DRIVERS OF FOOD FORM CONSUMPTION AND FOOD FORM'S IMPACT ON TOTAL CALORIE INTAKE

by

Suraksha Baral

(Under the Direction of William Secor)

ABSTRACT

Total calorie intake has implications on various epidemiological as well as economic analyses. The objective of this thesis is to investigate the factors affecting the consumption of different food forms as well as to assess the effect of consumption of food forms on total calorie intake. I analyze information from the 2017-18 NHANES data set, which involves a 24-hour dietary recall and questionnaire responses. Multinomial logistic regression is used to assess factors affecting the choice of food forms. Propensity score matching is used to measure the impact of food form choice on total calorie consumption. Socio-demographic variables such as gender, age, ethnicity, education and physical activity play important roles in the consumption of raw versus processed food forms. Most food types have a significant correlation with the calorie consumption, except for a few comparisons. The research provides initial insights to policymakers about the impact of food form on total calorie intake that may be linked to public health outcomes such as obesity.

INDEX WORDS: Consumption, food forms, NHANES, calorie

DRIVERS OF FOOD FORM CONSUMPTION AND FOOD FORM'S IMPACT ON TOTAL CALORIE INTAKE

by

SURAKSHA BARAL

B.S., Agriculture and Forestry University, Nepal, 2019

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirement for the degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2023

DRIVERS OF FOOD FORM CONSUMPTION AND FOOD FORM'S IMPACT ON TOTAL CALORIE INTAKE

by

SURAKSHA BARAL

Major Professor: William Secor

Committee: Gopinath Munisamy
Berna Karali

Electronic version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

DEDICATION

Dedicated to my beloved family.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. William Secor, for his unwavering support, encouragement, and guidance throughout my entire research journey. I am thankful for the knowledge and experience I have gained through our collaboration. I would like to extend my sincere appreciation to my committee members, Dr. Gopinath Munisamy and Dr. Berna Karali, for their insightful feedback and comments that helped me improve my thesis. I am also grateful to Dr. Benjamin Campbell for his continuous support and mentorship throughout my graduate studies. I would like to thank my friends Sudarshan Adhikari, Sushmita Bhatta, Rita Khadka, Laxmi Narayan Ojha, Subin Poudel, Aadesh Subedi, and Prabin Adhikari for their motivation and emotional support. Finally, I would like to give a special recognition to my parents and sisters, Apekshya and Aashma, without whom this achievement would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF APPENDICES

1 INTRODUCTION

People consume a variety of foods to fulfill their calorie and nutritional needs. These food choices are a function of hunger, food attractiveness, time constraints, food availability, familial influences, situational aspects, personality factors, and many more (Neumark-Sztainer et al., 1999). Additionally, these and other factors affect food-form choices. Some consumers prefer raw foods whereas the others prefer processed forms. Alternative food forms offer different levels of satiety (i.e., feeling full) to consumers (Pan & Hu, 2011). The consumption of various food forms differs according to the satisfaction the consumers gain from these forms (Flood-Obbagy & Rolls, 2009). Taken together, several factors govern food-form decisions, and these decisions have impacts on subsequent consumption decisions that may affect total food (i.e., calorie) intake.

This research seeks to answer two major questions. The first is to determine the factors influencing the consumption of various food forms. For this study, food forms in the following four food types are considered: peanut, almond, apple, and orange. Peanuts and almonds are among the most popular nuts in the United States (U.S.). The latest statistics show a record-breaking growth in per capita intake (NPB, 2021). In 2020, 299.34 million Americans consumed peanut butter (Statista, 2022a). Almonds has per capita consumption at 2.5 pounds per capita in 2020/21 (Statista, 2022b). Among fruits, Apples and oranges were the most preferred fruits in the U.S. in year 2019 with fresh fruits and juices being the most dominant forms of consumption (USDA, 2021a). Per capita consumption of fresh apple in the U.S. is in fluctuation year by year. However, the consumption remains considerably high, which was 16.18 pounds per capita in

2020/21 (Statista, 2021a). Statista (2021b) estimates that Americans consumed roughly 26.9 pounds of raw citrus fruit per person in the year 2020.

Individual foods that were consumed by respondents were categorized based on food codes provided by the Food Patterns Equivalent Database (USDA, 2021b). The nuts (almond and peanut) were classified as raw or processed. Raw forms included unaltered nuts like roasted or salted peanuts and almonds. Processed forms included items like peanut or almond butter, chocolate-covered nuts, and processed goods like almond or peanut flour. For the fruits (apple and orange), consumption was divided into raw and juice forms. To clarify, the raw form of fruits refers to both the whole form and cut-up pieces. In contrast, the processed form of fruits only includes juice form.

The food types of our interest have been studied in relation to various economic and health related researches in the U.S. In particular, research on peanut consumption has focused on health risks (Fakhri et al., 2022) and benefits (Ciftci & Suna, 2022). Research on almond consumption has focused on marketing and trade (Ajibade & Saghaian, 2022; Xu et al., 2022). Recent research on apples relate to consumer preferences on attributes (Meike et al., 2022) as well as economic analysis of certain apple production aspects (Zhang et al., 2022). Orange-related research in the U.S. have focused on its most prevalent food form (Juice) relating to health risks (Alhabeeb, et al., 2022; Cara et al., 2022). These nuts and fruits have rarely been studied based on the factors affecting their food forms and their relation to energy intake. I intend to fill this gap of research in the U.S. context with this study.

Our second research concern is to assess the effect of food forms' consumption on total calorie intake in the U.S. Calorie intake is often linked to the prevalence of weight gain or obesity (Hashem et al., 2022; Rosenheck, 2008). The association between calorie consumption and different health issues requires research that examines how various factors are linked to calorie intake, and how these factors may contribute to the development of health problems. Different food forms may affect total calorie intake by individuals. Processed drinks and foods make up the majority of the U.S. dietary source providing around 75% of daily caloric consumption in the country (Poti et al., 2015). In this context, it would be relevant to study the determinants contributing to consume various types of foods, which ultimately leads to intake energy in respective ways.

This study adds knowledge on the scant research held relating food forms to total calorie intake considering several factors. It will educate agribusinesses on the key factors affecting many product kinds they may manage and produce. This research helps health and economic policy makers, researchers, and other stakeholders directly or indirectly related to dimensions of food forms' consumption and energy intake.

The results suggest that there is a high proportion of non-consumers for all commodities (peanut, almond, apple, and orange), and consumption of raw and processed food forms varies by food type. Nuts are consumed more in processed form than raw, while the consumption of fruits in raw and juice form varies according to the fruit type. Raw apples are more commonly consumed than apple juice, while orange juice is more popular than raw oranges. The impact of demographic variables on consumption patterns also varies by food type, with some variables showing consistency within a particular food type but not across others. For example, age and certain ethnicities are associated with nut food form consumption, whereas age is not a consistent predictor of fruit food form consumption. These findings suggest that the choice of food form may be product-specific and depend on various factors.

The study also finds that, except for oranges, there is no significant difference in calorie consumption between raw and processed food forms. This indicates that not all processed food forms necessarily lead to increased calorie intake. In the case of oranges, however, juice drinkers consume more calories than those who eat raw oranges. Overall, these results suggest that factors such as food type, demographic variables, and calorie content should be considered when examining the impact of food form on consumption patterns.

Chapter 2 provides a review of pertinent literature. Subsequently, Chapter 3 explains the methods used in this study. In Chapter 4, the data used in the analysis is discussed. The results are then presented in Chapter 5. Finally, Chapter 6 concludes the thesis and makes suggestions for further study.

2 LITERATURE REVIEW

2.1 Introduction

The initial segment of this chapter introduces several food forms, alongside an exploration of the complex factors that influence food consumption decisions. Moreover, the chapter discusses research that examines calorie intake in general, as well as the impact of food types and socio-demographic factors on food consumption.

2.2 Food Consumption Factors

Different people have different options in our world, and this is most clearly visible in the food choices people make for themselves. The consumption of various food forms differs according to the satisfaction people gain from these forms (Flood-Obbagy & Rolls, 2009). Consumer perception on flavor of the foods is considered to help determine their food preferences (Lawless & Heymann, 2010). However, there are a variety of factors that influence people’s food choices, with hunger being one among them. Our food preferences are not only influenced by our bodies’ physiological or nutritional needs. There are also additional elements that affect eating preferences. The first is biological, such as hunger, appetite, and taste, which influence how people select various foods. The second one could be financial limitations like price, income, and availability that would affect the decision to buy. Variables like access, education, aptitude, and time could be the other factors. Hence, it is not only the flavor or physiological factor to predict consumer food preferences but it is a complex characterized by a number of elements including past information about the food, previous experiences, habit of consumption and attitude towards the food (Costell et al., 2010).

Culture in which individuals are brought up also attributes to their food forms preferences (Ludy & Mattes, 2012). Given our belief systems and the potential influence of culture, family, peers, and eating patterns, social variables may have a significant impact on the choice of food shape. Daily mood, stress, attitudes, beliefs, food knowledge, and other psychological factors may influence the meals chosen. A person's or a group of people's abilities to influence their food choices will differ depending on their age and stage of life. Therefore, not all demographic groups will respond well to a certain form of intervention to change food choice behavior. Instead, interventions must be tailored to various demographic segments, taking into account the many variables impacting each group's food preferences.

2.3 Food Forms and Total Calorie Intake

Research indicates that consumers with high calorie intake are found to have derived a higher energy percent from processed foods (Teo et al., 2022). The popularity and readily availability of processed food make their consumption higher. Additionally, food intake in the morning is quite satiating and can lower the total quantity consumed during the day, whereas food intake in the late evening misses satiating qualities and can increase daily intake (Castro, 2004). This shows the effect of time of consumption in the energy balance. Besides these, food form affects appetite and various physiological mechanisms necessary for maintaining energy balance (Dhillon et al., 2016). Recognizing the fundamental factors for people's food preferences may help policy makers while setting approaches to prevent obesity (Vabe & Hansen, 2014).

Our study has considered raw versus processed peanut and almond, and fresh versus juice apple and orange. There are numerous different ways to prepare peanuts, including roasted, salted, chocolate-covered, and as peanut butter. Varying kinds offer unique health advantages and different nutrient qualities. Peanuts have a healthy nutrient quality and high calorie content, making them most healthy when consumed in balance (Burgess, 2019). USDA (2018) also includes that peanuts are the great source of vitamin B where total amount of vitamin B1, B2, B3, B4, B6 and B9 are mentioned (in mg/100g) as 0.6, 0.1, 12.1, 52.5, 0.35 and 0.24 respectively. Total of 44.45 mg Tocopherols are also found per 100 gram of peanut (Silva et al., 2010). In this way, peanuts are the great source of macronutrients, minerals and vitamins (Hu et al., 2018). Raw peanuts include roasted as well as salted peanut in the whole raw form. Roasted salted peanut gives 170 Calories per serving (28 gram shelled). Roasted salted peanuts are added to the winter diets and considered a part of a good winter diet (NDTV, 2019).

Processed peanuts include peanut butter, chocolate made from peanut and many more which was not in the raw form. Peanut butter is a paste of peanut produced from crushed, powdered peanuts (Chang et al., 2013). A smooth style peanut butter contains 188 Calories per serving (2 tablespoon/32 grams). A peanut butter cookie (20 gram) contains 95 Calories (USDA, 2022). Per serving of a chocolate-peanut butter protein shake provides 402 calories, 26.1 g protein, 41.3 g carbohydrates and also other nutrients (Meyer, 2017). Peanut butter is also preferred with comparatively heavy foods; pasta and chicken (Bandurski, 2022). A serving (10 pieces) of chocolate coated peanut contains 208 Calories (USDA, 2022). According to USDA information, a cup (60 gram) gram serving of peanut flour provides 257 Calories. A peanut flour cookie provides 193 Calories per single serving (Thomson, 2010). Almond is a dense source of nutrients. The calorie content of almonds is 610 Kcal per 100 gram almond (Banerjee, 2020). Further, almonds consist of 21.6 g Carbohydrate, 21.2 g Protein and 49.9 g fat per 100 gram (Dube, 2022). Almond butter contains 96 Calories per serving (16g) (USDA, 2022). Almond cloud cookies consist of 36.2 Calories per serving. In case of Apple, medium sized apple contains about 100 Calories, 25-gram carbohydrates, 4 grams fiber and 10-gram sugar (Pathak, 2020). One cup of apple juice contains 117 calories, calories breaking down into 2% fat, 97% carbohydrates, and 1% protein (Fatsecret, 2021). Baked apple contains 191 Calories per serving (Bauer, 2021). A serving (1 piece; 125 gram) of apple pie consists 296 Calories (USDA, 2022). For the Oranges, 73 calories was found to be obtained from raw oranges and 8 ounces of orange juice has 110 calories, 2 grams of protein, 27 grams of carbs, and 0 grams of fat per serving (USDA, 2019).

2.4 Food Forms, Calorie Intake and Economics

Different articles have addressed the impact of demographic variables on the choice of the food people make for themselves. According to Drescher and Goddard (2008), lower education level has been related to lower quality diets. They looked upon the overall diet quality in Canada and found that higher income, age, female and higher education were positively associated with food diversity. The study by Pollard et al. (2002) investigates many food choice variables, underlines the intricate interplay of socio-demographic factors in shaping food choices, and underscores the significance of creating multimodal interventions to encourage healthy eating practices. Economics study has been interested in the connection between using food stamps, eating habits, and health effects. The study by Jilcott et al. (2011) examined the relationships between food stamp use and dietary practices, body mass index (BMI), and waist circumference among adult Americans by analyzing data from the National Health and Nutrition Examination Survey (NHANES) 2005–2010. In contrast to non-participants, the study indicated that food stamp recipients ate meals away from home less frequently. Even after correcting for demographic and socioeconomic characteristics, the study also discovered that food stamp participants had greater BMI and waist circumference than non-participants. Wansink and Chandon (2006) studied whether errors in predicting the calorie content of a meal were caused by the size of the meal or the individual's body size. Participants' body sizes ranged from normal to overweight, and meal portions were either little or high. According to the study, people might be underestimating how many calories are in their meals, especially if they are eating large meals. According to the study, overweight people may be more inclined to overestimate the number of calories in smaller meals, which may lead to overeating. In lean and obese young adults, Mourao et al. (2007) look into the effects of meal shape on hunger and calorie intake. The authors investigate the impact of consuming solid, liquid, or semi-solid foods on feelings of fullness and subsequent calorie consumption using information from the National Health and Nutrition Examination Survey (NHANES). According to earlier research, eating solid foods can result in better sensations of fullness and lower calorie consumption than eating liquid or semi-solid foods. The authors counter that this is because these studies have frequently concentrated on small sample sizes and have not taken into account how obesity may affect the association between meal type and appetite. They come to the conclusion that eating substantial meals might be a good approach to cut back on calories. In a study conducted by Wansink and Kim (2005), the impact of popcorn portion size on calorie consumption overall was investigated. Large or medium-sized popcorn buckets, filled with either fresh or stale popcorn, were distributed to participants. No matter how fresh the popcorn was, the results showed that individuals who received the larger buckets ingested more calories. In the study of Rolls et al. (2002), the impact of meal portion size on daily caloric intake was investigated. A variety of foods were offered to participants in either a large or small portion. According to the findings, those who received the larger amounts ate more calories than those who received the smaller quantities. The research by Rolls et al. (2007) examined the impact of big portion sizes on the overall number of calories consumed over the course of 11 days. A variety of foods were offered to participants in either a large or small portion. The results showed that throughout the course of the 11-day period, individuals who received the larger amounts ingested more calories. Mancino et al. (2009) studied the impact of food consumed away from home (FAFH) on diet quality using NHANES data. The authors conclude that FAFH has a detrimental influence on the quality of the diet as a whole, with the high calorie content of FAFH meals playing a significant role in this effect. The food and nutrient intake of people living in food-insecure families in the United States was investigated in the study by Ver Ploeg et al. (2009) using NHANES data. The researchers discovered that people from households with food insecurity consumed less of a number of critical nutrients, such as calcium, iron, and the vitamins A, C, and E. The study also discovered that people with food insecurity tended to favor nutrient-poor, high-energy diets. The effectiveness of eating outside the home on adult diet quality is investigated in the study by Todd et al. (2010) using NHANES data. Those who ate more meals away from home had poorer diet quality scores than adults who ate more meals at home, according to the study. The study also discovered that eating sit-down restaurant meals was linked to higher diet quality while consuming fast food was linked to lower diet quality.

One example of an economic study that used NHANES data and propensity score matching is a study by Basu et al. (2014) that examined the influence of the Supplemental Nutrition Assistance Program (SNAP) on medical results. They constructed a propensity score using NHANES data and took into account factors including participant income, age, and race/ethnicity, among others. The health outcomes of SNAP users and non-participants who shared comparable characteristics were then compared using propensity score matching. The findings revealed that SNAP use was linked to better self-reported health and a lower risk of obesity. A research conducted by Campbell et al. (2011) investigates how the National School

Lunch Program (NSLP) affects children's nutritional outcomes using information from the National Health and Nutrition Examination Survey (NHANES). They compare the food consumption of kids who took part in the NSLP with kids who did not, using a difference-in-differences technique together with propensity score matching. Their findings demonstrate that the NSLP has a beneficial effect on kids' intake of fruits and vegetables as well as the general caliber of their diet. More specifically, they discover that kids who take part in the NSLP consume 0.15 cups more fruits and vegetables daily and have a 4.5% higher Healthy Eating Index score than kids who do not. A study conducted by Carlson et al. (2014) sought to determine the association between overall food expense and diet quality. The Healthy Eating Index (HEI), the Nutrient Rich Foods Index (NRF), and the Alternative Healthy Eating Index were utilized by the researchers to measure the quality of the diet using NHANES data (AHEI). The study's findings indicated that there was only weak evidence for a relationship between total diet expense and diet quality. Diet cost and quality were correlated in some ways that were used to quantify diet quality, while in other cases, there was no correlation or a negative correlation.

3 METHODS

Two methods are used to answer the research questions in this thesis. First, a multinomial logit model is used to assess the impact of different factors on individual food form choices. Second, propensity score matching is used to estimate the impact of food forms on calorie consumption.

3.1 Multinomial Logit Model

The Multinomial Logit Model (MNL) is a well-known approach in economics for studying consumer behavior, as noted by Train (2009). This model can estimate the probability of selecting a particular option from a group of two or more alternatives using a set of explanatory variables. This technique has been widely utilized in empirical studies and can be used to analyze choice behavior across various domains.

According to Greene (2002), the MNL model equation is as follows:

$$Prob(Y_i = j|X) = \frac{e^{\beta'_j \mathbf{x_i}}}{\sum_{k=0}^{2} e^{\beta'_k \mathbf{x_i}}}, \quad j = 0,1,2 \text{ …………Equation (1)}$$

X are explanatory variables. In this study, these explanatory variables include gender, age, ethnicity, education, and aggregate activity per day. Y is a categorical dependent variable that represents different food forms. For all food product categories (e.g., peanuts), 0 is the category in which the consumer does not consume any of the food product. For the nut category, Y is 1 if the person consumed the raw form (i.e., almonds or peanuts) and 2 if the person consumed processed form (i.e., almond butter or peanut butter). In the case of fruits, Y is 1 is the person consumed the raw fruit and 2 if the person consumed the fruit in its juice form.

In order to examine many facets of consumer choice behavior, the MNL model has been widely employed in consumer economics research. The model, for instance, has been used to calculate the effects of changes in advertising, quality, and pricing on consumer demand for various goods and services (Ben-Akiva & Lerman, 1985). Moreover, customer preferences for various product features, such as brand, packaging, and nutritional value, have been estimated using the MNL model (Wansink & Sobal, 2007).

Multinomial logit models use coefficients to indicate the change in log odds of being in one category compared to a reference category, given a one-unit change in the predictor variable, while holding other variables constant (Greene, 2002). Although these coefficients help explain the relationship between predictor variables and outcomes, they are not always easy to interpret except for their sign. Marginal effects present results as changes in probabilities, which is more informative than odds ratios. They represent the change in probability of a particular category given a one-unit change in the predictor variable while holding other variables constant. Marginal effects are easier to interpret and important for predicting outcomes as they estimate changes in the outcome variable resulting from predictor variable changes (Perraillon, 2019). In our results, I have presented the average marginal effect of different socio-demographic variables on the consumption of raw and processed food forms as well as no consumption.

In our study, a multinomial logit regression, comparable to that of Arab et al. (2018), was employed to look at variables influencing the consumption of various meal forms. Three categories- raw, processed (juice for fruits), and not consumed, were used in the multinomial logit regression for the particular food forms. Finding the socio-demographic factors that influence the consumption of food in its various forms is made easy by the multinomial logit regression. I first evaluated the drivers that impact the food form consumption. I took age, gender, ethnicity, education, and aggregate activity per day, total calorie intake as the covariates that might impact the consumption of different food forms. I conducted four multinomial logit regression for peanut, almond, apple and orange. I use survey weights to make our result a reflection of the general U.S. population.

3.2 Propensity Score Matching

The analytical technique this study used is the propensity score matching (PSM) technique, a common method for estimating causal treatment effects. The empirical analysis first applies a logit model to obtain the propensity score to consumption of different food forms. Then, those scores were applied to match treated and control individuals that helps compare like-to-like individuals for all but their treatment. This gives us the significance of different food forms consumption on total calorie intake.

Propensity Score Matching (PSM) has gained popularity to measure the causal treatment effects in a broad range of academic fields of study (Caliendo & Kopeinig, 2008). It is a quasi-experimental approach where the investigator pairs each treated item with a non-treated item with identical features creating an artificial control group using statistical tools. This statistical technique compares a treatment item with one or more control items based on the propensity score of individual item. By minimizing selection bias, propensity score matching can improve causal assertions in quasi-experimental and observational experiments (Randolph et al., 2014). This technique enables researchers to match cases who have an identical chance of receiving the treatment after minimizing a high number of pretreatment factors into a single scalar expression (Adelson, 2013). Hence, propensity score matching is built on the concept that if a member of the intervention group is matched with a member of the comparison (control) group, both will have the same likelihood of falling into the intervention condition (i.e., the assumption is the same as in randomized group designs).

The likelihood that a unit in the merged sample of treated and untreated units gets the intervention, provided a set of observable factors, is known as the propensity score. The propensity score (or likelihood of involvement) will yield reliable matches for measuring the influence of a treatment if the researcher has access to all data pertinent to treatment and outputs. As a result, events may be compared just on the grounds of propensity scores instead of trying to match on all values of the factors (Heinrich et al., 2010).

To produce unbiased intervention outcomes by using propensity scores, two requirements must be met. The first requirement is that being allocated to a controlled or experimental treatment is not, on its own, a predictive feature. Additionally, all confounding factors must be known. Second, at the phase of choosing the intervention, there needs to be a genuine selection between experimental and control treatments for each item (Rosenbaum & Rubin, 1983). The propensity approach should ideally taken into account any covariates that may be confounding the intervention outcome. Three criteria must be met in order for anything to count as a confounder: initially, it has to be a covariate that was known before the treatment was assigned; secondly, it has to be capable of influencing the treatment choice; and finally, it has to be capable of influencing the case's outcome (Armitage & Colton, 2000).

In propensity score matching, after the propensity scores are calculated, they are matched between experimental and non-experimental groups. After matching by propensity score, the treatment effect should be examined to take into account the linked pairs. This analysis can take the form of a conditional logistic regression for binary outcomes, a stratified log-rank test, or a

Cox regression analysis segmented by paired combinations for time-to-event outcomes (Austin, 2008). The following is a selection of propensity score matching algorithms.

1. Nearest neighbor matching: According to their propensity scores, the nearest neighbor matching algorithm pairs each treated unit with the closest untreated unit. Due to its ease of use and effectiveness, this approach is extensively employed (Austin, 2011). Yet, if there are significant disparities between the propensity scores of the treated and control groups, it could result in an unbalanced study (Stuart, 2010).

2. Kernel-based matching: In this technique, treated and untreated units are matched in accordance with the distance between their propensity scores after the propensity score is estimated using a non-parametric kernel regression. Compared to nearest neighbor matching, this technique is more adaptable and can handle non-linear interactions between variables and the result (Stuart, 2010).

3. Radius matching: Within a given radius of their propensity scores, treated and untreated units are matched using the radius matching technique. When closest neighbor matching is unable to identify a good match, this method is more adaptable and can manage the situation (Ho et al., 2007).

4. Linear logit regression: With this technique, treated and untreated units are matched according to their projected propensity scores after computing the propensity score using a linear logistic regression model. It assumes that there is a linear relationship between covariates and the result, which may not always be the case (Austin, 2011).

In case of our study, I was solely interested in learning how calorie intake as a whole related to the consumption of particular food types. In order to analyze the impact of consuming various food types on the results of total calorie intake, it is therefore necessary to compare the treatment and control groups. The propensity score matching method was used. After the multinomial regression, as our main intention was to know the difference in total calorie intake, I used age, gender, ethnicity, education and aggregate activity per day as covariates which were the same variables taken in the multinomial logistic regression. Our outcome variable was total calorie intake. There were eight treatment variables, namely raw peanut, processed peanut, raw almond, processed almond, raw apple, apple juice, raw orange and orange juice. Control groups were different for different regression such as it was people not consuming these particular food forms for some regression and for other it was processed form of the food. Basically, I compared raw form consumption vs no consumption, processed form consumption vs no consumption and raw form consumption vs processed form consumption. The first step in propensity score matching is to acquire propensity scores using a logit model. As a result, I used a binary logit model to estimate the propensity scores for our research. Probability scores indicated how likely a person was to belong to a particular category. Furthermore, the findings of the propensity score matching are confined to the observations used in that model since it did not consider the survey weight, unlike the multinomial logit regression, which did. Therefore, the results of the propensity score matching cannot be generalized to the entire population.

Stata version 17 was used to estimate the multinomial logit model and psmatch2 propensity scores. Gender, age, ethnicity, education, and aggregate activity per day are explanatory variables. For various regressions, Y stands for the dependent treatment variable, which is either a raw peanut, a processed peanut, a raw almond, a processed almond, a raw apple, apple juice, or a raw orange and its juice.

One of our main concerns when completing the estimation was to ensure the validity of the propensity scores. I used pseudo-R2 as a comparative indication of model variance estimate.

After establishing the propensity score using the psmatch2 approach, matching based on the propensity scores was used to determine whether treatments caused substantial differences in our total calorie intake outcome. I applied a number of techniques to develop and evaluate the accuracy of our matching results. In order to ensure that the propensity scores from the control group fell between the minimum and maximum propensity scores from the treatment groups, the common support criterion was also implemented. Nearest neighbor (1) with replacement, Neighbor (1) without replacement, Kernel Matching, Radius Matching-caliper (0.1), Radius Matching-caliper (0.05), Radius Matching-caliper (0.01), and Local Linear Regression were the algorithms/methods under examination. After selecting the best matching approach, 500 replications were used to bootstrap the standard errors for significance testing, and the average treatment effects were calculated.

4 DATA

4.1 Design and Participants

The study uses data from the National Health and Nutrition Examination Survey (NHANES) dataset 2017-2018 that includes 24 hours dietary recall and follow-up data (CDC, 2022). Started in the early 1960s, NHANES program has been carried out as a set of studies that concentrate on various demographic groups or health-related issues in the U.S. It is a set of surveys to analyze the health and nutritional position of the citizens. The survey involves both interviews and physical testing, and is a key program of the National Center for Health Statistics (NCHS). NCHS is part of the Centers for Disease Control and Prevention (CDC), and is in charge of providing important health data for the entire country. One of the aspects of the NHANES interview is its dietary aspect. The United States populations' food, nutrients, food category, and dietary habits are described by NHANES dietary information to help develop and assess nutritional policies and initiatives. After taking into account daily fluctuation, average dietary consumption patterns may be determined. While preserving data quality and delivering timely data to assess the state of the nation's dietary pattern and health, NHANES stays open and adaptable to embrace changes. In conclusion, the NHANES gathers dietary information in the context of its wide, multifunctional purposes.

The study relied on dietary and demographic information obtained from the dataset. Nuts and fruits were intentionally chosen as the study's focal point due to their popularity in the U.S. and their prolonged shelf life, making them especially relevant to the research's focus on food form selection. Among the assortment of nuts, peanuts and almonds were specifically selected due to their prominence in the country. Likewise, apples and oranges were chosen as the fruits in the study, given their high rates of consumption. After the selection of different foods, I classified those foods into their different food forms. I used food codes provided by the Foods Patterns Equivalent Database to obtain the data regarding the different forms of peanut, almond, apple and orange (USDA, 2021b). The focus of this study was to investigate the factors that influence the selection of different food forms for a given product. To achieve this, I classified the commodities (peanut, almond, apple and orange) into food form categories. The nuts (peanut and almond) were categorized into two categories: raw and processed forms. The raw form included the whole form of the peanut and almond such as roasted peanut or peanut mix with salt, etc. that had not been transformed into other forms. The processed form included peanut or almond butter, chocolate made from these commodities, peanut or almond flour, and other processed products that utilize peanut or almond. For fruits under study (apple and orange), I classified their consumption into raw and juice forms, using food codes.

The primary research question of this study was to investigate the factors that influence the consumption of different food forms. To accomplish this, I analyzed the relationship between food form consumption and various socio-demographic variables. Specifically, I examined the respondents' age, gender, and total calorie intake, which were collected based on the NHANES dataset and I used them as presented on the dataset. However, I recoded or modified the education and ethnicity variables to make them more suitable for this study. To compare the effects of moderate and vigorous physical activity, I calculated aggregate activity per day by multiplying the activity level with standard MET scores for different activities, as provided by the NHANES dataset.

The study used various independent variables such as socio-economic demographics, dietary habits, and physical activity to analyze the dependent variable, which was food form. Food form was classified into three categories: no consumption, raw, and processed. The comparison was made between raw and processed forms for nuts, and between raw and juice forms for fruits. These specific foods were chosen due to their high consumption and popularity in the U.S. As our dependent variable is categorical and unordered, I used multinomial logistic regression four times; one for each commodity, namely peanut, almond, apple, and orange. Multinomial logit regression uses survey weights to weight the sample, ensuring that the survey data appropriately reflects the population from which it was collected.

Our study aimed to determine if there was a discrepancy in the total calorie intake between individuals who consumed a particular type of food and those who did not. I divided participants into two groups: a treatment group that consumed raw or processed peanuts, almonds, apples, apple juice, oranges, or orange juice, and a control group that did not consume or consumed a different form of the same food. Treatment groups must be precisely specified if biases are to be reduced and eliminated (Campbell et al., 2011). As a result, our treatment group consisted of those who consumed raw peanuts, processed peanuts, raw almonds, processed almonds, raw apples, apple juice, raw oranges, and orange juice.

In all of the regression analyses conducted for each food form, the outcome variable and covariates remain the same. The only difference between each regression is the treatment and control variable used. For example, in the case of peanuts, three different propensity matching analyses were performed. The first analysis compared the consumption of raw peanuts as the treatment variable with those who did not consume any peanuts as the control group. The second analysis used processed peanuts as the treatment variable and people who did not consume any peanuts as the control group. The third analysis compared raw peanuts as the treatment variable and processed peanuts as the control variable. Similar analyses were conducted for all the food forms included in the study. I analyzed the research's findings to compare the total calorie consumption of the treatment and control groups.

4.2 Covariates

Some of the variables included in this study were taken exactly the same way as given in the NHANES dataset. However, some other variables were recoded or modified as per the ease for this study. As an objective of this study, different socio-demographic variables were studied in relation to the food forms consumption. For this, food form was taken as dependent variable (0= no consumption, 1= raw, 2= processed) with the socio-economic demographics, dietary and physical activity as independent variables. The multinomial logistic regression was carried four times for four different commodities (peanut, almond, apple and orange). The independent variables used for this study are mentioned in the table below.

Table 1 Variables Used in the Analyses

Variable name	Description
Age	Age of the respondents in years
Total calorie intake	Total calorie consumption by the respondent in Kcal over two days of dietary recall
Aggregate activity MET score	Time spent doing some activity in a day (except sedentary activities). Sedentary activities includes time spent sitting on a day; the sitting may be at home, at school, coming from and going to somewhere in a bus or car, sitting at chairs, reading, watching TV, playing cards and working on a computer (does not include sleeping). The aggregate activity include

	vigorous-intensity activities at work on a typical day (vigorous-intensity activity causes large increases in breathing or heart rate and is done for at least 10 minutes continuously), moderate intensity work (causes small increases in breathing or heart rate and is done for at least 10 minutes continuously), transportation by walk or bicycle, vigorous recreational activities (leisure time activities such as sports) and moderate recreational activities. • MET score: As a variable in our study, the aggregate activity per day was multiplied by MET score. The NHANES dataset uses the Metabolic Equivalent of Task (MET) scores to quantify the level of physical activity. This makes it possible for researchers to assess the relative intensities of various activities and calculate the overall energy consumption for a specific time frame. The suggested MET scores for physical activity range from 4.0 for moderate work, leisure, and transportation activities to 8.0 for vigorous work and leisure activities.
Female	Gender of the respondent: 1 = Male and 2= Female
Ethnicity	Race and Hispanic origin information of the participant (recoded to make three categories: 1= Non-Hispanic White, 2= Mexican American or Other Hispanic, 3= Non-Hispanic Black or Other race)
Education	Highest grade or level of school the participant has completed (recoded to form two categories: 1= high school degree or less, 2= some college or college degree or higher)

4.3 Summary Statistics

4.3.1 Socio-Demographic Characteristics of the Participants

Table 2 shows the socio-demographic characteristics of the participants for the overall observation of 4,734. The observation represents a total population of 238,638,713. The table reveals that mean age of the participants was 48.36 years with standard deviation of 0.57. Similarly, the mean aggregate activity met score was 1075.41 with standard deviation of 39.53. The mean of total calorie intake over both dietary recall days for the given observations was 4155.35 Kcal with standard deviation of 44.77. Finally, besides these continuous variables, table 2 also presents the summary statistics of some categorical socio-demographic variables. The total number of observations for those variables was the same as the number of observations for continuous variables, i.e. 4734. Among the given number of observations, slightly less than half (47.74%) were male whereas the remaining were female. Similarly, Non-Hispanic White ethnic groups were 62.68 percent of the total participants under observation. Mexican American or Other Hispanic and Non-Hispanic Black or Other Race were found to be slightly lesser than (15.42%) and slightly higher than (21.89%) of one fifth of the total observation under study. Moreover, nearly two thirds (63.27%) of the participants under observation were some college or college degree or higher education holder, with the rest having the academic qualification of high school degree or lesser. The socio-demographic characteristics with respect to individual food forms are given in the appendices (Appendix 1-Appendix 12).

Table 2 Socio-Demographic Characteristics of the Participants

Variables	Mean/percentage	Standard deviation
Age (years)	48.358	0.567
Aggregate activity met score	1075.413	39.530

Total calorie intake	4155.351	44.771
Gender (%)		
Male	47.74	
Female	52.26	
Ethnicity (%)		
Non-Hispanic White	62.68	
Mexican American or Other Hispanic	15.42	
Non-Hispanic Black or Other Race	21.89	
Education (%)		
High School Degree or Less	36.73	
Some College or College Degree or Higher	63.27	

4.3.2 Product Consumption

Table 3 presents the percentages of different food forms of peanut, almond, apple and orange based on their respective observations. For peanut, the number of observations was 4,668, among which 5.16 percent consumed raw peanut, 13.31 percent consumed processed peanut and 75.53 percent did not consume peanut at all. The number of observations for almond was 4,707. Among this, 89.41 percent did not consume almond, 4.69 percent consumed raw almond and 5.9 percent consumed processed almond. Similarly, moving to the fruits, the total observation under study for apple was 4,691. Among the observations, 84.51 percent did not consume apple, 12.12 consumed raw apple and 3.37 consumed apple juice. Finally, among 4,678 observations for orange, 5.24 percent consumed raw orange and 13.14 percent consumed orange juice whereas 81.62 percent did not consume orange at all.

Table 3 Food Form Consumption Comparison Across Products

Product	No. of observations	Raw (%)	Processed (%)	None (%)
Peanut	4668	5.16	19.31	75.53
Almond	4707	4.69	5.90	89.41
Apple	4691	12.12	3.37	84.51
Orange	4678	5.24	13.14	81.62

5 RESULTS AND DISCUSSION

5.1 Factors Affecting the Food Form Consumption- Multinomial Logit

Table 4 delineates the factors affecting the peanut food form consumption in the U.S. based on NHANES dataset. The total number of observations for the factors affecting peanut food form consumption is 4,673, which represents the population of 234,414,910. The F-test p-value of 0.056 infers that the independent variables of the applied model reliably predict the dependent variable at the 10% significance level. For raw peanut, only age of the participant is found significant. The result shows that the probability of raw peanut consumption increases by 0.1 percentage points as the age of the participant increases by one year. The result is significant at 5% level of significance. He et al. (2005) determines age to be a significant factor for raw peanut (salted peanut, honey roasted peanut) consumption. However, the association is negatively linked in this case which is positive in our case, but this association is not significant for all forms.

For processed peanut, ethnicity of the participant is found significantly associated at 1% level of significance. The probability of processed peanut consumption is 12.4 percentage points lower when the participants are Mexican American or other Hispanic races compared to Non-Hispanic White respondents. For Non-Hispanic Black or other races, the probability of processed peanut consumption is 9.7 percentage points lower as compared to Non-Hispanic Whites. Except for ethnicity, other variables under consideration for the study are not significantly associated with the processed peanut consumption in our study. He et al. (2004) conducted a study on the factors affecting peanut butter consumption in the U.S. based on school students. They attribute the consumption of peanut butter, a highly popular peanut food form in America, to school level and their place of residence. In an another study by Jolly et al. (2005), age, race/ethnicity, convenience, information on fat content and income are some characteristics affecting the processed peanut (peanut butter) consumption. Hence, I can conclude that the factors of processed peanut consumption differ based on the place and context.

The ethnicity of the participant is also associated with no peanut consumption at 1% level of significance. The participant is more likely to be a non-consumer of peanut if s/he is Mexican American, other Hispanic, Non-Hispanic Black or other race as compared to Non-Hispanic White. For the participants who are Mexican American or other Hispanic, the probability of not consuming peanut is approximately 13 percentage points higher. On the other hand, if the participants are Non-Hispanic Black or other race, the probability of not consuming peanut is found to be greater by 8.4 percentage points. Despite not being socio-demographic factor, He et al. (2004) suggests that peanut butter allergy may act a role to limit peanut butter (processed peanut) consumption. The overall result under factors influencing peanut food form consumption indicates its dependence based on time, place and context.

Table 4 Marginal Effects from Multinomial Logistic Regression, Peanuts

Variable	Raw peanut	Processed peanut	None
Female	0.004	-0.025	0.021
S.E.	0.007	0.019	0.021
Age	0.001**	-0.0002	-0.0009
S.E.	0.0004	0.0007	0.0008
Ethnicity (Mexican American or Other Hispanic)	-0.006	-0.124***	0.130***

S.E.	0.013	0.029	0.030
Ethnicity (Non-Hispanic Black or Other Race)	0.013	-0.097***	0.084***
S.E.	0.017	0.020	0.022
Education	-0.013	0.006	0.007
S.E.	0.015	0.021	0.026
Aggregate activity	0.000002	-0.00001	0.000008
S.E.	0.000004	0.000007	0.000007
Intercept	-3.783***	-0.993**	

Number of observation: 4,673

F-test p-value: 0.056

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

Table 5 shows the factors that affect the food form consumption of almonds, which includes the factors associated with both raw and processed forms of almond, together with no almond consumption. The total number of observations for this analysis is 4,707. The F-test p-value of 0.015 infers that the independent variables of the applied model reliably predict the dependent variable at the 5% significance level. From the table, it can be seen that gender, age and education of the participants under observation have significant effects on raw almond consumption. These characteristics are significantly associated with raw almond consumption at the 1% level of significance. Looking at the result, the probability of consuming raw almonds is found to be 3.9 percentage points higher for female. Similarly, with increase in one year of age, the probability of consuming raw almond is found to increase by 0.08 percentage points. Moving ahead, education level of the participant is also significantly associated with the raw almond consumption. The probability of the participants being raw almond consumer is 3.9 percentage points higher when their education level is some college or college degree or higher, as compared to high school degree or less education. Similar to our result, Brown et al. (2014) also found the positive association of education level with the whole nuts (including almond) consumption based on their study in New Zealand. As almond is found to have several health advantages (Chen et al., 2006; Kamil & Chen, 2012; Lapsley & Huang, 2004; Mushtaq et al., 2015), those with higher levels of education may be more likely to consume it because they may be more likely to be aware of the health benefits of almond consumption.

Processed almond consumption is significantly associated with two variables (non-Hispanic Black or other race and education level) at the 5% level of significance both. The probability of participants to be processed almond consumer is found to be 2.2 percentage points lower if they are of non-Hispanic Black or other race. Moreover, for the participants with education level of some college or college degree or more, the probability of consuming processed almond is 3.7 percentage points higher.

Finally, from the point of no almond consumption, gender and education are found significant at 1% level of significance both. However, both have the negative association with no almond consumption. Looking at the association with gender, it is found that presence of female character has 6 percentage points lower probability of not consuming almonds. Again, having some college or college degree or more is associated with lesser probability of not being almond consumer by 7.6 percentage points. This association of nut consumption with education level can be compared with the study by Neale et al. (2020), which shows higher nut consumption to be associated with high education and income. The study also attributes the nut allergy and misconceptions about nut consumption on health to act as barriers of nut consumption.

Table 5 Marginal Effects from Multinomial Logistic Regression, Almonds

Variable	Raw almond	Processed almond	None
Female	0.039***	0.022	-0.060***
S.E.	0.010	0.013	0.015
Age	0.0008***	-0.0003	-0.0006
S.E.	0.0002	0.0004	0.0004
Ethnicity (Mexican American or Other Hispanic)	0.008	0.006	-0.014
S.E.	0.020	0.010	0.022
Ethnicity (Non-Hispanic Black or Other Race)	-0.002	-0.022**	0.024
S.E.	0.013	0.009	0.015
Education	0.039***	0.037**	-0.076***
S.E.	0.013	0.014	0.021
Aggregate activity	-0.000002	-0.0000003	0.000002
S.E.	0.000004	0.000004	0.000006
Intercept	-5.277***	-3.289***	
Number of observation: 4,707			
F-test p-value: 0.015			

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

Table 6 presents the factors affecting the apple food form consumption. The total number of observations and F-test p-value are 4,692 and 0.041 respectively. Among the variables hypothesized to affect the apple food form consumption, Mexican American or other Hispanic races and aggregate activity are found to be significant and positively associated with the raw apple consumption. The finding shows that there is more probability by 8.3 percentage points for the participants to consume raw apple if their ethnicity is Mexican American or other Hispanic as compared to non-Hispanic White, with significance at 1% level. Moreover, one additional unit of aggregate activity MET score is associated with increase in the probability of raw apple consumption by 0.0009 percentage points. This finding is significant at 10% level of significance.

For apple juice consumption, the significant variables are age, Mexican American or other Hispanic ethnicity, Non-Hispanic Black or other race and aggregate activity which are significant at 10%, 5%, 1% and 10% level of significance respectively. With one year increase in age, the probability of apple juice consumption is found to be decreased by 0.05 percentage points. In addition, the presence of Mexican American or other Hispanic ethnicity is associated with higher probability of apple juice consumption by 2.9 percentage points. The probability of participants to consume apple juice is 2.6 percentage points higher when they are Non-Hispanic Black or other race. Finally, an additional unit of aggregate activity MET score is associated with decrease in probability of consuming apple juice by the percentage point of 0.0006.

The table also reveals the factors that affect the no consumption of apple by the participants under observation. Based on the result, the probability of not consuming apple is 2.2 percentage points lower for female, with the significance at 10% level of significance. Similarly, if the ethnic group of the participants is Mexican American or other Hispanic, the probability of not consuming apple is 11.2 percentage points lower as compared to Non-Hispanic White. The result is significant at 1% level of significance. Lastly, at the 5% level of significance, it is found that there is 4.3 percentage points lower probability of not consuming apple for the participants who are Non-Hispanic Black or other race.

Table 6 Marginal Effects from Multinomial Logistic Regression, Apples

Variable	Raw apple	Apple juice	None
Female	0.017	0.005	-0.022*
S.E.	0.017	0.011	0.013
Age	0.00003	-0.0005*	0.0005
S.E.	0.0005	0.0003	0.0005
Ethnicity (Mexican American or Other Hispanic)	0.083***	0.029**	-0.112***
S.E.	0.024	0.010	0.024
Ethnicity (Non-Hispanic Black or Other Race)	0.017	0.026***	-0.043**
S.E.	0.017	0.008	0.019
Education	0.018	0.005	-0.023
S.E.	0.019	0.007	0.020
Aggregate activity	0.000009*	-0.000006*	-0.000004
S.E.	0.000005	0.000003	0.000005
Intercept	-2.401***	-2.932***	

Number of observation: 4,692

F- test p-value: 0.041

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

Table 7 shows the factors affecting the food forms consumption of orange. The total number of observations for this analysis is 4,682. The F-test p-value is 0.029. Starting with the raw orange, its consumption is found to be significantly associated with the gender, Mexican American or other Hispanic and Non-Hispanic Black or other race at 10%, 5% and 1% level of significance respectively. Presence of female character is associated with the probability of consuming raw orange by 1.4 percentage points. Moreover, presence of Mexican American or other Hispanic ethnicity is associated with 4.7 percentage points greater probability of consuming raw orange. In addition, when the ethnicity is Non-Hispanic Black or other race, the probability of consuming raw apple is 4.9 percentage points lower.

The table 7 also reveals the factors affecting the consumption of orange juice, giving only significant variable, i.e. age. An increase in one year of age is found to increase the probability of orange juice consumption by 0.2 percentage points and the result is significant at 1% level of significance.

The factors associated with no consumption of orange are also studied. The result reveals that the probability of not consuming orange is 3.6 percentage points higher for female participants. The finding is significant at 5% level of significance. Similarly, with increase in one year of age, the probability of not consuming orange is found to be decreased by 0.2 percentage points. The association is found significant at 1% level of significance. Presence of both ethnic groups under study is negatively associated with the no consumption of orange at 5% level of significance each. For Mexican American or other Hispanic ethnic group, their presence is associated with 7.7 percentage points lower probability of not consuming orange, as compared to Non-Hispanic White. Similarly, for the Non-Hispanic Black or other ethnic group, the probability of not consuming orange is found to be 6.1 percentage points lower compared to Non-Hispanic Whites.

Some research has studied the factors affecting fruit consumption in general were identified. For instance, household income and education level of the household head is found to be associated with higher fruit consumption in a Pakistan based study conducted by Amjad and Akbar (2023). A study in an educational setting in Nigeria concludes that the level of education of parents, marital status, nature of study based on time, and the availability of fruits are the influencing factors for fruit consumption (Obayelu et al., 2019).

Table 7 Marginal Effects from Multinomial Logistic Regression, Oranges

Variable	Raw orange	Orange juice	None
Female	-0.014*	-0.022	0.036**
S.E.	0.007	0.015	0.016
Age	0.0006	0.002***	-0.002***
S.E.	0.0004	0.0005	0.0005
Ethnicity (Mexican American or Other Hispanic)	0.047**	0.030	-0.077**
S.E.	0.017	0.029	0.030
Ethnicity (Non-Hispanic Black or Other Race)	0.049***	0.012	-0.061**
S.E.	0.015	0.022	0.026
Education	-0.014	0.008	0.006
S.E.	0.014	0.021	0.022
Aggregate activity	-0.00000004	-0.000002	0.000002

S.E.	0.000005	0.000008	0.000008
Intercept	-3.605***	-2.637***	

Number of observation: 4,682

F-test p-value: 0.0290

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

5.2 Food Form Impact on Calorie Consumption- Propensity Score Matching

Table 8 delineates the covariate balancing statistics for selecting the matching algorithm for peanut food form consumption impacts. The algorithms/methods under consideration are nearest neighbor (1) with replacement, neighbor (1) without replacement, kernel matching, radius matching- caliper (0.1), radius matching- caliper (0.05), radius matching- caliper (0.01) and local linear regression. The table presents three comparisons having each of these matching algorithms. The first comparison is between raw versus no peanut consumption whereas the second is between processed versus no peanut. Finally, I compare raw peanut with the processed peanut. For all the comparisons, pseudo-R2 value, along with bias reduction, R2 reduction and Chi2 reduction are calculated. The pseudo-R2 value, which estimates relative model variance measure, is 3.12, 2.30 and 5.77 for raw vs no peanut, processed vs no peanut, and raw vs processed peanut comparison groups respectively. It can be seen from the table that the biasness is reduced after matching for all the three comparisons except for neighbor (1) without replacement. Among the algorithms utilized, the radius matching with caliper 0.01 is chosen for raw versus no and processed versus no peanut, because it is found to reduce mean standardized bias, R-squared and Chi-squared by highest percentage when compared the after-matching condition with the before-matching condition. For the raw versus processed peanut, kernel matching, radius matching with caliper 0.05, and radius matching with caliper 0.01 have comparable percentages for mean standardized bias reduction. Hence, I select radius matching with caliper 0.01 to maintain uniformity throughout the comparisons under peanut.

Table 8 Covariate Balancing Statistics for Peanuts

Matching algorithm	Pseudo-R2	Bias reduction	R2 reduction	Chi2 reduction
Comparison: raw peanut vs no peanut				
Neighbor (1) with replacement	3.12	-64.92	-73.33	-90.17
Neighbor (1) no replacement	3.12	359.16	2480	-17.46
Kernel matching	3.12	-59.68	-76.67	-91.45
Radius matching- caliper (0.1)	3.12	-14.14	6.67	-61.6
Radius matching- caliper (0.05)	3.12	-56.54	-73.33	-90.06
Radius matching- caliper (0.01)	3.12	-95.29	-100	-99.89
Local linear regression	3.12	-64.92	-73.33	-90.17
Comparison: processed peanut vs no peanut				
Neighbor (1) with replacement	2.30	-62.93	-86.96	-93.42
Neighbor (1) no replacement	2.30	489.66	1408.7	67.43
Kernel matching	2.30	-75.0	-95.65	-97.29
Radius matching- caliper (0.1)	2.30	-44.83	-60.87	-79.21
Radius matching- caliper (0.05)	2.30	-67.24	-91.3	-96.26
Radius matching- caliper (0.01)	2.30	-92.24	-100	-99.72
Local linear regression	2.30	-62.93	-86.96	-93.42

Comparison: raw peanut vs processed peanut				
Neighbor (1) with replacement	5.77	-76.67	-96.49	-97.84
Neighbor (1) no replacement	5.77	208.67	163.16	-71.40
Kernel matching	5.77	-91.33	-100	-99.63
Radius matching- caliper (0.1)	5.77	-73.33	-94.74	-96.91
Radius matching- caliper (0.05)	5.77	-90.67	-100	-99.58
Radius matching- caliper (0.01)	5.77	-90.67	-100	-99.68
Local linear regression	5.77	-76.67	-96.49	-97.84

Table 9 shows the covariate balancing statistics for almond utilized in selecting the matching algorithm for food form comparison impacts. The comparisons in case of almond are raw almond versus no almond, processed versus no almond, and raw almond versus processed almond. The result shows that pseudo-R2 values for comparison between raw versus no almond, processed versus no almond, and raw versus processed almond are 2.35, 3.34 and 0.97 respectively. For raw versus no almond and raw versus processed almond, the greatest reduction in bias, R2 and Chi2 are found in radius matching with caliper 0.01. Being the most suitable method at reducing the mean standardized bias, the radius matching with caliper 0.01 is selected as the matching algorithm in this comparison. In case of comparison between processed versus no almond, neighbor (1) with replacement, radius matching with caliper 0.01, and local linear regression are found to have reduced the mean standardized bias by comparable percentages. Hence, to maintain uniformity within the almond food form comparisons, radius matching with caliper 0.01 is selected.

Table 9 Covariate Balancing Statistics for Almonds

Matching algorithm	Pseudo-R2	Bias reduction	R2 reduction	Chi2 reduction
Comparison: raw almond vs no almond				
Neighbor (1) with replacement	2.35	-71.62	-91.67	-97.35
Neighbor (1) no replacement	2.35			
Kernel matching	2.35	-40.54	-25.0	-74.43
Radius matching- caliper (0.1)	2.35	-35.14	-4.17	-67.66
Radius matching- caliper (0.05)	2.35	-37.84	-12.5	-71.31
Radius matching- caliper (0.01)	2.35	-89.86	-95.83	-99.26
Local linear regression	2.35	-71.62	-91.67	-97.35
Comparison: processed almond vs no almond				
Neighbor (1) with replacement	3.34	-84.35	-97.06	-99.27
Neighbor (1) no replacement	3.34	557.82	267.65	-98.27
Kernel matching	3.34	-42.18	-41.18	-79.6
Radius matching- caliper (0.1)	3.34	-11.56	38.24	-52.34
Radius matching- caliper (0.05)	3.34	-37.41	-32.35	-76.41
Radius matching- caliper (0.01)	3.34	-74.15	-91.18	-97.26
Local linear regression	3.34	-84.35	-97.06	-99.27
Comparison: raw almond vs processed almond				

Neighbor (1) with replacement	0.97	-10.47	-50.0	-48.66
Neighbor (1) no replacement	0.97	31.4	100	79.96
Kernel matching	0.97	-77.91	-90.0	-93.38
Radius matching- caliper (0.1)	0.97	-46.51	-60.0	-64.94
Radius matching- caliper (0.05)	0.97	-76.74	-90.0	-92.13
Radius matching- caliper (0.01)	0.97	-84.88	-100	-97.14
Local linear regression	0.97	-10.47	-50.0	-48.66

Table 10 presents the covariate balancing statistics used in selecting the matching method/algorithm in case of apple food form consumption impacts. This includes three comparisons: raw apple with no apple consumption, apple juice with no apple consumption, and raw apple with apple juice consumption. The highest bias reduction is accounted for radius matching with caliper 0.01 with 94.34 percentage reduction in bias in matched condition as compared to unmatched one. Being best at reducing mean standardized bias, R squared, and Chi-squared value, this algorithm is selected for the further analysis. Additionally, it is evident from the table that both the R2 and Chi2 reduction are greatest in radius matching with caliper 0.01. Moving to the next comparison i.e. apple juice versus no apple, the pseudo-R2 value is 2.23. Radius matching- caliper (0.01) provides the best balancing, with the reduction by 83.78 percent, 95.45 percent and 98.9 percent for bias reduction, R2 reduction and Chi2 reduction respectively. As being the most suitable algorithm at reducing mean standardized bias, the radius matching with caliper 0.01 is selected for further study. The last comparison under apple is raw apple versus apple juice, for which the pseudo-R2 value is 3.16. The bias reduction percentages for kernel matching, radius matching with caliper 0.05 and radius matching with caliper 0.01 are almost similar. Hence, to maintain uniformity in the comparisons under apple food forms, radius matching with caliper 0.01 is selected for further analysis.

Table 10 Covariate Balancing Statistics for Apples

Matching algorithm	Pseudo-R2	Bias reduction	R2 reduction	Chi2 reduction
Comparison: raw apple vs no apple				
Neighbor (1) with replacement	2.06	-85.85	-95.24	-98.3
Neighbor (1) no replacement	2.06	496.23	2152.38	121.63
Kernel matching	2.06	-76.42	-95.24	-96.63
Radius matching- caliper (0.1)	2.06	-10.38	-28.57	-65.69
Radius matching- caliper (0.05)	2.06	-70.75	-90.48	-95.8
Radius matching- caliper (0.01)	2.06	-94.34	-100	-99.77
Local linear regression	2.06	-85.85	-95.24	-98.3
Comparison: apple juice vs no apple				
Neighbor (1) with replacement	2.23	-66.22	-77.27	-92.75
Neighbor (1) no replacement	2.23	662.76	4445.45	-74.49
Kernel matching	2.23	-15.54	22.73	-60.47
Radius matching- caliper (0.1)	2.23	-2.03	63.64	-46.91
Radius matching- caliper (0.05)	2.23	-8.78	50	-52.42
Radius matching- caliper (0.01)	2.23	-83.78	-95.45	-98.9
Local linear regression	2.23	-66.22	-77.27	-92.75
Comparison: raw apple vs apple				

juice				
Neighbor (1) with replacement	3.16	-67.21	-81.25	-61.35
Neighbor (1) no replacement	3.16	80.87	15.63	-60.67
Kernel matching	3.16	-86.34	-96.88	-95.54
Radius matching- caliper (0.1)	3.16	-63.39	-84.38	-64.0
Radius matching- caliper (0.05)	3.16	-83.61	-96.88	-93.61
Radius matching- caliper (0.01)	3.16	-84.15	-96.88	-94.61
Local linear regression	3.16	-67.21	-81.25	-61.35

Table 11 presents the covariate balancing statistics which is used in matching algorithm selection for orange food form consumption impacts. According to this table, three comparisons are done for orange; the first, raw orange with no orange, the second, orange juice with no orange, and the third, raw orange with orange juice. Starting with the first comparison, the pseudo-R2 value is 2.78 for all the algorithms. Radius matching with caliper 0.01 is the algorithm reducing the highest percentage of mean standardized bias, R-squared and Chi-squared value. Similarly, the second comparison of the table is between orange juice and no juice consumption. The pseudo-R2 value is noted to be 1.05 which represents the relative measure of estimation of relative variance. As in most of the other comparisons, the best algorithm for further study is identified to be radius matching caliper 0.01 as it is the best at reducing the value of mean standardized bias, R-squared, and Chi-squared value. Finally, the pseudo-R2 value for the comparison between raw orange and orange juice is 2.50 for all the algorithms under consideration. The greatest reduction in bias, R2 and Chi2 values is accounted for radius matching with caliper 0.01, with the percentage reduction by 76.77, 96 and 98.32 respectively.

Being the best algorithm at reducing mean standardized bias, the radius matching with caliper 0.01 is selected for further analysis.

Table 11 Covariate Balancing Statistics for Oranges

Matching algorithm	Pseudo-R2	Bias reduction	R2 reduction	Chi2 reduction
Comparison: raw orange vs no orange				
Neighbor (1) with replacement	2.78	-62.83	-92.86	-96.99
Neighbor (1) no replacement	2.78	795.58	1925	-60.44
Kernel matching	2.78	-41.59	-71.43	-88.47
Radius matching- caliper (0.1)	2.78	-2.66	42.86	-43.22
Radius matching- caliper (0.05)	2.78	-35.40	-64.29	-85.92
Radius matching- caliper (0.01)	2.78	-76.11	-96.43	-98.78
Local linear regression	2.78	-62.83	-92.86	-96.99
Comparison: orange juice vs no orange				
Neighbor (1) with replacement	1.05	-24.66	-63.64	-84.34
Neighbor (1) no replacement	1.05	575.34	4590.91	390.36
Kernel matching	1.05	-53.42	-72.73	-86.97
Radius matching- caliper (0.1)	1.05	-2.74	18.18	-44.08
Radius matching- caliper (0.5)	1.05	-47.95	-63.64	-83.15
Radius matching- caliper (0.01)	1.05	-91.78	-100	-99.83
Local linear regression	1.05	-24.66	-63.64	-84.34

Comparison: raw orange vs orange juice				
Neighbor (1) with replacement	2.50	-71.72	-92.0	-94.54
Neighbor (1) no replacement	2.50	193.94	672	244.07
Kernel matching	2.50	-53.53	-92.0	-92.93
Radius matching- caliper (0.1)	2.50	-46.46	-84.0	-88.97
Radius matching- caliper (0.05)	2.50	-55.56	-92.0	-93.20
Radius matching- caliper (0.01)	2.50	-76.77	-96.0	-98.32
Local linear regression	2.50	-71.72	-92.0	-94.54

5.3 Average Treatment Effect (ATE) Estimation of Food Forms on Total Calorie Intake

Table 12 presents the average treatment effect of different food forms on total calorie intake. The comparison for the analysis are done in all the food types between three forms; raw versus no consumption, processed (juice in case of fruits) versus no consumption, and raw versus processed food form consumption. The analysis is performed after selecting the best propensity score matching algorithm for respective treatments and comparisons. Bootstrapped standard errors are reported using 500 replications. The table includes the coefficient values, standard error and significance for different food form treatments (giving several comparisons).

For peanut, three comparisons are made. First, the effect of raw peanut on calorie intake as compared to no peanut consumption is studied. The result suggests that raw peanut consumers consume 585.63 Kcal more energy in average compared to no peanut consumers. This result is significant at 1% level of significance as indicated by the p-value. Johnston et al. (2013) conducted a research based on U.S. to find out the impact of peanut versus grain bar consumption on various aspects including energy intake and weight loss. The foods were used in the meal as preloads. The findings of the research suggest that higher calorie intake is taken in case of peanut consumption as compared to the grain bar. Not only that, grain bar is associated with higher body weight loss as compared to peanut consumption. Peanut consumption resulting higher energy intake is in line with our result. However, a research by Barbour et al. (2014) concludes that consuming peanuts (Hi-oleic or regular) results in low calorie intake as compared to potato crisps, with the same perceived satisfaction from the foods.

The second comparison specifies the treatment as processed peanut and control as no peanut consumption. The results indicate that the consumption of processed peanut as compared to no consumption increases calorie intake by 507.58 Kcal in average. The result is significant at 1% level of significance. Reis et al. (2011) in their research based on Brazil presents their finding with one of the components being the difference in total calorie intake for peanut food form as treatment and usual meal without peanut as control. The research includes both the food forms of peanut which I have considered in my study: raw (raw peanut with skin and roasted peanut without skin in their case) and processed (ground roasted peanut without skin in their case), which are compared to the control group. What they find from the research is no significant difference on total calorie intake in relation to different peanut treatment and control groups.

Finally, the third comparison (i.e., raw peanut versus processed peanut) finds no significant calorie intake difference between consuming raw versus processed peanut food forms.

Table 12 also includes the average treatment effect of almond food form consumption on calorie intake. Similar to peanut, the almond also consists of three comparisons. Firstly, on raw versus no almond consumption, the finding suggests that consumption of raw almond is associated with increase in calorie intake with an average of 454.01 Kcal as compared to no almond consumption. The result is significant at 1% level of significance. While studying based on the roasted almond treatment comparing with cookies as control, Jung et al. (2018) finds that the total calorie intake is significantly increased with the consumption of almond. This result is in line with our finding. The almond consumption increases the calorie from the fat whereas reduces calorie from the carbohydrate. However, Hull et al. (2015) finds no significant difference in calorie intake between the whole, raw almond consumption condition and their normal diet condition in their study on healthy women. Moreover, Hollingworth et al. (2019), in their study on raw almond considering twenty-four hour calorie intake finds that the difference in raw almond consumption versus no energy control situation is not significant with the energy intake. It means that the consumption of raw almond is not different from the no energy control condition in relation to total calorie intake.

The results indicate that consuming processed almond consumption does not significantly affect calorie intake as evident from the average treatment effect analysis. The third comparison (i.e. raw almond versus processed almond) is also not significantly associated with the difference in total calorie intake.

Two treatments of apple; raw apple and apple juice, are compared with no apple consumption whereas the third comparison is between raw apple and apple juice. In case of raw apple versus no apple consumption effect on calorie intake difference, the association is positive; however not significant. However, several other articles have stated the association of raw apple with energy intake in the opposite direction. For instance, Flood-Obbagy and Rolls (2009) finds out that the consumption of whole fruit (apple taken for the research) is associated with lower energy intake in the subsequent meal as compared to the control condition. Moreover, De

Oliveira et al. (2008) states based on their research on women that the apple consumption results in the reduction of calorie intake.

On the other hand, the apple juice versus no apple consumption effect on calorie intake difference is positively associated and significant at 10% level of significance. The coefficient value indicates that apple juice consumers have higher calorie intake by 314.06 Kcal on average as compared to those not consuming apples.

The difference in calorie intake is not found significant between raw apple and apple juice consumption. However, studies have also shown that the raw apple consumption decreases calorie intake in the subsequent meal as compared to apple juice consumption (Flood-Obbagy & Rolls, 2009).

Table 12 also shows the average treatment effect of food forms of orange on calorie intake. The orange treatments, namely raw orange and orange juice are compared with no orange consumption. The third comparison is between raw orange and orange juice. The p-value for average treatment effect of raw orange consumption compared to no orange consumption on calorie intake indicates that the association is not statistically significant. Although generalizing as 'fruits' rather than orange, Guyenet (2019) in his review article suggests that the consumption of raw fruits is directed towards reducing the calorie intake, especially when the fruits are consumed before meal which doesn't match with our finding.

On the other hand, the p-value for the average treatment effect of orange juice consumption compared to no orange consumption on calorie intake indicates that the association is statistically significant at 1% level of significance. The coefficient value suggests that orange juice consumers have higher calorie intake by 602.93 kcal in average as compared to no orange consumers. Wang et al. (2012) conducted a research on impact of consuming orange juice on calorie intake in the U.S., based on the NHANES data 2003-2006. Interestingly, my research also uses the NHANES data from the year 2017-18. They also find that orange juice consumers have the higher calorie intake than non-consumers, which matches well with my finding. Moreover, DellaValle et al. (2005) conducted a research to find out whether consuming caloric or non-caloric beverage while having the lunch is associated with calorie intake. They keep orange juice as one of the caloric beverages for their study. Upon consuming the food with the caloric beverages, the calorie intake is found to increase as compared to no beverage consumption.

Lastly, the comparison of raw orange versus orange juice in relation to calorie intake demonstrates a negative and significant association. The coefficient value suggests that the consumption of raw orange is associated with lower calorie intake in average by 534.32 Kcal as compared to orange juice consumption. The result is significant at 1% level of significance.

Table 12 ATE of Food Forms on Total Calorie Intake (With Radius Matching- Caliper 0.01)

Food forms	Coefficient	Bootstrapped Standard Error	p-value
Raw vs no peanut	585.634	119.821	0.000
Processed vs no peanut	507.575	73.916	0.000
Raw vs processed peanut	56.737	121.143	0.640
Raw vs no almond	454.014	108.947	0.000
Processed vs no almond	153.218	109.095	0.160
Raw vs processed almond	197.009	145.138	0.175
Raw vs no apple	118.646	74.927	0.113
Apple juice vs no apple	314.057	162.119	0.053

Raw apple vs apple juice	-198.135	169.285	0.242
Raw vs no orange	79.627	105.623	0.451
Orange juice vs no orange	602.925	83.089	0.000
Raw orange vs orange juice	-534.315	118.978	0.000

6 CONCLUSION

The overall statistics of the survey suggests that more sample participants are female, Non-Hispanic White and with the education more than just a high school graduate. However, splitting these participants into sub groups on the basis of food form consumption gives varying results. While no consumption of food has the highest proportion in all the commodities (peanut, almond, apple and orange), food form consumption between raw and processed varies by food type. Nuts are consumed more in processed form compared to raw. Fruits in raw and juice form vary in consumption according to fruit type; apples are consumed more in raw form than juice, whereas orange juice is consumed more as compared to raw orange.

Socio-demographic variables gender, age, ethnicity, education and physical activity play important role in the consumption of raw versus processed food forms. However, the same variables are not significant for all the food forms of different food types, and at least one variable is significant in each food form of each food type. Education level is not found to be significantly associated with different food forms' consumption across all food types, except almond. In a similar way, physical activity is found significant in apple only, among all the food types. The demographic variables' impact varies by and within the food types. Some variables are consistent within food type, while others are not. For instance, the impact of age and some ethnicities are consistent across nuts' consumption; however, age is inconsistent across apple and orange consumption. This suggests that the choice of food form may be product specific and not consistent across food types.

The calorie intake for raw nuts consumers is significantly higher as compared to those who do not consume nuts. However, no significant difference in calorie consumption is observed between consumers of raw fruits and no fruit consumers. Processed food forms consumption when compared to no consumption of the respective foods, raise calorie consumption for all the products, except for almond. For nut products, raw consumption is associated with higher calorie intakes compared to processed forms, but the increase is not significant. For the fruits, juice consumption is associated with higher calorie intakes compared to raw forms. The higher calorie intake is significant for orange but not apple. This suggests that not all processed food forms raise calorie intake. This may be tied to how much is consumed in each food form or how the food is consumed (e.g., juice versus processed nut).

This research provides insights on the impact of socio-demographic factors with the food form consumption. This helps agribusinesses of different food forms to understand and explore the effects of demographic factors on the products they may handle and produce. This will also assist them in identifying the potential consumers of their products. Additional research may look into eating occasion and other external factors that may impact food form consumption. The current research has helped identify who is eating different food forms. This further research can help tell agribusinesses of when, how, and why consumers are eating their products in different food forms.

Moreover, health and economic policy makers may derive the potential relationship between food forms' consumption and health outcomes, such as diabetes, obesity, cholesterol, etc based on the relationship of food forms with calorie intake. Considering this, the future research should be focused on other health implications of food forms, such as obesity. The extension of the research can also be made considering other factors for food form consumption, such as nutrition, day of the week and food characteristics. Additional research should consider other food product categories that may also impact health outcomes.

REFERENCES

Adelson, J. L. (2013). Educational research with real-world data: Reducing selection bias with propensity score analysis. *Practical Assessment, Research, and Evaluation*, *18*(1), 15.

Ajibade, A., & Saghaian, S. (2022). U.S. Almond Exports and Retaliatory Trade Tariffs. *Sustainability*, *14*(11), 6409. https://doi.org/10.3390/su14116409

Alhabeeb, H., Sohouli, M. H., Lari, A., Fatahi, S., Shidfar, F., Alomar, O., ... & Abu-Zaid, A. (2022). Impact of orange juice consumption on cardiovascular disease risk factors: a systematic review and meta-analysis of randomized-controlled trials. *Critical reviews in food science and nutrition*, *62*(12), 3389-3402. https://doi.org/10.1080/10408398.2020.1865263

Amjad, M., & Akbar, M. (2023). The Association between Fruit and Vegetable Intake and Socioeconomic Factors in the Households of Pakistan Using Quantile Regression Model. *Social Work in Public Health*, *38*(3), 248-258. https://doi.org/10.1080/19371918.2022.2134249

Arab, L., Dhaliwal, S. K., Martin, C. J., Larios, A. D., Jackson, N. J., & Elashoff, D. (2018). Association between walnut consumption and diabetes risk in NHANES. *Diabetes/Metabolism Research and Reviews*, *34*(7), e3031. https://doi.org/10.1002/dmrr.3031

Armitage, P., & Colton, T. (2000). *Encyclopedia of epidemiologic methods*. John Wiley & Sons.

Austin, P. C. (2008). A report card on propensity-score matching in the cardiology literature from 2004 to 2006: results of a systematic review. *Circulation: Cardiovascular Quality and Outcomes*, *1*, 62-67.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, *46*(3), 399-424. https://doi.org/10.1080/00273171.2011.568786

Bandurski, K. (2022). *13 Foods That Are Better with a Scoop of Peanut Butter*. Taste of Home, https://www.tasteofhome.com/collection/what-to-eat-with-peanut-butter/#:~:text=Peanut%20butter%20and%20fresh%20fruits,%2C%20celery%2C%20bananas%20and%20more.

Banerjee, S. (2020). Nutritional Benefits of Various Nuts and Dry Fruits: The Natural Energizers. *Agriculture & Food*, *2*(11), 747-50.

Barbour, J. A., Howe, P. R., Buckley, J. D., Wright, G. C., Bryan, J., & Coates, A. M. (2014). Lower energy intake following consumption of Hi-oleic and regular peanuts compared with iso-energetic consumption of potato crisps. *Appetite*, *82*, 124-130. https://doi.org/10.1016/j.appet.2014.07.015

Basu, S., Seligman, H. K., Gardner, C., & Bhattacharya, J. (2014). Ending SNAP subsidies for sugar-sweetened beverages could reduce obesity and type 2 diabetes. *Health Affairs*, *33*(6), 1032-1039. https://doi.org/10.1377/hlthaff.2013.1246

Bauer, E. (2021). *Baked Apples.* Simply Receipe. https://www.simplyrecipes.com/recipes/baked_apples/

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.

Brown, R. C., Tey, S. L., Gray, A. R., Chisholm, A., Smith, C., Fleming, E., ... & Parnell, W. (2014). Patterns and predictors of nut consumption: Results from the 2008/09 New Zealand Adult Nutrition Survey. *British journal of nutrition*, *112*(12), 2028-2040. https://doi.org/10.1017/S0007114514003158

Burgess, L. (2019). *What are nutritional benefits of peanuts?* Medical News Today. https://www.medicalnewstoday.com/articles/325003#protein

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31-72.

Campbell, B. L., Nayga Jr, R. M., Park, J. L., & Silva, A. (2011). Does the National School Lunch Program improve children's dietary outcomes?. *American Journal of Agricultural Economics*, *93*(4), 1099-1130. https://doi.org/10.1093/ajae/aar031

Cara, K. C., Beauchesne, A. R., Wallace, T. C., & Chung, M. (2022). Effects of 100% Orange Juice on Markers of Inflammation and Oxidation in Healthy and At-Risk Adult Populations: A Scoping Review, Systematic Review, and Meta-analysis. *Advances in Nutrition*, *13*(1), 116-137. https://doi.org/10.1093/advances/nmab101

Carlson, A., Dong, D., & Lino, M. (2014). Association between Total Diet Cost and Diet Quality Is Limited. *Journal of Agricultural and Resource Economics*, *39*(1), 47–68. https://www.jstor.org/stable/44131314

Castro, J. M. (2004). The Time of Day of Food Intake Influences Overall Intake in Humans. *American Society for Nutritional Sciences*, 104-111.

CDC. (2022). *National Center for Health and Statistics: National Health and Nutrition Examination Survey- NHANES 2017-2018.* Centers for Disease Control and Prevention. https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017

Chang, A. S., Sreedharan, A., & Schneider, K. R. (2013). Peanut and peanut products: A food safety perspective. *Food Control*, *32*(1), 296-303.

Chen, C. Y., Lapsley, K., & Blumberg, J. (2006). A nutrition and health perspective on almonds. *Journal of the Science of Food and Agriculture*, *86*(14), 2245-2250. https://doi.org/10.1002/jsfa.2659

Çiftçi, S., & Suna, G. (2022). Functional Components of Peanuts (Arachis Hypogaea L.) and Health Benefits: A Review. *Future Foods*, 100140. https://doi.org/10.1016/j.fufo.2022.100140

Costell, E., Tárrega, A., & Bayarri, S. (2010). Food acceptance: the role of consumer perception and attitudes. *Chemosensory perception*, *3*(1), 42-50. https://doi.org/10.1007/s12078-009-9057-1

De Oliveira, M. C., Sichieri, R., & Mozzer, R. V. (2008). A low-energy-dense diet adding fruit reduces weight and energy intake in women. *Appetite*, *51*(2), 291-295. https://doi.org/10.1016/j.appet.2008.03.001

DellaValle, D. M., Roe, L. S., & Rolls, B. J. (2005). Does the consumption of caloric and non-caloric beverages with a meal affect energy intake?. *Appetite*, *44*(2), 187-193. https://doi.org/10.1016/j.appet.2004.11.003

Dhillon, J., Running, C. A., Tucker, R. M., & Mattes, R. D. (2016). Effects of food form on appetite and energy balance. *Food Quality and Preference*, *48*, 368-375. https://doi.org/10.1016/j.foodqual.2015.03.009

Drescher, L. S., & Goddard, E. W. (2008). *Comparing predictors of diet quality in Canada over time under consideration of altering food guides* (Working paper, staff paper series 08-

01). University of Alberta, Department of Resource Economics and Environmental Sociology. http://dx.doi.org/10.22004/ag.econ.52091

Dube, P. (2022). *Almonds- Benefits, Nutritional Value and Recipes*. HealthifyMe. https://www.healthifyme.com/blog/almonds/

Fakhri, Y., Omar, S. S., Mehri, F., Hoseinvandtabar, S., & Mahmudiono, T. (2022). Global systematic review and meta-analysis on prevalence and concentration of aflatoxins in peanuts oil and probabilistic risk assessment. *Reviews on Environmental Health*. https://doi.org/10.1515/reveh-2022-0075

Fatsecret. (2021). *Food database and calorie counter*. https://www.fatsecret.com/calories-nutrition/generic/apple-juice

Flood-Obbagy, J. E., & Rolls, B. J. (2009). The effect of fruit in different forms on energy intake and satiety at a meal. *Appetite*, *52*(2), 416-422. https://doi.org/10.1016/j.appet.2008.12.001

Greene, W. H. (2002). *Econometric Analysis* (Fifth edition). Prentice Hall.

Guyenet, S. J. (2019). Impact of whole, fresh fruit consumption on energy intake and adiposity: a systematic review. *Frontiers in nutrition*, 66. https://doi.org/10.3389/fnut.2019.00066

Hashem, A. M., Cifuentes, L., Calderon, G., Ricardo-Silgado, M. L., Gonzalez-Izundegui, D., Campos, A., ... & Acosta, A. (2022). Effect of caloric intake and macronutrient composition on intestinal cholesterol absorption and bile acids in patients with obesity. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *323*(3), G157-G164. https://doi.org/10.1152/ajpgi.00108.2022

Heinrich, C., Maffioli, A., & Vazquez, G. (2010). A primer for applying propensity-score matching. *Inter-American Development Bank.*

He, S., Fletcher, S. M., Chinnan, M. S., & Shi, Z. (2004). Factors affecting school students' consumption of peanut butter sandwiches. *Journal of Food Distribution Research*, *35*(1), 112-117.

He, S., Fletcher, S. M., & Rimal, A. (2005). Snack peanut consumption: type preference and consumption manners. *Journal of Food Distribution Research*, *36*(1), 79-85.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199-236. https://doi.org/10.1093/pan/mpl013

Hollingworth, S., Dalton, M., Blundell, J. E., & Finlayson, G. (2019). Evaluation of the influence of raw almonds on appetite control: satiation, satiety, hedonics and consumer perceptions. *Nutrients*, *11*(9), 2030. https://doi.org/10.3390/nu11092030

Hu, H., Liu, H., Shi, A., Liu, L., Fauconnier, M. L., & Wang, Q. (2018). The effect of microwave pretreatment on micronutrient contents, oxidative stability and flavor quality of peanut oil. *Molecules*, *24*(1), 62.

Hull, S., Re, R., Chambers, L., Echaniz, A., & Wickham, M. S. (2015). A mid-morning snack of almonds generates satiety and appropriate adjustment of subsequent food intake in healthy women. *European journal of nutrition*, *54*, 803-810. https://doi.org/10.1007/s00394-014-0759-z

Jilcott, S. B., Liu, H., DuBose, K. D., Chen, S., & Kranz, S. (2011). Food stamp participation is associated with fewer meals away from home, yet higher body mass index and waist circumference in a nationally representative sample. *Journal of nutrition education and behavior*, *43*(2), 110-115. https://doi.org/10.1016/j.jneb.2010.06.001

Johnston, C. S., Trier, C. M., & Fleming, K. R. (2013). The effect of peanut and grain bar preloads on postmeal satiety, glycemia, and weight loss in healthy individuals: an acute and a chronic randomized intervention trial. *Nutrition journal*, *12*, 1-10. https://doi.org/10.1186/1475-2891-12-35

Jolly, C. M., Hinds, M. J., Lindo, P., Ligeon, C., & Weiss, H. (2005). Consumers acceptance of peanut butter and products containing peanut butter in two southeastern states in the U.S. *Journal of International Food & Agribusiness Marketing*, *17*(2), 87-105. https://doi.org/10.1300/J047v17n02_05

Jung, H., Chen, C. Y. O., Blumberg, J. B., & Kwak, H. K. (2018). The effect of almonds on vitamin E status and cardiovascular risk factors in Korean adults: a randomized clinical trial. *European journal of nutrition*, *57*, 2069-2079. https://doi.org/10.1007/s00394-017-1480-5

Kamil, A., & Chen, C. Y. O. (2012). Health benefits of almonds beyond cholesterol reduction. *Journal of agricultural and food chemistry*, *60*(27), 6694-6702. https://doi.org/10.1021/jf2044795

Lapsley, K. G., & Huang, G. (2004). Health benefits of almonds. *Cereal foods world*, *49*(1), 6.

Lawless, H. T. & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices*, Springer.

Ludy, M., & Mattes, R.D. (2012). Comparison of sensory, physiological, personality, and cultural attributes in regular spicy food users and non-users. *Appetite*, *58*(1), 19-27. https://doi.org/10.1016/j.appet.2011.09.018

Mancino, L., Todd, J., & Lin, B. H. (2009). Separating what we eat from where: measuring the effect of food away from home on diet quality. *Food Policy*, *34*(6), 557-562. https://doi.org/10.1016/j.foodpol.2009.09.003

Meike, R., Dean, D. L., & Baird, T. (2022). Understanding Apple Attribute Preferences of U.S. Consumers. *Foods*, *11*(2), 166. https://doi.org/10.3390/foods11020166

Meyer, H. (2017). *Chocolate-Peanut Butter Protein Shake.* EatingWell.com. https://www.eatingwell.com/recipe/261428/chocolate-peanut-butter-protein-shake/#:~:text=Nutrition%20Facts&text=402%20calories%3B%20protein%2026.1g,784.7mg%3B%20sodium%20121.6mg.

Mourao, D. M., Bressan, J., Campbell, W. W., & Mattes, R. D. (2007). Effects of food form on appetite and energy intake in lean and obese young adults. *International journal of obesity*, *31*(11), 1688-1695. https://doi.org/10.1038/sj.ijo.0803667

Mushtaq, A., Khaliq, M., Saeed, A., Azeem, M. W., & Ghania, J. B. (2015). Almond (Purunus amygdalus L.): A review on health benefits, nutritional value and therapeutic applications. *International Journal of Chemical and Biochemical Sciences, 8*, 103-106.

NDTV. (2019). *Top reasons why peanuts should be your favourite snack this winter*. https://www.ndtv.com/health/tops-reasons-why-peanuts-should-be-your-favourite-snack-this-winter-2143886

Neale, E. P., Tran, G., & Brown, R. C. (2020). Barriers and facilitators to nut consumption: A narrative review. *International Journal of Environmental Research and Public Health*, *17*(23), 9127. https://doi.org/10.3390/ijerph17239127

Neumark-Sztainer, D., Story, M., Perry, C., & Casey, M. A. (1999). Factors influencing food choices of adolescents: findings from focus-group discussions with adolescents. *Journal* *of the American dietetic association*, *99*(8), 929-937. https://doi.org/10.1016/S0002-8223(99)00222-9

NPB. (2021). *Peanut Per Capita Consumption Breaks New Record for Second Year in a Row:* National Peanut Board. https://www.nationalpeanutboard.org/news/peanut-per-capita-consumption-breaks-new-record-for-second-year-in-row.htm?fbclid=IwAR1AVB1gIwIAP1uVVPmN01r38lO5AHJUMfrAt1KsEBJmz98iHqdWTgV0_sg

Obayelu, A. E., Ogunnaike, M. G., & Omotoso, F. K. (2019). Socioeconomic determinants of fruits consumption among students of the Federal University of Agriculture, Abeokuta, Ogun State, Nigeria. *International Journal of Fruit Science*, *19*(2), 211-220. https://doi.org/10.1080/15538362.2018.1528928

Pan, A., & Hu, F. B. (2011). Effects of carbohydrates on satiety: differences between liquid and solid food. *Current Opinion in Clinical Nutrition & Metabolic Care*, *14*(4), 385-390. 10.1097/MCO.0b013e328346df36

Pathak, M. (2020). *Apples*. WebMD. https://www.webmd.com/food-recipes/benefitsapples#:~:text=Apples%20are%20low%20in%20sodium,100%20calories

Perraillon, M. C. (2019). *Interpreting model estimates: marginal effects* [pdf]. *U*niversity of Colorado Anschutz Medical Campus. https://clas.ucdenver.edu/marcelo-perraillon/sites/default/files/attached-files/perraillon_marginal_effects_lecture_lisbon.pdf

Pollard, J., Kirk, S. L., & Cade, J. E. (2002). Factors affecting food choice in relation to fruit and vegetable intake: a review. *Nutrition research reviews*, *15*(2), 373-387. https://doi.org/10.1079/NRR200244

Poti, J. M., Mendez, M. A., Ng, S. W., & Popkin, B. M. (2015). Is the degree of food processing and convenience linked with the nutritional quality of foods purchased by U.S. households?. *The American journal of clinical nutrition*, *101*(6), 1251-1262. https://doi.org/10.3945/ajcn.114.100925

Randolph, J.J., Falbe, K., Manuel, A.K., & Balloun, J.L. (2014). A Step-by-Step Guide to Propensity Score Matching in R. *Practical Assessment, Research & Evaluation*, 19(18).

Reis, C. E. G., Bordalo, L. A., Rocha, A. L. C., Freitas, D. M. O., Da Silva, M. V. L., De Faria, V. C., ... & Alfenas, R. C. (2011). Ground roasted peanuts leads to a lower post-prandial glycemic response than raw peanuts. *Nutricion hospitalaria*, *26*(4), 745-751. DOI:10.3305/nh.2011.26.4.5024

Rolls, B. J., Morris, E. L., & Roe, L. S. (2002). Portion size of food affects energy intake in normal-weight and overweight men and women. *The American journal of clinical nutrition*, *76*(6), 1207-1213. https://doi.org/10.1093/ajcn/76.6.1207

Rolls, B. J., Roe, L. S., & Meengs, J. S. (2007). The effect of large portion sizes on energy intake is sustained for 11 days. *Obesity*, *15*(6), 1535-1543. https://doi.org/10.1038/oby.2007.182

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rosenheck, R. (2008). Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. *Obesity reviews*, *9*(6), 535-547. https://doi.org/10.1111/j.1467-789X.2008.00477.x

Silva, M. P., Martinez, M. J., Casini, C., & Grosso, N. R. (2010). Tocopherol content, peroxide value and sensory attributes in roasted peanuts during storage. *International journal of food science & technology*, *45*(7), 1499-1504.

Statista. (2021a). *Per capita consumption of fresh apples in the United States from 2000/01 to 2020/21*. https://www.statista.com/statistics/257167/per-capita-consumption-of-fresh-apples-in-the-us/?fbclid=IwAR3kyV2GVvqNVqUYonsse7rjpIh7Ncdd75bAnNPkciqqS_JNarf4XZQuKks

Statista. (2021b). *Per capita consumption of fresh citrus fruit in the United States from 2000 to 2021*. https://www.statista.com/statistics/257189/per-capita-consumption-of-fresh-citrus-fruit-in-the-us/?fbclid=IwAR0Q_N4BdHOlQk3B8YDQz6MDoai-dFYkYbCQnZoi2xx-6zjxQ2_OXb9QGqI

Statista (2022a). *Consumption of peanut butter in the U.S. 2020*. https://www.statista.com/statistics/281221/us-households-consumption-of-peanut-butter/?fbclid=IwAR0Ydhvw_OM5XljdMn9tb_caD5Tbm3IxM63VWcJXDPf1evhIw0REBg39ahE

Statista. (2022b). *Per capita almond consumption in the U.S. 2000/2001-2020/2021*. https://www.statista.com/statistics/1282808/us-almond-consumption-per-capita/?fbclid=IwAR2oAj0ee5f8xoxioWR7OLJEQOtHQKFuQK97l7m72OIiRr7-Gx8BV37ZgbE

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1. https://doi.org/10.1214%2F09-STS313

Teo, P. S., Tso, R., Dam, R. M., & Forde, C. G. (2022). Taste of Modern Diets: The Impacts of Food Processing on Nutrient Sensing and Dietary Energy Intake. *The Journal of Nutrition*, *152*(1), 200-210. https://doi.org/10.1093/jn/nxab318

Thomson, L. (2010). *Peanut Flour Cookies*. https://iheartvegetables.com/peanut-flour-cookies/

Todd, J. E., Mancino, L., & Lin, B. H. (2010). The impact of food away from home on adult diet quality. *USDA-ERS economic research report paper*, (90). https://dx.doi.org/10.2139/ssrn.1557129

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press. https://econpapers.repec.org/bookchap/oettbooks/emetr2.htm

USDA. (2018). *USDA food composition databases*. Food Data Central.

USDA. (2019). *Food Data Central*. U.S. Department of Agriculture, Agriculture Research Service. https://fdc.nal.usda.gov/fdc-app.html#/food-details/169917/nutrients

USDA. (2021a). *U.S. Per Capita Loss Adjusted Food Availability*. United States Department of Agriculture, Economic Research Service. https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=101944&fbclid=IwAR3ZHaNpHv6IJ0EcRBbYwXRtQ9_nKj1iukNvBdyP7yNN-BWg9tQlDQSJjfw

USDA. (2021b). *Food Patterns Equivalents Database*. Agricultural Research Service, U.S. Department of Agriculture. https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fped-overview/

USDA. (2022). *Food Data Central. U.*S. Department of Agriculture, Agricultural research Service. https://fdc.nal.usda.gov/

Vabe, M., & Hansen, H. (2014). The Relationship between Food Preferences and Food Choice: A theoretical Discussion. *International Journal of Business and Social Sciences, 5*(7), 145-157.

Ver Ploeg, M., Breneman, V., Farrigan, T., Hamrick, K., Hopkins, D., Kaufman, P., ... & Tuckermanty, E. (2009). *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to congress* (No. 2238-2019-2924). http://dx.doi.org/10.22004/ag.econ.292130

Wang, Y., Lloyd, B., Yang, M., Davis, C. G., Lee, S. G., Lee, W., ... & Chun, O. K. (2012). Impact of orange juice consumption on macronutrient and energy intakes and body composition in the U.S. population. *Public Health Nutrition*, *15*(12), 2220-2227. https://doi.org/10.1017/S1368980012000742

Wansink, B., & Chandon, P. (2006). Meal size, not body size, explains errors in estimating the calorie content of meals. *Annals of internal medicine*, *145*(5), 326-332. https://doi.org/10.7326/0003-4819-145-5-200609050-00005

Wansink, B., & Kim, J. (2005). Bad popcorn in big buckets: portion size can influence intake as much as taste. *Journal of nutrition education and behavior*, *37*(5), 242-245. https://doi.org/10.1016/S1499-4046(06)60278-9

Wansink, B., & Sobal, J. (2007). Mindless eating: The 200 daily food decisions we overlook. *Environment and Behavior*, *39*(1), 106-123. https://doi.org/10.1177/0013916506295573

Xu, P., Lone, T., & Torres, N. (2022). Market Integration and Price Discovery in California's Almond Marketing: A Vector Auto-Regressive (VAR) Approach. *International Journal of Business and Management*, *17*(9). https://doi.org/10.5539/ijbm.v17n9p43

Zhang, Z., Ampatzidis, Y., Fu, L., & Zhang, Z. (2022). Economic Analysis of an Apple Harvest and Infield Sorting Machine. In Z. Zhang, C. Igathinathane, Y. Wang, Y. Ampatzidis, & G. Liu (Eds.), *Mechanical Harvest of Fresh Market Apples* (pp. 105-121). Springer, Singapore. https://doi.org/10.1007/978-981-16-5316-2_7

APPENDICES

Appendix 1 Socio-Demographic Characteristics Based on No Peanut Consumption (n= 3673)

Variables	Mean/percentage	Standard deviation
Age (years)	47.61928	0.5458174
Total calorie intake	4088.38	49.85472
Aggregate activity MET score	1099.104	38.97645
Gender (%)		
Male	47.15	
Female	52.85	
Ethnicity (%)		
Non-Hispanic White	58.99	
Mexican American or Other Hispanic	17.54	
Non-Hispanic Black or Other Race	23.46	
Education (%)		
High School Degree or Less	37.52	
Some College or College Degree or Higher	62.48	

Appendix 2 Socio-Demographic Characteristics Based on Raw Peanut Consumption (n= 232)

Variables	Mean/percentage	Standard deviation
Age (years)	54.55358	2.404342
Total calorie intake	4341.708	215.1167

Aggregate activity MET score	996.1829	178.7003
Gender (%)		
Male	45.4	
Female	54.6	
Ethnicity (%)		
Non-Hispanic White	62.95	
Mexican American or Other Hispanic	12.18	
Non-Hispanic Black or Other Race	24.87	
Education (%)		
High School Degree or Less	43.1	
Some College or College Degree or Higher	56.9	

Appendix 3 Socio-Demographic Characteristics Based on Processed Peanut Consumption (n=763)

Variables	Mean/percentage	Standard deviation
Age (years)	49.05396	1.085738
Total calorie intake	4306.39	103.6783
Aggregate activity MET score	996.8034	53.91119
Gender (%)		
Male	50.3	
Female	49.7	
Ethnicity (%)		
Non-Hispanic White	75.54	

Mexican American or Other Hispanic	8.82	
Non-Hispanic Black or Other Race	15.64	
Education (%)		
High School Degree or Less	34.49	
Some College or College Degree or Higher	65.51	

Appendix 4 Socio-Demographic Characteristics Based on No Almond Consumption (n= 4290)

Variables	Mean/percentage	Standard deviation
Age (years)	48.18968	0.5715387
Total calorie intake	4162.112	49.86119
Aggregate activity MET score	1096.564	42.11154
Gender (%)		
Male	49.67	
Female	50.33	
Ethnicity (%)		
Non-Hispanic White	62.05	
Mexican American or Other Hispanic	15.47	
Non-Hispanic Black or Other Race	22.48	
Education (%)		
High School Degree or Less	38.91	
Some College or College Degree or Higher	61.09	

Appendix 5 Socio-Demographic Characteristics Based on Raw Almond Consumption (n= 197)

Variables	Mean/percentage	Standard deviation
Age (years)	53.6586	1.471336
Total calorie intake	4118.137	162.7183
Aggregate activity MET score	796.9889	84.29012
Gender (%)		
Male	26.16	
Female	73.84	
Ethnicity (%)		
Non-Hispanic White	66.96	
Mexican American or Other Hispanic	13.71	
Non-Hispanic Black or Other Race	19.33	
Education (%)		
High School Degree or Less	17.88	
Some College or College Degree or Higher	82.12	

Appendix 6 Socio-Demographic Characteristics Based on Processed Almond Consumption (n= 220)

Variables	Mean/percentage	Standard deviation
Age (years)	47.43933	1.730689
Total calorie intake	4105.099	124.0109
Aggregate activity MET score	1009.058	98.39258
Gender (%)		

Male	38.31	
Female	61.69	
Ethnicity (%)		
Non-Hispanic White	68.13	
Mexican American or Other Hispanic	16.52	
Non-Hispanic Black or Other Race	15.35	
Education (%)		
High School Degree or Less	22.15	
Some College or College Degree or Higher	77.85	

Appendix 7 Socio-Demographic Characteristics Based on No Apple Consumption (n= 3916)

Variables	Mean/percentage	Standard deviation
Age (years)	48.67879	.6259311
Total calorie intake	4159.514	52.23682
Aggregate activity MET score	1063.974	37.66517
Gender (%)		
Male	48.27	
Female	51.73	
Ethnicity (%)		
Non-Hispanic White	64.87	
Mexican American or Other Hispanic	13.75	
Non-Hispanic Black or Other Race	21.38	
Education (%)		

High School Degree or Less	36.93	
Some College or College Degree or Higher	63.07	

Appendix 8 Socio-Demographic Characteristics Based on Raw Apple Consumption (n= 617)

Variables	Mean/percentage	Standard deviation
Age (years)	47.2744	1.264299
Total calorie intake	4040.532	113.566
Aggregate activity MET score	1218.252	105.0622
Gender (%)		
Male	45.15	
Female	54.85	
Ethnicity (%)		
Non-Hispanic White	54.66	
Mexican American or Other Hispanic	23.24	
Non-Hispanic Black or Other Race	22.1	
Education (%)		
High School Degree or Less	35.33	
Some College or College Degree or Higher	64.67	

Appendix 9 Socio-Demographic Characteristics Based on Apple Juice Consumption (n=158)

Variables	Mean/percentage	Standard deviation
Age (years)	43.69587	
Total calorie intake	4557.652	

Aggregate activity MET score	861.8381	
Gender (%)		
Male	41.83	
Female	58.17	
Ethnicity (%)		
Non-Hispanic White	41.95	
Mexican American or Other Hispanic	24.75	
Non-Hispanic Black or Other Race	33.3	
Education (%)		
High School Degree or Less	34.85	
Some College or College Degree or Higher	65.15	

Appendix 10 Socio-Demographic Characteristics Based on No Orange Consumption (n=3766)

Variables	Mean/percentage	Standard deviation
Age (years)	47.64735	0.6163405
Total calorie intake	4080.943	48.82871
Aggregate activity MET score	1089.291	41.51896
Gender (%)		
Male	46.75	
Female	53.25	
Ethnicity (%)		
Non-Hispanic White	64.45	
Mexican American or Other Hispanic	14.39	

Non-Hispanic Black or Other Race	21.16
Education (%)	
High School Degree or Less	36.17
Some College or College Degree or Higher	63.83

Appendix 11 Socio-Demographic Characteristics Based on Raw Orange Consumption (n=334)

Variables	Mean/percentage	Standard deviation
Age (years)	50.22411	2.535868
Total calorie intake	4117.117	136.5797
Aggregate activity MET score	1065.672	177.5025
Gender (%)		
Male	53.86	
Female	46.14	
Ethnicity (%)		
Non-Hispanic White	42.85	
Mexican American or Other Hispanic	23.26	
Non-Hispanic Black or Other Race	33.89	
Education (%)		
High School Degree or Less	45.74	
Some College or College Degree or Higher	54.26	

Appendix 12 Socio-Demographic Characteristics Based on Orange Juice Consumption (n=578)

Variables	Mean/percentage	Standard deviation
Age (years)	52.079	0.7616598
Total calorie intake	4620.656	76.5594
Aggregate activity MET score	997.2612	96.77547
Gender (%)		
Male	51.33	
Female	48.67	
Ethnicity (%)		
Non-Hispanic White	62.09	
Mexican American or Other Hispanic	16.24	
Non-Hispanic Black or Other Race	21.67	
Education (%)		
High School Degree or Less	35.88	
Some College or College Degree or Higher	64.12	

Appendix 13 Multinomial Logistic Regression Results for Peanuts

Variable	Coefficient	S.E.	p-value
No peanut consumption	(Base outcome)		
Raw peanut consumption			
Female	0.053	0.149	0.730
Age**	0.024	0.010	0.034
Ethnicity (Mexican American or Other Hispanic)	-0.30	0.287	0.313

Ethnicity (Non-Hispanic Black or Other Race)	0.130	0.312	0.683
Education	-0.257	0.289	0.388
Aggregate Activity	0.00003	0.00008	0.744
Processed peanut consumption			
Female	-0.160	0.129	0.234
Age	0.00008	0.004	0.985
Ethnicity (Mexican American or Other Hispanic)***	-0.921	0.216	0.001
Ethnicity (Non-Hispanic Black or Other Race)***	-0.649	0.112	0.000
Education	0.023	0.141	0.871
Aggregate Activity	-0.00006	0.00005	0.215

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

Appendix 14 Multinomial Logistic Regression Results for Almonds

Variable	Coefficient	S.E.	p-value
No almond consumption	(Base outcome)		
Raw almond consumption			
Female***	0.984	0.213	0.000
Age***	0.018	0.005	0.003
Ethnicity (Mexican American or Other Hispanic)	0.176	0.416	0.678
Ethnicity (Non-Hispanic Black or Other Race)	-0.077	0.310	0.807
Education**	1.104	0.378	0.011
Aggregate Activity	-0.00004	.00008	0.662
Processed almond consumption			

Female	0.450	0.261	0.105
Age	-0.004	0.007	0.596
Ethnicity (Mexican American or Other Hispanic)	0.113	0.166	0.505
Ethnicity (Non-Hispanic Black or Other Race)**	-0.465	0.193	0.029
Education**	0.811	0.371	0.045
Aggregate Activity	-0.000007	0.00007	0.927

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

Appendix 15 Multinomial Logistic Regression Results for Apples

Variable	Coefficient	S.E.	p-value
No apple consumption	(Base outcome)		
Raw apple consumption			
Female	0.172	0.151	0.271
Age	-0.0004	0.005	0.939
Ethnicity (Mexican American or Other Hispanic)***	0.722	0.204	0.003
Ethnicity (Non-Hispanic Black or Other Race)	0.204	0.177	0.268
Education	0.180	0.184	0.346
Aggregate Activity*	0.00008	0.00005	0.094
Apple juice consumption			
Female	0.178	0.334	0.602
Age*	-0.016	0.008	0.056
Ethnicity (Mexican American or Other Hispanic)***	0.960	0.238	0.001
Ethnicity (Non-Hispanic Black or Other Race)***	0.812	0.246	0.005

Education	0.183	0.230	0.439
Aggregate Activity**	-0.0002	0.00008	0.039

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively

Appendix 16 Multinomial Logistic Regression Results for Oranges

Variable	Coefficient	S.E.	p-value
No orange consumption	(Base outcome)		
Raw orange consumption			
Female*	-0.309	0.146	0.051
Age	0.015	0.009	0.131
Ethnicity (Mexican American or Other Hispanic)***	0.958	0.255	0.002
Ethnicity (Non-Hispanic Black or Other Race)***	0.962	0.284	0.004
Education	-0.272	0.266	0.323
Aggregate Activity	-0.000003	0.00009	0.975
Orange juice consumption			
Female	-0.218	0.134	0.124
Age***	0.016	0.004	0.002
Ethnicity (Mexican American or Other Hispanic)	0.316	0.242	0.211
Ethnicity (Non-Hispanic Black or Other Race)	0.169	0.197	0.404
Education	0.055	0.183	0.770
Aggregate Activity	-0.00001	0.00007	0.837

Note: *, ** and *** = significant at 10%, 5% and 1% level of significance respectively