>
<tr>
<td>F/V Proportion</td>
<td>-6.4</td>
<td>-75.4, 62.6</td>
<td>0.79</td>
</tr>
<tr>
<td>Income</td>
<td>0.1</td>
<td>-0.4, 0.7</td>
<td>0.55</td>
</tr>
<tr>
<td>F/V Proportion</td>
<td>0.2</td>
<td>-48.4, 48.8</td>
<td>0.99</td>
</tr>
<tr>
<td>Education</td>
<td>0.2</td>
<td>-0.2, 0.5</td>
<td>0.23</td>
</tr>
</tbody>
</table>
CHAPTER 5

DISCUSSION

The purpose of this ecological, cross-sectional study which examined the fruit and vegetables purchased for use in the home in San Diego County’s six geographic services regions, was to determine if the variation could account for part of the variability in chronic disease rates, specifically hypertension (HTN). The study does not support the hypothesis that chronic disease rate variability is explained, in part, by the proportion of the food purchased for use at home that is comprised of fruits and vegetables. This was mainly because the proportions of fruits and vegetables are the same across the six regions that make up San Diego County. Studies that have looked at the relationship of fruit and vegetable intake as a protective factor for developing hypertension have found a relationship, for women, but once body mass index (BMI) was adjusted for the relationship was eliminated (Wang, Manson, Gaziano, Buring, & Sesso, 2012). The DASH diet, which is high in fruits and vegetables, has been found to help reduce blood pressure (Appel et al., 1997) but it is also low in salt which is another dietary component known to greatly affect blood pressure (Appel et al., 2011). Additionally, the data do not support the hypothesis that socioeconomic status (income and educational attainment) and race/ethnicity are associated with hypertension rates across the regions within San Diego County. This result could be due to the small sample size and subsequent low statistical power of the study.

KEY FINDINGS

The fruit and vegetable proportion was not significantly different across the six regions that make up San Diego County. This information is similar to that found by the Community Health Statistics Unit (CHSU) of the Health and Human services Agency (HHSA) of San Diego County. They divided all the food data from CEX into the five food groups with one additional group called “Empty Calories.” This group included sweetened, high fat, and processed grain items that do not follow the guidelines for a healthy diet as established by the USDA. The data for these six groups were similar across all sub-regional
areas of San Diego County, when evaluating proportions (County of San Diego, Health and Human Services Agency, Public Health Services, Community Health Statistics Unit, 2012).

Given that the socioeconomic and race/ethnicity variables, along with fruit/vegetable proportions were not associated with hypertension rates among San Diego County residents, this suggests that other factors are associated with the variability in hypertension rates across the regions. Further evaluation is warranted within the literature and research that has been done, to determine if other factors need further investigation.

**STRENGTHS**

The main strength of this study was that the data about purchased food was collected via a two week detailed diary rather than by a recall survey. This method attempts to minimize recall bias and social desirability bias. Additionally, the study is generalizable to the San Diego county population. Finally, since the study used population data, the hypertension rates were normally distributed which allowed the use of multiple linear regression.

**LIMITATIONS**

This study had several limitations. First and foremost, the sample size was small. Since the data available was at the regional level, there were only six cases available for evaluation and this has very low statistical power to detect any significant associations. It is possible that if the data was available at the household level, there would have been a significant difference between households and this would have further helped to find factors that are related to hypertension rates in San Diego County. Second, the data was at a population level and as such, the results can only be used at the population level and not at a household level. Ecological studies, or population studies, are often used to help generate hypothesis; they are most useful in identifying risk factors that are relatively homogeneous within a population but differ greatly between populations or change in time within a given population (Tominaga & Kuroishi, 1997). This data can be useful in following the diet of the population of the County of San Diego, over time (Golzarand et al., 2012). Third, the food data was initially available in dollars spent per household. It was converted to weight of food purchased. The USDA guidelines are for volume of food and weight does not directly relate to volume. This conversion could have over or underestimated the proportion of food
purchased in the form of fruit and vegetables. Next, since the data is for food purchased it does not take into consideration waste in preparation or non-consumed items. Lastly, the hypertension rates are self-reported and may not reflect the true rate of hypertension within the population – this data could have been intentionally, or unintentionally, reported inaccurately which would affect the percentages reported in the California Health Interview Survey.

**IMPLICATIONS**

The results of this study could have important implications for the future of public health in San Diego County. These data imply that there is little variation in the diets of the population of San Diego County, across regions. The interventions that have been designed, to date, do not appear to be working possibly because it is difficult to change individual consumer preferences (Zhao et al., 2011); people are not eating enough fruits and vegetables. The population may have become immune to the public service messages that are currently being used and new messages need to be designed, with more impact, to improve the diet of the San Diego County population.

**FUTURE DIRECTIONS**

More studies need to be done to evaluate the relationship between food purchased and food consumed. If the relationship is valid, as was partially shown by Cullen and French (Cullen et al., 2007; French et al., 2010) then this may be a better way to evaluate the diet of the population while minimizing self-report and social-desirability bias. Given the small sample size of this study, more studies should be done either at a larger scale, if the study is an ecological study, or in more detail, with data coming from households. I would suggest that the first step would be to complete a larger scale study given the availability of the data, from the CEX, for the entire nation. This may help to better determine the cause for the trend of increasing obesity prevalence across the United States (CDC, 2011).
BIBLIOGRAPHY


