CLUSTERS OF CANNABIS SMOKING IN UNITED STATES SECONDARY SCHOOLS: 1976-2013

By

Maria A. Parker

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Epidemiology – Master of Science

2015

ABSTRACT

CLUSTERS OF CANNABIS SMOKING IN UNITED STATES SECONDARY SCHOOLS: 1976-2013

By

Maria A. Parker

A prevailing epidemiological theory about occurrence of drug use among secondary school students is that use follows trends in perceived risk of drug-related harms. If so, one might expect occurrence and clustering of drug use to occur more often in schools with concurrent or previously low levels of drug risk perceptions. This thesis aims to estimate the degree to which cannabis use might be clustering among secondary school students in the United States and to investigate a hypothesis about the prediction from the senior class's cannabis risk perceptions in one school year to the occurrence of newly incident cannabis use in the next year's senior class. Each year from 1976-2013, roughly 16,000 12th graders in ~133 schools completed questionnaires with standardized survey items for the Monitoring the Future study. The statistical approach harnessed Alternating Logistic Regressions to derive pairwise odds ratio estimates (PWOR), with PWOR > 1 providing evidence of clustering and a possible 'contagion' process, as well as regression slopes to estimate effect of prior year risk perception on next year risk of initiating cannabis use. The PWOR estimate is consistent with modest clustering of cannabis use suggestive of within-school social sharing of cannabis or 'contagion' (PWOR = 1.11; 95% CI = 1.06, 1.16). Statistically robust regression slope estimates suggest a lower risk of becoming a newly incident user for each risk perception unit (OR = 0.10; 95% CI = 0.03, 0.33). The most important discovery might be that school-level risk perceptions of 12th graders in one year may account for occurrence of newly incident cannabis use among seniors the next year. A causal link can be confirmed by experimental manipulation of perceived risk via public health interventions.

ACKNOWLEDGEMENTS

I would like to express appreciation for Professor James C. Anthony and Professor Qing Lu for their mentorship, methodological guidance, and patience while writing this thesis. In addition, I would like to thank the Monitoring the Future (MTF) team for making the datasets available and specifically Deborah D. Kloska for explaining the intricacies of MTF data and Timothy Perry for enabling access to the variables of interest. Lastly, Dr. Randy Hillard has been instrumental in his approach to my epidemiological research. This work would not have been possible without these key individuals. The National Institute on Drug Abuse (T32DA021129; K05DA015799) and Michigan State University supported this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AIC – Akaike Information Criterion

ALR – Alternating Logistic Regressions

CI – Confidence Intervals

GEE – Generalized Estimating Equations

ICC – Intraclass Correlation

OR – Odds Ratio

MTF – Monitoring the Future

NSDUH – National Surveys on Drug Use and Health

PWOR – Pairwise Odds Ratio

QIC – Quasi-Likelihood Information Criterion

SAMHSA – Substance Abuse and Mental Health Services Administration

UNODC – United Nations Office on Drugs and Crime

US – United States

CHAPTER 1

INTRODUCTION

The study of disease clustering in space and time has been a focus of epidemiology dating back to John Snow and his cholera maps in 19th century England (1). Since Snow's work and Charles V. Chapin's invention of the secondary attack rate in the 1800s, epidemiology has developed new statistical tools to quantify the degree to which diseases (or health behaviors) cluster in space and time. Relying heavily on quantitative methods especially the Generalized Estimating Equations and Alternating Logistic Regressions (GEE; ALR), this thesis aims to explore the degree to which cannabis smoking clustering in secondary school students in the United States (US) and to investigate a hypothesis about the prediction from cannabis risk perceptions in one school year to the occurrence of newly incident cannabis use during the next school year.

1.1 Specific Aims

1. Drawing upon the ALR and large sample school survey data from the US, to estimate trends in school-level clustering of newly incident cannabis smoking among 12th graders during intervals of stability and change in the risk of becoming a cannabis user.
2. To answer a predictive (and possibly causal) question: To what extent is a 12th grader's risk of starting to smoke cannabis in a given year determined by cannabis risk perceptions of 12th graders in the prior school year?

1.2 Background

Anthony & Van Etten proposed five rubrics of epidemiology (2):

(i) Quantity: How many?

(ii) Location: Where?

(iii) Causes: Why?

(iv) Mechanisms: How?

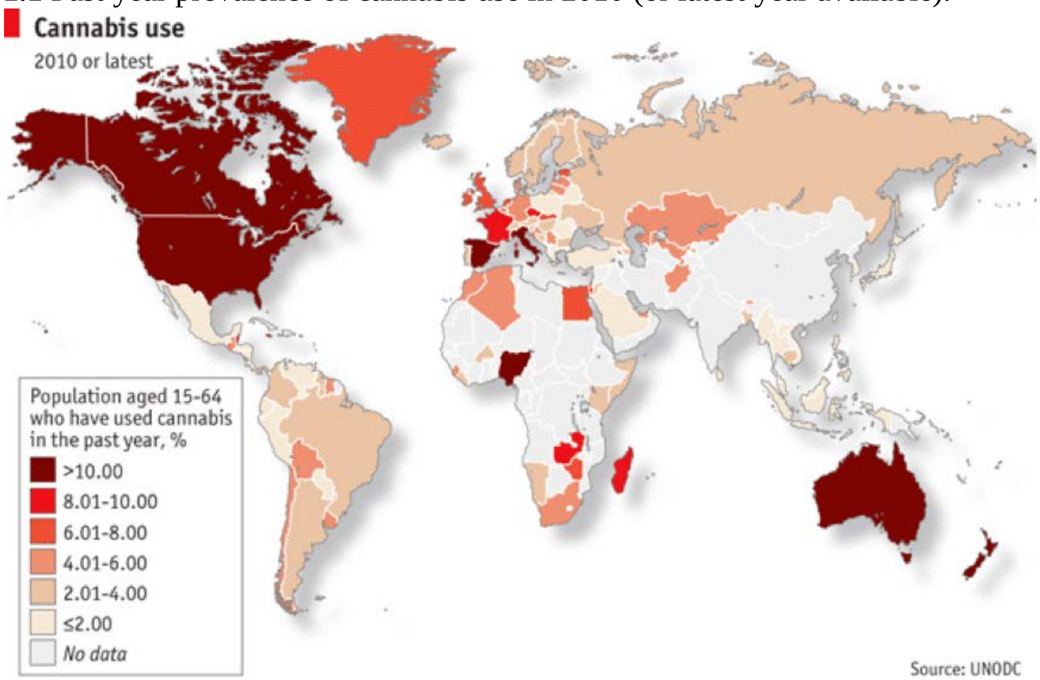
(v) Prevention and control: What can be done?

This thesis background is organized to reflect the use of epidemiology as a ‘lens’ in relation to these rubrics (2). As such, cannabis smoking in secondary school students of the US will be described in these five ways (i-v).

1.2.1 Quantity & Location

With respect to the rubric of quantity, cannabis (marijuana, hashish) is the most commonly used internationally regulated drug (3). In 2014, there were approximately 181.8 million current users worldwide (4). With respect to the rubric of location, annual prevalence is highest in the Americas, which is driven by consumption in North America. Oceania follows closely behind the Americas with high use in both New Zealand and Australia. In Asia, use is the lowest compared to other areas. Europe’s prevalence lies between that of Oceania and Asia (see Figure 1.1 (4,5).

Figure 1.1 Past year prevalence of cannabis use in 2010 (or latest year available).



According to estimates from the World Mental Health Surveys consortium report, the highest cumulative incidence (sometimes called lifetime prevalence) of cannabis use occurs in the US (42.4%) and the lowest is in China at 0.3% (6). In the most recent Substance Abuse and Mental Health Services Administration (SAMHSA) report for the US, there were an estimated 2.4 million newly incident cannabis users aged 12 and older in 2013 (began use in the past year). Of these newly incident users, 1.4 million began before age 18 (7).

In US high school students, according to the 2015 World Drug Report, cannabis use has been increasing in the past year (4). Based on the Monitoring the Future (MTF) study report dated December 2014, a nationally representative school based survey conducted over the past 38 years, approximately 48% of high school seniors had ever used cannabis in their lifetime (8). Most were recently active users: an estimated 37% 12th graders had used cannabis in the past year (prior to assessment date); 23% used in the past month (8). These MTF estimates are not too distant from corresponding estimates for 18-25 year olds, as surveyed for SAMHSA's 2013 National Surveys on Drug Use and Health (NSDUH), as shown in Table 1.1.

Table 1.1 Trends in prevalence of cannabis for individuals age 12 or older (in percent). Data from the National Survey on Drug Use and Health, United States 2013.[a]

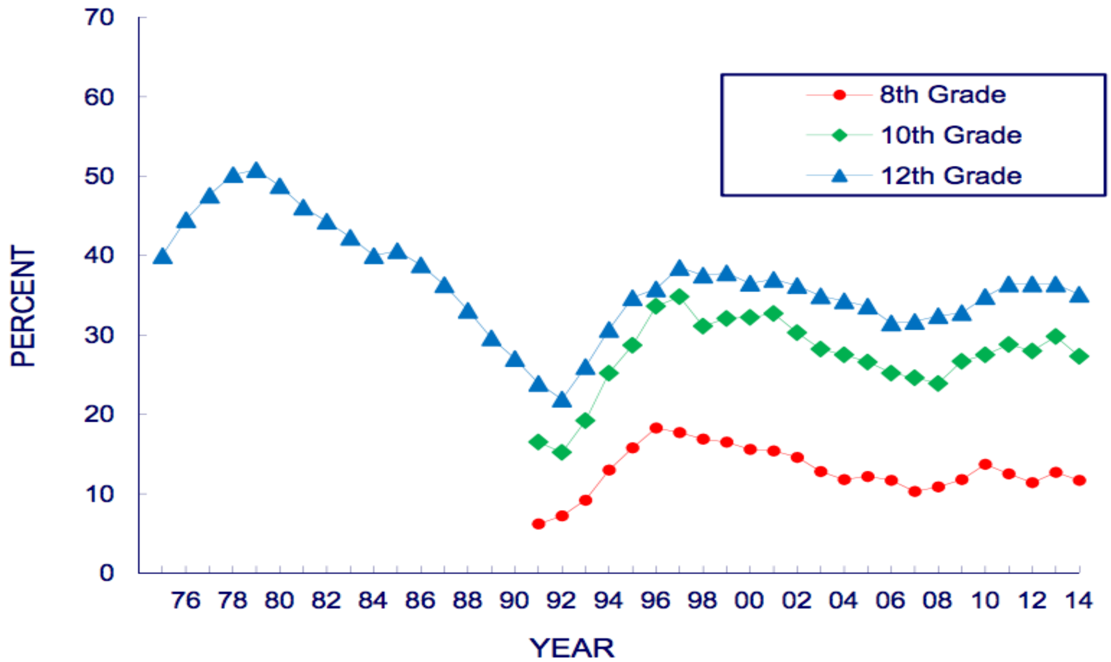
Time Period	Ages 12 or Older	Ages 12-17	Ages 18-25	Ages 26 or Older
Lifetime	43.7	16.4	51.9	45.7
Past year	12.6	13.4	31.6	9.2
Past month	7.5	7.1	19.1	5.6

[a] Adapted from http://www.drugabuse.gov/drugs-abuse/marijuana

In the United States, and in other countries (e.g., see Degenhardt et al., [11]), age is most commonly associated with cannabis use for both prevalence and incidence. For example, Table 1.1 shows how by and large lifetime prevalence increases across age strata. Nevertheless, the majority (56.6%) of newly incident users started using cannabis before 18 years old. This statistic has been similar for the past five years (7). Since 2002, the mean age of first use of cannabis was 17.5 years; for people who began their use prior to age 21, the average age was 16.2 years in 2013 (7).

In US students, annual prevalence of cannabis use increases by grade (see Figure 1.2; 8). Figure 1.2 also shows the time trend of cannabis annual prevalence estimates for US students. There is a noticeable peak in 1978-9 and in general a high prevalence between 1976-86 followed by declines until 1992, when prevalence doubled and then stabilized to a steady state. Cannabis incidence rates peaked around the same time as the annual prevalence at 21.0 per 1,000 potential new users, during the late 1970s (9). In 12-17 year olds, the average annual incidence rate was 6.1 (10). The incidence estimate during the trough after 1976 was 8.5 per 1,000 potential new users, mirroring that of the past year prevalence (9).

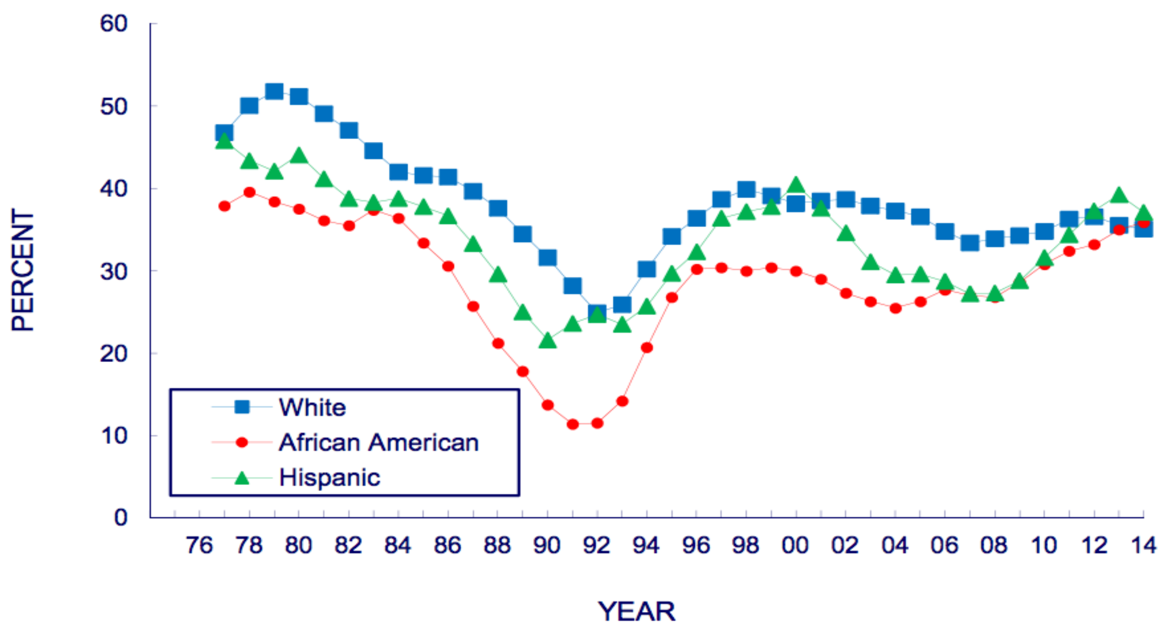
Figure 1.2 Trends in annual cannabis prevalence by grade. Data from the Monitoring the Future study, United States 1976-2014.



(A note about Figure 1.2 may be in order. The MTF study started its surveys of 8th and 10th graders in 1991. Therefore, there are no MTF estimates for these grades in the prior years.)

In probing other individual-level differences, males were more likely than females to be current users according to a recent NSDUH report (9.7% vs. 5.6%; 7). However, females have been more likely to use in past years. Recent studies have found that males are more likely to be prevalent cannabis users and females newly incident users (7,11,12). Generally, there are no sex differences in trends of cannabis use although over time in the US prevalence estimates for males tended to be larger (9). As for race/ethnicities, non-Hispanic Whites have been more likely to use in past years (13). Over time this facet of the epidemiology of cannabis use has changed with mixed race/ethnicity and non-White populations surpassing cannabis use estimates for Whites (7,9,14). Figure 1.3 shows that race/ethnicity differences seem to have disappeared in individuals attending school (8,13).

Figure 1.3 Twelfth grade trends in annual cannabis prevalence by race/ethnicity. Data from the Monitoring the Future study, United States 1976-2014.



In comparisons of public and private school students, past reports have shown less alcohol use and drug use in private schools (15). More recently, this gap has closed. Results from

2012 state more than half (54%) of private high school students attend schools that have drug use (16). This estimate is only 7% more than public drug using schools ten years before in 2002.

Within the US, regional variations in prevalence of cannabis use have been seen, but the top-ranked regions often change. For example, in 2013 the West had the highest prevalence of cannabis use among 12th graders, which happened to follow cannabis use patterns in all individuals 12 or older (7).

In general, population density (or urbanicity) of cannabis use is higher in larger urban/suburban areas versus rural areas (8). However, in 2014, the MTF report asserts that the top-ranked region for cannabis use was the Northeast. Although there are different divisions of metropolitan/non-metropolitan areas, this trend is similar in nationally representative samples (7).

1.2.2 Causes & Mechanisms

In epidemiology, working forward from John Snow's cholera studies and C.V. Chapin's measure of communicability of disease, the aim is to estimate the degree to which diseases (or health behaviors) cluster in space and time. Clustering could be possibly due to person-person spread of infections or between-person diffusion of innovations (behaviors, perceptions) (17).

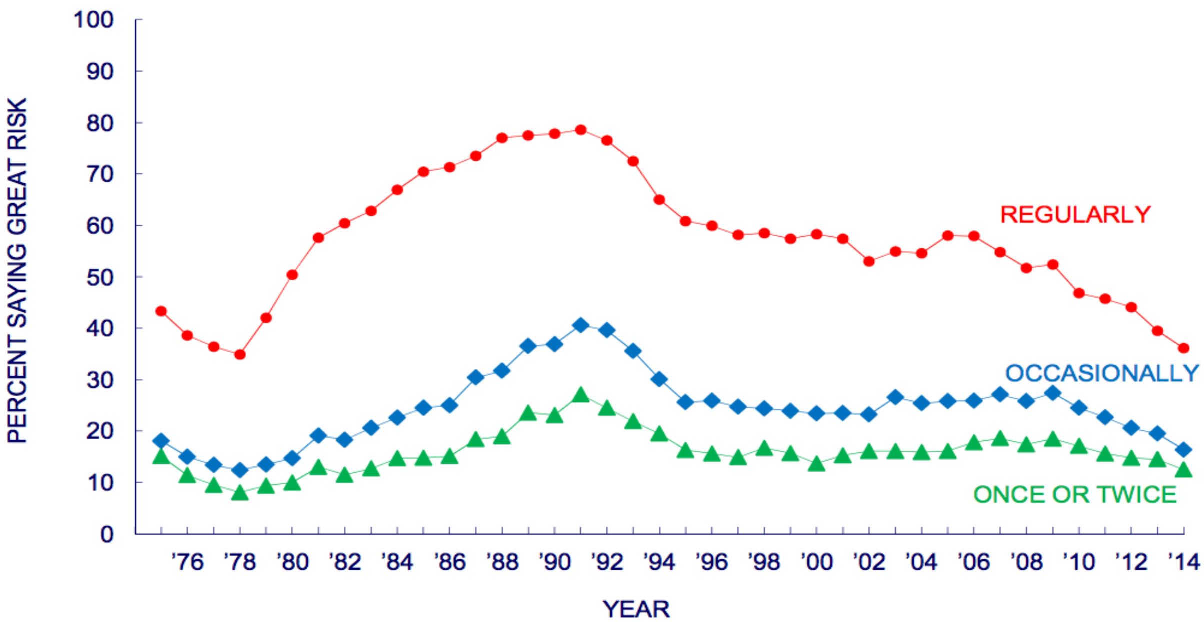
In research on drug use, the late Richard de Alarcon described person-to-person spread of heroin injection in his classic epidemiological studies 50 years ago (18). Later, Dishion and others introduced 'peer contagion' as an important National Institutes on Drug Abuse intervention research concept (19). Contagion has been described as a contextual effect of groups (20). With a focus on communicable disease, Susser and colleagues explain that prevalence affects the likelihood of an individual contracting the disease (20–22). Extending this research beyond transmissible infection, contagion also has been defined as the after effect of a prior effect, which encompasses behaviors, non-infectious disease, and other health outcomes (23).

Various reasons have been put forth to explain why some people use cannabis and others do not, with a range from micro-level genetic and epigenetic influences out toward macro-level societal or global influences such as the international psychotropic drug conventions. This thesis is focuses on school-level clustering of cannabis use, which stems from the idea of the 'contagion' concept. Cannabis and cocaine use clustering within neighborhoods of the US has been found (24,25). Cannabis and cocaine clustering within US communities is comparable to the magnitude of childhood diarrheal disease clustering among children in villages of the developing world (26). Although smaller in magnitude, underage drinking clustering in communities also has been found (27). Furthermore, preliminary findings suggest that the odds of drug use among school-attending youths increase when other youths use drugs in the same school (28). In this thesis research, clustering of cannabis smoking within schools should be expected, to the degree that there is social sharing of cannabis experience among student peers, perhaps with this contextual 'contagion' effect process such that an 'after effect' of one student's cannabis onset has later influence on the probability of other students' cannabis onsets. Other explanations of clustering put forth have included social networking, social norms, and self-efficacy (29,30).

In the late 20th century, social psychologists offered a now common theory about the rise and fall of drug epidemics. The basic idea is that drug use trends run in parallel with trend lines for changes in risk perception about drug use. Risk perception is a personal judgment made about a risk’s severity and its qualities (31). According to Bachman et al., risk perception is the number one predictor of drug use (32–35). The more perceived risk of harm, the less drug use there is

(36). This concept has been applied to cannabis smoking in secondary school students as observed in the MTF studies (32,33). Figure 1.4 shows trends of perceived risk of cannabis use among 12th graders, who were asked to rate harmfulness of using cannabis 'regularly,' with separate risk perception questions about using cannabis 'occasionally' and 'once or twice'). Comparing the trend in Figure 1.4 with that of cannabis prevalence among 12th graders from the same study in Figure 1.2, there is almost an inverse relationship.

Figure 1.4 Twelfth grade trends in perceived cannabis harmfulness by cannabis frequency. Data from the Monitoring the Future study, United States 1976-2014.



Additional extant theories explore the causes and mechanisms of cannabis use. The gateway concept has been a popular theme in drug use studies. More 'description' than 'theory', this theme emerged when fairly regular temporal sequences of drug use were observed. That is, initial alcohol and/or tobacco use leads to cannabis use and then cannabis use is followed by use of more toxic drugs such as cocaine and heroin, with toxicity defined in terms of risk of becoming drug dependence or suffering a drug overdose (37,38). In other words, Kandel and others describe cannabis as being a gateway drug for other internationally regulated drugs such as cocaine and heroin (37,39,40). Similarly, alcohol and/or tobacco have been shown to precede cannabis use and, therefore, are gateway drugs in their own right (40,41). While this idea has had much attention, there is considerable controversy surrounding it (42–45).

1.2.3 Prevention & Control

There is need for prevention and treatment of cannabis use and dependence, and other problems associated with cannabis use (e.g., driving under the influence). The number of individuals with cannabis problems outnumbers all other internationally regulated drugs (7). Although most people who start using cannabis never become dependent, there is not much evidence of control efforts in the published literature (46). If effective, primary prevention could stop people from beginning to start smoking cannabis and thereby reduce its incidence.

Past school based interventions have shown varying effectiveness (47–49). Family based approaches have shown prevention and reduction of developing problematic cannabis use (50,51). Recently, SAMHSA released a tool that provides summaries of prevention strategies and interventions for reducing youth cannabis use at the state and community levels (52). Future studies are needed on evaluation of more recent school-based prevention and intervention programs. Overall, clustering of newly incident cannabis smoking would suggest there is an opportunity for prevention/intervention which targets twelfth graders in the US.

1.3 Statistical Approach

Many epidemiological questions with correlated binary responses have been answered using Alternating Logistic Regressions (53,54). The nature of drug use is that it usually occurs in clusters (24,25,55,56). In this thesis research, due to shared environments, a pair of secondary school students sampled from the same school often will be more alike one another than two students from separate schools. Often marginal regression models have addressed within cluster association of this type. Implementing population-averaged generalized estimating equations (GEE) provides a way to deal with the correlation within clusters. ALR are an efficient second level GEE for the regression analysis of larger clustered binary data (57). Alternating Logistic Regressions (ALR) allow ‘population averaged’ modeling of the co-occurrence of drug use and initiation with large sample school data. ALR yield an odds ratio estimate that is easily understood (pairwise odds ratio; herein after, PWOR). Modeling newly incident cannabis use as unadjusted and adjusted can help identify student and school-level influences that may reduce and/or explain the magnitude of clustering. The ALR population average approach is appropriate from a public health perspective and can help target prevention and intervention programs based on these student and school-level influences.

CHAPTER 2

METHODS

2.1 Study Population & Sampling

The MTF study is a continuing study of US secondary school students and their behaviors and feelings about drug use and other social issues. This cross-sectional study is funded by the National Institute of Drug Abuse (part of the National Institute of Health) and conducted at the Survey Research Center in the Institute of Social Research at University of Michigan (8).

Each year between 1976 and 2013, the US MTF research team sampled and recruited approximately 135 public and private high schools for a nationally representative sample survey. Roughly 16,000 12th graders were assessed using institutional review board-approved group-administered self-report questionnaires every year. The study population was designated to include an accurate sample of US 12th graders each year. Students were presented with the same questions[1] for all 38 years to see how experiences have changed over time (58). Repeating these annual cross-sectional surveys over time allows an assessment of change across history in consistent age segments of the population, as well as among subgroups (8).

The sampling approach involved data collection in selected public and private high schools to provide a representative cross-section of 12th graders throughout the coterminous US each year. Each year, a multi-stage random sampling procedure was used to identify the study population. In stage one, geographical areas were selected. In the second stage, one or more schools in each geographical area were selected by accounting for size. The last stage selects classes within the schools. Up to 350 students may have been assessed in larger schools; in

[1] The Cross-Time Index shows questions for all grade 12 base year questionnaires from 1976-2010, sorted by subject area.

smaller schools, usually all of them were included. Schools were secured for two consecutive years; half of the schools are replaced every year. In this way, half of the schools are in the study for the second year and half are in the study for the first time. Weights are assigned and normalized so the weighted number of cases is equal the unweighted number of cases and to make up for any differences in selection probabilities. Few schools participate for only one year. School participation rates are 95% or above for all years (1976-2013). Replacement schools for schools that declined were matched geographically and by size (59).

Student participation rates averaged around 82-83% over all study years. About 1% of parents refused to let their child participate, less than 2% of students refused to complete the surveys, but most of the non-response was due to absenteeism (8). Others had missing or invalid responses to key study variables. For this reason, the effective sample size for the present investigation and the proportion of designated participants with useable data were n=9,417 for newly incident cannabis users and n=103,680 who had never used cannabis. After excluding those not asked about the key response variable (n=321,397) and those who had used cannabis in the past (n=83,203), this amounts to 58.2% of the entire 12th grade MTF sample between 1976-2013. All newly incident users had onset of cannabis use within the school year of survey assessment. Never users had never used cannabis in their lifetime.

2.2 Assessment & Measures

Twelfth grade students participated by completing 45-50 minute self-administered questionnaires in their normal classrooms during the spring each year. If this was not possible, the surveys may have occurred in a larger auditorium. Each MTF survey has six questionnaire forms that contain different content. Key drug use and demographic measures appeared in all forms (the entire sample of 16,000 12th graders is used). Other select forms included topics such as personal disapproval and perceived availability of various drugs. The minimum sample size for each form averaged around 2,300 each year (53). Standardized questionnaire items assessed background measures of interest (i.e., sex, race/ethnicity), grade of first cannabis use, as well as perceived risk of cannabis smoking. The ‘grade of first cannabis use’ question appeared on three of six forms, and the ‘perceived risk of cannabis smoking’ question was on five of six forms.

The key response variable in this study was about grade of first cannabis use. Of interest were 12th graders who began using marijuana in 12th grade (i.e., newly incident users). Newly incident use was measured via one survey question, “When (if ever) did you FIRST do each of the following things? Don’t count anything you took because a doctor told you to.” The selection of interest was “Try marijuana or hashish.” Answer choices were “Grade 6 or below; Grade 7; Grade 8; Grade 9 (Freshman); Grade 10 (Sophomore); Grade 11 (Junior); Grade 12 (Senior).” Seniors who first used before Grade 12 were excluded.

The main exposure of interest was perceived risk of cannabis smoking. Perceived risk was measured by three questionnaire items that use the same question root, “How much do you think people risk harming themselves (physically or in other ways), if they [try marijuana once or twice/smoke marijuana occasionally/smoke marijuana regularly]?” Answers included “No risk; Slight risk; Moderate risk; Great risk; Can’t say drug unfamiliar.” For the current study, a variable was created using only two of the questionnaire items by calculating the proportion of students who perceive “great risk” of: 1) ‘trying cannabis once or twice’ or 2) ‘smoking cannabis regularly in each school.

Other covariates under study included were at both the individual and school level. At the individual-level were sex (male vs. female), age in years, race/ethnicity (Non-Hispanic White, non-Hispanic black, Hispanic and other[2]), past alcohol use, and past tobacco cigarette use during the same period (prior to 12th grade). Past alcohol and past cigarette use were included because some degree of cannabis smoking clustering may have depended on the 12th graders' prior use following the 'gateway hypothesis' (37). School-level covariates of interest were region, population density, and whether or not it was public or private.

2.3 Data Analysis

The guiding conceptual model was one in which clustering of cannabis smoking within schools should be expected, to the degree that there is social sharing of cannabis experience among student peers. The plan for data analysis was organized in relation to standard "explore, analyze, explore" cycles, in which the first exploratory steps involve exploratory data analyses to shed light on the underlying distributions of the response variable and covariate of interest. In this work, precision of the study estimates were stressed with a focus on 95% confidence intervals (CI); p-values are presented as an aid to interpretation.

In the initial analysis step, the task was to describe the MTF participant sample of 12th grade users by demographic information, for which the statistical approach was estimating weighted and unweighted proportions. Next, ALR was used to estimate cannabis smoking clusters in schools. The null hypothesis was that all cannabis use occurring at random, with no underlying contagion processes. That is, sharing of drugs from student to student or use of drugs by each student within a school would not occur any more than would be the case if the only sharing of drugs was from student to student or use with students in peer groups aggregated across schools.

[2] Other included Asian, Native American/American Indian, Hawaiian/Pacific Islander, and other.

In subsequent analysis steps, the statistical approach involved creating year 'bins', as described for the history of stability and change in prevalence of cannabis use during the past 35-40 years. Previously published MTF 12th grade prevalence estimates guided how newly incident users were divided into cohorts, which reflected stability and change in occurrence of cannabis use (8). Cohorts of interest included: post-Vietnam high prevalence/incidence (1976-86), declines (1987-1992), rise and stabilizing (1993-2000), and steady state (2001-2013).

In addition, the entire sample of 12th graders was described by year cohort for all demographic information as well as for the response and outcome variables. Incidence rates were then estimated from 1976-2013.

2.4 Statistical Model

2.4.1 Alternating Logistic Regressions

The main analysis involved Generalized Estimating Equations (GEE; Alternating Logistic Regressions (ALR) specifically) to derive yearly PWOR estimates for evidence of 12th grade school-level newly incident cannabis smoking clusters each year. The GEE produces population-averaged estimates while considering correlation of the data (60). The ALR model uses first-order GEE when the outcome is binary to regress that outcome on covariates while simultaneously regressing the binary outcome in a school on others from the same schools (61). Unlike the traditional longitudinal approach, in this context, the clusters are secondary schools in the US.

The PWOR is defined in terms of possible pairs of individuals unlike the ordinary odds ratio, which is defined in terms of individuals. It measures the extent of clustering of an outcome among individuals. The PWOR can be described as a contextual ‘contagion’ effect (20,56). Unlike margin-sensitive alternatives (e.g., intraclass correlation coefficient), the PWOR does not depend upon prevalence/incidence of cannabis smoking. The clustering magnitude does not depend upon the marginal distributions. Also, in contrast to the intraclass correlation coefficient, the odds ratio estimate yielded from the ALR is easily understood by public health researchers and practitioners.

In this context, the PWOR reflects the odds of newly incident cannabis smoking for a 12th grader in a school given that another randomly chosen 12th grader from that school smokes relative to the odds if that randomly chosen 12th grader does not smoke. A PWOR > 1 provides evidence of co-occurring use or how many times more newly incident smoking occurs among 12th graders compared to what would be expected newly incident smoking were random. In other words, a PWOR > 1 indicates that the newly incident cannabis use of one 12th graders is statistically dependent upon the newly incident cannabis use of another randomly chosen 12th grader attending the same school, beyond the expectation of selecting random pairs of 12th graders and disregarding which school he/she attends. A PWOR=1 is null meaning no clustering. Procedures previously described guided this estimation of PWOR as an index of newly incident cannabis smoking within schools (26,57,62).

The PWOR is a specific parameter in the equation for the conditional expectation of cannabis smoking for a 12th grader, conditioning on the occurrence of newly incident cannabis smoking in another 12th grader chosen within the same school. Because only one level of clustering was of interest in this study, α only has one value (55). The association between pairs of 12th graders was modeled as follows:

$$\log\left(PWOR_{jk}\right) = \alpha_0 Z_{0jk}.$$

The logarithm of the PWOR is a function of an indicator variable that expresses whether a pair of 12th graders, *j* and *k*, are in the same school. Z_{jk} is a binary variable that takes value 1 when the pair belongs to the same school and takes value 0 otherwise (24).

As described in previously published work, the PWOR can also be described by examining a 2x2 table for outcomes that are paired (55). In this context, the rows correspond to whether the first 12th grader is a newly incident cannabis user or not and the columns correspond to whether the second 12th grader is a newly incident cannabis user (see Table 2.1). Each cell in the table has a probability of the *pair* to occur in each situation (i.e., both are newly incident cannabis smokers, a discordant pair, or neither are newly incident cannabis smokers). Similar to an ordinary odds ratio, the ratio of the four probabilities, p_{11}, p_{10}, p_{01}, p_{00}, is equal to the PWOR (24,55,57):

$$\frac{p_{11}p_{00}}{p_{10}p_{01}}$$

Table 2.1 Basic 2x2 Table for Estimation of Pairwise Odds Ratios.

	Second 12th grader in the pair	
First 12th grader in the pair	Newly incident cannabis smoker	Not a newly incident cannabis smoker
Newly incident cannabis smoker	p_{11}	p_{10}
Not a newly incident cannabis smoker	p_{01}	p_{00}

The resulting PWOR is comprised of a numerator, p_{11}/p_{01}, equal to the odds that both 12th graders in the pair are newly incident smokers, and a denominator, p_{10}/p_{00}, equal to the odds that one of the 12th graders is a newly incident smoker and the other is not. Taking a ratio of these two odds is equivalent to:

$$\frac{p_{11}/p_{01}}{p_{10}/p_{01}} = \frac{p_{11}p_{00}}{p_{10}p_{01}}$$

2.4.1.1 Estimation

ALR estimates PWOR for within school associations while simultaneously considering the dependence of newly incidence cannabis use on covariates. This method allows comparison of the PWOR across 12th graders in different schools over time. A logistic regression is *iteratively* used to control for potential school and student level variables:

$$\log\left(ODDS_j\right) = \beta_0 + \sum \beta_i X_i,$$

where X_is are the covariates for the jth student. The βs are the odds ratios for the risk of cannabis associated with the covariates. While accounting for the correlation of newly incident cannabis smoking within schools, the GEE method was used to estimate βs (24,26).

SAS version 9.4 'PROC GENMOD' with the LOGOR option on the REPEATED statement was used to estimate the within-school PWORs, regression coefficients, and robust standard errors for each year. The ALR algorithm alternates between a GEE and a logistic regression step until convergence. Each step updates the model: (i) using a first-order GEE, estimate β as a parameter in a marginal logistic regression for a given α; (ii) using a logistic regression of the outcome, estimate the OR parameter α for a given β. The GEE step is for the prevalence of the outcome and the logistic regression step is for the log odds ratio (57). ALR regression estimates are asymptotically normal. When the model converges, SAS provides regression parameter estimates for the mean (β), for the log odds ratios (α), the empirical standard errors, and their covariances (63).

2.4.1.2 Estimating Equations

The MTF data was obtained in clusters of secondary schools, with a binary outcome (newly incident cannabis smoking). Considering this data, for cluster i = 1, …, m, let $Y_i = (Y_{i1,\ldots,} Y_{in})'$ be a response vector with mean $E(Y_i) = \mu_i$, and let

$$A_i = Y_i - \mu_i, \qquad B_i = \text{cov}(Y_i), \quad C_i = \frac{\partial \mu_i}{\partial \beta}.$$

Let R_i denote the vector of residuals, S_i be the ${}_nC_2$ x ${}_nC_2$ diagonal matrix and T_i be the ${}_nC_2$ x q matrix. The following estimating equations are solved for β and α:

$$U_\beta = \sum_{i=1}^{m} C_i' B_i^{-1} A_i = 0 \qquad \text{and} \qquad U_\alpha = \sum_{i=1}^{m} T_i' S_i^{-1} R_i = 0.$$

Carey and colleagues have detailed these two unbiased estimating equations that the ALR estimates simultaneous solutions for (51).

2.4.1.3 Analysis Plan

First, an intercept only model was fitted to estimate associations of newly incident cannabis smoking within schools for each of the 38 years, 1976-2013. Because ALR estimates the PWOR and accommodates covariate adjustments, it was suspected that certain covariates might account for odds of each outcome of interest (here, odds of becoming a newly incident cannabis smoker) and/or might account for the magnitude of clustering. Initially sex and age were included separately in a model, then a model with both sex and age was estimated, and these models were evaluated.

Next, the one suspected causal determinant or covariate of central interest, risk perception cannabis use, was included. By comparing the unadjusted and adjusted estimates of the PWOR after adding covariates, the risk effect could be estimated (64). Perceived risk of trying once or twice and then perceived risk of regular cannabis smoking were introduced to the ALR model. Using schools that were surveyed two years in a row, the proportion of students who perceived great risk of cannabis use (either trying or regular smoking) from year 1 (t) were used in the model as a school-level covariate to predict incident use in year two (t+1).

After the year by year estimates, the most intriguing covariates of interest (after perceived risk) were introduced to a model for all years and year cohorts, past alcohol and past cigarette use. Both the odds ratios for the association between covariates (risk perception and past [alcohol/cigarette] use) and the outcome as well as the within-school PWOR adjusted for the covariates were obtained. Covariates for sex, age in years, race/ethnicity, and then school-level covariates (i.e., public vs. private, region, population density) were subsequently added with risk perception.

Although not obligatory to use, sample weights were included in all ALR models to adjust for oversampling of some demographic groups (see Appendix A for SAS code). Unweighted ALR model estimates were close to weighted model estimates. To test the equality of the PWOR in the contextual ALR model, Wald tests were performed.

2.4.2 Meta-Analysis

After estimation of year-specific PWOR, years were grouped in relation to stability and change in cannabis use trends (year cohorts) as explained in the Data Analysis section of the Methods. Meta-analysis was performed for each year cohort to examine PWOR for newly incident users for an intercept only model and a term for each school's level of cannabis risk perception at year t-1 was added to the regression model to estimate its prediction of cannabis onsets the next year (t). Meta-analysis is a quantitative method that summarizes the effects of several studies (65). In this context, it combines estimates rather than studies to create an overall summary estimate. Each year is weighted by the inverse of its variance. Natural logarithm estimates and lower and upper confidence limits were calculated before the meta-analysis was performed (66).

The two statistical models used to create the meta-analysis summary estimate are fixed-effects models and random-effects models. Fixed-effects models treat each parameter as fixed but as unknown and assume parameters are homogeneous (67). Random-effects models treat each parameter as if it were from a random population sample (67). In this study, Stata version

13 ‘metan’ were used. The command ‘metan’ uses a test of whether the summary effect measure is null and a test for heterogeneity is performed (68).

The test for heterogeneity numerically describes whether the true effect over all cohorts is the same; it is quantified using the I^2 measure (69).

$$I^2 = 100\% \times (\mathrm{Q} - \mathrm{df})/\mathrm{Q},$$

where Q is Cochran's heterogeneity statistic and df the degrees of freedom. Cochran's Q can be expressed by adding the squared deviations of each cohort's estimate from the meta-analysis summary estimate. Stata outputs p-values using a χ^2 distribution with k-1 degrees of freedom, where k is the number of cohorts (68). No heterogeneity occurs when I^2 is 0%, increasing heterogeneity is indicated by larger values (69).

When I^2 is small, its p-value is large, there is less cohort variation and a fixed-effects estimator should be used. When I^2 is large, its p-value is small, there is more cohort variation, and a random-effects estimator can be used. Following rules used previously, either the fixed- or random-effects estimator was used (see Table 2.2).

Table 2.2 Rule on whether to use the fixed- or random-effects meta-analysis summary estimator based on p-values for I^2.

	Type of estimator used
p-value<0.05	Random-effects
0.05<p-value<0.15	Both
p-value>0.05	Fixed-effects

After the initial meta-analyses were performed, for ease of interpretation, meta-analysis was completed to see exponentiated ALR parameter estimates (exp(α) = odds ratios) for perceived risk by year cohort.

2.4.3 Post-Estimation Exploratory Data Analyses

In post-estimation exploratory data analyses steps, the perceived risk variables were divided into quantiles to explore whether the intensity of risk perception (divided into four levels) in the year prior might disclose a non-linear pattern of association with newly incident cannabis smoking in the next year. ALR models by year cohort and perceived risk quantiles were estimated for trying once or twice as well as regular cannabis smoking.

Employing an adapted version of purposeful selection of covariates, all covariates were included in a multivariable model for all years and year cohorts (70). Covariates were only added using the screening criterion that the majority of univariable analysis and bivariate analysis estimates' p-values<0.20. The next step involved using a p-value of 0.05 as a retaining criterion. Model fit statistics (quasi-likelihood information criterion, QIC/QICu) were compared for the initial and reduced multivariable models (71). The Akaike information criterion (AIC) is not an appropriate to compare two models since GEE-based models are estimated without full likelihood specification. Each variable not selected with the initial retaining criterion was added back to the reduced model, using a Wald test for each covariate to compare changes in the values of the estimated coefficients, looking for a $\Delta\hat{\beta}$>20%. Lastly, the linearity assumption of age in years and perceived risk were checked due to their continuous nature (70). Note that consideration of the guiding conceptual model was utilized in every step of this post-hoc analysis (i.e., perceived risk was not removed due to its necessity based on the thesis specific aims).

CHAPTER 3

RESULTS

3.1 Characteristics of the Sample

Table 3.1 offers a description of the study sample; the MTF participants depicted can be regarded as a nationally representative sample of 12th grade users. Weighted percentages are close to unweighted percentages (results not shown in a table). In total there were 599,032 12th graders who participated in the MTF studies from 1976-2013. Some had never used cannabis and some were missing on the outcome variable (newly incident cannabis use). Overall, 113,097 12th graders were included in Table 3.1; there were 9,417 newly incident cannabis users and 103,680 never users. Approximately half of the sample was male (~45%). The mean age was 17.5 years. Distributions of age, sex, race/ethnicity and year cohorts appear similar for both groups of 12th graders. Figure 3.2 shows a flowchart of how students were selected for the final analytic sample.

Table 3.1 Selected characteristics of unweighted 12th grade newly incident cannabis users (n=9,417) and never users (n=103,680). Data from Monitoring the Future: Secondary School Students, United States 1976-2013.

Sample characteristics	Newly incident users	%[a]	Never users	%[a]
Sex				
Male	4,154	45.2	45,045	44.4
Female	5,047	54.9	56,349	55.6
Age (at interview)[b]	17.5	0.586	17.5	0.621
Race/ethnicity				
Non-Hispanic White	6,754	73.1	70,777	69.4
Non-Hispanic black	1,124	12.2	13,058	12.8
Hispanic	758	8.2	9,195	9.0
Other	604	6.5	9,005	8.8
Year cohorts				
1976-86 post-Vietnam high prevalence/incidence	2,918	31.0	25,988	25.1
1987-1992 declines	1,511	16.0	23,310	22.5
1993-2000 rise & stabilizing	2,325	24.7	24,671	23.8
2001-2013 steady state	2,663	28.3	29,711	28.7
Perceive great risk[b]				
For trying cannabis once or twice	0.146	0.088	0.171	0.010
For smoking cannabis regularly	0.595	0.166	0.637	0.161
Past alcohol use	7,415	81.3	57,089	62.5
Past cigarette use	3,859	60.5	22,194	29.6
School information				
Public	8,211	87.2	91,091	87.9
Private	1,206	12.8	12,589	12.1
Region				
Northeast	1,903	20.2	19,594	18.9
Midwest	2,614	27.8	27,880	26.9
South	3,025	32.1	35,931	34.7
West	1,875	19.9	20,275	19.6
Population Density				
Urban	2,992	31.8	31,677	30.6
Suburban	4,416	46.9	47,431	45.8
Rural	2,009	21.3	24,572	23.7

[a]Due to rounding, some percentages may not add to 100%.
[b]Mean with standard deviation.

Figure 3.2 Flowchart identifying newly incident 12th grade cannabis smokers. Data from the Monitoring the Future study, United States 1976-2013.

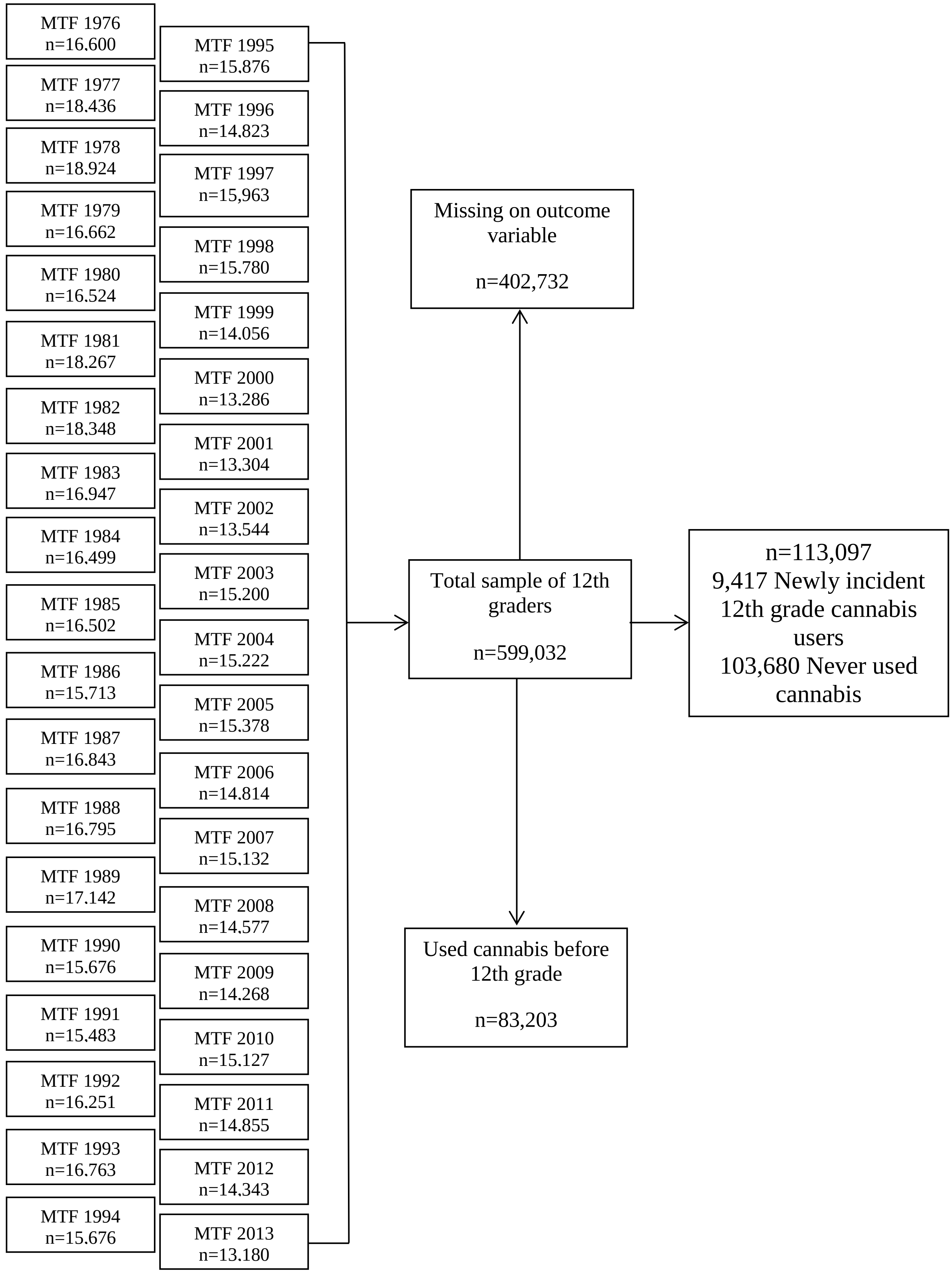


Table 3.2 shows a description of 12th graders by the main outcome and exposure variables. Here, all twelfth graders can be compared by cannabis experience as well as by risk perceptions of cannabis use.

Table 3.2. Distribution of 12th grader unweighted cannabis outcomes from 38 years of the Monitoring the Future study by year cohort (n=599,032), United States 1976-2013.[a]

	All years (n=599,032)		1976-1986 (n=189,422)		1987-1992 (n=98,190)		1993-2000 (n=122,476)		2001-2013 (n=188,944)	
	n	%[b]	n	%[b]	n	%[b]	n	%[b]	n	%[b]
Newly incident experience with cannabis	(n=196,300)		(n=60,696)		(n=39,503)		(n=43,004)		(n=53,097)	
First used in 12th grade	9,417	4.8	2,918	4.8	1,511	3.8	2,325	5.4	2,663	5.0
First used prior to 12th grade	83,203	42.4	31,790	52.4	14,682	37.2	16,008	37.2	20,723	39.0
Never used	103,680	52.8	25,988	42.8	23,310	57.4	24,671	57.4	29,711	56.0
Perceived risk of cannabis use										
Trying once or twice	(n=332,511)		(n=35,945)		(n=49,795)		(n=97,514)		(n=149,257)	
Great risk	53,399	16.0	4,151	11.6	11,233	22.6	15,798	16.2	22,217	14.9
Moderate risk	50,648	15.2	4,172	11.6	9,685	19.5	15,257	15.7	21,534	14.4
Slight risk	110,627	33.3	11,069	30.8	17,951	36.0	33,509	34.4	48,098	32.2
No risk	117,837	35.4	16,553	46.0	10,926	21.9	32,950	33.8	57,408	38.5
Smoking regularly	(n=331,895)		(n=35,918)		(n=49,751)		(n=97,355)		(n=148,871)	
Great risk	196,824	59.3	19,397	54.0	38,740	77.9	60,784	62.4	77,903	9.7
Moderate risk	71,992	21.7	8,861	24.7	7,435	14.9	20,711	21.3	34,985	14.5
Slight risk	39,077	11.8	5,161	14.4	2,205	4.4	10,095	10.4	21,616	23.5
No risk	24,002	7.2	2,499	7.0	1,371	2.8	5,765	5.9	14,367	52.3

[a] Data used for respondents with valid values; those missing were excluded.
[b] Due to rounding, percentages may not add to 100%.

Table 3.3 displays how both past users and newly incident users are similar and different on both individual-level variables displayed in Table 3.1 as well as school-level variables (i.e., public vs. private, region, and population density) by year cohort.

Table 3.3 Distribution of 12th grader unweighted characteristics from 38 years of the Monitoring the Future study by year cohort (n=599,032), United States 1976-2013.

	1976-1986 (n=189,422)		1987-1992 (n=98,190)		1993-2000 (n=122,476)		2001-2013 (n=188,944)	
	n	%[a]	n	%[a]	n	%[a]	n	%[a]
Sex								
Male	89,193	49.1	47,121	49.9	55,028	47.5	85,636	48.4
Female	92,313	50.9	47,393	50.1	60,789	52.5	91,146	51.6
Age (at interview)[b]	17.5	0.612	17.5	0.642	17.6	0.656	17.6	0.621
Race/ethnicity								
Non-Hispanic White	143,290	78.2	68,670	72.3	77,853	67.1	111,423	62.1
Non-Hispanic black	23,160	12.6	12,035	12.7	16,217	14.0	21,623	12.0
Hispanic	7,073	3.9	7,069	7.4	11,479	9.9	28,852	16.1
Other	9,814	5.4	7,205	7.6	10,449	9.0	17,517	9.8
Perceive great risk[b]								
For trying cannabis once or twice	0.117	0.092	0.215	0.105	0.162	0.082	0.152	0.077
For smoking cannabis regularly	0.546	0.191	0.775	0.095	0.625	0.114	0.536	0.121
Past alcohol use	50,320	86.4	31,700	85.0	31,757	76.8	35,327	69.4
Past cigarette use	3,368[c]	64.0	23,341	61.0	28,986	58.5	31,952	41.1
School information								
Public	164,998	87.1	87,442	89.0	108,384	88.5	164,319	87.0
Private	24,423	12.9	10,748	11.0	14,092	11.5	24,625	13.0
Region								
Northeast	45,634	24.1	21,317	21.7	27,103	22.1	39,754	21.0
Midwest	53,624	28.3	26,928	27.4	29,162	23.8	46,310	24.5
South	56,146	29.6	30,170	30.73	42,477	34.7	62,043	32.8
West	34,018	18.0	19,775	20.1	23,734	19.4	40,837	21.6
Population Density								
Urban	57,172	30.2	31,739	32.3	40,849	33.4	66,392	35.1
Suburban	85,936	45.4	47,102	48.0	55,804	45.6	86,084	45.6
Rural	46,314	24.5	19,349	19.7	25,823	21.1	36,468	19.3

[a] Due to rounding, percentages may not add to 100%.
[b] Mean with standard deviation.
[c] Between 1976-1985 no question on past cigarette use existed (58).

3.2 ALR Yearly PWOR Estimates

Figure 3.3 provides a fine-grained look at year to year variation in the PWOR as well as incidence of cannabis smoking among 12th graders each year. Shown are estimated PWOR and 95% CI linking cannabis use clustering within schools in 12th graders and estimated incidence for 12th grade cannabis use each year. For this unadjusted model, there is evidence of newly incident cannabis smoking clustering within schools between 1976-2013 for students who began use in 12th grade and had not previously used before in approximately a quarter of the years (e.g., 1986-7 and 1990-1992). Remember that the PWOR does not depend on the cannabis incidence rate.

Figure 3.3 Unadjusted estimated pairwise odds ratios (PWOR) and 95% confidence intervals for newly incident cannabis smoking clustering within schools in 12th graders. Data from the Monitoring the Future (MTF) study, United States 1976-2013.

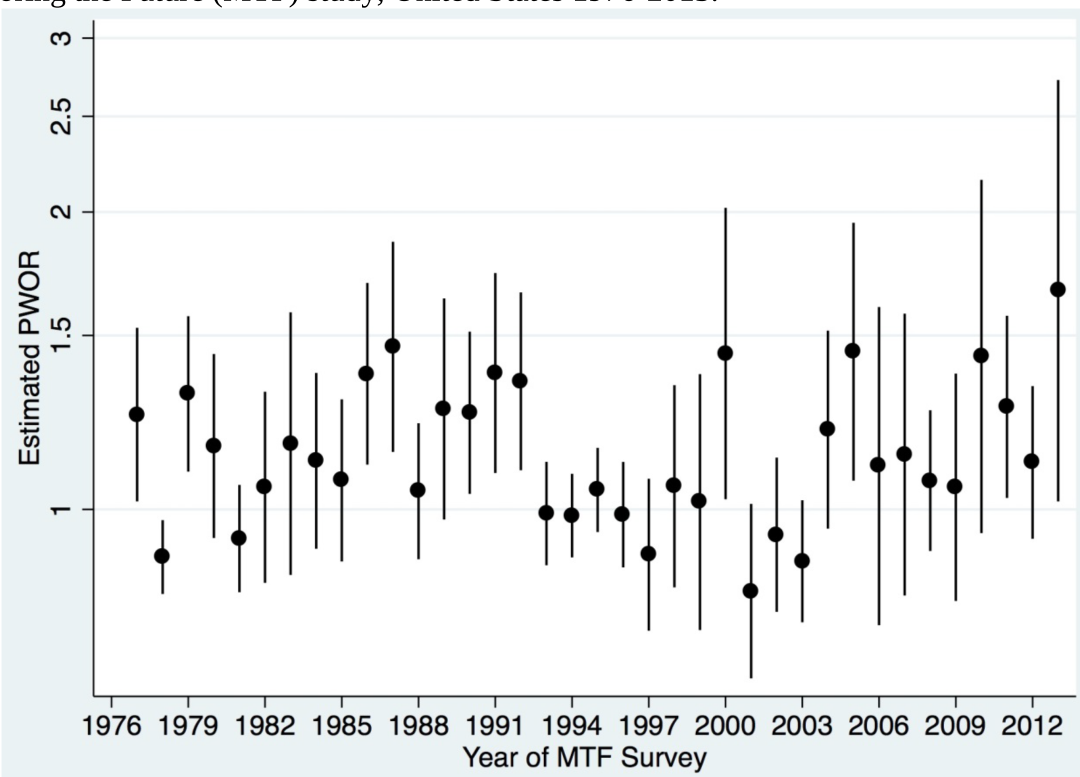
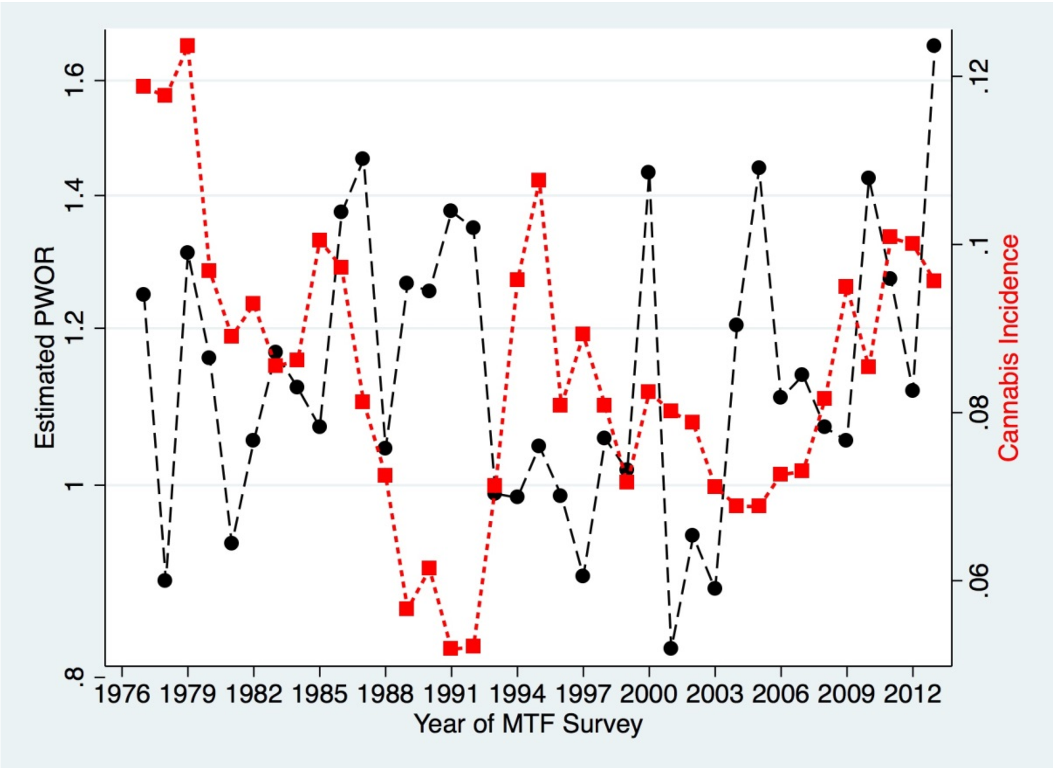


Figure 3.4 shows unadjusted estimated PWOR for newly incident cannabis smoking clustering within schools as well cannabis smoking incidence for 12th graders. For ease of understandability, 95% are absent from this figure. When cannabis incidence hits its lowest values, the PWOR point estimates are above unity.

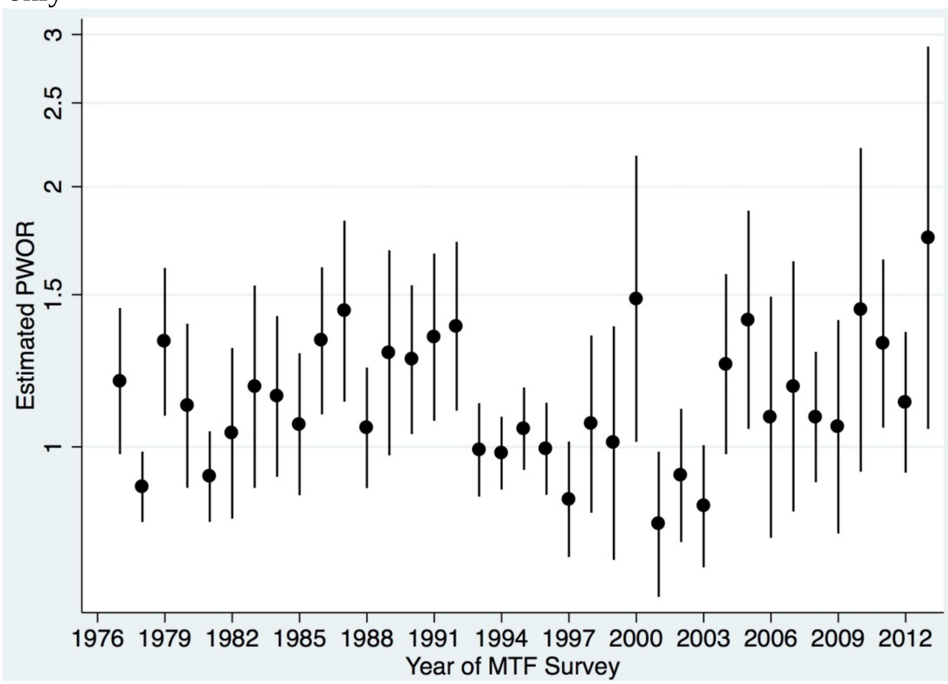
Figure 3.4 Estimated pairwise odds ratios (PWOR) for newly incident cannabis smoking clustering within schools among 12th graders (left). Estimated incidence rates for 12th grade cannabis smoking (right). Data from the Monitoring the Future (MTF) study, United States 1976-2013.



Adjusting by sex and age, PWOR estimates do not appreciably change. With few exceptions, neither covariate serves as a strong predictor year by year (p-value>0.05). As in the unadjusted model, many year specific PWOR and 95% CI are null. Figures 3.5 and 3.6 display within-school PWORs year by year with their 95% CI when sex and age were included in the model, separately (Figure 3.5) and then together (Figure 3.6).

Figure 3.5 Estimated pairwise odds ratios (PWOR) and 95% confidence intervals for newly incident cannabis smoking clustering within schools among 12th graders, with (a) sex and (b) age in the model. Data from the Monitoring the Future (MTF) study, United States 1976-2013.

(a) Sex only



(b) Age only

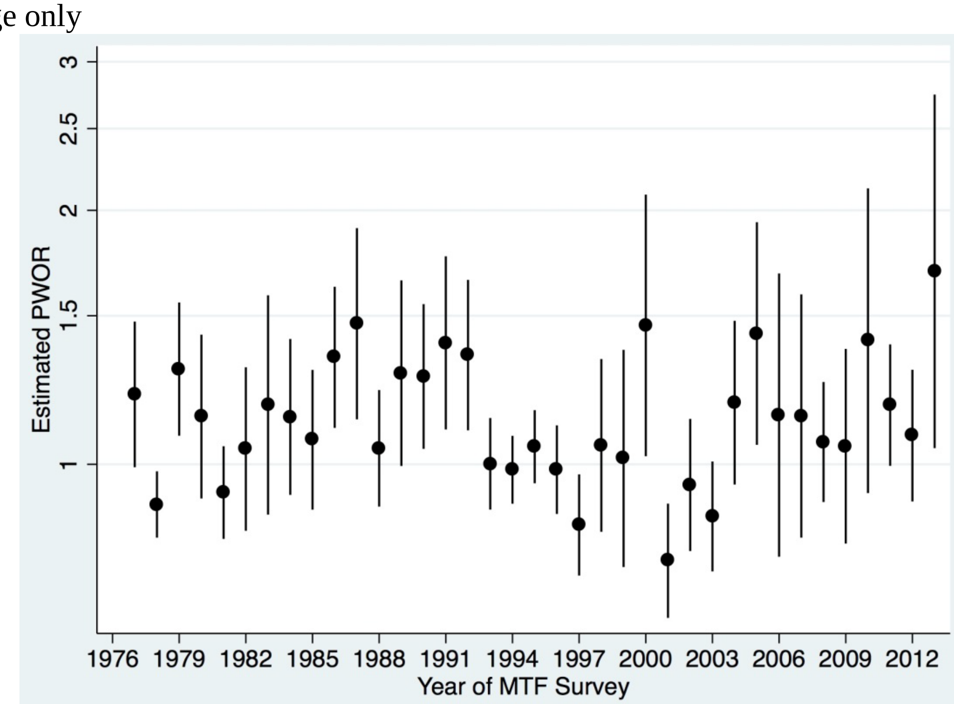
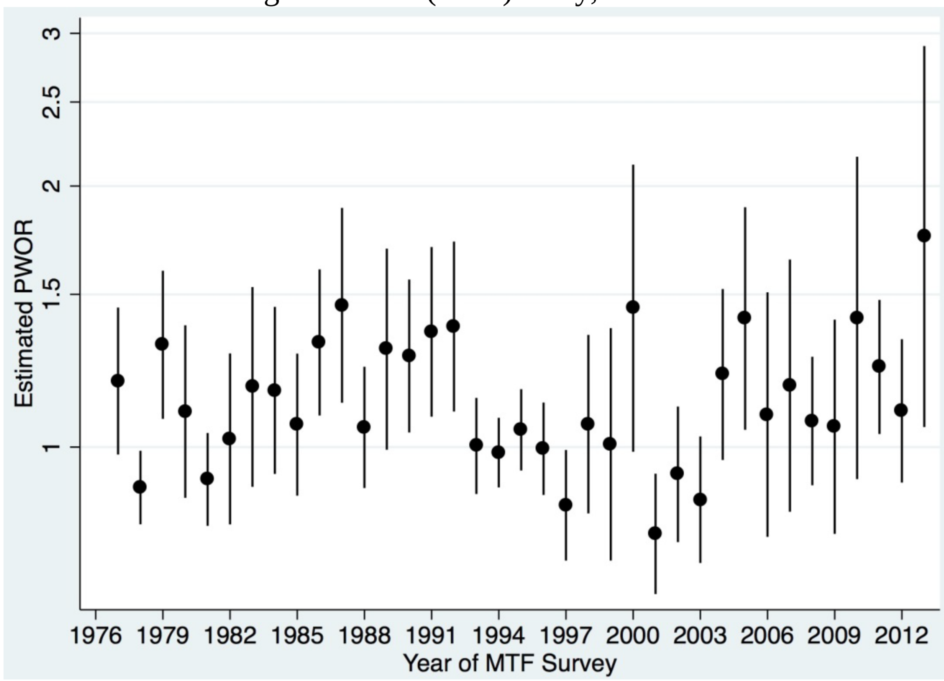


Figure 3.6 Estimated pairwise odds ratios (PWOR) and 95% confidence intervals for newly incident cannabis smoking clustering within schools in 12th graders, with age and sex in the model. Data from the Monitoring the Future (MTF) study, United States 1976-2013.



Results presented in Figure 3.7 depict estimated PWOR and 95% CI for newly incident cannabis use clustering within schools including risk estimates in the model. Even after including the prior year's risk perception for schools, the year specific PWOR and 95% CI are estimated around 1. Here, 12th grader cannabis risk perception at year 1 (t) predicting onsets of use at year 2 (t+1). Figure 3.7 shows a model that controls for perceived risk of trying cannabis once or twice; PWOR estimates with perceived risk of regular cannabis smoking were not noticeably different (see Figure 3.8). About a quarter of the PWOR estimates with perceived risk in the model were above unity (with either risk of trying once or twice or smoking cannabis regularly included as predictors).

Figures 3.7 and 3.8 provide a more detailed look at year to year variation in the PWOR among 12th graders each year with risk perception at time t is controlled as a covariate, which might help account for the observed clustering of newly incident users.

Figure 3.7 Estimated pairwise odds ratios (PWOR) and 95% confidence intervals for newly incident cannabis smoking clustering within schools among 12th graders with perceived risk of using cannabis once or twice in the model. Data from the Monitoring the Future study, United States 1976-2013.

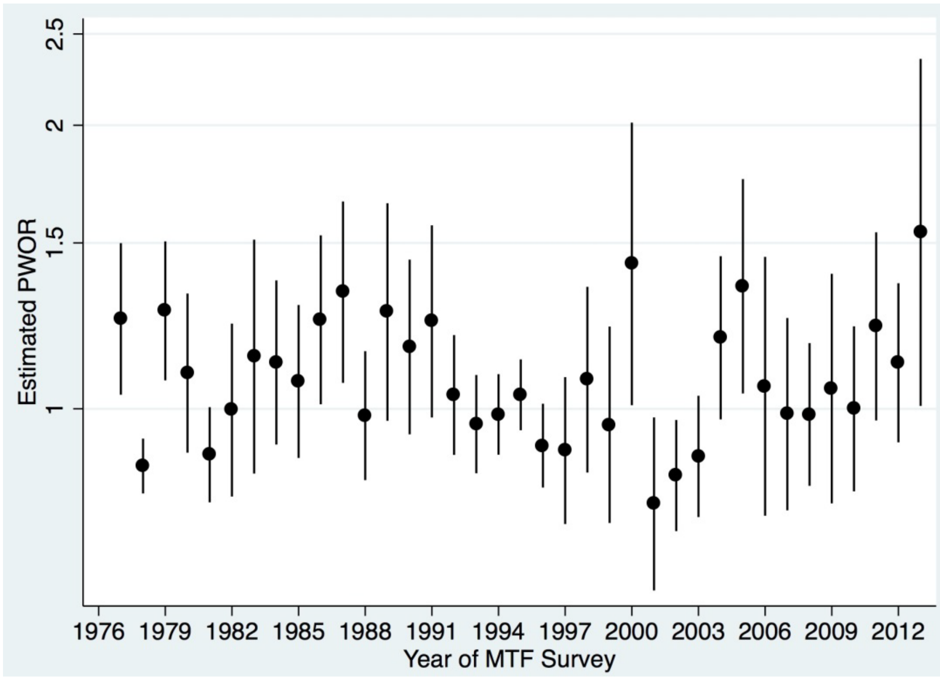
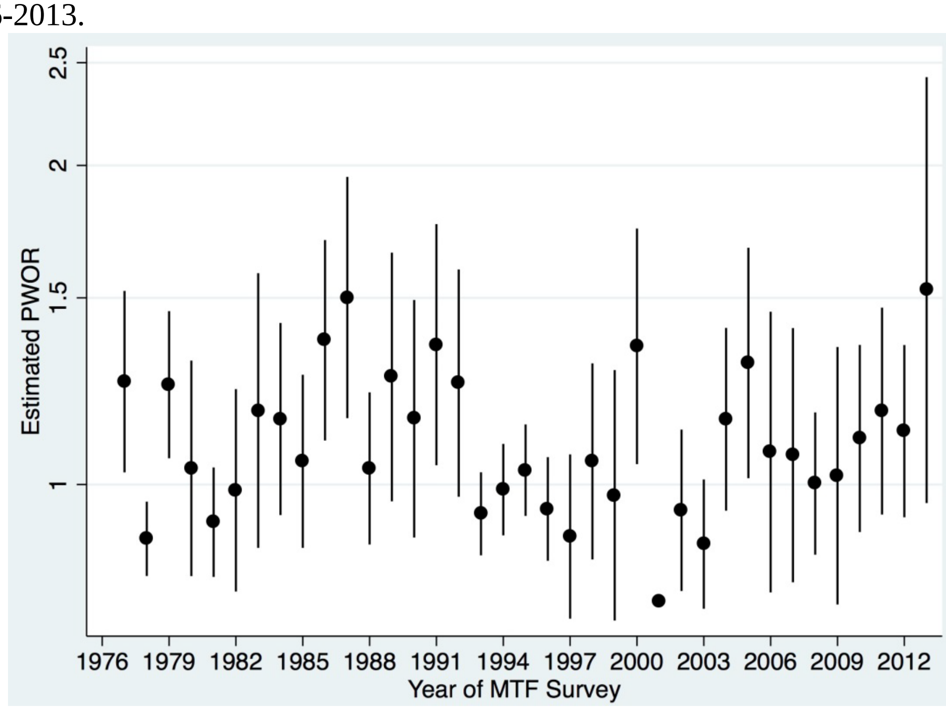


Figure 3.8 Estimated pairwise odds ratios (PWOR) and 95% confidence intervals for newly incident cannabis smoking clustering within schools among 12th graders with perceived risk of smoking cannabis regularly in the model. Data from the Monitoring the Future study, United States 1976-2013.



3.3 Bivariate Estimates

The next step of the data analysis plan involved running multiple bivariate ALR models with perceived risk and covariates of interest in year cohorts and for all years. After adding the additional covariates, the clustering estimates for newly incident cannabis smoking did not appreciably change from the model with risk perception only. However, the parameter estimates for past cigarette use and past alcohol use showed strong predicting power in their respective ALR models (p-value<0.05). Table 3.4 shows parameter estimates for sex, age, race/ethnicity, past alcohol use, past cigarette use, and then school-level covariates (i.e., public vs. private, region, population density). PWOR for all bivariate ALR models were similar. Table 3.5 displays PWOR estimates for each bivariate model.

Table 3.4 Alternating Logistic Regressions bivariate model parameter estimates (log odds, standard error) for 12th graders with perceived risk[a,b] and selected other covariates for all years and by year cohort[c]. Data from the Monitoring the Future study, United States 1976-2013.

	All years		1976-1986		1987-1992		1993-2000		2001-2013	
Model	1	2	1	2	1	2	1	2	1	2
Sex										
Male (reference)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Female	**-0.08 (0.03)**	**-0.08 (0.03)**	**-0.16 (0.07)**	**-0.17 (0.06)**	-0.04 (0.09)	-0.04 (0.08)	-0.07 (0.07)	-0.05 (0.07)	-0.04 (0.07)	-0.04 (0.07)
Age (at interview)	**-0.11 (0.03)**	**-0.13 (0.03)**	-0.08 (0.05)	-0.09 (0.05)	-0.05 (0.07)	**-0.09 (0.07)**	**-0.20 (0.06)**	**-0.21 (0.05)**	-0.09 (0.05)	**-0.11 (0.06)**
Race/ethnicity										
Non-Hispanic White (reference)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Non-Hispanic black	0.12 (0.06)	0.00 (0.06)	**0.26 (0.10)**	**0.25 (0.09)**	-0.26 (0.18)	**-0.45 (0.17)**	0.01 (0.12)	-0.15 (0.12)	0.19 (0.12)	0.10 (0.11)
Hispanic	0.05 (0.07)	-0.03 (0.07)	**0.43 (0.19)**	**0.42 (0.19)**	0.04 (0.19)	-0.03 (0.20)	-0.03 (0.15)	-0.10 (.15)	0.01 (0.10)	-0.05 (0.10)
Other	**-0.29 (0.07)**	**0.31 (0.07)**	-0.18 (0.16)	-0.13 (0.16)	**-0.45 (0.19)**	**-0.52 (0.20)**	**-0.43 (0.16)**	**-0.44 (0.16)**	-0.17 (0.12)	-0.21 (0.12)
Past alcohol use	**0.94 (0.05)**	**0.98 (0.04)**	**0.49 (0.08)**	**0.50 (0.08)**	**0.70 (0.13)**	**0.73 (0.13)**	**1.57 (0.11)**	**1.59 (0.11)**	**1.03 (0.08)**	**1.06 (0.07)**
Past cigarette use	**1.38 (0.04)**	**1.50 (0.05)**	**1.47 (0.20)**	**1.48 (0.10)**	**1.46 (0.10)**	**1.61 (0.09)**	**1.56 (0.08)**	**0.98 (0.05)**	**1.41 (0.08)**	**0.50 (0.08)**
School information										
Public (reference)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Private	-0.01 (0.07)	0.05 (0.07)	-0.01 (0.13)	-0.01 (0.13)	0.07 (0.15)	0.15 (0.15)	0.14 (0.14)	0.17 (0.12)	-0.20 (0.11)	-0.05 (0.12)
Region										
West (reference)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Northeast	-0.05 (0.07)	-0.01 (0.06)	-0.17 (0.12)	-0.20 (0.12)	0.03 (0.16)	0.08 (0.17)	0.18 (0.13)	0.20 (0.13)	-0.15 (0.11)	-0.10 (0.11)
Midwest	-0.09 (0.06)	-0.02 (0.06)	-0.20 (0.11)	-0.21 (0.11)	0.16 (0.15)	0.25 (0.16)	-0.01 (0.12)	0.05 (0.12)	-0.17 (0.11)	-0.06 (0.11)
South	0.01 (0.06)	-0.05 (0.06)	-0.20 (0.11)	**-0.22 (0.11)**	-0.14 (0.17)	-0.28 (0.17)	**0.25 (0.11)**	0.17 (0.10)	0.01 (0.10)	-0.02 (0.10)
Population Density										
Rural (reference)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Urban	**0.15 (0.06)**	**0.16 (0.06)**	**0.32 (0.12)**	**0.31 (0.12)**	0.19 (0.15)	0.20 (0.16)	0.03 (0.12)	0.00 (0.12)	0.20 (0.11)	**0.22 (0.11)**
Suburban	**0.16 (0.06)**	**0.18 (0.06)**	**0.29 (0.10)**	**0.27 (0.10)**	0.07 (0.15)	0.10 (0.15)	0.19 (0.11)	0.15 (0.11)	0.18 (0.11)	**0.22 (0.10)**

[a] Model 1 is a model with perceived 'great' risk of trying cannabis once or twice.
[b] Model 2 is a model with perceived 'great' risk of smoking cannabis regularly.
[c] Estimates in bold are statistically significant at the p-value<0.05 level.

Table 3.5 Alternating Logistic Regressions bivariate model pairwise alpha estimates for 12th graders with perceived risk[a,b] and selected other covariates for all years and by year cohort[c]. Data from the Monitoring the Future study, United States 1976-2013.

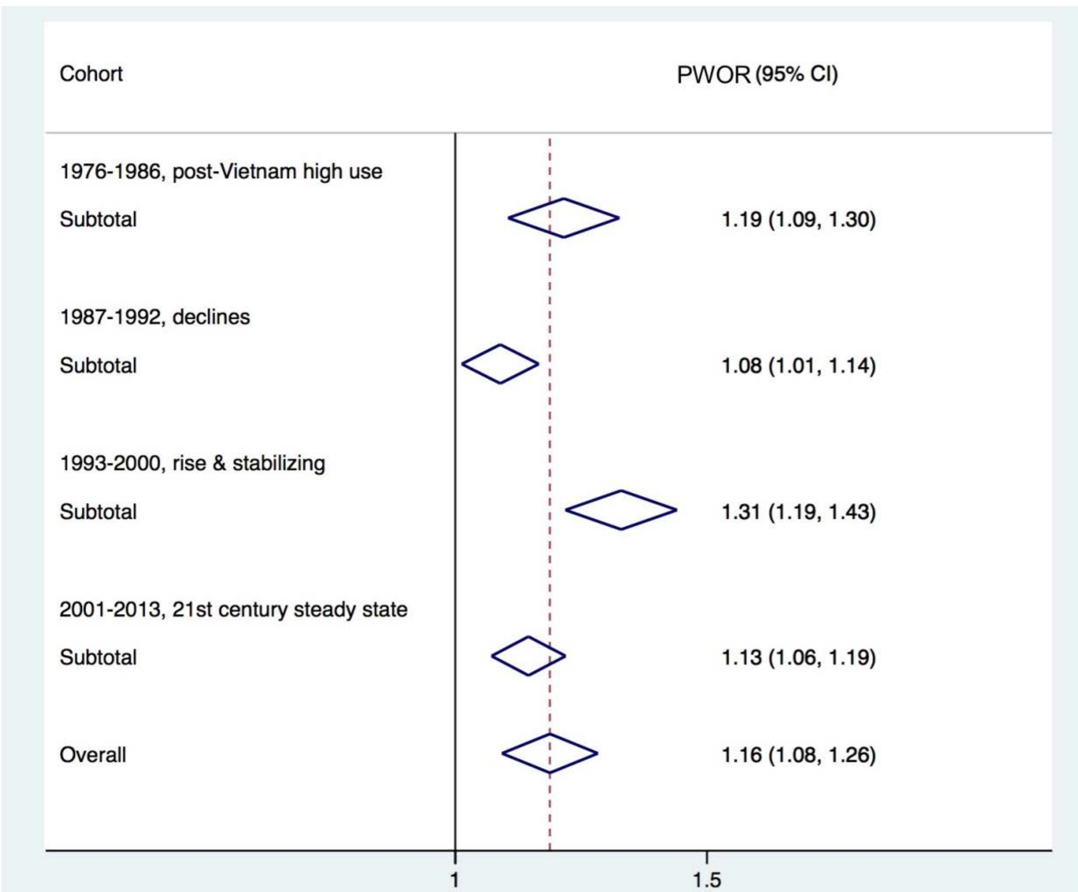
	All years		1976-1986		1987-1992		1993-2000		2001-2013	
Model	1	2	1	2	1	2	1	2	1	2
Sex	**0.13 (0.02)**	**0.13 (0.02)**	**0.10 (0.03)**	**0.10 (0.03)**	**0.18 (0.05)**	**0.23 (0.05)**	0.06 (0.03)	0.06 (0.03)	**0.12 (0.04)**	**0.13 (0.04)**
Age (at interview)	**0.12 (0.02)**	**0.13 (0.02)**	**0.10 (0.03)**	**0.10 (0.03)**	**0.18 (0.05)**	**0.22 (0.05)**	0.06 (0.03)	**0.06 (0.03)**	**0.10 (0.04)**	**0.11 (0.04)**
Race/ethnicity	**0.12 (0.02)**	**0.13 (0.02)**	**0.10 (0.03)**	**0.10 (0.03)**	**0.17 (0.04)**	**0.19 (0.04)**	0.05 (0.03)	0.05 (0.03)	**0.11 (0.04)**	**0.13 (0.04)**
Past alcohol use	**0.14 (0.02)**	**0.12 (0.02)**	**0.11 (0.03)**	**0.11 (0.03)**	**0.17 (0.04)**	**0.19 (0.04)**	**0.07 (0.03)**	0.06 (0.03)	**0.14 (0.05)**	**0.13 (0.04)**
Past cigarette use	**0.17 (0.03)**	**0.14 (0.02)**	0.20 (0.11)	**0.27 (0.11)**	**0.18 (0.05)**	**0.20 (0.05)**	**0.08 (0.04)**	0.07 (0.04)	**0.13 (0.05)**	**0.13 (0.05)**
School information										
Public/Private	**0.12 (0.02)**	**0.13 (0.02)**	**0.11 (0.13)**	**0.11 (0.03)**	**0.18 (0.04)**	**0.22 (0.05)**	0.05 (0.03)	0.05 (0.03)	**0.11 (0.04)**	**0.12 (0.04)**
Region	**0.12 (0.02)**	**0.13 (0.02)**	**0.11 (0.03)**	**0.11 (0.03)**	**0.16 (0.04)**	**0.20 (0.05)**	0.05 (0.03)	0.05 (0.03)	**0.11 (0.04)**	**0.12 (0.04)**
Population Density	**0.12 (0.02)**	**0.13 (0.02)**	**0.10 (0.03)**	**0.10 (0.03)**	**0.18 (0.04)**	**0.22 (0.05)**	0.05 (0.03)	0.05 (0.03)	**0.11 (0.04)**	**0.12 (0.04)**

[a] Model 1 is a model with perceived ‘great’ risk of trying cannabis once or twice.
[b] Model 2 is a model with perceived ‘great’ risk of smoking cannabis regularly.
[c] Estimates in bold are statistically significant at the p-value<0.05 level.

3.4 Meta-Analytic Estimates

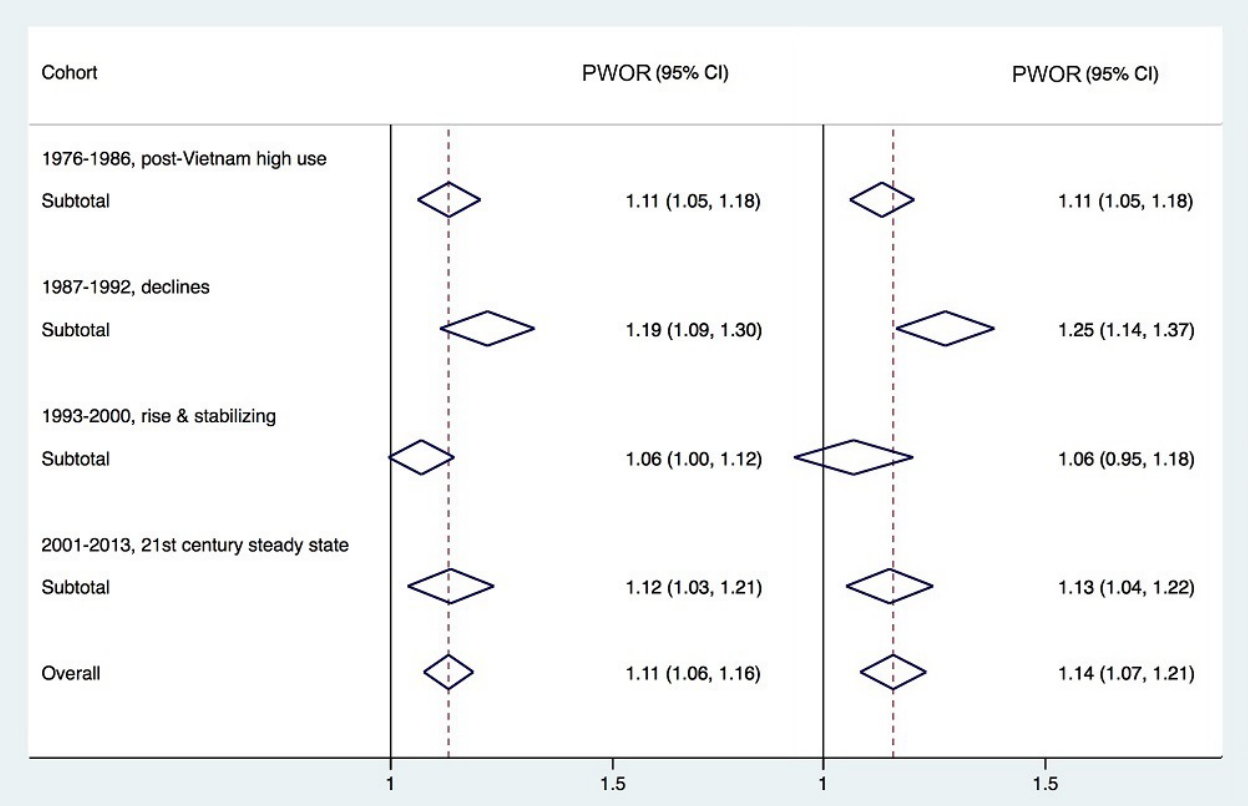
The main meta-analysis derived estimates of the study are presented in Figure 3.9. Pictured are meta-analysis PWOR estimates and 95% CI by year cohort for 12th grade newly incident cannabis use and clustering within schools for an intercept only model. The dashed line in the figure shows the meta-analytic summary estimate. For each year cohort that reflected stability and change in occurrence of cannabis use (described in Table 3.1), the four meta-analytic summary PWOR estimates are greater than unity. The meta-analysis summary PWOR is a 'random-effects' estimator (1.16; 95% CI = 1.08, 1.26; I^2 = 77.4%; p-value=0.004).

Figure 3.9 Meta-analysis derived unadjusted pairwise odds ratios (PWOR) and 95% confidence intervals (CI) for school-level clustering of newly incident cannabis smoking among 12th graders, binned by year cohorts. Data from the Monitoring the Future study, United States 1976-2013.



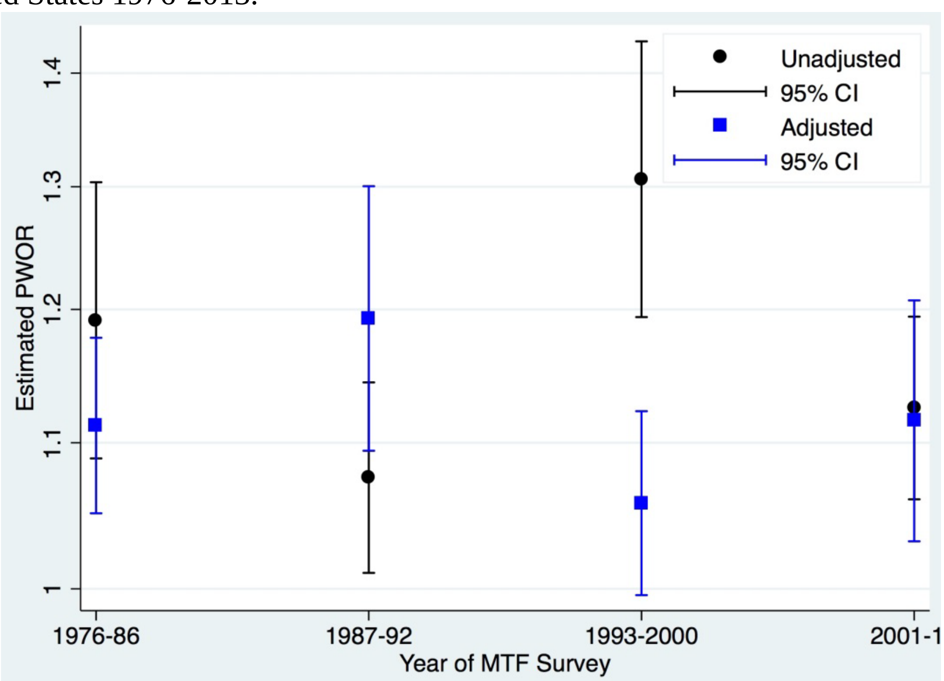
Similarly, Figure 3.10 shows meta-analysis estimates of PWOR and 95% CI for newly incident cannabis smoking clustering within schools in 12th grade when risk perceptions are included in the model. The overall meta-analytic summary PWOR estimates are greater than 1 consistent with school-level clustering of newly incident cannabis use with risk perception in the prior year of any given school predicting use in the second year. The meta-analysis summary PWOR for perceived risk of trying cannabis once or twice (left side of Figure 3.10) is a 'fixed-effects' estimator (1.11; 95% CI = 1.06, 1.16; I^2 = 42.2%; p-value=0.159). The meta-analysis summary PWOR for perceived risk of smoking cannabis regularly (right side of Figure 3.10) is a 'random-effects' estimator (1.14; 95% CI = 1.07, 1.21; I^2 = 53.8%; p-value=0.090). The corresponding 'fixed-effects' 95% CI = 1.09, 1.18.

Figure 3.10 Meta-analysis derived pairwise odds ratios (PWOR) and 95% confidence intervals (CI) for school-level clustering of newly incident cannabis use among 12th graders, binned by year cohorts and including perceived risk in the model: trying cannabis once or twice (left) and smoking cannabis regularly (right). Data from the Monitoring the Future study, United States 1976-2013.



In Figure 3.10, the meta-analysis PWOR estimate (dashed line) that borrows information from all stratified estimates is 1.11-14 (95% CI = 1.06, 1.21. For each year cohort described in Table 3.1, PWOR estimates are similar for both perceived risk variables. When the incidence was the highest between 1976-86, 10.1%, the PWOR = 1.11 (95% CI = 1.05-1.18. At incidence's lowest values from 1987-92, 6.1%, the PWOR was the strongest = 1.19 and 1.25 (95% CI = 1.09-1.37). From 1993-2000 when the incidence was 8.6%, the PWOR is the weakest and null, 1.06 (95% CI = 0.95-1.18). Last, at an incidence of 8.2% the PWOR = 1.12-3 (95% CI = 1.03-1.22). There is a slight dampening effect of the PWOR magnitude when perceived risk is included in the model for three of the four year cohorts. Figure 3.11 shows a comparison between the unadjusted model (intercept only) and an adjusted model (with perceived risk of trying cannabis once or twice).

Figure 3.11 Meta-analysis derived pairwise odds ratios (PWOR) and 95% confidence intervals (CI) for school-level clustering of newly incident cannabis use among 12th graders, binned by year cohorts comparing unadjusted and adjusted models[a]. Data from the Monitoring the Future study, Untied States 1976-2013.



[a]Unadjusted model has an intercept only and the adjusted model controls for risk perception of trying once or twice.

Depicted in Figure 3.12 are meta-analysis estimates of exponentiated parameter estimates and 95% CIs for risk perceptions of cannabis smoking for 12th grade in the ALR model. For each year cohort described in the table, all meta-analytic summary OR estimates are below unity. The overall meta-analytic summary OR estimates are less than 1, so 12th graders are less likely to be newly incident users for a ‘one unit increase’ in risk perception of the prior year of any given school predicting reduced use in the second year. Here a unit increase is going from 0% of students thinking regular cannabis smoking has great risk to 100%. This is evidence of risk perception as a predictor for regular cannabis smoking. The estimates for trying once or twice should be interpreted with caution as this perceived risk variable did not have any true zeroes (i.e., no schools had 0% of students who thought trying cannabis once or twice posed great risk). Table 3.6 shows corresponding ALR parameter estimates for perceived risk with 95% CI on the log odds scale.

Figure 3.12 Meta-analysis derived odds ratios (OR) and 95% confidence intervals (CI) for risk perception of (left) trying cannabis once or twice and smoking cannabis regularly (right) in the Alternating Logistic Regressions, binned by year cohorts. Data from the Monitoring the Future study, United States 1976-2013.

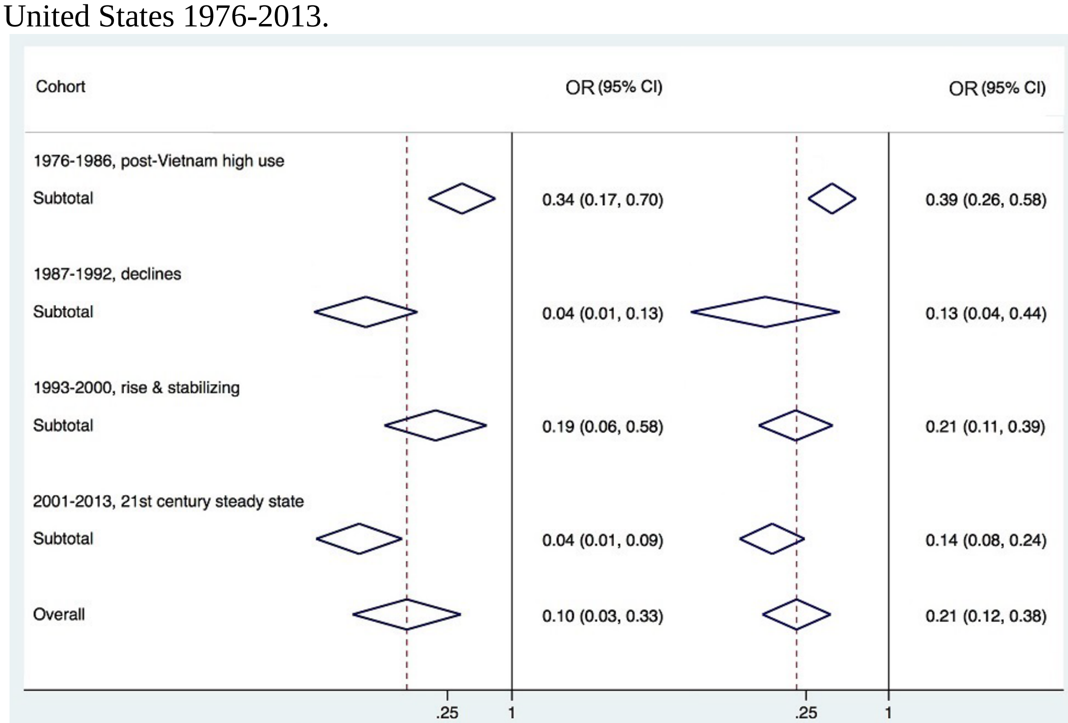


Table 3.6 Alternating Logistic Regressions parameter estimates and 95% confidence intervals (CI) for perceived risk of cannabis use for all years and binned by year cohorts. Data from the Monitoring the Future study, United States 1976-2013.

	Trying cannabis once or twice Log odds (95% CI)	Smoking cannabis regularly Log odds (95% CI)
All years	-2.38 (-2.85, -1.91)	-1.49 (-1.73, -1.25)
Year cohorts		
1976-1986	-1.07 (-1.79, -0.36)	-0.95 (-1.35, -0.55)
1987-1992	-3.14 (-4.25, -2.03)	-2.07 (-3.32, -0.82)
1993-2000	-1.64 (-2.74, -0.54)	-1.56 (-2.18, -0.94)
2001-2013	-3.28 (-4.20, -2.36)	-1.95 (-2.50, -1.41)

3.5 Post-Estimation Exploratory Data Analysis Steps

In dividing the perceived risk variables into quantiles and using the 4th quartile as the reference group, there is a gradient in PWOR. Generally, as perceived risk decreases the evidence of newly incident cannabis use increases (see Table 3.7).

Purposeful selection exploratory analyses to probe subgroup variation in the estimates disclosed neither sex differences in incident use nor any consistent differences by race/ethnicity over time. In bivariate analysis, the covariates for age in years, past alcohol use, past cigarette use, and population density satisfied the initial retaining criterion for backward elimination (p-value<0.20). A multivariable model with all these covariates revealed that age could be removed due to p-values>0.05, but the QIC/QICu were lower for a model with age versus not. Adding covariates not introduced to the initial multivariable model revealed that a model that additionally included sex performed the best (lowest QIC/QICu) for all models. Introducing squared terms for the continuous variables (i.e., age and perceived risk), none showed the quadratic term was required for any of the ALR models. No evidence of collinearity between past cigarette and past alcohol use was found. The final ALR model included terms for perceived risk, sex, age, past tobacco cigarette and alcohol use, and population density (Table 3.8).

Table 3.7 Estimated odds ratios (OR) and 95% confidence intervals (CI) by school as estimated by Alternating Logistic Regressions for all years and year cohorts[a]. Data from the Monitoring the Future study, United States 1976-2013.

	All years Overall p-value OR (95% CI)	1976-1986 Overall p-value OR (95% CI)	1987-1992 Overall p-value OR (95% CI)	1993-2000 Overall p-value OR (95% CI)	2001-2013 Overall p-value OR (95% CI)
Perceived risk of cannabis use (%)					
Trying cannabis once or twice	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
1st quartile	**1.73 (1.53, 1.96)**	**1.30 (1.07, 1.59)**	**2.21 (1.55, 3.15)**	1.38 (1.00, 1.90)	**1.98 (1.58, 2.48)**
2nd quartile	**1.64 (1.46, 1.84)**	**1.41 (1.13, 1.77)**	**1.74 (1.32, 2.30)**	**1.49 (1.22, 1.84)**	**1.66 (1.34, 2.05)**
3rd quartile	**1.36 (1.21, 1.54)**	1.13 (0.85, 1.50)	**1.54 (1.18, 2.00)**	1.20 (0.98, 1.45)	**1.51 (1.20, 1.88)**
4th quartile (reference)	1.00	1.00	1.00	1.00	1.00
Smoking cannabis regularly	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
1st quartile	**1.75 (1.57, 1.95)**	**1.48 (1.21, 1.80)**	1.27 (0.40, 4.09)	**1.51 (1.16, 1.98)**	**2.22 (1.69, 2.92)**
2nd quartile	**1.55 (1.37, 1.76)**	1.28 (0.99, 1.66)	1.29 (0.59, 2.79)	**1.41 (1.13, 1.75)**	**1.96 (1.49, 2.59)**
3rd quartile	**1.34 (1.20, 1.51)**	1.21 (0.95, 1.54)	**1.79 (1.38, 2.31)**	1.12 (0.93, 1.36)	**1.55 (1.15, 2.07)**
4th quartile (reference)	1.00	1.00	1.00	1.00	1.00

[a] Estimates in bold are statistically significant at the p-value<0.05 level.

Table 3.8 Analysis of Alternating Logistic Regression parameter estimates and 95% confidence intervals (CI) for all years. Data from the Monitoring the Future study, United States 1976-2013.

	Trying cannabis once or twice				Smoking cannabis regularly			
Parameter	Estimate	95% CI		p-value	Estimate	95% CI		p-value
Sex								
Male (reference)	0.00				0.00			
Female	-0.11	-0.20	-0.11	0.024	-0.11	-0.20	0.11	0.024
Age (at interview)	-0.12	-0.20	-0.12	0.002	-0.10	-0.17	-0.01	0.017
Past cigarette use	1.22	1.12	1.22	<0.001	1.13	1.02	-0.02	<0.001
Past alcohol use	0.78	0.65	0.78	<0.001	0.72	0.59	1.24	<0.001
Population Density								
Rural (reference)	0.00				0.00			
Urban	0.26	0.10	0.26	<0.001	0.31	0.15	0.47	0.001
Suburban	0.28	0.13	0.28	<0.001	0.30	0.15	0.45	<0.001
Perceived risk	-2.60	-2.97	-2.60	<0.001	-2.86	-3.47	-2.25	<0.001
Alpha	0.14	0.09	0.14	<0.001	0.20	0.15	0.26	<0.001

CHAPTER 4

DISCUSSION

This thesis presents the first quantitative estimates of school level clustering of cannabis smoking in the US. The main findings may be summarized succinctly. Overall, modest but tangible within-school clustering of cannabis smoking is seen, consistent with models for social sharing of cannabis experience among students.

School-level cannabis risk perceptions expressed by 12th graders of a school, as observed in one school year, help predict occurrence of newly incident cannabis use among 12th graders assessed in the next school year. The PWOR at time t+1 shifted downward with inclusion of several explanatory covariates and perceived risk at time t, which suggests that underlying mechanisms for social sharing or 'contagion' processes of newly incident cannabis use might be governed by ambient levels of perceived risk. The regression slope estimates indicate that prevailing levels of positive or negative risk perceptions the prior year may influence whether there is rising, falling, or stable risk of becoming a cannabis smoker the next year in the new class of 12th graders, in addition to influence on clustering of new use.

As for the size or magnitude of the observed clustering of newly incident cannabis use within schools, the PWOR estimates can be characterized as 'modest' relative to prior research on clustering of cannabis use in US neighborhoods. Nevertheless, even after covariate adjustment, there is evidence of cannabis clustering on par with the lower end of reported PWOR for clustering of childhood diarrheal disease in villages of the developing world, for which social sharing of infections and contagion processes clearly are at play – even when the PWOR is modest (26). Estimates of the quantitative clustering of cannabis smoking are based the largest US school survey.

Before detailed discussion of these results, several of the more important study limitations merit attention. Of central concern is MTF data are self-reported and the validity is affected by respondents' truthfulness, memory, and completeness. Social acceptability or fear of disclosure may affect responses about cannabis use (9). MTF data sometimes are incomplete or are not filled out correctly (i.e., scantron errors). However, this large sample school based study has been designed to be generalized to the US population of 12th graders, and the MTF research team has held methods of the surveys constant over the years. While it is cross-sectional, trends over time can be seen, but not in the same students as in a prospective/longitudinal study of individual students followed-up over time.. A small prospective German study recently showed how changes in risk perception predict changes in cannabis use (72). To date, there has not been a longitudinal study done examining perceived risk and newly incident drug use.

In addition, data are collected only from students present on the date of the survey and does not capture school dropouts. Frequent cannabis users may not regularly attend school or drop out. For 12th graders, approximately 9–15% drop out of high school before graduation (73). With respect to assessment of the key response variable, the incident use is first use and does not capture users who began use over the summer, which may be an important subpopulation of 12th graders.

The inclusion of more precisely worded survey questions would be useful. For example, the survey asks about harmfulness of cannabis use based on frequency of cannabis use (i.e., trying once or twice or smoking regularly). Responses are subjective and do not exhibit how the harm is viewed by the student (e.g., physical harm, legal harm, emotional harm, etc.) Assessment of variables herein may have suffered from the open interpretation.

With respect to the data analysis plan, the GEE-based model has a population averaged interpretation; inferences about a typical student cannot be made. ALR estimates with robust standard errors tend to be closer to the null than other statistical methods. For example, generalized linear mixed models can be used for multilevel binary data. This subject-specific approach is used for the intraclass correlation (ICC) that arises due to similarity of individuals within a cluster. ICC also measures the degree of clustering; it is “the extent to which members of a group resemble each other more than they resemble members of other groups” (65). However, it depends on the marginal distribution (e.g., occurrence of cannabis smoking in this research). ICC is a concept from linear regression that has no exact equivalent for logistic regression.

Notwithstanding limitations such as these, the study findings are of interest because to date this is the first nationally representative US study on cannabis clustering in secondary schools. The results from this study may have important implications in seeking to account for the clustering of cannabis use in high schools.

Year by year, as the incidence hits the lowest values, evident within-school clustering is seen over the 38-year period for 12th grade users who first used in 12th grade. School level clustering becomes evident at these troughs. Newly incident cannabis users are not occurring in isolation. This is consistent with clustering of infectious diseases and certain foodborne illnesses. Small clusters of cases are often found in geographical areas and not single cases.

Although modest, meta-analytic estimates for each year cohort show relatively stable clustering at the same magnitude of childhood diarrheal disease in Zambia households and underage drinking in communities (26,27). The magnitude of all school clustering reported herein is slightly smaller than Bobashev and Anthony (1998) observed in a study of concentration of marijuana use in US neighborhoods and Wells et al. (2009) found in areas of New Zealand. The similar effect size of a known communicable disease's geographic clustering and newly incident cannabis smoking lends credence to cannabis's contagion effect in US high schools (19,25).

In schools, the contagion model of drug use proposes that students who perceive risk of trying cannabis or smoking regularly may cluster due to exposure to experiences (including use and perceptions) by their fellow students (30). It was notable that perceived risk is a predictor of newly incident cannabis use among 12th graders for all year cohorts. The PWOR at year 2 (t+1) was not affected by inclusion of other covariates although risk perception from year 1 (t) is consistently a strong predictor of newly incident cannabis use.

Based on the literature, there was not a suspected difference of newly incident use between the majority of demographic subpopulations of 12th graders. Still, there was a marked age association among students in this sample. The negative sign on the estimated parameter for age suggests that there is less within school clustering of newly incident cannabis use for younger aged 12th graders. This may correspond with findings that state the peak age of first use is 18 (74).

Consistent with the gateway description, this study provides evidence of associations that link tobacco cigarette and alcohol use with the odds of becoming a newly incident cannabis user. Individuals who use alcohol and/or tobacco are more likely to subsequently use cannabis (e.g., Yamaguchi & Kandel, [35]).

At the school level, it was interesting that population density appears to have had a rather consistent effect estimate in the ALR model. Although urban and suburban areas did not differ statistically, in post-hoc analysis the schools in rural areas showed less within school clustering over the period under study. While this finding was not constant for all year cohorts, it is noteworthy that inclusion of the urban-rural indicator improved model fit.

In future research that builds from findings such as these, it may be possible to seek more definitive evidence on characteristics that might account for school-level variation in degree of clustering such as school characteristics and more individual level characteristics (e.g., classrooms and clarification of the harmfulness measure). Clustering might be more clear if MTF made it possible to know which students were in the same classrooms as well as what students mean when they perceive ‘great risk’ of cannabis use. Qualitative research on what it means when students say using cannabis is risky also is needed if the intent is to guide prevention programs that seek experimental manipulation of cannabis risk perceptions in order to prevent or delay onset of cannabis use.

As noted previously in this thesis report, the population averaged ALR has several strengths. PWOR estimates have an intuitive interpretation for quantifying the magnitude of within cluster association (the well-known OR). The ALR also performs well with larger clusters and correlated binary outcomes. Use of the empirical sandwich estimator means that regardless of the correlation structure, regression coefficients should not be numerically different. Most importantly, the PWOR does not depend on the prevalence/incidence of a disease or behavior (i.e., the marginal distribution of the outcome). Although ALR are not well known, applying this innovative statistical method to cannabis smoking clustering in schools was advantageous. This approach has been applied to other subpopulation/drug use combinations, and in this context allowed an opportunity for exploration of the ever so popular risk perception theory of drug use in secondary school students. In fact, this approach can be applied to the 10th graders from the MTF survey.

4.1 Conclusions

In conclusion, modest but noteworthy estimates of within-school clustering of newly incident cannabis use can be seen, and the regression slope estimates highlight the predictive importance of cannabis risk perceptions in 12th graders of one school year relative to occurrence of newly incident cannabis use in next cohort of the school's 12th graders. Other covariates help improved fit of the regression model, most notably, past tobacco cigarette and alcohol use. Besides deconstructing exactly what perceived risk means to 12th graders and creating a more appropriate metric for future research on perceived risk, experimental manipulations in order to shift perceived harmfulness of drug use will be needed. For example, randomized controlled trials designed to intervene at perceived risk might prove to be important in efforts to prevent or delay onset of cannabis incidence (i.e., beyond 12th grade). Socially shared attitudes and perceptions represent a potentially important causal influence on cannabis incidence and might open up new avenues for public health intervention or prevention program development.

APPENDIX

APPENDIX

SAS code for Alternating Logistic Regression models for all years and each cohort.

With one explanatory variable, risk perception of regular marijuana use

```
title 'All years';
proc genmod data=c.mtf12 descending;
where time EQ 2;
class school /param=ref;
model nican2= riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;

title '1976-1986';
proc genmod data=c.mtf12 descending;
where time EQ 2 and 1976<=Year<=1986;
class school /param=ref;
model nican2= riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;

title '1987-1992';
proc genmod data=c.mtf12may2015 descending;
where time EQ 2 and 1987<=Year<=1992;
class school /param=ref;
model nican2= riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;

title '1993-2000';
proc genmod data=c.mtf12 descending;
where time EQ 2 and 1993<=Year<=2000;
class school/param=ref;
model nican2= riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;

title '2001-2013';
proc genmod data=c.mtf12 descending;
where time EQ 2 and 2001<=Year<=2013;
class school/param=ref;
model nican2= riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;
```

With an additional explanatory variable for all years

```
title 'All years past cigarette';
proc genmod data=c.mtf12 descending;
where time EQ 2;
class school /param=ref;
model nican2= pastcig riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;

title 'All years sex';
proc genmod data=c.mtf12 descending;
where time EQ 2;
class school /param=ref;
model nican2= Gender riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;

title 'All years race';
proc genmod data=c.mtf12 descending;
where time EQ 2;
class school /param=ref;
model nican2= black hisp other riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;
```

With multiple explanatory variables for an example year cohort

```
title '1976-1986 past cigarette & population density';
proc genmod data=c.mtf12 descending;
where time EQ 2 and 1976<=Year<=1986;
class school pden/param=ref;
model nican2= pastcig pden riskmjg1/dist=bin type3;
repeated subject = school/sorted logor=exch covb;
weight SWeight;
run;
```

REFERENCES

REFERENCES

1. Snow J, University of Glasgow Library. On continuous molecular changes, more particularly in their relation to epidemic diseases [electronic resource] : being the oration delivered at the 80th anniversary of the Medical Society of London [Internet]. London : John Churchill; 1853 [cited 2015 Aug 12]. 48 p. Available from: http://archive.org/details/b21464819

2. Anthony, J.C., Van Etten, M.L. Epidemiology and its Rubrics. In: Comprehensive Clinical Psychology. Oxford, UK: Elsevier Science Publication; 1998. p. 355–90.

3. World Health Organization. WHO | Cannabis [Internet]. WHO. 2015 [cited 2015 Aug 13]. Available from: http://www.who.int/substance_abuse/facts/cannabis/en/

4. United Nations Office on Drugs and Crime. World Drug Report 2015. United Nations publication, Sales No. E.15.XI.6; 2015.

5. United Nations Office on Drugs and Crime. World Drug Report 2012. United Nations publication, Sales No. E.12.XI.1; 2012.

6. Degenhardt L, Chiu W-T, Sampson N, Kessler RC, Anthony JC, Angermeyer M, et al. Toward a Global View of Alcohol, Tobacco, Cannabis, and Cocaine Use: Findings from the WHO World Mental Health Surveys. PLoS Med. 2008 Jul 1;5(7):e141.

7. United States. Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings [Internet]. Rockville, MD: United States Department of Health and Human Services, Substance Abuse and Mental Health Services Administration. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality.; 2014. Report No.: (SMA) 14-4863. Available from: http://www.samhsa.gov/data/sites/default/files/NSDUHresultsPDFWHTML2013/Web/NSDUHresults2013.htm

8. Johnston LD, O'Malley PM, Bachman JG, Schulenberg, J. E., Meich, R. A. Monitoring the Future national survey results on drug use, 1975-2013: Volume I, Secondary school students. Institue for Social Research, The University of Michigan; 2014.

9. Gfroerer JC, Wu LT, Penne MA. Initiation of Marijuana Use: Trends, Patterns, and Implications (Analytic Series: A-17, DHHS Publication No. SMA 02-3711). Substance Abuse and Mental Health Services Administration, Office of Applied Studies; 2002.

10. United States. Summary of findings from the 2000 National Household Survey on Drug Abuse. DHHS Publication No. SMA 01-3549, NHSDA Series H-13. Substance Abuse and Mental Health Services Administration; 2000.

11. Degenhardt L, Chiu WT, Sampson N, Kessler RC, Anthony JC. Epidemiological patterns of extra-medical drug use in the United States: Evidence from the National Comorbidity Survey Replication, 2001–2003. Drug Alcohol Depend. 2007 Oct 8;90(2–3):210–23.

12. Seedall RB, Anthony JC. Risk estimates for starting tobacco, alcohol, and other drug use in the United States: Male–female differences and the possibility that “limiting time with friends” is protective. Drug Alcohol Depend. 2013 Dec;133(2):751–3.

13. Compton, W.M., Grant, B.F., Colliver, J.D., Glantz, M.D., Stinson, F.S. Prevalence of marijuana use disorders in the united states: 1991-1992 and 2001-2002. JAMA. 2004 May 5;291(17):2114–21.

14. Wu L-T, Brady KT, Mannelli P, Killeen TK. Cannabis use disorders are comparatively prevalent among nonwhite racial/ethnic groups and adolescents: A national study. J Psychiatr Res. 2014 Mar;50:26–35.

15. United States. Department of Education. Office of Educational Research and Improvement. Public and Private Schools: How Do They Differ? National Center for Education Statistics; 1997.

16. CASAColumbia. National Survey of American Attitudes on Substance Abuse XVII: Teens. The National Center on Addiction and Substance Abuse at Columbia University; 2012.

17. Rogers EM. Diffusion of Innovations, 4th Edition. Simon and Schuster; 2010. 550 p.

18. De Alarcon R. The spread of heroin abuse in a community. UN Bull Narc. 1969;(3):17–22.

19. Dishion TJ, Dodge KA. Peer Contagion in Interventions for Children and Adolescents: Moving Towards an Understanding of the Ecology and Dynamics of Change. J Abnorm Child Psychol. 2005 Jun;33(3):395–400.

20. Susser M. The logic in ecological: I. The logic of analysis. Am J Public Health. 1994 May;84(5):825–9.

21. Halloran ME, Struchiner CJ. Study designs for dependent happenings. Epidemiol Camb Mass. 1991 Sep;2(5):331–8.

22. Koopman JS, Longini IM, Jacquez JA, Simon CP, Ostrow DG, Martin WR, et al. Assessing risk factors for transmission of infection. Am J Epidemiol. 1991 Jun 15;133(12):1199–209.

23. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. Am J Public Health. 1998 Feb;88(2):216–22.

24. Bobashev GV, Anthony JC. Clusters of Marijuana Use in the United States. Am J Epidemiol. 1998 Dec 15;148(12):1168–74.

25. Petronis KR, Anthony JC. Perceived risk of cocaine use and experience with cocaine: do they cluster within US neighborhoods and cities? Drug Alcohol Depend. 2000 Jan 1;57(3):183–92.

26. Katz J, Carey VJ, Zeger SL, Sommer A. Estimation of Design Effects and Diarrhea Clustering within Households and Villages. Am J Epidemiol. 1993 Dec 1;138(11):994–1006.

27. Reboussin BA, Preisser JS, Song E-Y, Wolfson M. Geographic Clustering of Underage Drinking and the Influence of Community Characteristics. Drug Alcohol Depend. 2010 Jan 1;106(1):38.

28. Delva J, Bobashev G, González G, Cedeño M, Anthony JC. Clusters of drug involvement in Panama: results from Panama's 1996 National Youth Survey. Drug Alcohol Depend. 2000 Nov 10;60(3):251–7.

29. Ennett ST, Bauman KE, Hussong A, Faris R, Foshee VA, Cai L, et al. The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. J Res Adolesc. 2006 Jun 1;16(2):159–86.

30. Ali MM, Amialchuk A, Dwyer DS. The Social Contagion Effect of Marijuana Use among Adolescents. Hartl D, editor. PLoS ONE. 2011 Jan 10;6(1):e16183.

31. Slovic P. Perception of Risk. Science. 1987 Apr 17;236(4799):280–5.

32. Bachman JG, Johnston LD, O'Malley PM. Explaining recent increases in students' marijuana use: impacts of perceived risks and disapproval, 1976 through 1996. Am J Public Health. 1998 Jun 1;88(6):887–92.

33. Bachman JG, Johnston LD, O'Malley PM, Humphrey RH. Explaining the Recent Decline in Marijuana Use: Differentiating the Effects of Perceived Risks, Disapproval, and General Lifestyle Factors. J Health Soc Behav. 1988 Mar 1;29(1):92–112.

34. Kilmer JR, Hunt SB, Lee CM, Neighbors C. Marijuana use, risk perception, and consequences: Is perceived risk congruent with reality? Addict Behav. 2007 Dec;32(12):3026–33.

35. Lopez-Quintero C, Neumark Y. Effects of risk perception of marijuana use on marijuana use and intentions to use among adolescents in Bogotá, Colombia. Drug Alcohol Depend. 2010 Jun 1;109(1–3):65–72.

36. Bachman JG, Johnston LD, O'Malley PM. Explaining the Recent Decline in Cocaine Use among Young Adults: Further Evidence That Perceived Risks and Disapproval Lead to Reduced Drug Use. J Health Soc Behav. 1990 Jun 1;31(2):173–84.

37. Kandel DB, Yamaguchi K, Chen K. Stages of progression in drug involvement from adolescence to adulthood: further evidence for the gateway theory. J Stud Alcohol. 1992 Sep;53(5):447–57.

38. Kandel D, Kandel E. The Gateway Hypothesis of substance abuse: developmental, biological and societal perspectives. Acta Paediatr. 2015 Feb 1;104(2):130–7.

39. Ellickson PL, Hays RD, Bell RM. Stepping through the drug use sequence: longitudinal scalogram analysis of initiation and regular use. J Abnorm Psychol. 1992 Aug;101(3):441–51.

40. Yamaguchi K, Kandel DB. Patterns of drug use from adolescence to young adulthood: II. Sequences of progression. Am J Public Health. 1984 Jul;74(7):668–72.

41. Yamaguchi K, Kandel DB. Patterns of drug use from adolescence to young adulthood: III. Predictors of progression. Am J Public Health. 1984 Jul;74(7):673–81.

42. Anthony JC. Death of the "stepping-stone" hypothesis and the "gateway" model? Comments on Morral et al. Addict Abingdon Engl. 2002 Dec;97(12):1505–7.

43. Lynskey M. An alternative model is feasible, but the gateway hypothesis has not been invalidated: comments on Morral et al. Addict Abingdon Engl. 2002 Dec;97(12):1505–7.

44. Morral AR, McCaffrey DF, Paddock SM. Reassessing the marijuana gateway effect. Addict Abingdon Engl. 2002 Dec;97(12):1493–504.

45. Morral AR, McCafrey DF, Paddock SM. Evidence does not favor marijuana gateway effects over a common-factor interpretation of drug use initiation: responses to Anthony, Kenkel & Mathios and Lynskey. Addict Abingdon Engl. 2002 Dec;97(12):1509–10.

46. Anthony JC, Warner LA, Kessler RC. Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. Exp Clin Psychopharmacol. 1994;2(3):244–68.

47. Ellickson PL, Bell RM. Drug prevention in junior high: a multi-site longitudinal test. Science. 1990 Mar 16;247(4948):1299+.

48. Lynam DR, Milich R, Zimmerman R, Novak SP, Logan TK, Martin C, et al. Project DARE: no effects at 10-year follow-up. J Consult Clin Psychol. 1999 Aug;67(4):590–3.

49. McAlister A, Perry C, Killen J, Slinkard LA, Maccoby N. Pilot Study of Smoking, Alcohol and Drug Abuse Prevention. Am J Public Health. 1980 Jul;70(7):719.

50. Nelson SE, Van Ryzin MJ, Dishion TJ. Alcohol, marijuana, and tobacco use trajectories from age 12 to 24 years: demographic correlates and young adult substance use problems. Dev Psychopathol. 2015 Feb;27(1):253–77.

51. Vermeulen-Smit E, Verdurmen JE, Engels RC. The Effectiveness of Family Interventions in Preventing Adolescent Illicit Drug Use: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Clin Child Fam Psychol Rev. 2015 Sep;18(3):218–39.

52. United States. Strategies and Interventions to Prevent Youth Marijuana Use: An At-a-Glance Resource Tool. Substance Abuse and Mental Health Services Administration. Center for the Application of Prevention Technologies. Reference #HHSS277200800004C; 2015.

53. Preisser JS, Arcury TA, Quandt SA. Detecting patterns of occupational illness clustering with alternating logistic regressions applied to longitudinal data. Am J Epidemiol. 2003 Sep 1;158(5):495–501.

54. Wells JE, Degenhardt L, Bohnert KM, Anthony JC, Scott KM. Geographical clustering of cannabis use: Results from the New Zealand Mental Health Survey 2003-2004. Drug Alcohol Depend. 2009 Jan 1;99(1-3):309–16.

55. Bobashev GV, Anthony JC. Use of Alternating Logistic Regression in Studies of Drug-Use Clustering. Subst Use Misuse. 2000 Jan 1;35(6-8):1051–73.

56. Petronis KR, Anthony JC. A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. J Epidemiol Community Health. 2003 Nov 1;57(11):893–900.

57. Carey V, Zeger SL, Diggle P. Modelling multivariate binary data with alternating logistic regressions. Biometrika. 1993 Sep 1;80(3):517–26.

58. Bachman, J. G., Johnston, L. D., O'Malley, P. M. Monitoring the Future questionnaire responses from the nation's high school seniors. Institue for Social Research, The University of Michigan; 2010.

59. Bachman, J. G., Johnston, L. D., O'Malley, P. M., Schulenberg, J. E., Meich, R. A. The Monitoring the Future project after four decades: Design and procedures (Monitoring the Future Occasional Paper No. 82). Institue for Social Research, The University of Michigan; 2015.

60. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986 Apr 1;73(1):13–22.

61. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics. 1988 Dec;44(4):1049–60.

62. Katz J, Zeger SL, West KP, Tielsch JM, Sommer A. Clustering of Xerophthalmia within Households and Villages. Int J Epidemiol. 1993 Aug 1;22(4):709–15.

63. SAS Institute Inc. The GENMOD Procedure. In: SAS/STAT User's Guide. Cary, NC: SAS Institute Inc.; 1999. p. 1363–464.

64. Liang KY, Zeger SL. Regression Analysis for Correlated Data. Annu Rev Public Health. 1993;14(1):43–68.

65. Porta MS, Greenland S, Hernán M, Silva I dos S, Last JM. A Dictionary of Epidemiology. Oxford University Press; 2014. 377.

66. Aldworth, J, Chromy, JR, Davis, TR, Foster, MS, Packer, LE, Spagnola, K, et al. 2012 National Survey on Drug Use and Health: Statistical Inference Report. Substance Abuse and Mental Health Services Administration. RTI International; 2013.

67. Hedges LV, Vevea JL. Fixed-and random-effects models in meta-analysis. Psychol Methods. 1998;3(4):486.

68. Stata Corp. Stata Statistical Software: Release 13. College Station, TX: Stata Corp LP; 2013.

69. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep 6;327(7414):557–60.

70. Hosmer DW, Lemeshow S, Sturdivant RX. Wiley Series in Probability and Statistics : Applied Logistic Regression (3rd Edition) [Internet]. New York, NY, USA: John Wiley & Sons; 2013 [cited 2014 Aug 27]. Available from: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10677827

71. Hardin, J.W., Hilbe, J.M. Generalized Estimating Equations [Internet]. Second. 2012 [cited 2015 Aug 19]. 277 p. Available from: https://www.crcpress.com/Generalized-Estimating-Equations-Second-Edition/Hardin-Hilbe/9781439881132

72. Grevenstein D, Nagy E, Kroeninger-Jungaberle H. Development of risk perception and substance use of tobacco, alcohol and cannabis among adolescents and emerging adults: evidence of directional influences. Subst Use Misuse. 2015 Feb;50(3):376–86.

73. United States. Census Bureau. Elementary and Secondary Education: Completions and Dropouts - The 2012 Statistical Abstract. [Internet]. 2012 [cited 2015 Aug 18]. Available from: http://www.census.gov/compendia/statab/cats/education/elementary_and_secondary_education_completions_and_dropouts.html

74. Wagner FA, Anthony JC. From First Drug Use to Drug Dependence: Developmental Periods of Risk for Dependence upon Marijuana, Cocaine, and Alcohol. Publ Online 29 August 2001 Doi101016S0893-133X0100367-0. 2002 Apr 1;26(4):479–88.