Weeding Out Underreporting:
A Study of Trends in Reporting of Marijuana Consumption in the US

December 15, 2015

Maria Cuellar[1]
Carnegie Mellon University
School of Public Policy, Heinz College

Advisors: Amelia Haviland [2], Jonathan Caulkins[3], and Howard Seltman[4]

Abstract

When individuals respond to a survey about illicit drug consumption, they tend to underreport the amount of drugs they consume. To adjust for underreporting of marijuana use, researchers multiply the proportion of individuals who reported using by a constant factor, such as the ONDCP's 1.3 (Kilmer et al., 2014). Some researchers have suggested that different factors should be used for different groups, but no specific values have been suggested (Hser et al. (1999), Harrison et al. (2007), Fendrich and Vaughn (1994)). Although the current adjustments are simple, they do not account for changes in reporting over time. We determine whether reporting is changing over time by estimating changes in a quantity that is a proxy for changes in reporting, and propose a method to update the factors if there are changes in reporting over time. We used the data from 1979 to 2013 from the United States National Survey on Drug Use and Health, a nationally representative cross-sectional survey. We used the proportion of positive responses to the question, "Had you used marijuana by age 25?" across survey years as our outcome variable, and we estimated it for each survey year and birth cohort using three steps: 1) generation of synthetic birth cohorts, 2) domain estimation, and 3) weighted least squares regression. We found significant evidence of changes in reporting from 1979 to 1991, and no significant evidence of changes from 1992 to 2013. The currently used adjustment should remain, for now. Our results imply that since reporting did not change in 1992 to 2013, any change in reported marijuana use is not a product of a change in reporting, but a change in true use. In the future, if there are suspected changes in stigma, our method provides an inexpensive way to update the current adjustments. Additionally, our method is generalizable to other questions of sensitive nature, where stigma may affect reporting behavior.

[1]Email: mcuellar@cmu.edu. Current address: 5000 Forbes Ave. Pittsburgh, PA, USA
[2]Email: amelia@andrew.cmu.edu
[3]Email: caulkins@andrew.cmu.edu
[4]Email: hseltman@cmu.edu

1 Introduction

When individuals respond to a survey about illicit drug consumption, they tend to underreport the amount of drugs they consume (Wish et al. 1997) primarily due to stigma or social desirability bias (Fuchs, 2008). In a 2014 report from the Office of National Drug Control and Policy from the United States, Kilmer et al. say, "When surveying respondents about sensitive behaviors, underreporting is a perennial concern." Quantifying underreporting is critical in estimating the prevalence of marijuana consumption, which policy makers need to know to make decisions about topics as diverse as funding for public health programs, funding for drug enforcement, and legislation.

Underreporting is common in self-reported surveys on topics as varied as rape and sexual assault to HIV/AIDS (Kruttschnitt (2014), Harrison and Hughes (1997), Smith et al. (1999), Conway et al. (1989), Davis and Henderson (2003)). It is particularly common in surveys about drug use (Harrison and Hughes (1997); Ball (1967), Fendrich et al. (2004), Johnston and O'Malley (1997), Zullino et al. (2008)). Various authors have estimated the percentage of respondents that were honest about their marijuana use in self-reports and have found values ranging from 60% to 87.7% (Harrison et al. (2007), Gettman (2007), Kilmer and Pacula (2009), Fendrich et al. (2004), Hser et al. (1999), Kilmer et al. (2014)). These estimates have been translated into adjustments where the percentage of reported users is multiplied by the adjustment to correct for underreporting and obtain a more accurate percentage of users (the adjustments for the studies above range from to 1.15 to 1.6).

Although the adjustments provided by previous authors are simple and easy to use to correct for underreporting, they do not allow for heterogeneities in the sample. This is problematic because there is heterogeneity in self-reports for different demographic groups and types of drugs. For instance, younger respondents tend to overreport marijuana use while older respondents tend to underreport marijuana use (McAllister and Makkai (1991) and Kilmer and Pacula (2009)), African Americans are less likely to report their marijuana use than Whites, Hispanics, and Asians (Fendrich et al. (2004)), and opiate users have been found to overreport their drug use (Zullino et al. (2008)). Most troubling for today, however, is the heterogeneities that might arise from changes in reporting over time. The number of people consuming marijuana has been increasing in recent decades, particularly in developing countries, according to the United Nations Office on Drugs and Crime, World Drug Report (2008), and changes in prevalence have direct impacts on stigma (CITE), which is a main driver in underreporting. In the United States, the General Social Survey from 2014 showed that support for marijuana legalization has been rising since 1990 (see Figure 1). Thus, it is possible that individuals might be changing the way in which they report their marijuana use over time, which would imply that the adjustment factors suggested by the studies mentioned above might be particular to the year in which they were calculated. Indeed Hser et al.'s 1992 study suggests that underreporting might increase, but this study only compared two points in time. Until now, no one has estimated trends of marijuana reporting over time, which would allow researchers to have different adjustments for different calendar years.

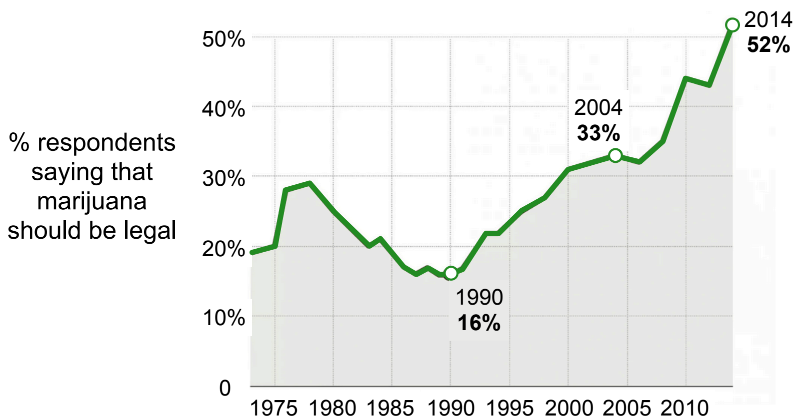

Figure 1: Results from a General Society Survey. The support for marijuana legalization has been rising since 1990.

Our main goal was to answer the question: Are individuals changing the way in which they report their marijuana use in surveys over time? To clarify, our aim was not to estimate the measurement error (i.e. the difference between the measured value and the true value), nor to propose using one of the adjustments proposed by previous authors. We intended to provide updates to all adjustments that could correct for changes in reporting over time. For this purpose, we examined the stability in reported ever-use of marijuana by age 25 in the National Survey on Drug Use and Health, a nationally representative cross-sectional household survey, from 1979 to 2013.

2 Data

The National Survey on Drug Use and Health (NSDUH)[5] is a nationwide survey of the civilian, non-institutionalized members of US households aged 12 and older. NSDUH is often used by policy makers, so improving its estimates might be useful for writing policy.

From 1971 to 1988 NSDUH was performed at least once every three years, and since 1990 it has been performed every year. Now, it covers the use of illicit drugs, the nonmedical use of prescription psychotherapeutic drugs, the use of alcohol and tobacco products, the dependence and abuse involving drugs and alcohol, mental health problems, and treatment of substance abuse and mental health problems.[6] Since 1991, approximately 70,000

[5]Before 2002, the National Survey on Drug Use and Health (NSDUH) was called the National Household Survey on Drug Abuse (NHSDA). The name was changed to reflect some methodology changes.

[6]At the time of this project, data were available from the 1979, 1982, 1985, 1988, and 1990–2013 collection

individuals were interviewed every year, and the first five surveys years have fewer than 10,000 respondents. Overall there are 1,067,408 cases available.

The NSDUH is implemented by in-person interviews conducted at the respondents' place of residence. An interviewer visits each selected residential unit to administer the interview using a laptop computer. Since 1999, the survey uses a combination of computer-assisted personal interviewing, conducted by an interviewer to obtain basic demographic information, and audio computer-assisted self-interviewing for most of the questions. These methods were introduced in 1999 and replaced paper-and-pencil questionnaires.

The NSDUH employs a multistage (stratified cluster) sample design for the selection of a representative sample from the population. It oversamples the younger population, and it provides sampling weights to adjust for unequal selection probabilities at the various stages of sample selection, including nonresponse.

Nonresponse to this survey is of particular concern because many of the drug-using behaviors measured have low prevalence in the population and estimations could be severely biased if nonrespondents are more likely to be users than respondents. Therefore, NSDUH introduced incentives. In the 2002–2013 surveys, the respondents received $30 USD for completing the full survey. This improved nonresponse rates as well as the individuals' decisions to take part in the survey.

For the NSDUH survey from 1979 to 2013, the average weighted interview response rate was 77.0%. (Note that these response rates reflect the original sample, not the public use file used in this paper).

3 Methodology

Our method had four steps:

i) imputation of the year of birth variable from the public-use data,

ii) domain estimation to derive the proportion of individuals who reported using by age 25 in each cohort and survey year

iii) stepwise regression with data splitting to select the model with the best fit,

iv) weighted least squares regression to determine changes in reporting over time.

We first we give a heuristic explanation of the methodology and then describe each of the four steps below. We estimate trends in reported marijuana lifetime-use by age 25 by tracking birth cohorts and comparing their responses over time. To describe this process, we refer to Figure 2. For each birth cohort, we compared the fraction of the individuals who reported use by age 25, out of all the individual in that birth cohort and that survey year. Because we can observe the same information in each survey, we can compare cohort-specific

---

years.

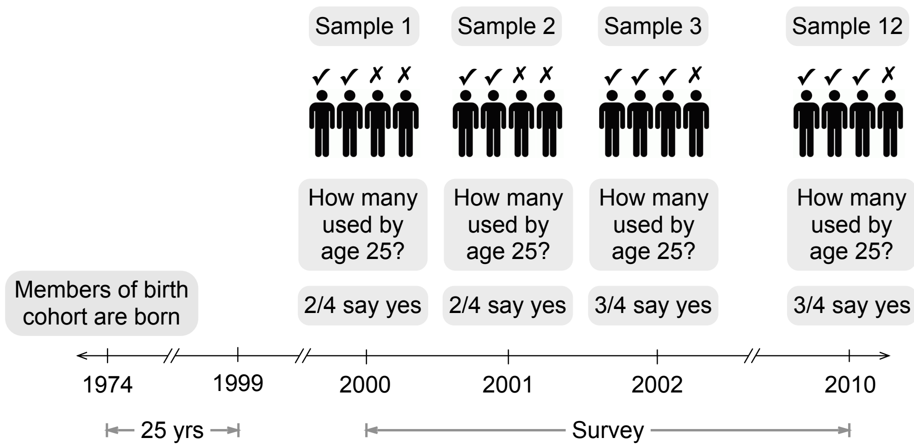

Figure 2: Example of our analysis for one birth cohort. The cohort born in 1974 is tracked throughout the survey years. In 2000 (after the individuals have turned 25), the survey shows that 2/4 of the cohort's members reported using by age 25, in 2001 the proportion is again 2/4, but in 2002 it is 3/4. We estimate this proportion until 2013 (our latest year of data). Differences in proportions reflect changes in reporting. We repeat this analysis for all eligible birth cohorts to obtain the trends in reporting.

responses about marijuana use year to year. The key idea is that since the individual respondents cannot change the reality about whether they had used marijuana by age 25, their true use does not change, and any changes seen in the surveys from 2002–2013 must reflect changes in reporting. If the proportions do not increase, then there are no changes in reporting over time, and if the proportions increase, then more individuals are reporting that they used marijuana by age 25 over time.

We ask about marijuana ever-usage by age 25 because 99% of the US population who used marijuana started doing so before age 25 to prevent having people not remember correctly the age at which they first used.

3.1 Imputation of the year of birth variable from the public-use data

A key issue that must be accounted for when comparing birth cohorts over time is that the population of the US is changing. For example, the population has a higher percentage of immigrants, it is aging as a whole, and some of the respondents from the earlier years might have passed away in the later years.

To account for these issues, we rely on the sampling weights provided by NSDUH. These weights reflect the number of people that each respondent represents at the time of the survey. The calculation of NSDUH person-level weights includes a calibration step that results in weights that are consistent with population control totals obtained from the U.S. Census Bureau (SAMHDA, 2014). These controls are based on the most recently available

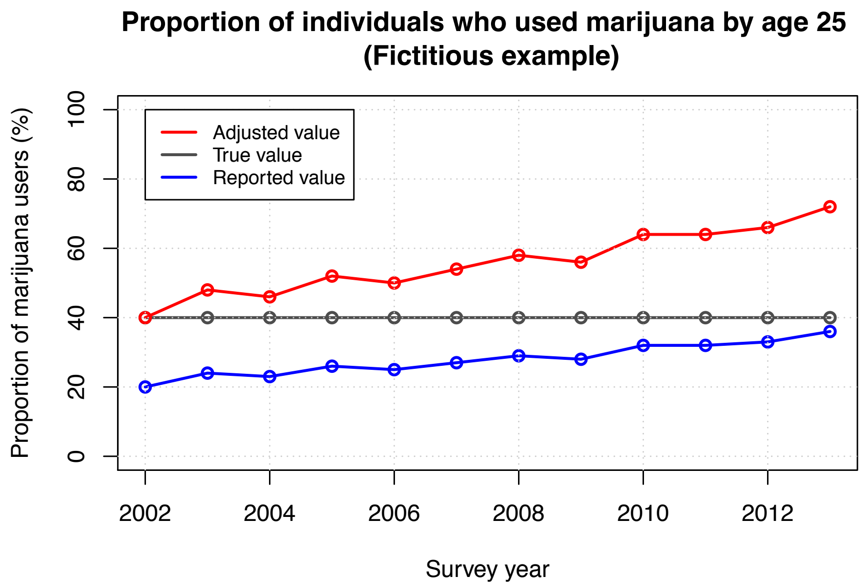

Figure 3: Ficticious example of adjustments for underreporting changing over time. If reporting did increase, and the adjustment stayed constant, then the adjusted value would become worse over time.

decennial census. The Census Bureau updates these control totals annually to account for population changes after the census. For the analysis weights in the NSDUH surveys from 2002 through 2010, the control totals were derived from the 2000 census data, and starting with the 2011 survey, the control totals were based on data from the 2010 census.

Another issue that must be accounted for is the changes over time in the likelihood of being included in our sample cohorts because of nonresponse. This is also accounted for by using the weights provided by NSDUH.

3.1.1 Using a synthetic cohort approach to track individuals over time

NSDUH is a cross-sectional survey, so it does not track individuals over time. Usually cross-sectional surveys only contain information about the time at which the survey was performed. However, NSDUH asks retrospective questions about marijuana use, and this allows for the comparison of response over time by using the synthetic cohort approach. Respondents are picked at random from the same population for each survey, so the individuals vary from year to year. This has the benefit of eliminating individual-specific fixed effects. However, calculations must be done as proportions of the total sample size by year, rather than absolute numbers, to account for changes in sample sizes among survey years.

Verbeek and Nijman 1992 (Verbeek and Nijman, 1992) show that the effects of ignoring the fact that only a synthetic panel is available is small if the cohort sizes are sufficiently large (over 100 observations) and if the true means within each cohort have a sufficient amount of time variation. Our cohort sizes are 951 on average, and the smallest one is 451. Thanks to its large sample size, NSDUH is an adequate candidate for this synthetic cohort approach.

To generate our synthetic cohorts, we needed each individual's year of birth. In order to protect the confidentiality of the sensitive information in the NSDUH survey, however, the "year of birth" variable is only provided in the restricted use file, which was not available for this study. However, NSDUH includes questions about the use of tobacco, alcohol, and other licit and illicit drugs, and these can be used to obtain a year of birth variable.

NSDUH provides two variables for each of the drugs in the survey, including alcohol and tobacco: "age of first use" and "year of first use". On average, from 2002 to 2013, 77% of the US population reported using at least one substance[7], and NSDUH required that they provide an age of first use for each substance. By using the year of birth, NSDUH obtained a logically imputed variable for the year of first use. Hence year of birth can be imputed for most respondents, including all the marijuana users, by subtracting the two provided quantities:

$$\text{year of birth} = \text{year of first use} - \text{age of first use}. \tag{1}$$

The following variables were used to generate the "year of birth" variable: Cigarettes, cigars, smokeless tobacco, chewing tobacco, snuff, alcohol, marijuana, cocaine, hallucinogens, LSD, PCP, ecstasy, inhalants, stimulants, methamphetamines. The "age of first use" variables were reported for 99% of the respondents who reported having used the drug. Once an individual has reported using the drug at some point, the computer software directs the individual to provide a response for the age of first use for that drug. Since the respondents only get the cash incentive when they have completed the survey, the item non-response rate is very low.

This method of obtaining "year of birth" by using Equation 1 is accurate up to a one year difference. It will be off by one year if the respondent's day of birth comes after the date in which the respondent first used marijuana. For example, if the responder was born on 06/01/1987 and first used on 01/01/2000, the respondent was 12 when he first used, but the calculation 2000−1987 gives 13. On the other hand, if the respondent's day of birth comes before the date of first use, the difference yields the correct age of first use. So, the year of birth variable used by this study is one year higher for half the population, in expectation. This does not affect the analysis because it is evenly distributed throughout the population.

The remaining 23% of respondents, i.e. the never-reporters, are excluded from this analysis. Our assumption is that the never-reporters do not have different trends in reporting

[7] This percentage is between 76% and 79% in every survey.

than the reporters, so excluding them from this analysis does not affect our results.

To verify the validity of this assumption, we determined that at least 90% of the respondents who reported using these drugs had used them for the first time by age 25. This ensures that the cohort samples that we compare in each survey year are drawn from the same distribution. If a large proportion of the individuals had used the drugs after age 25, this would mean that as the birth cohort grows older, we would include individuals who we had not included in earlier survey years, and thus we would have samples from different distributions. Thus, any effects seen from making comparisons between these groups would be confounded with the fact that the groups would be composed of different types of people.

3.2 Domain estimation to derive the proportion of individuals who reported using by age 25 in each cohort and survey year

3.2.1 Generating our outcome variable

Our outcome variable is 1 if the person used marijuana by age 25 and 0 otherwise. The NSDUH survey asks the question, "How old were you the first time you used marijuana or hashish?" It has been shown that asking about the age of first use yields more reliable results than asking about the year of first use (Harrison and Hughes, 1997).

From the response to this question and the "year of birth" variable, analysts at NSDUH calculate a "year of first use" variable, which is available in the public use file. Using this variable, we generated new indicator variables for the cohorts born in 1955 to 1975.

We chose 1955 to 1975 because before 1955 the sample sizes become smaller than 100 unweighted, and after 1975 the birth year would be too close to our survey period (1975 + 25 = 2000, and our survey period is 2002–2013). Our outcome variable is the following, for each individual:

$$\text{usedbyage25}_i = \begin{cases} 1 & \text{if age of first marijuana use} \leq 25 \\ 0 & \text{if year of first marijuana use} > 25 \text{ or year of first use is missing.} \end{cases}$$

3.2.2 Domain estimation

We used domain estimation to find our outcome variable. Each domain is a specific cohort in a specific survey year. Our outcome variable is the proportion of individuals from the birth cohort who reported using marijuana by age 25, for each survey year. This proportion is denoted by p_{cy},

$$p_{cy} = \frac{\sum_{i \in \text{Population domain}} y_i}{N_{cy}}, \tag{2}$$

where c is the birth cohort, y is the survey year, N_{cy} is the number of population units in cohort c and survey year y. The variable y_i is defined as follows:

$$y_i = \begin{cases} 1 & \text{if the respondent is in cohort } c \text{ and survey year } y\text{, and reported using marijuana by age 25} \\ 0 & \text{if the respondent is not in cohort } c \text{ and survey year } y. \end{cases}$$

$$x_i = \begin{cases} 1 & \text{if the respondent is in cohort } c \text{ and survey year } y \\ 0 & \text{if the respondent is not in cohort } c \text{ and survey year } y. \end{cases}$$

Our outcome variable $\hat{p}_{cy}$ has the following distribution:

$$\hat{p}_{cy} \sim N(p_{cy}, \mathrm{Var}(\hat{p}_{cy})), \tag{3}$$

where c is the birth cohort and y is the survey year.

A natural estimator of p_{cy} is

$$\hat{p}_{cy} = \frac{\sum_{i \in \text{Sample domain}} y_i}{n_{cy}}, \tag{4}$$

where n_{cy} is the number of sample units in cohort c and survey year y.

We could estimate $\hat{p}_{cy}$ simply by using the sample proportion and its standard error. However, until the sample is drawn, we do not know which individuals in the population belong to which cohort. That means the number of individuals who fall into each cohort is a random variable, with unknown value at the moment when the survey is designed (Lohr, 2009). In other words,

$$\mathrm{SE}(\hat{p}) \neq \sqrt{\left(\frac{N-n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}}. \tag{5}$$

If we calculate the variance using the right-hand side of Equation 5, we would be treating the subset of data in that subpopulation as a new survey. This approach gives the correct point estimates, but the incorrect standard errors.

3.2.3 Calculating the correct standard errors with domain estimation

To find the correct standard errors, we derived domain means as ratio estimators (Lohr, 2009).

The population ratio p_{cy} and its estimator $\hat{p}_{cy}$ are given by,

$$p_{cy} = \frac{\sum_{i=1}^{N_{cy}} y_i}{\sum_{i=1}^{N_{cy}} x_i} = \frac{\mu_y}{\mu_x}, \quad \hat{p}_{cy} = \frac{\sum_{i=1}^{n_{cy}} y_i}{\sum_{i=1}^{n_{cy}} x_i} = \frac{\bar{y}}{\bar{x}}. \tag{6}$$

$\hat{p}_{cy}$ is a biased estimator of p_{cy} because

$$E(\hat{p}_{cy}) = E\left(\frac{y}{x}\right) \neq \frac{E(\bar{y})}{E(\bar{x})} = \frac{\mu_y}{\mu_x} = p_{cy}. \tag{7}$$

However, for large sample sizes $\hat{p}_{cy}$ is approximately unbiased because the bias is of the order $O(\frac{1}{n})$. The sample sizes in our analysis are large (the smallest cohort has sample size 450, so $\hat{p}_{cy}$ is unbiased for the purposes of our study. Additionally, since $E(n_{cy})$ is large, the variance can be approximated by using a Taylor series expansion on the ratio $\hat{p}_{cy} = \bar{y}/\bar{x}$.

In large samples, the bias of $\hat{p}_{cy}$ is typically very small relative to its variance, so

$$\text{Var}(\hat{p}_{cy}) \approx \text{MSE}(\hat{p}_{cy}) = E\left[(\hat{p}_{cy} - p_{cy})^2\right].$$

To find the MSE of $\hat{p}_{cy}$ we get a Taylor approximation of $\hat{p}_{cy} - p_{cy}$. We express p_{cy} as a function of the population totals,

$$p_{cy} = \frac{t_y}{t_x} = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} \quad \text{and} \quad \hat{p}_{cy} = \frac{\hat{t}_y}{\hat{t}_x} = \frac{\overline{y}}{\overline{x}}.$$

To calculate the MSE, we Taylor expand $\hat{p}_{cy} - p_{cy}$ to first order:

$$\begin{aligned}
\hat{p}_{cy} - p_{cy} &\approx \frac{t_y}{t_x} + \frac{\partial}{\partial \hat{t}_x}\left[\frac{\hat{t}_y}{\hat{t}_x}\right]\Bigg|_{t_x,t_y}(\hat{t}_x - t_x) + \frac{\partial}{\partial \hat{t}_y}\left[\frac{\hat{t}_y}{\hat{t}_x}\right]\Bigg|_{t_x,t_y}(\hat{t}_x - t_x) - p_{cy} \\
&= -\frac{t_y}{t_x^2}(\hat{t}_x - t_x) + \frac{1}{t_x}(\hat{t}_y - t_y).
\end{aligned}$$

The expectation of the square of $\hat{p}_{cy} - p_{cy}$ is

$$\text{Var}(\hat{p}_{cy}) \approx E\left[(\hat{p}_{cy} - p_{cy})^2\right] \approx \frac{1}{t_x^2}\left[\frac{t_y}{t_x}V(\hat{t}_x) + V(\hat{t}_y) - 2\frac{t_y}{t_x}\text{Cov}(\hat{t}_x, \hat{t}_y)\right].$$

We then plug in estimators for the unknown quantities to obtain the estimated variance,

$$\widehat{\text{Var}}(\hat{p}_{cy}) = \sum_{j=1}^{k} V\left(\sum_{i \in S} w_i y_{ij}\right) + 2\sum_{j=1}^{k-1}\sum_{l=j+1}^{k}\text{Cov}\left(\sum_{i \in S} w_i y_{ij}, \sum_{i \in S} w_i y_{il}\right), \tag{8}$$

where w_i are survey weights.

We used Equation 8 to obtain the standard error corresponding to each proportion $\hat{p}_{cy}$ in our model.

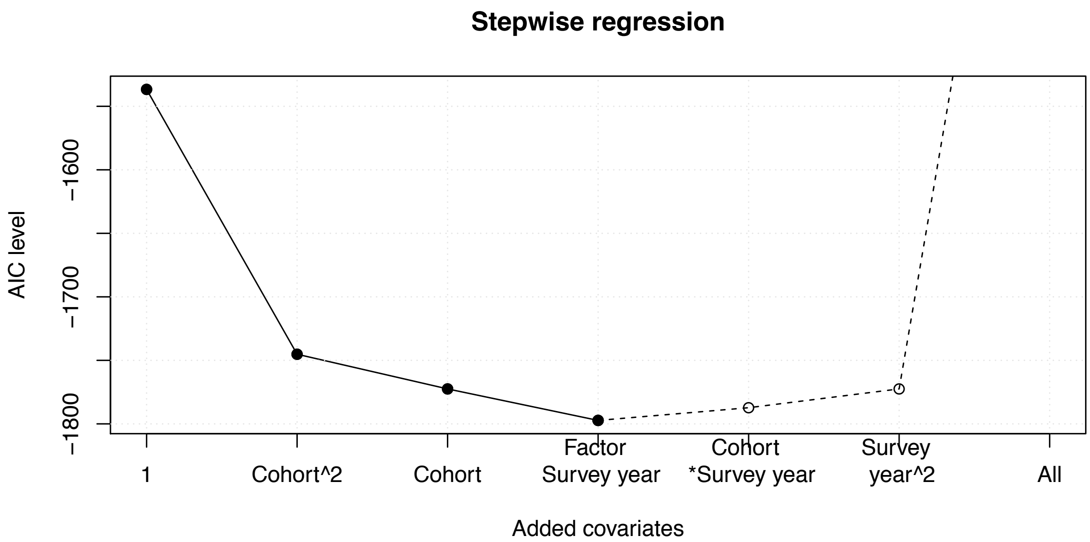

Figure 4: Results from the stepwise regression used to select the correct covariates. We ran the regression backward and forward, and found that the best model was the one from Equation 9.

3.3 Stepwise regression with data splitting to select the model with the best fit

3.3.1 Using model selection to find the correct set of covariates

In order to find the correct set of covariates for the model, we ran a stepwise regression in both directions. Stepwise regression is a semi-automated process of building a model by successively adding or removing variables from the candidate pool, depending on a certain criterion. It has the advantage of allowing consideration of more subsets, without the need for examining all $2k$ subsets of the variables in the covariate pool.

In this case, we used the AIC criterion, so our code selected the model with the lowest AIC. It is common to run stepwise regressions in both directions (instead of just forward or backward) to avoid locating local minima and missing the model with the best fit. The candidate covariates were the following: survey year, survey year^2, survey year^3, factor(survey year), cohort, cohort^2, cohort^3, cohor·survey year, factor(cohort).

It should be noted that if we had included survey year, cohort, and an interaction term between the two, the model would have been overfitted and there would have been no variation. So, we needed to include some combination of the terms determined by a statistical test.

In order to avoid the optimism of the training error when reducing estimated risk, we split the data into two halves at random, ran model selection on one half (train set) and fit the model on the other half (test set). We did data splitting 10 times and verified that the model did not change and the coefficients were essentially equal (i.e. they did not change signs or vary by more than 0.001 in either direction).

We did not run cross-validation because our sample size of the aggregate model was 252, which would result in very small sample sizes if running a k-fold cross validation, for example. Our process of splitting the data in half repeatedly essentially simulates cross validation.

3.4 Weighted least squares regression to determine changes in reporting over time

3.4.1 Accounting for weights

Since NSDUH is not a simple random sample, we also needed to account for the population weights in the survey. Let an individual's weight be $w_i = 1/\pi_i$. Then the estimated population total is $\hat{t}_y = \sum_{i=1}^{n_{cy}} w_i y_i$. In ratio estimation, $\hat{t}_{yp} = \frac{t_x}{\hat{t}_x}\hat{t}_y = \frac{t_x}{\hat{t}_x}\sum_{i=1}^{n_{cy}} w_i y_i$. Define $g_i = t_x/\hat{t}_x$. Then,

$$\hat{t}_{yp} = \sum_{i=1}^{n_{cy}} w_i g_i y_i.$$

The estimator $\hat{t}_{yp}$ is a weighted sum of the observations with weights $w_i^* = w_i g_i$. The weight adjustments calibrate the estimates on x, i.e. they force the estimated total for x to equal the known population total t_x since

$$\sum_{i=1}^{n_{cy}} w_i g_i y_i = t_x.$$

If we let $u = g_i \sum_{i=1}^{N_{cy}} (y_i - p_{cy} x_i)^2 = g_i e_i$, then,

$$\hat{V}(\bar{u}) = \left(1 - \frac{n_{cy}}{N_{cy}}\right) \frac{1}{n_{cy}(n_{cy}-1)} \sum_{i=1}^{n_{cy}} (u_i - \bar{u})^2 = \left(1 - \frac{n_{cy}}{N_{cy}}\right) \frac{s_{\hat{p}_{cy}^2}}{n} \left(\frac{t_x}{\hat{t}_x}\right)^2.$$

This process accounts for the complex survey design (stratification and clustering), including the proper weighting by the survey (probability) weights. We used the `survey` package in R to do these calculations.

By using this methodology, we obtained estimates of the proportions for each screening cohort. That is, we obtained estimates for the following $45 \times 12 = 540$ proportions:

$$\hat{p}_{cy}, \text{where } c = 1955, \ldots, 2000, \text{ and } y = 2002, \ldots, 2013.$$

3.4.2 Weighted least squares model

After doing model selection, we selected the following model:

$$p_{cy} = \beta_0 + \beta_1 c + \beta_2 c^2 + \sum_{k=2003}^{2013} \beta_k y_k + \varepsilon_{cy}, \text{ with weights} = \frac{1}{\widehat{se}^2}, \tag{9}$$

where p_{cy} is the proportion estimate, c is cohort, y is survey year, and se is standard error of each estimated proportion. Here, $\mathrm{E}(\varepsilon_{cy}) = 0$ and $\mathrm{Var}(\varepsilon_{cy}) = \sigma_{cy}^2$.

This model regresses "whether the individual reported use by age 25" on birth cohort and survey year. The estimated coefficients on the survey year variables represent the changes in reporting marijuana user over time.

We implemented a weighted least squares (WLS) model because weighting each data point accounts for the complex survey design. Usually, WLS is used to correct for heteroskedasticity, or unequal variance over certain covariate. In this case, we use the WLS approach in a slightly different way.

Some estimates will be more variable than others due to factors such as changing cohort sizes and changing stratification and clustering. Thus, using the standard errors that we found in the domain estimation we can give more weight in our calculations to the data points that we trust more and less weight to the points we trust less. In other words, the standard error for each point in our model acts as a placeholder for how much we trust that data point.

This aggregation allowed us to find an estimate of the consistency of individuals' responses over time independent of the individual's birth cohort. In fact, this model allowed us to explicitly find an estimate of the effect of belonging from different cohorts on whether the individuals' responses were consistent, which is an estimate of the cohort effect.

3.4.3 Finding the correct standard errors for factor covariates

Including survey year as a factor in the regression gave us flexibility for the effect of survey year on p_{cy}, but when comparing the coefficient estimates of all survey years, we had to be careful with the standard errors. Although the effect of survey year 2003 on p_{cy} might be $\beta_0 + \beta_4$, the standard error of the effect is not just $\mathrm{se}(\beta_0)+\mathrm{se}(\beta_4)$ because the variables are not independent, and from basic statistics,

$$V(aX + bY + c) = a^2V(X) + b^2V(Y) + 2ab\mathrm{Cov}(X, Y) \neq V(aX) + V(bY). \tag{10}$$

Thus, we obtain the variance-covariance matrix from the model outcome. Let M be the variance-covariance matrix of the sample coefficients β. M is a $p \times p$ symmetric square matrix, where p is the number of coefficients in the model including the intercept, which in our case is $p = 14$. The diagonal values are the variances of the β's and the off-diagonal values are the covariances of the $\beta's$, $\mathrm{Cov}(\beta_i, \beta_j)$. Thus, from Equation 10, the variances of the coefficients are

$$\begin{aligned}
\hat{V}(\beta_0) &= M_{1,1} \\
\hat{V}(\beta_3) &= M_{3,3} + M_{1,1} + 2 * M_{1,3} \\
\hat{V}(\beta_4) &= M_{4,4} + M_{1,1} + 2 * M_{1,4} \\
&\dots \\
\hat{V}(\beta_{13}) & M_{13,13} + M_{1,1} + 2 * M_{1,13}.
\end{aligned}$$

We can then use these to produce confidence intervals at each coefficient.

3.5 Our assumptions

As a summary, our analysis depends on the following assumptions:

1. The non-responders would not have different reporting trends than the responders if they responded (only the trend in reporting over time needs to be the same, not the use levels).

2. The proportion of those who do use some substance but report no use does not change as people age.

3. Mortality is uncorrelated with marijuana use, so when we account for changing population this does not invalidate our results.

4. In each survey year the sample from each birth cohort is a representative sample of the members of the target population in that birth cohort.

4 Results

4.1 Weighted least squares model results

NOTE: THE FOLLOWING IS STILL IN ROUGH DRAFT FORM.

Table 2 shows the results from the weighted least squares regression from Equation 9, and Figure **??** shows a visual representation of the survey year coefficients as fitted values.

The first result is that the coefficients for survey year do not seem to increase, which contradicts our hypothesis. In fact, it seems that there are two flat levels: the estimates for 2002–2004 and the estimates for 2005–2013.

The first three years of the survey seem higher than the rest. There was a large design change in 2005, in which NSDUH changed their clustering from using clusters that they designed to using census tracts. This design change might have resulted in seeing higher proportions in the earlier years.

More careful revision with a Tukey Honest Significance Difference test in Figure 5 shows that the only year that is significantly different from all the rest is 2002. The survey from 2012 is also significantly different from 2003 and 2004. But these are the only proportions that are statistically significantly different from each other.

The survey from 2002 was the first one in which NSDUH gave out cash incentives to the whole sample (that is, to the respondents who filled out the complete survey). It was also the first survey to include many new questions and to implement improved data collection quality control procedures. In addition, new population data was available from the 2000 census, so the weights in 2002 are new and different than the weights in 2001. The survey might have had "growing pains" that could have caused the 2002 survey to be different than the 2003 to 2013 surveys.

In Table 2 it can be seen that the residual standard error is low, which means that the model has a good fit. The survey year variable was included in the model as a set of indicator variables, so if we add the estimated coefficients from the model output to the intercept, we obtained the fitted proportions $\hat{p}_{cy}$. The unweighed and weighted sample sizes used in this analysis are shown in table 3.

Figure 6 shows the same result as Figure **??**, but disaggregated by birth cohort. The darker red indicates lower proportions and the lighter yellow indicates higher proportions. It can be noted that there does not seem to be an increase from left to right, which would indicate that our hypothesis was right. There does seem to be a cohort effect in which earlier cohorts are less likely to report their marijuana use.

4.2 Diagnostic tests for the weighted least squares model

We ran the model in Equation 9 and verified that the regular least squares assumptions were not violated. Figures 8 and 9 in the Appendix shows our diagnostics plots.

The Q-Q plot shows that the points lie almost exactly on the diagonal line, indicating that our residuals are approximately normally distributed. The standardized residuals are pretty evenly distributed in all directions (up, down, right, and left) and about the zero residual line. There are around seven points outside the boundary lines, but they are evenly distributed so they do not invalidate the results. Indeed, the results hold if we run the regression with and without the outliers. The studentized residuals vs. fitted values also look evenly distributed and are all within the adequate levels, which were adjusted with the Bonferroni correction. The Cook's distance shows that no point is near 1, so there are no significant outliers. The leverage and residuals vs. leverage plots show that there are no clear outliers. Overall, the diagnostics show that our modeling assumptions hold.

Additional diagnostic plots for the model in Equation 9. The standardized residuals are evenly distributed when plotted vs. the indicator survey year variables and the linear cohort variable. The standardized residuals vs. the square and cubic cohort terms are right-skewed, which would suggest that the outcome variable might need to be log-transformed. However, when it is log-transformed, the model has a worse fit. Therefore, we left the

	WLS model
(Intercept)	0.38^{***}
Survey year 1980	-0.02
Survey year 1991	-0.10^{***}
Survey year 1992	-0.14^{***}
Survey year 1993	-0.13^{**}
Survey year 1994	-0.13^{**}
Survey year 1995	-0.12^{**}
Survey year 1996	-0.13^{*}
Survey year 1997	-0.18^{***}
Survey year 1998	-0.14^{*}
Survey year 1999	-0.18^{**}
Survey year 2000	-0.18^{**}
Survey year 2001	-0.17^{*}
Survey year 2002	-0.15^{*}
Survey year 2003	-0.13
Survey year 2005	-0.14
Survey year 2004	-0.16
Survey year 2005	-0.15
Survey year 2006	-0.19
Survey year 2007	-0.15
Survey year 2008	-0.15
Survey year 2009	-0.19
Survey year 2010	-0.19
Survey year 2011	-0.21
Survey year 2012	-0.19
Survey year 2013	-0.19
poly(birth_cohort, 2)1	2.13^{***}
poly(birth_cohort, 2)2	-0.16^{***}
I(age^2)	0.00
R^2	0.95
Adj. R^2	0.94
Num. obs.	260
RMSE	1.03

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Statistical models

	WLS model	
(Intercept)	0.527	***
Survey year 2003	−0.008	
Survey year 2004	−0.005	
Survey year 2005	−0.024	***
Survey year 2006	−0.030	***
Survey year 2007	−0.032	***
Survey year 2008	−0.019	**
Survey year 2009	−0.031	***
Survey year 2010	−0.030	***
Survey year 2011	−0.036	***
Survey year 2012	−0.042	***
Survey year 2013	−0.022	***
Cohort	0.036	***
Cohort^2	−0.004	***
Cohort^3	0.000097	***
R^2	0.82	
Residual standard error	0.0004	
Num. obs.	252	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Results from the weighted least squares model in Equation 9. The coefficients of the survey year indicator variables all seem to be different from the 2002 baseline term. However, the only year that has a significantly different proportion of reported users is 2002 (see Figure 5).

model as it is in Equation 9.

4.3 Updates to the currently used adjustment

4.4 Sensitivity checks

4.4.1 Choosing "used by age 25" as our outcome variable

If our model yielded significantly different results if we switched the outcome variable to be a few years earlier or later (e.g. used by age 23 or 27), this would suggest that our model is highly dependent to the age 25.

We generated two new outcome variables: "used by age 23" and "used by age 27", and ran the model from Equation 9. We verified that the coefficients on the survey year indicator variables do not change by more than 0.001 in either direction. This means that our results are robust to choosing different ages for our outcome variable.

Unweighted	Unweighted	Weighted	Weighted percentage
Total respondents in public use file	664,548	2,985,724,671	100%
Avg. respondents per year	55,379	248,810,389	8.3%
Avg. respondents per cohort	941	7,887,460	0.2%

Table 3: Number of respondents per survey. NSDUH oversamples the youth, so the later cohorts tend to have a larger sample size than the earlier ones.

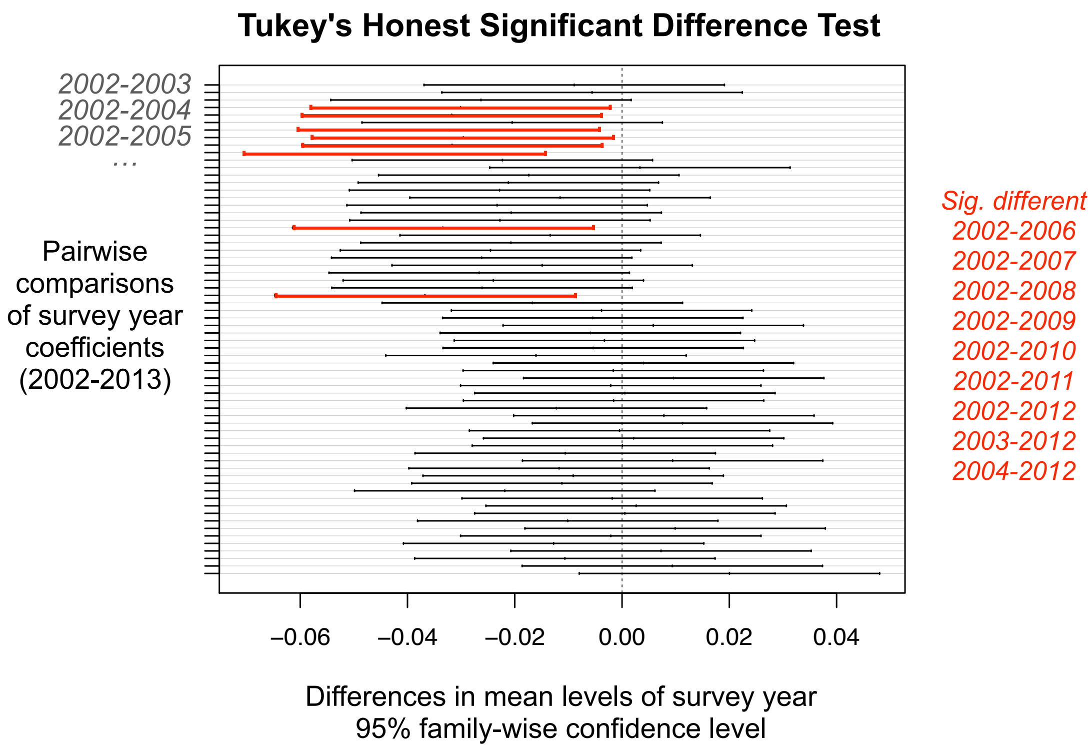

Figure 5: Results from Tukey Honest Significant Differences test. The y-axis is the pairwise comparison of all the survey years, from 2002 to 2013. The bars that cross the zero line shows that the estimates for those two years not significantly different at the 95% confidence level, and the ones that do not are significantly different. The 2002 estimates are significantly different from the 2005–2013 estimates, and 2012 is significantly different from 2003 and 2004. Since 2002 is the first year in which cash incentives were offered for taking the survey, it is likely that this caused some differences compared to 2003–2013. Otherwise, there was no significant change in the estimate between years.

5 Discussion

We estimated the consistency of responses about ever-use of marijuana by age 25 over time in the National Survey on Drug Use and Health and found that responses are indeed

Table 4: My caption

Year	Update to adjustment	e.g. If 1.33 was selected in 2000
1979	0.72	0.96
1982	0.74	0.98
1985	0.76	1.01
1988	0.78	1.04
1990	0.80	1.06
1991	0.82	1.09
1992	0.84	1.12
1993	0.86	1.14
1994	0.88	1.17
1995	0.90	1.20
1996	0.92	1.22
1997	0.94	1.25
1998	0.96	1.28
1999	0.98	1.30
2000	1.00	1.33
2001	1.02	1.36
2002	1.04	1.38
2003	1.06	1.41
2004	1.08	1.44
2005	1.10	1.46
2006	1.12	1.49
2007	1.14	1.52
2008	1.16	1.54
2009	1.18	1.57
2010	1.20	1.60
2011	1.22	1.62
2012	1.24	1.65
2013	1.26	1.68

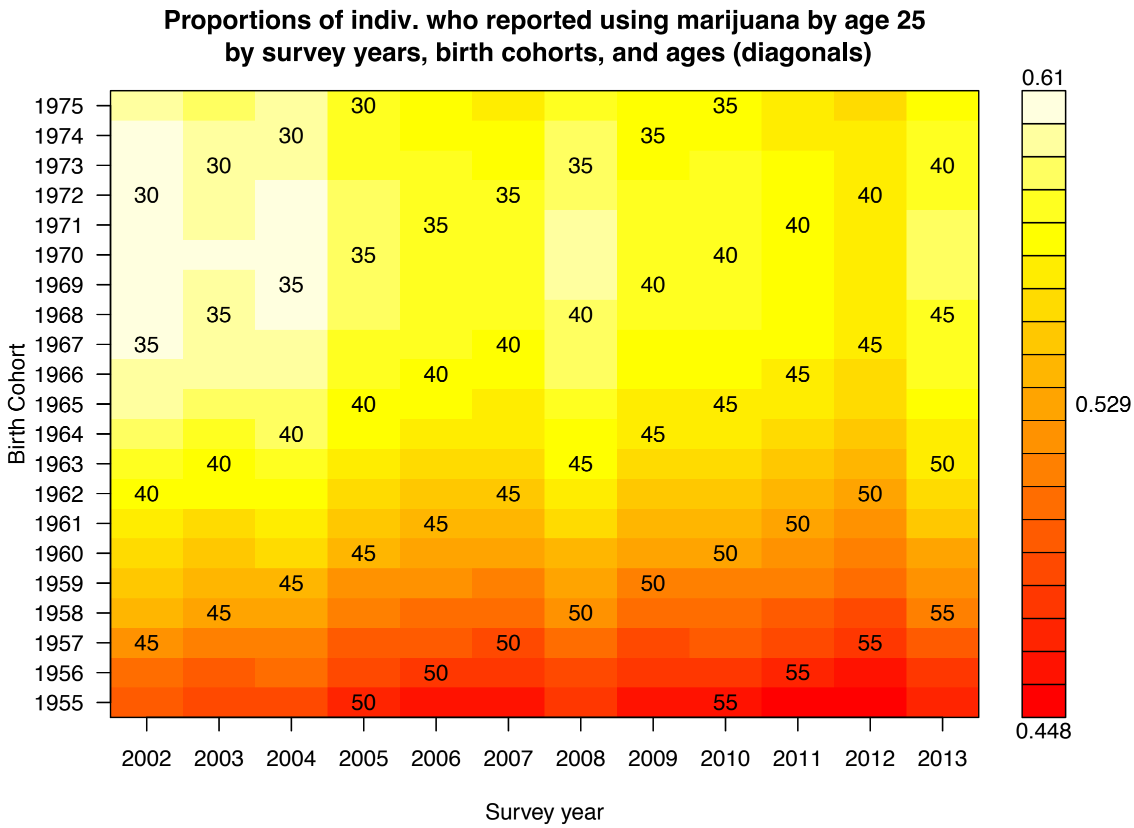

Figure 6: Image plot of the proportion of individuals who reported using marijuana by age 25. The vertical direction shows period effects, the horizontal direction shows cohort effects, and the diagonal direction shows age effects. There is a clear cohort effect: Later cohorts report their use less. The lighter spot in surveys 2002–2004 for cohorts 1957–1963 could be due to changes in stratification in 2005. The fact that the colors are similar over the vertical lines suggests that there is no period effect. That means that individuals have not changed their reported use between 2002 and 2013, which contradicts our hypothesis.

consistent. In other words, we have shown that individuals from the same birth cohort were not more likely to report their marijuana use in 2013 than in 2002.

This has three primary implications for policy. First, the current adjustment is adequate in being constant. Although we do not know whether the adjustments used by various researchers are correctly inflating the proportions of reported users to get the true number of users, we know that the adjustment used should be constant in time, at least from 2002 to 2013. Second, we can run our model as a monitoring system to determine if and when the adjustment needs to be revised. And third, we have developed a low-cost way to update the adjustment required to translated reported marijuana prevalence to true marijuana prevalence. Previously, researchers would need to perform costly physical tests on the population or extrapolate alcohol and tobacco reporting trends even though these drugs and marijuana differ in many ways. Now, we can find the changes in reporting cheaply and quickly by using our method. In addition, this method could be adapted to other drugs. Figure 7 shows a preliminary result of our model application to cocaine data in the NSDUH survey. It is well documented (Harrison and Hughes, 1997) that stigma about cocaine use dropped dramatically in the 1980's. In the early 1980's using cocaine was seen as an upper-class behavior, and later crack cocaine was introduced. This caused cocaine to become a highly stigmatized and lower-class drug.

Our method compares the proportion of individuals who reported using cocaine by age 25. Our method shows that the proportion of individuals who reported using cocaine by age 25 drops throughout the 1980's, while accounting for some significant survey design changes. We pooled 10 birth cohorts to obtain more power (fewer people use cocaine than marijuana).

Further examination is needed to find trends in reporting other drugs in different time periods.

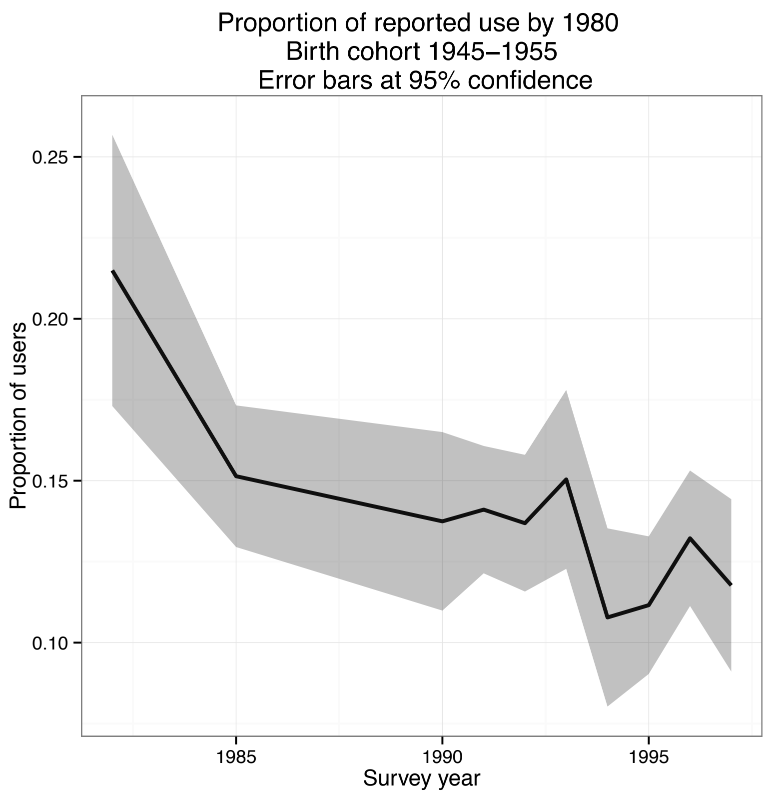

Figure 7: Preliminary plot of our method applied to cocaine data in 1980–1990 NSDUH. Our method shows that the proportion of individuals who reported using cocaine by age 1980 drops throughout the 1980's, despite some significant survey design changes.

6 Appendix

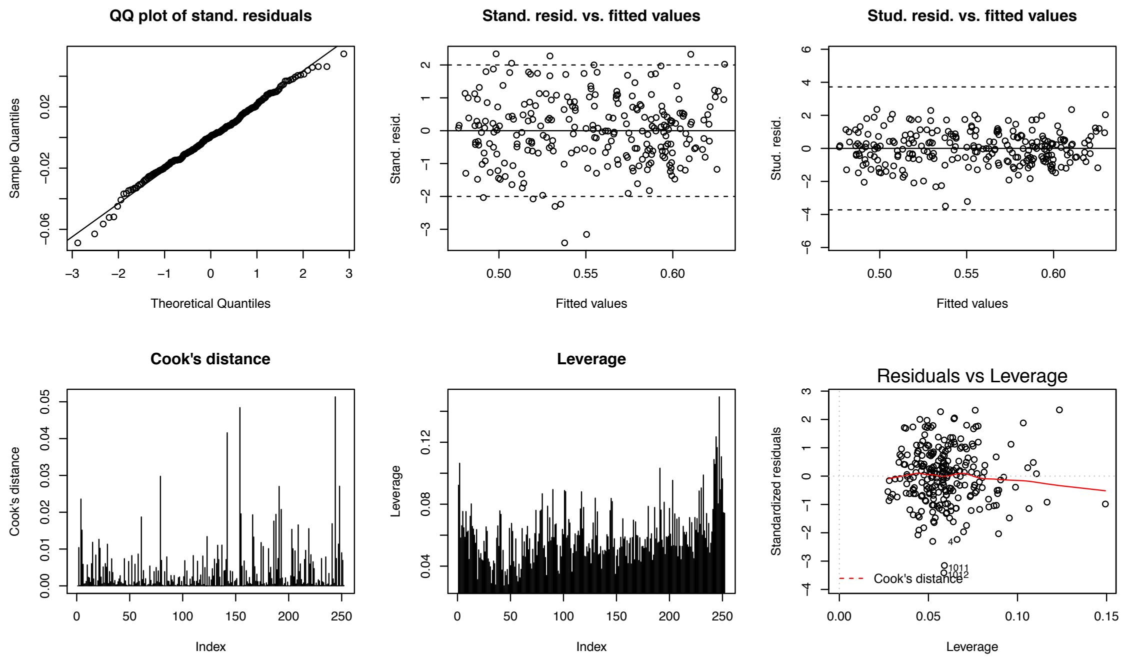

Figure 8: Diagnostic plots for the model in Equation 9. The Q-Q plot shows that the points lie almost exactly on the diagonal line, indicating that our residuals are approximately normally distributed. The standardized residuals are pretty evenly distributed in all directions (up, down, right, and left) and about the residuals=0 line. There are around seven points outside the boundary lines, but they are evenly distributed so they do not invalidate the results. Indeed, the results hold if we run the regression with and without the outliers. The studentized residuals vs. fitted values also look evenly distributed and are all within the adequate levels, which were adjusted with the Bonferroni correction. The Cook's distance shows that no point is near 1, so there are no significant outliers. The leverage and residuals vs. leverage plots show that there are no clear outliers. Overall, the diagnostics show that our modeling assumptions hold.

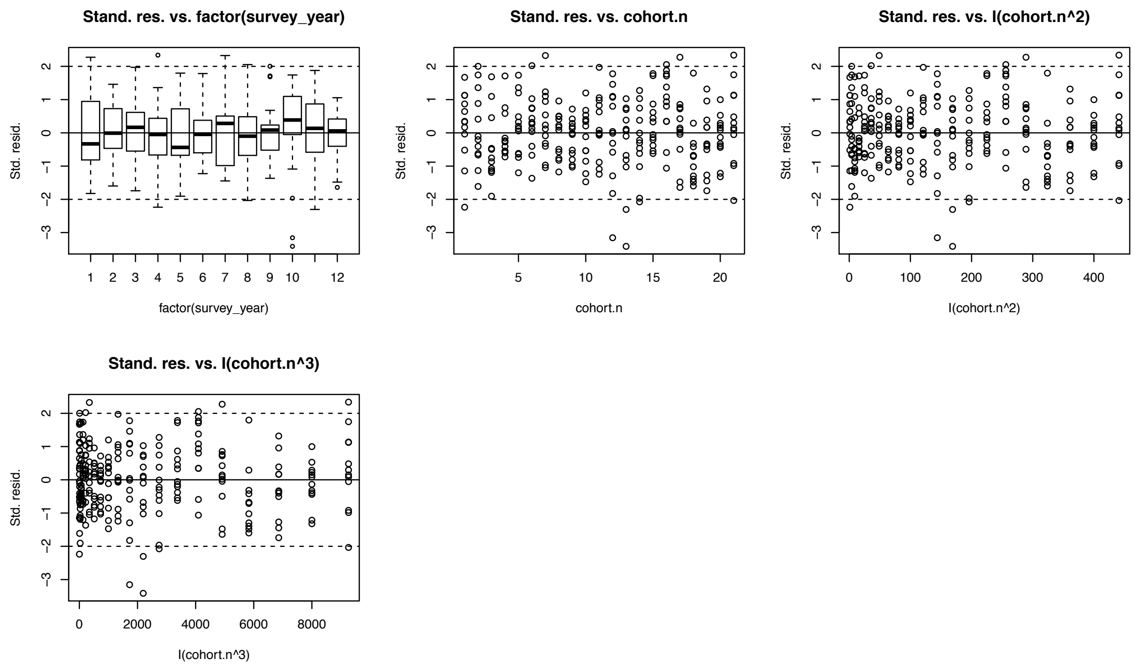

Figure 9: Additional diagnostic plots for the model in Equation 9. The standardized residuals are evenly distributed when plotted vs. the indicator survey year variables and the linear cohort variable. The standardized residuals vs. the square and cubic cohort terms are right-skewed, which would suggest that the outcome variable might need to be log-transformed. However, when it is log-transformed, the model has a worse fit. Therefore, we left the model as it is in Equation 9.

References

Ball, J. C. (1967), "The Reliability and Validity of Interview Data Obtained from 59 Narcotic Drug Addicts," *American Journal of Sociology*, 72, 650–654.

Conway, G., ColleyNiemeyer, B., Pursley, C., Cruz, C., Burt, S., Rion, P., and Jr., C. H. (1989), "Underreporting of AIDS cases in South Carolina, 1986 and 1987," *Journal of the American Medical Association*, 262, 2859–63.

Davis, R. C. and Henderson, N. J. (2003), "Willingness to Report Crimes: The Role of Ethnic Group Membership and Community Efficacy," *Crime and Delinquency*.

Fendrich, M., Johnson, T., Wislar, J., Hubbell, A., and Spiehler, V. (2004), "The utility of drug testing in epidemiological research: results from a general population survey," *Addiction*, 99, 197208.

Fendrich, M. and Vaughn, C. M. (1994), "iminished Lifetime Substance Use Over Time: An Inquiry Into Differential Underreporting," *The Public Opinion Quarterly*, 58, 96–123.

Fuchs, M. (2008), *Encyclopedia of Survey Research Methods*, Sage Publications, Inc., chap. Underreporting, p. 924.

Gettman, J. (2007), "Lost Taxes and Other Costs of Marijuana Laws," Available at http://norml.drugsense.org/downloads/Gettman_Lost_Taxes.pdf, Last accessed: 03/15/2015.

Gile, K. J. and Handcock, M. S. (2010), "Respondent-driven sampling: An assessment of current methodology," *Sociological Methodology*, 40, 285–327.

Harrison, L. and Hughes, A. (1997), "The Validity of Self-Reported Drug Use: Improving the Accuracy of The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates," *NIDA Research Monograph*, 167, 1–514.

Harrison, L. D., Martin, S. S., Enev, T., and Harrington, D. (2007), "Comparing Drug Testing and Self-Report of Drug Use among Youths and Young Adults in the General Population. A report from the Substance Abuse and Mental Health Services Administration (SAMHSA)," Available at http://calabria.dronet.org/comunicazioni/news/drugTest.pdf, Last accessed: 03/15/2015.

Hser, Y., Anglin, M. D., and Chou, C. (1992), "Reliability of retrospective self-report by narcotics addicts," *Psychological Assessment*, 4, 207–213.

Hser, Y., K, M. M., and Boyle (1999), "Validity of selfreport of drug use among STD patients, ER patients, and arrestees," *American Journal of Drug and Alcohol Abuse*, 25, 81–91.

Johnston, L. D. and O'Malley, P. M. (1997), "The recanting of earlier reported drug use by young adults. The Recanting of Earlier Reported Drug Use by Young Adults," *NIDA Research Monograph 167*, 167, 59–80.

Kilmer, B., Everingham, S., Caulkins, J., Midgette, G., Pacula, R., Reuter, P., Burns, R., Han, B., and Lundberg, R. (2014), "What America's Users Spend on Illegal Drugs: 2000-2010. A report for the Office of National Drug Control Policy (ONDCP)," Available at http://www.whitehouse.gov/sites/default/files/ondcp/policy-and-research/wausid_results_report.pdf, Last accessed: 03/15/2015.

Kilmer, B. and Pacula, R. (2009), "Estimating the Size of the Global Drug Market: A Demand-Side Approach," Report 2. TR-711-EC. RAND Corporation, Santa Mon- ica, CA. Available from http://www.rand.org/pubs/technicalreports/TR711, last accessed April 14, 2011.

Kruttschnitt (2014), *Estimating the Incidence of Rape and Sexual Assault*, The National Academies Press.

Lohr, S. L. (2009), *Sampling Design and Analysis*, Cengage Learning.

McAllister, I. and Makkai, T. (1991), "Correcting for the Underreporting of Drug Use in Opinion Surveys," *The International journal of the addictions*, 26, 945.

NSDUH (2014), "National Survey on Drug Use and Health, 2013 Public Use File Codebook. Substance Abuse and Mental Health Services Administration," Available at http://www.icpsr.umich.edu/icpsrweb/SAMHDA/sda, Last accessed 03/06/2015.

Pacula, R. L., Kilmer, B., Wagenaar, A., Alexander, C., Chaloupka, Frank, J., and Caulkins, J. P. (2014), "Developing Public Health Regulations for Marijuana: Lessons From Alcohol and Tobacco," *American Journal of Public Health*, 104, 1021–1028.

SAMHDA (2014), "Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings," Online at http://www.samhsa.gov/data/sites/default/files/NSDUHresultsPDFWHTML2013/Web/NSDUHresults2013.pdf. Last accessed 04-17-2015.

Smith, L., Adler, N., and Tschann, J. (1999), "Underreporting sensitive behaviors: the case of young women's willingness to report abortion," *Health Psychology*, 18, 37–43.

Tourangeau, R. (2007), "Sensitive Questions in Surveys," *Psychological Bulletin*, 133, 859 – 883.

Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M., and Bates, N. (2014), *Hard-to-Survey Populations*, Cambridge University Press.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge University Press.

Verbeek, M. and Nijman, T. (1992), "Can Cohort Data be Treated as Genuine Panel Data?" *Empirical Economics*, 17, 9–23.

Zullino, D., Krenz, S., Eap, C. B., Benguettat, D., and Khan, R. (2008), "Over- and Underreporting of Recent Drug Use in Subjects Entering an Inpatient Detoxification Unit," *European Journal of Medical Research*.