AN ANALYSIS OF SURVEY DATA TO DETERMINE SIGNIFICANT RISK FACTORS ASSOCIATED WITH ADOLESCENT MARIJUANA USE THROUGH UTILIZATION OF SAMPLE WEIGHTING METHODS

by

Kelsey Kimiko Baker

B.A. in Mathematics, Saint Martin's University, 2012

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Kelsey Kimiko Baker

It was defended on

November 14, 2014

and approved by

Thesis Advisor:
Jeanine M. Buchanich, PhD, Research Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Ada Youk, PhD, Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Marnie Bertolet, PhD, Assistant Professor
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

Jeanine M. Buchanich, M.Ed., Ph.D.

AN ANALYSIS OF SURVEY DATA TO DETERMINE SIGNIFICANT RISK FACTORS ASSOCIATED WITH ADOLESCENT MARIJUANA USE THROUGH UTILIZATION OF SAMPLE WEIGHTING METHODS

Kelsey Kimiko Baker, M.S.

University of Pittsburgh, 2014

ABSTRACT

This investigation seeks to identify factors associated with adolescent marijuana use in the 30 days prior to survey response collection in the 2012 National Survey on Drug Use and Health (NSDUH). Both inverse probability weighted and unweighted backwards elimination multivariate logistic regression modeling techniques were used to determine these factors. Final models compared the magnitude of the difference between odds ratios, the selection of final variables, the statistical significance of selected variables, and the overall fit of the models to determine whether or not we believed a weighted model was more appropriate for this type of complex sampling survey data.

Our analysis showed that age, tendency towards risky behavior, importance of religious beliefs, academic grades, cigarette use, and alcohol consumption were significant predictors of marijuana use. In addition, the odds of marijuana use in those who smoke cigarettes and consume alcohol are much higher than the odds in those who do not partake in either.

The public health significance of this study is that the results can be used to help public health officials understand the risk factors that affect an adolescent's decision to use marijuana. This insight would allow them to collaborate with policy makers to more accurately identify at risk teens and allow for avoidance, earlier detection, and treatment strategies.

The assumptions of logistic regression were met, but few model diagnostics were available for the weighted model due to the lack of appropriate statistical diagnostics in the Stata statistical software. However, based on our results, we believe the weighted model, which incorporates the complex sampling methods used in the data collection, is more sufficient for our data. Although the available diagnostics revealed similar results for both models, we saw notable differences in the odds ratios for race and academic grades, which leads us to believe that weights are a necessary component of the model.

Keywords: National Survey Data, Sampling Weights, Logistic Regression, Marijuana, Adolescent Drug Use, Methodological Comparisons, Alcohol, Cigarette Smoking

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my academic and thesis adviser, Dr. Jeanine Buchanich, the members of my thesis committee, Dr. Ada Youk and Dr. Marnie Bertolet, and one of my academic and professional mentors, Dr. Roslyn Stone, for their encouragement and support through the production of my thesis, as well as, throughout my academic career in the university's Graduate School of Public Health.

I would also like to thank my family (Dave, Leslie, and Morgan Baker) for all of their unconditional love, support, and confidence throughout all of my academic endeavors over the years.

1.0 INTRODUCTION

Marijuana use among both adults and adolescents has been a growing topic of public health interest over the past few decades. Following the legal authorization of personal use in both Colorado State and Washington State, along with increasing political pressure nationwide, quantification of the physical and mental health effects of the drug are needed. It is assumed that greater availability to the public will increase use. Further, although the drug is only legally permitted to be used by adults 21 and older, there is unease about the access adolescents (12 – 17 years of age) will have to the substance (Wallach, 2014).

According to the National Institute on Drug Abuse, marijuana is one of the most commonly used illicit substances in the United States ("Drug Facts: Marijuana", 2014). Between 2007 and 2012, approximately 4.5 million people (ages 12 and up) began using marijuana, expanding the national total to 18.9 million marijuana users ("Drug Facts: Nationwide Trends", 2014). This growth is in contrast to most other illicit drugs for which use has remained constant or even declined ("Drug Facts: Nationwide Trends", 2014).

In adolescents, comparable increases in marijuana use have been documented. Although a decrease in marijuana use was noted between the late 1990's and the mid-to-late 2000's, this trend did not hold. Since then, increases in use in the past 30 days by 8^{th}, 10^{th}, and 12^{th} graders

have been reported ("Drug Facts: High School and Youth Trends", 2014; Youth Risk Behavior Surveillance System, 2014; "Marijuana", 2014). For 12th graders, daily use has grown from 5% to 6.5% between 2008 and 2013 ("Drug Facts: High School and Youth Trends", 2014; "Marijuana", 2014).

Researchers have attributed this increasing trend in marijuana use to a decrease in the perceived risk of the drug. In the past, drug use patterns have shown that as perceived risk of a drug decreases, use of that drug increases, and vice versa ("Drug Facts: High School and Youth Trends", 2014). It is believed that this change in adolescent perception of marijuana may be connected to the growing medical use and the move towards the legalization of the drug ("Drug Facts: High School and Youth Trends", 2014). These public actions give the perception that marijuana is a "safe" drug.

In order to prevent further escalation in adolescent marijuana use, it is important for public health professionals to understand which risk factors increase a minor's probability of trying and/or abusing the drug. By identifying these elements, doctors and treatment centers can apply specialized patient care to their practice, and public health professionals can work with policy makers to create and apply preventative measures and implement the development of treatment programs within the appropriate communities.

1.1 MARIJUANA

Marijuana refers to the dried form of the plant *Cannabis sativa* ("Drug Facts: Marijuana", 2014; "Marijuana: MedlinePlus, 2014). The main, mind-altering chemical contained in the drug is delta-9-tetrahydrocannabinol (THC); this is the substance that produces the consumers' "high" ("Drug Facts: Marijuana", 2014). The Food and Drug Administration (FDA) approved the solitary form of the chemical compound THC (not marijuana as a whole) for medicinal purposes only. It is primarily used for cancer patients undergoing chemotherapy and HIV/AID patients with severe weight loss for pain relief and increased appetite ("Drug Facts: Marijuana", 2014; "Marijuana: MedlinePlus", 2014).

Although often referred to as "pot" or "weed", many of its other common names are "cannabis", "ganja", "grass", and "hash"/"hashish" ("Marijuana: MedlinePlus, 2014). In its dried form, marijuana can be rolled up and smoked like a cigarette or packed into and smoked via a basic pipe, water bong, hookah, or vaporizer, among others ("Drug Facts: Marijuana", 2014). It can also be cooked in food and ingested in the form of an "edible" or mixed and consumed as a form of tea ("Drug Facts; Marijuana", 2014; "Marijuana: MedlinePlus, 2014).

One of the main concerns surrounding adolescent use of marijuana is the drug's effect on brain development ("Drug Facts; Marijuana", 2014). Heavy use is shown to have a lasting effect on the user's memory and the brain's ability to process information; one study showed that people who began heavy use of the drug in adolescence lost an average of eight IQ points by age 38 (Meier, 2012). This cognitive decline was not seen in users who did not begin until adulthood (Meier, 2012).

Besides potential damage to intellectual development, marijuana has been shown to impair the cardiopulmonary system and harm the user's mental health ("Drug Facts; Marijuana", 2014). Like tobacco cigarettes, the smoke from marijuana irritates the lungs of the user. This kind of regular irritation can lead to daily coughing, recurrent acute chest illness, and increased risk of lung infections ("Drug Facts; Marijuana", 2014). The heart is also affected shortly after. The consumer will experience a 20-100% rise in heart rate as well as a 4.8-fold increased risk of heart attack within the first hour of inhalation ("Drug Facts; Marijuana", 2014). Additionally, the span of mental health damage ranges from psychotic reaction, to depression, to impaired judgment during and/or following a "high" ("Drug Facts; Marijuana", 2014; Johns, 2001; Arseneault, 2004). The association and direction of the relationship between marijuana and mental health effects is not fully understood but it has been linked to depression, anxiety, suicidal thoughts, and personality disturbances among adolescents ("Drug Facts; Marijuana", 2014; Johns, 2001; Arseneault, 2004). Some of these responses may depend on genetics, age at first use, and/or mental health status preceding use ("Drug Facts; Marijuana", 2014).

Facts regarding the effect of marijuana on the adolescent brain and body, support from previous research, and personal speculation aided in the assembly of the risk factors of interest. Studies reported a number of factors that are associated with the initiation and/or continuation of drug use. These variables include academic grades, parental relationships, peer relationships, mental health, risk-seeking behavior, religious commitment, alcohol use, personal experiences, age, gender, and race (Newcomb, 1986; Kilpatrick, 2000; Nation, 2006; McCabe, 2007). For example, one study found that "African Americans, but not Hispanics or Native Americans, were

at approximately 1/3 the risk of substance abuse as Caucasians" (Kilpatrick, 2000). The same study found that adolescents living with family members who had drug or alcohol problems had an increased risk for substance abuse (Kilpatrick, 2000). Another study determined that "antisocial peers and delinquent behavior were the strongest predictors of substance abuse" (Nation, 2006). Newcomb et al. summarize a number of different studies, and report that a variety of factors are "implicated in the initiation and maintenance of adolescent drug use" (Newcomb, 1986). Some of these factors include "peer use, poor grades in school, poor relationship with parents, depression, psychological distress, sensation seeking and the desire for novel and unusual experiences, lack of religious commitment, disruptive life events, and early use of alcohol" (Newcomb, 1986). Furthermore, another study revealed differences in marijuana use by gender. For all age groups, marijuana use was "significantly" higher among men than women (McCabe, 2007). Alcohol consumption and cigarette use have been considered "gateway" drugs for marijuana use (Newcomb, 1986; Kilpatrick; 2000).

Although we gathered a large list of risk factors from the literature, none of the previous studies reported finding all of these factors as significant predictors of marijuana use in a cumulative model. Table 1 lists identified risk factors associated with marijuana use.

Table 1. Previously identified risk factors

DEMOGRAPHICS
Age
Gender
Race
LIFE EXPERIENCES
Academic grades
Parental Relationships
Peer Relationships
Mental Health
Risk-Seeking Behavior
Religious Commitment
Alcohol Use
Personal Experiences

Using the information from the literature, we hope to find supplementary evidence for the previously identified risk factors as well as new support for a relationship between marijuana use and an adolescent's socioeconomic status (SES), overall health, and cigarette use to generate new hypotheses for future research.

1.2 REGRESSION

1.2.1 Binary Outcomes

Binary outcomes are those that can be divided into two separate, mutually exclusive and exhaustive outcomes. The sum of the probability of each of the two outcomes must be equal to

one. This outcome generally indicates whether or not a particular result of interest is true or false or whether there was a success or failure in a trial (Chatterjee, 2012). For example, the use of marijuana in the past 30 days could be a binary outcome.

1.2.2 Unweighted Logistic Regression

Logistic regression can be used with a binary outcome. It allows identification of significant risk factors associated with the outcome of interest.

The assumptions of logistic regression include the following (Chatterjee, 2012; Vittinghoff, 2005; Agresti, 2002; Steyerberg, 2009):

1. *Random sampling* – The sample is collected randomly from the population of interest so that the resulting sample is representative of the population from which it was drawn.
2. *Independent observations* – Each observation (person) in the study is independent of all others. Observations are not matched or the result of multiple measurements. One observation does not depend on the results of another observation for any reason.
3. *Discrete (binary) outcome variable* – The outcome can only take on a value of 0 or 1. These outcome categories must be mutually exclusive and exhaustive meaning that all cases must fall into one category or the other.
4. *Correct model specification* – The model includes and excludes the appropriate independent variables.
5. *Lack of multicollinearity* – There is no correlation among the independent variables in the model.

6. *Linearity* – There is a linear relationship between the independent covariates and the log odds of the outcome.
7. *Additivity* – Interaction terms are considered in the model, and significant interactions are included in the final model.

In the logistic model, a value of 0 is assigned to an observation if the conditions of the outcome are false. Otherwise, the observation is assigned a value of 1 indicating that the conditions are met. Because it takes on the values of 0 and 1, this outcome (Y_i) is considered dichotomous, and it follows a Bernoulli distribution with the following probability density function (Bain, 1992):

$$P[Y_i = y_i] = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

where π indicates the probability that the outcome equals one (or, for our interest, the probability that a participant used marijuana in the past 30 days).

In the following equation, the x_i's signify the independent variables (Chatterjee, 2012; Agresti, 2002):

$$\pi_i = \frac{\exp(\alpha + \beta_i x_i)}{1 + \exp(\alpha + \beta_i x_i)}$$

In order to satisfy the linearity assumption, the probability is written in terms of the log odds (Chatterjee, 2012; Agresti, 2002):

$$logit(\pi_i) = log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \beta_i x_i$$

Using this transformation, it is possible for independent variables to take on values extending from negative infinity to positive infinity and still be integrated in the model with a result remaining within the (0,1) range (Chatterjee, 2012). Each of the β's can be interpreted in the

following fashion: A one-unit increase in x_i results in an increase in the log-odds of Y_i by a value of β.

Further, notice that if we exponentiate the above equation, we get the following (Chatterjee, 2012):

$$\frac{\pi_i}{1 - \pi_i} = e^{\alpha + \beta_i x_i}$$

This produces the odds ratio (the probability of an event divided by one minus the probability of an event). If our independent variable of interest is cigarette use in the past month, the odds ratio can be interpreted in the following manner: The odds of marijuana use in smokers are $\frac{\pi_i}{1-\pi_i}$ times the same odds in non-smokers.

1.2.3 Weighted Logistic Regression

Weighted modeling is often used in cases where data is collected via survey sampling, which generally includes a combination of stratification, clustering, and sampling steps in order to collect a final sample that is representative of the entire population of interest. Sampling in this way allows the data to be used to make more accurate inferences and estimates about the population of interest.

Weighted logistic regression is a type of modeling that can be used when working with binary outcomes. The weighted model allows the use of sample data (a representative sample from the population of interest) as opposed to complete data (data from all individuals from the population of interest) on the entire population to calculate unbiased estimates for outcomes in the entire

population represented by the sample. National surveys assign a final sample weight for each participant. For the i^{th} respondent, the given weighted value, w_i, indicates the number of sampling units from the target population each participant represents. This is also known as the inverse probability of selection for the i^{th} respondent.

$$w_i = \frac{1}{P[Person\ i\ would\ be\ included\ in\ the\ sample]}$$

The sum of each of these weighted values, w_i, equals the total estimated size of the target population (USDHHS: Codebook, 2012).

$$\sum_i w_i = estimated\ size\ of\ target\ population$$

where the summation is over all respondents in the survey (USDHHS: Codebook, 2012). The weighted sample average is calculated in the following manner:

$$E_w[Y] = \frac{\sum_i w_i y_i}{\sum_i w_i}$$

where $E_w[Y]$ is the weighted average of the outcome Y, y_i is the value of Y for participant i, and w_i is the probability weight corresponding to person i (Vittinghoff, 2005; Korn, 1999; Lohr, 2010). In order to satisfy the linearity assumption, the probability for the weighted model can also be written in terms of the log odds:

$$\text{logit}[\pi_{w_i} \mid w_i] = \frac{\sum_i \left\{ w_i \left[log\left(\frac{\pi_{w_i}}{1 - \pi_{w_i}} \right) \right] \right\}}{\sum_i w_i} = \frac{\sum_i \{ w_i [\alpha + \beta_i x_i] \}}{\sum_i w_i}$$

$$\text{logit}[\pi_{w_i} \mid w_i] = \frac{1}{\sum_i w_i} \left\{ \sum_i [w_i \alpha + w_i \beta_i x_i] \right\}$$

Exponentiating the above equation gives the following:

$$\frac{\pi_{w_i}}{1-\pi_{w_i}} = \exp\left[\frac{\sum_i\{w_i[\alpha + \beta_i x_i]\}}{\sum_i w_i}\right] = \exp\left\{\frac{1}{\sum_i w_i}\left[\sum_i [w_i\alpha + w_i\beta_i x_i]\right]\right\}$$

where $\frac{\pi_{w_i}}{1-\pi_{w_i}}$ is the weighted odds ratio.

There is controversy in the literature over when it is appropriate to use weighted modeling techniques when using survey data (Korn, 1999). Some analysts believe that if a complex sampling design is used, weighting in the model is unnecessary as the probability of selection was accounted for in the sampling process (Chun, 1997; Wissoker; Winship, 1994). Others consider it important to utilize a weighted model when analyzing cross-national comparative surveys due to the differences in population sizes between countries or when the objective of the study is to make inferences about the population (Skinner, 2012). Moreover, there are some who argue that both weighted and unweighted models should be fit and compared to determine which is more appropriate based on the magnitude of the difference (Dumouchel, 1983; Hinkins, 2009). As there is no standard rule for model weighting and because we are not making national comparisons or population inferences, we will use both weighted and unweighted models and attempt to determine which is more appropriate based on the magnitude of the differences between odds ratio estimates and overall model fit. Section 2.7 will give a detailed description of the statistical methods.

1.3 STATEMENT OF THE PROBLEM

Due to increasing marijuana use among adolescents in the past few years, it is of utmost importance that we understand which factors contribute to their decision to initiate and continue use of the substance. Understanding these elements would allow appropriate intervention efforts to be put into place in order to prevent the detrimental health effects associated with marijuana use in the youth of the United States.

Furthermore, we are interested in establishing whether a sample weighted or unweighted model is more appropriate for these data by determining the reliability of an unweighted model when complex sampling techniques are used. Thus, we pose the questions:

1. What factors are significant predictors of adolescent marijuana use in the past 30 days?
2. Is sample weighted or unweighted modeling more appropriate for these data?

2.0 METHODS

2.1 NATIONAL SURVEY ON DRUG USE AND HEALTH, 2012

The United States Department of Health and Human Services, the Substance Abuse and Mental Health Services Administration, and the Center for Behavioral Health Statistics and Quality fund the National Survey on Drug Use and Health (NSDUH). The survey is conducted annually in compliance with Section 505 of the Public Health Service Act, which requires annual collection of data on the level and patterns of substance use ("About the Survey"). These organizations use the survey responses to measure prevalence and determine factors associated with drug use in the United States (USDHHS: Overview, 2014; USDHHS: Codebook, 2012). The survey collects data on a sample of the non-institutionalized, civilian population. This excludes members of active-duty military and persons in institutional group quarters such as hospitals, prisons, nursing homes, and treatment centers (USDHHS: Codebook, 2012). Less than 2% of the population is estimated to fall into one these omitted categories.

The 2012 sample was selected using a multi-level, stratified random sampling technique further described in Section 2.4: Survey Sampling.

All information for the 2012 survey was collected via computer-assisted personal interviewing and audio computer-assisted self-interviewing. The personal interviewing process covered demographic information such as gender, race/ethnicity, age, marital status, educational background, employment status, and household composition. Respondents then moved on to the self-interviewing phase. Questions during this segment covered more sensitive information such as illicit drug use and private behaviors. The self-interviewing phase was used in hopes that participants would feel more comfortable being honest in their responses to sensitive material (USDHHS: Overview, 2014; USDHHS: Codebook, 2012; Clark, 2013).

Participants older than 18 gave informed consent during a preliminary interview stage of the selection process. Those participants who were 12-17 years old gave assent. This means that these respondents were included in the study only if their parent/guardian gave consent for the household to be involved in the process, and then the adolescent also provided verbal agreement (Clark, 2013).

2.2 DATA MANAGEMENT

One of the major obstacles faced during this study was narrowing down the vast quantity of available variables into a meaningful number of representative ones. In total, there was data collected on 11 "core substances" which included tobacco, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives. Within each of these individual sections, over 50 questions could be asked regarding first time of use, frequency of use, and time since last use, among others. Responses were also collected in great

detail for prior substance abuse, substance treatment, social environment, youth experiences, household composition, income and insurance, and demographics, among others. After the participant answered all questions, investigators created a substantial number of new variables that provided summaries of multiple responses.

After much time spent reading through the codebook, understanding the format of the survey, and scanning the literature for previously reported relationships, 18 variables of interest were selected from the 3,120 that were collected and/or created.

Depending on the age of the participant and their answers to survey questions in earlier sections, respondents would automatically be opted out of future sections. For example, adults were limited to adult or general sample sections, so they did not respond to any questions within the "Youth Experiences" section or any other adolescent-only units. Furthermore, if any respondent, adult or otherwise, answered that they had "Never Used" a drug, they would be automatically opted out of the any further questions regarding that drug's use.

In the student version of Stata 13, a maximum of 2,048 variables can be loaded. Therefore, the data files were manipulated within SPSS, removing unnecessary variables, and converted back to a Stata file. Data files were thoroughly checked to ensure that the conversion process had not eliminated or changed any responses; errors were detected by comparison to the original data and corrected as necessary. This was especially important for values denoting "Refused to Answer", "Don't Know", or "Skipped".

After restricting the dataset to adolescents (ages 12 – 17), we created our outcome variable by constructing an indicator for marijuana use in the past 30 days from the responses to the number of times the participant used marijuana in the past 30 days and if they had ever used marijuana. Those who responded that they had "Never Used Marijuana" or "Did Not Use Marijuana in the Past 30 Days" were given a value of 0 ("No") for the indicator variable. Those who had used anywhere from 1 to 30 days in the past 30 days were given a value of 1 ("Yes") for the indicator term. If the question was left blank or a response of "Don't Know" or "Refused" was given, the answer was coded as missing. Because the outcome takes on the values of 0 ("No") and 1 ("Yes"), it is dichotomous and meets the requirements for logistic regression.

An indicator variable for cigarette use in the past 30 days was created using responses to the number of times the participant smoked all or part of a cigarette in the past 30 days and if they had ever smoked all or part of a cigarette. Likewise, an indicator variable for alcohol consumption in the past 30 days was created using responses to the number of times the participant consumed any kind of alcohol in the past 30 day and if they had ever consumed even a drop of alcohol.

For all other independent covariates, if the question was left blank, coded by survey investigators as "Other Missing", or a response of "Don't Know" or "Refused" was given, the observation was assigned missing.

The race variable was pooled from seven groups into four groups in order to have an adequate sample within each group: Non-Hispanic White, Non-Hispanic Black/African American,

Hispanic, and Non-Hispanic Other (Native American/AK Native, Native HI/Other Pacific Islander, Asian, More Than One Race). Similarly, the self-reported health variable was merged from five levels into four levels: Excellent, Very Good, Good, and Fair/Poor.

2.3 SURVEY QUESTIONS

The NSDUH is conducted annually in the United States to collect a variety of information on drug use, mental health, demographics, and a collection of life experiences. Data on frequency of use is collected for an assortment of drugs under the categories of tobacco, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives (USDHHS: Codebook, 2012). These groupings include both street drugs and prescription substances. Other questions include the participant's perceived risk and availability for each drug and whether treatment was ever received.

Adolescents were asked about their social environment including how often they moved in the past few years, importance of religious views, academic experiences, interest in school, and frequency of aggressive behaviors towards themselves or someone else. They also provided information regarding their home life including number of residents in their home, whether both parents lived in the household, and degree of parental activity in the child's development. Information regarding the child's mental health status was primarily related to whether the child received mental health treatment at any facility in the past year.

The variables selected to determine which characteristics of an adolescent's life would put him or her at higher risk of using marijuana in the past 30 days are listed in Table 2.

For the questions regarding the presence of a mother and/or father figure in the household, participants were asked about the composition of their home. If they indicated that a female lived with them, the interviewer displayed a show-card of 14 female options and asked the child to specify how they would rank each cohabitant. Responses for mother included biological mother, stepmother, foster mother, and adoptive mother (USDHHS: Specification, 2012). Equivalently, if the participant indicated that a male lived in the household, they were given a show-cared of 14 male options and gave them a ranking. Responses for father included biological father, stepfather, foster father, and adoptive father (USDHHS: Specification, 2012).

Due to the age of our sample of interest, it was difficult to determine how to incorporate SES information into the model. The survey allowed another household member 18 or older to act as a proxy for income and insurance information (USDHHS: Specification, 2012). Participants were asked if they would like to use a proxy at the beginning of the section covering income and insurance information. If a proxy was utilized, they joined the participant in responding for the income and insurance section only. Over 85% of the adolescents opted to use a proxy for these questions (USDHHS: Codebook, 2012).

Some examples of government assistance programs contained in the corresponding question include Supplemental Security Income (SSI) and food stamps among others; and some examples

of health insurance coverage contained in the corresponding question include Medicare, Medicaid, military health care, and private health insurance among others.

A comparison of total family income to the poverty thresholds determined by the U.S. Census Bureau for each distinct family composition establishes each individual's relationship to the Federal poverty level. "To be at 100% of the poverty threshold is equivalent to having a family income that is the same as the poverty threshold. A poverty level less than 100% indicates having a family income less than the poverty threshold and therefore defined by the Federal government as living in poverty. A poverty level greater than 100% indicates having a family income greater than the poverty threshold" (USDHHS: Codebook, 2012).

Table 2. Survey question variables

GENERAL DEMOGRAPHICS
Age
Gender
Race / Hispanicity
County Metro Status
SURVEY QUESTIONS
Overall Health
Tendency Towards Risky Behavior
Importance of Religious Beliefs
Academic Grades
Positive Reinforcement from Parents
Presence of a Mother in Household
Presence of a Father in Household
Total Family Income
Comparison to Federal Poverty Level
Use of Any Government Assistance Programs
Coverage by Any Health Insurance
Use of Cigarettes in the Past 30 Days
Use of Alcohol in the Past 30 Days
OUTCOME
Use of Marijuana in the Past 30 Days

2.4 SURVEY METHODS

It is important in any data collection process to avoid as much bias as possible and to accumulate a representative sample of the population of interest. When a survey is used to collect data, it is imperative to have an understanding of who will be participating and what part of the population they represent. To do this, the NSDUH used a multi-step stratification and sampling process to accumulate their sample. Using the steps in this process, a weighted value (inverse probability weight) was assigned to each respondent, which signified the number of people whom particular participant represented within the entire population (USDHHS: Codebook, 2012). Applying these sampling techniques avoids the possible underestimation of the standard errors (SE) of the predictors in the regression models. Sampling weights were used instead of frequency weights to avoid discovery of false statistically significant results (Korn, 1999; Chatterjee, 2012).

2.5 SURVEY SAMPLING

In order to collect a representative sample, investigators used many different levels of stratification and sampling to appropriately narrow down the population. Overall, they attempted to sample an equal number of people from three main age groups: 12-17 year olds, 18-25 year olds, and 26 years or older. Table 3 gives a summary of the stage delineated sample design, and Figure 1 visually depicts the sampling design.

For Stage 1, they stratified the United States into each of the 50 states plus the District of Columbia. It was determined that each of the eight largest states (California, Florida, Illinois,

Michigan, New York, Ohio, Pennsylvania, and Texas) would have a target sample of 3,600 participants and the remaining 43 states (including the District of Columbia) would each have a target sample of 900 participants. All 50 states plus Washington D.C. were then stratified into state sampling regions (SSR's) of equal population size within each state for a total of 900 SSR's per state (Morton, 2013; USDHHS: Overview, 2014; USDHHS: Codebook, 2012).

Each SSR was divided by census tract. Within each census tract, investigators hoped to obtain a minimum of 100 or 150 dwelling units (DU's) depending on rural or urban area status, respectively (USDHHS: Codebook, 2012). If the minimum requirement was not met, smaller census tracts were combined with adjacent tracts. (Note that the result of merging two contiguous census tracts is simply a new census tract.) Once the minimum requirement was met or exceeded, a sample of 48 census tracts were selected per SSR with a selection probability proportional to the size (PPS) of the population with oversampling due to SES and race/ethnicity composition (USDHHS: Codebook, 2012; Morton, 2013).

Because census tracts generally exceeded the minimum requirement, they were divided into census blocks, which can be a visible feature (street, railroad, etc.) or a nonvisible boundary (city, town, etc.) (USDHHS: Codebook, 2012). Neighboring census blocks were then aggregated to create an area segment. For Stage 2, using a random sampling method, one area segment was selected PPS from each of the census tracts, and then randomly assigned to survey year and data collection quarter (Morton, 2013; USDHHS: Overview, 2014; USDHHS: Codebook, 2012).

Once area segments were assigned, field investigators created lists of all eligible DU's within each segment. In Stage 3, they determined the minimum number of DU's needed in order to meet the target sample size, and systematically sampled from the addresses based on a random start point plus an interval. In Stage 4, researchers then visited the selected households to determine the number of available participants, and randomly sampled 0, 1, or 2 residents per household with oversampling for age and size of state (Morton, 2013). Preliminary personal interviews were conducted with a head of the household (18 or older) to determine whether or not a DU maintained eligibility and to collect consent for the family's inclusion in the study (Morton, 2013; USDHHS: Overview, 2014; USDHHS: Codebook, 2012).

Table 3. Stage delineated sample design

STAGE 1
Stratification by State Stratification by State Sampling Region (SSR's) *(Chosen to be approximately the same size)* Sampling of Census Tracts from SSR's *(PPS with oversampling due to average SES and race/ethnicity composition)*
STAGE 2
Sampling of Area Segments from Census Tracts *(PPS)*
STAGE 3
Sampling of Dwelling Units (DU's) from Area Segments *(Systematically based on random start plus interval)*
STAGE 4
Sampling of Persons (0, 1, or 2) from DU's *(Simple random sample with oversampling for age and state size – large or small)*

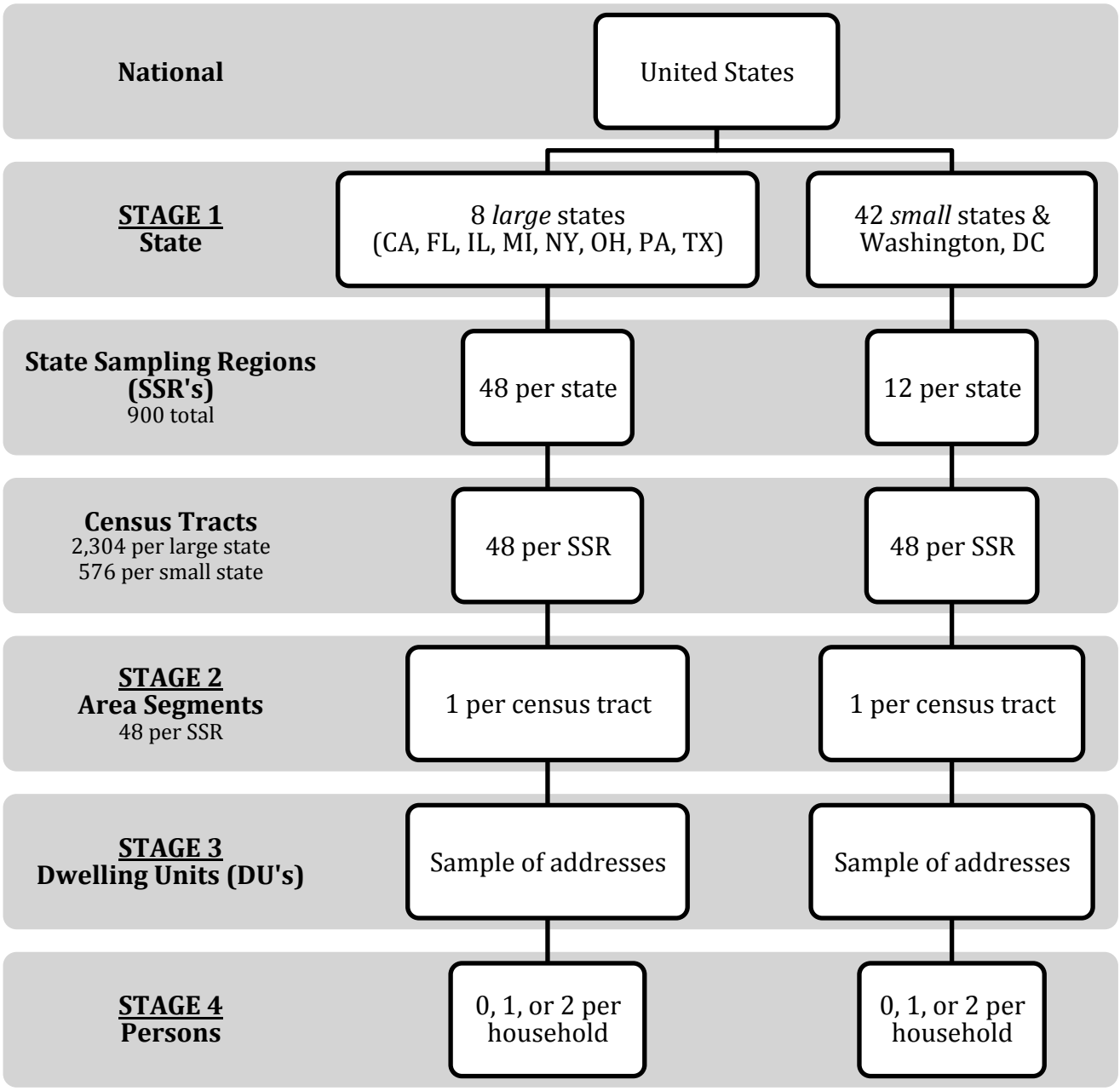

Figure 1. Flow chart of sampling design

2.6 SAMPLING WEIGHTS

Investigators for the NSDUH created an inverse probability weight (sampling weight) for each individual based on the product of 15 components (Table 4) (Morton, 2013, USDHHS: Codebook, 2012). These components account for the probability of selection at each stage and sub-stage of the selection process (Morton, 2013).

As shown in Table 4, the first ten components are related to dwelling selection and the other five components are related to the selection of the individual within the dwelling. Weighting Step #1 and Step #2 account for the probability of census tract and area segment selection, respectively, based on Stage 1 and Stage 2 of the sampling design (Morton, 2013). Step #3 makes an adjustment for the probability of assignment to data collection quarter, and Step #4 makes an adjustment for adding or sub-sampling dwelling units. Addition of a dwelling unit may occur if a new home was built and was unaccounted for in the initial tally of eligible dwelling units. A dwelling may be sub-sampled if more than one family lives in a dwelling unit (i.e. an apartment). Step #5 accounts for the probability of dwelling unit selection based on Stage 3 of the sampling design (Morton, 2013). The final five steps of the dwelling unit-level weighting component are standard adjustments. They make modifications for additions to the number of eligible DU's available for sampling (Step #6), removal of number of eligible DU's (Step #7), failure to sample eligible DU's (Step #8), changes in expected sampling size (Step #9), and occurrence of extreme weights due to sampling and previous weight adjustments (Step #10) (Morton, 2013). These final adjustments help match adjusted weights to census data, which is the "gold standard" for population estimates.

Within the person-level weight components (Table 4), Step #11 accounts for the probability of person selection within a dwelling unit based on Stage 4 of the sampling design (Morton, 2013). In Step #12, an adjustment is made for changes in actual demographics of sampled persons compared to expected demographics (Morton, 2013). Steps #13, #14, and #15 are standard adjustments made for failure to sample eligible persons or refusal of eligible persons to participate, changes in expected sample size, and occurrence of extreme weights due to sampling and previous weight adjustments, respectively (Morton, 2013). These final adjustments help match adjusted weights to census data, which is the "gold standard" for population estimates.

Although many different weighting variables are available, for our purposes, we are only considering the overall weight assigned to each individual. For example, based on the data and available weights a 14-15 year old Hispanic female living in a large metropolitan area represents between 5 and 11,794 people depending on area segment, census tract, state sampling region, and state.

Table 4. Sampling weight components

DWELLING UNIT-LEVEL DESIGN WEIGHT COMPONENT
#1: Inverse probability of selecting census tract
#2: Inverse probability of selecting segment
#3: Quarter segment weight adjustment
#4: Sub-segmentation inflation adjustment
#5: Inverse probability of selecting dwelling unit
#6: Inverse probability of added/sub-sampled dwelling unit
#7: Dwelling unit release adjustment
#8: Dwelling unit non-response adjustment
#9: Dwelling unit post-stratification adjustment
#10: Dwelling unit extreme weight adjustment
PERSON-LEVEL DESIGN WEIGHT COMPONENT
#11: Inverse probability of selecting a person within a dwelling unit
#12: Selected person post-stratification to roster adjustment
#13: Person-level non-response adjustment
#14: Person-level post-stratification adjustment
#15: Person-level extreme weight adjustment

*Adapted from Morton, 2013

Figure 2 shows that the final person level weights have a truncated normal distribution (as negative weights can not exist) meaning that the weight adjustments were performed correctly. No extreme weighted values can be seen.

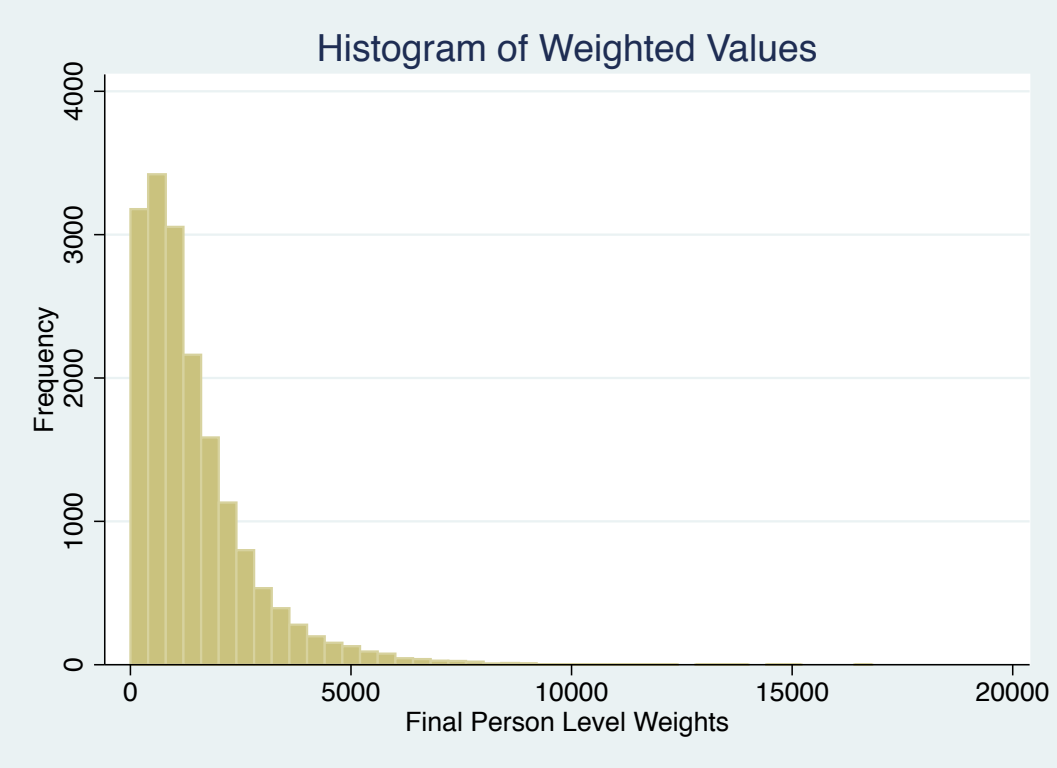

Figure 2. Histogram of final person level weights

2.7 STATISTICAL METHODS

The questions of interest in Table 2 were either taken directly from the survey, or are a recoded combination of multiple questions in order to supply a greater amount of information in one answer.

We tabulated the unweighted demographic information based on marijuana use in the past 30 days to determine the characteristics of the sample. A tabulation of participants' responses to our selected questions of interest was also created.

Univariate unweighted and weighted logistic regression models were fitted for the covariates of interest to determine their relationship to the outcome of interest (marijuana use in the past 30 days). These covariates included age, gender, race/hispanicity, county metro status, overall health, tendency towards risky behavior, importance of religious beliefs, academic grades, parents' frequency of positive reinforcement, presence of a mother figure in the household, presence of a father figure in the household, total family income, use of any government assistance program, comparison to the federal poverty level, health insurance coverage, cigarette use in the past 30 days, and alcohol consumption in the past 30 days.

We also explored multivariate unweighted and sample weighted logistic models. Using the backwards-elimination selection method with a semi-conservative p-value to remove (pr) of $\alpha=0.1$, we produced our two models of interest. The age, gender, and race/hispanicity demographic variables were forced into each model at every level regardless of statistical significance, based on the literature ("Drug Facts: High School and Youth Trends", 2014; Youth Risk Behavior Surveillance System, 2014; "Marijuana", 2014; McCabe, 2007). [Note that hispanicity refers to the race and ethnicity of the participant, which incorporates Hispanic background]. Both the unweighted and weighted models were considered complete when all variables in the model had p-values less than $\alpha=0.1$ (excluding the demographic variables previously mentioned). We did not make any adjustments for multiple comparisons.

The final models were tested to determine if the model assumptions were met. Based on the methods for data collection we assumed an independent, random sample, and the creation of our outcome variable confirmed a discrete, binary outcome. For the unweighted model, calculation of the variance inflation factors (VIF's) allowed us to determine if multicollinearity (correlation among independent variables) was an issue. Goodness of fit was determined using the Hosmer-Lemeshow Goodness-of-Fit test, and the linktest was used to determine selection of meaningful predictors and appropriate model specification. Furthermore, we calculated and plotted residuals, influential points, and leverage points to evaluate outlying, influential, and high leverage points, respectively (Agresti, 2002; Vittinghoff, 2005, Chatterjee, 2012; Hinkins, 2009; Valiant).

The literature supported use of the same diagnostic procedures for the weighted model as we used for the unweighted model (Valiant). Unfortunately, these diagnostic procedures are not yet available for sample-weighted models in the Stata software. Therefore, we were only able to calculate the VIF's to determine the presence of multicollinearity, and the linktest to determine appropriate model specification.

By comparing odds ratios, model diagnostics, and overall fit, we determined whether a weighted model was necessary and/or more appropriate for these data (Dumouchel, 535; Hilkins, 2207-2210). The magnitude of the difference between the odds ratios in the models was determined. If there appeared to be any notable differences in the odds ratios, the weighted model would be more applicable to the data. The uncentered VIF calculations were compared for each model to determine if one model selected variables that resulted in a multicollinearity issue. If one model

had a multicollinearity issue, it would indicate that the alternate model should be selected as a more suitable fit. The results of the linktest allowed us to compare the overall fit of each model. If the linktest was statistically significant for one model but not for the other, it would indicate that the other model was the more appropriate fit for the data.

We re-ran the final unweighted model removing points with the highest standardized residual values and again removing points with the highest influential values.

3.0 RESULTS

Any participants who did not know if they had used marijuana in the past 30 days, refused to answer the question, or left the question blank were removed from the sample (n = 193) as their answers did not supply the necessary information.

3.1 DESCRIPTIVE TABLATION OF DATA

3.1.1 Sample Demographics by Marijuana Use

In our unweighted tabulation (Table 5) of demographic information, about 35% more (32.3% vs. 67.1%) 16-17 year olds and 30% fewer (34.4% vs. 4.5%) 12-13 year olds used marijuana in the past 30 days. No prominent differences were seen by gender, race/hispanicity, and county metro status. The age groups were approximately evenly split for marijuana non-users, while marijuana users had much larger differences by age group. The groups were approximately evenly split by gender. More than half of the participants were white for both marijuana users and marijuana non-users. Non-metro was the least represented county metro status for both groups.

Table 5. Demographics (Unweighted)

	Total n = 17206	**No MJ** n = 15937	**MJ** n = 1269
Age – n(%)			
12 – 13	5531	5474 (34.4%)	57 (4.5%)
14 – 15	5681	5320 (33.4%)	361 (28.5%)
16 – 17	5994	5143 (32.3%)	851 (67.1%)
Gender – n(%)			
Male	8686	8031 (50.4%)	655 (51.6%)
Female	8520	7906 (49.6%)	614 (48.4%)
Race/Hispanicity – n(%)			
White	9779	9046 (56.8%)	733 (57.8%)
Black/African American	2281	2118 (13.3%)	163 (12.8%)
Hispanic	3418	3176 (19.9%)	242 (19.1%)
Other	1728	1597 (10.0%)	131 (10.3%)
County Metro/Non-Metro Status – n(%)			
Large Metro	7660	7084 (44.5%)	576 (45.4%)
Small Metro	5844	5402 (33.9%)	442 (34.8%)
Non-Metro	3702	3451 (21.7%)	251 (19.8%)

3.1.2 Survey Question Responses by Marijuana Use

Survey question responses by marijuana use in the last 30 days are shown in Table 6. Nearly half of adolescents who used marijuana responded that they "Sometimes" or "Always" like to test themselves by doing something risky compared to only about a quarter of adolescents who did not use marijuana. Most adolescents for both marijuana users and non-users responded that they "Agreed" that their religious beliefs were a very important part of their life. Furthermore, almost 19% (33.4% vs. 14.7%) fewer participants who "Strongly Agreed" that their religious beliefs were a very important part of their life decided to partake in the use of the drug. A higher percent of those with an income less than or equal to one times the poverty threshold (i.e. living in poverty) and those who participated in any kind of government assistance program used

marijuana. At least 85% of adolescents reported that there was a mother figure in the household, while only about 65% reported that a father figure was in the household.

For marijuana non-users, “B+, B, B-” was the highest reported academic grading average followed by “A+, A, A-”, “C+, C, C-”, “D or Less”, and “Non-Letter Grades”, while for marijuana users, “B+, B, B-” was the highest reported average, but “C+, C, C-” was the second highest followed by “A+, A, A-”, “D or Less”, and “Non-Letter Grades”.

Over 50% of marijuana non-users reported that their parents “Always” told them they were proud of something they had done in the past 12 months. On the other hand, only about 37% of marijuana users reported this answer.

Large differences were found in marijuana use in both cigarette use and alcohol consumption in the past 30 days. About 42% (4.0% vs. 45.6%) more adolescents who smoked all or part of a cigarette in the past 30 days also used marijuana. Similarly, about 52% (8.3% vs. 60.8%) more adolescents who consumed alcohol in the past 30 days also used marijuana.

Table 6. Survey questions considered as risk factors - n(%)

How often do you like to test yourself by doing something a little risky?

	Never	Seldom	Sometimes	Always	Missing
No MJ	5917 (37.1%)	5329 (33.4%)	3943 (24.7%)	596 (3.7%)	152 (1.0%)
MJ	158 (12.5%)	390 (30.7%)	536 (42.2%)	181 (14.3%)	4 (0.3%)

Your religious beliefs are a very important part of your life.

	Strongly Disagree	Disagree	Agree	Strongly Agree	Missing
No MJ	1652 (10.4%)	2247 (14.1%)	6322 (39.7%)	5316 (33.4%)	400 (2.2%)
MJ	322 (25.4%)	296 (23.3%)	436 (34.4%)	187 (14.7%)	28 (2.2%)

What were your grades for the last semester or grading period you completed?

	A+, A, A- Average	B+, B, B- Average	C+, C, C- Average	D or < D Average	Non-Letter Grades	Missing
No MJ	4903 (30.8%)	6063 (38.0%)	2657 (16.7%)	609 (3.8%)	443 (2.8%)	1262 (7.9%)
MJ	195 (15.4%)	435 (34.3%)	407 (32.1%)	156 (12.3%)	31 (2.4%)	45 (3.6%)

Table 6. Continued

During the past 12 months, how often did your parents tell you they were proud of something you had done?

	Always	Sometimes	Seldom	Never	Missing
No MJ	9020 (56.6%)	4680 (29.4%)	1568 (9.8%)	595 (3.7%)	74 (0.5%)
MJ	474 (37.4%)	427 (33.7%)	241 (19.0%)	118 (9.3%)	9 (0.7%)

Is there a mother in the household?

	Yes	No	Missing
No MJ	14658 (92.0%)	1267 (8.0%)	12 (0.1%)
MJ	1088 (85.7%)	179 (14.1%)	2 (0.2%)

Is there a father in the household?

	Yes	No	Missing
No MJ	11655 (73.1%)	4267 (26.8%)	15 (0.1%)
MJ	829 (65.3%)	438 (34.5%)	2 (0.2%)

How would you rate your overall health?

	Excellent	Very Good	Good	Fair / Poor	Missing
No MJ	5477 (34.4%)	6728 (42.2%)	3170 (20.0%)	561 (3.5%)	1 (< 0.0%)
MJ	281 (22.1%)	519 (40.9%)	373 (29.4%)	96 (7.6%)	0 (0.0%)

Table 6. Continued

What is the total income of your family?

	< $20,000	**$20,000 - $49,999**	**$50,000 - $74,999**	**≥ $75,000**	**Missing**
No MJ	2895 (18.2%)	4961 (31.1%)	2763 (17.3%)	5318 (33.4%)	0 (0.0%)
MJ	288 (22.7%)	401 (31.6%)	211 (16.6%)	369 (29.1%)	0 (0.0%)

Have you participated in one or more government assistance programs?

	Yes	**No**	**Missing**
No MJ	4359 (27.4%)	11578 (72.7%)	0 (0.0%)
MJ	434 (34.2%)	835 (65.8%)	0 (0.0%)

How do you compare to the Federal poverty threshold?

	Income ≤ 1x Threshold	**Income >1x & ≤ 2x Poverty Threshold**	**Income > 2x Poverty Threshold**	**Missing**
No MJ	3607 (22.6%)	3453 (21.7%)	8877 (55.7%)	0 (0.0%)
MJ	321 (25.3%)	290 (22.9%)	658 (51.9%)	0 (0.0%)

Table 6. Continued

Does any type of health insurance cover you?

	Yes	No	Missing
No MJ	14837 (93.1%)	880 (5.5%)	220 (1.4%)
MJ	1180 (93.0%)	76 (6.0%)	13 (1.0%)

Have you smoked part of all of a cigarette in the past 30 days?

	Yes	No	Missing
No MJ	629 (4.0%)	15285 (95.9%)	23 (0.1%)
MJ	578 (45.6%)	680 (53.6%)	11 (0.9%)

Have you ever, even once, in the past 30 days had a drink of any type of alcoholic beverage?

	Yes	No	Missing
No MJ	1329 (8.3%)	14280 (89.6%)	328 (2.1%)
MJ	771 (60.8%)	457 (36.0%)	41 (3.2%)

3.2 LOGISTIC MODELS

3.2.1 Unweighted Logistic Regression

Results of the unweighted univariate logistic regression models can be seen in Table 7. In an unadjusted model, age was the only demographic variable with a statistically significant relationship with marijuana use in the past 30 days ($p < 0.0001$). The odds of marijuana use in 14-15 year olds and 16-17 year olds are about 6.5 and 16 times the odds in 12-13 year olds, respectively. Gender, race/hispanicity, and county metro status did not reveal any statistically significant relationships with marijuana use in their respective univariate models.

Level of self-reported overall health, tendency towards risky behavior, importance of religious beliefs, academic grades, parental relationship with the adolescent, and presence of a mother and/or father figure in the household were personal experience variables that had highly statistically significant relationships with marijuana use ($p < 0.0001$). The odds of marijuana use in adolescents who ranked their overall health as "Fair / Poor" was just over three times the same odds as those who selected "Excellent" health ($p < 0.0001$). Moreover, the odds of marijuana use in those who received a "D Average or Less" in their last academic grading period was almost six and a half times the same odds as those who received an "A+, A, or A- Average" ($p < 0.0001$).

Coverage by any type of health insurance was the only SES variable that did not reveal a statistically significant relationship with marijuana use ($p = 0.50$); all other SES variables had statistically significant relationships. These included total family income ($p < 0.001$),

participation in any government assistance program ($p < 0.001$), and Federal poverty level ranking ($p = 0.02$).

The odds of marijuana use in those who drank alcohol or smoked all or part of a cigarette in the past 30 days were about 18 and 20 times the same odds in those who did not partake in alcohol or cigarette use, respectively ($p < 0.0001$).

Table 7. Results of unweighted univariate logistic regression

Predictor	Odds Ratio	*p-value	95% CI
Age			
Baseline: 12-13 yrs.			
14-15 yrs.S	6.52		[4.92 , 8.64]
16-17 yrs.	15.89	***< 0.0001***	[12.12 , 20.83]
Gender			
Baseline: Male			
Female	0.95	0.40	[0.85 , 1.07]
Race / Hispanicity			
Baseline: White			
Black/African American	0.95		[0.80 , 1.13]
Hispanic	0.94		[0.81 , 1.09]
Other	1.01	0.82	[0.83 , 1.23]
County Metro Status			
Baseline: Large Metro			
Small Metro	1.01		[0.88 , 1.14]
Non-Metro	0.89	0.29	[0.77 , 1.04]
Overall Health			
Baseline: Excellent			
Very Good	1.50		[1.29 , 1.75]
Good	2.29		[1.95 , 2.69]
Fair / Poor	3.34	***< 0.0001***	[2.60 , 4.27]
Tendency Towards Risky Behavior			
Baseline: Never			
Seldom	2.74		[2.27 , 3.31]
Sometimes	5.09		[4.24 , 6.11]
Always	11.37	***< 0.0001***	[9.04 , 14.31]
Importance of Religious Beliefs			
Baseline: Strongly Agree			
Agree	1.96		[1.65 , 2.34]
Disagree	3.74		[3.10 , 4.53]
Strongly Disagree	5.54	***< 0.0001***	[4.59 , 6.69]
Grades from Last Grading Period			
Baseline: A+, A, A-			
B+, B, B-	1.80		[1.52 , 2.14]
C+, C, C-	3.85		[3.23 , 4.60]
D or Less	6.44		[5.13 , 8.08]
Non-Letter Grades	1.76	***< 0.0001***	[1.19 , 2.60]

Table 7. Continued

Parents Said They Were Proud			
Baseline: Always			
Sometimes	1.74		[1.52 , 1.99]
Seldom	2.92		[2.48 , 3.45]
Never	3.77	***< 0.0001***	[3.03 , 4.69]
Mother in Household			
Baseline: Yes			
No	1.90	***< 0.0001***	[1.61 , 2.25]
Father in Household			
Baseline: Yes			
No	1.44	***< 0.0001***	[1.28 , 1.63]
Total Family Income			
Baseline: < $20,000			
$20,000 - $49,999	0.81		[0.69 , 0.95]
$50,000 - $74,999	0.77		[0.64 , 0.92]
≥ $75,000	0.70	***< 0.001***	[0.59 , 0.82]
Use of Govt. Assistance Program			
Baseline: Yes			
No	0.72	***< 0.0001***	[0.64 , 0.82]
Poverty Level			
Baseline: Income ≤ 1x Threshold			
Income >1x & ≤ 2x Threshold	0.94		[0.78 , 1.11]
Income > 2x Threshold	0.83	***0.02***	[0.72 , 0.96]
Health Insurance Coverage			
Baseline: Yes			
No	1.09	0.50	[0.85 , 1.38]
Cigarette Use in Past 30 Days			
Baseline: No			
Yes	20.66	***< 0.0001***	[18.02 , 23.68]
Alcohol Use in Past 30 Days			
Baseline: No			
Yes	18.13	***< 0.0001***	[15.94 , 20.62]

****Bold and italicized*** values indicate significance at the $\alpha = 0.05$ level

In order to produce our final unweighted multivariate logistic regression model, our predefined process of elimination and stopping rule was used. [Intermediate models can be seen in Table 15 in Appendix A]. The removed variables were, in order of elimination, coverage by any type of health insurance, comparison to Federal poverty level, and use of any government assistance program.

The results of the final unweighted multivariate logistic regression model (Table 8) indicated that the odds of marijuana use in the past 30 days in 16 and 17 year olds is almost 5.5 times the same odds in 12 and 13 year olds ($p < 0.0001$). The trend showed increasing odds with age. The same increasing trend in odds of marijuana use was seen with an adolescent's tendency towards risky behavior. The odds of drug use in participants who "Always" test themselves with risky behavior were over three times the odds of use in those who "Never" test themselves with risky behavior ($p < 0.0001$). Furthermore, lower academic grades were strongly associated with an increase in odds of marijuana use ($p < 0.0001$). Controlling for total family income, the other SES variables (health insurance, poverty level, and government assistance) were no longer statistically significant. In the corresponding univariate models, government assistance and poverty level were statistically significant at the $\alpha = 0.05$ level ($p < 0.001$, $p = 0.02$, respectively), while health insurance was not ($p = 0.50$).

Similar to the univariate results, we saw the largest increase in odds of marijuana use for adolescents who smoked cigarettes or consumed alcohol in the past 30 days compared to those who did not. Specifically, the odds of marijuana use in those who drank alcohol or smoked all or

part of a cigarette in the past 30 days are almost 6 and 7 times the same odds in those who did not engage, respectively ($p < 0.0001$), controlling for other variables.

Additionally, as in the unweighted univariate model, gender was not a statistically significant predictor in the multivariate model at the $\alpha = 0.05$ level, but was included in the model based on information from the literature and personal assumption. Presence of a father figure was also not statistically significant at the $\alpha = 0.05$ level in the final model, but it maintained a position in the model based on our stopping rule of $\alpha = 0.1$.

Table 8. Results of unweighted backwards-elimination logistic regression (Final Model)

Predictor	Odds Ratio	*p-value	95% CI
Age			
Baseline: 12-13 yrs.			
14-15 yrs.	3.52		[2.50 , 4.94]
16-17 yrs.	5.42	***< 0.0001***	[3.89 , 7.55]
Gender			
Baseline: Male			
Female	0.95	0.52	[0.82 , 1.11]
Race / Hispanicity			
Baseline: White			
Black/African American	1.78		[1.40 , 2.27]
Hispanic	1.02		[0.83 , 1.26]
Other	1.44	***< 0.0001***	[1.12 , 1.85]
County Metro Status			
Baseline: Large Metro			
Small Metro	0.91		[0.77 , 1.08]
Non-Metro	0.64	***< 0.001***	[0.52 , 0.79]
Overall Health			
Baseline: Excellent			
Very Good	1.16		[0.96 , 1.40]
Good	1.44		[1.17 , 1.79]
Fair / Poor	1.59	***< 0.01***	[1.13 , 2.25]
Tendency Towards Risky Behavior			
Baseline: Never			
Seldom	1.71		[1.36 , 2.15]
Sometimes	2.13		[1.70 , 2.67]
Always	3.22	***< 0.0001***	[2.38 , 4.37]
Importance of Religious Beliefs			
Baseline: Strongly Agree			
Agree	1.43		[1.16 , 1.77]
Disagree	2.05		[1.61 , 2.60]
Strongly Disagree	2.75	***< 0.0001***	[2.17 , 3.50]
Grades from Last Grading Period			
Baseline: A+, A, A-			
B+, B, B-	1.23		[1.00 , 1.51]
C+, C, C-	1.87		[1.49 , 2.35]
D or Less	2.53		[1.86 , 3.44]
Non-Letter Grades	2.12	***< 0.0001***	[1.31 , 3.42]

Table 8. Continued

Parents Said They Were Proud			
Baseline: Always			
Sometimes	1.18		[1.00 , 1.41]
Seldom	1.33		[1.07 , 1.65]
Never	1.61	***< 0.01***	[1.20 , 3.42]
Mother in Household			
Baseline: Yes			
No	1.35	***0.01***	[1.07 , 1.69]
Father in Household			
Baseline: Yes			
No	1.18	0.07	[0.99 , 1.41]
Total Family Income			
Baseline: < $20,000			
$20,000 - $49,999	0.72		[0.58 , 0.89]
$50,000 - $74,999	0.89		[0.69 , 1.15]
≥ $75,000	0.94	***< 0.01***	[0.74 , 1.21]
Cigarette Use in Past 30 Days			
Baseline: No			
Yes	5.79	***< 0.0001***	[4.85 , 6.99]
Alcohol Use in Past 30 Days			
Baseline: No			
Yes	6.84	***< 0.0001***	[5.85 , 8.00]
Constant	0.002	***< 0.001***	[0.001 , 0.003]

****Bold and italicized*** values indicate significance at the $\alpha = 0.05$ level

Multiple model diagnostic procedures were performed to determine the fit of the unweighted multivariate model. Calculation of the uncentered VIF's (Table 9) revealed values notably less than 10 for all variables indicating that multicollinearity (correlation between two or more predictors) was not a cause for concern. The results of the Hosmer-Lemeshow Goodness-of-Fit test (Table 15) stated that the model was very well fit to the data ($p > 0.99$). Information from the linktest (Table 15) confirmed that relevant predictors were selected for the final model and that there was no specification error ($p < 0.001$).

Table 9. Uncentered VIF's for final unweighted logistic regression model

Variable	VIF
Age	
Baseline: 12-13 yrs.	
14-15 yrs.	2.08
16-17 yrs.	2.37
Gender	
Baseline: Male	
Female	1.95
Race / Hispanicity	
Baseline: White	
Black/African American	1.36
Hispanic	1.42
Other	1.17
County Metro Status	
Baseline: Large Metro	
Small Metro	1.71
Non-Metro	1.48
Overall Health	
Baseline: Excellent	
Very Good	2.21
Good	1.64
Fair / Poor	1.14
Tendency Towards Risky Behavior	
Baseline: Never	
Seldom	2.09
Sometimes	1.93
Always	1.21
Importance of Religious Beliefs	
Baseline: Strongly Agree	
Agree	2.15
Disagree	1.50
Strongly Disagree	1.41
Grades from Last Grading Period	
Baseline: A+, A, A-	
B+, B, B-	2.26
C+, C, C-	1.75
D or Less	1.24
Non-Letter Grades	1.09

Table 9. Continued

Parents Said They Were Proud	
Baseline: Always	
Sometimes	1.59
Seldom	1.26
Never	1.12
Mother in Household	
Baseline: Yes	
No	1.12
Father in Household	
Baseline: Yes	
No	1.61
Total Family Income	
Baseline: < $20,000	
$20,000 - $49,999	2.24
$50,000 - $74,999	1.75
≥ $75,000	2.53
Cigarette Use in Past 30 Days	
Baseline: No	
Yes	1.35
Alcohol Use in Past 30 Days	
Baseline: No	
Yes	1.45
Mean VIF	*1.65*

Diagnostic plots can be seen in the four figures below. A scatter plot of the standardized Pearson residuals (Figure 3) revealed a few potential outliers. The three values above ten were not cause for great concern, but they were worth investigating. Similarly, the plot of deviance residuals (Figure 4) showed only slight variation in the residuals indicating that outlying values were likely not a problem. The plot of leverages (Figure 5) shows a few points that may have high values. Generally, points greater than 2k/n (where k is the number of predictors and n is the number of observations in the model) signify high leverage. For this model, 2k/n = 0.004. Using this cut point, the vast majority of values would be considered high leverage. Instead, the general pattern of the values was considered and a cut off of 0.025 was used. Figure 6 displays a plot of the influence values. As a general rule, values greater than one are characterized as high influence points. Based on the values alone, none of the points have high leverage. On the other hand, based on the pattern in the plot, the handful of points with DBETA values greater than 0.1 should be investigated to determine their true influence on the model.

Re-running the final model removing points with the highest standardized residual values and again removing points with the highest influential values, showed that, in both cases, there was almost no change in our estimates and no change at all in their level of statistical significance indicating that our model is not sensitive to these points. [See Table 18 in Appendix C and Table 19 in Appendix D for these additional models.]

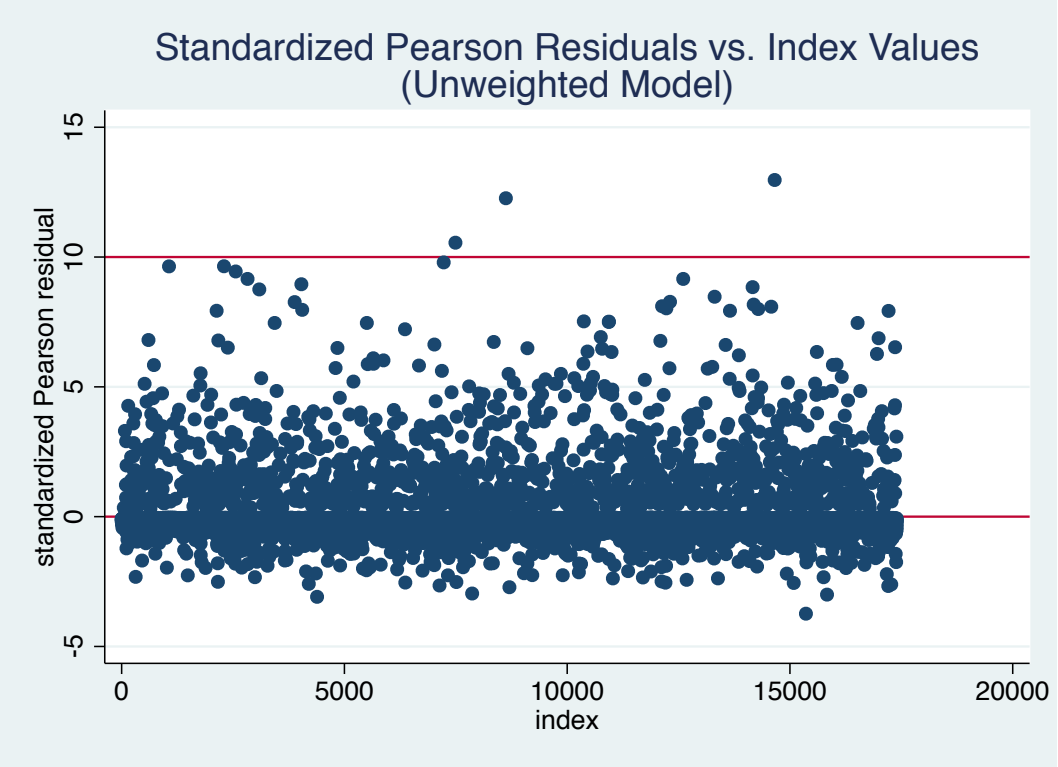

Figure 3. Scatter plot of standardized Pearson residuals (Unweighted)

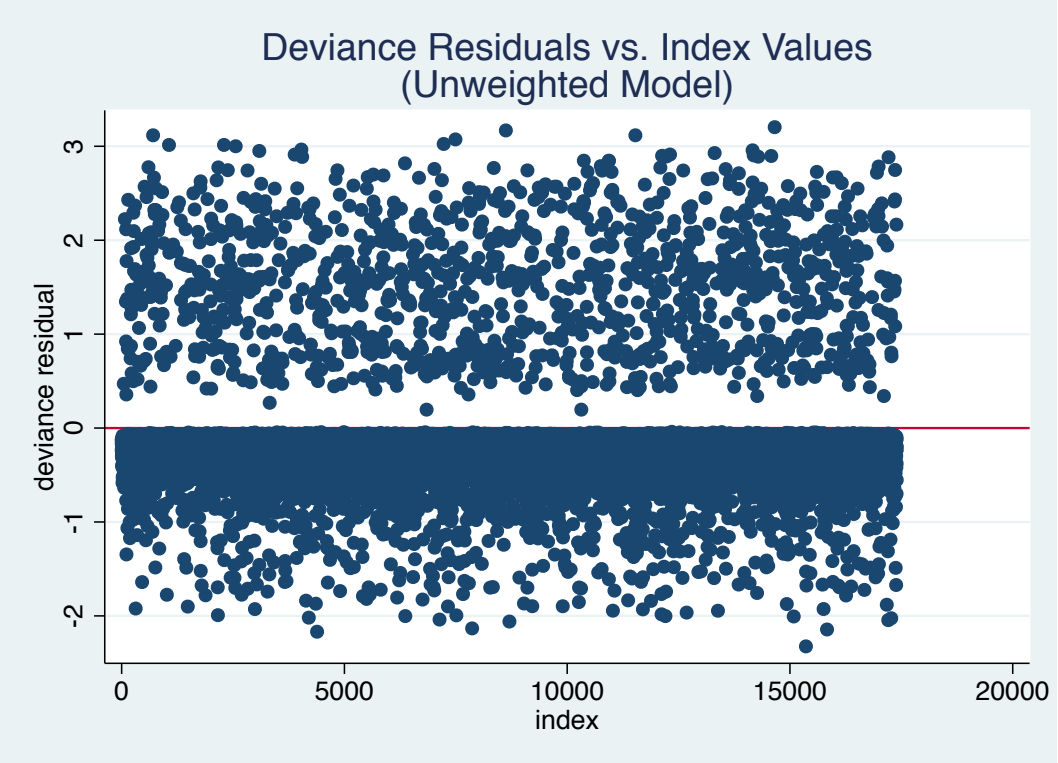

Figure 4. Scatter plot of deviance residuals (Unweighted)

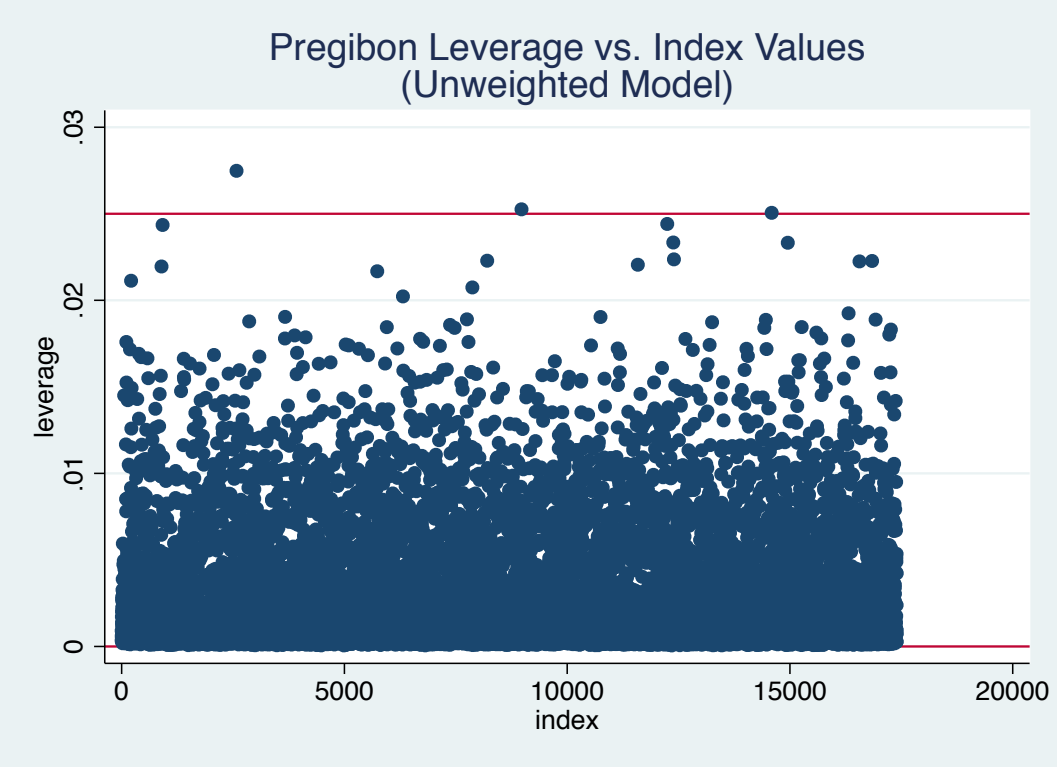

Figure 5. Scatter plot of Pregibon leverage (Unweighted)

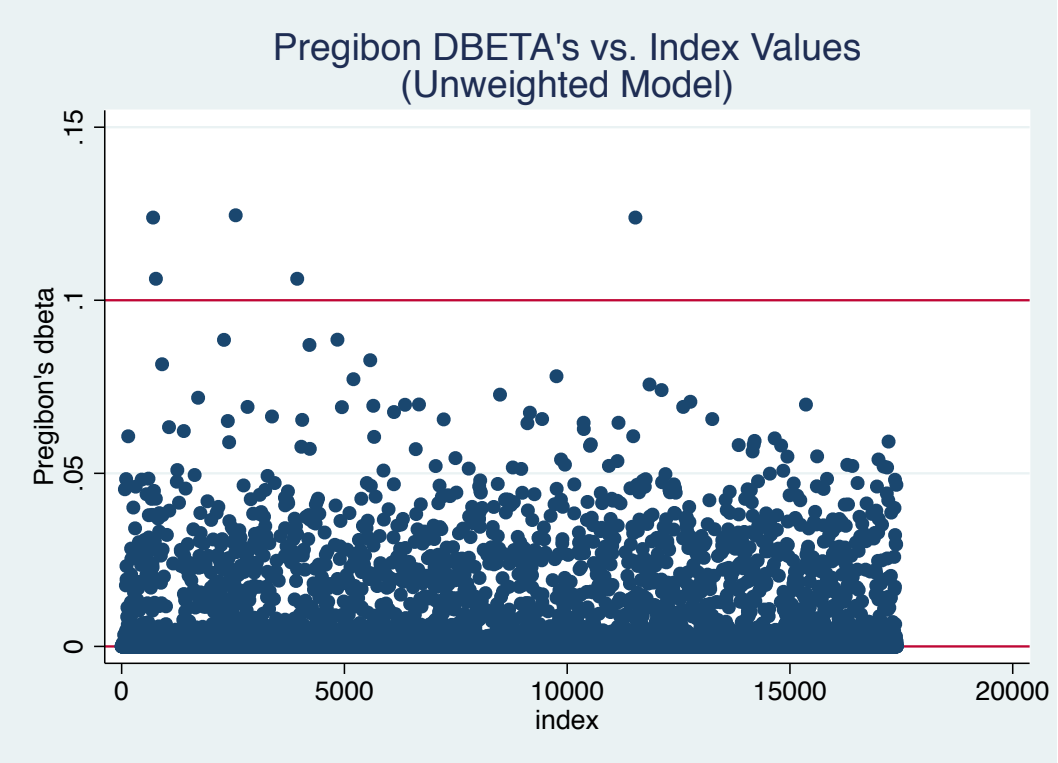

Figure 6. Scatter plot of Pregibon DBETA's (Unweighted)

3.2.2 Weighted Logistic Regression

The sample weighted univariate logistic regression models (Table 10) revealed the same statistically significant ($\alpha = 0.05$) variables found in the unweighted univariate models, with one exception. In an unadjusted, weighted model, race/hispanicity was a statistically significant predictor of marijuana use in the past 30 days. This relationship was not seen in the unweighted model. In general, the odds ratios in each of the univariate weighted models only varied slightly from those in the univariate unweighted models.

Table 10. Results of weighted univariate logistic regression

Predictor	Odds Ratio	*p-value	95% CI
Age			
Baseline: 12-13 yrs.			
14-15 yrs.	6.86		[4.57 , 10.29]
16-17 yrs.	17.58	***< 0.0001***	[11.88 , 26.00]
Gender			
Baseline: Male			
Female	0.90	0.20	[0.77 , 1.06]
Race / Hispanicity			
Baseline: White			
Black/African American	0.94		[0.74 , 1.20]
Hispanic	0.85		[0.68 , 1.05]
Other	0.63	***< 0.01***	[0.48 , 0.82]
County Metro Status			
Baseline: Large Metro			
Small Metro	1.08		[0.91 , 1.28]
Non-Metro	0.87	0.16	[0.70 , 1.08]
Overall Health			
Baseline: Excellent			
Very Good	1.46		[1.19 , 1.79]
Good	2.28		[1.83 , 2.84]
Fair / Poor	3.11	***< 0.0001***	[2.22 , 4.37]
Tendency Towards Risky Behavior			
Baseline: Never			
Seldom	2.64		[2.04 , 3.43]
Sometimes	4.96		[3.86 , 6.35]
Always	10.77	***< 0.0001***	[7.85 , 14.76]
Importance of Religious Beliefs			
Baseline: Strongly Agree			
Agree	1.88		[1.48 , 2.39]
Disagree	3.45		[2.68 , 4.44]
Strongly Disagree	5.09	***< 0.0001***	[3.94 , 6.57]
Grades from Last Grading Period			
Baseline: A+, A, A-			
B+, B, B-	1.74		[1.37 , 2.20]
C+, C, C-	3.82		[3.00 , 4.86]
D or Less	6.80		[4.95 , 9.33]
Non-Letter Grades	0.98	***< 0.0001***	[0.59 , 1.63]

Table 10. Continued

Parents Said They Were Proud			
Baseline: Always			
Sometimes	1.84		[1.53 , 2.22]
Seldom	3.25		[2.60 , 4.07]
Never	3.68	***< 0.0001***	[2.70 , 5.00]
Mother in Household			
Baseline: Yes			
No	1.99	***< 0.0001***	[1.59 , 2.48]
Father in Household			
Baseline: Yes			
No	1.46	***< 0.0001***	[1.23 , 1.72]
Total Family Income			
Baseline: < $20,000			
$20,000 - $49,999	1.01		[0.81 , 1.26]
$50,000 - $74,999	0.80		[0.62 , 1.03]
≥ $75,000	0.76	***0.01***	[0.61 , 0.95]
Use of Govt. Assistance Program			
Baseline: Yes			
No	0.76	***< 0.01***	[0.64 , 0.89]
Poverty Level			
Baseline: Income ≤ 1x Threshold			
Income >1x & ≤ 2x Threshold	1.06		[0.84 , 1.34]
Income > 2x Threshold	0.84	***0.03***	[0.69 , 1.02]
Health Insurance Coverage			
Baseline: Yes			
No	1.15	0.44	[0.81 , 1.65]
Cigarette Use in Past 30 Days			
Baseline: No			
Yes	21.69	***< 0.0001***	[17.98 , 26.16]
Alcohol Use in Past 30 Days			
Baseline: No			
Yes	17.57	***< 0.0001***	[14.72 , 20.98]

****Bold and italicized*** values indicate significance at the $\alpha = 0.05$ level

The same model building criteria was used for the creation of the weighted multivariate model as for the unweighted version. [Intermediate models can be seen in Table 17 in Appendix B]. The removed variables were, in order of elimination, coverage by any type of health insurance, total family income, comparison to the Federal poverty level, and use of any government assistance program. Contrary to the unweighted model, none of the SES variables remained in the final weighted model. Their corresponding weighted univariate models showed that total family income ($p = 0.01$), government assistance ($p < 0.01$), and poverty level ($p = 0.03$) were statistically significant at the $\alpha = 0.05$ level, while health insurance was not ($p = 0.44$).

The results of the final weighted model (Table 11) revealed very similar results to what was seen in the final unweighted model. Gender was not a statistically significant predictor in the multivariate model at the $\alpha = 0.05$ level ($p = 0.47$), but was included in the model based on information from the literature and personal assumption. Increased age was strongly associated with increased odds of marijuana use ($p < 0.0001$). Likewise, an increased tendency towards risky behavior was also associated with increased odds of marijuana use. The odds of marijuana use in the past 30 days for adolescents who "Strongly Disagree[d]" that their religious beliefs were important to them were more than 2.5 times the same odds as those who "Strongly Agree[d]" ($p < 0.0001$). Although weighting attenuated the odds ratios for cigarette use and alcohol consumption, the results were still extremely statistically significant. Specifically, the odds of marijuana use in those who drank alcohol or smoked all or part of a cigarette in the past 30 days were more than 5.5 and 6 times the same odds in those who did not engage, respectively ($p < 0.0001$).

In general, we saw approximately the same results in both the weighted and unweighted multivariate models with only slight differences in the odds ratios and statistically significant predictors.

Table 11. Results of weighted backwards elimination logistic regression (Final Model)

Predictor	Odds Ratio	*p-value	95% CI
Age			
Baseline: 12-13 yrs.			
14-15 yrs.	3.32		[2.07 , 5.34]
16-17 yrs.	5.37	***< 0.0001***	[3.35 , 8.59]
Gender			
Baseline: Male			
Female	0.93	0.47	[0.76 , 1.14]
Race / Hispanicity			
Baseline: White			
Black/African American	1.72		[1.23 , 2.41]
Hispanic	0.98		[0.72 , 1.33]
Other	1.03	***0.01***	[0.73 , 1.45]
County Metro Status			
Baseline: Large Metro			
Small Metro	1.01		[0.81 , 1.27]
Non-Metro	0.62	***< 0.01***	[0.46 , 0.83]
Overall Health			
Baseline: Excellent			
Very Good	1.06		[0.83 , 1.37]
Good	1.39		[1.04 , 1.87]
Fair / Poor	1.71	***0.04***	[1.00 , 2.90]
Tendency Towards Risky Behavior			
Baseline: Never			
Seldom	1.73		[1.25 , 2.39]
Sometimes	2.17		[1.60 , 2.96]
Always	3.10	***< 0.0001***	[2.02 , 4.76]
Importance of Religious Beliefs			
Baseline: Strongly Agree			
Agree	1.30		[0.98 , 1.74]
Disagree	1.81		[1.33 , 2.44]
Strongly Disagree	2.64	***< 0.0001***	[1.93 , 3.60]
Grades from Last Grading Period			
Baseline: A+, A, A-			
B+, B, B-	1.11		[0.82 , 1.50]
C+, C, C-	1.84		[1.33 , 2.54]
D or Less	2.78		[1.82 , 4.25]
Non-Letter Grades	1.10	***< 0.0001***	[0.57 , 2.14]

Table 11. Continued

Parents Said They Were Proud			
Baseline: Always			
Sometimes	1.38		[1.08 , 1.76]
Seldom	1.51		[1.11 , 2.05]
Never	1.48	***0.01***	[0.93 , 2.35]
Mother in Household			
Baseline: Yes			
No	1.61	***< 0.01***	[1.17 , 2.20]
Father in Household			
Baseline: Yes			
No	1.26	***0.04***	[1.00, 1.59]
Cigarette Use in Past 30 Days			
Baseline: No			
Yes	5.68	***< 0.0001***	[4.30 , 7.49]
Alcohol Use in Past 30 Days			
Baseline: No			
Yes	6.03	***< 0.0001***	[4.79 , 7.59]
Constant	0.002	***< 0.001***	[0.001 , 0.003]

****Bold and italicized*** values indicate significance at the $\alpha = 0.05$ level

Although the literature indicated that the same diagnostic procedures should be used for the weighted multivariate model as used for the unweighted model, these techniques are not available for weighted models in the Stata statistical software. Due to this, it was not possible to evaluate the residual, leverage, and influence values, or perform the Hosmer-Lemeshow Goodness-of-Fit test on the weighted model. Multicollinearity was not an issue based on VIF values less than 10 for all variables (Table 12). Information from the linktest (Table 15) confirmed that relevant predictors were selected for the final model and there was no specification error ($p < 0.001$).

Table 12. Uncentered VIF's for final weighted logistic regression model

Variable	VIF
Age	
Baseline: 12-13 yrs.	
14-15 yrs.	1.98
16-17 yrs.	2.22
Gender	
Baseline: Male	
Female	1.80
Race / Hispanicity	
Baseline: White	
Black/African American	1.37
Hispanic	1.43
Other	1.14
County Metro Status	
Baseline: Large Metro	
Small Metro	1.46
Non-Metro	1.27
Overall Health	
Baseline: Excellent	
Very Good	2.07
Good	1.59
Fair / Poor	1.13
Tendency Towards Risky Behavior	
Baseline: Never	
Seldom	1.89
Sometimes	1.81
Always	1.17
Importance of Religious Beliefs	
Baseline: Strongly Agree	
Agree	2.06
Disagree	1.44
Strongly Disagree	1.35
Grades from Last Grading Period	
Baseline: A+, A, A-	
B+, B, B-	2.18
C+, C, C-	1.67
D or Less	1.23
Non-Letter Grades	1.08

Table 12. Continued

Parents Said They Were Proud	
Baseline: Always	
Sometimes	1.59
Seldom	1.25
Never	1.11
Mother in Household	
Baseline: Yes	
No	1.11
Father in Household	
Baseline: Yes	
No	1.46
Cigarette Use in Past 30 Days	
Baseline: No	
Yes	1.33
Alcohol Use in Past 30 Days	
Baseline: No	
Yes	1.45
Mean VIF	*1.52*

4.0 CONCLUSIONS

4.1 DISCUSSION

We constructed two primary models in order to analyze which risk factors are associated with marijuana use in the past 30 days. Both the sample weighted and unweighted multivariate logistic models revealed similar results, although it was difficult to make a complete comparison between the two models due to the lack of availability of some diagnostic procedures for the weighted model in the Stata statistical software.

In general, the models shared similar results. A direct comparison of the final models is shown in Table 14. Odds ratios for most variables in the unweighted and weighted models were within about 0.2 of each other. The main distinction was in the level of statistical significance for each of the variables between the models. One of the most obvious disparities was seen in the race/hispanicity variable. In the unweighted model, this variable is highly statistically significant, while in the weighted model, it is only significant at the $\alpha = 0.05$ level, and a difference of more than 0.4 is seen between the unweighted and weighted model for the "Other" race group. This may be due to misrepresentation of each of the race groups in the unweighted model. The response options for the associated question were "Non-Hispanic White", "Non-Hispanic Black/African American", "Non-Hispanic Native American/AK Native", "Non-Hispanic Native

HI/Other Pacific Islander", "Non-Hispanic Asian", "Non-Hispanic more than one race", and "Hispanic". Because the survey specified "non-Hispanic" for all but one option, they may have had a higher response rate for the "Hispanic" group than they would have if "non-Hispanic" were not a stipulation. Tabulation of the recoded race variable used in our models showed an unexpectedly high number of "Hispanic" participants in comparison to the other groups (White [n = 9863], Black [n = 2328], Hispanic [n = 3457], Other [n = 1751]). With "White" as a baseline, the odds ratios seen in both models for "Black/African Americans" and "Hispanics" are 1.78 and 1.02 in the unweighted model compared to 1.72 and 0.98 in the weighted model, respectively. These estimates do not differ by more than 0.06. On the other hand, a noticeable difference is seen in the "Other" race group with odds ratios of 1.44 and 1.03 for the unweighted and weighted models, respectively. Again, this is likely due to the formulation of the question.

Table 13 shows a tabulation of the distribution of each race in the unweighted and weighted models. Since the values are similar, it indicates that the weights are not making a large adjustment. The United States Census Bureau allows for selection of multiple races, which makes it difficult to compare the percentages from the survey to national values from the census (USA QuickFacts, 2014).

Table 13. Tabulation of unweighted and weighted race distribution

Race	Unweighted	Weighted
White	9,779 (56.8%)	1,3624,884 (55.3%)
Black / African American	2,281 (13.3%)	3,458,690.3 (14.0%)
Hispanic	3418 (19.9%)	5,425,288.5 (22.0%)
Other	1,728 (10.0%)	2,133,079.9 (8.7%)
TOTAL	17,206 (100.0%)	24,641,943 (100.0%)

In order to resolve the differences in interpretation that likely arose from this question, survey investigators should allow multiple responses for race and ethnicity. If "Hispanic" was its own separate response and could be combined such that a participant could respond with "Hispanic" and "White" or "Black/African American", more complete information would be obtained.

Another interesting difference was that the presence of a mother figure but not a father figure in the household was statistically significant in the unweighted multivariate model ($p = 0.01$, OR = 1.35 and $p = 0.07$, OR = 1.18, respectively). The weighted model showed that both the presence of a mother and a father figure were statistically significant in the model but at different levels of significance ($p < 0.01$, OR = 1.61 and $p = 0.04$, OR = 1.26, respectively). In the weighted model, the increase in odds of marijuana use if a father figure is not present is not even as much as the increase in the odds if a mother figure is not present in the either model.

In addition, we saw a difference between the weighted and unweighted models in the odds of marijuana use among students who received "Non-Letter Grades". Based on the unweighted model, the odds of marijuana use among these students are 2.12 times the odds of those who received an "A+, A-, or A Average", second only to those who received a "D Average or Less". On the other hand, in the weighted model, the odds of marijuana use among "Non-Letter Grade" students is only 1.10 times the odds of those who received an "A+, A-, or A Average", the lowest odds of all other grading levels. This difference could be the result of the broad spectrum of students not receiving letter grades, which may include adolescents who are homeschooled or enrolled in a Montessori or alternative education program. If these findings are to be explored

further, survey investigators would need to gather more information on the students who receive "Non-Letter Grade[s]" to better define this section of the sample.

Also, gender was not statistically significant at any level of significance in either model. Both models had odds ratios fairly close to one, which would indicate the same odds of marijuana use between males and females.

In both models, an increase in odds of marijuana use was associated with an increase in age, a decrease in overall health, an increased tendency towards risky behavior, and a decrease in importance of religious beliefs. For age, we saw odds ratios of 3.52 and 3.32 in 14-15 year olds and 5.42 and 5.37 in 16-17 year olds in the unweighted and weighted models, respectively. The increased odds of marijuana use in adolescents who reported "Very Good" health was an average of about 0.5 less than those who reported "Fair / Poor" health in both models. Furthermore, there in an average increase in the odds of marijuana use of about 1.5 between those who "Seldom" test themselves with risky behavior compared to those who "Always" test themselves in both models. Adolescents who felt they "Strongly Disagree[d]" with the statement that religious beliefs were an important part of their lives had the greatest increase in odds among any other response, followed by "Disagree" and "Agree".

Each model showed the greatest increase in odds of marijuana use with cigarette use and alcohol consumption. Further, alcohol consumption showed the greatest increase in odds of marijuana use among all other covariates for both models. In the unweighted model, the odds of marijuana use in adolescents who consume alcohol are almost seven times the odds in those who do not.

Cigarette use had the second highest increase in odds overall. For both models, the odds of marijuana use in adolescents who smoked all or part of a cigarette in the past 30 days were almost six times the odds in those who did not. This provides support for the idea that alcohol and cigarettes may be "gateway" drugs.

In terms of diagnostics, there were almost identical results between the two models. Despite the lack of residual, leverage, and influence values in the weighted models, we were able to compare the VIF's and the linktests. All VIF's in both models were considerably less than 10. In fact, both models had VIF's less than 3 with mean VIF's 1.52 and 1.65 for the weighted and unweighted models, respectively.

We were initially interested in including information about the participants' mental health, but none of the sections, questions, or related variables seemed to appropriately convey an overall mental health status. Although a series of standardized depression symptom questions were covered, access to this section depended on responses in previous sections resulting in over 80% missing values for the sample. Additionally, mental health treatment information was collected, but use of treatment is not fully indicative of a particular state of mental health. If a treatment variable were used to indicate mental health, it would assume that anyone diagnosed with mental health illness searched out and obtained treatment, which is not the case. Furthermore, many people with mental illness go undiagnosed. A standardized mental health questionnaire would have been required for this survey to grasp the scope of mental illness within adolescents. These types of questionnaires collect information on number of psychological symptoms, which can be

pooled together to determine mental standing based on a scale of symptoms. One of the most commonly used surveys is used for determination of the total number of depressive symptoms.

Table 14. Comparison of final unweighted and weighted backwards elimination logistic regression models

Variable	Unweighted Final Model Odds Ratio [95% Confidence Interval]	Weighted Final Model Odds Ratio [95% Confidence Interval]
Age		
Baseline: 12-13 yrs.	****	****
14-15 yrs.	3.52	3.32
	[2.50 , 4.94]	[2.07 , 5.34]
16-17 yrs.	5.42	5.37
	[3.89 , 7.55]	[3.35 , 8.59]
Gender		
Baseline: Male		
Female	0.95	0.93
	[0.82 , 1.11]	[0.76 , 1.14]
Race / Hispanicity		
Baseline: White	****	*
Black/African American	1.78	1.72
	[1.40 , 2.27]	[1.23 , 2.41]
Hispanic	1.02	0.98
	[0.83 , 1.26]	[0.72 , 1.33]
Other	1.44	1.03
	[1.12 , 1.85]	[0.73 , 1.45]
County Metro Status		
Baseline: Large Metro	***	**
Small Metro	0.91	1.01
	[0.77 , 1.08]	[0.81 , 1.27]
Non-Metro	0.64	0.62
	[0.52 , 0.79]	[0.46 , 0.83]
Overall Health		
Baseline: Excellent	**	*
Very Good	1.16	1.06
	[0.96 , 1.40]	[0.83 , 1.37]
Good	1.44	1.39
	[1.17 , 1.79]	[1.04 , 1.87]
Fair / Poor	1.59	1.71
	[1.13 , 2.25]	[1.00 , 2.90]

Table 14. Continued

Tendency Towards Risky Behavior		
Baseline: Never	****	****
Seldom	1.71	1.73
	[1.36 , 2.15]	[1.25 , 2.39]
Sometimes	2.13	2.17
	[1.70 , 2.67]	[1.60 , 2.96]
Always	3.22	3.10
	[2.38 , 4.37]	[2.02 , 4.76]
Importance of Religious Beliefs		
Baseline: Strongly Agree	****	****
Agree	1.43	1.30
	[1.16 , 1.77]	[0.98 , 1.74]
Disagree	2.05	1.81
	[1.61 , 2.60]	[1.33 , 2.44]
Strongly Disagree	2.75	2.64
	[2.17 , 3.50]	[1.93 , 3.60]
Grades from Last Grading Period		
Baseline: A+, A, A-	****	****
B+, B, B-	1.23	1.11
	[1.00 , 1.51]	[0.82 , 1.50]
C+, C, C-	1.87	1.84
	[1.49 , 2.35]	[1.33 , 2.54]
D or Less	2.53	2.78
	[1.86 , 3.44]	[1.82 , 4.25]
Non-Letter Grades	2.12	1.10
	[1.31 , 3.42]	[0.57 , 2.14]
Parents Said They Were Proud		
Baseline: Always	**	*
Sometimes	1.18	1.38
	[1.00 , 1.41]	[1.08 , 1.76]
Seldom	1.33	1.51
	[1.07 , 1.65]	[1.11 , 2.05]
Never	1.61	1.48
	[1.20 , 3.42]	[0.93 , 2.35]
Mother in Household		
Baseline: Yes	*	**
No	1.35	1.61
	[1.07 , 1.69]	[1.17 , 2.20]

Table 14. Continued

Father in Household		
Baseline: Yes		*
No	1.18	1.26
	[0.99 , 1.41]	[1.00, 1.59]
Total Family Income		
Baseline: < $20,000	**	
$20,000 - $49,999	0.72	
	[0.58 , 0.89]	
$50,000 - $74,999	0.89	---
	[0.69 , 1.15]	
≥ $75,000	0.94	
	[0.74 , 1.21]	
Cigarette Use in Past 30 Days		
Baseline: No	****	****
Yes	5.79	5.68
	[4.85 , 6.99]	[4.30 , 7.49]
Alcohol Use in Past 30 Days		
Baseline: No	****	****
Yes	6.84	6.03
	[5.85 , 8.00]	[4.79 , 7.59]
Constant		
	***	***
	0.002	0.002
	[0.001 , 0.003]	[0.001 , 0.003]

*Significant at 0.05
**Significant at < 0.01
***Significant at < 0.001
****Significant at < 0.0001

Table 15. Comparison of final unweighted and weighted backwards-elimination logistic regression model diagnostics

	Unweighted Model	Weighted Model
n	15136	15136
Pseudo R^2	0.37	0.36
Hosmer-Lemeshow GOF	> 0.99	---
Linktest [p-value of _hatsq]	< 0.001	< 0.001

Our models reinforce many of the results stated in the literature. As Newcomb, Nation, and Kilpatrick found in their work, our results strengthened the argument that academic grades, parental relationships, risk seeking behavior, religious commitment, alcohol use, personal experiences, and race are factors associated with the initiation and/or continuation of marijuana use. Our data also support information about marijuana and nationwide drug use trends published by the National Institute on Health. In particular, our models highlighted the increased odds of marijuana use with increased adolescent age. This relationship with age may be related to the decrease in perceived risk of the drug that researchers believe is associated with amplified use in the past decade. This may be due to the increased access following the legalization of medical and personal use of marijuana.

The relationships seen in our research allow us to have a better understanding of which adolescents are at higher risk of using marijuana. Prevention efforts should primarily target adolescents who are already experimenting with alcohol consumption and cigarettes. Further, there should be a focus on students who struggle academically and/or have a tendency towards risky behavior. Based on the Monitoring the Future study, older adolescents reported a decrease in the perceived risk of marijuana in the past decade, and 81% of 12^{th} graders reported that it was "Fairly Easy" or "Very Easy" to obtain the drug (Johnston, 2010). Thus, a strategy for

intervention efforts should be put in place particularly for older adolescents who are more likely to have used marijuana than their younger counterparts. Early prevention and treatment efforts are of utmost importance as the United States is reported to have a large "treatment gap" for all drug related issues. In 2012, about 8.9% of Americans (23.1 million) required treatment, but only about 1% (2.5 million) received the necessary care ("Drug Facts: Nationwide Trends", 2014). The inability to fully evaluate treatment in the study is an unfortunate limitation. Because marijuana is the most commonly used illicit drug, early prevention and treatment protocols could help close this gap.

The minor differences between the final weighted and unweighted models are not surprising based on the controversy in the literature. Although a complex sampling design was incorporated in the survey, we believe sample weighting in the modeling process is still necessary. Formation of some questions resulted in some significant differences between the two models, and two different models were selected when weighting was utilized. Based on our criteria for determining the better-fit model, we found that the magnitude of the difference for some odds ratios was rather large which indicated selection of the weighted model may be more appropriate. The variance estimates for both models were quite similar; there was only a difference of about 0.2 for any variable between the weighted and unweighted model. Neither the VIF's nor the linktests revealed any differences based on whether a weighted model was used. Both models had VIF's much lower than 10 and both linktests were statistically significant. Thus, on the basis of the magnitude of the difference of the odds ratios, we select the weighted models as the more appropriate fit for these data.

4.2 LIMITATIONS

Multiple factors determine to whom these results can be generalized. One main concern is incomplete model specification because of the lack of a meaningful mental health status variable. Furthermore, gender was a binary variable (male/female) that did not take into account how each individual may have self-identified him or herself. Availability of more responses would have allowed investigators to capture differences between gender minority (transgender/gender nonconforming and have a gender different from their sex at birth) and cisgender (gender identity or expression matching their assigned sex at birth) adolescents. One study found that gender minority youth had increased odds of marijuana used compared to cisgender youth (Reisner, 2014).

The exclusion criteria for entrance into the study somewhat affect the generalizability of the results, as well (USDHHS: Codebook, 2012). Since people in institutionalized group quarters were excluded from the study, people in the hospital or in treatment centers were not included. Therefore, those receiving treatment for drug abuse at the time of the study were not taken into account.

Bias related to parental consent for an adolescent to be included in the study may have occurred in the sampling process of data collection. Those parents who chose not to allow their children to be involved in the survey may be fundamentally different than those who did, which likely resulted in a difference in adolescents sampled.

Honesty and ability to correctly recall information is always an element that must be considered when working with survey or interview data (USDHHS: Codebook, 2012). More specifically, full understanding of what each question asks is a potential issue worth considering when working with an adolescent sample. For example, in terms of our risk variable, we do not know how each participant interpreted the temporality of the question. Some responders may be more impulsive while under the influence of marijuana versus when they are sober, but we do not have any indication of that based on the survey question. Likewise, it is difficult to gauge how people will respond to self-report variables such as ranking of their overall health. For two people in the same health standing, one may report "Very Good" health while the other may report "Fair". The matter of perception is important to understand in these types of questions. For the sake of our interest, though, perceived health may be more valuable than physical health. Moreover, participants may have experienced language barriers. The survey was only offered in English and Spanish, but participants were drawn from a variety of ethnic backgrounds. Thus, the assumption is made that all participants are sufficiently literate in one of the two available languages.

In terms of model assumptions, we assumed independence, but due to the sampling methods up to two people could have been sampled from one household. Thus, if two people (siblings) were sampled from one household, they were likely biologically related, and a potential violation of the independence assumption. The chances of this happening are not ignorable since adolescents and young adults were oversampled. Furthermore, some dependencies were induced due to the hierarchical sampling design. These dependencies are failures in the independence assumption that would not have occurred if a true random sampling technique had been incorporated in the design.

Large sample sizes for both the weighted and unweighted models may have had an effect on the statistical significance found. The unweighted model had 15,136 observations; the weighted model had the same number of observations though they represented 21,679,225 people in the population. One concern of very large sample sizes is a great increase in power, which may allow for discovery of false significant results. Because we saw differences in significance between the two models, and not all variables were statistically significant we do not believe that this was an issue.

With any type of survey, we also encounter the issue of cross-sectional versus longitudinal data (USDHHS: Codebook, 2012). Because each participant is interviewed at one point in time, the results can only be interpreted as the prevalence of marijuana use at that one time point. This may cause slight bias in the outcome. For instance, participants who had only used marijuana, smoked cigarettes, and/or consumed alcohol once in their lives and do not plan to partake in any of those acts again may have responded positively for this survey. Due to the timing of the survey and the coding of the variables, though, if their one time of use was in the past 30 days they are considered "users" for the sake of the model. Because the same sample is not surveyed every year, there is no follow up information on any of the participants that would allow for some kind of longitudinal analysis.

4.3 FUTURE WORK

This analysis generated interesting topics for potential future consideration. Further model specification would be a valuable addition to the current results. Consideration of questions/variables about each adolescent's family history of drug abuse and mental health and incorporation of some kind of gold-standard mental health rating would be extraordinarily beneficial based on supporting outcomes seen in the literature. Reevaluation of which variables should be forced into each model might result in different models, as well. In particular, all of our models both weighted and unweighted for univariate and multivariate analysis did not show a statistically significant association between gender and marijuana use in the past 30 days. Thus, it would be worth using model-building criteria that allowed gender to be removed or to create gender specific models. Further, this analysis did not investigate the additivity assumption, which considers interactions between variables. Doing so in future models may attain different results.

It would also be interesting to include a variable that takes into account other drug use within the participant's household. Specifically, does parental or sibling drug use have a stronger association with an adolescent's marijuana use? Does it depend what types of drugs the adolescent is exposed to by peers or parental figures?

Utilization of mixed models would allow for incorporation of both fixed and random effects in the model that would be useful in accounting for the clustering that resulted from the sampling design.

Although it is not currently part of this survey and the associated data, a longitudinal analysis would be advantageous in understanding long-term trends in adolescent marijuana use and grasping the relationship between drug use and the time-varying risk factors.

Finally, creation of similar models for other illicit street and prescription drug use would allow comparison of risk factors across a variety of drugs including cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives, and all of these classifications could be investigated using the same survey and data set. If the resources were to become available, risk factors for each of the outcomes could be compared by country, as well.

4.4 CONCLUDING REMARKS

Irrespective of which model was used (weighted or unweighted), similar results were achieved. We saw that in both models, age, race, county metro status, self-reported overall health, tendency towards risky behavior, importance of religious views, academic grades, positive reinforcement from parents, presence of a mother figure in the household, cigarette use, and alcohol consumption were statistically significant predictors of marijuana use in the past 30 days. Initial univariate analysis revealed that many of the SES variable were statistically significant factors in marijuana use, but when adjusting for all other elements this relationship was no longer seen. Despite the complex sampling technique used for the survey, we determined that the weighted model is more appropriate for the data compared to the unweighted model.

4.5 PUBLIC HEALTH IMPACT

In order to prevent further escalation in adolescent marijuana use, it is important for public health professionals to understand which risk factors increase a minor's probability of trying and/or abusing the drug. By identifying these elements, doctors and treatment centers can apply specialized patient care to their practice, and public health professionals can work with policy makers to create and apply preventative measures and implement the development of treatment programs within the appropriate communities.

APPENDIX A: INTERMEDIATE UNWEIGHTED BACKWARDS-ELIMINATION LOGISTIC REGRESSION MODELS

Model 1 is the full model. Each successive model in the next model is the backwards-elimination sequence (pr = 0.1). All models contain the Age, Gender, and Race/Hispanicity variables.

Table 16. Unweighted backwards-elimination logistic regression models

Covariates	**Model 1**		**Model 2**		**Model 3**	
	OR [95% CI]	*p-value	OR [95% CI]	*p-value	OR [95% CI]	*p-value
Age						
Baseline: 12-13 yrs.						
14-15 yrs.	3.54 [2.51 , 5.00]		3.52 [2.50 , 4.94]		3.52 [2.51 , 2.95]	
16-17 yrs.	5.50 [3.93 , 7.69]	***< 0.0001***	5.45 [3.92 , 7.60]	***< 0.0001***	5.454 [3.92 , 7.60]	***< 0.0001***

Table 16. Continued

Gender						
Baseline: Male						
Female	0.94		0.95		0.95	
	[0.81 , 1.10]	0.45	[0.82 , 1.11	0.52	[0.82 , 1.11]	0.54
Race / Hispanicity						
Baseline: White						
Black/African American	1.79		1.75		1.74	
	[1.40 , 2.29]		[1.37 , 2.24]		[1.37 , 2.23]	
Hispanic	1.05		1.03		1.02	
	[0.85 , 1.30]		[0.83 , 1.27]		[0.83 , 1.26]	
Other	1.46		1.44		1.43	
	[1.13 , 1.89]	***<0.0001***	[1.12 , 1.85]	***< 0.0001***	[1.11 , 1.85]	***< 0.0001***
County Metro Status						
Baseline: Large Metro						
Small Metro	0.91		0.92		0.91	
	[0.77 , 1.09]		[0.77 , 1.09]		[0.77 , 1.08]	
Non-Metro	0.63		0.64		0.63	
	[0.51 , 0.78]	***< 0.001***	[0.51 , 0.78]	***< 0.001***	[0.51 , 0.78]	***< 0.001***
Overall Health						
Baseline: Excellent						
Very Good	1.15		1.16		1.16	
	[0.95 , 1.38]		[0.96 , 1.40]		[0.96 , 1.40]	
Good	1.46		1.44		1.44	
	[1.18, 1.81]		[1.17 , 1.78]		[1.17 , 1.78]	
Fair / Poor	1.58		1.58		1.59	
	[1.12 , 1.24]	***< 0.01***	[1.12 , 2.24]	***< 0.01***	[1.13 , 2.22]	***< 0.01***

Table 16. Continued

Tendency Towards Risky Behavior						
Baseline: Never						
Seldom	1.68		1.71		1.72	
	[1.33 , 2.11]		[1.36 , 2.16]		[1.37 , 2.16]	
Sometimes	2.10		2.13		2.13	
	[1.68 , 2.64]		[1.70 , 2.66]		[1.70 , 2.66]	
Always	3.20		3.22		3.21	
	[2.35 , 4.35]	***< 0.0001***	[2.37 , 4.36]	***< 0.0001***	[2.37 , 4.35]	***< 0.0001***
Importance of Religious Beliefs						
Baseline: Strongly Agree						
Agree	1.43		1.43		1.43	
	[1.16 , 1.78]		[1.16 , 1.77]		[1.16 , 1.77]	
Disagree	2.06		2.03		2.04	
	[1.62 , 2.62]		[1.60 , 2.58]		[1.61 , 2.59]	
Strongly Disagree	2.82		2.75		2.75	
	[2.21 , 3.58]	***< 0.0001***	[2.16 , 3.50]	***< 0.0001***	[2.16 , 3.49]	***< 0.0001***
Grades from Last Grading Period						
Baseline: A+, A, A-						
B+, B, B-	1.22		1.22		1.22	
	[0.99 , 1.51]		[0.99 , 1.51]		[0.99 , 1.51]	
C+, C, C-	1.85		1.86		1.86	
	[1.47 , 2.33]		[1.48 , 2.33]		[1.48 , 2.33]	
D or Less	2.56		2.52		2.51	
	[1.88 , 3.49]		[1.86 , 3.43]		[1.85 , 3.41]	
Non-Letter Grades	2.10		2.10		2.10	
	[1.30 , 3.39]	***< 0.0001***	[1.30 , 3.39]	***< 0.0001***	[1.30 , 3.40]	***< 0.0001***

Table 16. Continued

Parents Said They Were Proud						
Baseline: Always						
Sometimes	1.19		1.18		1.18	
	[1.00 , 1.42]		[1.00 , 1.41]		[0.99 , 1.41]	
Seldom	1.36		1.32		1.32	
	[1.10 , 1.70]		[1.07 , 1.65]		[1.06 , 1.64]	
Never	1.66		1.62		1.62	
	[1.24 , 2.23]	***< 0.01***	[1.21 , 2.18]	***< 0.01***	[1.21 , 2.17]	***< 0.01***
Mother in Household						
Baseline: Yes						
No	1.36		1.36		1.36	
	[1.08 , 1.72]	***< 0.01***	[1.08 , 1.71]	***< 0.01***	[1.08 , 1.71]	***< 0.01***
Father in Household						
Baseline: Yes						
No	1.16		1.17		1.17	
	[0.97 , 1.39]	0.12	[0.98 , 1.40]	0.08	[0.98 , 1.40]	0.08
Total Family Income						
Baseline: < $20,000						
$20,000 - $49,999	0.65		0.66		0.76	
	[0.47 , 0.89]		[0.49 , 0.91]		[0.61 , 0.95]	
$50,000 - $74,999	0.95		0.91		0.97	
	[0.61 , 1.48]		[0.58 , 1.41]		[0.73 , 1.28]	
≥ $75,000	1.03		0.98		1.04	
	[0.66 , 1.62]	***< 0.001***	[0.63 , 1.53]	***< 0.01***	[0.79 , 1.36]	***< 0.01***

Table 16. Continued

Use of Govt. Assistance Program						
Baseline: Yes						
No	0.84		0.85		0.86	
	[0.69 , 1.03]	0.09	[0.70 , 1.04]	0.12	[0.71 , 1.05]	0.14
Poverty Level						
Baseline: Income ≤ 1x Threshold						
Income >1x & ≤ 2x Threshold	1.31		1.26		---	---
	[0.95 , 1.79]		[0.92 , 1.72]			
Income > 2x Threshold	1.07		1.09			
	[0.71 , 1.60]	0.14	[0.73 , 1.63]	0.26		
Health Insurance Coverage						
Baseline: Yes			---	---	---	---
No	1.02					
	[0.74 , 1.40]	0.90				
Cigarette Use in Past 30 Days						
Baseline: No						
Yes	5.68		5.70		5.75	
	[4.76 , 6.79]	***< 0.0001***	[4.78 , 6.80]	***< 0.0001***	[4.83 , 6.86]	***< 0.0001***
Alcohol Use in Past 30 Days						
Baseline: No						
Yes	6.88		6.89		6.85	
	[5.87 , 8.06]	***< 0.0001***	[5.89 , 8.06]	***< 0.0001***	[5.85 , 8.01]	***< 0.0001***
Constant	0.002		0.002		0.002	
	[0.001 , 0.003]	***< 0.001***	[0.001 , 0.003]	***< 0.001***	[0.001 , 0.003]	***< 0.001***

****Bold and italicized*** values indicate significance at the α = 0.05 level

APPENDIX B: INTERMEDIATE WEIGHTED BACKWARDS-ELIMINATION LOGISTIC REGRESSION MODELS

Model 1 is the full model. Each successive model in the next model is the backwards-elimination sequence (pr = 0.1). All models contain the Age, Gender, and Race/Hispanicity variables.

Table 17. Weighted backwards-elimination logistic regression models

Covariates	**Model 1**		**Model 2**		**Model 3**		**Model 4**	
	OR [95% CI]	*p-value	OR [95% CI]	*p-value	OR [95% CI]	*p-value	OR [95% CI]	*p-value
Age								
Baseline: 12-13 yrs.								
14-15 yrs.	3.29 [2.03 , 8.98]		3.34 [2.08 , 5.37]		3.33 [2.07 , 5.36]		3.32 [2.07 , 5.35]	
16-17 yrs.	5.54 [3.42 , 8.98]	***< 0.0001***	5.44 [3.39 , 8.74]	***< 0.0001***	5.43 [3.38 , 8.72]	***< 0.0001***	5.41 [3.37 , 8.69]	***< 0.0001***

Table 17. Continued

Gender								
Baseline: Male								
Female	0.93		0.93		0.93		0.93	
	[0.75 , 1.14]	0.46	[0.76 , 1.15]	0.51	[0.76 , 1.14]	0.49	[0.76 , 1.14]	0.50
Race / Hispanicity								
Baseline: White								
Black/African American	1.69		1.67		1.66		1.66	
	[1.21 , 2.37]		[1.20 , 2.34]		[1.19 , 2.32]		[1.19 , 2.33]	
Hispanic	1.03		0.97		0.96		0.96	
	[0.76 , 1.39]		[0.72 , 1.32]		[0.71 , 1.30]		[0.70 , 1.31]	
Other	1.05		1.02		1.02		1.02	
	[0.74 , 1.50]	***0.02***	[0.72 , 1.46]	***0.02***	[0.72 , 1.45]	***0.02***	[0.72 , 1.44]	***0.02***
County Metro Status								
Baseline: Large Metro								
Small Metro	1.02		1.02		1.01		1.01	
	[0.81 , 1.28]		[0.81 , 1.28]		[0.80 , 1.26]		[0.80 , 1.26]	
Non-Metro	0.59		0.61		0.60		0.61	
	[0.43 , 0.79]	***< 0.001***	[0.45 , 0.83]	***< 0.01***	[0.45 , 0.81]	***< 0.01***	[0.45 , 0.82]	***< 0.01***
Overall Health								
Baseline: Excellent								
Very Good	1.05		1.07		1.07		1.07	
	[0.81 , 1.35]		[0.83 , 1.39]		[0.83 , 1.38]		[0.83 , 1.38]	
Good	1.43		1.40		1.39		1.39	
	[1.06 , 1.92]		[1.03 , 1.88]		[1.03 , 1.87]		[1.03 , 1.86]	
Fair / Poor	1.73		1.72		1.71		1.70	
	[1.01 , 2.96]	***0.03***	[1.01 , 2.93]	***0.04***	[1.01 , 2.91]	***0.04***	[1.00 , 2.88]	0.05

Table 17. Continued

Tendency Towards Risky Behavior								
Baseline: Never								
Seldom	1.69		1.73		1.74		1.74	
	[1.22 , 2.34]		[1.25 , 2.39]		[1.26 , 2.40]		[1.26 , 2.41]	
Sometimes	2.18		2.18		2.18		2.18	
	[1.60 , 2.96]		[1.60 , 2.96]		[1.60 , 2.96]		[1.60 , 2.97]	
Always	3.04		3.08		3.11		3.12	
	[1.97 , 4.67]	***< 0.0001***	[2.01 , 4.72]	***< 0.0001***	[2.03 , 4.76]	***< 0.0001***	[2.03 , 4.78]	***< 0.0001***
Importance of Religious Beliefs								
Baseline: Strongly Agree								
Agree	1.31		1.30		1.30		1.30	
	[0.98 , 1.75]		[0.97 , 1.73]		[0.97 , 1.74]		[0.98 , 1.74]	
Disagree	1.79		1.79		1.80		1.80	
	[1.32 , 2.43]		[1.32 , 2.42]		[1.33 , 2.43]		[1.33 , 2.43]	
Strongly Disagree	2.75		2.63		2.64		2.64	
	[2.01 , 3.77]	***< 0.0001***	[1.93 , 3.60]	***< 0.0001***	[1.93 , 3.60]	***< 0.0001***	[1.93 , 3.61]	***< 0.0001***
Grades from Last Grading Period								
Baseline: A+, A, A-								
B+, B, B-	1.11		1.11		1.10		1.11	
	[0.82 , 1.51]		[0.82 , 1.50]		[0.82 , 1.49]		[0.82 , 1.49]	
C+, C, C-	1.83		1.85		1.81		1.82	
	[1.31 , 2.57]		[1.32 , 2.58]		[1.30 , 2.52]		[1.31 , 2.51]	
D or Less	2.88		2.82		2.75		2.75	
	[1.85 , 4.47]		[1.82 , 4.35]		[1.79 , 4.25]		[1.80 , 4.21]	
Non-Letter Grades	1.10		1.10		1.07		1.08	
	[0.56 , 2.15]	***< 0.0001***	[0.56 , 2.13]	***< 0.0001***	[0.55 , 2.09]	***< 0.0001***	[0.55 , 2.12]	***< 0.0001***

Table 17. Continued

Parents Said They Were Proud								
Baseline: Always								
Sometimes	1.38		1.38		1.38		1.37	
	[1.08 , 1.76]		[1.08 , 1.76]		[1.08 , 1.76]		[1.08 , 1.76]	
Seldom	1.55		1.50		1.50		1.50	
	[1.14 , 2.11]		[1.09 , 2.04]		[1.10 , 2.04]		[1.10 , 2.04]	
Never	1.51		1.49		1.48		1.48	
	[0.95 , 2.41]	***0.01***	[0.93 , 2.37]	***0.02***	[0.93 , 2.35]	***0.02***	[0.93 , 2.35]	***0.02***
Mother in Household								
Baseline: Yes								
No	1.65		1.65		1.61		1.61	
	[1.19 , 2.27]	***< 0.01***	[1.21 , 2.27]	***< 0.01***	[1.18 , 2.21]	***< 0.01***	[1.17 , 2.20]	***< 0.01***
Father in Household								
Baseline: Yes								
No	1.24		1.27		1.23		1.22	
	[0.97 , 1.60]	0.09	[0.99 , 1.64]	0.06	[0.97 , 1.56]	0.09	[0.97 , 1.55]	0.09
Total Family Income								
Baseline: < $20,000								
$20,000 - $49,999	0.95		1.02					
	[0.60 , 1.51]		[0.65 , 1.61]		---	---	---	---
$50,000 - $74,999	1.23		1.20					
	[0.67 , 2.27]		[0.65 , 2.19]					
≥ $75,000	1.39		1.35					
	[0.74 , 2.61]	0.33	[0.72 , 2.52]	0.59				

Table 17. Continued

Use of Govt. Assistance Program								
Baseline: Yes								
No	0.84		0.85		0.86		0.87	
	[0.63 , 1.11]	0.22	[0.64 , 1.13]	0.26	[0.65 , 1.14]	0.30	[0.68 , 1.11]	0.26
Poverty Level								
Baseline: Income ≤ 1x Threshold								
Income >1x & ≤ 2x Threshold	1.19		1.16		1.18		---	---
	[0.76 , 1.86]		[0.75 , 1.81]		[0.83 , 1.67]			
Income > 2x Threshold	0.83		0.86		1.04			
	[0.48 , 1.43]	0.17	[0.50 , 1.47]	0.27	[0.74 , 1.47]	0.58		
Health Insurance Coverage								
Baseline: Yes								
No	0.96		---	---	---	---	---	---
	[0.58 , 1.58]	0.86						
Cigarette Use in Past 30 Days								
Baseline: No								
Yes	5.36		5.59		5.56		5.62	
	[4.05 , 7.10]	***< 0.0001***	[4.24 , 7.37]	***< 0.0001***	[4.22 , 7.33]	***< 0.0001***	[4.26 , 7.42]	***< 0.0001***
Alcohol Use in Past 30 Days								
Baseline: No								
Yes	6.23		6.08		6.06		6.04	
	[4.95 , 7.85]	***< 0.0001***	[4.84 , 7.65]	***< 0.0001***	[4.83 , 7.62]	***< 0.0001***	[4.80 , 7.60]	***< 0.0001***
Constant	0.002		0.002		0.002		0.002	
	[0.001 , 0.004]	***< 0.001***	[0.001 , 0.004]	***< 0.001***	[0.001 , 0.004]	***< 0.001***	[0.001 , 0.004]	***< 0.001***

****Bold and italicized*** values indicate significance at the α = 0.05 level

APPENDIX C: UNWEIGHTED MODELS REMOVING OUTLIERS AND INFLUENTIAL POINTS

The "Original Model" is the final unweighted logistic regression model. The model "Removing Outliers" is the final unweighted logistic regression model removing standardized residual values greater than 10. The model "Removing Influential Points" is the final unweighted logistic regression model removing influential points with DBETA values greater than 0.1.

Table 18. Unweighted models removing outliers and influential points

Covariates	**Original Model**		**Removing Outliers**		**Removing Influential Points**	
	OR [95% CI]	*p-value	OR [95% CI]	*p-value	OR [95% CI]	*p-value
Age						
Baseline: 12-13 yrs.						
14-15 yrs.	3.52 [2.50 , 4.94]		3.69 [2.61 , 5.22]		3.59 [2.55 , 5.06]	
16-17 yrs.	5.42 [3.89 , 7.55]	***< 0.0001***	5.67 [4.04 , 7.95]	***< 0.0001***	5.48 [3.92 , 7.65]	***< 0.0001***

Table 18. Continued

Gender						
Baseline: Male						
Female	0.95		0.95		0.96	
	[0.82 , 1.11]	0.52	[0.82 , 1.11]	0.53	[0.83 , 1.12]	0.60
Race / Hispanicity						
Baseline: White						
Black/African American	1.78		1.77		1.78	
	[1.40 , 2.27]		[1.39 , 2.26]		[1.40 , 2.27]	
Hispanic	1.02		1.02		1.02	
	[0.83 , 1.26]		[0.82 , 1.25]		[0.83 , 1.26]	
Other	1.44		1.44		1.39	
	[1.12 , 1.85]	***< 0.0001***	[1.12 , 1.86]	***< 0.0001***	[1.08 , 1.79]	***< 0.0001***
County Metro Status						
Baseline: Large Metro						
Small Metro	0.91		0.91		0.90	
	[0.77 , 1.08]		[0.77 , 1.08]		[0.76 , 1.07]	
Non-Metro	0.64		0.63		0.63	
	[0.52 , 0.79]	***< 0.001***	[0.51 , 0.77]	***< 0.001***	[0.51 , 0.78]	***< 0.001***
Overall Health						
Baseline: Excellent						
Very Good	1.16		1.16		1.15	
	[0.96 , 1.40]		[0.96 , 1.40]		[0.95 , 1.39]	
Good	1.44		1.44		1.42	
	[1.17 , 1.79]		[1.17 , 1.78]		[1.15 , 1.76]	
Fair / Poor	1.59		1.60		1.59	
	[1.13 , 2.25]	***< 0.01***	[1.13 , 2.26]	***< 0.01***	[1.13 , 2.24]	***< 0.01***

Table 18. Continued

Tendency Towards Risky Behavior						
Baseline: Never						
Seldom	1.72		1.74		1.73	
	[1.36 , 2.15]		[1.38 , 2.19]		[1.38 , 2.18]	
Sometimes	2.13		2.16		2.14	
	[1.70 , 2.67]		[1.72 , 2.71]		[1.70 , 2.68]	
Always	3.22		3.22		3.18	
	[2.38 , 4.37]	***< 0.0001***	[2.37 , 4.38]	***< 0.0001***	[2.34 , 4.32]	***< 0.0001***
Importance of Religious Beliefs						
Baseline: Strongly Agree						
Agree	1.43		1.43		1.42	
	[1.16 , 1.77]		[1.15 , 1.77]		[1.15 , 1.76]	
Disagree	2.05		2.05		2.05	
	[1.61 , 2.60]		[1.62 , 2.61]		[1.61 , 2.60]	
Strongly Disagree	2.75		2.76		2.73	
	[2.17 , 3.50]	***< 0.0001***	[2.17 , 3.52]	***< 0.0001***	[2.15 , 3.48]	***< 0.0001***
Grades from Last Grading Period						
Baseline: A+, A, A-						
B+, B, B-	1.23		1.21		1.23	
	[1.00 , 1.51]		[0.98 , 1.50]		[1.00 , 1.51]	
C+, C, C-	1.87		1.87		1.84	
	[1.49 , 2.35]		[1.49 , 2.35]		[1.46 , 2.31]	
D or Less	2.53		2.53		2.52	
	[1.86 , 3.44]		[1.86 , 3.44]		[1.86 , 3.43]	
Non-Letter Grades	2.12		2.13		2.01	
	[1.31 , 3.42]	***< 0.0001***	[1.31 , 3.44]	***< 0.0001***	[1.23 , 3.27]	***< 0.0001***

Table 18. Continued

Parents Said They Were Proud						
Baseline: Always						
Sometimes	1.18		1.19		1.17	
	[1.00 , 1.41]		[1.00 , 1.42]		[0.98 , 1.39]	
Seldom	1.33		1.34		1.32	
	[1.07 , 1.65]		[1.08 , 1.66]		[1.06 , 1.64]	
Never	1.61		1.62		1.62	
	[1.20 , 3.42]	***< 0.01***	[1.21 , 2.18]	***< 0.01***	[1.20 , 2.17]	***< 0.01***
Mother in Household						
Baseline: Yes						
No	1.35		1.36		1.34	
	[1.07 , 1.69]	***0.01***	[1.08 , 1.70]	***< 0.01***	[1.07 , 1.68]	***0.01***
Father in Household						
Baseline: Yes						
No	1.18		1.18		1.18	
	[0.99 , 1.41]	0.07	[0.99 , 1.41]	0.06	[0.99 , 1.41]	0.07
Total Family Income						
Baseline: < $20,000						
$20,000 - $49,999	0.72		0.72		0.72	
	[0.58 , 0.89]		[0.58 , 0.89]		[0.58 , 0.89]	
$50,000 - $74,999	0.89		0.90		0.89	
	[0.69 , 1.15]		[0.70 , 1.16]		[0.69 , 1.15]	
≥ $75,000	0.94		0.94		0.92	
	[0.74 , 1.21]	***< 0.01***	[0.74 , 1.21]	***< 0.01***	[0.72 , 1.18]	***< 0.01***

Table 18. Continued

Cigarette Use in Past 30 Days						
Baseline: No						
Yes	5.79		5.81		5.85	
	[4.85 , 6.99]	***< 0.0001***	[4.88 , 6.93]	***< 0.0001***	[4.91 , 6.97]	***< 0.0001***
Alcohol Use in Past 30 Days						
Baseline: No						
Yes	6.84		6.87		6.90	
	[5.85 , 8.00]	***< 0.0001***	[5.88 , 8.04]	***< 0.0001***	[5.90 , 8.08]	***< 0.0001***
Constant	0.002		0.002		0.002	
	[0.001 , 0.004]	***< 0.001***	[0.001 , 0.003]	***< 0.001***	[0.001 , 0.003]	***< 0.001***

****Bold and italicized*** values indicate significance at the $\alpha = 0.05$ level

APPENDIX D: WEIGHTED MODELS REMOVING OUTLIERS AND INFLUENTIAL POINTS

The "Original Model" is the final weighted logistic regression model. The model "Removing Outliers" is the final weighted logistic regression model removing standardized residual from the final unweighted logistic regression model with values greater than 10. The model "Removing Influential Points" is the final weighted logistic regression model removing influential points from the final unweighted logistic regression model with DBETA values greater than 0.1.

Table 19. Weighted models removing outliers and influential points

Covariates	**Original Model**		**Removing Outliers**		**Removing Influential Points**	
	OR [95% CI]	*p-value	OR [95% CI]	*p-value	OR [95% CI]	*p-value
Age						
Baseline: 12-13 yrs.						
14-15 yrs.	3.32 [2.07 , 5.34]		3.41 [2.10 , 5.52]		3.36 [2.08 , 5.43]	
16-17 yrs.	5.37 [3.35 , 8.59]	***< 0.0001***	5.49 [3.40 , 8.87]	***< 0.0001***	5.37 [3.34 , 8.65]	***< 0.0001***

Table 19. Continued

Gender						
Baseline: Male						
Female	0.93		0.93		0.93	
	[0.76 , 1.14]	0.47	[0.76 , 1.14]	0.48	[0.76 , 1.15]	0.52
Race / Hispanicity						
Baseline: White						
Black/African American	1.72		1.72		1.73	
	[1.23 , 2.41]		[1.23 , 2.41]		[1.24 , 2.42]	
Hispanic	0.98		0.98		0.98	
	[0.72 , 1.33]		[0.72 , 1.33]		[0.72 , 1.34]	
Other	1.03		1.03		1.01	
	[0.73 , 1.45]	***0.01***	[0.73 , 1.46]	***0.01***	[0.71 , 1.43]	***0.01***
County Metro Status						
Baseline: Large Metro						
Small Metro	1.01		1.01		1.00	
	[0.81 , 1.27]		[0.81 , 1.27]		[0.80 , 1.25]	
Non-Metro	0.62		0.61		0.61	
	[0.46 , 0.83]	***< 0.01***	[0.45 , 0.83]	***< 0.01***	[0.45 , 0.83]	***< 0.01***
Overall Health						
Baseline: Excellent						
Very Good	1.06		1.07		1.06	
	[0.83 , 1.37]		[0.83 , 1.38]		[0.82 , 1.37]	
Good	1.39		1.40		1.37	
	[1.04 , 1.87]		[1.04 , 1.88]		[1.02 , 1.84]	
Fair / Poor	1.71		1.71		1.70	
	[1.00 , 2.90]	***0.04***	[1.01 , 2.91]	***0.04***	[1.00 , 2.90]	0.06

Table 19. Continued

Tendency Towards Risky Behavior						
Baseline: Never						
Seldom	1.73		1.73		1.74	
	[1.25 , 2.39]		[1.25 , 2.40]		[1.26 , 2.41]	
Sometimes	2.17		2.18		2.18	
	[1.60 , 2.96]		[1.60 , 2.97]		[1.60 , 2.97]	
Always	3.10		3.07		3.00	
	[2.02 , 4.76]	***< 0.0001***	[2.00 , 4.72]	***< 0.0001***	[1.95 , 4.62]	***< 0.0001***
Importance of Religious Beliefs						
Baseline: Strongly Agree						
Agree	1.30		1.30		1.30	
	[0.98 , 1.74]		[0.97 , 1.74]		[0.97 , 1.74]	
Disagree	1.81		1.80		1.81	
	[1.33 , 2.44]		[1.33 , 2.44]		[1.34 , 2.44]	
Strongly Disagree	2.64		2.64		2.60	
	[1.93 , 3.60]	***< 0.0001***	[1.93 , 3.60]	***< 0.0001***	[1.90 , 3.56]	***< 0.0001***
Grades from Last Grading Period						
Baseline: A+, A, A-						
B+, B, B-	1.11		1.10		1.11	
	[0.82 , 1.50]		[0.82 , 1.49]		[0.82 , 1.49]	
C+, C, C-	1.84		1.84		1.81	
	[1.33 , 2.54]		[1.33 , 2.54]		[1.31 , 2.50]	
D or Less	2.78		2.78		2.78	
	[1.82 , 4.25]		[1.82 , 4.25]		[1.82 , 4.25]	
Non-Letter Grades	1.10		1.10		1.05	
	[0.57 , 2.14]	***< 0.0001***	[0.57 , 2.14]	***< 0.0001***	[0.54 , 2.06]	***< 0.0001***

Table 19. Continued

Parents Said They Were Proud						
Baseline: Always						
Sometimes	1.38		1.38		1.36	
	[1.08 , 1.76]		[1.08 , 1.77]		[1.06 , 1.74]	
Seldom	1.51		1.51		1.50	
	[1.11 , 2.05]		[1.11 , 2.06]		[1.10 , 2.05]	
Never	1.48		1.48		1.49	
	[0.93 , 2.35]	***0.01***	[0.93 , 2.36]	***0.01***	[0.93 , 2.37]	***0.02***
Mother in Household						
Baseline: Yes						
No	1.61		1.61		1.61	
	[1.17 , 2.20]	***< 0.01***	[1.17 , 2.21]	***< 0.01***	[1.17 , 2.21]	***< 0.01***
Father in Household						
Baseline: Yes						
No	1.26		1.26		1.27	
	[1.00 , 1.59]	***0.04***	[1.00 , 1.59]	***0.04***	[1.01 , 1.60]	***0.04***
Cigarette Use in Past 30 Days						
Baseline: No						
Yes	5.68		5.69		5.74	
	[4.30 , 7.49]	***< 0.0001***	[4.31 , 7.50]	***< 0.0001***	[4.35 , 7.58]	***< 0.0001***
Alcohol Use in Past 30 Days						
Baseline: No						
Yes	6.03		6.04		6.09	
	[4.79 , 7.59]	***< 0.0001***	[4.80 , 7.61]	***< 0.0001***	[4.83 , 7.66]	***< 0.0001***
Constant	0.002		0.002		0.002	
	[0.001 , 0.003]	***< 0.001***	[0.001 , 0.003]	***< 0.001***	[0.001 , 0.003]	***< 0.001***

****Bold and italicized*** values indicate significance at the $\alpha = 0.05$ level

APPENDIX E: VARIABLE DESCRIPTION AND CODING

Table 20. Variable description and coding

Variable Coding	Description
IRSEX	Gender
CATAG7	Categorical Age (7 Level with 3 Relevant)
NEWRACE2	Race/Hispanicity
COUTYP2	County Metro/Non-Metro Status
HEALTH	Overall Health
RKFQRSKY	Likes to Test Self By Doing Risky Things
YERLGIMP	Religious Beliefs are Very Important
YELSTGRD	Grades for Last Semester/Grading Period Completed
YEPPROUD	Parents Said They Were Proud (Past 12 Months)
IMOTHER	Mother in Household
IFATHER	Father in Household
INCOME	Total Family Income (4 Levels)
POVERTY2	Poverty Level
GOVTPROG	Use of Government Assistance Program Indicator
ANYHLTI2	Use of Any Health Insurance Indicator
ALCEVER	Ever Had Alcoholic Drink
ALCDAYS	# of Days Had 1 or More Drinks in Past 30 Days
ALCDAYSIND	Alcohol Use in Past 30 Days Indicator
CIGEVER	Ever Smoke Cigarette
CIG30USE	# of Days Smoked Cigarettes in the Past 30 Days
CIG30IND	Cigarette Use in Past 30 Days Indicator
MJEVER	Ever Used Marijuana/Hashish
MJDAY30A	# of Days Used Marijuana/Hashish in the Past 30 Days
MJDAY30IND	Marijuana Use in Past 30 Days Indicator
ANALWT_C	Inverse Probability Sample Weights

BIBLIOGRAPHY

"About the Survey." *National Survey on Drug Use and Mental Health*. Web. 06 Nov. 2014. <https://nsduhweb.rti.org/respweb/project_description.html>.

Agresti, Alan. *Categorical Data Analysis*. 2nd ed. New York: Wiley-Interscience, 2002. Print.

Arseneault, Louise, PhD, Mary Cannon, MRCPsych, John Witton, BSc, and Robin M. Murray, FRCPsych. "Causal Association between Cannabis and Psychosis: Examination of the Evidence." The British Journal of Psychiatry 184.2 (2004): 110-17. Web. 11 Nov. 2014.

Bain, Lee J., and Max Engelhardt. *Introduction to Probability and Mathematical Statistics*. 2nd ed. Pacific Grove, CA: Duxbury/Thomson Learning, 1992. Print.

Chatterjee, Samprit, and Ali S. Hadi. *Regression Analysis by Example*. 5th ed. Hoboken, NJ: Wiley, 2012. Print.

Chun, Barbara. "Overview of Issues in Regression Analysis Using Complex Survey Data." *SCC Annual Meeting: Proceedings of the Survey Methods Sections* (1997). Web. 22 Oct. 2014.

Clark, Christine, Lee E. Coffey, David C. Hunter, Ilona Johnson, Shuangquan Liu, Megan C. Livengood, Lauren M. Machingo, Peilan C. Martin, Allison C. McKamey, Katherine B. Morton, Susan K. Myers, James Pratt, Beth H. Riggsbee, Dustin Williams, and Hilary Zelko. "2012 National Survey on Drug Use and Health: Data Collection Final Report." *Substance Abuse and Mental Health Services Administration (SAMHSA)*. RTI International, September 2013. Web. 15 October 2014.

"Drug Facts: High School and Youth Trends." *National Institute on Drug Abuse (NIDA)*. January 2014. Web. 11 October 2014. <http://www.drugabuse.gov/publications/drugfacts/high-school-youth-trends>.

"Drug Facts: Marijuana." *National Institute on Drug Abuse (NIDA)*. January 2014. Web. 11 October 2014. <http://www.drugabuse.gov/publications/drugfacts/marijuana>.

"Drug Facts: Nationwide Trends." *National Institute on Drug Abuse (NIDA)*. January 2014. Web. 11 October 2014. <http://www.drugabuse.gov/publications/drugfacts/nationwide-trends>.

Dumouchel, William H., and Greg J. Duncan. "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples."*Journal of the American Statistical Association* 78.383 (1983): 535. Web. 22 Oct. 2014.

Hinkins, Susan, Edward Mulrow, and Fritz Scheuren. "Visualization of Complex Survey Data: Regression Diagnostics." *JSM - Section on Survey Research Methods* (2009): 2206-218. Web. 22 Oct. 2014.

Johns, Andrew. "Psychiatric Effects of Cannabis." The British Journal of Psychiatry 178.2 (2001): 116-22. Web. 11 Nov. 2014.

Johnston, Lloyd D., Patrick M. O'Malley, Jerald G. Schulenberg, and John E. Schulenberg. "Monitoring the Future: National Results on Adolescent Drug Use. Overview of Key Findings." *ERIC: Institute of Education Sciences*. National Institutes of Health (NIH), May 2010. Web. 3 Nov. 2014.

Kilpatrick, Dean G., Ron Acierno, Benjamin Saunders, Heidi S. Resnick, Connie L. Best, and Paula P. Schnurr. "Risk Factors for Adolescent Substance Abuse and Dependence: Data from a National Sample." *Journal of Consulting and Clinical Psychology* 68.1 (2000): 19-30. Web.

Korn, Edward Lee., and Barry I. Graubard. *Analysis of Health Surveys*. New York: Wiley, 1999. Print.

Lohr, Sharon L. *Sampling: Design and Analysis*. Boston, MA: Brooks/Cole, 2010. Print.

"Marijuana." *National Institute on Drug Abuse (NIDA)*. March. 2014. Web. 14 October 2014. <http://www.drugabuse.gov/drugs-abuse/marijuana>.

"Marijuana: MedlinePlus." *U.S National Library of Medicine*. U.S. National Library of Medicine, 15 April 2014. Web. 13 October 2014. <http://www.nlm.nih.gov/medlineplus/marijuana.html>.

Mccabe, Sean Esteban, Michele Morales, James A. Cranford, Jorge Delva, Melnee D. Mcpherson, and Carol J. Boyd. "Race/Ethnicity and Gender Differences in Drug Use and Abuse Among College Students."*Journal of Ethnicity in Substance Abuse* 6.2 (2007): 75-95. Web. 21 Oct. 2014.

Meier, M. H., A. Caspi, A. Ambler, H. Harrington, R. Houts, R. S. E. Keefe, K. Mcdonald, A. Ward, R. Poulton, and T. E. Moffitt. "Persistent Cannabis Users Show Neuropsychological Decline from Childhood to Midlife." *Proceedings of the National Academy of Sciences* 109.40 (2012): E2657-2664. Web. 10 Nov. 2014.

Morton, Katherine B., Keilan C. Martin, Bonnie E. Shook-Sa, James R. Chromy, and Erica L. Hirch. "2012 National Survey on Drug Use and Health: Sample Design Report." *Substance Abuse and Mental Health Services Administration (SAMHSA)*. RTI International, January 2013. Web. 13 October 2014.

Nation, Maury, and Craig Anne Heflinger. "Risk Factors for Serious Alcohol and Drug Use: The Role of Psychosocial Variables in Predicting the Frequency of Substance Use Among Adolescents." *The American Journal of Drug and Alcohol Abuse* 32.3 (2006): 415-33. Web.

Newcomb, Michael D., PhD, Ebrahim Maddahian, PhD, and P. M. Bentler, PhD. "Risk Factors for Drug Use among Adolescents: Concurrent and Longitudinal Analyses." *American Journal of Public Health* 76.5 (1986): 525-31. Web. 12 October 2014. <http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.76.5.525>.

Reisner, Sari L., Emily A. Greytak, Jeffrey T. Parsons, and Michele L. Ybarra. "Gender Minority Social Stress in Adolescence: Disparities in Adolescent Bullying and Substance Use by Gender Identity." *The Journal of Sex Research* (2014): 1-14. Web. 11 Nov. 2014.

Skinner, Chris, and Ben Mason. "Weighting in the Regression Analysis of Survey Data with a Cross-national Application." *Canadian Journal of Statistics* 40.4 (2012): 697-711. Web. 22 Oct. 2014.

Steyerberg, Ewout W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009.

United States Department of Health and Human Services (USDHHS), Substance Abuse and Mental Health Services Administration, and Center for Behavioral Health Statistics and Quality. "National Survey on Drug Use and Health: CAI Specification for Programming (English), 2012 (ICPSR 34933)." *Substance Abuse and Mental Health Services Administration (SAMHSA)*. RTI International, October 2011. Web. 14 October 2014.

United States Department of Health and Human Services (USDHHS), Substance Abuse and Mental Health Services Administration, and Center for Behavioral Health Statistics and Quality. "National Survey on Drug Use and Health (NSDUH) - Overview, 2012 (ICPSR 34933)." *Substance Abuse and Mental Health Services Administration (SAMHSA)*. 6 October 2014. Web. 13 October 2014. <http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/34933>.

United States Department of Health and Human Services (USDHHS), Substance Abuse and Mental Health Services Administration, and Center for Behavioral Health Statistics and Quality. "National Survey on Drug Use and Health: Public Use File Codebook, 2012 (ICPSR 34933)." *Substance Abuse and Mental Health Services Administration (SAMHSA)*. RTI International, 30 October 2013. Web. 14 October 2014.

"USA QuickFacts from the US Census Bureau." United States Census Bureau. N.p., 8 July 2014. Web. 11 Nov. 2014. <http://quickfacts.census.gov/qfd/states/00000.html>.

Valiant, Richard, Jianzhu Li, and Dan Liao. "Regression Diagnostics for Survey Data." Web. 22 Oct. 2014.

Vittinghoff, Eric, David V. Glidden, Stephen C. Shiboski, and Charles E. McCulloch. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer, 2005. Print.

Wallach, Phillip A., and John Hudak. "Legal Marijuana: Comparing Washington and Colorado." *The Brookings Institution*. 8 July 2014. Web. 14 October 2014. <http://www.brookings.edu/blogs/fixgov/posts/2014/07/08-washington-colorado-legal-marijuana-comparison-wallach-hudak>.

Winship, Christopher, and Larry Radbill. "Sampling Weights and Regression Analysis." *Sociological Methods & Research* 23.2 (1994): 230-57. Web. 22 Oct. 2014.

Wissoker, Douglass. "Notes on Weighting in Regression". The Urban Institute. Web. 22 Oct. 2014.

Youth Risk Behavior Surveillance System. "Trends in the Prevalence of Marijuana, Cocaine, and Other Illegal Drug Use: National YRBS 1991-2011." *Centers for Disease Control and Prevention*. 12 June 2014. Web. 14 October 2014. <http://www.cdc.gov/healthyyouth/yrbs/pdf/us_drug_trend_yrbs.pdf>.